Surface Temperatures of the Arctic Oceans

Nielsen-Englyst, Pia

Publication date:
2023

Document Version
Publisher's PDF, also known as Version of record

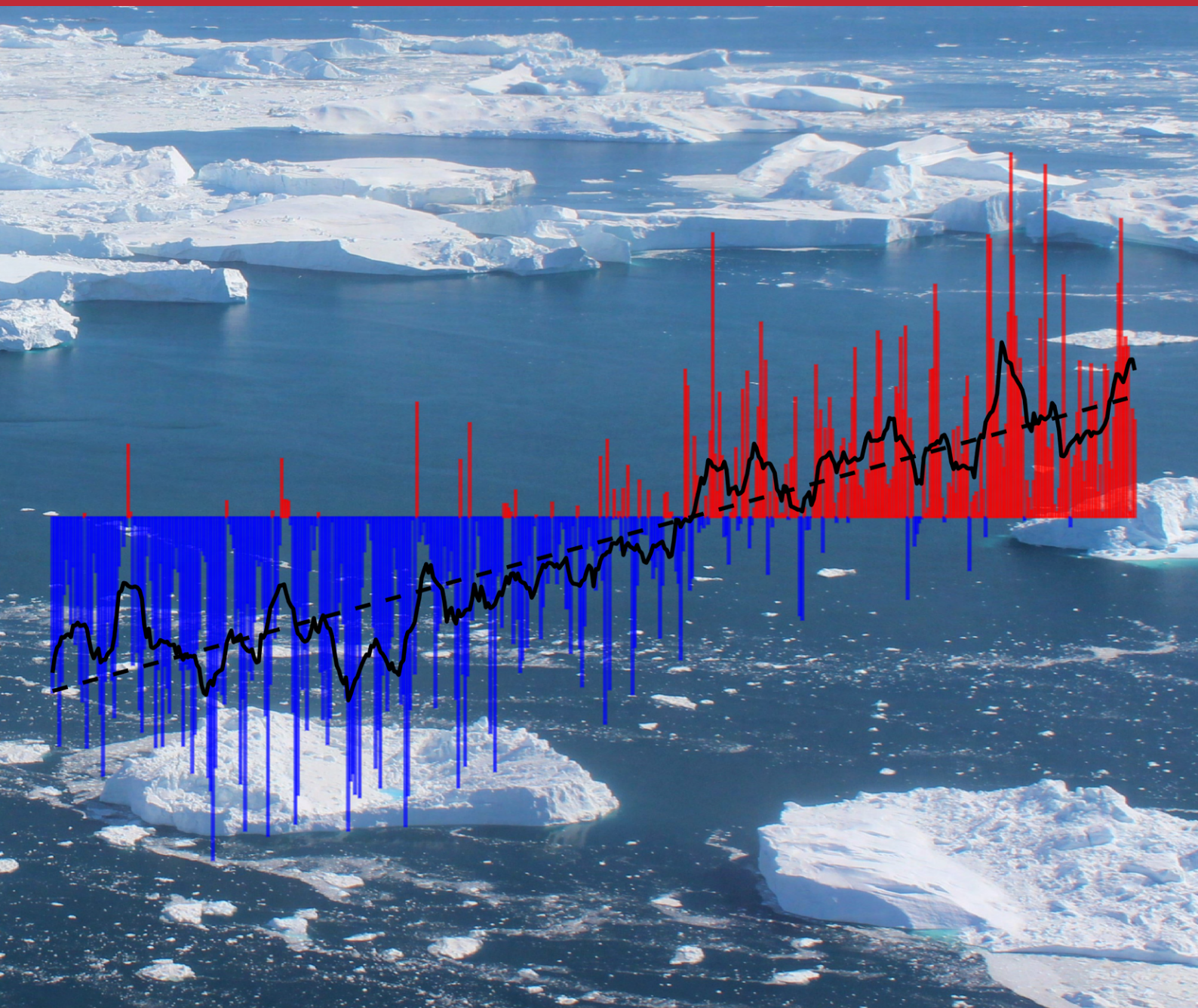
DTU Space
National Space Institute

Surface Temperatures of the Arctic Oceans

PhD Thesis

Pia Nielsen-Englyst

July, 2023



Surface Temperatures of the Arctic Oceans

PhD Thesis
July, 2023

By
Pia Nielsen-Englyst

Cover Picture (graph): Monthly (bars) and yearly (solid curve) mean anomalies of the combined sea and sea-ice surface temperature (1982-2021) of the Arctic (>58°N) using data from Nielsen-Englyst et al., 2023b

Cover Picture (photo): Sea ice along the coast of northwestern Greenland
Photo Credit: Pia Nielsen-Englyst

Published by: DTU Space, Microwaves & Remote Sensing, Ørsteds Plads, Building 348, 2800 Kgs. Lyngby Denmark
www.space.dtu.dk

Preface

This PhD thesis is submitted in partial fulfillment for obtaining at the PhD Degree at the Technical University of Denmark (DTU). The work described in the PhD thesis has been conducted at the National Space Institute, DTU Space, at the Microwaves and Remote Sensing division and at the National Centre for Climate Research (NCKF) at the Danish Meteorological Institute (DMI). The PhD study was supervised by Rasmus T. Tonboe and Leif T. Pedersen from DTU Space, and Jacob L. Høyer from DMI.

The PhD project was funded by DTU-Space and NCKF at DMI. Part of the project has also received funding through the Copernicus Marine Environment Monitoring Service (CMEMS), the European Space Agency Climate Change Initiative for Sea Surface Temperature (ESA-CCI SST) grant no. 4000109848/13/I-NB and the EU Horizon 2020 (EUSTACE) grant no. 640171.

Pia Nielsen-Englyst
Copenhagen

Acknowledgements

I would like to express my sincere gratitude to my supervisors at the Technical University of Denmark (DTU), Leif T. Pedersen and Rasmus T. Tonboe for their guidance, support and encouragement in relation to my PhD study. A special thanks to my supervisor at Danish Meteorological Institute (DMI), Jacob Høyer, for his valuable supervision, inspiration, motivation and endless support throughout my PhD studies.

I would also like to thank my colleagues at DTU and DMI as well as the co-authors for the scientific discussions and contributions. Also thank you for creating a very pleasant and joyful work environment and for all the support and kindness.

I would also like to express my gratitude to Sandra Castro for providing me the opportunity to spend time abroad at University of Colorado Boulder. Thank you to both Sandra Castro and Gary Wick for valuable discussions and for opening their home to my husband Mathias and me, and providing us with a memory for life.

Last but not least, I would also like to thank my family and friends for their unwavering support, patience and encouragement. In particular, I would like to thank Mathias for his love, support and for taking care of all the practical stuff at home allowing me to focus on the PhD study and being a mom. And to my two kids, Lucas and Elias, for their sincere love, laughter and smiles that gave me extra strength to keep pushing through tough days, especially in the final month.

I am truly blessed to have so many wonderful people in my life.

Abstract

The Arctic is warming faster than any other region of the world (known as Arctic amplification) leading to rapid and widespread changes, which transform the Arctic environment with far-reaching consequences. Despite much attention, existing observational datasets, reanalyses and climate models show large uncertainties in Arctic surface temperatures and limited consensus on the magnitude of the Arctic amplification. The difference and uncertainties mainly arise in the Arctic oceans where many clouds, the mix of open water and sea ice, and the sparse *in situ* network challenge an accurate and absolute surface temperature estimation.

Each of these challenges are considered in this PhD study with the overall aim to provide more accurate and consistent sea surface temperature (SST) and sea ice surface temperature (IST) estimates in the Arctic, and thereby improving the understanding, characterization and monitoring of the Arctic warming and amplification.

The frequent and persistent cloud cover in the Arctic limits the extent to which SST can be retrieved from thermal infrared satellite sensors. Therefore, this PhD study explores the capability of using passive microwave (PMW) observations to retrieve SST and improve the SST estimates in the Arctic. Multiple PMW SST retrieval algorithms have been developed, analysed and validated and the first European PMW SST climate data record has been generated. To prepare for the future Copernicus Imaging Microwave Radiometer (CIMR) satellite mission, this study also investigates the impact of using different frequency channels in SST retrievals, with promising results for the proposed CIMR constellation. The impact of including the PMW SST observations in the Arctic surface temperature estimation has been evaluated and substantial improvements are seen. The results are expected to become even better in the future with the launch of CIMR, which will enable SST retrievals at lower uncertainties and much closer to the coasts and sea ice.

Due to the mix of open ocean and sea ice (and the temporal varying sea ice coverage) the most consistent way to monitor the Arctic surface temperature change is by integrating SST and IST estimates. This PhD study presents the first gap-free infrared satellite-based climate data record (1982-2021) of combined sea and sea-ice surface temperatures in the Arctic (>58°N), which can be used as a consistent indicator for climate monitoring. It shows that the combined sea and sea-ice surface temperature has increased by ∼4.5°C from 1982 to 2021, with a peak warming of ∼10°C in the northeastern Barents Sea.

To supplement the sparse *in situ* network, the satellite-observed ISTs have been used to estimate near-surface air temperature (T2m) over sea ice. The satellite-derived T2m estimates provide much better spatial coverage than the *in situ* observations and show improved performance compared to ECMWF's most recent reanalysis (ERA5). The satellite-derived IST and T2m estimates provide an important supplement to the existing *in situ* observations and have a large potential to be used for assimilation, evaluating and improving global surface temperature reconstructions, atmospheric reanalyses and climate models in the Arctic. Initial efforts show that the satellite-derived surface temperatures can improve our physical understanding and guide future developments in global climate models and atmospheric reanalyses in the Arctic.

Resumé

I Arktis stiger temperaturen meget hurtigere end i resten af verden (kendt som arktisk forstærkning), hvilket fører til hurtige og omfattende forandringer, der transformerer det arktiske miljø med vidtrækkende konsekvenser. På trods af megen opmærksomhed viser eksisterende observationelle datasæt, reanalyser og klimamodeller store usikkerheder i arktiske overfladetemperaturer og stor uenighed om størrelsen af den arktiske forstærkning. Forskellene og usikkerhederne opstår hovedsageligt i de arktiske havområder, hvor mange skyer, blandingen af åbent vand og havis, og det sparsomme *in situ* netværk udfordrer bestemmelsen af absolutte og nøjagtige overfladetemperaturer.

I denne ph.d.-undersøgelse tages hver af disse udfordringer i betragtning med det formål at udlede mere nøjagtige og konsistente hav- og havis-overfladetemperaturer i Arktis og dermed forbedre forståelsen, karakteriseringen og overvågningen af den arktiske opvarmning og forstærkning.

Det hyppige og vedvarende skydække i Arktis begrænser i vid udstrækning, bestemmelsen af havtemperaturer fra termisk infrarøde satellit-sensorer. I dette studie undersøges muligheden for at bruge passive mikrobølge (PMW) observationer til at bestemme havoverfladetemperatur (SST) og forbedre eksisterende SST-estimater i Arktis. Forskellige PMW SST algoritmer er blevet udviklet, analyseret og valideret, og det første europæiske PMW-SST-klimadatasæt er blevet produceret. For at forberede os på den fremtidige Copernicus Imaging Microwave Radiometer (CIMR) satellit-mission undersøges effekten af at anvende forskellige frekvenskanaler i SST algoritmerne med lovende resultater for den forudsatte CIMR konstellation. Effekten af at inkludere PMW SST-observationer til bestemmelsen af den Artiske overfladetemperatur er blevet undersøgt, og der ses betydelige forbedringer. Resultaterne forventes at blive endnu bedre i fremtiden med opsendelsen af CIMR, som vil gøre det muligt at estimere havtemperaturer med meget lavere usikkerheder og meget tættere på kysterne og havisen.

På grund af kombinationen af åbent hav og havis (og den tidslige variation i havisdække) er den mest konsistente måde, at overvåge temperatur-ændringen i de arktiske havområder at integrere hav- og havis-overfladetemperaturer. Denne ph.d.-undersøgelse præsenterer det første infrarøde satellit-baserede klimadatasæt (uden huller ved skydække) af kombineret hav- og havis-overfladetemperaturer i Arktis (>58°N), der kan bruges som en konsistent indikator i klimaovervågningen. I følge det nye datasæt, er den kombinerede overfladetemperatur steget med ∼4.5°C fra 1982 til 2021 med en maksimal opvarmning på ∼10°C i det nordøstlige Barentshav.

For at supplere det sparsomme *in situ* netværk undersøges potentialet i at bruge de satellit-observerede havis-temperaturer til at estimere lufttemperatur (T2m) over havis. De satellit-baserede T2m giver meget bedre geografisk dækning end *in situ* observationerne og udviser større nøjagtighed sammenlignet med ECMWF's nyeste reanalyse (ERA5). De satellit-baserede overfladetemperaturer er et vigtigt supplement til de sparsomme *in situ* observationer og har desuden et stort potentiale til at blive brugt til assimilering, evaluering samt forbedring af eksisterende overfladetemperatur estimater. De foreløbige resultater viser, at de satellit-baserede overfladetemperaturer kan forbedre vores fysiske forståelse og guide fremtidige udviklinger i globale klimamodeller og reanalyser i Arktis.

List of Acronyms

Abbreviations	Short for
AASTI	Arctic and Antarctic ice and sea Surface Temperatures from thermal Infrared
AMAP	Arctic Monitoring and Assessment Programme
AMOC	Atlantic Meridional Overturning Circulation
AMSR-E	Advanced Microwave Scanning Radiometer - EOS
AMSR2	Advanced Microwave Scanning Radiometer 2
ASIP	Automated Sea Ice Products
BAOE	Bias Aware Optimal Estimation
BEST	Berkeley Earth Surface Temperatures
BT	Brightness Temperature
CALIOP	Cloud-Aerosols Lidar and the Infrared Pathfinder
CARRA	Copernicus Arctic Regional Reanalysis
CCI	Climate Change Initiative
CCMP	Cross-Calibrated Multi-Platform
CDR	Climate Data Record
CIMR	Copernicus Imaging Microwave Radiometer
CMEMS	Copernicus Marine Environment Monitoring Service
CMIP	Coupled Model Intercomparison Project
CRREL	Cold Regions Research and Engineering Laboratory
DMI	Danish Meteorological Institute
DTU	Technical University of Denmark
ECMWF	European Centre for Medium-Range Weather Forecasts
ECV	Essential Climate Variable
EOS	Earth Observing System
ERA-I	ERA-Interim
ESA	European Space Agency
EUMETSAT	European Organisation for the Exploitation of Meteorological Satellites
EUSTACE	EU Surface Temperature for All Corners of Earth
GCOM-W	Global Change Observation Mission – Water
GCOS	Global Climate Observing System
GHRSST	Group for High Resolution Sea Surface Temperature
GISS	Goddard Institute for Space Studies
GISTEMP	GISS Surface Temperature Analysis
GrIS	Greenland Ice Sheet
IPCC	Intergovernmental Panel on Climate Change
IR	Infrared
IST	Sea Ice Surface Temperature
ITU	International Telecommunication Union
JAXA	Japan Aerospace Exploration Agency
LSBoost	Least-Squares Boosting
LSTM	Land Surface Temperature Monitoring
MODIS	Moderate Resolution Imaging Spectroradiometer
ML	Machine Learning
MMD	Multi-sensor Matchup Dataset

MOHC	Met Office Hadley Centre
NASA	National Aeronautics and Space Administration
NN	Neural Network
NP	North Pole
NOAA	National Oceanic and Atmospheric Administration
NWP	Numerical Weather Prediction
OE	Optimal Estimation
OI	Optimal Interpolation
OSISAF	Ocean and Sea Ice Satellite Application Facility
OSTIA	Operational Sea Surface Temperature and Ice Analysis
PMW	Passive Microwave
RE	Regression
REMSS/RSS	Remote Sensing Systems
RMSE	Root Mean Squared Difference
RFI	Radio Frequency Interference
RTM	Radiative Transfer Model
SI	International System of Units
SIC	Sea Ice Concentration
SLSTR	Sea and Land Surface Temperature Radiometer
SMHI	Swedish Meteorological and Hydrological Institute
SSS	Sea Surface Salinity
SST	Sea Surface Temperature
T2m	Near-surface air temperatures
TCLW	Total Column Liquid Water
TCWV	Total Column Water Vapour
TRISHNA	Thermal InfraRed Imaging Satellite for High-resolution Natural resource Assessment
VIIRS	Visible Infrared Imaging Radiometer Suite
WS	Wind Speed
XGBoost	Extreme Gradient Boosting

Contents

CHAPTER 1

Introduction

The Arctic is warming more rapidly than any other region of the world, a phenomenon known as Arctic amplification (AMAP, 2021; Overland et al., 2019; Pithan & Mauritsen, 2014), which arise due to a number of amplifying feedback mechanisms (Goosse et al., 2018). The fast rising Arctic temperatures lead to rapid and widespread changes in snow cover, sea ice, land ice (glaciers and ice sheets), permafrost, precipitation and extreme events (Box et al., 2019; IPCC, 2021; Stroeve & Notz, 2018). These changes are transforming the Arctic environment with far-reaching consequences e.g. global sea level rise and opening of new shipping routes including the opportunities and risks associated with this (AMAP, 2021). Most projections by climate models indicate that future warming will also be amplified in the Arctic (Holland & Landrum, 2021) and some scenarios indicate that the Arctic Ocean may be entirely ice free in summer by ∼2050 (Notz & Community, 2020). This makes the Arctic an important indicator of climate change.

Both sea surface temperature (SST) and sea ice surface temperature (IST) are categorized as essential climate variables (ECVs) by the Global Climate Observing System (GCOS), and observing these with high accuracy is crucial for monitoring, understanding, characterizing and predicting the Arctic climate system and its change (GCOS, 2022). Accurate Arctic surface temperatures are critical in various elements of the climate system determining the exchange of heat between the surface and atmosphere, ice growth and melt processes, and effecting the ocean circulation, ocean carbon dioxide uptake, as well as the ecosystems (AMAP, 2021; Behrenfeld et al., 2006; Bentamy et al., 2017; Key et al., 1997; Maykut, 1986; O'Carroll et al., 2019). Passing specific temperature thresholds may also result in the passing of "tipping points" (leading to large and often irreversible changes) e.g. in the Arctic sea ice, Greenland Ice Sheet (GrIS), Atlantic meridional overturning circulation (AMOC), permafrost and ecosystems (Lenton et al., 2008; Wang et al., 2023). Finally, Arctic surface temperatures play a crucial role in weather and sea ice forecasts through assimilation into ocean, sea ice and atmospheric models (Carton & Giese, 2008; Oke et al., 2008; Rasmussen et al., 2018; Rayner et al., 2003).

Despite much attention, existing observational datasets, atmospheric reanalyses and global climate models show large differences and uncertainties in the Arctic surface temperatures (e.g. Davy & Outten, 2020; IPCC, 2021; Lenssen et al., 2019; Wang et al., 2019) and limited consensus on the magnitude of Arctic amplification (Rantanen et al., 2022). The differences and uncertainties mainly arise in the Arctic oceans which consist of open ocean, sea ice and a marginal ice zone (defined as the transitional zone between the open sea and pack ice). In this region, persistent cloud cover, the mix of open water and sea ice (and the

temporal varying sea ice coverage), and the sparse *in situ* network challenge accurate and absolute estimation and monitoring of the Arctic sea and sea-ice surface temperatures.

The surface temperature has routinely been measured *in situ* from meteorological stations and other platforms over sea ice, which usually provide near-surface air temperatures at 1–2 m height (e.g. Nielsen-Englyst et al., 2019; Richter-Menge et al., 2006; referred to as T2m in the following), and over the ocean using ships, drifting/moored buoys, Argo floats and other platforms which provide SST at varying depths (Atkinson et al., 2014; Good et al., 2013; Kennedy, 2014; Rayner et al., 2006; Roemmich et al., 2009; Woodruff et al., 2011). However, the extreme environment and the poor accessibility make *in situ* observations challenging and sparse in the Arctic (Centurioni et al., 2019; Donlon et al., 2012). The best way to get a spatially broad coverage of the Arctic is through satellite remote sensing. Thermal infrared (IR) sensors are very useful for ocean and ice (or snow) covered surfaces and are heavily used to provide spatially detailed maps of SST and IST.

Since the early 1980s, IR satellite sensors have routinely been used to observe SST with a typical spatial resolution of 1–4 km and uncertainties of about 0.2–0.4°C (e.g. Donlon et al., 2007; Embury et al., 2012; Gladkova et al., 2016; Merchant et al., 2019; Reynolds et al., 2002). The IR SST products are derived using satellite observations in the atmospheric windows at wavelengths of 3.5–4.1 μm (mid-IR) and/or 9.5–13 μm (thermal IR). In both IR bands, the radiance is partially attenuated by the atmosphere due to water vapor, CO_2, CH_4, NO_2, and aerosols, with the main contribution from water vapor, which is very variable in both time and space (Saunders & Edwards, 1989). This requires an atmospheric correction and several methods have been proposed (e.g. Barton, 1995; Emery et al., 1994; Kilpatrick et al., 2001; Kilpatrick et al., 2015; McMillin, 1975; Minnett, 1990; Walton, 1988; Wick et al., 1992). Insufficient representation of the atmospheric attenuation is one of the main sources of uncertainty in IR SST retrievals in the Arctic (Castro et al., 2008; Donlon et al., 2007; O'Carroll et al., 2019). Several other factors also complicate IR SST retrievals in the Arctic e.g. the limited opportunities for matchups with *in situ* observations (in particular during summer when sun glint contaminates the mid-IR wavelengths) to determine the coefficients for the atmospheric correction algorithms and for assessing the accuracy of the retrieved SSTs (Centurioni et al., 2019; Donlon et al., 2012; Minnett et al., 2019; O'Carroll et al., 2019). Moreover, cloud cover introduces intolerable errors in the IR SST retrievals, and cloud-contaminated pixels should be identified and excluded prior to the retrieval of SST from IR measurements. Therefore, the accuracy of IR SSTs is very dependent on the availability of a good cloud screening algorithm (e.g. Kilpatrick et al., 2001; Liu & Minnett, 2016; Merchant et al., 2005; Vázquez-Cuervo et al., 2004), in particular in persistent cloudy regions such as the Arctic (Curry et al., 1996; Eastman & Warren, 2010; Jia & Minnett, 2020). However, accurate cloud screening is difficult in the Arctic where persistent cloud cover, long twilight, and in some cases similar temperature of the cloud tops and the ocean challenge cloud screening algorithms in accurate identification of clouds (Curry et al., 1996; Shupe et al., 2011).

The long periods without surface coverage from IR observations make passive microwave (PMW) observations an important supplement (despite a poorer spatial resolution) since these are not prevented by non-precipitating clouds (Donlon et al., 2009; Ulaby et al., 1986; Wentz et al., 2000). The first global accurate PMW SST retrievals became available with the launch of the Advanced Microwave Scanning Radiometer – Earth Observing System (AMSR-E) on the NASA Aqua satellite (Chelton & Wentz, 2005; Kawanishi et al., 2003) in 2002, which was followed by the currently operational AMSR2 on the Japan Aerospace Exploration Agency (JAXA) GCOM-W1 satellite, launched in May 2012. At frequencies between 4 and 11 GHz, the vertically polarized brightness temperature of the ocean has an

appreciable sensitivity to SST. In addition to SST, the brightness temperatures depend on the sea surface roughness (usually parameterized in terms of the near-surface wind speed and direction) and the atmospheric temperature and moisture profiles (Ulaby et al., 1986; Wentz & Meissner, 2000), but these effects can usually be accounted for in the SST retrieval e.g. given simultaneous measurements at multiple frequencies and polarizations (e.g. Meissner & Wentz, 2012; Nielsen-Englyst et al., 2018; Wentz & Meissner, 2000; Wentz & Meissner, 2007).

In general, the retrieval algorithms can be divided into two categories: statistical algorithms and physically based algorithms. The statistical algorithms are the most common and these are developed through comparisons of satellite measured brightness temperatures and collocated *in situ* observations and model data (e.g. Chang et al., 2015; Shibata, 2006). The physically based algorithms use radiative transfer models (RTMs) to simulate the top of atmosphere brightness temperatures using instrumental (azimuth/earth incidence angles, frequency and polarisation) and environmental (SST, sea surface salinity, wind speed/direction, water vapour density, liquid water density, pressure, and atmospheric profiles of temperature) information. The simulated brightness temperatures can then be used to determine the coefficients in regression-based algorithms (Meissner & Wentz, 2012) or alternatively the RTM can be inverted to retrieve SST (and other parameters) from the satellite observed brightness temperatures e.g. using optimal estimation (OE; Nielsen-Englyst et al., 2018).

The PMW SST retrievals provide almost complete daily global coverage only prevented in regions with rain, strong winds, sun-glitter, Radio Frequency Interference (RFI), and/or proximity (typically within ∼100 km) to land and sea ice, where the impact of side-lobe contamination becomes a significant error (Gentemann, 2014; Gentemann & Hilburn, 2015). PMW SSTs typically have uncertainties of 0.4–0.5°C and spatial resolutions of 50–60 km (Gentemann, 2014; Gentemann et al., 2010; Nielsen-Englyst et al., 2018; Shibata, 2006; Wentz et al., 2000). The existing PMW SST algorithms have typically been derived for individual sensors, limited time periods or using sparse *in situ* data for tuning and validation. However, the complementary characteristics of PMW SSTs and the sustained measurements over a time frame that is now of sufficient length for climate studies, make them a potentially important supplement to existing satellite IR SSTs for development of climate time series, especially in cloudy regions such as the Arctic (O'Carroll et al., 2019). The potential is even greater since the series of PMW imagers has now been sustained for the future with AMSR3 planned by JAXA (Kasahara et al., 2020) as well as the Copernicus Imaging Microwave Radiometer (CIMR) prepared by European Space Agency (ESA). CIMR is a polar mission designed to provide high-accuracy, high resolution PMW observations to enable retrievals of SST and other surface parameters with lower uncertainty and at a higher spatial resolution than what is possible with the existing PMW missions (Donlon, 2020).

The radiometer channel configuration of CIMR is not identical but has channels in common with the AMSR missions and for that reason, it is important to assess the expected impact of using the CIMR channel configuration compared to previous missions. Theoretical information content studies have been conducted to assess the expected CIMR performance for SST and the impacts from including different channels (Kilic et al., 2018; Pearson et al., 2019; Prigent et al., 2013), but the ocean forward models typically used in these studies show large differences and significant disagreement with observations in particular for high wind speeds and cold waters (Kilic et al., 2019a). To supplement the existing theoretical studies, it is important to evaluate PMW SST retrieval algorithms against a common *in situ* reference dataset when assessing the impact of different channel selections.

The large potential of using high latitude PMW observations, with frequent updates from previous, current and future PMW missions, makes it important to investigate how IR and PMW SST retrievals should be combined in the Arctic. The SST retrievals are usually averaged, interpolated and combined into different data levels, with level 4 (L4) being gridded and gap-free fields. A large number of global L4 SST products are available using a variety of different satellite observations (both IR and PMW) and some including *in situ* observations (e.g. Donlon et al., 2012; Good et al., 2020; Merchant et al., 2019; Reynolds et al., 2007; Roberts-Jones et al., 2012). The existing L4 SST analyses perform fairly uniformly globally, but significant differences and large uncertainties are found in the Arctic (Castro et al., 2016; Dash et al., 2012; Vazquez-Cuervo et al., 2022), where the extreme environmental conditions, sparse *in situ* network, persistent cloud cover and the varying length of the sunlit part of the day round the year, complicate accurate SST retrievals (Donlon et al., 2009; Høyer et al., 2012; Minnett et al., 2019; Wang & Key, 2005). A number of specialized high latitude algorithms have been developed and shown improved performance in the Arctic compared to the global analyses (Jia & Minnett, 2020; Vincent et al., 2008a; Vincent et al., 2008b). More accurate SST observations and improved PMW data coverage in the Arctic have been identified by the SST community as being of high priority for future SST developments and research (O'Carroll et al., 2019).

Few global L4 SST products do not report SST in sea ice covered regions, while most products provide a foundation temperature, assumed to be at the freezing point of seawater. However, using foundation SSTs in sea ice regions does not represent the surface temperature since large temperature gradients can exist in the ice and snow (Tonboe et al., 2011). Different methods have been used to generate proxy SSTs from sea ice concentrations in the marginal ice zone (Banzon et al., 2020), which in some cases are blended with the closest open water SSTs (Rayner et al., 2003; Reynolds et al., 2002; Reynolds et al., 2007). Blending SST and IST is recognised as the most appropriate approach for determining surface temperatures in the marginal ice zone (Minnett et al., 2019).

IST can be derived from thermal emission at wavelengths in the IR or PMW domain (Lavergne et al., 2022). The IR sensors observe the radiation from the upper micrometers of the snow or sea ice surface (i.e. the skin surface), while the PMW sensors observe the radiation from the snow pack, snow-ice interface or even the solid sea ice depending on the frequency, salinity, temperature and snow/ice metamorphism (e.g. Tonboe, 2010; Tonboe et al., 2011; Ulaby et al., 1986). The penetration into the snow and sea-ice cover, and a varying sea ice emissivity complicate IST retrievals from PMW observations (Comiso et al., 2003). For that reason, ISTs from IR sensors are more widely used, although PMW sensors are considerably less sensitive to clouds. ISTs have been retrieved from IR satellite temperature observations since 1982 at daily (with gaps) and monthly temporal resolution (Dybkjær et al., 2012; Hall et al., 2004; Key & Haefliger, 1992; Key et al., 1997; Key et al., 2013; Liu et al., 2018; Maslanik et al., 1997). Cloud screening are more challenging over ice (compared to open ocean) due to the similar temperatures of the sea ice and cloud tops (Dybkjær et al., 2012; Wang & Key, 2005). As a result, IST retrievals usually have larger uncertainties than SST retrievals, and undetected clouds or atmospheric ice/water typically result in a cold bias in IST products (Dybkjær et al., 2012; Hall et al., 2004).

Traditionally, the sea and sea-ice surface temperature datasets have been derived and analysed individually. However, estimating climate trends for the Arctic open-ocean and sea ice separately is complicated due to the large seasonal and annual variability of the sea ice cover (Bulgin et al., 2020; Comiso, 2003). Using the L4 SST analyses, including foundation SST estimates below sea ice, to determine Arctic warming (e.g. as done in Chen et al., 2019; IPCC, 2021) should also be done with caution as the use of foundation

SSTs (usually fixed at the freezing point) in sea ice covered regions suppress the observed warming compared to the use of air temperatures, with differences up to 0.15°C in the global monthly mean anomalies (Rohde & Hausfather, 2020).

Some IR satellite-based Arctic surface temperature products combine SST and IST (Dybkjær et al., 2012; Vincent et al., 2008b), and a few products provide land surface temperatures as well (Comiso, 2003; Comiso & Hall, 2014; Dodd et al., 2019). These are all clear-sky surface temperature estimates i.e. with gaps due to clouds. A gap-free combined SST and IST product has been produced within the Copernicus Marine Environment Monitoring Service (CMEMS) providing near-real-time surface temperature estimates of the Arctic open ocean, sea ice and marginal ice zone (doi: http://dx.doi.org/10.48670/moi-00130). However, no satellite-based gap-free combined SST and IST reanalysis or climate data record (CDR) are currently available, but it could potentially provide a consistent climate indicator, which is important for studying climate trends in the Arctic.

Traditionally, the Arctic warming has instead been assessed using global gridded near-surface temperature products (such as HadCRUT, BEST, NOAAGlobalTemp and GISTEMP), which are currently based only on *in situ* observed near-surface air temperatures over sea ice (and land), and SST observations (usually both satellite and *in situ*) over open ocean (e.g. Lenssen et al., 2019; Morice et al., 2021; Rohde and Hausfather, 2020; Vose et al., 2012). They are among the longest instrumental records of temperature and have routinely been used in climate assessments (e.g. AMAP, 2021; IPCC, 2021; Rantanen et al., 2022). However, their utility is limited in applications that require high temporal and/or spatial resolution and absolute temperatures, since the datasets are presented at monthly, coarse spatial resolution, and often expressed in terms of temperature anomalies (temperatures relative to some reference period), rather than in terms of absolute temperature.

The use of SST instead of air temperatures over the ocean is mainly because SST observations from satellites and *in situ* are far more numerous than marine air temperature measurements (Kennedy, 2014), which have been in recent decline (Berry & Kent, 2017). However, the combination of near-surface air temperatures over sea ice and SST over open ocean also complicates the surface temperature trend calculations since the changing sea ice cover leads to shifts between the two variables, and artefacts of the varying techniques may thereby be introduced into the data record, which is clearly undesirable in the context of climate monitoring (IPCC, 2021; Richardson et al., 2018).

In addition to the global gridded observational datasets, dynamical reanalyses, which combine numerical weather prediction (NWP) models with a variety of observational data (e.g. Dee et al., 2011; Gelaro et al., 2017; Hersbach et al., 2020; Kobayashi et al., 2015) are increasingly being used for climate monitoring and to force ocean and sea ice models. These alternative sources of near-surface temperatures can provide information in regions that are not well represented in traditional observational datasets. Examples of these include the ERA5 (Hersbach et al., 2020) (and its predecessor ERA-Interim, ERA-I, Dee et al., 2011), JRA-55 (Kobayashi et al., 2015), and MERRA-2 (Gelaro et al., 2017). However, the sparse *in situ* network causes large deviations and uncertainties in the Arctic in existing global reanalyses (Davy & Outten, 2020; IPCC, 2021; Lindsay et al., 2014; Marquardt Collow et al., 2020; Simmons et al., 2017; Wang et al., 2019; Wesslén et al., 2014) as well as in the global gridded observational datasets (Cowtan & Way, 2014; Lenssen et al., 2019; Rapaić et al., 2015). As recognized by Marquardt Collow et al. (2020), more *in situ* and remote sensing observations as well as a better use of existing satellite observations are needed in order to represent the characteristics of the Arctic. Likewise, the Arctic

Monitoring and Assessment Programme (AMAP) emphasizes the need for Arctic and international science institutions and governments to address key data gaps e.g. by using satellite observations to gather data from difficult-to-reach areas of the Arctic in order to expand the monitoring and documentation of Arctic climate change (AMAP, 2021).

1.1 Motivation and research objectives

The aim of this PhD is to improve the surface temperature estimation and monitoring of the Arctic oceans. The PhD study focuses on the challenges related to the persistent cloud cover, the mix of open water and sea ice (and the temporal varying sea ice coverage) and the sparse *in situ* network, which all complicate accurate and absolute estimation and monitoring of the sea and sea-ice surface temperatures in the Arctic. To guide the work of this PhD study, five main research objectives have been established.

To deal with the persistent Arctic cloud cover, this PhD study investigates the potential of using PMW sensors to fill in the large and persistent data gaps (due to clouds) in Arctic IR SSTs. This is investigated through the following research objectives:

1. Develop/improve and compare algorithms to retrieve SST from PMW satellite observations.
2. Assess the impact of using different channel selections in the PMW SST retrieval algorithms using *in situ* observations as reference.
3. Examine how to best combine IR and PMW SSTs, and assess the impact of including PMW SST observations in an Arctic gap-free SST analysis.

To address the challenges related to the varying sea ice cover, this PhD study investigates how sea and sea-ice parameters can be integrated in a gap-free analysis to increase the consistency in the surface temperature monitoring of the Arctic oceans.

4. Generate a gap-free combined sea and sea-ice surface temperature CDR of the Arctic

To supplement the sparse *in situ* network, this PhD study examines the potential of using satellite-observed ISTs to improve the near-surface air temperature estimation of the Arctic (through an improved spatial coverage).

5. Use the satellite-observed ISTs to derive near-surface air temperatures over ice surfaces in the Arctic

1.2 Thesis outline

This PhD thesis builds upon a number of appended publications (Section 1.3) and the reader is referred to these for detailed information on each of the studies carried out during this PhD study.

To address the first research objective, Chapter 2 presents the work that has been done to develop and improve SST retrievals from PMW observations using an optimal estimation approach (Section 2.1), a regression model (Section 2.2, Paper I) and machine learning techniques (Section 2.3, Paper II). The second research objective is addressed in Section 2.4, which provides an assessment of the impact of using different radiometer channels in the retrievals (Paper III). The chapter ends with a comparison and discussion of the strengths and weaknesses of the PMW SST retrieval algorithms in Section 2.5.

To address the fourth research question, Chapter 3 investigates how SST observations can be integrated with sea ice parameters in the Arctic. For this, efforts have first been put into the generation of an accurate and consistent long-term sea ice concentration

product. This is described in Section 3.1 and used in the generation of an Arctic gap-free IR-based sea and sea-ice surface temperature CDR (Section 3.2, Paper IV). An assessment of including PMW SSTs in the Arctic gap-free CDR is provided in (Section 3.3, Paper V) to address the third research objective. To address the last research objective, Chapter 3 also presents a method which can be used to derive near-surface air temperatures from satellite-observed ice surface temperatures (Section 3.4.1, Paper VI) as well as an example of using the derived product for model evaluation (Section 3.5, Paper VII). The chapter concludes with a discussion of the results and their contributions to the monitoring and characterisation of the Arctic warming and amplification (Section 3.6).

Finally, in Chapter 4, the findings of this PhD study is summarized and the overall contribution to the surface temperature mapping and monitoring of the Arctic oceans is provided together with an outlook on future work.

1.3 Publications

This section provides an overview of the publications I have contributed to during my PhD study. For each publication, I have included a percentage of my estimated contribution to the given publication. Only publication I-VII are appended to this thesis.

Paper I Alerskans, Emy, Jacob L. Høyer, Chelle L. Gentemann, Leif Toudal Pedersen, **Pia Nielsen-Englyst**, Craig Donlon. 2020. "Construction of a climate data record of sea surface temperature from passive microwave measurements", *Remote Sensing of Environment*, Volume 236, 111485, ISSN 0034-4257.

My contribution: 10%

This work was done in parallel and extension of the work presented in Nielsen-Englyst et al. (2018) (see end of this section), where my focus was on the optimal estimation algorithm. My contribution to this paper was limited to simulating brightness temperatures used for the sensitivity experiments, discussing/interpreting the results and proofreading the paper.

Paper II Alerskans, Emy, Ann-Sofie P. Zinck, **Pia Nielsen-Englyst**, Jacob L. Høyer. 2022. "Exploring machine learning techniques to retrieve sea surface temperatures from passive microwave measurements", *Remote Sensing of Environment*, Volume 281, 113220, ISSN 0034-4257.

My contribution: 10%

I was not involved in the model development, and my contribution was limited to the sensitivity experiments, interpretation and general discussion of the results, as well as writing and proofreading the paper.

Paper III **Nielsen-Englyst, Pia**, Jacob L. Høyer, Emy Alerskans, Leif Toudal Pedersen, Craig Donlon. 2021. "Impact of channel selection on SST retrievals from passive microwave observations", *Remote Sensing of Environment*, Volume 254, 112252, ISSN 0034-4257.

My contribution: 85%

Most of this paper and the underlying experiments, analysis, interpretation and writing was done by me with supervision from co-authors. One exception is the regression algorithm retrievals, which were conducted by Emy.

Paper IV **Nielsen-Englyst, Pia**, Jacob L. Høyer, Wiebke M. Kolbe, Gorm Dybkjær, Thomas Lavergne, Rasmus Tage Tonboe, Sotirios Skarpalezos, Ioanna Karagali. 2023. "A

combined sea and sea-ice surface temperature climate dataset of the Arctic, 1982–2021", *Remote Sensing of Environment*, Volume 284, 113331, ISSN 0034-4257.

My contribution: 80%

For this paper, I did most of the work with supervision from co-authors, but part of the actual running of the CDR was conducted during my maternity leave by Jacob and Wiebke.

Paper V **Nielsen-Englyst, Pia**, Jacob L. Høyer, Ioanna Karagali, Wiebke M. Kolbe, Rasmus T. Tonboe, Leif T. Pedersen. 2023. "Impact of microwave observations on the estimation of Arctic sea surface temperature". Submitted. *In review – Remote Sensing of Environment*

My contribution: 90%

For this paper, I did almost everything myself with supervision from co-authors. One exception is the spectral analysis part, which was done by Ioanna.

Paper VI **Nielsen-Englyst, Pia**, Jacob L. Høyer, Kristine S. Madsen, Rasmus T. Tonboe, Gorm Dybkjær, Sotirios Skarpalezos. 2021. "Deriving Arctic 2 m air temperatures over snow and ice from satellite surface temperature measurements". *The Cryosphere*, 15(7), 3035-3057.

My contribution: 80%

For this paper, I compiled and quality checked *in situ* data, developed and tested the regression models, validated/analyzed the results and wrote the manuscript with supervision from co-authors. The uncertainty estimation was conducted by Kristine and Jacob.

Paper VII Tian Tian, Shuting Yang, Jacob L. Høyer, **Pia Nielsen-Englyst**, Suman Singha. 2023. "Concerns on benchmarking climate models in the Arctic." *Preprint submitted to Nature Geoscience - Correspondence*

My contribution: 15%

I have provided the satellite-derived near-surface air temperatures, performed the validation of those and the ERA-5, and contributed to the interpretation and discussion of the results as well as to the writing of the manuscript.

Paper VIII Ponsoni, Leandro, Mads H. Ribergaard, **Pia Nielsen-Englyst**, Tore Wulf, Jørgen Buus-Hinkler, Matilde B. Kreiner, Till A. S. Rasmussen. 2023. "Greenlandic sea ice products with a focus on an updated operational forecast system", *Frontiers in Marine Science*, 10, 2296-7745, doi: https://doi.org/10.3389/fmars.2023.979782

My contribution: 5%

I have generated and provided the sea ice concentration product (DMI-SIC), wrote the corresponding data section and contributed to the interpretation and discussion of the results (i.e. comparisons with model output and sea ice charts).

Paper IX Rayner, Nick A., Renate Auchmann, Janette Bessembinder, Stefan Brönnimann, Yuri Brugnara, Francesco Capponi, Laura Carrea, Emma M. A. Dodd, Darren Ghent, Elizabeth Good, Jacob L. Høyer, John J. Kennedy, Elizabeth C. Kent, Rachel E. Killick, Paul van der Linden, Finn Lindgren, Kristine S. Madsen, Christopher J. Merchant, Joel R. Mitchelson, Colin P. Morice, **Pia Nielsen-Englyst**, Patricio F. Ortiz, John J. Remedios, Gerard van der Schrier, Antonello A. Squintu, Ag

Stephens, Peter W. Thorne, Rasmus T. Tonboe, Tim Trent, Karen L. Veal, Alison M. Waterfall, Kate Winfield, Jonathan Winn, and R. Iestyn Woolway. 2020. "The EUSTACE Project: Delivering Global, Daily Information on Surface Air Temperature", *Bulletin of the American Meteorological Society*, 101, 11, E1924-E1947, doi: https://doi.org/10.1175/BAMS-D-19-0095.1

My contribution: <5%

For this study, the satellite-derived near surface air temperatures for the ice sheets and sea ice developed in Paper VI were used, but except from deriving those, I only contributed with proofreading the manuscript.

Paper X Karagali, Ioanna, Magnus B. Suhr, Ruth Mottram, **Pia Nielsen-Englyst**, Gorm Dybkjær, Darren Ghent, and Jacob L. Høyer. 2022. "A new Level 4 multi-sensor ice surface temperature product for the Greenland Ice Sheet." *The Cryosphere*, 16.9, 3703-3721. doi: https://doi.org/10.5194/tc-16-3703-2022

My contribution: 10%

I have contributed to the validation using PROMICE and IceBridge observations, and to the interpretation and discussion of the results as well as the writing and proofreading the paper.

It should be noted that this PhD thesis will reference two other first-author publications namely:

Nielsen-Englyst, Pia, Jacob L. Høyer, Leif Toudal Pedersen, Chelle L. Gentemann, Emy Alerskans, Tom Block, and Craig Donlon. 2018. "Optimal Estimation of Sea Surface Temperature from AMSR-E" *Remote Sensing*, 10, no. 2: 229, doi: https://doi.org/10.3390/rs10020229

Nielsen-Englyst, Pia, Jacob L. Høyer, Kristine S. Madsen, Rasmus T. Tonboe, Gorm Dybkjær, Emy Alerskans. 2019. "In situ observed relationships between snow and ice surface skin temperatures and 2 m air temperatures in the Arctic". *The Cryosphere*, 13(3), 1005-1024, doi: https://doi.org/10.5194/tc-13-1005-2019

The first one was written and published before my PhD start. The second one presents work that was initiated before my PhD start and continued during my PhD as external work (including the paper writing and review). For those reasons, they cannot officially count as PhD publications and the work should not be counted in the evaluation.

The same holds for the following publication, which was also written before my PhD start:

Rasmussen, Till A. S., Jacob L. Høyer, Darren Ghent, Claire E. Bulgin, Gorm Dybkjær, Mads H. Ribergaard, **Pia Nielsen-Englyst**, Kristine S. Madsen, 2018. "Impact of Assimilation of Sea-Ice Surface Temperatures on a Coupled Ocean and Sea-Ice Model". *Journal of Geophysical Research: Oceans*, 123(4), 2440-2460. doi: https://doi.org/10.1002/2017JC013481

Lastly, Appendix H contains an unpublished report which was submitted as a part of a 3-week PhD course at DTU, *02910 Computational Data Analysis*. The course was taken as part of the PhD degree, and the report should not be counted in the evaluation.

CHAPTER 2

Sea Surface Temperature from Microwaves

In the Arctic, frequent and persistent cloud cover limits the extent to which SST can be retrieved from IR satellite sensors. Therefore, there is a large potential in using the almost all-weather capability of PMW observations to retrieve SST in this region to supplement the clear-sky IR SSTs. The PMW sensors now offer a long and consistent time series, which has been sustained for the future with AMSR3 and CIMR in the pipeline. To prepare for AMSR3 and CIMR, and to fully benefit from the existing PMW satellite missions, there is a need to further develop and improve algorithms to retrieve SST from PMW satellite observations.

This chapter summarizes the PMW SST retrieval algorithm development conducted during this PhD study. Multiple PMW SST retrieval algorithms have been developed, tested and compared including an optimal estimation (OE) algorithm (Nielsen-Englyst et al., 2018) (Section 2.1), a regression (RE) algorithm (Alerskans et al., 2020) (Paper I, Section 2.2), and a number of machine learning (ML) models (Alerskans et al., 2022) (Paper II, Section 2.3). They are all global retrieval algorithms, but have been derived with a particular focus on the Arctic. To get a better understanding of the various types of retrieval algorithms and to prepare for future PMW satellite missions the impact of using different radiometer channels in the retrieval algorithms has also been investigated. This work is presented in Section 2.4 (Nielsen-Englyst et al., 2021a) (Paper III). Section 2.5 compares the different types of retrievals algorithms and discusses the strengths and weaknesses of each of them. Moreover, suggestions are provided for future work with a focus on improving the Arctic PMW SSTs and preparing for future PMW satellite missions.

The foundation for the retrieval work, is the the ESA Climate Change Initiative (CCI) Multi-sensor Matchup Datasets (MMDs), which have been used both for development and validation of the PMW SST retrieval algorithms. They consist of pairs of *in situ* SSTs (drifters and Argo floats) and satellite observations (AMSR-E and AMSR2) matched with wind speed (WS), total column water vapour (TCWV), total cloud liquid water (TCLW), SST, and sea ice concentration from the ERA-I and ERA5 reanalyses as well as Cross-Calibrated Multi-Platform (CCMP) gridded surface vector winds, and sea surface salinity from the GLOBAL-REANALYSIS-PHY-001-030 reanalysis product, provided by the CMEMS. All MMDs have been quality controlled and filtered to reduce effects from rain, diurnal warming, sun glitter, land, sea ice, and RFI. For more information on the MMDs and the quality checks, the reader is referred to Nielsen-Englyst et al. (2018) as

well as Alerskans et al. (2020), Alerskans et al. (2022) and Nielsen-Englyst et al. (2021a) (Paper I-III). The MMDs have been powerful tools in the development, assessment and validation of the different retrieval algorithms.

2.1 Optimal Estimation

This section provides a brief description of the optimal estimation (OE) algorithm presented in Nielsen-Englyst et al. (2018), which was developed before this PhD study started. It is included here to provide a complete review of the PMW SST retrieval algorithms used in this PhD study.

The OE algorithm was developed within the ESA-CCI SST project and used to retrieve subskin SST using AMSR-E birghtness temperatures from 2010. The algorithm is based on a forward model, which predicts the top-of-atmosphere PMW brightness temperatures that should be measured by the individual channels given knowledge of the physical state of the atmosphere and ocean. The forward model in Nielsen-Englyst et al. (2018) is based on the RTM described in Wentz and Meissner (2000), which consists of atmospheric absorption model for oxygen, water vapor and cloud liquid water and a sea surface emissivity model that determines the emissivity as a function of SST, sea surface salinity, sea surface wind speed and direction. In OE, the geophysical parameters are retrieved by inverting the forward model using the observed brightness temperatures, *a priori* information (from ERA-I in this case) about the state of the atmosphere and ocean, and the related uncertainties, to constrain the retrievals (Nielsen-Englyst et al., 2018; Rodgers, 2000). Nielsen-Englyst et al. (2018) shows that it is essential to update the first guess atmospheric and oceanic state variables and to perform several (typically three to four) iterations (Figure 4). The OE method directly provides an uncertainty estimate for each retrieval based on the uncertainties in the measurements, forward model and in the *a priori* state vector. In Nielsen-Englyst et al. (2018) an additional uncertainty indicator is setup based on a scaled root-mean-squared difference of the simulated versus observed brightness temperatures (RMSE TB). It turned out to be an efficient quality indicator of the OE SST retrievals (Figure 7, Nielsen-Englyst et al., 2018) and will be referred to as the theoretical uncertainty in the following.

A validation against drifting buoys shows a mean difference of 0.02°C and a standard deviation of 0.47°C when considering the 64% of the matchups, where the RMSE TB is less than 0.5 K (Table 3, Nielsen-Englyst et al., 2018). The corresponding mean theoretical uncertainty is estimated to 0.48°C, including the *in situ* and sampling (point to satellite footprint) uncertainties. The OE framework also directly provides the sensitivities of the retrieved parameters to the true state (through the averaging kernel). The mean SST sensitivity is found to be 0.50 with the smallest sensitivities in high latitudes and the largest in the equatorial region (Figure 10, Nielsen-Englyst et al., 2018), which is in agreement with Gentemann et al. (2010), and consistent with the fact that the sensitivity of the brightness temperatures to SST is smaller for cold waters, especially for the 10.65 GHz channels (Nielsen-Englyst et al., 2021a; Prigent et al., 2013) (Paper III, Figure 2).

In general, the results from Nielsen-Englyst et al. (2018) are very encouraging and demonstrate that the OE algorithm can provide reliable PMW SSTs as a valuable supplement to statistical retrieval algorithms. The forward model from Nielsen-Englyst et al. (2018) has been used to simulate the brightness temperatures, which are used in the sensitivity calculations performed in Alerskans et al. (2020) (Paper I, Section 2.2) and Alerskans et al. (2022) (Paper II, Section 2.3), and the OE algorithm used in Nielsen-Englyst et al. (2021a) (Paper III, Section 2.4) uses a setup very close to the on presented in Nielsen-Englyst et al. (2018).

2.2 Regression

This section briefly summarizes the work presented in Alerskans et al. (2020) (Paper I), where a statistical regression-based SST retrieval algorithm is developed and used to generate a consistent CDR of PMW SST using AMSR-E and AMSR2 observations. The level 2 (L2) algorithm consists of a wind speed (WS) retrieval algorithm and SST retrieval algorithm. The WS algorithm is inspired by the NOAA AMSR2 WS retrieval algorithm (Chang et al., 2015) and consists of two steps: a "global" regression algorithm for which an initial estimate of the WS is retrieved using one set of regression coefficients for all WS, and a "local" regression algorithm, trained to perform well in restricted WS bins, using the initial WS estimate for bin selection. As for WS, the SST algorithm consists of a global regression algorithm and local regression algorithm with the latter consisting of localized latitude algorithms for ascending and descending orbits. The localized algorithms are implemented to account for the non-linearities in the relationship between the brightness temperature and the SST and WS, respectively. The SST algorithm uses the brightness temperatures, Earth incidence angle and the relative angle between satellite azimuth angle and wind direction.

All regression coefficients are determined using drifting buoy observations and the least-squares method. Separate regression coefficients have been derived for AMSR-E and AMSR2. To obtain a more consistent retrieval between AMSR-E and AMSR2, the AMSR2 7.3 GHz channels are not used in the retrievals. All SST retrievals come with an uncertainty estimate, which is based on a global regression model. A new and effective RFI filter has also been developed by comparing the above presented retrieval with alternative retrieval algorithms excluding the 10 GHz and 18 GHz channels, respectively (Alerskans et al., 2020) (Paper I, Figure 4+5).

The sensitivity of the retrieved SST to true changes in SST has been estimated, using simulated brightness temperatures (by the forward model from Section 2.1) in the retrieval algorithm, and an average sensitivity of 0.90 is found. The validation against independent drifters (comparable for Argo) shows mean differences of –0.02°C and 0.002°C and standard deviations of 0.46°C and 0.45°C for AMSR-E and AMSR2, respectively, using the quality levels 4+5 (Alerskans et al., 2020) (Paper I, Table 6), with the largest standard deviations in high latitudes and in regions with large SST gradients (Alerskans et al., 2020) (Paper I, Figure 6+9). The validation results are comparable to or even better than found for existing PMW SST retrievals (Gentemann, 2014; Gentemann & Hilburn, 2015; O'Carroll et al., 2008). The corresponding mean modelled uncertainties are 0.45°C and 0.44°C including the *in situ* and sampling uncertainties. The analysis and extensive validation of SST, and the corresponding uncertainties presented in Alerskans et al. (2020) (Paper I) demonstrate a consistent and reliable multi-satellite PMW SST CDR that facilitates the uptake within the many applications of SSTs. Section 3.3 and Nielsen-Englyst et al. (2023a) (Paper V) provides an example of one application where the PMW SST is ingested into an IR-based gap-free sea and sea-ice surface temperature CDR of the Arctic.

2.3 Machine Learning

Machine learning (ML) models may supplement or improve the existing PMW SST retrieval algorithms through their higher flexibility and ability to recognize complex patterns in data. The first attempt to use ML techniques for PMW SST retrievals is presented in Alerskans et al. (2022) (Paper II). In parallel to this, I participated in a 3-week PhD Course, *Computational Data Analysis* at DTU-Compute in which I also worked with ML models for PMW SST retrievals using a smaller subset of the MMD (the validation subset from Paper II). This section briefly summarizes the work presented in Alerskans et al.

(2022) (Paper II) and the work conducted during the PhD course (Appendix H).

Alerskans et al. (2022) (Paper II) presents two types of ML models, which are used to retrieve PMW SST from AMSR-E brightness temperatures and other associated data available from the MMD (Alerskans et al., 2022) (Paper II, Figure 2). The first algorithm is an Extreme Gradient Boosting (XBG) model, which is a decision tree-based model, and the second algorithm is a multi-layer perceptron neural network (NN). For a detailed description of the ML models the reader is referred to Alerskans et al. (2022) (Paper II). The regression (RE) algorithm from Alerskans et al. (2020) (Paper I) is used as benchmark to evaluate the performance of the two ML algorithms. The validation against independent drifter observations shows mean differences of 0.01°C, 0.01°C and –0.02°C and standard deviations of 0.36°C, 0.50°C and 0.55°C for XGB, NN and RE, respectively (Table 4, Paper II). Simulated brightness temperatures (by the forward model from Nielsen-Englyst et al., 2018) have been used in each of the retrieval algorithms to estimate the SST sensitivities, which are found to be 0.78, 0.88 and 0.90 for XGB, NN and RE, respectively. Hence, the lower standard deviation of XGB is accompanied by a lower SST sensitivity, which may be a result of over-fitting in certain areas. This demonstrates the importance of including the sensitivity when evaluating different retrieval algorithms. The NN shows good performance in terms of both standard deviation and sensitivity.

During the PhD course, four ML models were investigated based on the validation subset of Alerskans et al. (2022) (Paper II), which was divided randomly into three (further) subsets used for training, testing and validating of the ML models (Appendix H). The ML models consist of a bootstrap aggregation (Bagging) model, random forest model, Least-squares boosting (LSBoost) model, and a neural network (see Appendix H for more details). The neural network is very similar to the one in Alerskans et al. (2022) (Paper II), but with minor (improving) changes e.g. from leaving out the wind direction and satellite azimuth angle (but keeping the relative angle between the satellite azimuth angle and the wind direction). In this comparison, the neural network also demonstrates superior performance when considering both standard deviation (0.45°C) and sensitivity (0.92) (Table 1, Appendix H). All models show increased standard deviations and decreased sensitivities in the Arctic (Figure 1–3 and Table 2A, Appendix H). In particular, the tree-based (i.e. bagging, random forest and the least squares boosting) models show decreased performance for extremes (e.g. strong winds and cold waters), while the neural network manages to compensate better for these effects (Figure 2A, Appendix H).

The results presented in Alerskans et al. (2022) (Paper II) and in Appendix H are very promising and indicate that ML models have a great potential to retrieve SST from PMW observations.

2.4 Channel Selection

This section summarizes the work done in Nielsen-Englyst et al. (2021a) (Paper III) to investigate the impact of using different radiometer channel selections for PMW SST retrievals using independent *in situ* observations as reference. This work is done to assess the optimal channel selection for PMW SST retrievals to prepare for future satellite missions (with a specific focus on CIMR, which has a slightly different channel configuration compared to AMSR-E and AMSR2), but also to investigate the effects of other scenarios e.g. where one or more channels fail during operations or in a case where future International Telecommunication Union (ITU) frequency allocations contaminate some of the specific bands used for PMW SST retrievals. Previously, theoretical information content studies have been conducted to assess the impacts from using different frequency channels (e.g. Kilic et al., 2018; Prigent et al., 2013). However, the large discrepancies among the

forward models (Kilic et al., 2019a) makes it important supplement these studies, with an assessment including *in situ* observations as reference. The use of *in situ* observations for validation may furthermore inform us about deficiencies and strengths of the different retrieval algorithms.

In Nielsen-Englyst et al. (2021a) (Paper III), two PMW SST retrieval algorithms are used namely the OE algorithm from Nielsen-Englyst et al. (2018) (Section 2.1) and the regression algorithm from Alerskans et al. (2020) (Paper I, Section 2.2). Unfortunately, the ML models were not developed at the time of this study. In the OE setup, we use an inflated diagonal covariance matrix element for SST to increase the sensitivity of the retrieved SST to true changes in SST (increases from 0.50 to 0.99) and to minimize the dependence on first guess. In the RE setup, the global-derived coefficients are used to ensure a consistent comparison of the different channel selections and to minimize the effects from localized algorithm coefficient derivations. This explains the inflated standard deviations in this paper compared to Nielsen-Englyst et al. (2018) and Alerskans et al. (2020) (Paper I). In this study, the focus in not on the absolute performances but on the relative differences between the channel selections and the retrieval algorithms.

In general, it is found that the retrieval performance increases (as expected) when more channels are included (Nielsen-Englyst et al., 2021a) (Paper III, Figure 5). Including more channels allow a better representation of the range of different observing conditions (Nielsen-Englyst et al., 2021a) (Paper III, Figure 6). The two retrieval algorithms agree that using a three-channel configuration of the 6, 10 and 18 GHz provides better SST retrievals than using the 6, 10, and 23 GHz configuration. This is demonstrated for all seasons and geographical regions (Nielsen-Englyst et al., 2021a) (Paper III, Figure 11). In general, it is found that withholding observations from the 23 and 36 GHz channels has the least impact on the SST retrieval performance and that the CIMR like channel configuration (excluding the 23 GHz channels) performs very well when compared to an AMSR-E configuration using both types of retrieval algorithm. This study also demonstrates a very good performance of the 6, 10, and 18 GHz channel combination. The limited impact of the 36 GHz channels supports the channel selection for CIMR, since the CIMR 36 GHz channels will likely only contribute with limited information to the SST retrievals, due to elevated NEdT for the 36 GHz (Donlon, 2020).

Nielsen-Englyst et al. (2021a) (Paper III) also demonstrates some of the strengths and weaknesses of the two types of retrieval algorithms. These will be discussed in the following section (Section 2.5), which also includes a comparison and discussion of the ML models.

2.5 Discussion

In this PhD study, multiple global PMW SST retrieval algorithms have been developed, validated and compared to improve the understanding and performance of PMW SSTs with a particular focus on the Arctic, where cloud clover severely limits the SST coverage from IR sensors. This section summarizes the main findings and discusses the strengths and weaknesses of the different kinds of retrieval algorithms. Moreover, suggestions for future work and priorities related to Arctic PMW SST retrievals are provided.

Generally, the retrieval algorithms can be grouped into two primary categories: physically based (or RTM-based) algorithms (e.g. OE) and statistical algorithms (e.g. RE+ML). The main advantage of the RTM-based algorithms is that the physical processes are evident in each step of the retrieval, and therefore, using RTM simulations can provide useful insights into the physics of the retrieval. Both the RE algorithm and the ML models fall into the second category of retrieval algorithms, which rely on matches of *in situ* and

satellite brightness temperatures as well as some environmental and instrument information depending on the design of the retrieval. The understanding of the physics behind the retrieval e.g. gained from RTM simulations, is important for the design and input selection of the RE algorithm as well as the feature selection in ML models.

In contrast to the OE algorithm, the ML and RE algorithms require a large and representative dataset to train the models and to derive the relevant retrieval coefficients. A drawback from this is that the retrievals can only match conditions where *in situ* measurements are available. In conditions that are not (well) represented by *in situ* measurements, the ML and RE algorithms have limited basis for algorithm/model training (and validation), and thus, we can only rely on their extrapolation capabilities in these conditions. This can be a problem in under-represented regions, such as the Arctic, where the environmental conditions are moreover changing fast and thereby out-dating the representativeness of the training datasets at an equally fast pace. Despite the high latitudes being under-represented, Nielsen-Englyst et al. (2021a) (Paper III) found that the RE algorithm outperforms OE in high-latitudes, in particular during winter (Nielsen-Englyst et al., 2021a) (Paper III, Figure 11), and better represent the range of different observing conditions compared to OE (Nielsen-Englyst et al., 2021a) (Paper III, Figure 6).

To provide accurate SST retrievals, the OE algorithm requires an accurate RTM and the satellite sensor to be well characterised and calibrated. Differences in simulated and observed brightness temperatures can normally be attributed to errors in the RTM or measurement errors such as imperfect calibration, channel contamination etc (Minnett et al., 2019). Typically, *ad-hoc* corrections have been used to bias-correct the simulated brightness temperatures to better match the observations as in Nielsen-Englyst et al. (2018) and Nielsen-Englyst et al. (2021a) (Paper III). This is in contrast to the statistical approach where these effects to some extent are empirically accounted for in the coefficients of the RE algorithm and in the design of the ML models, and thus, reducing the need for understanding the instrument characteristics and calibration.

Both the OE and RE algorithm make a series of initial assumptions about how the measured brightness temperatures are related to the SST and other physical properties (as well as proxies for these), and are constrained to the established relationships during the retrieval. In ML models, on the other hand, there are much fewer assumptions about the functional form of how the input features are related to the SST. This flexibility of ML models enables the development of much more complex functional forms that may better approximate the unknown instrument effects as well as the actual physical processes that underlie the emission of microwave radiation and the transfer through the atmosphere, and thus provide a more accurate SST retrieval.

In contrast to the RE and ML algorithms, the OE algorithm uses *a priori* information to constrain the retrievals. In situations with limited information in a given set of channels, the OE algorithm will put more weight on the first guess (in this case from ERA-I) and thereby still be able to provide reliable SST retrievals (although at a lower sensitivity). This is demonstrated in Nielsen-Englyst et al. (2021a) (Paper III, Figure 5+6), where the OE algorithm shows a relatively stable performance for the different channel subsets, while the RE algorithm clearly shows reduced performance in cases where few channels are included.

Besides providing insights into the physics, the OE framework simultaneously retrieves SST, WS, TCLW and TCWV, and offers several options for assessing the quality and sensitivity of the individual retrievals. These outputs can provide valuable information on pixel level about the retrieval quality, and can be used directly to identify and discard

erroneous SST retrievals (e.g. resulting from rain scattering, side-lobe contamination from land/ice, extreme wind, atmospheric attenuation and emission, sun-glint, RFI, imperfect calibration etc.). An example of this was seen in ESA CCI SST project, where we examined (mainly before my PhD start) the impact by atmospheric components such as cloud cover, aerosols and cloud ice and water content using observations from the active sensor CALIOP (Cloud-Aerosols Lidar and the Infrared Pathfinder) and MODIS from 2010 (Høyer et al., 2019). It was found that cloud cover and aerosols had no impact on the OE and RE SST retrievals, while deep convective clouds (and the related total ice water content) showed a degradation in the retrieval performance. The RE uncertainty was not able to capture the effect, while the OE algorithm showed great potential for identifying and screening the observations influenced by deep convective clouds, both by using the retrieved TCLW and the theoretical OE uncertainty. The capability of the theoretical OE uncertainty is also seen in Nielsen-Englyst et al. (2021a) (Paper III, Figure 5), where it provides an equally good uncertainty indicator for the OE and RE algorithm. The higher flexibility of ML models may allow them to account for part of the contamination e.g. from deep convective clouds and RFI. An example of this is seen in Alerskans et al. (2022) (Paper II, Figure 4), where the two ML models seem to be more robust against RFI compared to the RE algorithm, e.g. near Ascension Island, which is known for both ground-based and satellite-based RFI (Gentemann & Hilburn, 2015).

The largest standard deviations against *in situ* measurements are generally observed in high latitudes (i.e. cold waters), conditions with strong winds, and in regions with large spatial SST gradients, with the latter being partly attributed to the difference in spatial sampling of the satellites and buoy measurements (Alerskans et al., 2020; Castro et al., 2012) (Paper I). The increased uncertainties in cold waters are related to the reduced sensitivity to SST for the low frequency channels as seen in Nielsen-Englyst et al. (2021a) (Paper III, Figure 2a) as well as in Kilic et al. (2018) and Prigent et al. (2013). In Nielsen-Englyst et al. (2021a) (Paper III, Figure 6a), the OE algorithm clearly shows elevated uncertainties for cold and rough waters for all channel combinations. This is in contrast to the RE algorithm, for which increased uncertainties in cold and rough waters are most pronounced for the two-frequency combinations without 18 GHz, while more (>2) frequencies provide enough information to balance the lower sensitivity in cold and rough waters (Nielsen-Englyst et al., 2021a) (Paper III, Figure 6b).

Part of the uncertainty in cold waters may also arise in the modelling of the higher frequency channels, where the sensitivity to SST to a large extent originates from in cold waters (with decreasing brightness temperatures for increasing SST) (Nielsen-Englyst et al., 2021a) (Paper III, Figure 2). As suggested in Nielsen-Englyst et al. (2021a) (Paper III), this effect could possibly be reduced to some extent by including the covariances (off-diagonals) between the geophysical parameters in the OE algorithm. Since these are currently not well known, future work should aim at estimating these in a similar way as done in Merchant et al. (2020) for the IR domain. In Merchant et al. (2020) a bias correction and the error covariance parameters are retrieved as a part of the state vector using matches of satellite observations to *in situ* reference measurements (refereed to as Bias Aware OE, BAOE). The aim is to bring the OE framework closer to optimality which is characterised by well-known error covariances and on the errors in the prior, measurements and RTM being zero mean (Merchant et al., 2008). This work has been initiated in this PhD study, where a set of simulated brightness temperatures have been used to test the BAOE framework instead of the actual satellite observations. To replace the satellite observations, simulated brightness temperatures are calculated by running the forward model once with the prior state vector from ERA-I. In the BAOE retrieval,

simulated brightness temperatures have been used with bias and/or noise added to one or more channels. The BAOE setup succeeded in reproducing both the noise and bias (individually and simultaneously) added to the brightness temperatures (individually and simultaneously for the channels) in a test case, where the simulated brightness temperatures were not updated for each iteration. More work is needed to extend the current OE implementation to be bias-aware, which however, is an important step towards inter-sensor consistency in a future OE PMW SST CDR.

The RE PMW SST CDR shows good inter-sensor consistency (Alerskans et al., 2020) (Paper I, Figure 12), and this can likely also be obtained for the ML models if the sensors are referenced to the same *in situ* network (as is the case for the RE algorithm). This approach, however, has the risk of introducing biases related to inadequacies in the observational network. In Alerskans et al. (2020) (Paper I), this effect is attempted to be minimized e.g. by ensuring a balanced geographical distribution, and it appears small as the validation against Argo floats confirms the validation results against drifting buoys. Future work should aim at establishing and maintaining SI-traceability (i.e. traceability of the measurement standards and measuring instruments to the International System of Units, SI) of global, independent and fully characterized *in situ* measurements (Fiducial Reference Measurements, FRM) for satellite-observed PMW SSTs as has been established for SST retrievals from IR satellite sensors (e.g. Barton et al., 2004; Minnett, 2010; Theocharous et al., 2019) to support both the algorithm development and validation.

From above review of the different retrieval algorithms developed/tested in this PhD study, it is evident that each type of retrieval algorithm has its own strengths and weaknesses. The choice of retrieval algorithm should be based on careful consideration of the particular situation i.e. the retrieval requirements, availability of *in situ* observations, performance timeliness requirements, instrument characterization, forward model accuracy, availability of auxiliary information from the satellite observations, and stability/consistency.

Both the RE and ML algorithms can easily be adapted to CIMR, taking into account the improved NEdTs and higher spatial resolution at low frequencies as well as the addition of the 1.4 GHz (L-band) channels. The OE algorithm requires an extension of our forward model to include 1.4 GHz channels, which requires that salinity is also retrieved, and ideally a fully BAOE implementation to retrieve bias correction and error covariance parameters.

Future development could consider combining retrievals e.g. by using RTM simulations to determine regression coefficients and/or retrieval uncertainties, or using the OE retrieved state as input to the RE and/or ML models to obtain higher consistency (compared to using external auxiliary information e.g. from ERA5/ERA-I). This could potentially include the development of an integrated ocean and sea ice OE processor (similar to Melsheimer et al., 2009; Pedersen, 1994; Scarlat et al., 2017, although their focus were on sea ice parameters) to obtain higher consistency. However, an integrated accurate retrieval requires improvements to the existing snow and sea-ice emissivity models, but initially, it could be considered to use the information from the integrated retrieval to identify open water, and use a separate SST retrieval in these regions (e.g. using the SST from the integrated retrieval as first guess). This would still increase the consistency, and could potentially enable PMW SST retrievals closer to the sea ice. Other methods could also be explored in order to allow SST retrievals closer to the sea ice and coasts e.g. using similar methods as used for salinity retrievals in Meissner and Manaster (2021) and Olmedo et al. (2017).

Considering the ML models developed/tested in this PhD study, the neural network seems most promising considering both the validation and sensitivity results. Moreover, it is

flexible, easy to implement and train, and it is able to capture the non-linear relations. Future work, should investigate the performance against *in situ* observations when using different subsets of the features from Alerskans et al. (2022) (Paper II) in the retrieval (to supplement the feature analysis) in order to get a better understanding of different model designs as well as identifying the optimal features to be included for PMW SST retrievals.

All algorithms have been developed to retrieve subskin SST, but derived for cases not affected by diurnal warming (since these matchups were filtered out). This enabled the use of SST observations from drifting buoys (with a nominal depth of 20 cm) and Argo observations (with a depth of 5 m) for testing and validating the retrieval algorithms, since the subskin and depth temperatures are the same on average in the absence of diurnal warming (Gentemann & Minnett, 2008; Minnett & Kaiser-Weiss, 2012). Due to the high SST sensitivity, SST can also be retrieved during daytime conditions with diurnal warming, but accurate validation requires a method or model to account for the subskin to 20 cm or 5 m differences. Such a model is likewise required in order to combine SST retrievals made at different depths (Donlon et al., 2007). To ease the user uptake and facilitate blended SST products, the PMW SST CDR from Alerskans et al. (2020) (Paper I) has been adjusted using such a model to represent the temperature at 20 cm depth (available at http://gws-access.ceda.ac.uk/public/esacci-sst/PMW2.0_release/AMSR/L2P) for consistency with the ESA-CCI IR SST retrievals (Embury et al., 2012; Merchant et al., 2019).

The PMW SST algorithm development and analysis in this chapter demonstrate the strengths and weaknesses of the retrieval algorithms and the impact of using different channels selections. These are important steps to prepare for future PMW satellite missions and to fully benefit from the past and current PMW satellite sensors. In this relation, the first European PMW SST CDR has also been produced and demonstrated consistent and promising validation results. Overall, the findings of this chapter indicate that PMW SST retrievals have a large potential to supplement the IR SSTs in the Arctic. This will be investigated and discussed in the following chapter (Section 3.3) and in Nielsen-Englyst et al. (2023a) (Paper V).

CHAPTER 3

Arctic Surface Temperatures

One of the key indicators of climate change is the rapid decline in Arctic sea ice extent (e.g. Cavalieri & Parkinson, 2012; Onarheim et al., 2018; Stroeve & Notz, 2018) with the largest downward trend observed during the end of the melt season in September (Serreze & Meier, 2019). Due to the temporally varying sea ice coverage, the most consistent way to monitor the surface warming of the Arctic oceans is by combining sea and sea-ice surface temperatures. This is in contrast to the existing global gridded observational datasets, which typically combine SST with near-surface air temperatures (T2m) in sea ice covered regions.

This chapter investigates how SST observations can be integrated with sea ice parameters in a multi-sensor gap-free (L4) combined sea and sea-ice surface temperature climate dataset for the Arctic (>58°N). This work was initiated with an assessment of the consistency between satellite-observed SST and sea-ice products, and the construction of a consistent and accurate sea ice concentration (SIC) field of the Arctic (Section 3.1), which is important for the generation of a combined L4 SST and IST product. Section 3.2 summarizes the development, generation and validation of the first IR satellite-based L4 SST/IST CDR (including uncertainties) of the Arctic presented in Nielsen-Englyst et al. (2023b) (Paper IV). Since, IR satellite observations are obstructed by clouds, there is a large potential in including the almost all-weather PMW SSTs in the Arctic. Therefore, Section 3.3 describes the work that has been done towards combining IR and PMW SSTs in order to improve the Arctic SST estimation (Nielsen-Englyst et al., 2023a) (Paper V).

Currently, no satellite observed ISTs are included in the existing global gridded observational datasets (or dynamical reanalyses) over sea ice, and the estimates are thus based on a very limited number of *in situ* observations leading to large uncertainties. To compensate for the lack of *in situ* measurements, Section 3.4 brings an investigation of how the satellite-observed ISTs can be used to derive T2m in the Arctic (Nielsen-Englyst et al., 2021b) (Paper VI). The satellite-derived T2m product provides multiple usages e.g. for assimilation, evaluation and as a supplement to existing surface temperature estimates. Section 3.5 provides an example of one application in which the satellite-derived T2m estimates are used for benchmarking ERA5 and climate models in the Arctic (Tian et al., 2023) (Paper VII). Finally, Section 3.6 provides a discussion of the results, limitations, applications and suggestions for future work.

3.1 Consistent Sea Ice Product

With the sea ice diminishing and becoming more fragmented, consistency between SST and sea ice variables becomes more important. Traditionally SST and sea ice products have been developed individually to fulfill the requested user requirements over time. Several studies have compared the different SIC records and evaluated their performances using information from ice charts and ice drift observations (e.g. Cavalieri & Parkinson, 2012; Kern et al., 2019; Onarheim et al., 2018; Serreze & Meier, 2019; Stroeve & Notz, 2018). Similarly, different SST products have been compared in several previous studies and compared to *in situ* observations in the Arctic (e.g. Castro et al., 2016; Dash et al., 2012; Vazquez-Cuervo et al., 2022).

In this PhD study, the consistency between satellite-observed sea ice and SST products is investigated (Nielsen-Englyst et al., 2023b) (Paper IV), with the overall aim to derive an accurate and consistent SIC product (referred to as DMI-SIC), which is the basis for both SST and IST retrievals as well as for the generation of a combined L4 SST and IST product of the Arctic.

The DMI-SIC is generated using SIC information from the EUMETSAT OSI-SAF global SIC products OSI-450 and its interim extension OSI-430b (hereafter referred to as OSI-450 for the whole time series), which are based on coarse resolution (30–60km) PMW satellite data, and the ESA CCI programme SICCI-25km product, which is derived from medium-resolution (15–25 km) PMW satellite data. These are resampled onto a 0.05° regular latitude longitude grid and compared to the ESA CCI SSTs (Merchant et al., 2019; Merchant et al., 2009) from 1982 to 2010, and the Operational Sea Surface Temperature and Sea Ice Analysis (OSTIA; Donlon et al., 2012; Good et al., 2020) from 2011 and onwards. The largest number of potentially inconsistent cases with "warm" ($> 3°C$) SST and at the same time sea ice (SIC>15%), are found in the marginal ice zone and coastal areas during summer. The coarser resolution OSI-450 product shows more of these cases due to more pronounced land and sea-ice spill-over contamination compared to SICCI-25km (Kern et al., 2022; Lavergne et al., 2019).

The DMI-SIC uses the SICCI-25km whenever it is available and OSI-450 otherwise. To improve OSI-450 and to increase the consistency of the merged SIC product, a number of filters have been applied using information from the higher resolution SST products. Nielsen-Englyst et al. (2023b) (Paper IV) provides the details of the SIC filters and the generation of the DMI-SIC. Figure 3+4 in Nielsen-Englyst et al. (2023b) (Paper IV) show examples of the improved consistency, and reduced spurious sea ice from land contamination in OSI-450, which is obtained after filtering. As an effect of the coarse resolution PMW sensors, both SICCI-25km and OSI-450 have coastal challenges in the Baltic region, and for that reason we use a high resolution SIC product based on manual ice charts from the SMHI (1982–2011) and the CMEMS 1 km SIC fields (2012–present) in this region (Høyer & Karagali, 2016).

The derived DMI-SIC feeds directly into the L4 Arctic SST/IST CDR described in Section 3.2 and Nielsen-Englyst et al. (2023b) (Paper IV). It can also be used for assimilation as well as model evaluations of which two examples are provided in the following section.

3.1.1 Assimilation and Evaluation

The consistent DMI-SIC product described in Section 3.1 and Nielsen-Englyst et al. (2023b) (Paper IV), has turned out to be useful for multiple applications. Firstly, an earlier (and slightly different) version of it has been ingested into the Copernicus Arctic Regional Reanalysis (CARRA) which covers two domains of the European Arctic with 2.5 km horizontal resolution from 1991 to present, with regular updates planned to be provided

until June 2025 (Yang et al., 2020). It is the plan to include the updated DMI-SIC from Nielsen-Englyst et al. (2023b) (Paper IV, Section 3.1) in the second generation of the Arctic reanalysis, CARRA2, which will have a pan-Arctic coverage. This is under development and the production is planned to start in 2024, with release of the full data set (∼1991–2025) in 2026 (https://climate.copernicus.eu/copernicus-arctic-regional-reanalysis-service).

Secondly, the derived DMI-SIC has been used to evaluate the capability of the improved and recently-launched DMI sea ice operational forecast system (DMI-HYCOM-CICE) to predict the sea ice edge (Ponsoni et al., 2023) (Paper VIII). The DMI-SIC is included as one out of three different observational references with the other two being manual ice charts and Automated Sea Ice Products (ASIP). In that relation, an inter-comparison of the three observational references is also conducted to provide an estimate of the observational uncertainty. In general, the DMI-SIC shows very good agreement with manual ice charts except from off the southeastern Greenlandic coast during summer, where the DMI-SIC reports lower SICs than the manual ice charts (Ponsoni et al., 2023) (Paper VIII, Figure 10+11). Part of this discrepancy may be linked to melt ponds being interpreted as open water in the microwave observations. In general, the daily DMI-SIC proved to be very valuable for identifying and evaluating the impact of different changes made to the forecast system.

3.2 Sea and Sea-Ice Surface Temperature Reanalysis

Due to the varying location and extent of sea ice, the most consistent way to monitor the surface temperature change of the Arctic oceans is by blending sea and sea-ice surface temperatures. This section summarizes the development, construction and validation of the first L4 Arctic (>58°N) combined SST and IST CDR (1982-2021), which is presented in Nielsen-Englyst et al. (2023b) (Paper IV). The underlying algorithm combines multiple IR satellite observations and performs a statistical optimal interpolation (OI) to obtain daily gap-free fields with a 0.05 degrees resolution. Each daily L4 SST/IST field is accompanied by an uncertainty estimate, which is a direct output of the OI processing scheme.

The L4 SSTs represent the daily mean temperature at a depth of 20 cm (as the SST input data; Merchant et al., 2019). Validation of the L4 SST shows mean differences of 0.01°C, 0.04°C and 0.04°C and standard deviations of 0.54°C, 0.56°C and 0.51°C, against *in situ* measurements from drifting buoys, moored buoys and Argo floats, respectively (Nielsen-Englyst et al., 2023b) (Paper IV, Table 2). The drifting buoys have been used for a static mean bias correction (+0.16°C) of the L4 SST analysis, while moored buoys and Argo floats provide independent observations.

The L4 ISTs are clear-sky averages and a clear-sky bias correction of +0.85°C (following Nielsen-Englyst et al., 2019) has been added over sea ice (>70%) and linearly scaled towards 0°C over open ocean (using the DMI-SIC from Section 3.1 as scaling factor). The L4 ISTs have been compared with KT-19 measurements from IceBridge flights (IAKST1B v2) with a mean difference of 1.52°C and standard deviation of 3.12°C, as well as with T2m measurements from the North Pole (NP) ice drifting stations, ECMWF distributed buoys and CRREL buoys, with mean differences of –2.35°C, –3.21°C and –2.87°C and standard deviations of 3.12°C, 3.34°C and 3.36°C, respectively (Nielsen-Englyst et al., 2023b) (Paper IV, Table 2). The physical difference between IST and T2m introduces a temperature difference, which may be up to several degrees (Nielsen-Englyst et al., 2019), which is also included in these results. This effect is further investigated in Nielsen-Englyst et al. (2021b) (Paper VI) where a method is also developed to derive T2m from satellite-observed IST.

The resulting combined surface temperature product covering ocean, sea ice and the marginal ice zone is the first of its kind and it can be used as a consistent climate indicator for monitoring day-to-day variations as well as long-term climate trends in the Arctic oceans. The combined sea and sea-ice surface temperature of the Arctic (>58°N) has increased with 4.5°C for the period 1982–2021, with a peak warming in the north-eastern Barents Sea of around 10°C (Nielsen-Englyst et al., 2023b) (Paper IV, Figure 14). The exceptional warming of the Barents Sea is in agreement with most recent findings (AMAP, 2021; Isaksen et al., 2022; Rantanen et al., 2022), and is likely related to the rapid diminishing winter sea ice extent in this region (Isaksen et al., 2022; Onarheim & Årthun, 2017).

3.3 Impact of Microwave Sea Surface Temperatures

The frequent and persistent cloud cover in the Arctic limits the extent to which SST can be retrieved from IR satellite sensors. Due to the almost all-weather capability of PMW sensors, there is a large potential of supplementing the IR SSTs with PMW SSTs in the Arctic. Therefore, it is important to investigate how to best combine IR and PMW SST retrievals in the Arctic in a gap-free analysis. This is done in Nielsen-Englyst et al. (2023a) (Paper V), which uses the IR-based L4 Arctic SST/IST CDR developed in (Nielsen-Englyst et al., 2023b) (Paper IV) as baseline and investigates the impact of including observations from the ESA-CCI PMW SST CDR, which was developed and generated in Alerskans et al. (2020) (Paper I).

The different characteristics (e.g. footprint, spatial resolution and coverage) of IR and PMW observations need to be taken into account in order to combine the two data sets properly (Castro et al., 2016). Different methods have been tested using one year (2015) of the data. The inclusion of PMW SSTs shows improved validation results against drifter SSTs for all tested methods (Nielsen-Englyst et al., 2023a) (Paper V, Table 1). The PMW SSTs show lower mean differences against drifters compared to the IR SSTs, which in general are too cold. For climate analyses, it is important to correct for systematic biases in the IR and PMW SSTs relative to each other. For consistency with time periods without Arctic PMW SSTs, the PMW SSTs have been adjusted to IR SSTs using a dynamic high-latitude bias correction adapted from Høyer et al. (2014) to generate a consistent IR and PMW Arctic SST CDR for the ESA-CCI PMW SST CDR period (2002-2017).

The inclusion of PMW SST reduces the standard deviations from 0.54°C, 0.55°C and 0.47°C to 0.47°C, 0.54°C and 0.41°C against drifters, moorings and Argo floats, respectively (Nielsen-Englyst et al., 2023a) (Paper V, Table 2). The improvements are seen in almost all regions including those already covered by IR-observations (Paper V, Figure 11b), but with the largest improvements in IR data sparse regions (Paper V, Figure 12b). The mean theoretical uncertainty decreases by 0.08°C, which is in good agreement with the observed reduction in standard deviation against drifters. Generally, the results are very promising and they will very likely improve even further in the future with the launch of AMSR3 and CIMR, which are highly complementary. CIMR will enable PMW SST retrievals with lower uncertainties and much closer to coasts and sea ice (where the largest uncertainties are found) than possible today, while the combination of AMSR3 and CIMR will provide unprecedented coverage and revisit time of the Arctic oceans.

3.4 Near-Surface Air Temperatures from Satellites

Daily Arctic near-surface air temperatures (T2m) derived from satellite observations have a large potential to increase the information and quality of the existing gridded observational surface temperature estimates and to supplement contemporary reanalyses. Therefore,

Nielsen-Englyst et al. (2021b) (Paper VI) investigates the potential of using satellite ISTs to derive T2m over ice surfaces based on the understanding of the empirical relationships between *in situ* observed skin temperatures and T2m from Nielsen-Englyst et al. (2019). The work presented in Nielsen-Englyst et al. (2019) and in Nielsen-Englyst et al. (2021b) (Paper VI) is part of the EU Horizon 2020–funded EUSTACE project (2015–2019, www.eustaceproject.org).

Nielsen-Englyst et al. (2021b) (Paper VI) introduces a methodology for estimating T2m over the Arctic sea ice and the Greenland Ice Sheet (GrIS) based on daily clear-sky (L3) observations from the Arctic and Antarctic Ice Surface Temperatures from thermal Infrared satellite sensors (AASTI) version 1 (v1), covering the period 2000-2009. For each day, the daily mean T2m and a corresponding uncertainty are provided on a 0.25 degree regular latitude-longitude grid. The T2m satellite derivation is based on a linear regression model tuned against *in situ* observed T2m. Different models were examined and a model using the satellite skin temperature combined with a seasonal variation was selected to predict corresponding T2m. Separate regression coefficients were derived and used for the Arctic sea ice and the GrIS.

The satellite based T2m show mean differences of 0.30°C and 0.35°C and standard deviations of 3.45°C and 3.18°C for the GrIS and Arctic sea ice, respectively. This is similar or better performance compared to ERA-I and its successor ERA5 despite that these reanalyses actively assimilate available *in situ* observations. Both ERA-I and ERA5 suffer from warm biases in the Arctic with mean differences of 2.03°C (3.41°C) and 2.19°C (1.14°C) for ERA5 (ERA-I) over the GrIS and Arctic sea ice, respectively.

Days with too few observations (due to clouds) to resolve the daily temperature cycle are excluded from the analysis and the satellite-derived T2m dataset has gaps in those places. Whenever, the satellite-derived T2m is available, it provides an estimate of the daily averaged (all-sky) T2m since it has been regressed towards all-sky *in situ* observations. Figure 15a in Nielsen-Englyst et al. (2021b) (Paper VI) shows an increasing coverage over time period (2000-2009), and seasonal minimum coverage over Arctic sea ice towards the end of the summer (Figure 15b, Paper VI). Despite the gaps, the satellite-derived L3 T2m provides a much better spatial coverage than the sparse *in situ* network in the Arctic.

The gaps can be filled using various methods e.g. using data assimilation in atmospheric models, optimal interpolation (as in Nielsen-Englyst et al., 2023b) (Paper IV) or other statistical techniques. In the EUSTACE project, *in situ* observed T2m and satellite-derived L3 T2m estimates (over land, lakes, ocean, and ice), including uncertainty estimates, were combined using an advanced statistical model to generate a global and gap-free daily analysis of surface air temperatures from 1850 to 2015 (Rayner et al., 2020) (Paper IX). The satellite-derived T2m product developed in Nielsen-Englyst et al. (2021b) (Paper VI) is used directly as input to the EUSTACE surface air temperature analysis for the GrIS and the Arctic sea ice presented in (Rayner et al., 2020) (Paper IX). The combination of *in situ* measurements and satellite-derived temperatures provides a much more complete global dataset of air temperature compared to global gridded observational datasets. It is complementary to products from dynamical reanalyses and it moreover extents further back in time than many of the dynamical reanalyses.

3.4.1 Gap-free Reanalysis

Nielsen-Englyst et al. (2023b) (Paper IV) also explores the potential of using the regression model and coefficients from Nielsen-Englyst et al. (2021b) (Paper VI) to derive gap-free T2m over the Arctic sea ice using the satellite ISTs from the L4 SST/IST CDR. In this context, the clear-sky bias correction of the L4 IST is omitted since a clear-sky bias correc-

tion is already incorporated in the IST-T2m regression model, which has been regressed towards all-sky *in situ* T2m observations. It should be noted that this approach, using L4 IST to estimate L4 T2m, differs from the approach in Rayner et al. (2020) (Paper IX), where L4 T2m is generated from L3 T2m (+*in situ* observations).

Using the satellite-derived L4 T2m estimates based on Nielsen-Englyst et al. (2023b), Nielsen-Englyst et al. (2021b) (Paper IV+VI), the mean differences are reduced to 0.65°C, –0.45°C and –0.17°C when compared to T2m measurements from the NP ice drifting stations, ECMWF distributed buoys and CRREL buoys, respectively, while the standard deviations remain more or less the same.

These initial results indicate that it is possible to derive reliable T2m over sea ice based on the satellite-observed L4 ISTs. The satellite-derived L4 T2m product provides an important supplement to the sparse *in situ* air temperature network in the Arctic and to the existing model-based air temperatures. It has a large potential to be used for assimilation, global surface temperature reconstructions or for evaluation of global reanalyses and climate models. One example of the latter is provided in the following section (Section 3.5), which summarizes the work presented in Tian et al. (2023) (Paper VII), while additional applications are discussed in Section 3.6.

3.5 Benchmark for Climate Models and Reanalyses

Traditionally, global reanalyses have been used for evaluating climate models (e.g. Bock et al., 2020; Davy & Outten, 2020; IPCC, 2021). However, it is evident that many global reanalyses have issues in simulating the extreme cold surface temperatures in large parts of the Arctic Ocean with warm biases of ∼5–10°C (Batrak & Müller, 2019). As recognized in Batrak and Müller (2019), the widespread use of global reanalyses for model validation, initialisation of prediction systems, forcing of ocean and sea-ice models, etc., makes it crucial to take into account this warm bias.

This concern is also addressed in Tian et al. (2023) (Paper VII) in relation to the evaluation of global reanalyses and climate models in the Arctic. Tian et al. (2023) (Paper VII) uses the monthly mean satellite-derived T2m estimates from Nielsen-Englyst et al. (2023b) (Paper IV) and Section 3.4.1 as benchmark for evaluation of ERA5 and the latest generation (phase 6) of the Coupled Model Inter-comparison Project (CMIP6) historical simulations.

In agreement with the previous findings (Batrak & Müller, 2019; Wang et al., 2019), Tian et al. (2023) (Paper VII, Figure 1) shows a widespread warm ERA5 in the central Arctic Ocean, when compared with the satellite-derived T2m. In contrast, the CMIP6 shows a very good agreement with the satellite-derived T2m in the central Arctic Ocean. Both ERA5 and CMIP6 are colder than the satellite-derived T2m towards the sea ice margins, with the largest deviations in the CMIP6, which are likely explained by an overestimation of the CMIP6 sea ice extent, which is possible explained by an underestimation of the northward ocean heat transport (Shu et al., 2020).

It is also evident that ERA5 has a warm offset throughout the period (1982-2020), while the CMIP6 ensemble mean is in close agreement with satellite-derived T2m for the duration of the CMIP6 historical simulations (1982-2014) in the central Arctic Ocean (SIC>70%). In this region, it was also found that both ERA5 and CMIP6 underestimate the T2m trend (by –0.23°C decade^{-1} and –0.18°C decade^{-1}, respectively) compared to the satellite-derived T2m estimate (Tian et al., 2023) (Paper VII, Figure 1).

These results highlight the importance of having accurate and absolute reference fields for

evaluating and assessing the current as well as the next generation of dynamical reanalyses and global climate models (CMIP7) in the Arctic oceans. This and other applications of the satellite-derived T2m (and IST), as well as limitations to the current satellite-derived T2m estimates, will be discussed in the following section (Section 3.6).

3.6 Discussion

The work presented in this chapter clearly demonstrates the advantages of combining IR and PMW SSTs, to compensate each other's deficiencies and complement each other, in the generation of an Arctic L4 SST/IST product. The inclusion of high quality PMW SSTs provides valuable information which results in a general and substantial reduction in standard deviation against drifting buoys and Argo floats (Nielsen-Englyst et al., 2023a) (Paper V, Table 3). The improvements are seen in almost all regions, including those already covered by IR sensors (Paper V, Figure 11b), but it is also found that the reduction in standard deviation increases as a function of the number of days since the last IR observation was available in that particular grid cell (Paper V, Figure 12b). The results emphasize that PMW sensors are an important part of the constellation of SST sensors for monitoring the Arctic.

Considering the merged IR and PMW Arctic L4 SST/IST product there are still some challenges to be considered in future development, and these are outlined in the following.

Generally, the largest L4 SST standard deviations and L4 SST uncertainties are found close to the coasts and sea ice (Nielsen-Englyst et al., 2023a) (Paper V, Figure 11a+13a). In these regions, sparse SST coverage from both PMW and IR sensors limits the extent to which accurate SSTs can be estimated. Moreover, large spatial SST gradients usually exist in these regions e.g. along the marginal ice zone (Carvalho & Wang, 2020). Part of the elevated uncertainties in these regions will likely be mitigated in the future with the launch of higher resolution sensors both in the IR and PMW domain. High resolution coastal IR SST may be possible in the future with the planned Thermal InfraRed Imaging Satellite for High-resolution Natural resource Assessment (TRISHNA) mission (Buffet et al., 2021) and the ESA Copernicus Land Surface Temperature Monitoring (LSTM) mission (Koetz et al., 2019), while the improved spatial resolution from CIMR will enable PMW SST retrievals much closer to the coasts and sea ice than possible with previous and current PMW radiometers. Future work should focus on improving the validation close to the sea ice e.g. by using Saildrone observations (Gentemann et al., 2020; Jia et al., 2022; Vazquez-Cuervo et al., 2022) and potentially improving the retrievals in this region. As discussed in Section 2.5, recent studies have developed methods to reduce the land and sea ice contamination in salinity retrievals to improve retrievals closer to coasts and sea ice (Meissner & Manaster, 2021; Olmedo et al., 2017), and these methods may also be applicable to SST. Moreover, a combined and simultaneous OE retrieval of ocean and sea-ice parameters may also enable PMW SST retrievals closer to the sea ice. Both of these opportunities could be investigated in the future.

In Nielsen-Englyst et al. (2023a) (Paper V, Figure 12b), it is found that regions rarely observed by IR sensors (>60 days without IR observations) during fall and early winter on average are provided with a too warm (first guess) SST. This is clearly seen for the IR-only run, but much less pronounced when including the more frequent PMW SSTs. This is explained by the setup of the OI scheme, where the previous day's analysis is used as first guess in long periods in grid cells where the neighboring observations are too far away to provide any weight in the OI scheme. Since, most of existing and widely used global L4 SST products are also based on OI (e.g. Castro et al., 2016; Vazquez-Cuervo et al., 2022) and using previous day's analysis as first guess (e.g. NOAA OI SST and

REMSS MW-IR SST), it is likely that they also suffer from the same artefact. This is of course undesirable and should be addressed in future updates of the Arctic L4 SST/IST reanalysis as well as new Arctic and global L4 SST analyses in particular when only IR observations are used, but also in combined IR and PMW analyses where the problem likely persists close to coasts and sea ice.

In the current configuration of the merged IR and PMW Arctic L4 SST/IST reanalysis, the L3 PMW SSTs have been adjusted to L3 IR SSTs for consistency with time periods when no Arctic PMW SSTs are available (Nielsen-Englyst et al., 2023a) (Paper V). While the IR SSTs are generally too cold, the PMW SSTs show a very good agreement with *in situ* observations. Future work should consider using the PMW SST observations to adjust the cold Arctic IR SSTs (extended to times when no PMW SSTs are available) e.g. by using CIMR as a reference. In this relation, future work could also explore the potential of joint retrievals, where the L2 IR data are tied to the L2 PMW data, e.g. through direct referencing or through a common *in situ* reference data set, to increase the consistency before entering the L4 generation.

As discussed in Nielsen-Englyst et al. (2023b) (Paper IV), the IR-based L4 SST meets the CDR requirement for SST stability (GCOS, 2022; Ohring et al., 2005), but at the time of the Paper IV writing, no stability requirements were available for IST. However, in 2022, the IST was included as an ECV by GCOS, and corresponding stability requirements were established. These are identical the GCOS stability requirements for land surface temperature i.e. 0.3°C decade^{-1} for the threshold stability and 0.1°C decade^{-1} for the target stability (GCOS, 2022), which are also referenced to in Nielsen-Englyst et al. (2023b) (Paper IV). Considering the NP validation (which spans the longest period), the L4 IST meets the target stability of a CDR according to GCOS (2022). Meeting the required stability is important for studying and assessing long-term climate variability and change (Minnett et al., 2020; Ohring et al., 2005). Nevertheless, as emphasized in Nielsen-Englyst et al. (2023b) (Paper IV), the limited availability of high-quality long-term reference *in situ* IST observations (with known and controlled stability) limits the extent to which satellite ISTs can be evaluated for meeting the CDR requirements. Future work should focus towards establishing and maintaining FRMs to support IST algorithm development and validation, but also improving the capacity of the established (but sparse) *in situ* network that can be used for validation of the satellite IST products. Future work should also aim at providing *in situ* observations of the marginal ice zone, including activities that are capable of providing estimates of the spatial variability and radiometric signals, such as flight campaigns, to quantify the uncertainties, assess the performance of the surface temperature estimates and potentially improve them.

Future work should also aim at improving the L4 IST e.g. by including other available satellite IST products such as the MODIS IST (Hall et al., 2004), VIIRS IST (Key et al., 2013; Liu et al., 2015), and a potential future SLSTR IST product. This PhD study has only been concerned with IR ISTs, but future work could also investigate the potential use of PMW ISTs, which has been increasingly available at daily temporal resolution based on the vertically polarized 6.9-GHz channel AMSR-data (e.g. Comiso et al., 2003; Kilic et al., 2019b; Le Traon et al., 2015). In contrast to the IR sensors which measure the skin surface temperature, the derived emitting layer temperature from PMW sensors are closer to the physical temperature of the snow/ice interface (Tonboe, 2010; Tonboe et al., 2011; Ulaby et al., 1986). Combining IR and PMW ISTs is thus a challenging task and requires a model or method to account for the large temperature gradients in the snow during winter (e.g. Comiso et al., 1989; Comiso et al., 2003).

Currently, Arctic surface temperatures are heavily under-sampled in terms of both satellite and *in situ* observations, causing large uncertainties in current gridded observational datasets, dynamical renanalyses as well as in global climate models (Cowtan & Way, 2014; Lenssen et al., 2019; Morice et al., 2012; Rapaić et al., 2015). Both satellite and *in situ* observations are subject to uncertainties and limitations, and there is a large potential in combining the available information. Currently no satellite-observed surface temperatures are included over the sea ice in reanalyses and global gridded observational datasets, and the T2m estimates are thus derived using only a very limited number of *in situ* observations. The results of Nielsen-Englyst et al. (2021b) (Paper VI) demonstrate a great potential of using L3 satellite IST to derive daily L3 T2m over the GrIS and the Arctic sea ice to supplement the existing *in situ* network and to improve the surface temperature estimation of the Arctic. Moreover, the method developed in Nielsen-Englyst et al. (2021b) (Paper VI) proved to be surprisingly good to derive L4 T2m estimates over sea ice, when applied to the L4 ISTs from Nielsen-Englyst et al. (2023b) (Paper IV). Nevertheless, there are some limitations of the methodology that should be mentioned. First of all, the regression model and coefficients are derived using an earlier version of the AASTI data (v1) than used in the L4 SST/IST CDR (AASTI v2), which nevertheless is based on the same satellite observations and algorithm. Secondly, the AASTI v1 data only covers the period 2000-2009, which means that the regression model is only trained on a limited portion of the CDR period, while being applied to both previous and later decades experiencing different conditions (e.g. on average colder (warmer) temperatures before (after) the AASTI v1 period). Another limitation is the sparse number of observations used to train the model in particular over sea ice. Repeating the analysis using AASTI v2, which covers 1982-2019, will increase the number of observations substantially and allow more robust IST-T2m relationships to be established. As already mentioned, future efforts should also be put into expanding the current *in situ* network in particular over sea ice, which would also increase the number of matchups available to determine the relationships.

Other studies have shown strong but varying IST-T2m dependencies on the wind speed for the GrIS and sea ice (e.g. Adolph et al., 2018; Nielsen-Englyst et al., 2019). However, in contrast to our expectations, including the wind speed (from ERA-I) as predictor in the regression model only showed limited improvements in the validation against *in situ* observations. As discussed in Nielsen-Englyst et al. (2021b) (Paper VI), this is likely explained by the quality of the ERA-I winds not being adequate for the relationship model. Accurately representing the Arctic surface winds is a challenging task due to the high resolution needed to represent the governing mechanisms (e.g. Steeneveld, 2014; Sterk et al., 2013). More accurate wind speeds, e.g. from the regional high resolution CARRA product (and the next generation CARRA2 with pan-Arctic coverage), may be able to improve the regression model and hence the satellite-derived T2m estimates in the future.

Both the combined sea and sea-ice surface temperatures (Nielsen-Englyst et al., 2023b) (Paper IV) and the corresponding regression-derived T2m (Nielsen-Englyst et al., 2023b; Nielsen-Englyst et al., 2021b) (Paper IV+VI) are valuable supplements to model-based air temperatures to be used for global surface temperature reconstructions, model evaluation or assimilation into ocean, sea-ice and atmospheric models. The potential of assimilating satellite-observed IST into a coupled ocean and sea-ice model was investigated in Rasmussen et al. (2018), and the experiments demonstrated an improved timing of the snow melt onset after assimilation. Tian et al. (2023) (Paper VII) provides an example of how the satellite-derived T2m can be used as benchmark when evaluating global reanalyses and climate models in the Arctic. It is evident that CMIP6 agrees very well with the satellite-

derived T2m compared to ERA5 (which is too warm) in the central Arctic Ocean, while CMIP6 (and ERA5 to a smaller degree) underestimate the surface temperatures in the marginal ice zone. As stated in Section 3.5, the colder temperatures of CMIP6 in the marginal ice zone is likely related to an underestimation of the fast decline in Arctic sea ice extent (Shu et al., 2020), which indicates that the climate models cannot fully capture the response of the Arctic sea ice to the changes in the external forcing (Shen et al., 2021).

Batrak and Müller (2019) uses the winter L3 ISTs from Nielsen-Englyst et al. (2023b) (Paper IV) as benchmark when evaluating several contemporary global reanalyses (i.e. ERA5, MERRA-2, JRA-55 and NCEP-2). All reanalyses show warm biases of 5–10°C in the central Arctic Ocean, when using the L3 IST as reference. The warm biases (and differences among the reanalyses) are attributed to a misrepresentation of the conductive heat flux through the sea ice and snow due to missing snow layer and misrepresentation of sea ice thickness (Batrak & Müller, 2019). The warm bias in ERA-5 and its predecessor, ERA-I, is also seen in Nielsen-Englyst et al. (2021b) (Paper VI, Figure 11) and for ERA-5 in Tian et al. (2023) (Paper VII) where the ERA T2m estimates are compared with the satellite-derived T2m. Recently, Zampieri et al. (2023) used the satellite-observed ISTs (L2 AASTI v1) to correct the winter clear-sky temperature bias over the Arctic sea ice in the ERA5 and JRA-55 reanalyses using a machine learning based correction, which learns from the satellite ISTs. The bias correction turned out to be efficient in correcting skin temperature bias affecting the current generation of atmospheric reanalyses and at the same time increasing the physical understanding of the bias and the ice system, when assessing the spatial and temporal evolution of the correction (Zampieri et al., 2023). These applications demonstrate that the use of the satellite-derived surface temperatures can be very informative and guide future development efforts to improve climate models as well as reanalyses in the Arctic.

The model deficiency of atmospheric reanalyses (such as ERA5) will likely be addressed in the future using fully coupled modelling systems and by assimilating new types of observations (Davy & Outten, 2020). CARRA already took the first step in this direction where the snow on top of sea ice is modeled more accurately. It is part of the CARRA project plan to use the satellite-derived T2m estimates from Nielsen-Englyst et al. (2023b) (Paper IV) for validation of the next generation of the Arctic regional reanalysis, CARRA2, with pan-Arctic coverage. It should also be noted that the satellite-derived T2m estimates are independent of NWP models, and are accompanied with validated well-characterized estimates of uncertainty, and thus, a combination of NWP models and the satellite-derived T2m could possible lead to even better T2m estimates in the future. The many applications of the satellite-derived IST and T2m estimates emphasize the value of having more complete and accurate absolute benchmark fields in the Arctic.

The combined Arctic L4 SST/IST CDR also presents a new possibility for a more consistent surface temperature monitoring compared to existing global gridded observational datasets that usually combine SST with T2m estimates over sea ice. Because of the rapid decline in sea ice extent some regions which were previously mapped as a T2m are now shifted to SST in these products. An artefact of the varying variable and technique is thereby introduced into the data record, which is clearly undesirable in the context of climate monitoring (IPCC, 2021). This in particular affects the analyses using anomalies in which a switch from T2m to SST causes a switch to a warmer climatology, resulting in an underestimation of the warming compared to analyses using absolute temperatures (IPCC, 2021; Richardson et al., 2018).

In addition to being more consistent, the Arctic L4 SST/IST also provides much more

detailed spatial information over the sea ice covered regions, compared to the existing global surface temperature datasets, which only rely on *in situ* observations in these regions. The increased information content facilitates studies on physical processes like sea-ice melt and refreeze, melt ponds, leads and associated sea-ice-air heat fluxes, as well as studies concerned with the regional patterns and trends in the observed warming.

Recent studies have found that the Arctic warming of about twice the global average, as previously reported by numerous studies (e.g. Graversen et al., 2008; Jansen et al., 2020; Screen & Simmonds, 2010; Walsh, 2014), is a significant underestimate of the Arctic warming e.g. the recent AMAP report found an Arctic temperature increase three times the global average (AMAP, 2021) and Rantanen et al. (2022) reported a warming rate of about four times the global average using several observational datasets (i.e. ERA5, BEST, HadCRUT5 and GISTEMP), but with large differences in the regional trends in the Arctic Ocean (Marquardt Collow et al., 2020; Rantanen et al., 2022). The varying spatial coverage and different ways to account for the incomplete sampling can likely explain part of the diversity in the estimates of the Arctic warming. Future work should include a comprehensive and detailed comparison and common validation of the satellite-derived estimates and the existing datasets in the Arctic to get a better understanding of the differences and to guide future developments.

The recent study by Zampieri et al. (2023) shows that applying the bias correction based on satellite-observed ISTs, leads to an increased warming trend (for the region north of 70°N) of +0.16°C decade^{-1} for ERA5. This is in good agreement with the findings in Tian et al. (2023) (Paper VII), where ERA5 also showed a weaker warming trend compared to the satellite-derived T2m (of –0.23°C decade^{-1}) over the central Arctic Ocean (DMI-SIC>70%). If the ERA5 modelled T2m is replaced by the satellite-derived T2m estimates over Arctic sea ice, the estimated Arctic warming increases from 0.79°C decade^{-1} to 0.89°C decade^{-1} and the corresponding Arctic amplification increases from 3.8 to 4.2, with the Arctic being defined as the area north of 66.5°N as in Rantanen et al. (2022). According to climate model simulations, incomplete data coverage suppresses the reported Arctic warming with about 16% when considering HadCRUT4 data availability (Richardson et al., 2018). This may explain part of the suppressed warming trend in ERA-5 when compared to the satellite-derived surface temperature trends as seen in Tian et al. (2023) (Paper VII) and Zampieri et al. (2023). This should be investigated in greater detail in the future. Nevertheless, these results indicate that the Arctic warming is even now underestimated in ERA-5, and they demonstrate the value of having spatially detailed satellite-derived surface temperatures of the Arctic sea ice. Including (or simply learning from) the satellite-derived surface temperatures may improve the estimation of the Arctic warming and thereby also improve the representation of the Arctic amplification in reanalyses, observational data sets and global climate models.

Future work could consider expanding the satellite-derived surface temperature estimates to global coverage, and possible including other large ice-surfaces such as the ice sheets e.g. as demonstrated for the GrIS in Karagali et al. (2022) (Paper X).

CHAPTER 4

Conclusion

This PhD study is concerned with many of the challenges that arise when estimating and monitoring the sea and sea-ice surface temperatures in the Arctic. These include the limitations related to a persistent cloud cover, the mix of open water and sea ice (and the temporally varying sea ice cover) and the sparse *in situ* network, which all complicate an accurate and absolute surface temperature mapping and monitoring of the Arctic oceans.

The persistent Arctic cloud cover limits the extent to which SST can be retrieved from thermal infrared satellite sensors. To improve the observational coverage of Arctic, this PhD study investigates the potential of retrieving SST from PMW satellite sensors, which are less impacted by clouds than the infrared sensors. Multiple PMW SST retrieval algorithms have been developed and evaluated, and the first European global Level 2 PMW SST CDR has been produced providing almost complete daily global coverage. To prepare for the future CIMR satellite mission, this PhD study also assesses the use of different radiometer frequency channels in the PMW SST retrievals, with promising results for the proposed CIMR constellation. Moreover, it has been investigated how to best combine infrared and PMW satellite SSTs in an Arctic gap-free (L4) analysis. The inclusion of PMW SSTs showed substantial improvements in the spatial observational coverage as well as in the Arctic SST estimates, and the results are expected to become even better in the future with the launch of AMSR3 and CIMR. CIMR will enable PMW SST retrievals at lower uncertainties and better spatial resolution than what is possible today, while the combination of AMSR3 and CIMR will provide unprecedented coverage and revisit time. The PMW SST algorithm development, radiometer channel assessment as well as the L4 integration assessment are all important steps in the preparation for future satellite missions, such as AMSR3 and CIMR, and will allow for improved monitoring of the Arctic.

Due to the varying sea ice coverage, the most consistent way to monitor the surface temperature change of the Arctic oceans is by combining SST and IST estimates. This PhD study presents the first infrared satellite-based Arctic (>58°N) L4 combined SST and IST CDR covering the period 1982-2021. The daily accurate and absolute L4 SST/IST CDR presents a new possibility for consistent surface temperature monitoring of day-to-day variations and well as long-term climate trends in the Arctic, and it provides much more detailed spatial information over sea ice, compared to the existing global surface temperature datasets. It shows that the combined sea and sea-ice surface temperature of the Arctic has increased with 4.5°C for the period 1982–2021, with a peak warming in the northeastern Barents Sea of around 10°C, which is likely associated with the fast decline

in sea ice cover in this region.

The satellite-observed ISTs provide an important supplement to the sparse *in situ* network due to the superior spatial coverage. To compensate for the lack of *in situ* measurements, efforts have been done to relate the satellite ISTs to near-surface air temperature (T2m), which is the variable measured *in situ* and usually used in temperature assessments of global reanalyses, *in situ* based observational datasets, climate models as well as for forcing field in sea ice and ocean models. The results show that it is possible to derive surprisingly good T2m estimates above sea ice based on satellite observations, with improved performance compared to ECMWF's most recent reanalysis (ERA5). The satellite-derived IST and T2m estimates provide an important supplement to the existing *in situ* network and it has a large potential to be used for assimilation, evaluating and improving global surface temperature reconstructions, atmospheric reanalyses as well as climate models. Initial efforts show that an evaluation against the satellite-derived surface temperatures can improve our physical understanding and guide future development efforts in global climate models as well as reanalyses in the Arctic.

This PhD study contributes to an overall more accurate and more consistent surface temperature mapping and long-term climate monitoring of the Arctic oceans, which enables a better estimation and representation of the Arctic warming and amplification.

4.1 Outlook

The knowledge and insights gained during this PhD study led to the discovery of new challenges, possibilities and ideas for future work and development in order to further improve the surface temperature estimation and monitoring of the Arctic. These were presented in Section 2.5 in relation to the PMW SST retrieval science, and in Section 3.6 in relation to a complete surface temperature estimation of the Arctic oceans. The main recommendations for future work and development are summarized here:

- Improving the observational *in situ* network in the Arctic

  Future work should focus on improving and expanding the existing (but sparse) *in situ* network in the Arctic. For example by establishing and maintaining fiducial reference measurements of IST and PMW SST. Efforts should be made to provide *in situ* observations in the marginal ice zone, including activities that provide estimates of the spatial variability and the radiometric signals such as flight campaigns, to quantify the uncertainties, assess the performance, and potentially improve the surface temperature estimates in this region.

- Improving and integrating surface temperature retrievals in the Arctic

  Future work should aim at improving the PMW SST estimates e.g. in proximity to sea ice and coasts. Future work could investigate the use of optimal estimation for simultaneously retrieval of sea and sea-ice parameters to improve the consistency and possibly enable PMW SST retrievals closer to the sea ice. To improve the surface temperature estimation of sea ice, the potential of using snow-ice-interface PMW ISTs to supplement the skin IR ISTs in a gap-free analysis, could also be investigated.

- Evaluating the impact of using the satellite-derived surface temperatures in the estimation of the Arctic warming and amplification

  Future work should include a comparison and common validation of the satellite-derived L4 SST/IST and SST/T2m data sets and global gridded observational

datasets, dynamical reanalyses and global climate models in the Arctic. In addition, the impact of assimilating or learning from the satellite-derived surface temperatures in the estimation of the Arctic warming and amplification could also be assessed. Finally, it could be considered to expand the satellite-derived Arctic L4 SST/IST and SST/T2m climate datasets to global coverage and possible other ice surfaces, such as the ice sheets.

Bibliography

Adolph, A. C., Albert, M. R., & Hall, D. K. (2018). Near-surface temperature inversion during summer at summit, greenland, and its relation to modis-derived surface temperatures. *The Cryosphere*, *12*(3), 907–920.

Alerskans, E., Høyer, J. L., Gentemann, C. L., Pedersen, L. T., Nielsen-Englyst, P., & Donlon, C. (2020). Construction of a climate data record of sea surface temperature from passive microwave measurements. *Remote Sensing of Environment*, *236*, 111485.

Alerskans, E., Zinck, A.-S. P., Nielsen-Englyst, P., & Høyer, J. L. (2022). Exploring machine learning techniques to retrieve sea surface temperatures from passive microwave measurements. *Remote Sensing of Environment*, *281*, 113220.

AMAP. (2021). Arctic climate change update 2021: Key trends and impacts. summary for policy-makers.

Atkinson, C. P., Rayner, N. A., Kennedy, J. J., & Good, S. A. (2014). An integrated database of ocean temperature and salinity observations. *Journal of Geophysical Research: Oceans*, *119*(10), 7139–7163.

Banzon, V., Smith, T. M., Steele, M., Huang, B., & Zhang, H.-M. (2020). Improved estimation of proxy sea surface temperature in the arctic. *Journal of Atmospheric and Oceanic Technology*, *37*(2), 341–349.

Barton, I. J. (1995). Satellite-derived sea surface temperatures: Current status. *Journal of Geophysical Research: Oceans*, *100*(C5), 8777–8790.

Barton, I., Minnett, P., Maillet, K., Donlon, C., Hook, S., Jessup, A., & Nightingale, T. (2004). The miami2001 infrared radiometer calibration and intercomparison. part ii: Shipboard results. *Journal of Atmospheric and Oceanic Technology*, *21*(2), 268–283.

Batrak, Y., & Müller, M. (2019). On the warm bias in atmospheric reanalyses induced by the missing snow over arctic sea-ice. *Nat. Commun.*, *10*(1), 4170.

Behrenfeld, M. J., O'Malley, R. T., Siegel, D. A., McClain, C. R., Sarmiento, J. L., Feldman, G. C., Milligan, A. J., Falkowski, P. G., Letelier, R. M., & Boss, E. S. (2006). Climate-driven trends in contemporary ocean productivity. *Nature*, *444*(7120), 752–755.

Bentamy, A., Piolle, J.-F., Grouazel, A., Danielson, R., Gulev, S., Paul, F., Azelmat, H., Mathieu, P., von Schuckmann, K., Sathyendranath, S., et al. (2017). Review and assessment of latent and sensible heat flux accuracy over the global oceans. *Remote Sensing of Environment*, *201*, 196–218.

Berry, D. I., & Kent, E. C. (2017). Assessing the health of the in situ global surface marine climate observing system. *International Journal of Climatology*, *37*(5), 2248–2259.

Bock, L., Lauer, A., Schlund, M., Barreiro, M., Bellouin, N., Jones, C., Meehl, G., Predoi, V., Roberts, M., & Eyring, V. (2020). Quantifying progress across different cmip phases with the esmvaltool. *125*(21), e2019JD032321.

Box, J. E., Colgan, W. T., Christensen, T. R., Schmidt, N. M., Lund, M., Parmentier, F.-J. W., Brown, R., Bhatt, U. S., Euskirchen, E. S., Romanovsky, V. E., et al. (2019). Key indicators of arctic climate change: 1971–2017. *Environmental Research Letters*, *14*(4), 045010.

Buffet, L., Gamet, P., Maisongrande, P., Salcedo, C., & Crebassol, P. (2021). The tir instrument on trishna satellite: A precursor of high resolution observation missions in the thermal infrared domain. *International Conference on Space Optics—ICSO 2020*, *11852*, 300–310.

Bulgin, C. E., Merchant, C. J., & Ferreira, D. (2020). Tendencies, variability and persistence of sea surface temperature anomalies. *Scientific reports*, *10*(1), 7986.

Carton, J. A., & Giese, B. S. (2008). A reanalysis of ocean climate using simple ocean data assimilation (soda). *Monthly weather review*, *136*(8), 2999–3017.

Carvalho, K., & Wang, S. (2020). Sea surface temperature variability in the arctic ocean and its marginal seas in a changing climate: Patterns and mechanisms. *Global and Planetary Change*, *193*, 103265.

Castro, S. L., Wick, G. A., & Emery, W. J. (2012). Evaluation of the relative performance of sea surface temperature measurements from different types of drifting and moored buoys using satellite-derived reference products. *Journal of Geophysical Research: Oceans*, *117*(C2).

Castro, S. L., Wick, G. A., Jackson, D. L., & Emery, W. J. (2008). Error characterization of infrared and microwave satellite sea surface temperature products for merging and analysis. *Journal of Geophysical Research: Oceans*, *113*(C3).

Castro, S. L., Wick, G. A., & Steele, M. (2016). Validation of satellite sea surface temperature analyses in the beaufort sea using uptempo buoys. *Remote Sensing of Environment*, *187*, 458–475.

Cavalieri, D. J., & Parkinson, C. L. (2012). Arctic sea ice variability and trends, 1979–2010. *The Cryosphere*, *6*(4), 881–889.

Centurioni, L. R., Turton, J., Lumpkin, R., Braasch, L., Brassington, G., Chao, Y., Charpentier, E., Chen, Z., Corlett, G., Dohan, K., et al. (2019). Global in situ observations of essential climate and ocean variables at the air–sea interface. *Frontiers in Marine Science*, *6*, 419.

Chang, P., Jelenak, Z., & Alsweiss, S. (2015). Algorithm theoretical basis document: Gcom-w1/amsr2 day-1 edr version 1.0.

Chelton, D. B., & Wentz, F. J. (2005). Global microwave satellite observations of sea surface temperature for numerical weather prediction and climate research. *Bulletin of the American Meteorological Society*, *86*(8), 1097–1116.

Chen, J.-L., Kang, S.-C., Meng, X.-H., & You, Q.-L. (2019). Assessments of the arctic amplification and the changes in the arctic sea surface. *Advances in Climate Change Research*, *10*(4), 193–202.

Comiso, J., Grenfell, T., Bell, D., Lange, M., & Ackley, S. (1989). Passive microwave in situ observations of winter weddell sea ice. *Journal of Geophysical Research: Oceans*, *94*(C8), 10891–10905.

Comiso, J. C. (2003). Warming trends in the arctic from clear sky satellite observations. *Journal of Climate*, *16*(21), 3498–3510.

Comiso, J. C., Cavalieri, D. J., & Markus, T. (2003). Sea ice concentration, ice temperature, and snow depth using amsr-e data. *IEEE Transactions on Geoscience and Remote Sensing*, *41*(2), 243–252.

Comiso, J. C., & Hall, D. K. (2014). Climate trends in the arctic as observed from space. *Wiley Interdisciplinary Reviews: Climate Change*, *5*(3), 389–409.

Cowtan, K., & Way, R. G. (2014). Coverage bias in the hadcrut4 temperature series and its impact on recent temperature trends. *Quarterly Journal of the Royal Meteorological Society*, *140*(683), 1935–1944.

Curry, J. A., Schramm, J. L., Rossow, W. B., & Randall, D. (1996). Overview of arctic cloud and radiation characteristics. *Journal of Climate*, *9*(8), 1731–1764.

Dash, P., Ignatov, A., Martin, M., Donlon, C., Brasnett, B., Reynolds, R. W., Banzon, V., Beggs, H., Cayula, J.-F., Chao, Y., et al. (2012). Group for high resolution sea surface temperature (ghrsst) analysis fields inter-comparisons—part 2: Near real time web-based level 4 sst quality monitor (l4-squam). *Deep Sea Research Part II: Topical Studies in Oceanography*, *77*, 31–43.

Davy, R., & Outten, S. (2020). The arctic surface climate in cmip6: Status and developments since cmip5. *Journal of Climate*, *33*(18), 8047–8068.

Dee, D. P., Uppala, S. M., Simmons, A., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M., Balsamo, G., Bauer, d. P., et al. (2011). The era-interim reanalysis: Configuration and performance of the data assimilation system. *Quarterly Journal of the royal meteorological society*, *137*(656), 553–597.

Dodd, E., Veal, K. L., Ghent, D., van den Broeke, M., & Remedios, J. J. (2019). Toward a combined surface temperature data set for the arctic from the along-track scanning radiometers. *Journal of Geophysical Research: Atmospheres*, *124*(13), 6718–6736.

Donlon, C. (2020). *Copernicus imaging microwave radiometer (cimr) mission requirements document, version 4* (ref. ESA-EOPSM-CIMR-MRD-3236). European Space Agency. Noordwijk, The Netherlands.

Donlon, C., Casey, K., Gentemann, C., LeBorgne, P., Robinson, I., Reynolds, R., Merchant, C., Llewellyn-Jones, D., Minnett, P., Piolle, J., et al. (2009). Successes and challenges for the modern sea surface temperature observing system. *Community White Paper for OceanObs*, *9*, 1–9.

Donlon, C., Robinson, I., Casey, K., Vazquez-Cuervo, J., Armstrong, E., Arino, O., Gentemann, C., May, D., LeBorgne, P., Piollé, J., et al. (2007). The global ocean data assimilation experiment high-resolution sea surface temperature pilot project. *Bulletin of the American Meteorological Society*, *88*(8), 1197–1214.

Donlon, C. J., Martin, M., Stark, J., Roberts-Jones, J., Fiedler, E., & Wimmer, W. (2012). The operational sea surface temperature and sea ice analysis (ostia) system. *Remote Sensing of Environment*, *116*, 140–158.

Dybkjær, G., Tonboe, R., & Høyer, J. (2012). Arctic surface temperatures from metop avhrr compared to in situ ocean and land data. *Ocean Science*, *8*(6), 959–970.

Eastman, R., & Warren, S. G. (2010). Interannual variations of arctic cloud types in relation to sea ice. *Journal of Climate*, *23*(15), 4216–4232.

Embury, O., Merchant, C. J., & Corlett, G. K. (2012). A reprocessing for climate of sea surface temperature from the along-track scanning radiometers: Initial validation, accounting for skin and diurnal variability effects. *Remote Sensing of Environment*, *116*, 62–78.

Emery, W. J., Yu, Y., Wick, G. A., Schluessel, P., & Reynolds, R. W. (1994). Correcting infrared satellite estimates of sea surface temperature for atmospheric water vapor attenuation. *Journal of Geophysical Research: Oceans*, *99*(C3), 5219–5236.

GCOS, G., World Meteorological Organization. (2022). The 2022 gcos implementation plan. technical report gcos-244.

Gelaro, R., McCarty, W., Suárez, M. J., Todling, R., Molod, A., Takacs, L., Randles, C. A., Darmenov, A., Bosilovich, M. G., Reichle, R., et al. (2017). The modern-era retrospective analysis for research and applications, version 2 (merra-2). *Journal of climate*, *30*(14), 5419–5454.

Gentemann, C. L. (2014). Three way validation of modis and amsr-e sea surface temperatures. *Journal of Geophysical Research: Oceans*, *119*(4), 2583–2598.

Gentemann, C. L., & Hilburn, K. A. (2015). In situ validation of sea surface temperatures from the gcom-w 1 amsr 2 rss calibrated brightness temperatures. *Journal of Geophysical Research: Oceans*, *120*(5), 3567–3585.

Gentemann, C. L., & Minnett, P. J. (2008). Radiometric measurements of ocean surface thermal variability. *Journal of Geophysical Research: Oceans*, *113*(C8).

Gentemann, C. L., Wentz, F. J., Brewer, M., Hilburn, K., & Smith, D. (2010). Passive microwave remote sensing of the ocean: An overview. *Oceanography from Space*, 13–33.

Gentemann, C., Scott, J. P., Mazzini, P. L., Pianca, C., Akella, S., Minnett, P. J., Cornillon, P., Fox-Kemper, B., Cetinić, I., Chin, T. M., et al. (2020). Saildrone: Adaptively sampling the marine environment. *Bulletin of the American Meteorological Society*, *101*(6), E744–E762.

Gladkova, I., Ignatov, A., Shahriar, F., Kihai, Y., Hillger, D., & Petrenko, B. (2016). Improved viirs and modis sst imagery. *Remote Sensing*, *8*(1), 79.

Good, S., Fiedler, E., Mao, C., Martin, M. J., Maycock, A., Reid, R., Roberts-Jones, J., Searle, T., Waters, J., While, J., et al. (2020). The current configuration of the ostia system for operational production of foundation sea surface temperature and ice concentration analyses. *Remote Sensing*, *12*(4), 720.

Good, S. A., Martin, M. J., & Rayner, N. A. (2013). En4: Quality controlled ocean temperature and salinity profiles and monthly objective analyses with uncertainty estimates. *Journal of Geophysical Research: Oceans*, *118*(12), 6704–6716.

Goosse, H., Kay, J. E., Armour, K. C., Bodas-Salcedo, A., Chepfer, H., Docquier, D., Jonko, A., Kushner, P. J., Lecomte, O., Massonnet, F., et al. (2018). Quantifying climate feedbacks in polar regions. *Nature communications*, *9*(1), 1919.

Graversen, R. G., Mauritsen, T., Tjernström, M., Källén, E., & Svensson, G. (2008). Vertical structure of recent arctic warming. *Nature*, *451*(7174), 53–56.

Hall, D. K., Key, J. R., Casey, K. A., Riggs, G. A., & Cavalieri, D. J. (2004). Sea ice surface temperature product from modis. *IEEE transactions on geoscience and remote sensing*, *42*(5), 1076–1087.

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., et al. (2020). The era5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, *146*(730), 1999–2049.

Holland, M. M., & Landrum, L. (2021). The emergence and transient nature of arctic amplification in coupled climate models. *Frontiers in Earth Science*, *9*, 719024.

Høyer, J. L., & Karagali, I. (2016). Sea surface temperature climate data record for the north sea and baltic sea. *Journal of Climate*, *29*(7), 2529–2541.

Høyer, J. L., Karagali, I., Dybkjær, G., & Tonboe, R. (2012). Multi sensor validation and error characteristics of arctic satellite sea surface temperature observations. *Remote Sensing of Environment*, *121*, 335–346.

Høyer, J. L., Le Borgne, P., & Eastwood, S. (2014). A bias correction method for arctic satellite sea surface temperature observations. *Remote Sensing of Environment*, *146*, 201–213.

Høyer, J. L., Tian, T., Alerskans, E., Nielsen-Englyst, P., Pedersen, L. T., Embacher, S., Block, T., & Gentemann, C. L. (2019). Quantification of atmospheric influence on passive microwave observations. https://eu-west-1.climate.esa.int/media/documents/SST_CCI-WP100-DMI-201-Issue-1-signed.pdf

IPCC. (2021). *Climate change 2021: The physical science basis. contribution of working group i to the sixth assessment report of the intergovernmental panel on climate*

change (V. Masson-Delmotte, P. Zhai, A. Pirani, S. Connors, C. Péan, S. Berger, N. Caud, Y. Chen, L. Goldfarb, M. Gomis, M. Huang, K. Leitzell, E. Lonnoy, J. Matthews, T. Maycock, T. Waterfield, O. Yelekçi, R. Yu, & B. Zhou, Eds.). Cambridge University Press, Cambridge, United Kingdom; New York, NY, USA, 2391 pp.

Isaksen, K., Nordli, Ø., Ivanov, B., Køltzow, M. A., Aaboe, S., Gjelten, H. M., Mezghani, A., Eastwood, S., Førland, E., Benestad, R. E., et al. (2022). Exceptional warming over the barents area. *Scientific reports*, *12*(1), 1–18.

Jansen, E., Christensen, J. H., Dokken, T., Nisancioglu, K. H., Vinther, B. M., Capron, E., Guo, C., Jensen, M. F., Langen, P. L., Pedersen, R. A., et al. (2020). Past perspectives on the present era of abrupt arctic climate change. *Nature Climate Change*, *10*(8), 714–721.

Jia, C., & Minnett, P. J. (2020). High latitude sea surface temperatures derived from modis infrared measurements. *Remote sensing of environment*, *251*, 112094.

Jia, C., Minnett, P. J., Szczodrak, M., & Izaguirre, M. (2022). High latitude sea surface skin temperatures derived from saildrone infrared measurements. *IEEE Transactions on Geoscience and Remote Sensing*.

Karagali, I., Barfod Suhr, M., Mottram, R., Nielsen-Englyst, P., Dybkjær, G., Ghent, D., & Høyer, J. L. (2022). A new level 4 multi-sensor ice surface temperature product for the greenland ice sheet. *The Cryosphere*, *16*(9), 3703–3721.

Kasahara, M., Kachi, M., Inaoka, K., Fujii, H., Kubota, T., Shimada, R., & Kojima, Y. (2020). Overview and current status of gosat-gw mission and amsr3 instrument. *Sensors, Systems, and Next-Generation Satellites XXIV*, *11530*, 1153007.

Kawanishi, T., Sezai, T., Ito, Y., Imaoka, K., Takeshima, T., Ishido, Y., Shibata, A., Miura, M., Inahata, H., & Spencer, R. (2003). The advanced microwave scanning radiometer for the earth observing system (amsr-e), nasda's contribution to the eos for global energy and water cycle studies. *IEEE Transactions on Geoscience and Remote Sensing*, *41*(2), 184–194.

Kennedy, J. J. (2014). A review of uncertainty in in situ measurements and data sets of sea surface temperature. *Reviews of Geophysics*, *52*(1), 1–32.

Kern, S., Lavergne, T., Notz, D., Pedersen, L. T., Tonboe, R. T., Saldo, R., & Sørensen, A. M. (2019). Satellite passive microwave sea-ice concentration data set intercomparison: Closed ice and ship-based observations. *The Cryosphere*, *13*(12), 3261–3307.

Kern, S., Lavergne, T., Pedersen, L. T., Tonboe, R. T., Bell, L., Meyer, M., & Zeigermann, L. (2022). Satellite passive microwave sea-ice concentration data set intercomparison using landsat data. *The Cryosphere*, *16*(1), 349–378.

Key, J., & Haefliger, M. (1992). Arctic ice surface temperature retrieval from avhrr thermal channels. *Journal of Geophysical Research: Atmospheres*, *97*(D5), 5885–5893.

Key, J. R., Collins, J. B., Fowler, C., & Stone, R. S. (1997). High-latitude surface temperature estimates from thermal satellite data. *Remote Sensing of Environment*, *61*(2), 302–309.

Key, J. R., Mahoney, R., Liu, Y., Romanov, P., Tschudi, M., Appel, I., Maslanik, J., Baldwin, D., Wang, X., & Meade, P. (2013). Snow and ice products from suomi npp viirs. *Journal of Geophysical Research: Atmospheres*, *118*(23), 12–816.

Kilic, L., Prigent, C., Aires, F., Boutin, J., Heygster, G., Tonboe, R. T., Roquet, H., Jimenez, C., & Donlon, C. (2018). Expected performances of the copernicus imaging microwave radiometer (cimr) for an all-weather and high spatial resolution estimation of ocean and sea ice parameters. *Journal of Geophysical Research: Oceans*, *123*(10), 7564–7580.

Kilic, L., Prigent, C., Boutin, J., Meissner, T., English, S., & Yueh, S. (2019a). Comparisons of ocean radiative transfer models with smap and amsr2 observations. *Journal of Geophysical Research: Oceans*, *124*(11), 7683–7699.

Kilic, L., Tonboe, R. T., Prigent, C., & Heygster, G. (2019b). Estimating the snow depth, the snow–ice interface temperature, and the effective temperature of arctic sea ice using advanced microwave scanning radiometer 2 and ice mass balance buoy data. *The Cryosphere*, *13*(4), 1283–1296.

Kilpatrick, K., Podesta, G., & Evans, R. (2001). Overview of the noaa/nasa advanced very high resolution radiometer pathfinder algorithm for sea surface temperature and associated matchup database. *Journal of Geophysical Research: Oceans*, *106*(C5), 9179–9197.

Kilpatrick, K. A., Podestá, G., Walsh, S., Williams, E., Halliwell, V., Szczodrak, M., Brown, O., Minnett, P., & Evans, R. (2015). A decade of sea surface temperature from modis. *Remote Sensing of Environment*, *165*, 27–41.

Kobayashi, S., Ota, Y., Harada, Y., Ebita, A., Moriya, M., Onoda, H., Onogi, K., Kamahori, H., Kobayashi, C., Endo, H., et al. (2015). The jra-55 reanalysis: General specifications and basic characteristics. *Journal of the Meteorological Society of Japan. Ser. II*, *93*(1), 5–48.

Koetz, B., Berger, M., Blommaert, J., Del Bello, U., Drusch, M., Duca, R., Gascon, F., Ghent, D., Hoogeveen, J., Hook, S., et al. (2019). Copernicus high spatio-temporal resolution land surface temperature mission: Mission requirements document. *ESA: Noordwijk, The Netherlands*.

Lavergne, T., Kern, S., Aaboe, S., Derby, L., Dybkjaer, G., Garric, G., Heil, P., Hendricks, S., Holfort, J., Howell, S., et al. (2022). A new structure for the sea ice essential climate variables of the global climate observing system. *Bulletin of the American Meteorological Society*, *103*(6), E1502–E1521.

Lavergne, T., Sørensen, A. M., Kern, S., Tonboe, R., Notz, D., Aaboe, S., Bell, L., Dybkjær, G., Eastwood, S., Gabarro, C., et al. (2019). Version 2 of the eumetsat osi saf and esa cci sea-ice concentration climate data records. *The Cryosphere*, *13*(1), 49–78.

Le Traon, P.-Y., Antoine, D., Bentamy, A., Bonekamp, H., Breivik, L., Chapron, B., Corlett, G., Dibarboure, G., DiGiacomo, P., Donlon, C., et al. (2015). Use of satellite observations for operational oceanography: Recent achievements and future prospects. *Journal of Operational Oceanography*, *8*(sup1), s12–s27.

Lenssen, N. J., Schmidt, G. A., Hansen, J. E., Menne, M. J., Persin, A., Ruedy, R., & Zyss, D. (2019). Improvements in the gistemp uncertainty model. *Journal of Geophysical Research: Atmospheres*, *124*(12), 6307–6326.

Lenton, T. M., Held, H., Kriegler, E., Hall, J. W., Lucht, W., Rahmstorf, S., & Schellnhuber, H. J. (2008). Tipping elements in the earth's climate system. *Proceedings of the national Academy of Sciences*, *105*(6), 1786–1793.

Lindsay, R., Wensnahan, M., Schweiger, A., & Zhang, J. (2014). Evaluation of seven different atmospheric reanalysis products in the arctic. *Journal of Climate*, *27*(7), 2588–2606.

Liu, Y., & Minnett, P. J. (2016). Sampling errors in satellite-derived infrared sea-surface temperatures. part i: Global and regional modis fields. *Remote sensing of environment*, *177*, 48–64.

Liu, Y., Dworak, R., & Key, J. (2018). Ice surface temperature retrieval from a single satellite imager band. *Remote Sensing*, *10*(12), 1909.

Liu, Y., Key, J., Tschudi, M., Dworak, R., Mahoney, R., & Baldwin, D. (2015). Validation of the suomi npp viirs ice surface temperature environmental data record. *Remote Sensing*, *7*(12), 17258–17271.

Marquardt Collow, A. B., Cullather, R. I., & Bosilovich, M. G. (2020). Recent arctic ocean surface air temperatures in atmospheric reanalyses and numerical simulations. *Journal of Climate*, *33*(10), 4347–4367.

Maslanik, J., FowLER, C., Key, J., Scambos, T., Hutchinson, T., & Emery, W. (1997). Avhrr-based polar pathfinder products for modeling applications. *Annals of Glaciology*, *25*, 388–392.

Maykut, G. A. (1986). *The surface heat and mass balance*. Springer.

McMillin, L. M. (1975). Estimation of sea surface temperatures from two infrared window measurements with different absorption. *Journal of geophysical research*, *80*(36), 5113–5117.

Meissner, T., & Manaster, A. (2021). Smap salinity retrievals near the sea-ice edge using multi-channel amsr2 brightness temperatures. *Remote sensing*, *13*(24), 5120.

Meissner, T., & Wentz, F. J. (2012). The emissivity of the ocean surface between 6 and 90 ghz over a large range of wind speeds and earth incidence angles. *IEEE Transactions on Geoscience and Remote Sensing*, *50*(8), 3004–3026.

Melsheimer, C., Heygster, G., Mathew, N., & Pedersen, L. T. (2009). Retrieval of sea ice emissivity and integrated retrieval of surface and atmospheric parameters over the arctic from amsr-e data. *Journal of the Remote Sensing Society of Japan*, *29*(1), 236–241.

Merchant, C. J., Embury, O., Bulgin, C. E., Block, T., Corlett, G. K., Fiedler, E., Good, S. A., Mittaz, J., Rayner, N. A., Berry, D., et al. (2019). Satellite-based time-series of sea-surface temperature since 1981 for climate applications. *Scientific data*, *6*(1), 1–18.

Merchant, C. J., Saux-Picart, S., & Waller, J. (2020). Bias correction and covariance parameters for optimal estimation by exploiting matched in-situ references. *Remote Sensing of Environment*, *237*, 111590.

Merchant, C., Harris, A., Maturi, E., & MacCallum, S. (2005). Probabilistic physically based cloud screening of satellite infrared imagery for operational sea surface temperature retrieval. *Quarterly Journal of the Royal Meteorological Society: A journal of the atmospheric sciences, applied meteorology and physical oceanography*, *131*(611), 2735–2755.

Merchant, C., Le Borgne, P., Marsouin, A., & Roquet, H. (2008). Optimal estimation of sea surface temperature from split-window observations. *Remote Sensing of Environment*, *112*(5), 2469–2484.

Merchant, C., Le Borgne, P., Roquet, H., & Marsouin, A. (2009). Sea surface temperature from a geostationary satellite by optimal estimation. *Remote Sensing of Environment*, *113*(2), 445–457.

Minnett, P. J. (2010). The validation of sea surface temperature retrievals from spaceborne infrared radiometers. *Oceanography from Space: Revisited*, 229–247.

Minnett, P. J., Kilpatrick, K. A., Podestá, G. P., Evans, R. H., Szczodrak, M. D., Izaguirre, M. A., Williams, E. J., Walsh, S., Reynolds, R. M., Bailey, S. W., et al. (2020). Skin sea-surface temperature from viirs on suomi-npp—nasa continuity retrievals. *Remote Sensing*, *12*(20), 3369.

Minnett, P. (1990). The regional optimization of infrared measurements of sea surface temperature from space. *Journal of Geophysical Research: Oceans*, *95*(C8), 13497–13510.

Minnett, P., Alvera-Azcárate, A., Chin, T., Corlett, G., Gentemann, C., Karagali, I., Li, X., Marsouin, A., Marullo, S., Maturi, E., et al. (2019). Half a century of satellite remote sensing of sea-surface temperature. *Remote Sensing of Environment*, *233*, 111366.

Minnett, P., & Kaiser-Weiss, A. (2012). Group for high resolution sea-surface temperature discussion document: Near-surface oceanic temperature gradients. https://www.ghrsst.org/wp-content/uploads/2021/04/SSTDefinitionsDiscussion.pdf

Morice, C. P., Kennedy, J. J., Rayner, N. A., & Jones, P. D. (2012). Quantifying uncertainties in global and regional temperature change using an ensemble of observational estimates: The hadcrut4 data set. *Journal of Geophysical Research: Atmospheres*, *117*(D8).

Morice, C. P., Kennedy, J. J., Rayner, N. A., Winn, J., Hogan, E., Killick, R., Dunn, R., Osborn, T., Jones, P., & Simpson, I. (2021). An updated assessment of near-surface temperature change from 1850: The hadcrut5 data set. *Journal of Geophysical Research: Atmospheres*, *126*(3), e2019JD032361.

Nielsen-Englyst, P., Høyer, J. L., Alerskans, E., Pedersen, L. T., & Donlon, C. (2021a). Impact of channel selection on sst retrievals from passive microwave observations. *Remote Sensing of Environment*, *254*, 112252.

Nielsen-Englyst, P., Høyer, J. L., Karagali, I., Kolbe, W. M., Tonboe, R. T., & Pedersen, L. T. (2023a). Impact of microwave observations on the estimation of arctic sea surface temperatures. *Preprint submitted to Remote Sensing of Environment*.

Nielsen-Englyst, P., Høyer, J. L., Kolbe, W. M., Dybkjær, G., Lavergne, T., Tonboe, R. T., Skarpalezos, S., & Karagali, I. (2023b). A combined sea and sea-ice surface temperature climate dataset of the arctic, 1982–2021. *Remote Sensing of Environment*, *284*, 113331.

Nielsen-Englyst, P., Høyer, J. L., Madsen, K. S., Tonboe, R., Dybkjær, G., & Alerskans, E. (2019). In situ observed relationships between snow and ice surface skin temperatures and 2 m air temperatures in the arctic. *The Cryosphere*, *13*(3), 1005–1024.

Nielsen-Englyst, P., Høyer, J. L., Madsen, K. S., Tonboe, R. T., Dybkjær, G., & Skarpalezos, S. (2021b). Deriving arctic 2 m air temperatures over snow and ice from satellite surface temperature measurements. *The Cryosphere*, *15*(7), 3035–3057.

Nielsen-Englyst, P., L Høyer, J., Toudal Pedersen, L., L Gentemann, C., Alerskans, E., Block, T., & Donlon, C. (2018). Optimal estimation of sea surface temperature from amsr-e. *Remote Sensing*, *10*(2), 229.

Notz, D., & Community, S. (2020). Arctic sea ice in cmip6. *Geophysical Research Letters*, *47*(10), e2019GL086749.

O'Carroll, A. G., Armstrong, E. M., Beggs, H. M., Bouali, M., Casey, K. S., Corlett, G. K., Dash, P., Donlon, C. J., Gentemann, C. L., Høyer, J. L., et al. (2019). Observational needs of sea surface temperature. *Frontiers in Marine Science*, *6*, 420.

O'Carroll, A. G., Eyre, J. R., & Saunders, R. W. (2008). Three-way error analysis between aatsr, amsr-e, and in situ sea surface temperature observations. *Journal of atmospheric and oceanic technology*, *25*(7), 1197–1207.

Ohring, G., Wielicki, B., Spencer, R., Emery, B., & Datla, R. (2005). Satellite instrument calibration for measuring global climate change: Report of a workshop. *Bulletin of the American Meteorological Society*, *86*(9), 1303–1314.

Oke, P. R., Brassington, G. B., Griffin, D. A., & Schiller, A. (2008). The bluelink ocean data assimilation system (bodas). *Ocean modelling*, *21*(1-2), 46–70.

Olmedo, E., Martínez, J., Turiel, A., Ballabrera-Poy, J., & Portabella, M. (2017). Debiased non-bayesian retrieval: A novel approach to smos sea surface salinity. *Remote Sensing of Environment*, *193*, 103–126.

Onarheim, I. H., & Årthun, M. (2017). Toward an ice-free barents sea. *Geophysical Research Letters*, *44*(16), 8387–8395.

Onarheim, I. H., Eldevik, T., Smedsrud, L. H., & Stroeve, J. C. (2018). Seasonal and regional manifestation of arctic sea ice loss. *Journal of Climate*, *31*(12), 4917–4932.

Overland, J., Dunlea, E., Box, J. E., Corell, R., Forsius, M., Kattsov, V., Olsen, M. S., Pawlak, J., Reiersen, L.-O., & Wang, M. (2019). The urgency of arctic change. *Polar Science*, *21*, 6–13.

Pearson, K., Good, S., Merchant, C. J., Prigent, C., Embury, O., & Donlon, C. (2019). Sea surface temperature in global analyses: Gains from the copernicus imaging microwave radiometer. *Remote Sensing*, *11*(20), 2362.

Pedersen, L. T. (1994). Merging microwave radiometer data and meteorological data for improved sea ice concentrations. *EARSeL Advances in Remote Sensing*, *3*(2-XII), 81–89.

Pithan, F., & Mauritsen, T. (2014). Arctic amplification dominated by temperature feedbacks in contemporary climate models. *Nature geoscience*, *7*(3), 181–184.

Ponsoni, L., Ribergaard, M. H., Nielsen-Englyst, P., Wulf, T., Buus-Hinkler, J., Kreiner, M. B., & Rasmussen, T. A. S. (2023). Greenlandic sea ice products with a focus on an updated operational forecast system. *Frontiers in Marine Science*, *10*, 979782.

Prigent, C., Aires, F., Bernardo, F., Orlhac, J.-C., Goutoule, J.-M., Roquet, H., & Donlon, C. (2013). Analysis of the potential and limitations of microwave radiometry for the retrieval of sea surface temperature: Definition of microwat, a new mission concept. *Journal of Geophysical Research: Oceans*, *118*(6), 3074–3086.

Rantanen, M., Karpechko, A. Y., Lipponen, A., Nordling, K., Hyvärinen, O., Ruosteenoja, K., Vihma, T., & Laaksonen, A. (2022). The arctic has warmed nearly four times faster than the globe since 1979. *Communications Earth & Environment*, *3*(1), 1–10.

Rapaić, M., Brown, R., Markovic, M., & Chaumont, D. (2015). An evaluation of temperature and precipitation surface-based and reanalysis datasets for the canadian arctic, 1950–2010. *Atmosphere-Ocean*, *53*(3), 283–303.

Rasmussen, T. A., Høyer, J. L., Ghent, D., Bulgin, C. E., Dybkjær, G., Ribergaard, M. H., Nielsen-Englyst, P., & Madsen, K. S. (2018). Impact of assimilation of sea-ice surface temperatures on a coupled ocean and sea-ice model. *Journal of Geophysical Research: Oceans*, *123*(4), 2440–2460.

Rayner, N., Parker, D. E., Horton, E., Folland, C. K., Alexander, L. V., Rowell, D., Kent, E. C., & Kaplan, A. (2003). Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century. *Journal of Geophysical Research: Atmospheres*, *108*(D14).

Rayner, N. A., Auchmann, R., Bessembinder, J., Brönnimann, S., Brugnara, Y., Capponi, F., Carrea, L., Dodd, E. M., Ghent, D., Good, E., et al. (2020). The eustace project: Delivering global, daily information on surface air temperature. *Bulletin of the American Meteorological Society*, *101*(11), E1924–E1947.

Rayner, N. A., Brohan, P., Parker, D., Folland, C. K., Kennedy, J. J., Vanicek, M., Ansell, T., & Tett, S. (2006). Improved analyses of changes and uncertainties in sea surface temperature measured in situ since the mid-nineteenth century: The hadsst2 dataset. *Journal of Climate*, *19*(3), 446–469.

Reynolds, R. W., Rayner, N. A., Smith, T. M., Stokes, D. C., & Wang, W. (2002). An improved in situ and satellite sst analysis for climate. *Journal of climate*, *15*(13), 1609–1625.

Reynolds, R. W., Smith, T. M., Liu, C., Chelton, D. B., Casey, K. S., & Schlax, M. G. (2007). Daily high-resolution-blended analyses for sea surface temperature. *Journal of climate*, *20*(22), 5473–5496.

Richardson, M., Cowtan, K., & Millar, R. J. (2018). Global temperature definition affects achievement of long-term climate goals. *Environmental Research Letters*, *13*(5), 054004.

Richter-Menge, J. A., Perovich, D. K., Elder, B. C., Claffey, K., Rigor, I., & Ortmeyer, M. (2006). Ice mass-balance buoys: A tool for measuring and attributing changes in the thickness of the arctic sea-ice cover. *Annals of Glaciology*, *44*, 205–210.

Roberts-Jones, J., Fiedler, E. K., & Martin, M. J. (2012). Daily, global, high-resolution sst and sea ice reanalysis for 1985–2007 using the ostia system. *Journal of Climate*, *25*(18), 6215–6232.

Rodgers, C. D. (2000). *Inverse methods for atmospheric sounding: Theory and practice* (Vol. 2). World scientific.

Roemmich, D., Johnson, G. C., Riser, S., Davis, R., Gilson, J., Owens, W. B., Garzoli, S. L., Schmid, C., & Ignaszewski, M. (2009). The argo program: Observing the global ocean with profiling floats. *Oceanography*, *22*(2), 34–43.

Rohde, R. A., & Hausfather, Z. (2020). The berkeley earth land/ocean temperature record. *Earth System Science Data*, *12*(4), 3469–3479.

Saunders, R., & Edwards, D. (1989). Atmospheric transmittances for the avhrr channels. *Applied optics*, *28*(19), 4154–4160.

Scarlat, R. C., Heygster, G., & Pedersen, L. T. (2017). Experiences with an optimal estimation algorithm for surface and atmospheric parameter retrieval from passive microwave data in the arctic. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *10*(9), 3934–3947.

Screen, J. A., & Simmonds, I. (2010). The central role of diminishing sea ice in recent arctic temperature amplification. *Nature*, *464*(7293), 1334–1337.

Serreze, M. C., & Meier, W. N. (2019). The arctic's sea ice cover: Trends, variability, predictability, and comparisons to the antarctic. *Annals of the New York Academy of Sciences*, *1436*(1), 36–53.

Shen, Z., Duan, A., Li, D., & Li, J. (2021). Assessment and ranking of climate models in arctic sea ice cover simulation: From cmip5 to cmip6. *Journal of Climate*, *34*(9), 3609–3627.

Shibata, A. (2006). Features of ocean microwave emission changed by wind at 6 ghz. *Journal of oceanography*, *62*(3), 321–330.

Shu, Q., Wang, Q., Song, Z., Qiao, F., Zhao, J., Chu, M., & Li, X. (2020). Assessment of sea ice extent in cmip6 with comparison to observations and cmip5. *Geophysical Research Letters*, *47*(9), e2020GL087965.

Shupe, M. D., Walden, V. P., Eloranta, E., Uttal, T., Campbell, J. R., Starkweather, S. M., & Shiobara, M. (2011). Clouds at arctic atmospheric observatories. part i: Occurrence and macrophysical properties. *Journal of Applied Meteorology and Climatology*, *50*(3), 626–644.

Simmons, A., Berrisford, P., Dee, D., Hersbach, H., Hirahara, S., & Thépaut, J.-N. (2017). A reassessment of temperature variations and trends from global reanalyses and monthly surface climatological datasets. *143*(702), 101–119.

Steeneveld, G.-J. (2014). Current challenges in understanding and forecasting stable boundary layers over land and ice. *Frontiers in Environmental Science*, *2*, 41.

Sterk, H., Steeneveld, G., & Holtslag, A. (2013). The role of snow-surface coupling, radiation, and turbulent mixing in modeling a stable boundary layer over arctic sea ice. *Journal of Geophysical Research: Atmospheres*, *118*(3), 1199–1217.

Stroeve, J., & Notz, D. (2018). Changing state of arctic sea ice across all seasons. *Environmental Research Letters*, *13*(10), 103001.

Theocharous, E., Fox, N. P., Barker-Snook, I., Niclòs, R., Santos, V. G., Minnett, P., Göttsche, F. M., Poutier, L., Morgan, N., Nightingale, T., et al. (2019). The 2016 ceos infrared radiometer comparison: Part ii: Laboratory comparison of radiation thermometers. *Journal of Atmospheric and Oceanic Technology*, *36*(6), 1079–1092.

Tian, T., Yang, S., Høyer, J., Nielsen-Englyst, P., & Singha, S. (2023). Concerns on benchmarking climate models in the arctic. *Preprint submitted to Nature Geoscience - Correspondence*.

Tonboe, R. T. (2010). The simulated sea ice thermal microwave emission at window and sounding frequencies. *Tellus A: Dynamic Meteorology and Oceanography*, *62*(3), 333–344.

Tonboe, R. T., Dybkjær, G., & Høyer, J. L. (2011). Simulations of the snow covered sea ice surface temperature and microwave effective temperature. *Tellus A: Dynamic Meteorology and Oceanography*.

Ulaby, F. T., Moore, R. K., & Fung, A. K. (1986). *Microwave remote sensing: Active and passive. volume 3- from theory to applications*.

Vazquez-Cuervo, J., Castro, S. L., Steele, M., Gentemann, C., Gomez-Valdes, J., & Tang, W. (2022). Comparison of ghrsst sst analysis in the arctic ocean and alaskan coastal waters using saildrones. *Remote Sensing*, *14*(3), 692.

Vázquez-Cuervo, J., Armstrong, E. M., & Harris, A. (2004). The effect of aerosols and clouds on the retrieval of infrared sea surface temperatures. *Journal of climate*, *17*(20), 3921–3933.

Vincent, R., Marsden, R., Minnett, P., & Buckley, J. (2008a). Arctic waters and marginal ice zones: 2. an investigation of arctic atmospheric infrared absorption for advanced very high resolution radiometer sea surface temperature estimates. *Journal of Geophysical Research: Oceans*, *113*(C8).

Vincent, R., Marsden, R., Minnett, P., Creber, K., & Buckley, J. (2008b). Arctic waters and marginal ice zones: A composite arctic sea surface temperature algorithm using satellite thermal data. *Journal of Geophysical Research: Oceans*, *113*(C4).

Vose, R. S., Arndt, D., Banzon, V. F., Easterling, D. R., Gleason, B., Huang, B., Kearns, E., Lawrimore, J. H., Menne, M. J., Peterson, T. C., et al. (2012). Noaa's merged land–ocean surface temperature analysis. *Bulletin of the American Meteorological Society*, *93*(11), 1677–1685.

Walsh, J. E. (2014). Intensified warming of the arctic: Causes and impacts on middle latitudes. *Global and Planetary Change*, *117*, 52–63.

Walton, C. C. (1988). Nonlinear multichannel algorithms for estimating sea surface temperature with avhrr satellite data. *Journal of Applied Meteorology and Climatology*, *27*(2), 115–124.

Wang, C., Graham, R. M., Wang, K., Gerland, S., & Granskog, M. A. (2019). Comparison of era5 and era-interim near-surface air temperature, snowfall and precipitation over arctic sea ice: Effects on sea ice thermodynamics and evolution. *The Cryosphere*, *13*(6), 1661–1679.

Wang, S., Foster, A., Lenz, E. A., Kessler, J. D., Stroeve, J. C., Anderson, L. O., Turetsky, M., Betts, R., Zou, S., Liu, W., et al. (2023). Mechanisms and impacts of earth system tipping elements. *Reviews of Geophysics*, *61*(1), e2021RG000757.

Wang, X., & Key, J. R. (2005). Arctic surface, cloud, and radiation properties based on the avhrr polar pathfinder dataset. part ii: Recent trends. *Journal of Climate*, *18*(14), 2575–2593.

Wentz, F. J., & Meissner, T. (2000). *Algorithm Theoretical Basis Document(ATBD): AMSR Ocean Algorithm (Version 2)* (RSS Tech. Proposal 121599A-1). Remote Sensing Systems. Santa Rosa, CA.

Wentz, F. J., Gentemann, C., Smith, D., & Chelton, D. (2000). Satellite measurements of sea surface temperature through clouds. *Science*, *288*(5467), 847–850.

Wentz, F. J., & Meissner, T. (2007). Supplement 1: Algorithm theoretical basis document for amsr-e ocean algorithms. *NASA: Santa Rosa, CA, USA*, *30*.

Wesslén, C., Tjernström, M., Bromwich, D., De Boer, G., Ekman, A. M., Bai, L.-S., & Wang, S.-H. (2014). The arctic summer atmosphere: An evaluation of reanalyses using ascos data. *Atmospheric Chemistry and Physics*, *14*(5), 2605–2624.

Wick, G. A., Emery, W. J., & Schluessel, P. (1992). A comprehensive comparison between satellite-measured skin and multichannel sea surface temperature. *Journal of Geophysical Research: Oceans*, *97*(C4), 5569–5595.

Woodruff, S. D., Worley, S. J., Lubker, S. J., Ji, Z., Eric Freeman, J., Berry, D. I., Brohan, P., Kent, E. C., Reynolds, R. W., Smith, S. R., et al. (2011). Icoads release 2.5: Extensions and enhancements to the surface marine meteorological archive. *International journal of climatology*, *31*(7), 951–967.

Yang, X., Schyberg, H., Palmason, B., Bojarova, J., Box, J., Pagh Nielsen, K., et al. (2020). C3s arctic regional reanalysis-full system documentation.

Zampieri, L., Arduini, G., Holland, M., Keeley, S., Mogensen, K., Shupe, M. D., & Tietsche, S. (2023). A machine learning correction model of the winter clear-sky temperature bias over the arctic sea ice in atmospheric reanalyses. *Monthly Weather Review*.

APPENDIX A

Paper I: Construction of a climate data record of sea surface temperature from passive microwave measurements

This appendix contains a full copy of Paper I:

- Alerskans, Emy, Jacob L. Høyer, Chelle L. Gentemann, Leif Toudal Pedersen, **Pia Nielsen-Englyst**, Craig Donlon. 2020. "Construction of a climate data record of sea surface temperature from passive microwave measurements", *Remote Sensing of Environment*, Volume 236, 111485, ISSN 0034-4257.

Remote Sensing of Environment 236 (2020) 111485

Contents lists available at ScienceDirect

Remote Sensing of Environment

journal homepage: www.elsevier.com/locate/rse

Construction of a climate data record of sea surface temperature from passive microwave measurements

Emy Alerskans[a,*], Jacob L. Høyer[a], Chelle L. Gentemann[b], Leif Toudal Pedersen[c], Pia Nielsen-Englyst[a,c], Craig Donlon[d]

[a] Danish Meteorological Institute, Lyngbyvej 100, DK-2100, Copenhagen Ø, Denmark
[b] Earth and Space Research, Seattle, WA, 98121, USA
[c] DTU-Space, Technical University of Denmark, DK-2800, Lyngby, Denmark
[d] European Space Agency/European Space Research and Technology Centre (ESA/ESTEC), 2201, AZ, Noordwijk, the Netherlands

ARTICLE INFO

Keywords:
Remote sensing
Microwave
AMSR-E
AMSR2
Sea surface temperature
Climate data record

ABSTRACT

A statistical regression-based microwave sea surface temperature (SST) retrieval algorithm has been developed within the European Space Agency Climate Change Initiative (ESA-CCI) SST project. The retrieval algorithm was used to generate a climate data record (CDR) of passive microwave (PMW) SST from the Advanced Microwave Scanning Radiometer – Earth Observing System (AMSR-E) and its follow-on instrument AMSR2 for the period June 2002–October 2017. Multisensor Matchup Datasets (MMDs), which includes satellite orbital data collocated with in situ and auxiliary data, was used to derive consistent algorithms for AMSR-E and AMSR2. The retrieval algorithms consist of wind speed (WS) retrievals and SST retrievals, with corresponding uncertainty retrievals. The WS retrieval consists of a two-step regression model, where the second step is a localized algorithm, trained to perform well over restricted WS intervals. A two-step multiple linear regression retrieval with localized algorithms is used to retrieve SST. The first-stage algorithm is trained to perform well over restricted latitude intervals for ascending and descending orbit, respectively, whereas the second-stage retrieval uses localized SST and WS algorithms. Furthermore, a new and effective method for detecting and screening for Radio Frequency Interference (RFI) was developed.

Validation of the PMW SSTs against drifter in situ SSTs shows an overall bias of −0.02 K for quality level (QL) 4 and 5 AMSR-E retrievals with a standard deviation of 0.46 K. Validation results for QL 4 and 5 AMSR2 retrievals against drifter in situ SSTs give a bias of 0.002 K and a standard deviation of 0.45 K. The corresponding mean modelled SST uncertainties, including in situ and sampling uncertainties, are estimated to 0.45 K for QL 4 and 5 AMSR-E retrievals and 0.44 K for QL 4 and 5 AMSR2 retrievals. Validation against near-surface temperature measurements from Argo floats yielded comparable results, confirming the drifting-buoy validation. The validation results demonstrate a consistent PMW SST CDR with accurate SST observations and reliable uncertainty estimates.

1. Introduction

Global Sea Surface Temperatures (SSTs) are important inputs for Numerical Weather Prediction (NWP) (Brasnett and Colan, 2016), coupled models (Liang et al., 2017), ocean models (Le Traon et al., 2015; Yang et al., 2015), SST analyses (Donlon et al., 2012), oceanographic research (Gentemann et al., 2003), air-sea interaction research (Monzikova et al., 2017; Ning et al., 2018), and of broad use in understanding changes to the marine and ecological environment (Chacko, 2017; Ishizaki, 2014). While infrared (IR) satellite SSTs have been available since 1981, the satellite microwave radiometer SST record began in 1997 with the Tropical Rainfall Measuring Mission (TRMM) Microwave Imager (TMI, (e.g. Wentz et al., 2000)) and then continued (with global data) with the launch of the Advanced Microwave Scanning Radiometer for EOS (AMSR-E, (e.g. Chelton and Wentz, 2005)) and other follow-on instruments. SSTs may be obtained from IR observations at a spatial resolution of typically 1–4 km with low uncertainty (e.g. Embury et al., 2012; Gladkova et al., 2016), but IR retrievals are not possible in the presence of clouds and may be compromised in the presence of atmospheric aerosols (Merchant et al., 2006, 1999) and are sensitive to the vertical distribution of water vapor (Le Borgne et al., 2011; Minnett, 1986). Furthermore, the near-IR

* Corresponding author.
E-mail address: ea@dmi.dk (E. Alerskans).

https://doi.org/10.1016/j.rse.2019.111485
Received 6 April 2019; Received in revised form 24 September 2019; Accepted 17 October 2019
Available online 19 November 2019
0034-4257/

channels used by e.g. MODIS to retrieve SST are susceptible to solar contamination (Brown and Minnett, 1999). Passive microwave (PMW) SSTs are at a lower resolution (approximately 50 km) but can be retrieved in all weather conditions except rain. PMW SST data are not used in the presence of Radio Frequency Interference (RFI) or typically within ~100 km from sea ice or land where the impact of side-lobe contamination in the vicinity of a typically large thermal contrast becomes a significant source of error. Another disadvantage of PMW SST retrievals is their dependence on wind speed (e.g. Gentemann et al., 2004). The influence of PMW SST on water vapor is more readily minimized than in the IR, PMW retrievals are designed to account for the influence of liquid water path, and retrievals are not affected by aerosols to which the PMW wavelengths are transparent (e.g. Ulaby et al., 1981). Assuming calibration stability over a mission lifetime, these important features of PMW SSTs mean they have great value for developing time series for climate applications and as a complement to traditional satellite IR SST climate data records (CDRs).

PMW SST retrieval algorithms have been developed by several groups and differ in their approaches. There are statistical regression algorithms, developed through comparisons between measured brightness temperatures and collocated in situ SSTs (Chang et al., 2015; Shibata, 2006), physically-based regression retrieval algorithms developed through radiative transfer modeling (Meissner and Wentz, 2012), and optimal estimation algorithms developed using an iterative forward-model, environmental data, and in situ observations (Nielsen-Englyst et al., 2018). The statistical and physically-based regression algorithms commonly use a multi-stage regression to account for non-linearities in the brightness temperatures' dependence on retrieved variables. The existing algorithms have typically been derived for one satellite at the time and for a limited time period or with a limited in situ observational database for the tuning and validation.

In this paper we describe a statistical regression-based SST retrieval algorithm, which has been developed within the European Space Agency Climate Change Initiative (ESA-CCI) SST project (Merchant et al., 2014) to generate a PMW SST CDR. The retrieval algorithm has been derived consistently for AMSR-E and AMSR2 using a large amount of in situ observations. All SST retrievals are accompanied by an uncertainty value and extensive validation has been performed for both the retrievals and the uncertainty estimates. The approach taken in this paper ensures a consistent and reliable multi-satellite SST CDR based only upon microwave observation that facilitates the uptake within the many applications of SSTs. The PMW SST CDR is available from the Centre for Environmental Data Archival (CEDA) at http://gws-access.ceda.ac.uk/public/esacci-sst/PMW2.0_release/AMSR/L2P/.

The paper is structured first with a description of the satellite, in situ and auxiliary data in Section 2, as well as the matchup database used for algorithm development and validation. Thereafter, the retrieval algorithms are described in Section 3. The validation results are presented in Section 4. Section 5 discusses the results and Section 6 contains conclusions and ideas for future work.

2. Data

2.1. In situ observations

In this study we use in situ measured sea surface temperature observations from the International Comprehensive Ocean-Atmosphere DataSet (ICOADS) version 2.5.1 (Woodruff et al., 2011), and measurements from the Met Office Hadley Center Ensembles dataset version 4.2.0 (EN4 (Good et al., 2013),). Measurements from drifting buoys are the main source of observations but measurements from Argo free-drifting profiling floats (Argo, 2018; Roemmich et al., 2009) have also been used. The drifting buoys measure the sea surface temperature at a depth of about 20 cm in calm water with an accuracy of approximately 0.2 °C (O'Carroll et al., 2012, 2008). The drifting buoy sea surface temperatures have been quality checked, as described in Atkinson et al. (2014). The quality flags are provided with the data. For Argo floats, the uppermost temperature measurement has been used, which is measured at a typical depth of about 5 m (Gille, 2008), with an accuracy of 0.002 °C (Abraham et al., 2013; Kennedy, 2014). The quality control of the Argo sea surface temperatures is described in (Good et al., 2013). Sea surface temperatures from drifting buoys and Argo floats have previously been used for algorithm development and validation studies (see e.g. Embury et al., 2012; Høyer et al., 2012; Merchant et al., 2012; Nielsen-Englyst et al., 2018; Udaya Bhaskar et al., 2009).

2.2. AMSR-E data

The Advanced Microwave Scanning Radiometer for EOS (AMSR-E) is a twelve-channel, six-frequency microwave radiometer supplied by the Japan Aerospace Exploration Agency (JAXA) for the National Aeronautics and Space Administration's (NASA's) Earth Observation System Aqua platform. It measures brightness temperatures at 6.9, 10.7, 18.7, 23.6, 36.5 and 89.0 GHz at both horizontal and vertical polarization and at an Earth incidence angle of approximately 55°. A 1.6 m offset parabolic reflector antenna collects the microwave radiation and results in spatial resolutions ranging from approximately 60 km at 6.9 GHz to approximately 5 km at 89.0 GHz. The 89.0 GHz channels are sampled every 5 km, all other channels every 10 km. The dataset used in the present study consists of spatially resampled brightness temperatures at the resolution of the 6.9 GHz channel (Ashcroft and Wentz, 2013) and covers the period June 2002 to October 4, 2011.

2.3. AMSR2 data

The Advanced Microwave Scanning Radiometer 2 (AMSR2) is an instrument on JAXA's Global Change Observation Mission 1st – Water (GCOM-W1) platform. GCOM-W1 was launched in May 2012 and began collecting data on July 4, 2012. AMSR2 measures brightness temperatures at 6.9, 7.3, 10.7, 18.7, 23.6, 36.5 and 89.0 GHz at both horizontal and vertical polarization and at an Earth incidence angle of approximately 55°. Note the addition of the 7.3 GHz channels relative to AMSR-E, added for RFI mitigation. AMSR2 uses a 2 m offset parabolic antenna to obtain a spatial resolution ranging from 4 km at 89.0 GHz to approximately 50 km at 6.9 and 7.3 GHz. The 89.0 GHz channels are sampled every 5 km, all other channels every 10 km. We use the *Dataset of Brightness Temperature Modified Using the Antenna Pattern Matching Technique* (Maeda et al., 2016) which contains similar spatially resampled brightness temperatures to the AMSR-E dataset.

2.4. Auxiliary data

Several different datasets have been included in the retrieval and analysis to improve and interpret the retrieved parameters. Information of Total Column Water Vapor (TCWV), Total Cloud Liquid Water (TCLW), surface winds and sea ice concentration has been used from the atmospheric reanalysis ERA-Interim (Dee et al., 2011), which has a spatial resolution of 79 km. An additional surface wind speed (WS) dataset was included from the Cross-Calibrated Multi-Platform (CCMP) gridded surface vector winds, which is a product that combines many different satellite and in situ observations (Atlas et al., 2011). The product used here is version 2.0, which has a spatial resolution of 0.25°.

2.5. Matchup database

2.5.1. ESA CCI multisensor matchup dataset

The calibration and validation of the retrieval algorithm have been performed using the Multisensor Matchup Datasets (MMDs), which have been developed within the ESA-CCI SST project and contain pairs of in situ and satellite observations that are within 20 km and 4 h from each other. Subregions of 21 by 21 AMSR-E and AMSR2 pixels have

been extracted around the central matchup positions (corresponding to approximately ± 100 km from the central matchup) and the auxiliary information have been extracted for the central points for a 5x5 subregion (corresponding to approximately ± 160 km from the central matchup). Separate MMDs have been created for the AMSR-E and the AMSR2 datasets, but they follow the same specifications in terms of data content and format. For more information on the generation of MMDs, see (Block et al., 2018; Nielsen-Englyst et al., 2018). The AMSR-E MMD includes matchups for the period June 2002–October 2011 and the AMSR2 MMD includes matchups for the period July 2012–December 2016.

2.5.2. Training and validation subsets

The performance of the retrieval algorithm is closely linked to the quality of the satellite observations and auxiliary data used. It is therefore essential to exclude erroneous matchups from the training data. The following paragraphs describe the quality control procedures implemented during algorithm development.

The quality control flags for the AMSR-E and AMSR2 data ("satellite scan quality" and "6–36 GHz channel quality") were used to check the quality of the satellite pixels. If either of these flags indicated bad data, the corresponding matchups were flagged as erroneous. In addition, matchups with brightness temperatures outside the accepted range (0–320 K) were flagged. For valid oceanographic retrievals, the v-pol brightness temperature should always be larger than the h-pol. Hence, to remove obviously bad observations, data were flagged if the difference between the h-pol and v-pol brightness temperature for the 18–36 GHz channels was negative. This check was only performed for the 18–36 GHz channels since they are the channels for which the atmospheric contribution is largest. Furthermore, to exclude low-quality brightness temperature observations, the spatial standard deviation of the 23 V and H and the 36 V and H brightness temperatures were calculated in the 21 by 21 pixel extract surrounding each pixel. Different approaches were used for AMSR-E and AMSR2 data. For AMSR-E matchups, data were flagged as unusable if the standard deviation of the above-mentioned channels exceeded 55, 35, 25 and 25 K, respectively. Due to differences in the distribution of brightness temperatures, the same limits were not applicable for AMSR2 matchups. Thus, to perform a comparable flagging for AMSR2, the 1% of AMSR2 matchups with the highest standard deviation were considered to be of low quality and flagged accordingly.

The quality of the in situ and auxiliary data was also considered. Matchups with an in situ quality control flag indicating bad data were flagged as erroneous. Moreover, matchups with an in situ or ERA-Interim SST less than −2 °C or greater than 34 °C were excluded. Similarly, matchups with ERA-Interim or CCMP WS greater than 20 ms^{-1} were also flagged. Together, the above-mentioned checks on the satellite, in situ and auxiliary data constitute quality control checks (denoted "Quality control checks" in Table 1) which ensure that the input data are of sufficient quality for the algorithm development and validation.

Further checks are necessary as both atmospheric and surface effects can contaminate the signal and lead to erroneous retrievals. Sea ice and land affect the retrieval due to antenna side-lobe contamination. The satellite land/ocean flag and ERA-Interim sea ice fraction was used to flag matchups. If land was detected within the 21x21 pixel extract surrounding each pixel, the matchup was flagged. Correspondingly matchups for which sea ice was detected within the 5x5 ERA-Interim pixel extract surrounding each pixel were flagged. Diurnal warming effects were accounted for by flagging daytime AMSR-E matchups with wind speeds less than 4 ms^{-1}. For AMSR2, the wind speed limit was increased to 6 ms^{-1} as inspection of the results revealed that diurnal warming effects were not completely removed when using the lower limit of 4 ms^{-1}. The lower limit of 4 ms^{-1} was kept for AMSR-E in order to retain as many valid matchups as possible. Furthermore, precipitating clouds compromise the PMW retrievals and these effects must therefore be excluded. To account for contamination due to rain, data were flagged if the 18 GHz v-pol brightness temperature exceeded 240 K. Sun glitter contamination was accounted for by flagging data with a sun glint angle of less than 25°. To avoid contamination due to RFI, two different approaches were used. For AMSR-E, ground-based and space-based RFI was masked out using Table 2 in Gentemann and Hilburn (2015) together with observation location and geostationary reflection longitude and latitude, respectively. AMSR2 has an additional channel at 7.3 GHz, which is specifically designed for detection of RFI. Both the 6.9 GHz and 7.3 GHz channels on AMSR2 suffer from RFI contamination, however, the geographical distributions are different, making it possible to use the two channels for RFI detection. Based on this, the absolute difference between the 6.9 and 7.3 GHz channels (both polarizations) was used to mask out RFI if the difference exceeded 3 K. However, the latter of these methods does not completely mask out all RFI and a large signal could still be seen around Ascension Island. To be sure to mask out all RFI-contaminated data, all matchups in the area around Ascension Island (24°W to 6°W longitude and 18°S to 2°S latitude) were masked out for the AMSR2 matchups. Additional steps to further remove RFI-contaminated data will be introduced in Section 3.4. As a last quality control, obviously erroneous in situ SSTs were excluded using a 3-σ filter, which removed all matchups that deviated more than three standard deviations from the mean difference between the ERA-Interim reanalysis and in situ SST differences.

Finally, to obtain a more latitudinally representative dataset, the number of matchups per degree of latitude was restricted. Individual limits were imposed for the different years due to the widely varying number of matchups amongst years. In addition, a limit on the number of matchups per year was imposed. Since each year consists of very different numbers of matchups, a limit of 1 million matchups per year was imposed to get a more temporally representative dataset. However, if the number of filtered matchups was less than 700,000 for a year, the even-out-data-by-year restriction was set to 70% of the total number of good matchups to ensure that data from years with fewer matchups are represented in all subset (see Fig. 1). The summary statistics for the

Table 1
The number of matchups after filtering and the percentage of matchups the filtering step removes. The statistics are shown for both the AMSR-E and AMSR2 datasets. The footnotes indicate to which subset the percentage of matchups removed is relative to; [1] to all MMD matchups ("No filter"); [2] to the number of matchups which passed the "Quality control checks" step; [3] to the number of matchups which passed the "All above checks" step; [4] to the number of matchups which passed the "All checks" step; and [5] to the number of matchups which passed the "Even-out-by-latitude" step.

	AMSR-E		AMSR2	
Filter	N	% removed	N	% removed
No filter	40,480,306	–	27,796,093	–
Quality control checks [1]	34,340,715	15.2	25,658,424	7.7
Rain [2]	34,088,030	0.7	25,114,008	2.1
Sun glint [2]	32,269,911	6.0	24,286,797	5.3
RFI [2]	31,832,433	7.3	25,295,077	1.4
Land [2]	31,285,364	8.9	22,790,471	11.2
Sea ice [2]	30,852,455	10.2	22,627,223	11.8
Diurnal warming [2]	30,034,321	12.5	18,975,850	26.0
All above checks [1]	22,319,646	44.9	14,821,726	47.5
3-σ filter on reanalysis and in situ *SST diff.* [3]	21,972,506	1.6	14,588,260	1.6
All checks [1]	21,972,506	45.7	14,588,260	47.5
Even-out-by-latitude [4]	16,322,299	22.9	9,880,565	32.3
Even-out-by-year [5]	8,460,292	50.0	4,918,012	50.2
Total [1]	8,460,292	79.1	4,918,012	82.3

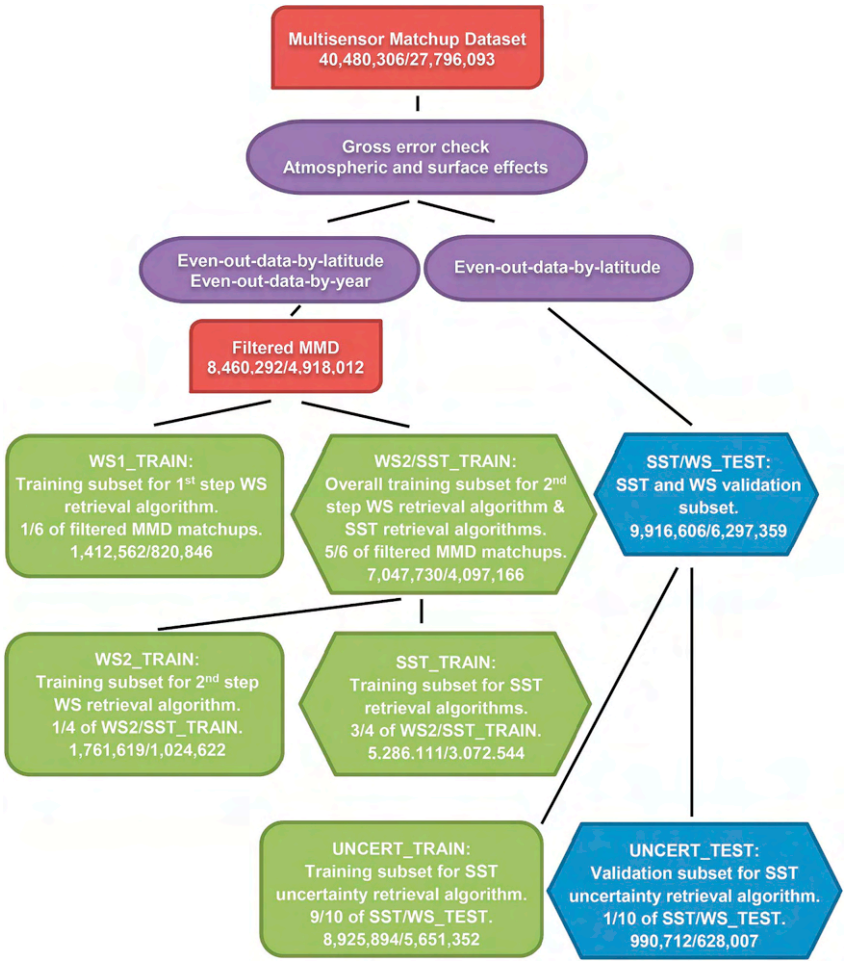


Fig. 1. Schematic of the subset division for the AMSR-E and AMSR2 MMDs. The numbers denote the total number of matchups in each AMSR-E/AMSR2 subset.

different filtering processes and their effect on the number of matchups are listed in Table 1. The quality control checks and the checks to account for atmospheric and surface effects are performed for all matchups. The latitudinal and temporal filtering, however, are only performed for selected subsets of the two MMDs (see Fig. 1 and description of the subset division).

To obtain an independent validation, the MMDs are divided into 7 subsets each (see Fig. 1);

- WS1_TRAIN: training subset for step 1 of the WS retrieval algorithm;
- WS2/SST_TRAIN: overall training subset for step 2 of the WS retrieval algorithm and the SST retrieval algorithm (divided into separate training subsets subsequently);
- WS2_TRAIN: training subset for step 2 of the WS retrieval algorithm;
- SST_TRAIN: training subset for the SST retrieval algorithms;
- SST/WS_TEST: combined validation subset for the SST retrieval algorithms and the WS retrieval algorithms;
- UNCERT_TRAIN: training subset for the SST uncertainty retrieval algorithm; and
- UNCERT_TEST: validation subset for the SST uncertainty retrieval algorithm.

Fig. 1 shows the filtering procedures and subsequent subset division of the AMSR-E and AMSR2 MMDs. The “Filtered MMD”-box represents the filtered MMDs where erroneous data have been excluded and the subsets have been made latitudinally and temporally representative. As shown in Fig. 1, the filtered MMD is then divided into two subsets; WS1_TRAIN and WS2/SST_TRAIN, through random selection of matchups. These subsets are used for training of the first step of the WS retrieval algorithm and for the second step of the WS algorithm and the SST algorithm, respectively (see Section 3.1 and 3.2). One sixth of the filtered MMD matchups are used for WS1_TRAIN and the rest for the combined training subset, WS2/SST_TRAIN. Following this, the WS2/SST_TRAIN subset is divided into two subsets; WS2_TRAIN (1/4 of WS2/SST_TRAIN) and SST_TRAIN (3/4 of WS2/SST_TRAIN). The division is again performed through random selection of matchups. The WS2_TRAIN subset is used to train the second step of the WS algorithm, whereas the SST_TRAIN subset is used for training of the SST retrieval algorithm. The fifth subset, SST/WS_TEST, is constructed from the matchups removed by the application of the even-out filters on the good data when constructing the WS and SST training subsets (see Fig. 1). Imposing a restriction on the number of matchups per degree of latitude for these data gives us the SST/WS_TEST subset, which is used as a validation subset for the retrieved SSTs and WSs. An upper limit on the number of matchups per year is not imposed for the SST/WS_TEST subset, since as many valid matchups as possible should be used for the validation. The total number of matchups in the AMSR-E and AMSR2 SST/WS_TEST subsets is 9,916,606 and 6,279,359, respectively. Finally, the SST/WS_TEST subset is divided, through random selection of matchups, into two subsets for training and validation of the SST uncertainty retrieval; UNCERT_TRAIN (9/10 of SST/WS_TEST) and UNCERT_TEST (1/10 of SST/WS_TEST).

The geographical distribution of matchups per square kilometer and the latitudinal distribution of drifter matchups for the combined AMSR-E and AMSR2 SST/WS_TEST subsets are shown in Fig. 2. The number of matchups per year is shown in Fig. 3. Fig. 2 demonstrates the spatial

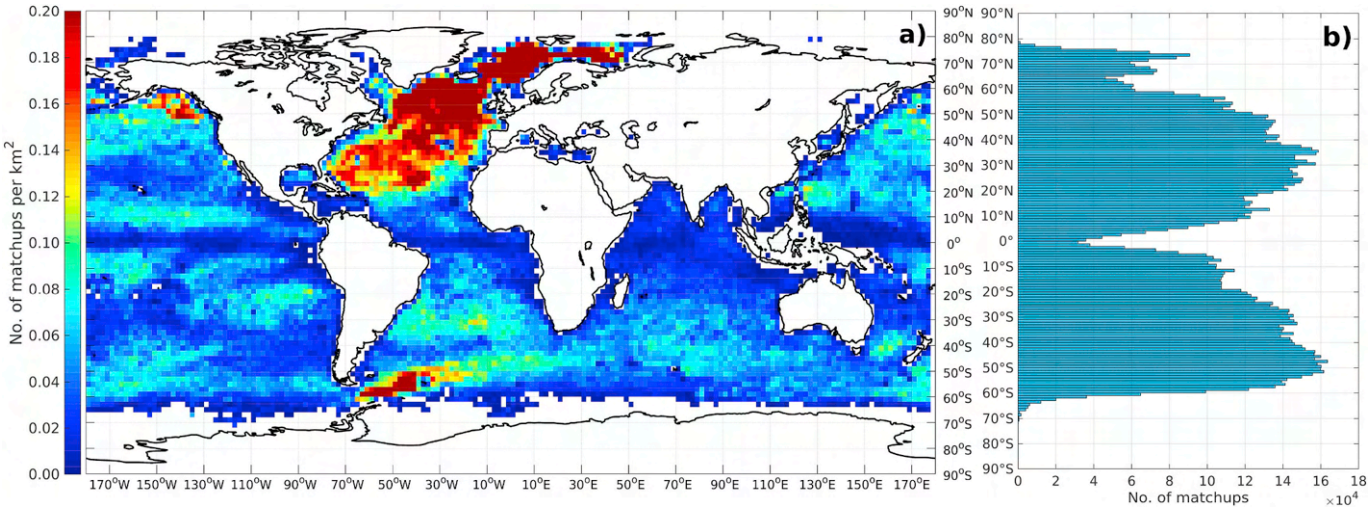


Fig. 2. a) Geographical distribution of drifter matchups per square kilometer; and b) latitudinal distribution of drifter matchups for the combined AMSR-E and AMSR2 SST/WS_TEST subsets. The geographical statistics have been gridded using a grid size of 2°, with a minimum of 50 matchups per grid cell.

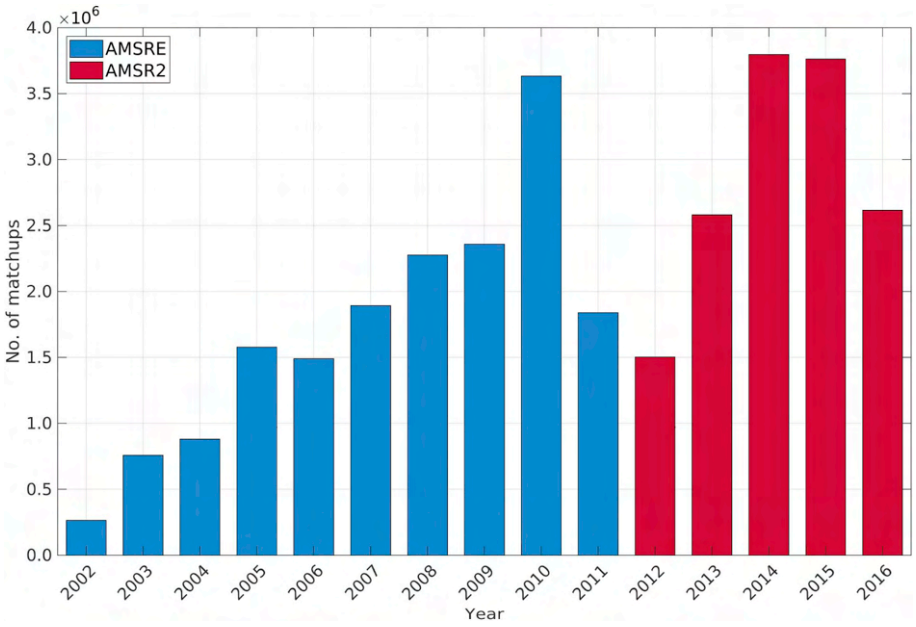


Fig. 3. Number of drifter matchups per year for the AMSR-E and AMSR2 SST/WS_TEST subsets.

distribution of the in situ matchup database, with the North Atlantic Ocean having the largest number of collocations and the Indian Ocean the least. The annual distribution of matchups (Fig. 3) shows few matchups in 2002 and another decrease in 2012. In 2002, AMSR-E data does not start until June, and the instrument failed in October 2011. In 2012, AMSR2 data was first available from July.

3. Algorithm development

The regression retrieval algorithms described here have been developed within the ESA-CCI SST project to retrieve subskin SST from AMSR-E and AMSR2 and to generate a climate data record of PMW SSTs. The MMDs used for algorithm development and validation only covers the period June 2002 to December 2016, whereas the PMW SST CDR covers an extended period, from June 2002 to December 2017. Two retrieval algorithms have been developed; one for wind speed (WS) and one for SST. A two-step multiple linear regression model is used to retrieve WS given satellite brightness temperatures, NWP reanalysis fields and CCMP data. In the first stage, a global algorithm is used, whereas in the second stage, localized algorithms, here for restricted wind speed intervals, are used. The SST retrieval algorithm described here is a two-step multiple linear regression model with localized retrieval algorithms. In the first stage, the algorithm is trained to perform well over restricted latitude intervals for ascending and descending orbits, respectively, whereas in the second stage, the applied algorithms are localized for restricted SST and WS intervals, using the first stage retrievals. The localized algorithms are used to address the non-linearity problem arising from the non-linear relationship between the geophysical parameters in question (SST or WS) and brightness temperature.

3.1. WS retrieval algorithm

A global regression model is used in the first stage of the wind speed retrieval process to retrieve an initial estimate of wind speed (WS_a). The definition of “global” is here taken to mean that only one set of regression coefficients is used for all wind speeds. The wind speed retrieval algorithm is inspired by the National Oceanic and Atmospheric

Administration (NOAA) AMSR2 wind speed retrieval algorithm (Chang et al., 2015) and expresses wind speed in terms of brightness temperature (T_B) and Earth incidence angle (θ_{EIA})

$$WS_a = a_0 + \sum_{i=1}^{10} (a_{1i} t_i + a_{2i} t_i^2) + a_3 \theta \quad (1)$$

where

$$t_i = T_{Bi} - 150 \quad \text{for all channels except the 23.6 GHz channels} \quad (2)$$

$$t_i = \ln(290 - T_{Bi}) \quad \text{for the 23.6 GHz channels} \quad (3)$$

$$\theta = \theta_{EIA} - 55 \quad (4)$$

The coefficients a_0, a_1, a_2 and a_3 are regression coefficients, denoted $\mathbf{B_{global}}$, the summation index i represents the summation over 10 brightness temperature channels; 6.9, 10.7, 18.7, 23.6 and 36.5 GHz (vertical and horizontal polarization), and T_{Bi} denotes the brightness temperature for the ith channel. To obtain a more consistent WS retrieval between AMSR-E and AMSR2, the AMSR2 7.3 GHz brightness temperature channel is not used in the retrieval. The regression coefficients are obtained through training on the WS1_TRAIN subset, using the least-squares method. Two sets of $\mathbf{B_{global}}$ regression coefficients are obtained; one for AMSR-E and one for AMSR2.

The relationship between brightness temperature and wind speed is non-linear and the first-stage retrieval is not able to represent these non-linearities. Hence, a second-stage retrieval needs to be performed where these non-linearities are taken into account. This is done by using localized wind speed retrieval algorithms. With "localized" we mean that the algorithm is trained to perform well over different subsets of a parameter, in this case over fixed wind speed reference intervals. Hence, one set of regression coefficients, $\mathbf{B_{WS}}$, is obtained for each reference interval, using the least-squares method. The coefficients are derived through training on subsets of the WS2_TRAIN subset, consisting of a minimum of 50 matchups each. For bins consisting of less than 50 matchups, the coefficients from the closest bin are used. To avoid discontinuities in the retrievals, the wind speed bins were defined with a 50% overlap during training. Two sets of $\mathbf{B_{WS}}$ regression coefficients are obtained; one for AMSR-E and one for AMSR2.

The localized wind speed algorithms are defined for fixed wind speeds in the interval 0 to 20 ms^{-1}, with a bin size of 1 ms^{-1}. The upper limit was chosen as 20 ms^{-1} to ensure enough matchups for a robust fit for the regression. This gives a total of 20 localized wind speed algorithms. When performing the retrieval, regression coefficients are selected from the correct wind speed bin based on the retrieved WS from the first-stage retrieval (WS_a). Like the first-stage retrieval, brightness temperature and incidence angle are used to retrieve the second-stage wind speed

$$WS_{rk} = b_{0k} + \sum_{i=1}^{10} (b_{1ik} t_i + b_{2ik} t_i^2) + b_{3k} \theta \quad (5)$$

where k denotes the reference wind speed bin, ranging from 0 to 20 ms^{-1}, and the coefficients b_0, b_1, b_2 and b_3 are regression coefficients, denoted $\mathbf{B_{WS}}$. The final retrieved wind speed, WS_r, is obtained through linear interpolation between the wind speed obtained for the current WS bin and for the closest neighboring WS bin. For wind speeds outside the defined bins the regression coefficient for the closest wind speed bin is used.

3.2. SST retrieval algorithm

In the first stage of the SST retrieval algorithm, a preliminary estimate of SST (SST_a) is retrieved using a regression model with localized latitude algorithms for ascending and descending orbits. This means that the algorithm is trained to perform well over fixed reference latitudes and for ascending and descending orbit, respectively. The algorithm is separately fit for latitude to account for the latitudinal dependence of the geographic distribution of SST. The partitioning of ascending and descending orbits is to account for the differing quality of daytime (ascending) and nighttime (descending) data. One set of regression coefficients, $\mathbf{B_{LAT,ORB}}$, is therefore obtained for each latitude and for ascending and descending orbits, respectively. The localized algorithms are derived for reference latitudes in the interval −72 to 82°, with a bin size of 2°, and descending (0) or ascending (1) orbit. This gives a total of 156 localized latitude and orbit algorithms. As with the localized WS algorithms, the latitude bins were defined with a 50% overlap during training to avoid discontinuities. For the retrieval process, regression coefficients are selected from the correct latitude and orbit bin using satellite latitude (φ_{LAT}) and orbit. The SST retrieval algorithm is inspired by the Remote Sensing System (RSS) AMSR-E SST retrieval algorithm (Wentz and Meissner, 2007). SST is expressed in terms of brightness temperature (T_B), Earth incidence angle (θ_{EIA}), wind speed (WS_r) and the relative angle between satellite azimuth angle and wind direction (ϕ_{REL})

$$SST_{alm} = c_{0lm} + \sum_{i=1}^{12} (c_{1ilm} t_i + c_{2ilm} t_i^2) + c_{3lm} \theta + c_{4lm} WS_r + \sum_{j=1}^{2} (c_{5jlm} \cos(j\phi_{REL}) + c_{6jlm} \sin(j\phi_{REL})) \quad (6)$$

where

$$t_i = T_{Bi} - 150 \quad \text{for all channels except the 23.6 GHz channels} \quad (7)$$

$$t_i = -\ln(290 - T_{Bi}) \quad \text{for the 23.6 GHz channels} \quad (8)$$

$$\theta = \theta_{EIA} - 55 \quad (9)$$

and l denotes the reference latitude, ranging from −72 to 82°, and m denotes the reference orbit (0 for descending and 1 for ascending orbit). The coefficients c_0, c_1, c_2, c_3, c_4, c_5 and c_6 are regression coefficients, denoted $\mathbf{B_{LAT,ORB}}$, the summation index i represents the summation over 12 brightness temperature channels; 6.9, 10.7, 18.7, 23.6, 36.5 and 89.0 GHz (each with dual polarization), and T_{Bi} denotes the brightness temperature for the ith channel. As with the WS retrieval, the AMSR2 7.3 GHz brightness temperature channel is not used in the SST retrieval (neither the preliminary nor the final retrieval step). This is to obtain a more consistent retrieval between AMSR-E and AMSR2. The regression coefficients are obtained through training on subsets of the SST_TRAIN subset, consisting of a minimum of 100 matchups each. The bins consisting of less than 100 matchups were assigned the coefficients from the closest bin. Two sets of $\mathbf{B_{LAT,ORB}}$ regression coefficients are obtained; one for AMSR-E and one for AMSR2. The least-squares method is used to derive the coefficients. To avoid discontinuities in the retrieval, the final preliminary retrieved SST is obtained through linear interpolation of the SST retrieved for the current latitude and orbit bin and the SST retrieved using the closest latitude and orbit bin.

Wind speed influences the sea surface roughness, which affects the emissivity and therefore also the brightness temperature (Hollinger, 1971; Meissner and Wentz, 2002). These effects are very complex and to account for this non-linear impact on the brightness temperature, a second-stage SST retrieval is performed. Here, a regression model with localized SST and WS algorithms is used to retrieve SST. The localized algorithms are defined for reference SSTs in the interval −2 to 34 °C, with a bin size of 2 °C, and reference wind speeds in the interval 0 to 20 ms^{-1}, with a bin size of 2 ms^{-1}. This gives a total of 209 localized SST and WS algorithms. The SST and WS bins were defined with a 1 ms^{-1} and 1 °C overlap, respectively, during training to avoid discontinuities in the retrieval. When performing the retrieval, the correct regression coefficients are found based on retrieved wind speed (WS_r) and first-stage retrieved SST (SST_a). The second-stage SST retrieval algorithm is formulated in the same way as the first-stage retrieval.

Brightness temperature, Earth incidence angle, retrieved wind speed and the relative angle between satellite azimuth angle and wind direction are used to retrieve SST

$$SST_{rnp} = d_{0np} + \sum_{i=1}^{12}(d_{1inp}t_i + d_{2inp}t_i^2) + d_{3np}\theta + d_{4np}WS_r + \sum_{j=1}^{2}(d_{5jnp}\cos(j\phi_{REL}) + d_{6jnp}\sin(j\phi_{REL})) \quad (10)$$

where n denotes the reference SST, ranging from -2 to 34 °C, and p denotes the reference wind speed, ranging from 0 to 20 ms^{-1}. The regression coefficients d_0, d_1, d_2, d_3, d_4, d_5 and d_6 are referred to as $\mathbf{B_{SST,WS}}$. The localized algorithms are trained on subsets of the SST_TRAIN subset, consisting of a minimum of 100 matchups each, and the coefficients are obtained using the least-squares method. For bins consisting of less than 100 matchups, the coefficients from the closest bin are used. Separate $\mathbf{B_{SST,WS}}$ regression coefficients are obtained for AMSR-E and AMSR2 due to training on different SST_TRAIN subsets. The final retrieved SST, SST_r, is obtained through bi-linear interpolation between the SST obtained for the current SST and WS bin and for the three closest neighboring bins.

3.3. SST uncertainty retrieval algorithm

An important part of a climate data record is the inclusion and validation of realistic uncertainties (Merchant et al., 2017). These facilitate the use of the SSTs within ocean and atmosphere models. In this study, we have chosen to follow the approach within the ESA-CCI SST project (Bulgin et al., 2016; Rayner et al., 2015), where the total uncertainty is a combination of three uncertainty components, each with different characteristics. The total uncertainty for the retrieved SST, ε_{SSTr}, is thus divided into three independent components; a random component (ε_{random}), a local systematic component (ε_{local}), and a global systematic component (ε_{global}). The total uncertainty is given by

$$\varepsilon_{SSTr} = \sqrt{\varepsilon_{random}^2 + \varepsilon_{local}^2 + \varepsilon_{global}^2} \quad (11)$$

These three components have been chosen due to their different behavior when aggregating the observations. When aggregating in time or space, it is assumed that ε_{random} will reduce as $1/\sqrt{N}$, where N is the number of observations. Similarly, ε_{local} will reduce as $1/\sqrt{N^*}$ where N^* is the effective number of observations, taking into account a synoptic timescale of 2–5 days and spatial distances of 500–1000 km. The ε_{global} component is assumed to be systematic and not to be reduced, even for large spatial and temporal averaging scales.

Both the local systematic uncertainty component and the random uncertainty component are obtained through the use of a global regression model. The global systematic uncertainty component, on the other hand, is assumed to be small and therefore set to zero. The variables in the algorithms for the local systematic and random uncertainty components have been selected through analysis of the validation results and express the uncertainty in terms of retrieved SST (SST_r), retrieved wind speed (WS_r), latitude (φ_{LAT}), and solar zenith angle (θ_{SZA})

$$\varepsilon_{SST} = e_0 + e_1 SST_r + e_2 SST_r^2 + e_3 WS_r + e_4 WS_r^2 + e_5\theta_{SZA} + e_6\theta_{SZA}^2 + \sum_{p=1}^{4}\left(e_{7p}\cos\left(\frac{\varphi_{LAT}}{p}\right) + e_{8p}\sin\left(\frac{\varphi_{LAT}}{p}\right)\right) \quad (12)$$

where the coefficients e_0, e_1, e_2, e_3, e_4, e_5, e_6, e_7 and e_8 are regression coefficients, determined through training on the UNCERT_TRAIN subset. Two sets of regression coefficients are obtained for AMSR-E and two for AMSR2; one for the local systematic uncertainty component ($\mathbf{B_{local}}$) and one for the random (uncorrelated) uncertainty component ($\mathbf{B_{rnd}}$).

The uncorrelated uncertainty, $\mathbf{B_{rnd}}$, is used to represent the uncertainty due to radiometric noise, which is represented by the noise equivalent differential temperature (NEdT). To estimate the uncertainty on the retrieved SST due to uncorrelated effects, an NEdT on the T_bs of 0.1 K (Wentz and Meissner, 2000) was propagated through the retrieval algorithm and a new set of SSTs was generated ($SST_{r,rnd}$). Hereafter, a pre-binning was performed using the UNCERT_TRAIN subset, where data was binned for retrieved SST, retrieved WS, latitude, and solar zenith angle. The intervals and bin sizes used for the pre-binning are shown in Table 2. For each bin consisting of more than 50 matchups, two standard deviation estimates were calculated;

Table 2
Pre-binning intervals and bin sizes for the uncertainty retrieval training dataset, UNCERT_TRAIN.

Variable	Bin size	Min	Max
Retrieved SST	2 °C	−1 °C	33 °C
Retrieved WS	2 ms^{-1}	1 ms^{-1}	19 ms^{-1}
Latitude	10°	−85°	85°
Solar zenith angle	15°	7.5°	172.5°

- $\sigma_{\Delta SSTr}$: the standard deviation of the SST_r minus in situ SST differences; and
- $\sigma_{\Delta SSTr,rnd}$: the standard deviation of the SST_r minus $SST_{r,rnd}$ differences.

The first standard deviation, $\sigma_{\Delta SSTr}$, is used to represent local effects on the total uncertainty and includes drifter uncertainty and sampling effects, whereas the second standard deviation, $\sigma_{\Delta SSTr,rnd}$, is used to represent random and uncorrelated effects. To obtain regression coefficients for the random uncertainty component, the retrieval algorithm was trained against $\sigma_{\Delta SSTr,rnd}$. For the local systematic uncertainty component, the retrieval algorithm was trained against the part of the uncertainty attributed only to local variations, σ_{local}. Hence both sampling effects and drifter uncertainty needed to be removed from $\sigma_{\Delta SSTr}$, in order to only get the local variations. The drifter uncertainty was set to 0.2 K whereas the sampling effect was calculated as a function of latitude. The sampling effect in this context was assumed to be primarily spatial (Høyer et al., 2012) and has been estimated by calculating the pixel-to-footprint variability for one year of GHRSST Level 4 DMI_OI Global Foundation Sea Surface Temperature Analysis with a spatial resolution of 0.05° (see e.g. Dash et al., 2012; Høyer et al., 2014; Høyer and She, 2007).

3.4. Improved RFI mask

Radio Frequency Interference (RFI) is a rapidly increasing problem for geophysical SST and WS retrievals using 6.9, 10.7 and 18.7 GHz. While these are protected frequencies for scientific use (International Telecommunication Union, 2012), the bandwidth of these channels exceed the protected bands. RFI can be divided into three categories (in order of occurrence); space-based ocean-reflected, satellite-to-satellite interference and ground-based RFI. The main source of surface reflected space-based RFI is media broadcasts (TV and radio) from geostationary satellites and satellite downlinks, which affect mostly the 6.9, 10.7 and 18.7 GHz channels. Satellite-to-satellite RFI is a growing problem with an unclear solution. Ground-based RFI is related to land-based microwave link communication systems on oil rigs, near cities, and in regions with military activities and mostly affects the 6.9 GHz channel.

Gentemann and Hilburn (2015) developed an RFI mask which uses observation location and geostationary reflection latitude and longitude to check for RFI contamination. This method was used as an initial RFI mask to screen for potentially contaminated RFI matchups in the data filtering process (Section 2.5.2). Here we present an alternative RFI filtering method, based on two additional SST retrieval algorithms. The

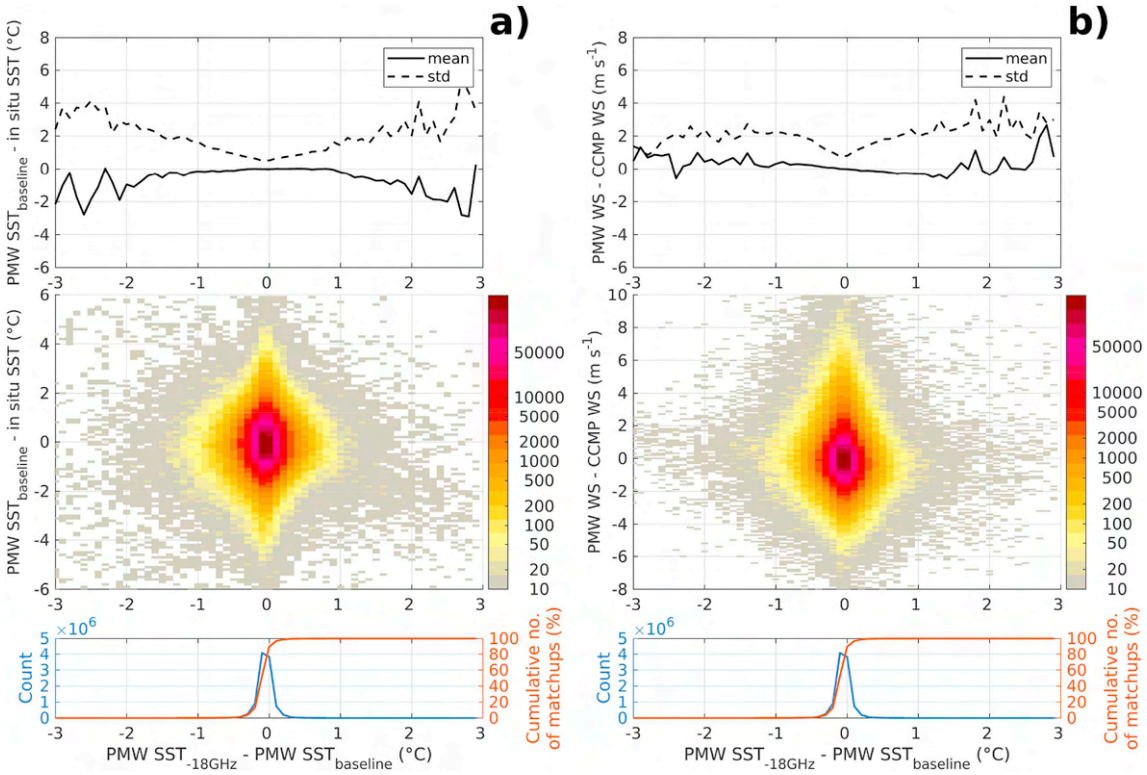


Fig. 4. a) Baseline retrieved AMSR-E PMW SST minus in situ SST as a function of the difference between the −18GHz retrieved AMSR-E PMW SST and the baseline retrieved AMSR-E PMW SST; and b) retrieved AMSR-E PMW WS minus CCMP WS as a function of the difference between the −18GHz retrieved AMSR-E PMW SST and the baseline retrieved AMSR-E PMW SST. Top panel shows the mean (solid) and standard deviation (dashed) of differences for each bin. The heat plot in the mid panel shows the distribution of matchups, and the bottom panel shows the number of matchups (blue) and the cumulative number of matchups (orange) in each bin. A minimum of 10 matchups per bin was used for the statistics calculation. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

two new algorithms are formulated exactly as the baseline algorithm with the exception that one excludes the 10 GHz channels (−10 GHz algorithm) and the other excludes the 18 GHz channels (−18 GHz algorithm). As for the baseline retrieval algorithm, the two-step regression model is used to retrieve SST for the −10GHz and −18GHz algorithms. Even though the 6.9 GHz channels are the ones that are most affected by RFI, we do not exclude them because this could introduce a strong SST dependency in the filter due to the large temperature variation in the sensitivity of the 10 GHz channel (Gentemann et al., 2010). Fig. 4a) shows the performance of the baseline retrieved AMSR-E PMW SST as a function of the difference between the AMSR-E PMW SST retrieved using the −18GHz algorithm and the baseline retrieved AMSR-E PMW SST. Similarly, Fig. 4b) shows the performance of the retrieved PMW WS versus CCMP WS against the difference between the AMSR-E PMW SST retrieved using the −18GHz algorithm and the baseline retrieved AMSR-E PMW SST.

The standard deviation of baseline retrieved PMW SST minus in situ SST increases as the difference between the −18GHz retrieved PMW SST and the baseline retrieved PMW SST grows, with the largest values found at the tails of the distribution. Furthermore, the magnitude of the mean also increases with increasing difference between the −18GHz retrieved PMW SST and the baseline retrieved PMW SST. The behavior of the dependence of the SST difference against −10GHz retrieved PMW SST minus baseline retrieved PMW SST is similar and is therefore not shown. Similarly, the results for the AMSR2 retrievals resemble those for the AMSR-E retrievals and are therefore not included here. Fig. 4b) shows a similar behavior for the mean and standard deviation with increased (absolute) values with increased difference between the −18GHz retrieved PMW SST and the baseline retrieved PMW SST. The results for the WS difference against the −10GHz retrieved PMW SST difference and the results for AMSR2 resemble the results shown in Fig. 4b) and are therefore not included.

Based on these results indicating that the performance of the SST retrieval is linked with the differences between the baseline SST retrieval and the −10 GHz and −18GHz, respectively, a new combined RFI and quality control mask is proposed. The new mask uses a 3-σ filter to detect RFI and erroneous data and flag if any of the following expressions are true

$$|(SST_{r,baseline} - SST_{r,-10GHz}) - \mu_{-10GHz}| > 3\sigma_{-10GHz} \quad (13)$$

$$|(SST_{r,baseline} - SST_{r,-18GHz}) - \mu_{-18GHz}| > 3\sigma_{-18GHz} \quad (14)$$

where $SST_{r,-10GHz}$, $SST_{r,-18GHz}$ and $SST_{r,\ baseline}$ are the final retrieved SST using the −10GHz, −18 GHz and baseline algorithms, respectively. μ_{-10GHz} and μ_{-18GHz} denote the mean of the difference $SST_{r,-10GHz} - SST_{r,baseline}$ and $SST_{r,-18GHz} - SST_{r,basesline}$, respectively, whereas σ_{-10GHz} and σ_{-18GHz} denote the standard deviation of the corresponding differences. Hereafter, when referring to the baseline retrieved PMW SST we will drop the reference to the retrieval algorithm and simply write PMW SST. The mean and standard deviation of differences used for the proposed RFI mask are shown in Table 3.

Fig. 5a) and b) show the geographical distribution of the gridded mean and standard deviation of the combined AMSR-E and AMSR2

Table 3
Mean and standard deviation of differences for retrieved SSTs using the −10GHz and −18GHz algorithm minus baseline retrieved SST for AMSR-E and AMSR2.

Sensor	μ_{-10GHz} (K)	μ_{-18GHz} (K)	σ_{-10GHz} (K)	σ_{-18GHz} (K)
AMSR-E	0.0024	0.0071	0.192	0.138
AMSR2	−0.0087	0.0043	0.170	0.130

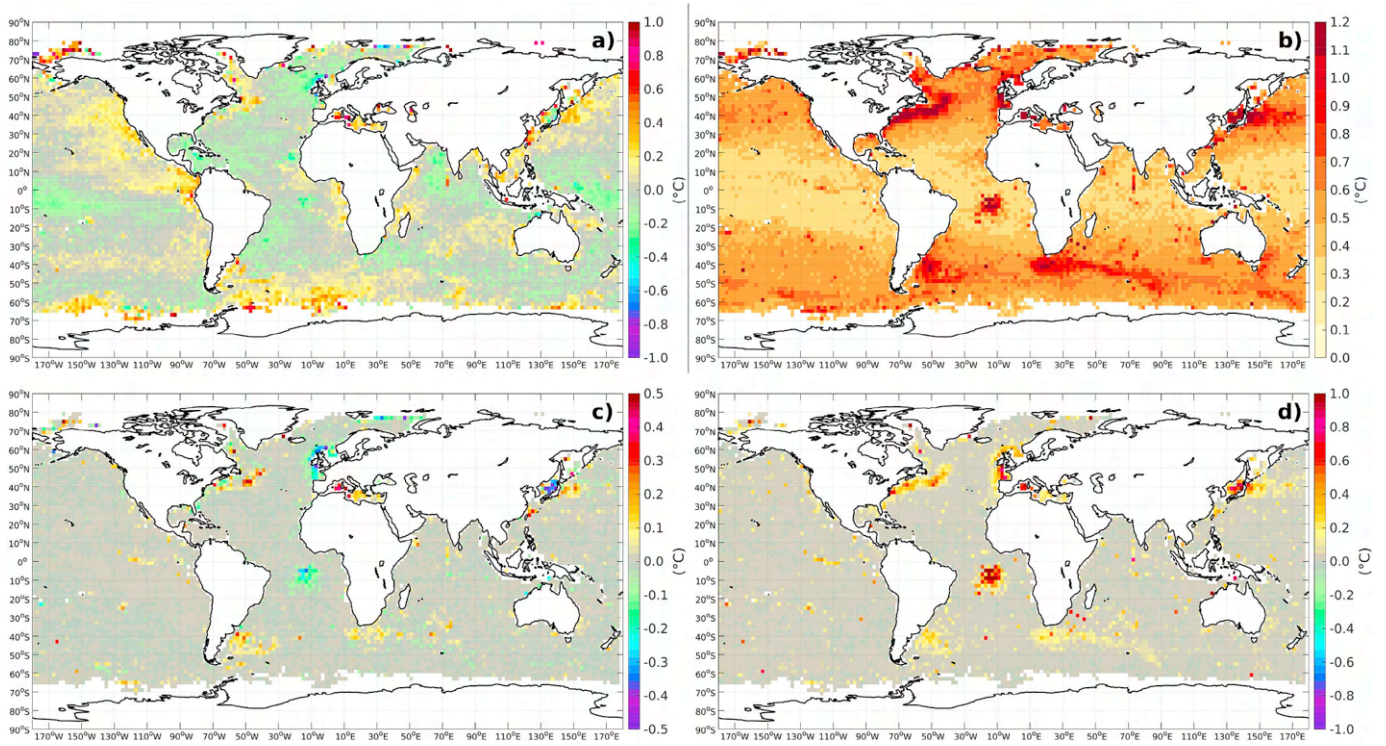


Fig. 5. Top panel shows the geographical distribution of a) mean and b) standard deviation of the SST bias (AMSR-E and AMSR2 PMW SST minus in situ SST) without application of an RFI mask. Bottom panel shows the distribution of c) the difference between the mean of the SST bias (AMSR-E and AMSR2 PMW SST minus in situ SST) without RFI mask minus the mean the SST bias with the proposed RFI mask; and d) the difference between the standard deviation of the SST bias (AMSR-E and AMSR2 PMW SST minus in situ SST) without RFI mask minus the standard deviation of the SST bias with the proposed RFI mask. The statistics have been gridded using a grid size of 2°, with a minimum of 50 matchups per grid cell.

Table 4
Mean and standard deviation of PMW SST minus in situ SST, number of matchups removed by the RFI mask and the total number of matchups left in the drifter validation subset, SST/WS_TEST, after all checks have been performed (see Section 2.5.2). Additional filtering was performed after application of the RFI mask to exclude obvious erroneous retrievals. Retrievals with a PMW SST outside the accepted range (−2 to 34 °C) or with a PMW WS outside the accepted range (0 to 20 ms^{-1}) were therefore excluded.

Sensor	RFI mask	Mean (°C)	Std (°C)	No. of matchups removed by RFI filter	No. of matchups
AMSR-E	No RFI mask	−0.012	0.54	–	10,067,979
	3-σ filter	−0.0099	0.52	154,344	9,916,606
AMSR2	No RFI mask	0.0027	0.54	–	6,384,638
	3-σ filter	0.0067	0.51	107,096	6,297,359

PMW SST minus in situ SST without application of an RFI mask. The region with a strong negative bias and a high standard deviation over Western Europe is a well-known region with large RFI contamination, both from space-based and ground-based RFI sources. The effect of RFI in this region here results in colder PMW SSTs compared to in situ SSTs. Ascension Island is another well-known region which is strongly influence by ground-based RFI, as is most evident in the increased standard deviation but also in the slightly negative mean. Yet another region which is well-known for being contaminated by RFI is the Mediterranean Sea, where the RFI contamination here results in warmer PMW SSTs compared to in situ SSTs. The application of the proposed RFI mask successfully removes RFI in all of the above-mentioned regions, as is evident in Fig. 5c) and d). In addition to removing RFI-contaminated data, the combined RFI and quality control mask also removes those matchups that deviates significantly (more than three standard deviations) from the mean. For this reasons, matchups, which most likely are not RFI-contaminated but otherwise possibly erroneous, are also removed from e.g. the dynamical ocean regions (such as the Gulf Stream, the Kuroshio and the Agulhas current). For comparison, Table 4 shows the performance of the SST retrieval algorithm with and without application of the new RFI mask. Application of the proposed RFI mask results in a decrease of the global standard deviation of the PMW SST minus in situ SST difference by 0.02 and 0.03 K for AMSR-E and AMSR2, respectively, compared to without application of an RFI mask. Hereafter, the new RFI mask has been used instead of the initial mask.

4. Results

The filtering steps described in Section 2.5.2 and the RFI mask described in Section 3.4 only ensures that the input data to the retrieval algorithm is of sufficient quality. It is also important to ensure that the retrievals used in the validation are of sufficient quality and to flag erroneous output. Therefore, additional filtering was performed where PMW SST retrievals outside the accepted range (−2 to 34 °C) and PMW WS retrievals outside the accepted range (0 to 20 ms^{-1}) were flagged as erroneous. The results shown in this section include the filtering steps outlined in Section 2.5.2, but with the new RFI mask instead of the initial one, as well as the checks on the retrieval output as outlined above. For the generation of the climate data record, the PMW SST retrievals follow the GHRSST GDS 2.0 data specification (GHRSST Science Team, 2010) for L2P and each retrieval has been assigned a

Table 5
Definition of quality levels and rules for assignment.

Quality Level (QL)	Definition	Rules
0	No data	
1	Bad data	Quality controls and various checks for atmospheric and surface effects (see Section 2.5.2)
2	Worst-quality useable data	• SST uncertainty ≥ 1 • Proximity to sea ice • Proximity to land
3	Low quality	0.5 < SST uncertainty < 1
4	Acceptable quality	0.35 < SST uncertainty ≤ 0.5
5	Best quality	SST uncertainty ≤ 0.35

quality level (QL) from 0 (no data) to 5 (best quality) to indicate the quality of the individual retrievals. Table 5 shows the definition of the different QLs and according to what rules they are assigned.

The quality controls and various checks for surface and atmospheric effects referred to in the definition of QL 1 are the same as the filtering procedures defined in Section 2.5.2 (new RFI mask instead of initial mask), with few exceptions. The QL 1 check for the PMW SST CDR retrievals does not include any of the in situ SST-related checks, nor does it include a check for diurnal warming effects. On the other hand, it includes an additional check where the retrieved SST is compared to a background SST (ERA-Interim). If the retrieved SST deviates more than 10 °C from the background, the retrieval is assigned QL 1. QL 2 PMW SST retrievals in the CDR are assigned based on the total SST retrieval uncertainty and on an extended sea ice and land mask, which is based on the proximity of the retrieval to sea ice and land, respectively. In the production of the climate data record, the extended sea ice mask is based on the ERA-Interim sea ice cover product and if any sea ice is detected within ± 200 km the retrieval is assigned QL 2. Similarly, if any land is detected within ± 100 km, the retrieval is also assigned QL 2. Note that in the final production of the CDR, a larger ice and land mask was applied than what was feasible using the MMD due to the validation results shown later.

The regression retrieval has been run for the AMSR-E and AMSR2 validation subsets (SST/WS_TEST) defined in Section 2.5.2. The overall summary statistics of the PMW SSTs against drifter in situ SSTs for different QLs are shown in Table 6. The AMSR-E QL 3–5 PMW SSTs give a bias of −0.01 K and standard deviation of 0.52 K when compared against in situ SSTs. By contrast, the QL 4 and 5 AMSR-E retrievals give SSTs with a bias of −0.02 K and a standard deviation of 0.46 K. Comparing the AMSR2 QL 3–5 PMW SSTs against in situ SSTs gives a bias of 0.007 K and a standard deviation of 0.51 K. The mean and standard deviation of the AMSR2 QL 4 and 5 PMW SST retrievals are smaller, 0.002 K and 0.45 K, respectively.

To obtain a completely independent validation, the retrieved SSTs were also validated against SSTs from Argo floats. The Argo floats are not as numerous as the drifters and therefore, the AMSR-E and AMSR2 Argo validation subsets only consist of 148,895 and 154,715 matchups, respectively. The comparison of AMSR-E and AMSR2 QL 4 and 5 retrievals against Argo SSTs shows a slightly better performance than the comparison with drifters. The AMSR-E QL 4 and 5 retrievals give a bias of −0.009 K and a standard deviation of 0.44 K when compared against Argo floats. Similarly, comparing the AMSR2 QL 4 and 5 PMW SSTs against Argo floats gives a bias of 0.01 K and a standard deviation of 0.43 K.

Table 7
Performance of the WS retrieval algorithm for AMSR-E and AMSR2 against CCMP WS. The table shows the mean and standard deviation of PMW WS minus CCMP WS and number of matchups contributing to the statistics for different subsets.

Sensor	QL	Mean (ms^{-1})	Std (ms^{-1})	No. of matchups
AMSR-E	3	−0.02	1.07	2,763,087
	4	−0.03	0.80	4,399,894
	5	−0.01	0.69	2,753,625
	3–5	0.002	0.86	9,916,606
	4–5	0.01	0.76	7,153,519
AMSR2	3	−0.003	1.09	1,729,073
	4	0.02	0.81	2,549,348
	5	−0.003	0.68	2,000,938
	3–5	−0.002	0.86	6,279,359
	4–5	−0.002	0.76	4,550,286

The retrieved PMW WSs are validated against CCMP WSs and the overall summary statistics for different QLs are shown in Table 7. The performance of the WS retrieval algorithm is very similar for AMSR-E and AMSR2 with only minor differences. The AMSR-E QL 3–5 PMW WS retrievals give a bias of 0.002 ms^{-1} and a standard deviation of 0.86 ms^{-1}. The corresponding results for AMSR2 QL 3–5 give retrieved WSs with a bias of −0.002 ms^{-1} and standard deviation of 0.86 ms^{-1}. Comparing AMSR-E QL 4 and 5 WS retrievals against CCMP WSs gives a bias of 0.01 ms^{-1} and a standard deviation of 0.76 ms^{-1}. The mean and standard deviation of the AMSR2 QL 4 and 5 WS retrievals are comparable, −0.002 ms^{-1} and 0.76 ms^{-1}, respectively.

Fig. 6 shows the geographical distribution of the gridded mean and standard deviation of the AMSR-E and AMSR2 PMW SST minus drifter in situ SST based on the combined AMSR-E and AMSR2 SST/WS_TEST subsets, consisting of in total 16,195,965 QL 3–5 drifter matchups. The distribution of the mean shows a positive bias for high latitudes, especially in the southern hemisphere. This has been confirmed to be linked to undetected sea ice. Furthermore, a warm bias is seen for the west coast of America and the east coast of Asia. A negative bias is seen for e.g. the Pacific warm pool area and the Arabian Sea. No clear latitudinal dependence can be discerned for the mean bias. However, for the geographical distribution of the standard deviation a latitudinal dependence is observed. Lower standard deviations are found at low latitudes and higher standard deviations at higher latitudes. Furthermore, higher standard deviation is observed in the dynamical ocean regions, such as the Kuroshio Current, the Gulf Stream Extension and the Agulhas Current. These are highly dynamical regions with large SST gradients over smaller scales. When the large satellite footprints (43x75 km and 35x62 km for the

Table 6
Performance of the SST retrieval algorithm for AMSR-E and AMSR2 drifter and Argo subsets. The table shows the mean and standard deviation of PMW SST minus in situ SST and number of matchups contributing to the statistics for different subsets.

Sensor	QL	Mean Drifter (K)	Std Drifter (K)	No. of matchups	Mean Argo (K)	Std Argo (K)	No. of matchups
AMSR-E	3	0.02	0.64	2,763,087	0.007	0.62	39,939
	4	−0.01	0.51	4,399,894	−0.002	0.50	60,398
	5	−0.03	0.37	2,753,625	−0.02	0.36	48,558
	3–5	−0.01	0.52	9,916,606	−0.004	0.50	148,895
	4–5	−0.02	0.46	7,153,519	−0.009	0.44	108,956
AMSR2	3	0.02	0.64	1,729,073	0.03	0.62	37,273
	4	0.006	0.52	2,549,348	0.02	0.51	57,343
	5	−0.003	0.35	2,000,938	0.004	0.34	60,089
	3–5	0.007	0.51	6,279,359	0.02	0.48	154,705
	4–5	0.002	0.45	4,550,286	0.01	0.43	117,432

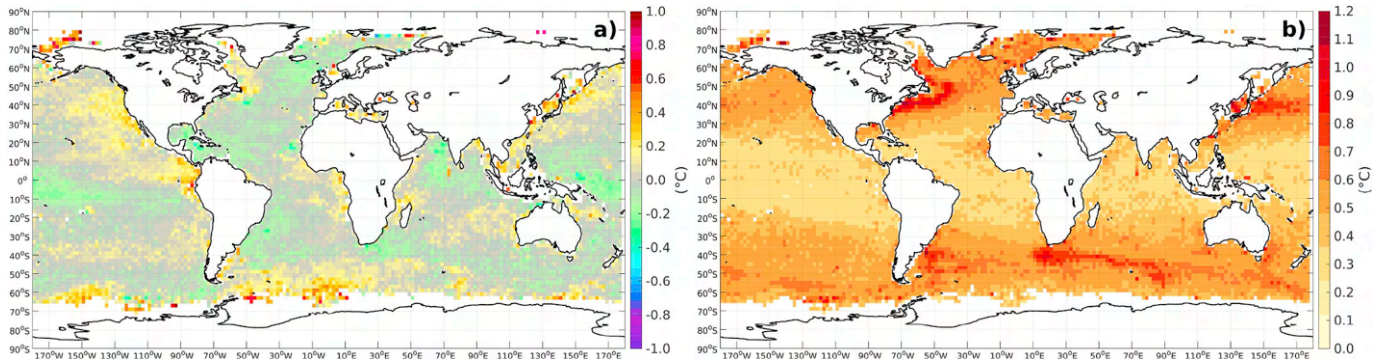


Fig. 6. The geographical distribution of a) mean and b) standard deviation of QL 3–5 AMSR-E and AMSR2 PMW SST minus drifter in situ SST. The statistics have been gridded using a grid size of 2°, with a minimum of 50 matchups per grid cell.

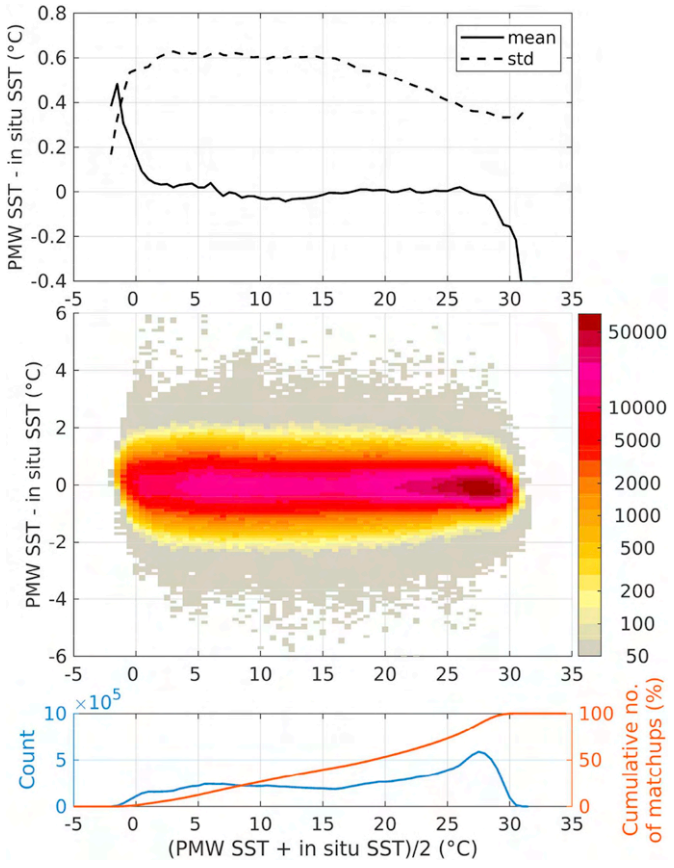


Fig. 7. AMSR-E and AMSR2 PMW SST minus drifter in situ SST as a function of average SST ((PMW SST + in situ SST)/2). Top panel shows the mean (solid) and standard deviation (dashed) of differences for each bin. The heat plot in the mid panel shows the distribution of matchups, and the bottom panel shows the number of matchups (blue) and the cumulative number of matchups (orange) in each bin. A minimum of 50 matchups per bin was used for the statistics calculation. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

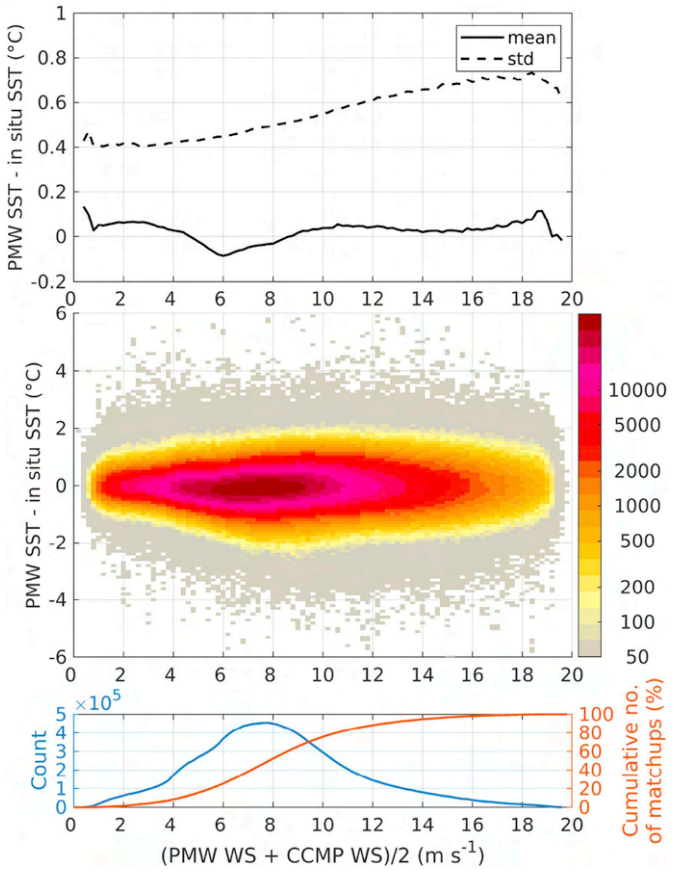


Fig. 8. AMSR-E and AMSR2 PMW SST minus drifter in situ SST as a function of average WS ((PMW WS + CCMP WS)/2). Top panel shows the mean (solid) and standard deviation (dashed) of differences for each bin. The heat plot in the mid panel shows the distribution of matchups, and the bottom panel shows the number of matchups (blue) and the cumulative number of matchups (orange) in each bin. A minimum of 50 matchups per bin was used for the statistics calculation. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

AMSR-E and AMSR2 6.9 GHz resolution, respectively) are compared against in situ SST observations, which are point measurements, the SST differences are enhanced. Hence, the higher standard deviations in these regions are not a measure of the quality of the PMW SST retrievals but rather related to the large SST variability in these regions.

The performance of the PMW SST against drifter in situ SST as a function of SST, wind speed and latitude has been investigated (Figs. 7–9). A dependence can be seen for both cold and warm SSTs. The warm bias for cold SST (SSTs < 0 °C) was investigated and found to be related to sea ice contamination, which was accounted for by extending the sea ice masking, when generating the final CDR. The reason for the cold bias for warm SST (SST > 29 °C) is not yet understood, but might be related to contamination from rain not being discarded. The standard deviation, on the other hand, decreases with warmer SST. Wind speed affects the sea surface roughness, which impacts the emissivity of the ocean and therefore also the brightness temperature in the retrievals. The dependence of the retrieved SST on wind speed therefore reflects the dependence on sea surface roughness. The binned statistics of the performance against wind speed shows a cold bias for wind

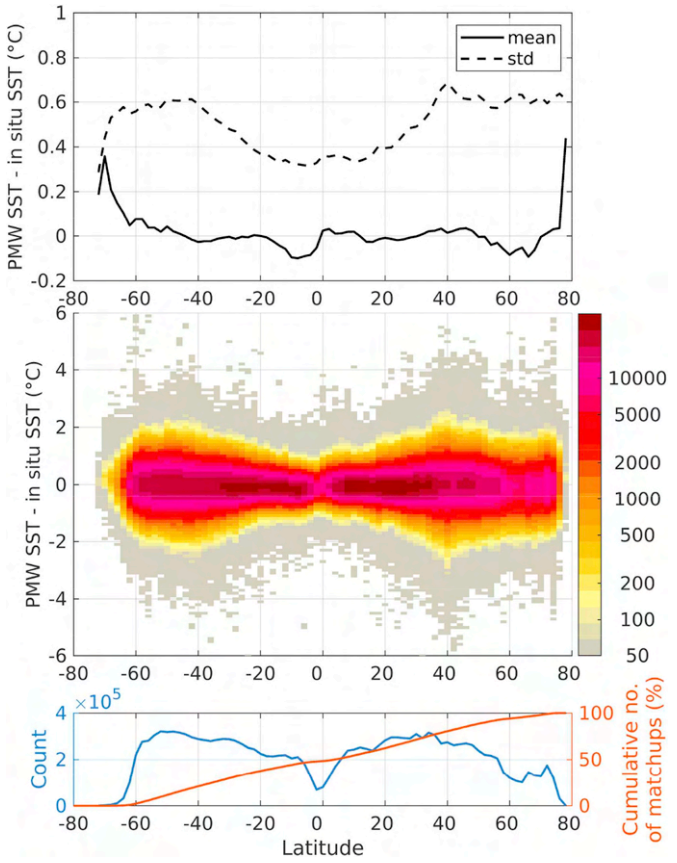


Fig. 9. AMSR-E and AMSR2 PMW SST minus drifter in situ SST as a function of latitude. Top panel shows the mean (solid) and standard deviation (dashed) of differences for each bin. The heat plot in the mid panel shows the distribution of matchups, and the bottom panel shows the number of matchups (blue) and the cumulative number of matchups (orange) in each bin. A minimum of 50 matchups per bin was used for the statistics calculation. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

speeds in the range 5–8 ms^{-1} and otherwise a small warm bias for both high and low wind speeds. The standard deviation is lowest for low wind speeds and increases for increasing wind speeds. As was evident in the geographical distributions of the gridded statistics (Fig. 6) a latitudinal dependence is seen for the SST performance against in situ SSTs. The overall variations are coupled to the latitudinal variations in both the SST and wind speeds. Lower standard deviation is found at lower latitudes and higher standard deviation is found for higher latitudes. A cold bias is seen for matchups located between 15°S and the equator, which corresponds to the cold bias seen for the Pacific warm pool area. The negative bias seen for matchups centered around 60°N corresponds to the cold bias seen just south of Iceland, in the North Atlantic Ocean. Furthermore, a warm bias is seen for matchups at the higher latitudes. This has been confirmed to be an effect of sea ice contamination.

4.1. Sensitivity

An important characteristic of an SST retrieval is the SST sensitivity, which measures the change in retrieved SST per unit change in true SST (Merchant et al., 2009). Sensitivity can be estimated via simulation and is a useful discriminator between SST retrieval methods (Petrenko et al., 2014) The SST sensitivity ideally is 1 $K K^{-1}$, however, there are several geophysical factors affecting the microwave retrieval, such as water vapor, cloud water and surface roughness (quantified through wind speed) that may lower sensitivity to the true SST variations.

The sensitivity of the retrieved SST with respect to changes in true sea surface temperature was investigated. A subset of AMSR-E versus drifter matchups for the year 2010, consisting of 4,642,710 good drifter matchups, was used to generate two sets of simulated brightness temperatures. The simulated brightness temperatures were generated using an updated version of the forward model developed by Wentz and Meissner (2000), as described in Nielsen-Englyst et al. (2018). Both sets use WS, TCLW and TCWV input from the ERA-Interim reanalysis and SST input from drifting buoys. The first set was simulated with in situ SST plus 1 °C (SST_{+1}), while the second set of brightness temperatures was simulated with in situ SST minus 1 °C (SST_{-1}), with NWP reanalysis input held constant for both sets. The resulting brightness temperatures were subsequently propagated through the regression retrieval algorithm to derive the corresponding SST_{-1} and SST_{+1} estimates. The

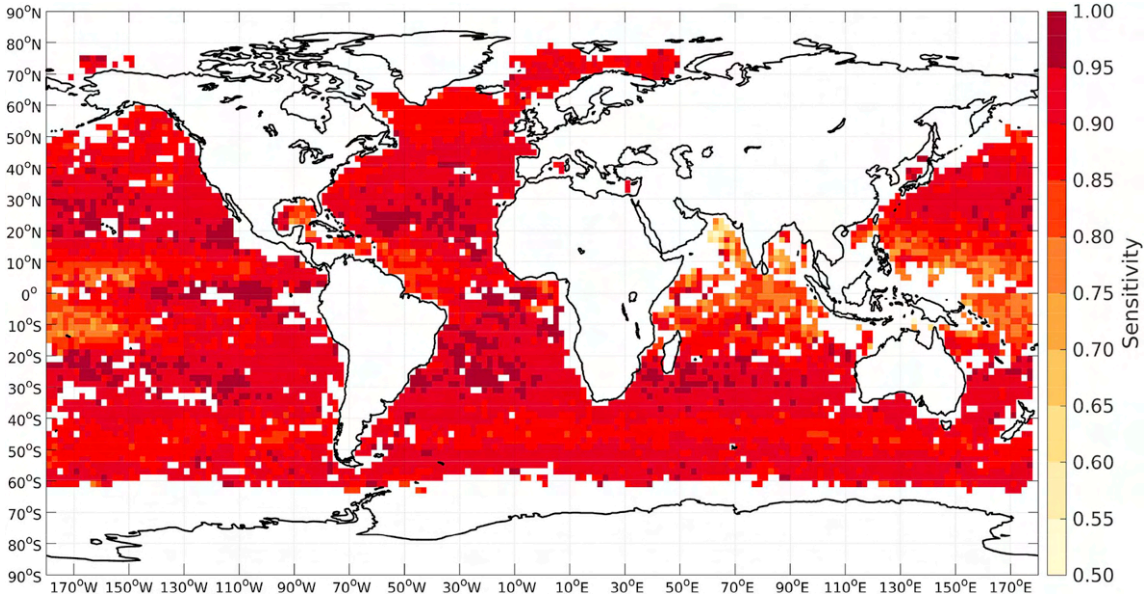


Fig. 10. The geographical distribution of sensitivity with respect to changes in SST for an AMSR-E subset consisting of matchups from year 2010. The statistics have been gridded using a grid size of 2°, with a minimum of 50 matchups per grid cell.

sensitivity was then calculated based on the retrieved SSTs and the overall mean sensitivity was estimated to 0.90.

The geographical distribution of sensitivity is shown in Fig. 10. The sensitivity is mostly larger than 0.85 but there are areas with lower sensitivity, such as the Pacific warm pool area and the Arabian Sea, with minimum sensitivities just below 0.5. The Pacific warm pool area is characterized by high TCWV content, whereas the Arabian Sea is a relatively dry area in comparison. There are also areas of relatively lower sensitivity, such as south of Greenland and in a belt around Antarctica. These regions are characterized by high wind speeds, persistent cloud cover and rain. The dependence of the sensitivity on TCWV, TCLW and WS was investigated (not shown). A small dependence for very high TCWV was found but nothing for TCLW or WS. The geographical distribution of sensitivity for the initial estimate of SST (i.e. the latitude and ascending/descending retrieval), was also investigated to see if an explanation for the low sensitivity could be found. In addition, a global regression model, using the same regression algorithm, but with only one set of regression coefficients for all data, was used to estimate the sensitivity. For the latitude and ascending/descending retrieval, lower sensitivities were found in a narrow band just north of the Equator, as well as for high latitudes (not shown). The global model exhibited a mean sensitivity of 0.92, with higher sensitivities for low latitudes and lower sensitivities for high latitudes (not shown). Based on this, we reason that the low sensitivities seen for the two-stage retrieval algorithm are primarily an artefact of the localized retrieval tuning for subsets of the data. The binning performed in both the first-stage and second-stage retrievals might result in bins with very small SST variability. For these bins, e.g. wind speed and water vapor might vary more and thus the regression algorithm will instead correct for those contributions. Hence, the sensitivity of the algorithm to changes in true SST will be lower.

4.2. SST uncertainty

Validation of the derived uncertainty is essential for the usefulness of the uncertainty estimate (Bulgin et al., 2016; Merchant et al., 2017). Here, the satellite SST retrievals versus drifter in situ observations have been used for validating the total uncertainty estimates. Only independent drifter matchups have been used for the validation. Fig. 11 shows the observed AMSR-E and AMSR2 PMW SST uncertainty against drifters versus the modelled uncertainty, i.e. the estimated PMW SST uncertainty (see Section 3.3). The ideal uncertainty is given by the dashed lines, which in addition to the satellite SST uncertainty, also includes the drifter in situ uncertainty and sampling effects. The drifter uncertainty is estimated to 0.2 K whereas the sampling effect is defined as a latitude dependent function. As shown in Høyer et al. (2012), the sampling effect of these matchup windows is primarily spatial, with only a small temporal component. Based on this, the sampling effect has been estimated by calculating the pixel-to-footprint variability for one year of GHRSST Level 4 DMI_OI Global Foundation Sea Surface Temperature Analysis (Dash et al., 2012; Høyer et al., 2014; Høyer and She, 2007). Subtracting these estimates, the mean retrieved AMSR-E and AMSR2 SST uncertainty is estimated to 0.42 K and 0.41 K, respectively. The validation results for the observed satellite SST uncertainties show good agreement with the retrieved uncertainties (see Fig. 11).

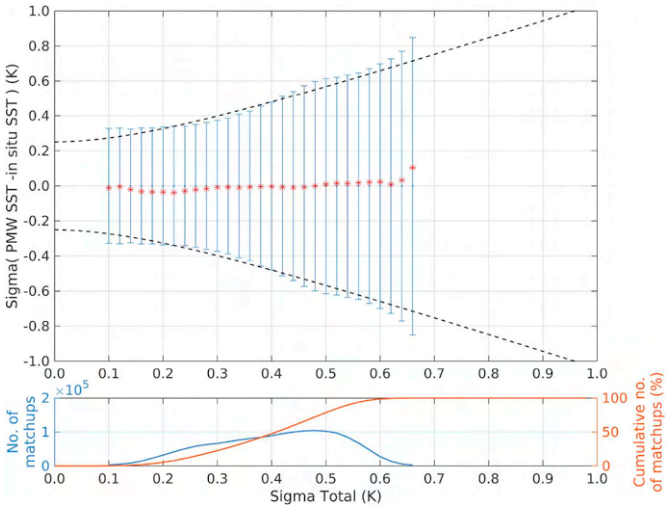


Fig. 11. AMSR-E and AMSR2 PMW SST uncertainty validation against drifter in situ SST. The top panel shows the observed PMW SST minus drifter in situ SST uncertainty versus the modelled uncertainty, i.e. the retrieved PMW SST uncertainty. The dashed lines indicate the ideal uncertainty, in which drifter SST uncertainty and sampling errors have been included. The red asterisks denote the mean bias and the solid blue lines mark one standard deviation of the PMW SST minus drifter in situ SST difference for each 0.02 K bin. The bottom panel shows the number of matchups (blue) and the cumulative percentage of matchups (red) per bin. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

4.3. Temporal consistency

The temporal stability in a climate data record is essential for later use and analysis. As the algorithms for the two satellite datasets have been derived using the same reference in situ dataset (drifting buoys), we expect that there are no large inconsistencies between the two datasets. Fig. 12 shows the seasonal averaged independent validation statistics. It is clear from the figure that the performance of the two

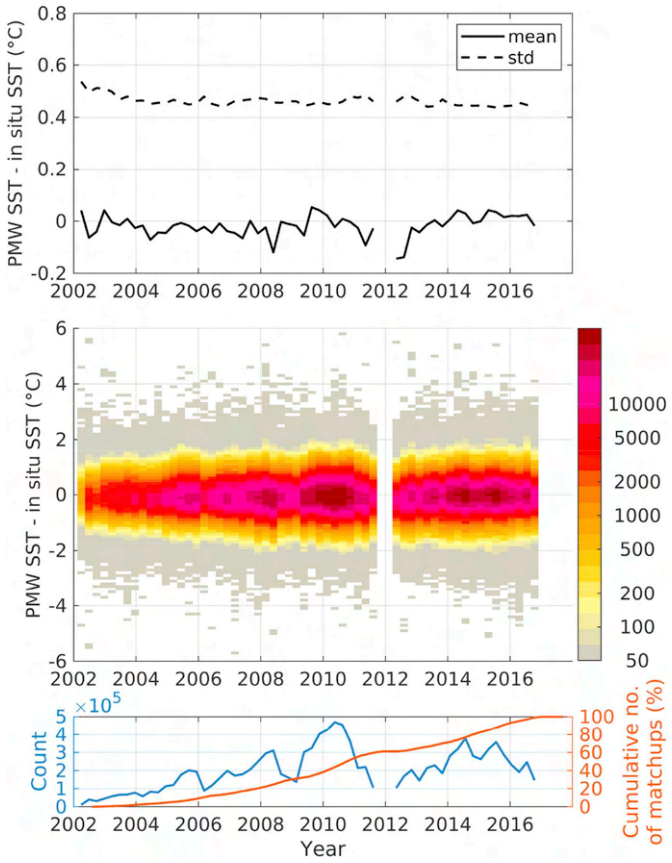


Fig. 12. AMSR-E and AMSR2 QLs 4 and 5 PMW SST minus drifter in situ SST as a function of time (season). Top panel shows the mean (solid) and standard deviation (dashed) of differences for each bin. The heat plot in the mid panel shows the distribution of matchups, and the bottom panel shows the number of matchups (blue) and the cumulative number of matchups (orange) in each bin. A minimum of 50 matchups per bin was used for the statistics calculation. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

datasets is very similar. One excursion is the beginning of the AMSR2 period, where we see a cold bias larger than 0.1°. The reason for this is not clear to us. RFI and geographical sampling effects were investigated but not found to be responsible for this deviation. It is, however, possible that the cold bias seen at the beginning of the AMSR2 record could be due to problems with the calibration of the AMSR2 instrument.

5. Discussion

In the retrieval presented above, we train and test our wind speeds against the CCMP winds that are based upon observations from passive and active Microwaves (see e.g. Atlas et al., 2011). This implies that information from the AMSR-E and AMSR2 are already included in the CCMP and that it cannot be regarded as a truly independent wind estimate. However, since the CCMP also includes information from other sensors, such as in situ observations and models, we decided to use this product for the algorithm validation, as it gave significantly better SST results, when we used the CCMP winds than using the ERA-Interim wind speeds.

The global SST validation results with a standard deviation of 0.46 and a bias of −0.02 for QL 4 and 5 retrievals are comparable to or even better than previous validations of AMSR-E PMW SST retrievals (Gentemann, 2014; O'Carroll et al., 2008). Validation results for QL 4 and 5 AMSR2 retrievals give SSTs with a bias of 0.002 K and a standard deviation of 0.45 K. Gentemann and Hilburn (2015) reported a standard deviation of 0.55 K and a bias of −0.04 K for validation of AMSR2 against in situ buoy measurements for the same period (2012–2016). We have also derived an uncertainty for each PMW SST retrieval and the uncertainty validation results against in situ measurements indicate a good agreement, with a mean uncertainty of 0.42 K and 0.41 K for AMSR-E and AMSR2 PMW retrievals, respectively. The mean of the modelled uncertainties, including the in situ uncertainty and sampling effects are estimated to 0.45 K and 0.44 K for AMSR-E and AMSR2 QLs 4 and 5 retrievals, respectively.

The performance of the SST retrieval algorithm shows a latitudinal dependence in the satellite – in situ standard deviation, with higher values for higher latitudes. This is in line with the findings reported in Gentemann (2014). Furthermore, our results show an increase in standard deviation in the dynamical ocean regions, which are regions with large mesoscale activity and large SST variability over smaller scales. This is not believed to be related to the performance of the algorithm, but arise from elevated spatial sampling errors that are introduced in dynamical regions due to the large satellite footprint compared to the pointwise in situ observation. Similar latitudinal dependences was reported by Nielsen-Englyst et al. (2018).

It was evident from Figs. 6 and 9 that we saw some sea ice contamination in the high latitudes. Due to the construction of the MMD, we were only able to mask out ice within 160 km in the algorithm development and validation, which is probably too low for the sea ice product used here. The effect on the global validation statistics presented here is limited but when generating the full CDR, an ice mask distance of 200 km was chosen to minimize the impact from sea ice. These residual effects from the sea ice could probably be reduced using an embedded sea ice retrieval, where the sea ice information is based upon a retrieval using the observations and not relying on an external product. No coastal effects were seen in the validation, indicating that the coastal mask was adequate.

The performance of the PMW SST retrievals is comparable to the good infrared satellite SST retrievals, in particular when the sampling errors are taken into account. The difference in footprint size between the IR (1 km) and the PMW SSTs (40–70 km) will introduce a larger sampling difference. Furthermore, cloud contamination has a significant impact on infrared SST retrievals since IR satellite instruments are unable to see the surface through clouds (e.g. Jones et al., 1996; Reynolds et al., 2002). In addition, biasing from both aerosols (Diaz et al., 2001; e.g. Vásquez-Cuervo et al., 2004) and through water vapor attenuation (Emery et al., 1994) affect the IR retrievals. PMW SSTs do not suffer from contamination due to either of these factors, but they are limited by precipitating clouds. The availability of PMW SSTs in regions characterized by persistent cloud cover should therefore be much higher than the availability of IR SSTs. Table 1 indicates the global coverage, defined as the number of good matchups in relation to all matchups and assuming that the full matchup dataset represents daily coverage. However, the filtering procedure described in Section 2.5.2 includes flagging of not only the quality of the satellite data but also flagging based on in situ and auxiliary data, as well as flagging to remove diurnal warming effects. By excluding the filters which do not pertain to the quality of satellite data, a global coverage of PMW SSTs can be obtained. For the tropical Indian Ocean (80–100°E,15°S-20°N), which is a region dominated by clouds (Rossow, 1993) and heavy rain (Arkin and Janowiak, 1993), the coverage was found to be 71%, meaning that 71% of the matchups can be used. The performance of the PMW SSTs against drifter in situ observations in this region was found to be −0.04 ± 0.36 K. The majority of matchups were of QL 5, with only some QL 4 data. This is in line with findings reported by Guan and Kawamura (2003), where the annual mean availability of PMW SSTs in the tropical Indian Ocean was estimated to 73%. The availability of IR SSTs in the same region was reported to be 35% (Guan and Kawamura, 2003).

Nielsen-Englyst et al. (2018) reported sensitivities ranging from 0.4 to up to 0.6, with the lowest values found for high latitudes. Similar results were found by Gentemann et al. (2010). In contrast, we have reported sensitivities ranging from just below 0.5 up to close to 1.0, with the lowest sensitivities found at low latitudes. The latitudinal dependency was opposite expectations because of the temperature dependency of the T_b sensitivity to SST changes, where warm SSTs have higher T_bs sensitivity than cold waters (Prigent et al., 2013). The sensitivity to changes in true SST for IR retrievals should be close to 1 (i.e. perfect sensitivity) (Merchant et al., 2009). Merchant et al. (2009) reported an average sensitivity of 0.93, with minimum sensitivities in areas of high TCWV such as the equatorial Atlantic and Pacific Oceans, in particular the Pacific Warm Pool area. Similarly, Embury and Merchant (2012) reported sensitivities close to unity. The binning performed in both the first-stage and the second-stage retrieval is thought to be the reason for this difference in latitudinal dependency. When performing the regression, we divide the data into different bins. If the SST variability within a bin is small, the algorithm corrects for other factors, such as wind speed (through its impact on surface roughness), water vapor, etc. instead of SST. Hence, the sensitivity to changes in true SST will be lower.

The CDR generated here showed a good inter-sensor consistency, as they have been referenced to the same in situ observational network. The use of in situ observations for SST algorithm developments has a risk of introducing biases related to inadequacies in the observational network. This effect was attempted to be minimized with the geographical even-out-by-latitude and appeared to be small for our algorithm, as the Argo validation confirms the validation results from drifting buoys. However, the approach of using physically based retrievals is preferred for the IR retrievals as discussed in (Merchant et al., 2008; Merchant and Le Borgne, 2004) and have also shown promising results for AMSR-E (Nielsen-Englyst et al., 2018). Future developments could include forward modelling in the inter-sensor adjustment and algorithm developments.

A High Priority Candidate Mission (HPCM) called the Copernicus Imaging Microwave Radiometer (CIMR (Donlon, 2019),) is now being studied at the European Space Agency. CIMR proposes a conically scanning radiometer having a swath > 1900 km and will include channels at 1.4 GHz (~60 km), 6.9 and 10.65 GHz (< 15 km) 18.8 GHz (5–6 km) and 36.5 GHz (4–5) km on the same platform in a high inclination dawn-dusk orbit coordinated with the MetOp-SG(B) (offering opportunities for synergy with the Microwave Radiometer (MWI) and Scatterometer (SCA)). Our algorithms can easily be adapted to the

CIMR frequencies, taking into account the enhanced spatial resolution, improved NEdT and addition of an L-band channel that can help in wind speeds much greater than 20 ms^{-1}.

6. Conclusions

Retrieval algorithms have been developed with the purpose of generating an SST climate data record from the AMSR-E and AMSR2 passive microwave satellite observations for the period June 2002–October 2017. The algorithms include WS and SST retrievals, with corresponding SST uncertainty retrievals and a new and effective method for detecting and removing RFI contamination. The results show an overall bias and standard deviation for QLs 4 and 5 AMSR-E and AMSR2 PMW SSTs against drifter in situ SSTs of -0.02 ± 0.46 K and 0.002 ± 0.45 K, respectively. These results are comparable or even better than previous validations of both AMSR-E and AMSR2 SST retrievals (Gentemann, 2014; Gentemann and Hilburn, 2015; O'Carroll et al., 2008). The modelled SST uncertainty, including in situ uncertainty and sampling effects, are estimated to 0.45 K and 0.44 K, respectively. These retrieved SST uncertainties are, to our knowledge, a new feature and distinguishes the presented PMW SST CDR from other PMW SST products. The thorough validation results against independent in situ observations demonstrate that the algorithms developed here generate a consistent climate data record, with very good performance and reliable uncertainty estimates. Furthermore, the SST sensitivity of 0.9 for the algorithm ensures retrievals that are able to represent the true variability in the SST. The PMW SST CDR is publically available at http://gws-access.ceda.ac.uk/public/esacci-sst/PMW2.0_release/AMSR/L2P/.

To conclude, the PMW SST CDR has been shown to provide accurate and consistent SST retrievals which can be used in non-precipitating conditions for global monitoring and assessment of the oceans. With the uncertainties in the funding for future microwave satellite missions, it is therefore highly recommended that a fully operational multi-frequency passive microwave satellite mission, such as CIMR, should be part of the future satellite constellation to monitor the world oceans.

Author contributions

Emy Alerskans, Jacob L. Høyer and Chelle L. Gentemann conceived and designed the experiments. Emy Alerskans, Jacob L. Høyer and Pia Nielsen-Englyst performed the experiments. Emy Alerskans, Jacob L. Høyer, Chelle L. Gentemann, Leif Toudal Pedersen and Craig Donlon analyzed the data.

Declaration of competing interest

The authors declare no conflict of interest.

Acknowledgements

The authors wish to thank Christopher J. Merchant for valuable comments to the paper. The retrieval algorithms and the datasets have been developed within the European Space Agency's Climate Change Initiative for Sea Surface Temperature project under grant number 4000109848/13/I-NB. The PMW SST climate data record is available from the Centre for Environmental Data Archival (CEDA) at http://gws-access.ceda.ac.uk/public/esacci-sst/PMW2.0_release/AMSR/L2P/. CCMP Version-2.0 vector wind analyses are available at www.remss.com. ICOADS data are available at https://icoads.noaa.gov/. EN4 version 4.2.0 is available at https://www.metoffice.gov.uk/hadobs/en4/. The resampled L2A swath data product AMSR-E V12 is produced by RSS and distributed by NASA's National Snow and Ice Data Center (NSIDC). Data are available at https://nsidc.org/data/ae_l2a. The AMSR2 L1R version 2 swath data product is produced by JAXA and is available from the Globe Portal System (G-Portal) at https://gportal.jaxa.jp/gpr/. ERA-Interim reanalysis data is available at https://www.ecmwf.int/en/forecasts/datasets/reanalysis-datasets/era-interim.

References

Abraham, J.P., Baringer, M., Bindoff, N.L., Boyer, T., Cheng, L.J., Church, J.A., Conroy, J.L., Domingues, C.M., Fasullo, J.T., Gilson, J., Goni, G., Good, S.A., Gorman, J.M., Gouretski, V., Ishii, M., Johnson, G.C., Kizu, S., Lyman, J.M., Macdonald, A.M., Minkowycz, W.J., Moffitt, S.E., Palmer, M.D., Piola, A.R., Reseghetti, F., Schuckmann, K., Trenberth, K.E., Velicogna, I., Willis, J.K., 2013. A review of global ocean temperature observations: implications for ocean heat content estimates and climate change. Rev. Geophys. 51, 450–483. https://doi.org/10.1002/rog.20022.

Argo, 2018. Argo Float Data and Metadata from Global Data Assembly Centre (Argo GDAC). SEANOE.

Arkin, P.A., Janowiak, J.E., 1993. Tropical and subtropical precipitation. In: Gurney, R.J., Foster, J.L., Parkinson, C.L. (Eds.), Atlas of Satellite Observations Related to Global Change. Cambridge University Press, pp. 165–180.

Ashcroft, P., Wentz, F.J., 2013. AMSR-E/Aqua L2A Global Swath Spatially-Resampled Brightness Temperatures. Version 3.

Atkinson, C.P., Rayner, N.A., Kennedy, J.J., Good, S.A., 2014. An integrated database of ocean temperature and salinity observations. J. Geophys. Res. Ocean. 119, 7139–7163. https://doi.org/10.1002/2014JC010053.

Atlas, R., Hoffman, R.N., Ardizzone, J., Leidner, S.M., Jusem, J.C., Smith, D.K., Gombos, D., 2011. A cross-calibrated, multiplatform ocean surface wind velocity product for meteorological and oceanographic applications. Bull. Am. Meteorol. Soc. 92, 157–174. https://doi.org/10.1175/2010BAMS2946.1.

Block, T., Embacher, S., Merchant, C.J., Donlon, C., 2018. High-performance software framework for the calculation of satellite-to-satellite data matchups (MMS version 1.2). Geosci. Model Dev 11, 2419–2427. https://doi.org/10.5194/gmd-11-2419-2018.

Brasnett, B., Colan, D.S., 2016. Assimilating retrievals of sea surface temperature from VIIRS and AMSR2. J. Atmos. Ocean. Technol. 33, 361–375. https://doi.org/10.1175/JTECH-D-15-0093.1.

Available atBrown, O.B., Minnett, P.J., 1999. MODIS Infrared Sea Surface Temperature Algorithm - Algorithm Theoretical Basis Document Version 2.0. : https://modis.gsfc.nasa.gov/data/atbd/atbd_mod25.pdf.

Bulgin, C.E., Embury, O., Corlett, G., Merchant, C.J., 2016. Independent uncertainty estimates for coefficient based sea surface temperature retrieval from the Along-Track Scanning Radiometer instruments. Remote Sens. Environ. 178, 213–222. https://doi.org/10.1016/j.rse.2016.02.022.

Chacko, N., 2017. Chlorophyll bloom in response to tropical cyclone Hudhud in the Bay of Bengal: bio-Argo subsurface observations. Deep-Sea Res. Part I Oceanogr. Res. Pap. 124, 66–72.

Chang, P.S., Jelenak, Z., Alsweiss, S., 2015. Algorithm Theoretical Basis Document: GCOM-W1/AMSR2 Day-1 EDR Version 1.0. Available at:https://www.star.nesdis.noaa.gov/jpss/documents/ATBD/ATBD_AMSR2_Ocean_EDR_v2.0.pdf.

Chelton, D.B., Wentz, F.J., 2005. Global microwave satellite observations of sea surface temperature for numerical weather prediction and climate research. Bull. Am. Meteorol. Soc. 86, 1097–1115. https://doi.org/10.1175/BAMS-86-8-1097.

Dash, P., Ignatov, A., Martin, M., Donlon, C., Brasnett, B., Reynolds, R.W., Branzon, V., Beggs, H., Cayula, J.-F., Chao, Y., Grumbine, R., Maturi, E., Harris, A., Mittaz, J., Sapper, J., Chin, T.M., Vazquez-Cuervo, J., Armstrong, E.M., Gentemann, C., Cummings, J., Piolle, J.F., Autret, E., Roberts-Jones, J., Ishizaki, S., Høyer, J.L., Poulter, D., 2012. Group for high resolution sea surface temperature (GHRSST) analysis fields inter-comparisons - Part 2: near real time web-based level 4 SST quality monitor (L4-SQUAM). Deep Sea Res. Part II Top. Stud. Oceanogr. 77.

Dee, D.P., Uppala, S.M., Simmons, A.J., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M.A., Balsamo, G., Bauer, P., Bechtold, P., Beljaars, A.C.M., van de Berg, L., Bidlot, J., Bormann, N., Delsol, C., Dragani, R., Fuentes, M., Geer, A.J., Haimberger, L., Healy, S.B., Hersbach, H., Hólm, E.V., Isaksen, L., Kållberg, P., Köhler, M., Matricardi, M., McNally, A.P., Monge-Sanz, B.M., Morcrette, J.J., Park, B.K., Peubey, C., de Rosnay, P., Tavolato, C., Thépaut, J.N., Vitart, F., 2011. The ERA-Interim reanalysis: configuration and performance of the data assimilation system. Q. J. R. Meteorol. Soc. 137, 553–597. https://doi.org/10.1002/qj.828.

Diaz, J.P., Arbelo, M., Expósito, F.J., Podestá, G., Prospero, J.M., Evans, R., 2001. Relationship between errors in AVHRR-derived sea surface temperature and the TOMS Aerosol Index. Geophys. Res. Lett. 28, 1989–1992.

Donlon, C.J., 2019. Mission Requirements Document for the Copernicus Imaging Microwave Radiometer, v2.0, March 2019. available from: the European Space Agency at. https://cimr.eu/sites/cimr.met.no/files/documents/CIMR-MRD-v2.0-20190305-ISSUED_0.pdf.

Donlon, C.J., Stark, M.M., Roberts-Jones, J., Fiedler, E., Wimmer, W., 2012. The operational sea surface temperature and sea ice analysis (OSTIA) system. Remote Sens. Environ. 116, 140–158.

Embury, O., Merchant, C.J., 2012. A reprocessing for climate of sea surface temperature from the along-track scanning radiometers: a new retrieval scheme. Remote Sens. Environ. 116, 47–61. https://doi.org/10.1016/j.rse.2010.11.020.

Embury, O., Merchant, C.J., Corlett, G.K., 2012. A reprocessing for climate of sea surface temperature from the along-track scanning radiometers: initial validation, accounting for skin and diurnal variability effects. Remote Sens. Environ. 116, 62–78. https://doi.org/10.1016/j.rse.2011.02.028.

Emery, W.J., Yu, Y., Wick, G.A., Schluessel, P., Reynolds, R.W., 1994. Correcting infrared satellite estimates of sea surface temperature for atmospheric water vapor attenuation. J. Geophys. Res. Ocean. 99, 5219–5236.

Gentemann, C.L., 2014. Three way validation of MODIS and AMSR-E sea surface temperatures. J. Geophys. Res. Ocean. 119, 2583–2598. https://doi.org/10.1002/2013JC009716.
Gentemann, C.L., Donlon, C.J., Stuart-Menteth, A., Wentz, F.J., 2003. Diurnal signals in satellite sea surface temperature measurements. Geophys. Res. Lett. 30. https://doi.org/10.1029/2002GL016291.
Gentemann, C.L., Hilburn, K., 2015. In situ validation of sea surface temperatures from the GCOM-W1 AMSR2 RSS calibrated brightness temperatures: validation of RSS GCOM-W1 SST. J. Geophys. Res. Ocean. 120, 3567–3585. https://doi.org/10.1002/2014JC010574.
Gentemann, C.L., Meissner, T., Wentz, F.J., 2010. Accuracy of satellite sea surface temperatures at 7 and 11 GHz. IEEE Trans. Geosci. Remote Sens. 48, 1009–1018. https://doi.org/10.1109/TGRS.2009.2030322.
Gentemann, C.L., Wentz, F.J., Mears, C.A., Smith, D.K., 2004. In situ validation of Tropical Rainfall Measuring Mission microwave sea surface temperatures. J. Geophys. Res. Ocean. 109, 1–9. https://doi.org/10.1029/2003JC002092.
GHRSST Science Team, 2010. The Recommended GHRSST Data Specification (GDS) 2.0, Document Revision 5. pp. 123. Available from: the GHRSST International Project Office, 2011. https://www.ghrsst.org/about-ghrsst/governance-documents/.
Gille, S.T., 2008. Decadal-scale temperature trends in the southern hemisphere ocean. J. Clim. 21, 4749–4765. https://doi.org/10.1175/2008JCLI2131.1.
Gladkova, I., Ignatov, A., Shahriar, F., Kihai, Y., Hillger, D., Petrenko, B., 2016. Improved VIIRS and MODIS SST imagery. Remote Sens. 8, 79.
Good, S.A., Martin, M.J., Rayner, N.A., 2013. EN4 : quality controlled ocean temperature and salinity profiles and monthly objective analyses with uncertainty estimates. J. Geophys. Res. Ocean. 118, 6704–6716. https://doi.org/10.1002/2013JC009067.
Guan, L., Kawamura, H., 2003. SST availabilities of satellite infrared and microwave measurements. J. Oceanogr. 59, 201–209.
Hollinger, J.P., 1971. Passive microwave measurements of sea surface roughness. IEEE Trans. Geosci. Electron. 9, 165–169.
Høyer, J.L., Karagali, I., Dybkjær, G., Tonboe, R., 2012. Multi sensor validation and error characteristics of Arctic satellite sea surface temperature observations. Remote Sens. Environ. 121, 335–346. https://doi.org/10.1016/j.rse.2012.01.013.
Høyer, J.L., Le Borgne, P., Eastwood, S., 2014. A bias correction method for Arctic satellite sea surface temperature observations. Remote Sens. Environ. 146, 201–213.
Høyer, J.L., She, J., 2007. Optimal interpolation of sea surface temperature for the North sea and baltic sea. J. Mar. Syst. 65, 176–189.
International Telecommunication Union, 2012. The Radio Regulations, 2012 Ed. Geneva, Switzerland. http://en.wikipedia.org/wiki/ITU_Radio_Regulations.
Ishizaki, S., 2014. The State of the Western North Pacific in the First Half of 2013, vol. 22. PICES Press, pp. 40.
Jones, M.S., Saunders, M.A., Guymer, T.H., 1996. Reducing cloud contamination in ATSR averaged sea surface temperature data. J. Atmos. Ocean. Technol. 13, 492–506.
Kennedy, J.J., 2014. A review of uncertainty in in situ measurements and data sets of sea surface temperature: in situ SST uncertainty. Rev. Geophys. 52, 1–32. https://doi.org/10.1002/2013RG000434.
Le Borgne, P., Roquet, H., Merchant, C.J., 2011. Estimation of sea surface temperature from the Spinning Enhanced Visible and Infrared Imager, improved using numerical weather prediction. Remote Sens. Environ. 115, 55–65.
Le Traon, P.Y., Antoine, D., Bentamy, A., Bonekamp, H., Breivik, L.A., Chapron, B., Corlett, G., Dibarboure, G., Digiacomo, P., Donlon, C.J., Faugère, Y., Font, J., Girard-Ardhuin, F., Gohin, F., Johannessen, J.A., Kamachi, M., Lagerloef, G., Lambin, J., Larnicol, G., Le Borgne, P., Leuliette, E., Lindstrom, E., Martin, M.J., Maturi, E., Miller, L., Mingsen, L., Morrow, R., Reul, N., Rio, M.H., Roquet, H., Santoleri, R., Wilkin, J., 2015. Use of satellite observations for operational oceanography: recent achievements and future prospects. J. Oper. Oceanogr. 8, 12–27. https://doi.org/10.1080/1755876X.2015.1022050.
Liang, X., Yang, Q., Nerger, L., Losa, S.N., Zhao, B., Zheng, F., Zhang, L., Wu, L., 2017. Assimilating Copernicus SST data into a pan-arctic ice-ocean coupled model with a local SEIK filter. J. Atmos. Ocean. Technol. 34, 1985–1999. https://doi.org/10.1175/JTECH-D-16-0166.1.
Maeda, T., Taniguchi, Y., Imaoka, K., 2016. GCOM-W1 AMSR2 level 1R product: dataset of brightness temperature modified using the antenna pattern matching Technique. IEEE Trans. Geosci. Remote Sens. 54, 770–782. https://doi.org/10.1109/TGRS.2015.2465170.
Meissner, T., Wentz, F., 2002. An updated analysis of the ocean surface wind direction signal in passive microwave brightness temperatures. IEEE Trans. Geosci. Remote Sens. 40, 1230–1240.
Meissner, T., Wentz, F.J., 2012. The emissivity of the ocean surface between 6 and 90 GHz over a large range of wind speeds and Earth incidence angles. IEEE Trans. Geosci. Remote Sens. 50, 3004–3026. https://doi.org/10.1109/TGRS.2011.2179662.
Merchant, C.J., Embury, O., Le Borgne, P., Bellec, B., 2006. Saharan dust in night-time thermal imagery: detection and reduction of related biases in retrieved sea surface temperature. Remote Sens. Environ. 104, 15–30.
Merchant, C.J., Embury, O., Rayner, N.A., Berry, D.I., Corlett, G.K., Lean, K., Veal, K.L., Kent, E.C., Llewellyn-jones, D.T., Remedios, J.J., Saunders, R., 2012. A 20 year independent record of sea surface temperature for climate from Along-Track Scanning Radiometers. J. Geophys. Res. Ocean. 117. https://doi.org/10.1029/2012JC008400.
Merchant, C.J., Embury, O., Roberts-Jones, J., Fiedler, E., Bulgin, C.E., Corlett, G.K., Good, S., McLaren, A., Rayner, N., Morak-Bozzo, S., Donlon, C., 2014. sea Surface temperature datasets for climate applications from phase 1 of the european space agency climate change initiative (SST CCI). Geosci. Data J. 1, 179–191. https://doi.org/10.1002/gdj3.20.
Merchant, C.J., Harris, A.R., Murray, M.J., Zavody, A.M., 1999. Toward the elimination of bias in satellite retrievals of skin sea surface temperature 1. Theory, modeling and inter-algorithm comparison. J. Geophys. Res. Ocean. 104, 23565–23578. https://doi.org/10.1029/1999JC900105.
Merchant, C.J., Harris, A.R., Roquet, H., Le Borgne, P., 2009. Retrieval characteristics of non-linear sea surface temperature from the Advanced Very High Resolution Radiometer. Geophys. Res. Lett. 36. https://doi.org/10.1029/2009GL039843.
Merchant, C.J., Le Borgne, P., 2004. Retrieval of sea surface temperature from space, based on modeling of infrared radiative transfer: capabilities and limitations. J. Atmos. Ocean. Technol. 21, 1734–1746. https://doi.org/10.1175/JTECH1667.1.
Merchant, C.J., Le Borgne, P., Marsouin, A., Roquet, H., 2008. Optimal estimation of sea surface temperature from split-window observations. Remote Sens. Environ. 112, 2469–2484. https://doi.org/10.1016/j.rse.2007.11.011.
Merchant, C.J., Paul, F., Popp, T., Ablain, M., Bontemps, S., Defourny, P., Hollmann, R., Lavergne, T., Laeng, A., de Leeuw, G., Mittaz, J., Poulsen, C., Povey, A.C., Reuter, M., Sathyendranath, S., Sandven, S., Sofeiva, V.F., Wagner, W., 2017. Uncertainty information in climate data records from Earth observation. Earth Syst. Sci. Data 9, 511–527. https://doi.org/10.5194/essd-9-511-2017.
Minnett, P.J., 1986. A numerical study of the effects of anomalous North Atlantic atmospheric conditions on the infrared measurement of sea surface temperature from space. J. Geophys. Res. Ocean. 91, 8509–8521.
Monzikova, A.K., Kudryavtsev, V.N., Reul, N., Chapron, B., 2017. On the upper ocean response to tropical cyclones: satellite microwave observation. In: Progress in Electromagnetics Research Symposium-Fall (PIERS-FALL). IEEE, pp. 2437–2444.
Nielsen-Englyst, P., Høyer, J.L., Pedersen, L.T., Gentemann, C.L., Alerskans, E., Block, T., Donlon, C., 2018. Optimal estimation of sea surface temperature from AMSR-E. Remote Sens. 10. https://doi.org/10.3390/rs10020229.
Ning, J., Qing, X., Wang, T., Zhang, S., 2018. Upper ocean response to super typhoon soudelor revealed by different SST products. In: IGARSS 2018-2018 IEEE International Geoscience and Remote Sensing Symposium. IEEE.
O'Carroll, A.G., August, T., Le Borgne, P., Marsouin, A., 2012. The accuracy of SST retrievals from Metop-A IASI and AVHRR using the EUMETSAT OSI-SAF matchup dataset. Remote Sens. Environ. 126, 184–194.
O'Carroll, A.G., Eyre, J.R., Saunders, R.W., 2008. Three-way error analysis between AATSR, AMSR-E, and in situ sea surface temperature observations. J. Atmos. Ocean. Technol. 25, 1197–1207. https://doi.org/10.1175/2007JTECHO542.1.
Petrenko, B., Ignatov, A., Kihai, Y., Stroup, J., Dash, P., 2014. Evaluation and selection of SST regression algorithms for JPSS VIIRS. J. Geophys. Res. Atmos. 119, 4580–4599.
Prigent, C., Aires, F., Bernardo, F., Orlhac, J.C., Goutoule, J.M., Roquet, H., Donlon, C., 2013. Analysis of the potential and limitations of microwave radiometry for the retrieval of sea surface temperature: definition of MICROWAT, a new mission concept. J. Geophys. Res. Ocean. 118, 3074–3086.
Rayner, N., Good, S., Block, T., 2015. SST CCI Product User Guide, Project Document. SST_CCI-PUG-UKMO-201. http://www.esa-sst-cci.org/PUG/documents.htm.
Reynolds, R.W., Rayner, N.A., Smith, T.M., Stokes, D.C., Wang, W., 2002. An improved in situ and satellite SST analysis for climate. J. Clim. 15, 1609–1625.
Roemmich, D., Johnson, G.C., Riser, S., Davis, R., Gilson, J., Owens, W.B., Garzoli, S.L., Schmid, C., Ignaszewski, M., 2009. The Argo program: observing the global ocean with profiling floats. Oceanography 22, 34–43. https://doi.org/10.5670/oceanog.2009.36.
Rossow, W.B., 1993. Clouds. In: Gurney, R.J., Foster, J.L., Parkinson, L. (Eds.), Atlas of Satellite Observations Related to Global Change. Cambridge University Press, pp. 141–163.
Shibata, A., 2006. Features of ocean microwave emission changed by wind at 6 GHz. J. Oceanogr. 62, 321–330. https://doi.org/10.1007/s10872-006-0057-3.
Udaya Bhaskar, T.V.S., Rahman, S.H., Pavan, I.D., Ravichandran, M., Nayak, S., 2009. Comparison of AMSR-E and TMI sea surface temperature with Argo near-surface temperature over the Indian Ocean. Int. J. Remote Sens. 30, 2669–2684. https://doi.org/10.1080/01431160802555796.
Ulaby, F.T., Moore, R.K., Fung, A.K., 1981. Microwave Remote Sensing: Active and Passive. Volume 1-Microwave Remote Sensing Fundamentals and Radiometry.
Vásquez-Cuervo, J., Armstrong, E.M., Harris, A., 2004. The effect of aerosols and clouds on the retrieval of infrared sea surface temperatures. J. Clim. 17, 3921–3933. https://doi.org/10.1175/1520-0442(2004)017<3921:TEOAAC>2.0.CO;2.
Wentz, F.J., Gentemann, C., Smith, D., Chelton, D., 2000. Satellite measurements of sea surface temperature through clouds. Science 288, 847–850.
Wentz, F.J., Meissner, T., 2007. Supplement 1 Algorithm Theoretical Basis Document for AMSR-E Ocean Algorithms. RSS Technical Report 051707. Remote Sensing Systems, Santa Rosa, CA. http://images.remss.com/papers/rsstech/2007_051707_Wentz_AMSR_Ocean_Algorithm_Version_2_Supplement1_ATBD.pdf.
Wentz, F.J., Meissner, T., 2000. Algorithm Theoretical Basis Document (ATBD): AMSR Ocean Algorithm (Version 2), RSS Tech. Proposal 121599A-1. Remote Sensing Systems, Santa Rosa, CA. http://images.remss.com/papers/rsstech/2000_121599A-1_Wentz_AMSR_Ocean_Algorithm_ATBD_Version2.pdf.
Woodruff, S.D., Worley, S.J., Lubker, S.J., Ji, Z., Eric Freeman, J., Berry, D.I., Brohan, P., Kent, E.C., Reynolds, R.W., Smith, S.R., Wilkinson, C., 2011. ICOADS Release 2.5: extensions and enhancements to the surface marine meteorological archive. Int. jounral Climatol. 31, 951–967. https://doi.org/10.1002/joc.2103.
Yang, C.-S., Kim, S.-H., Ouchi, K., Back, J.-H., 2015. Generation of high resolution sea surface temperature using multi-satellite data for operational oceanography. Acta Oceanol. Sin. 34, 74–88. https://doi.org/10.1007/s13131-015-0694-8.

APPENDIX B

Paper II: Exploring machine learning techniques to retrieve sea surface temperatures from passive microwave measurements

This appendix contains a full copy of Paper II:

- Alerskans, Emy, Ann-Sofie P. Zinck, **Pia Nielsen-Englyst**, Jacob L. Høyer. 2022. "Exploring machine learning techniques to retrieve sea surface temperatures from passive microwave measurements", *Remote Sensing of Environment*, Volume 281, 113220, ISSN 0034-4257.

Remote Sensing of Environment 281 (2022) 113220

Contents lists available at ScienceDirect

Remote Sensing of Environment

journal homepage: www.elsevier.com/locate/rse

Exploring machine learning techniques to retrieve sea surface temperatures from passive microwave measurements

Emy Alerskans [a,*,1], Ann-Sofie P. Zinck [b,1], Pia Nielsen-Englyst [c,a], Jacob L. Høyer [a]

[a] Danish Meteorological Institute, Copenhagen, Denmark
[b] IMAU, Utrecht University, Utrecht, The Netherlands
[c] DTU-Space, Technical University of Denmark, Lyngby, Denmark

ARTICLE INFO

Edited by Menghua Wang

Keywords:
Remote sensing
Passive microwave
Sea surface temperature
Machine learning
AMSR-E

ABSTRACT

Two machine learning (ML) models are investigated for retrieving sea surface temperature (SST) from passive microwave (PMW) satellite observations from the Advanced Microwave Scanning Radiometer – Earth Observing System (AMSR-E) and auxiliary data, such as ERA5 reanalysis data. The first model is the Extreme Gradient Boosting (XBG) model and the second is a multilayer perceptron neural network (NN). The performance of the two ML algorithms is compared to that of an existing state-of-the-art regression (RE) retrieval algorithm.

The performance of the three algorithms is assessed using independent in situ SSTs from drifting buoys. Overall, the three models have similar biases; 0.01, 0.01 and −0.02 K for the XGB, NN and RE, respectively. The XGB model performs best with respect to standard deviation; 0.36 K. While the NN model performs slightly better than the RE model with respect to standard deviation, 0.50 and 0.55 K, respectively, the RE model is found to be more sensitive to changes in the in situ SST. Moreover, the XGB model is the least sensitive with an overall sensitivity of 0.78, compared to 0.90 for the RE model and 0.88 for the NN model.

The good performance of the two ML algorithms compared to the state-of-the-art RE algorithm in this initial study demonstrates that there is a large potential in the use of ML algorithms for the retrieval of SST from PMW satellite observations.

1. Introduction

Sea surface temperature (SST) is an essential climate variable (Bojinski et al., 2014) used in various applications such as climate monitoring (e.g. Merchant et al., 2019), numerical weather prediction (NWP; Chelton and Wentz, 2005; Brasnett and Colan, 2016), ocean and coupled models (Le Traon et al., 2015; Yang et al., 2015; Liang et al., 2017) and in the understanding of air–sea interactions (Monzikova et al., 2017; Ning et al., 2018). SST has been measured in situ for more than 150 years, initially from ships and oceanographic profiles and later from moored and drifting buoys (Rayner et al., 2006). SST retrieved from Earth-orbiting satellites is a crucial supplement to the in situ network due to the more complete temporal and spatial coverage from satellites (Minnett et al., 2019). Thermal infrared (IR) satellite observations have been available since 1981, but these observations are biased from aerosols and limited by their inability to observe the surface through clouds (Merchant et al., 1999, 2006). Observations from passive microwave (PMW) sensors are widely recognised as an important supplement to IR observations since PMW observations of the surface are not prevented by non-precipitating clouds and the impact from aerosols is small (Wentz and Meissner, 2000; Chelton and Wentz, 2005). They are, however, impacted by precipitation (Gentemann et al., 2010) and sun glint contamination, which increases the swath gaps (Gentemann and Hilburn, 2015). The first global accurate PMW SST data using the 6 GHz channels became available in 2002 from the Advanced Microwave Scanning Radiometer – Earth Observing System (AMSR-E; Kawanishi et al., 2003; Chelton and Wentz, 2005), carried onboard the National Aeronautics and Space Administration's (NASA's) Earth Observation System Aqua platform. AMSR-E ceased normal operations in October 2011 and was followed by the currently operational AMSR2 on the Global Change Observing Mission (GCOM-W1; Maeda et al., 2015), launched in May 2012. An AMSR2 follow-on mission (AMSR3) is planned by Japan Aerospace Exploration Agency (JAXA) (Maeda et al., 2020) and the Copernicus Imaging Microwave Radiometer (CIMR) is prepared by the European

* Corresponding author.
E-mail address: ea@dmi.dk (E. Alerskans).
[1] Previous address: Physics of Ice, Climate, and Earth, Niels Bohr Institute, University of Copenhagen, Copenhagen, Denmark.

https://doi.org/10.1016/j.rse.2022.113220
Received 23 December 2021; Received in revised form 7 August 2022; Accepted 9 August 2022
Available online 5 September 2022

Space Agency (ESA) as a part of the Copernicus Expansion Program of the European Union (http://www.cimr.eu/; Donlon, 2020).

Different PMW SST retrievals have been developed and refined over the years using different frequency channels and approaches. Two types of retrieval algorithms have generally been used to retrieve SST from PMW observations; statistical algorithms and physically based algorithms. The most common approach to generate PMW SST products is by using a statistical retrieval algorithm (e.g. Shibata, 2006; Wentz and Meissner, 2007; Gentemann et al., 2009; Chang et al., 2015; Alerskans et al., 2020). Statistical retrieval algorithms are developed by comparisons of satellite measured brightness temperatures and collocated and temporally matched in situ observations and model data, such as atmospheric and oceanic reanalysis data of wind speed, atmospheric water vapour content and SST. The second type of retrieval algorithm uses a radiative transfer model (RTM) to simulate the top of atmosphere brightness temperatures. This approach requires instrument information (azimuth/earth incidence angles, frequency and polarisation) and environmental information (SST, sea surface salinity, wind speed/direction, water vapour density, liquid water density, pressure, and atmospheric profiles of temperature). Optimal estimation (OE) theory is an example of an approach that makes use of an RTM (forward model). In OE models, the RTM is inverted in order to retrieve SST from satellite measured brightness temperatures (Nielsen-Englyst et al., 2018). The inversion is performed using a priori information about the ocean and atmosphere (and corresponding uncertainties) to constrain the retrievals.

The OE retrieval allows for indication of measurement errors, such as imperfect calibration and channel contamination (Minnett et al., 2019). This also means that the performance of OE algorithms is constrained by the accuracy of the RTM as well as the representativeness of the observation and prior error covariances (Merchant et al., 2020). Moreover, measurement errors require ad-hoc corrections to the geophysical retrievals in the OE type of algorithms (Meissner and Wentz, 2012; Nielsen-Englyst et al., 2018). In contrast, statistically based algorithms may account for some of the measurement errors through the coefficient derivation process, but they are limited by the established statistical relationships between the variables. Hence, both physical and statistical models make a series of considerable assumptions about the nature of the radiative transfer process, which are provided directly by the RTM in physical models, whereas statistical models rely on established assumptions of how the geophysical quantities can be used as proxies for the actual physical processes that influence the surface emissivity and the radiative transfer through the atmosphere.

Machine learning (ML) models may improve or supplement existing retrieval algorithms through their higher flexibility and capability of recognising meaningful patterns and structures in complex problems (Lee et al., 2017; Azodi et al., 2020). Compared to both physical models and statistical models, there are much fewer assumptions about the functional form of how the geophysical quantities are related to the predicted quantity in ML models. This may allow development of complex functional forms that more closely approximate the actual physical processes and thereby provide a more accurate SST retrieval. ML models may also be a good alternative in situations where observation characteristics and the structure of the uncertainty components are not well known, e.g. during a commissioning phase of a new instrument such as CIMR. Therefore, there is a need for insight into the performance of different ML models for retrieving SST. Until recently, the use of ML techniques has been very limited within the field of SST retrievals, but investigations using ML to improve the accuracy of SST algorithms is listed as one of the priority recommendations provided by the SST community (O'Carroll et al., 2019). There has been an increasing amount of research applying ML techniques to specific parts of retrieval algorithms, such as for cloud detection (Paul and Huntemann, 2021), bias correction (Saux Picart et al., 2018), error estimation (Kumar et al., 2021), identification of eddies (Moschos et al., 2020) and ocean extremes (Prochaska et al., 2021). A recent study also used ML techniques to retrieve daily cloud-free IR SSTs from the MODIS Aqua sensor (Sunder et al., 2020). In addition, ML techniques have also been used for retrieval of other satellite-derived geophysical variables, such as soil moisture (Rodriguez-Fernandez et al., 2015) and precipitation (Sanò et al., 2016, 2018).

In this paper two types of ML SST retrieval techniques have been assessed and compared against an existing state-of-the-art statistical regression model retrieval algorithm. The first is the decision tree-based algorithm Extreme Gradient Boosting (XGBoost, here XGB; Chen and Guestrin, 2016), and the second is a multilayer perceptron (MLP) neural network (NN; Haykin, 1999; Nielsen, 2015). These methods differ in architecture and represent two of the main ML categories; decision-trees and neural networks. The XGB is a relatively new algorithm which has shown good performance for retrieval and bias correction of other geophysical variables (e.g. Just et al., 2018, 2020; Liu et al., 2021). The MLP is a fully-connected feed-forward NN and is one of the simplest and most used neural network architectures.

The paper is structured with a description of the dataset, as well as pre-processing and dataset splitting in Section 2. This is followed by a presentation of the three retrieval algorithms and model optimisation of the two ML algorithms in Section 3. The results are presented in Section 4 and discussed in Section 5 before the final concluding remarks are provided in Section 6.

2. Data

2.1. ESA CCI Multisensor Matchup Dataset (MMD)

The ESA climate change initiative (CCI) Multisensor Matchup Dataset (MMD), described in Nielsen-Englyst et al. (2018) and Alerskans et al. (2020), is the basis for this work. The MMD consists of quality controlled in situ measured SST observations from the International Comprehensive Ocean-Atmosphere DataSet (ICOADS) version 2.5.1 (Woodruff et al., 2011) and the Met Office Hadley Centre (MOHC) Ensembles dataset version 4.2.0 (EN4; Good et al., 2013). Brightness temperatures from the AMSR-E Level 2 A (L2 A) swath data product, AMSR-E V12 (Ashcroft and Wentz, 2013), spatially re-sampled to the 6.9 GHz resolution (75 × 43 km), are also included. The in situ and satellite observations are matched by imposing a maximal geodesic distance of 20 km and a maximal time difference of 4 h. The MMD includes matchups from the period June 2002–October 2011.

Additional data included in the MMD are information from both the ERA-Interim reanalysis (Dee et al., 2011) and the ERA5 reanalysis (Hersbach et al., 2020) on SST, total column water vapour (TCWV), total cloud liquid water (TCLW), wind speed (WS) and sea ice concentration (SIC). Sea surface salinity (SSS) from the GLOBAL-REANALYSIS-PHY-001-030 reanalysis product, provided by the Copernicus Marine Environment Monitoring Service (CMEMS; http://marine.copernicus.eu) is also included in the MMD. Additional wind data from the Cross-Calibrated Multi-Platform (CCMP) gridded surface wind vector product (Atlas et al., 2011) version 2.0 was included. The additional data were collocated in time and space with the MMD matchups using the nearest neighbour interpolation. For a list of the MMD variables extracted for this study and considered as input features to the two ML retrieval algorithms see Table 1.

2.2. Pre-processing

To ensure an accurate derivation of the retrieval algorithms erroneous in situ, satellite and auxiliary data are excluded. The quality of the brightness temperatures were assessed using the L1 AMSR-E instrument quality flags and low-quality data were excluded. Moreover, brightness temperatures outside the accepted range (0–320 K) were flagged. In addition, data were excluded if the difference between vertical (V) and horizontal (H) polarisations for the 18–36 GHz brightness temperatures were negative, as this indicates invalid oceanographic

Table 1
MMD variables considered as input features to the ML retrieval algorithms. The asterisk marks the features used in the XGB model (see Section 3.2).

Feature	Acronym	Source
AMSR-E orbit (asc./desc.)	orbit	AMSR-E
Latitude*	lat	AMSR-E
Longitude*	lon	AMSR-E
Solar zenith angle	solza	AMSR-E
Satellite zenith angle*	satza	AMSR-E
Satellite azimuth angle	sataz	AMSR-E
Sun glint angle	sga	AMSR-E
Brightness temperature, channel 6V*	tb6V	AMSR-E
Brightness temperature, channel 6H*	tb6H	AMSR-E
Brightness temperature, channel 10V*	tb10V	AMSR-E
Brightness temperature, channel 10H*	tb10H	AMSR-E
Brightness temperature, channel 18V	tb18V	AMSR-E
Brightness temperature, channel 18H	tb18H	AMSR-E
Brightness temperature, channel 23V	tb23V	AMSR-E
Brightness temperature, channel 23H	tb23H	AMSR-E
Brightness temperature, channel 36V*	tb36V	AMSR-E
Brightness temperature, channel 36H*	tb36H	AMSR-E
Brightness temperature, channel 89V*	tb89V	AMSR-E
Brightness temperature, channel 89H*	tb89H	AMSR-E
Wind speed*	WS	ERA5
Wind direction	ϕ_W	ERA5
Relative angle between sataz and ϕ_W^*	ϕ_{REL}	ERA5/AMSR-E
Total column water vapour*	TCWV	ERA5
Cloud liquid water*	CLWT	ERA5

retrievals. To exclude low-quality brightness temperatures possibly contaminated by e.g. rain and sea ice, an additional quality control check for the AMSR-E 23 and 36 GHz brightness temperatures were performed. The spatial standard deviation was calculated over a 21 × 21 pixel subregion around each matchup in order to exclude matchups with an anomalously high standard deviation. Data were flagged if the standard deviations of the 23 V and H and 36 V and H channels were larger than 55, 35, 25 and 25 K, respectively. These thresholds were chosen based on the distribution of the spatial standard deviations in order to exclude matchups in the end tails, as they have a very high and anomalous spatial standard deviation and therefore likely are contaminated. The chosen thresholds resulted in exclusion of less than 1% of the matchups. Low quality in situ data and matchups with an in situ or ERA5 SST outside the range −2–34 °C were also excluded, where the lower limit of −2 °C is used in order to exclude matchups potentially contaminated by sea ice. Furthermore, matchups with an ERA5 WS greater than 20 ms^{-1} were also flagged. The upper wind speed limit is based on the fact that extreme surface roughness and the existence of foam on the surface caused by high wind speeds impact the brightness temperatures and make the SST retrievals uncertain (Kilic et al., 2018). Together, these checks constitute the gross error checks in Table 2, which are performed to remove obviously erroneous satellite, in situ and auxiliary data. To exclude matchups that might have been contaminated due to atmospheric or surface effects, additional checks were performed. Matchups contaminated by sea ice or land were excluded using the AMSR-E land/ocean flag and the ERA5 sea ice fraction. To account for contamination due to rain, matchups were removed if the 18 V GHz brightness temperature was greater than 240 K. Sun glitter contamination was avoided by excluding matchups with a sun glint angle less than 25°. Diurnal warming effects were accounted for by excluding daytime matchups with ERA5 WS less than 4 ms^{-1}. Matchups potentially contaminated by ground-based and space-based RFI were excluded using observation location and reflection longitude and latitude according to Table 2 in Gentemann and Hilburn (2015). Lastly, obviously erroneous in situ SSTs were removed using a 3-sigma filter, based on the mean difference between ERA5 and in situ SSTs. To ensure a balanced and latitudinally representative dataset, such that the models are trained and validated on data not only from a few specific regions in which in situ observations are dense, the number of matchups per latitude degree was restricted. As the number of

Table 2
Number of matchups remaining after each check and the percentage of matchups each check removes. The percentages removed for checks 1–8 plus the summary checks ("Checks 1–7", "Checks 1-8" and "All checks 1–9") are with respect to all matchups. The percentage of matchups removed by the SST 3σ-filter, on the other hand, is with respect to "Checks 1–7" and the even-out-by-latitude check is with respect to "Checks 1–8".

Filter	No. of matchups	Percentage of matchups removed (%)
(0) All matchups (no filter)	40,480,306	–
(1) Gross error checks	31,070,944	23.24
(2) Rain	401,42,612	0.83
(3) Sun glint	38,145,778	5.77
(4) RFI	37,387,456	7.64
(5) Land	34,839,510	13.93
(6) Sea ice	31,390,993	22.45
(7) Diurnal warming	37,311,599	7.83
Checks 1–7	19,397,886	52.08
(8) SST 3σ-filter	18,999,399	2.05
Checks 1–8	18,999,399	53.07
(9) Even-out-by-latitude	15,316,989	19.38
All checks 1–9	15,316,989	61.16

matchups increase with time, the restriction is temporally dependent with different number of matchups allowed for different years. For 2002, which is the year with fewest matchups, a maximum of 2,000 matchups per latitude degree were allowed, whereas for 2011, which is the year with the most matchups, a maximum limit of 20,000 matchups per latitude degree was used. The percentage of matchups removed and the total number of matchups left after each filtering check is shown in Table 2. Furthermore, the geographical distribution of satellite versus drifting buoy matchups after filtering for the validation dataset is shown in Fig. 1.

The MMD is divided into six subsets in order to perform all steps on independent data. A random splitting of the data is performed such that all datasets retain the same distribution for each variable. The number of matchups in each subset, as well as the percentage of data with respect to the filtered dataset, is indicated in parentheses.

1. Training dataset: used for training the NN and XGB models (6,126,795/40.0%).
2. Test dataset: used for evaluating the performance of the ML models during training (1,021,133/6.7%).
3. Hyperparameter optimisation dataset: used for optimising model hyperparameters (see Section 3.4; 3,063,398/20.0%).
4. Feature selection dataset: used for selecting input variables for the NN and XGB models (see Section 3.1; 1,021,133/16.7% of the training dataset).
5. Validation dataset: used for validating the performance of the NN, XGB and RE models (5,105,663/33.3%).
6. Sensitivity dataset: used for the estimating the SST sensitivity of the NN, XGB and RE models (1,021,133/20.0% of the validation dataset).

Another important part of the pre-processing step in order to ensure a good performance for the ML retrieval algorithms is data normalisation (Kotsiantis et al., 2006; Huang et al., 2020). Normalisation of the data is a transformation of the data in order to transform the data to the same scale. Different methods can be used for normalisation of the data. Here, we have used quantile transform normalisation. This method uses quantile information in order to transform the data to follow a uniform distribution and as it reduces the impact from outliers it is a robust normalisation method. It has previously been used for feature normalisation within satellite-based applications and for classification (Ferreira et al., 2019; Sismanidis et al., 2021). Other popular methods include min–max normalisation and standardisation. It should be mentioned that no universal normalisation method exists

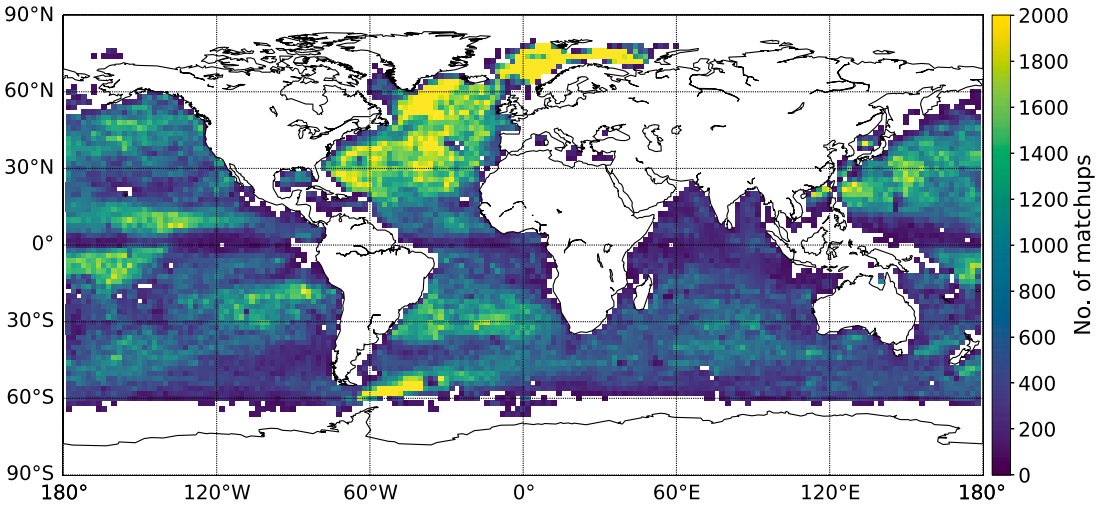


Fig. 1. Spatial distribution of satellite matchups with drifting buoys for the validation dataset. The statistics have been calculated on a 2 × 2 degree grid with a minimum of 50 matchups per grid cell.

and that the performance of models might vary depending on the normalisation method and the problem.

3. Retrievals

In this section the three models used for retrieving SST will be introduced. First the selection of input variables (also called input features) is presented, followed by a description of the two ML models, XGB and NN, and the optimisation thereof. Lastly, the state-of-the-art regression retrieval model used as a benchmark is presented.

3.1. Feature selection

Table 1 shows the 24 features that were extracted from the MMD and considered as input to the two ML models. To exclude redundant features and only select the important ones, such that the dimensionality of the input data is reduced and the risk of overfitting likewise is reduced (Goodfellow et al., 2016), a feature importance analysis was performed in order to obtain the explanatory power of each input feature. The analysis was based on the SHapley Additive exPlanations (SHAP) values (Lundberg and Lee, 2017). SHAP uses shapley values (Shapley Ll, 1953), which are based on cooperative game theory and are used in many state-of-the-art feature attribution methods (Ribeiro et al., 2016; Shrikumar et al., 2016, 2017). SHAP is based on the idea that the performance of all possible combinations of input features should be considered when determining the importance of a single feature on a single prediction. To determine the importance of each feature, the ML model to be used (here XGB and NN) is trained for each combination of input features and the marginal contribution of each feature is evaluated. The marginal contribution is defined as the difference between the performance of the model which includes the feature to be assessed and the model in which the feature is excluded. The marginal contribution of a feature is therefore obtained by considering the difference between all models in which this feature is present and all models in which it is excluded. From this, the average contribution of a single feature can be obtained. Based on this, an importance is assigned to each feature for each prediction and from this the average explanatory power of each feature can be estimated. For a further explanation of SHAP see Lundberg and Lee (2017).

The SHAP analysis is performed on the XGB and NN base models, i.e. the models with default settings. For XGB the default settings are given by its python implementations using scikit-Learn (Pedregosa et al., 2011), whereas for the NN model the corresponding default parameters were used, with the exception of number of hidden layers

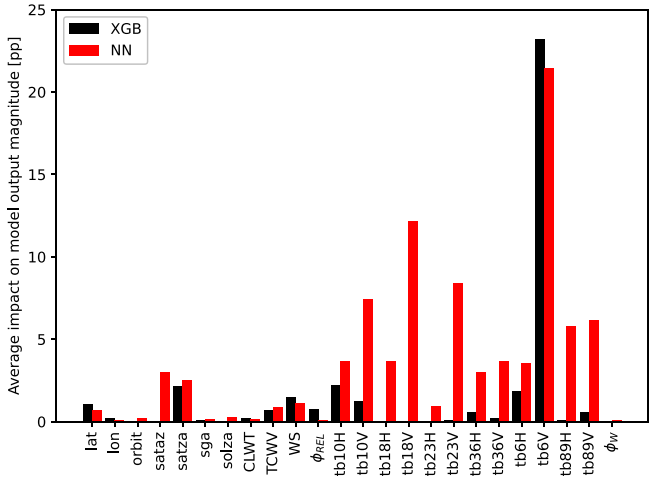


Fig. 2. SHAP feature importance analysis for XGB (black) and NN (red) with the average magnitude of impact on the model prediction (in percentage points) on the ordinate. Each bar represents the importance of a single input feature.

and neurons, which were chosen as 2 layers and 20 and 15 neurons, respectively (see Section 3.3). Therefore, the SHAP feature importance analysis might be slightly different after the ML model settings have been optimised. However, performing the optimisation using all features might also yield different results than performing it on the subset of selected features. As the purpose of the feature analysis is to estimate the explanatory power of the input features and perform a feature selection we therefore perform the SHAP analysis before optimising the models. Fig. 2 shows the SHAP values for each input feature for the two models. In both models the most dominant feature is the vertical polarisation of the 6 GHz brightness temperature (tb6V), on average changing the predicted values by 23 percentage points (pp) and 21 pp for XGB and NN, respectively. Other than tb6V, the SHAP values of the different input features differ greatly between the two models. Furthermore, the general magnitudes of importances are very different for the two models, with lower average importances in the XGB model. Therefore, the choice is made to keep all features in the NN model and only reduce the number of features in the XGB model by using a threshold of 0.1 pp. The features with an average impact on the XGB model predictions higher than this threshold are therefore included in the XGB model. These are marked with an asterisk in Table 1.

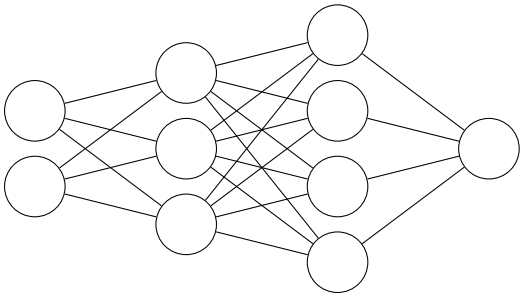


Fig. 3. Example of a neural network with an input layer consisting of two input neurons, two hidden layers with three and four neurons, respectively, and an output layer consisting of one neuron.

3.2. XGBoost

Extreme Gradient Boosting (XGBoost, here referred to as XGB) is a supervised machine learning model for when working with tabulated data. It has shown to provide state-of-the-art results on both classification and regression problems (Zhang et al., 2017; Liu et al., 2020). Here, a brief introduction of XGB is given and for a more detailed description the reader is referred to e.g. Chen and Guestrin (2016).

XGB is a so-called tree-based ML model, which means that it consists of decision trees (Breiman et al., 1984). A decision tree divides the input data into different regions with separate parameters for each region, such that the structure of the model resembles that of a tree. In XGB, trees are built sequentially, which allows a tree to learn from previous trees through a method called gradient boosting (Friedman, 2001). XGB is based on extreme gradient boosting, which is a highly scalable gradient boosting method with a sophisticated sparsity-aware algorithm for parallel tree learning (Chen and Guestrin, 2016). In gradient boosting algorithms, prior knowledge of trees and splitting is used to build better trees, since every tree is validated as it is built. Thus, each new tree will be better than the previous. All trees then contribute to the final prediction through a weighted average. The extreme gradient boosting algorithm uses a learning rate (also called shrinkage) to update the trainable model parameters in the same way as a neural network (Chen and Guestrin, 2016).

A well-known issue with machine learning is the risk of overfitting, which means that the model becomes good only at predicting data from the training dataset and performs poorly when presented to unseen data, i.e. the generalisation ability of the model becomes poor (Goodfellow et al., 2016). The problem of overfitting in tree-based models is a well-studied topic, with several different methods suggested for preventing overfitting, such as pre-pruning, post-pruning and early stopping (Esposito et al., 1997; Ying, 2019). To avoid overfitting in the XGB model early stopping, based on the mean absolute difference (MAD) metric of a test dataset, is used in the training of the model.

3.3. Neural network (NN)

Neural networks (NNs) are inspired by the functionality of the neural system and are one of the most well-known ML techniques for supervised learning. Here, a short description of the neural network used in this study is given. For a more comprehensive and detailed description of neural networks see e.g. LeCun et al. (2015), Nielsen (2015) and Goodfellow et al. (2016).

Fig. 3 shows an illustration of an NN, which consists of an input layer, two hidden layers and an output layer. Each layer consists of one or more neurons (also called nodes or units). It is through the input layer that the NN receives its input and the output layer produces the output of the model. The number of neurons corresponds to the number of inputs and outputs, respectively. In Fig. 3, the NN receives two inputs and as the output layer only consists of one neuron, only one output is produced. The layers in between the input and output layers are called hidden layers as they are neither input nor output layers, but are hidden in between. The connections between neurons in the different layers each has a weight associated with it, which indicates the weight given to the respective input information. Furthermore, the connections between the neurons in the hidden layer(s) are associated with an activation function. The purpose of the activation function is to introduce non-linearity to the system, as well as to allow for variable importance and to introduce an on–off behaviour in the response of the model to the input data. In this study, we have used the multilayer perceptron model, which is a fully-connected feed-forward NN, applying the backpropagation method (Hecht-Nielsen, 1992) during training. The NN retrieval algorithm used in this study is implemented using the TensorFlow interface (Abadi et al., 2015). As for the XGB model, early stopping is applied to ensure that the NN is not overfitting.

3.4. Optimisation

ML models have two types of parameters; (i) parameters which the ML model estimates during the training process; and (ii) hyperparameters, which need to be assigned prior to the model training. These hyperparameters can be tuned in order to improve the performance of the model. This is done through a process called hyperparameter optimisation (HPO). There exist several methods for performing HPO, two of them being through gridded and randomised searches (Liashchynskyi and Liashchynskyi, 2019; Yang and Shami, 2020). The gridded search offers a thorough scan of the entire desired parameter space, whereas the randomised search only scans a fraction of the desired parameter space, based on the chosen hyperparameter distribution function, thereby decreasing the computational cost.

HPO for the XGB model is performed using the scikit-learn RandomizedSearchCV. For the NN model, on the other hand, the scikit-learn GridSearchCV was used. The reason for this was that it was difficult to define the parameter space and obtain a satisfactory performance with the randomised algorithm for the NN model. Hence, a gridded search was performed instead. The parameter space was easier to define for the XGB model, which is why a randomised algorithm was used in order to reduce the computational cost. To ensure that the models are not overfitting to the training data in the HPO, k-fold cross-validation (e.g. Grimm et al., 2017; Berrar, 2018) in five folds is used. The hyperparameters considered, their prior distributions and the final value for each hyperparameter obtained from the HPO is shown in Table 3. It should be noted, that the entire possible hyperparameter space has not been investigated. Searching for more combinations in an extended space, might alleviate the strong link currently seen between the prior distributions and their final chosen values.

3.5. Regression model

The statistical regression model retrieval algorithm described in Alerskans et al. (2020) is used as benchmark in order to compare the performance of the ML retrievals. The regression (RE) model consists of a two-stage WS regression model followed by a two-step SST retrieval regression model. The first step uses a global algorithm to retrieve an initial estimate of wind speed. In the second step, these initial estimates are used to derive localised retrieval algorithms. Both steps in the WS retrieval algorithm use AMSR-E brightness temperatures and are regressed against CCMP wind speeds. The SST retrieval algorithm applies localised retrievals for both steps. In the first step, regression coefficients are derived locally for fixed latitude intervals and ascending and descending passes, respectively, whereas the second step uses localised SST and WS algorithms. Both steps in the SST retrieval algorithm use AMSR-E brightness temperatures, Earth incidence angle, retrieved wind speeds from the WS retrieval algorithm, and the relative angle between satellite azimuth angle and wind direction. For more information on the RE model, see Alerskans et al. (2020).

Table 3
The hyperparameters optimised for the XGB and NN models, their prior distributions and final values obtained through the HPO. The prior distributions for the XGB randomised HPO include a uniform distribution, with the minimum and maximum values specified, a Poisson distribution, with the expected separation indicated, and a normal (Gaussian) distribution, with the mean and standard deviation indicated. For the NN, the gridded HPO search intervals are shown.

	Hyperparameter	Prior distribution	Final value
XGB	Number of gradient boosted trees	Poisson(100)	103
	Maximum tree depth	Poisson(25)	22
	Minimum number of incidences in a final leaf	uniform(1,5)	3
	Subsampling[a]	norm(0.6,0.1)	0.58
	Subsampling by tree[b]	norm(0.6,0.1)	0.7
	Subsampling by level[c]	norm(0.6,0.1)	0.63
	Learning rate[d]	norm(0.1,0.03)	$8.5 \cdot 10^{-2}$
	Hyperparameter	Search range	Final value
NN	Number of hidden layers	1–3	2
	Number of neurons in each hidden layer	[15,20,25,30]	20 (1st layer) 15 (2nd layer)
	Activation function in the hidden layers	[ReLU, tanh]	tanh
	Optimiser[e]	[Adam, SGD]	Adam
	Initial learning rate[f]	0.0001–0.01	$8 \cdot 10^{-4}$

[a]Fraction used to randomly select a subset of training data.
[b]Fraction used to randomly select features to train each tree.
[c]Fraction used to randomly select a subsample of the features for every new depth level reached in a tree.
[d]The rate at which the trainable model parameters are updated during the training process.
[e]Algorithm by which the weights are optimised in order to minimise a loss function.
[f]Adam uses an adaptive learning rate, hence the initial value of the learning rate is optimised here.

The RE model was developed using a previous MMD version, in which ERA-Interim data was used instead of ERA5 data, as it had not been produced yet. The subsetting for the RE model is therefore different from the two ML retrieval algorithms, also due to different needs for number of subsets. Therefore, the same data are not used for training of the RE and ML models. The two training dataset are, however, representative of each other and the RE and ML models are therefore trained on similar data. The RE model is, on the other hand, validated on the same subset as the ML models. However, this means that the validation of the RE model is likely not performed on completely independent data as some matchups in the RE training dataset likely are included in the validation dataset. However, as validation on the same data makes the results more comparable the RE model was validated on the same subset as the two ML models.

4. Results

4.1. Overall

The two ML models and the RE retrieval algorithm have been run for the validation dataset introduced in Section 2.2. The overall performance of the three retrieval algorithms, as validated against drifter in situ SSTs (SST_{insitu}), is shown in Table 4. The overall bias of the retrieved AMSR-E PMW SSTs is 0.01, 0.01 and −0.02 K for the XGB, NN and RE models, respectively. The standard deviation of the retrieved PMW SSTs versus drifter in situ SSTs is 0.36, 0.50 and 0.55 K for the XGB, NN and RE models, respectively. The XGB retrieval algorithm performs the best, with a small bias and lowest standard deviation, whereas the NN and RE retrievals perform more similarly, where the NN model has a slightly smaller standard deviation. Overall, the XGB model shows better performance with respect to the other verification metrics as well and the NN and RE models show more similar overall results, with the NN model performing slightly better.

Fig. 4 shows the geographical distribution of mean and standard deviation of retrieved minus in situ SSTs for the XGB, NN and RE models. For the XGB model, only few areas have biases and these are generally small, as can be seen in Fig. 4(a). At higher latitudes, especially in the Southern Ocean, areas with a slight warm bias can be seen. Small cold biases, on the other hand, can be seen for e.g. the Arabian Sea and the Pacific warm pool area. The corresponding results for the NN model (Fig. 4(c)) show more and larger areas with both warm and cold biases. Most notable are the areas of large warm biases in the Southern Ocean. The XGB model also show a warm bias for some of these areas, although not a as wide-spread nor as large. Cold biases can be seen for the NN model for e.g. the higher northern latitudes and close to the tip of South America, as well as for the Arabian Sea. The RE model (Fig. 4(e)) also has a larger warm bias in the high latitudes, especially for the southern hemisphere, which was confirmed to be linked to undetected sea ice. This area is the same area where both the XGB and NN models also exhibit warm biases. In addition, areas of large warm biases are seen for the west coast of North and Central America. Furthermore, the RE model also exhibits a similar cold bias for the Arabian Sea and the Pacific warm pool area, much like the two ML models, however more pronounced and wide-spread. Otherwise, areas of both warm and cold biases can be seen. In general, no clear latitudinal pattern for the spatial distribution of bias can be seen for any of the models.

Table 4
Overall performance of the three retrieval algorithms. The table shows the mean difference (MD), standard deviation of the difference (STD), mean absolute difference (MAD), mean squared difference (MSD) and the R^2 score of retrieved minus in situ SST. The overall sensitivity of the three models to changes in situ SST (see Section 4.3) is shown as well.

	NN	XGB	RE
MD $[K]$	0.01	0.01	−0.02
MSD $[K^2]$	0.25	0.13	0.30
MAD $[K]$	0.37	0.24	0.42
STD $[K]$	0.50	0.36	0.55
R^2	0.997	0.998	0.996
Sensitivity	0.88	0.78	0.90

The geographical distribution of standard deviation, on the other hand, shows a clear latitudinal pattern for all three models. Higher standard deviations are seen for the higher latitudes, and lower standard deviations are found at lower latitudes. Furthermore, all three models exhibit higher standard deviations for the dynamical ocean regions, such as the Gulf Stream extension, the Aghulas Current and the Kuroshio Current, as well as off the east coast of Argentina. These dynamical ocean regions are areas with large SST gradients over smaller scales. Comparing the retrievals from AMSR-E, which has a resolution of several kilometres, with a point observation in one of these regions will therefore add to the discrepancies, as discussed in Alerskans et al. (2020) and Nielsen-Englyst et al. (2018). Higher standard deviations are therefore expected for the dynamical ocean regions and is not

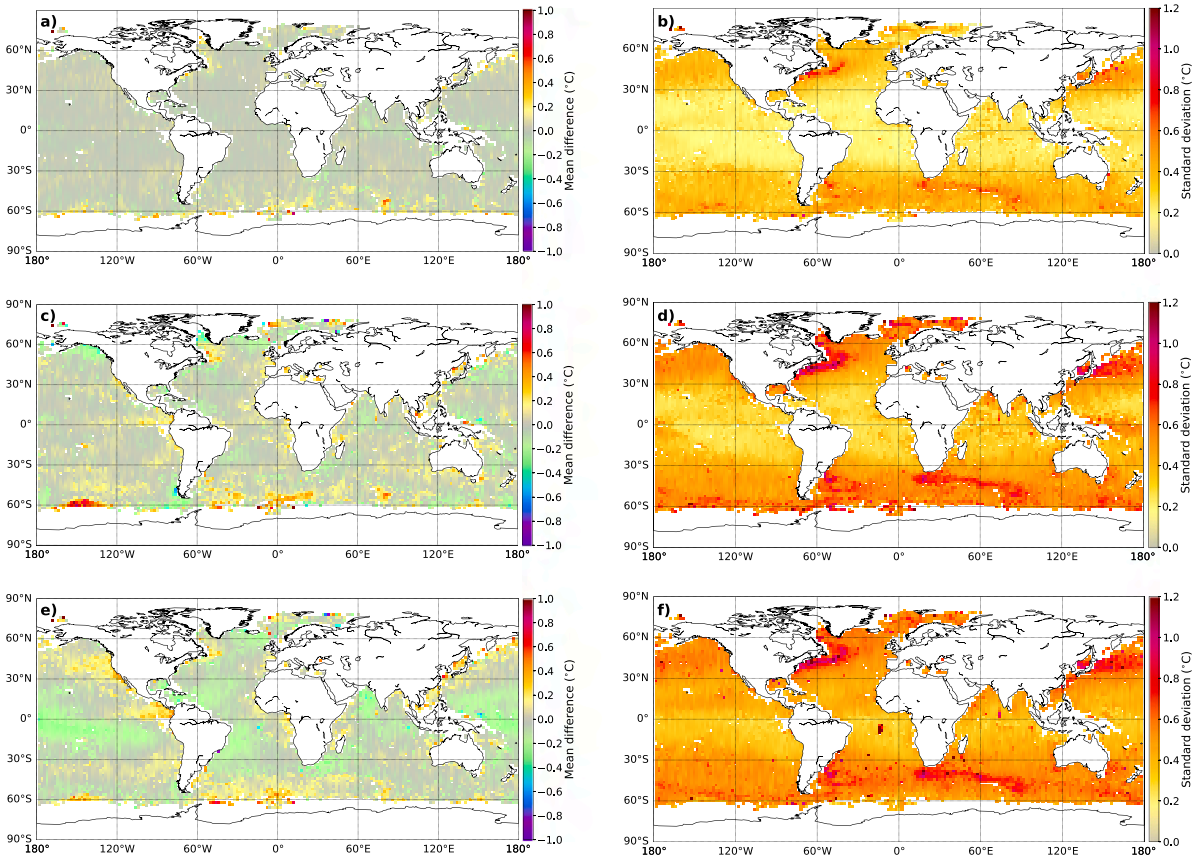


Fig. 4. The geographical distribution of mean and standard deviation of SST_r minus in situ SST for XGB (a) and b)), NN (c) and d)) and RE (e) and f)). The statistics have been calculated on a 2 × 2 degree spatial grid with a minimum of 50 matchups per grid cell.

necessarily an indication of the quality of the retrievals. Overall, the XGB model shows smaller standard deviations, whereas the magnitude of the standard deviation for the NN and RE models are more similar. However, local areas with high standard deviations are seen for the RE model. Most notably is the relatively larger area of higher standard deviation in the South Atlantic. Neither the NN nor the XGB model shows such high standard deviations for this area, although they seem to have locally slightly larger standard deviations for the same area.

4.2. Dependencies

To further investigate the performance of the three retrievals, the dependency of the retrieved minus in situ SST as a function of in situ SST and WS is shown in Fig. 5. The dependence of the retrieved SSTs on wind speed reflects the change in the sea surface roughness and hence the emissivity of the ocean. It should be noted here that the wind speed used for the XGB and NN models is the ERA5 wind speed, whereas the wind speed used for the RE model is the CCMP wind speed, which is also what was used in the RE retrieval algorithm (Alerskans et al., 2020).

The binned statistics for retrieved SST minus in situ SST as a function of in situ SST (Fig. 5a) show a warm bias for cold SSTs (SST_{insitu} < 1 °C) and a cold bias for warm SSTs (SST_{insitu} > 30 °C) for all three models. The standard deviation can be seen to decrease with increasing SST, except for very warm SSTs where a sharp increase can be seen, at least for the two ML models. This is the same interval for which the cold bias is seen. Overall, all three models show similar biases. In general, the XGB model has slightly lower standard deviation, whereas the NN and RE models both have similar standard deviations. However, for the edges of the SST interval, all three models exhibit a sharp increase in standard deviation for very cold temperatures, whereas for very warm temperatures a large increase in standard deviation is seen only for the two ML models.

Fig. 5b shows no significant dependence of the retrieved SST for the XGB and NN models as a function of WS with respect to bias. Only a small bias can be seen for high wind speeds for both models. For the RE model, on the other hand, a small bias can be seen for wind speeds of around 4–8 ms^{-1}, as well as for high wind speeds. The standard deviation increases with increasing wind speed for all three models, most notably for the NN and RE models which show standard deviations of up to almost 1 ms^{-1} for very high wind speeds. Overall, the XGB model has smaller standard deviations, with the NN and RE models exhibiting larger standard deviations.

4.3. Sensitivity

The SST sensitivity is a measure of the change in retrieved SST per unit change in the true SST (Merchant et al., 2009). Ideally, the SST sensitivity is 1 K K^{-1}, however, several geophysical factors can have an impact on the sensitivity, such as water vapour, cloud water and sea surface roughness. Here, a modified version of the forward model developed by Wentz and Meissner (2000) (Nielsen-Englyst et al., 2018) is used to estimate the SST sensitivity of each of the retrieval algorithms. The forward model relates the relevant geophysical factors to brightness temperatures, and the sensitivities to the geophysical factors show good agreement with those found by Prigent et al. (2013) using the fast radiative transfer model, RTTOV (Nielsen-Englyst et al., 2021). Two sets of brightness temperature simulations were performed for the sensitivity subset. The first set used ERA5 TCLW, TCWV and WS input together with modified drifting buoy SSTs, where 1 °C was added (SST_{+1}). The second set, on the other hand, used the same ERA5 data, but now together with modified drifting buoy SSTs where

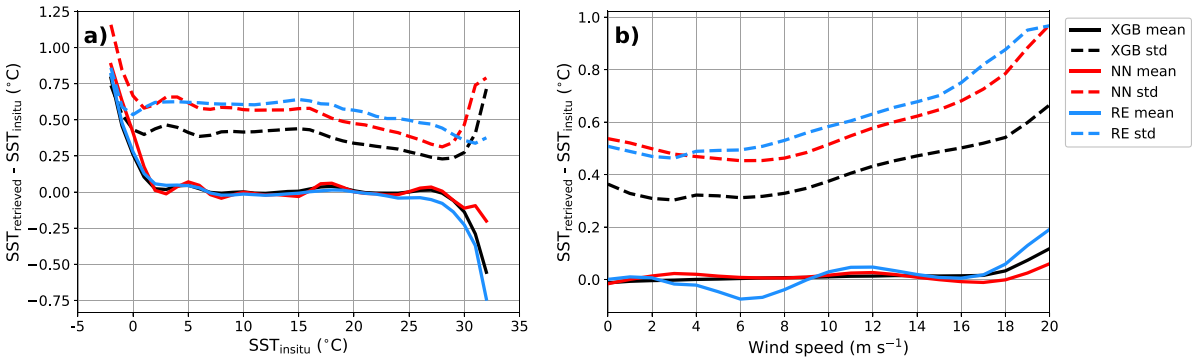


Fig. 5. Retrieved SST minus in situ SST as a function of (a) in situ SST and (b) wind speed. Solid lines show the mean and dashed lines show the standard deviation for the XGB (black), NN (red), and RE (blue) retrieval algorithms. A minimum of 50 matchups were used for the statistics calculations. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

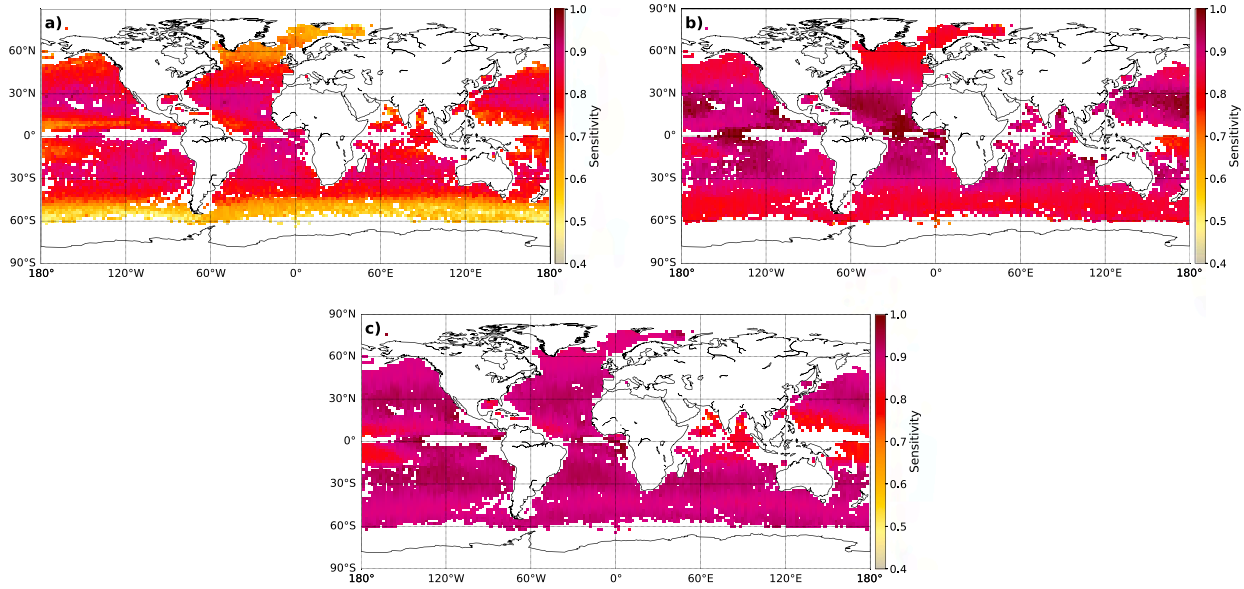


Fig. 6. The geographical distribution of sensitivity with respect to changes in SST for (a) the XGB, (b) the NN, and (c) the RE retrieval algorithms. The statistics have been calculated on a 2 × 2 degree spatial grid with a minimum of 50 matchups per grid cell.

1 °C was subtracted (SST$_{-1}$). These two subsets of simulated brightness temperatures were propagated through the retrievals to obtain new SSTs - SST$_{r,+1}$ and SST$_{r,-1}$. The sensitivity was then calculated based on these new SST retrievals, such that the sensitivity is given by (SST$_{r,+1}$ − SST$_{r,+1}$)/2, which ideally should be 1 as the two retrieved SSTs ideally should differ by 2 °C. The average sensitivity for the XGB, NN and RE models were found to be 0.78, 0.88 and 0.90, respectively.

Fig. 6 shows the geographical distribution of sensitivity for the three models. Both the XGB and NN models have higher sensitivities for lower latitudes and smaller sensitivities for higher latitudes. Areas with relatively lower sensitivities can be seen in the Pacific warm pool area as well as in the Arabian Sea. The two ML models show the same geographical patterns in the sensitivity results, however, overall the sensitivity for the NN model is higher than for the XGB model. The RE model shows some similar geographical dependencies as the other two models, such as lower sensitivities for the Pacific warm pool area and the Arabian Sea, where minimum sensitivities of 0.50 can be found. Overall, higher sensitivities are mainly found for lower latitudes, however, areas with lower sensitivities are also present at lower latitudes. Furthermore, high sensitivities are also found for higher latitudes. Hence, the same clear latitudinal pattern as for the other two models is not present for the RE model. Overall, the sensitivity of the RE model is slightly higher compared to the NN model, especially for the higher latitudes.

The dependency of the sensitivity on in situ SST is shown in Fig. 7. Here, a clear dependence can be seen, with lower sensitivities for colder SSTs and higher sensitivities for warm SSTs. However, a sharp decrease in sensitivity can be seen for very warm SSTs for all three models. The XGB model shows lowest sensitivity for all SSTs, with minimum sensitivities of less than 0.5 seen for the very cold SSTs. The NN model can be seen to have consistently higher sensitivities than the XGB model; more than 0.2 for colder SSTs where the largest difference can be seen. The RE model, however, has consistently higher sensitivities than the NN model, except for very warm SSTs. Furthermore, the RE model does not exhibit the same dependence on SST as for the other two models. Instead, a more consistent sensitivity with SST can be seen, except for very warm SSTs for which the sensitivity drops significantly.

5. Discussion

The XGB model provides the lowest bias and standard deviation with a mean bias of 0.01 K and standard deviation of 0.36 K. The NN and RE models have biases of 0.01 and −0.02 K and standard deviations of 0.50 and 0.55 K, respectively. The results obtained here for all three models are comparable to, and in the case of the XGB model even better than, previous validation results of AMSR-E PMW SST retrievals. O'Carroll et al. (2008) report a bias of 0.02 K and a standard deviation of 0.46 K, whereas Gentemann (2014) obtained

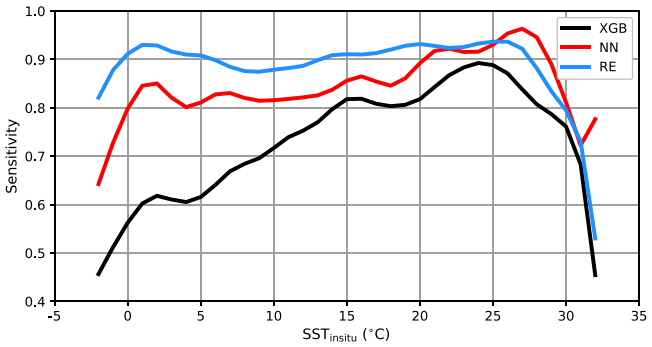


Fig. 7. Sensitivity as a function of in situ SST for the XGB (black), NN (red), and RE (blue) retrieval algorithms. A minimum of 50 matchups were used for the statistics calculations. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

retrieved SSTs with a bias of −0.05 K and a standard deviation of 0.48 K. Both Nielsen-Englyst et al. (2018) and Alerskans et al. (2020) report similar validation results with biases ±0.02 K and standard deviations of around 0.46 K.

All three retrieval algorithms exhibit warm biases for higher latitudes, especially in the Southern Ocean. The areas close to the poles have previously been confirmed to be partly related to sea ice contamination (Alerskans et al., 2020). Furthermore, it is also believed that wind induced effects could play a role in the larger biases. Fig. 5 showed an increase in bias for strong wind speeds and since the Southern Ocean is characterised by very strong wind speeds (Young, 1999) it is likely that the retrievals are effected. Furthermore, the RE model, and the NN model to some extent, have a cold bias in the Arabian Sea, as well as in the Pacific warm pool area. These regions are characterised by warm temperatures and the RE model especially has a cold bias for very warm SSTs (Fig. 5). However, neither the NN nor the XGB models show the same cold bias for the Pacific warm pool area, which indicates that the RE model is impacted by other factors and at the moment it is still unclear why this large cold bias is seen.

A latitudinal distribution of the standard deviation of the retrieved minus in situ SSTs is seen for all three models, with lower values for low latitudes and higher values for high latitudes. Gentemann (2014) reported similar results, with lowest standard deviation between 40°S and 40°N and higher values for increasing latitude. The PMW SST validation results of Alerskans et al. (2020) and Nielsen-Englyst et al. (2018) both show the same latitudinal dependence as shown in Fig. 4. The greater variability near the poles can be attributed to the presence of sea ice, which can contaminate the microwave observations (see e.g. Alerskans et al., 2020) and the fact that the sensitivity of the brightness temperature to SST is lower for colder temperatures compared to warmer temperatures (Prigent et al., 2013). Higher standard deviations are also seen for the dynamical ocean regions, with a very similar pattern reported in Nielsen-Englyst et al. (2018) and Alerskans et al. (2020). This increase is not believed to be due to the performance of the models but rather due to sampling errors in these large variability regions. The large AMSR-E footprint (43 × 75 km for the 6.9 GHz resolution) is compared against point measurements from drifting buoys. The maximum allowed geodesic distance of 20 km and maximum allowed temporal difference of 4 h between the matchups of AMSR-E and buoy data will contribute to the discrepancies seen. A high-resolution IR based SST reference has been used to calculate the variability within an AMSR-E microwave footprint and a similar pattern in the variability was found in the dynamical regions. This indicates that the increased standard deviation seen in Fig. 4 is related to the larger variability found in these regions and sampling errors, and not to poor model performances.

A higher standard deviation is also seen for the RE model for a small area in the South Atlantic Ocean. In addition, a slightly larger cold bias is seen as well. Neither the NN nor the XGB models show the same clear pattern, although for the XGB a slight increase in standard deviation can be seen. This area is a known region with strong RFI (Gentemann and Hilburn, 2015). The pre-processing of the data included an RFI mask, which has previously been applied and successfully removed RFI (Gentemann and Hilburn, 2015; Nielsen-Englyst et al., 2018). Alerskans et al. (2020) showed that it is also possible to exclude RFI contaminated matchups by comparing the baseline retrieved SSTs to additionally retrieved SSTs, for which the 10 GHz and 18 GHz channels were excluded. A similar filter could be applied here. Neither the NN nor the XGB models show the same pronounced increase in standard deviation, which indicates that they might not be as sensitive to RFI contamination. Further work is needed to investigate this.

The dependence of the retrieved minus in situ SST on in situ SST (Fig. 5) shows elevated bias and standard deviations for very warm and very cold SSTs for all three models. The behaviour for cold SSTs is believed to be partly due to sea ice contamination, as previously discussed. Furthermore, it is well-known that ML models have a hard time predicting extreme values (Ribeiro and Moniz, 2020). If not enough training data for a certain range, e.g. cold SSTs, are included, the ML models will have a hard time learning how to predict these cold SSTs. In both the NN and XGB, the optimisation is performed based on the minimisation of a loss function. As this loss function measures the average performance of the model across the domain of the target variable, the most abundant cases will have the largest impact on the model performance. Rare cases will have an almost negligible effect and the performance of the model for these cases will therefore suffer (Ribeiro and Moniz, 2020). As such, the ML models will have a hard time retrieving SSTs for the very cold and very warm SSTs, as there are not many matchups for these cases. In addition, extrapolation of predictions for data outside the training data ranges poses a problem for ML models (Xu et al., 2020). To improve the results of the ML models for the extreme ends, more training data is needed for these cases. Another option would be to modify the loss function by scaling it during the training process such that a wrong prediction of the rarer cases would have a larger impact on the model. The uneven distribution of the training data could therefore perhaps be partly offset. Yet another possibility is to train the ML models in a similar way as the RE model. By training separate instances of the ML models for e.g. the very cold and very warm SSTs a better performance might be obtained. The RE model is also seen to perform worse for the very cold and very warm SSTs. In this case, it is most likely related to the training of the model. As the last step of the RE model uses local SST and wind speed retrievals, the training data was binned into SST and wind speed bins. For the extreme ends, there are not many training examples and a minimum number of matchups for each bin was required in order to obtain robust statistics. Therefore, if there were not enough matchups in a bin, the regression coefficient from the closest SST and wind speed bin is used instead. Hence, for the very cold and very warm SSTs, as well as for the very high wind speeds, the regression coefficients are obtained from nearby bins, which might not accurately represent the relationship for the current bin.

All three models show an increase in standard deviation with wind speed, as well as a slightly higher bias for very high wind speeds. The increased uncertainty in retrieved SST for larger wind speeds is well known (see e.g. Alerskans et al. (2020)). It is likely to be related to the surface roughness and the physical characteristics that the sensitivity of the brightness temperature to wind speed increases as the wind speed (i.e. surface roughness) increases and when white foam appears on the surface (Kilic et al., 2018).

The geographical pattern of sensitivities for the NN and XGB models, with higher sensitivities for lower latitudes and lower sensitivities for higher latitudes, is similar to what was reported by Nielsen-Englyst et al. (2018) using an OE algorithm. As the sensitivity of the brightness temperatures to SST decreases with colder SSTs (Prigent et al., 2013; Nielsen-Englyst et al., 2021) higher latitudes are expected to

be associated with lower sensitivities and lower latitudes with higher sensitivities. Gentemann et al. (2009) reported sensitivities of 0.39 for SSTs of 0 °C and 0.65 for SSTs of 30 °C. The geographical distribution of sensitivity for the two ML models are therefore consistent with the expected distribution. The RE model, on the other hand, does not show such a clear latitudinal pattern. There are regions with both relatively higher and lower sensitivities found at lower latitudes. As discussed in Alerskans et al. (2020), the absence of a clear latitudinal dependence is thought to be related to the retrieval algorithm itself, more specifically to the binning performed. The RE model is trained on binned data such that separate regression coefficients are obtained for each bin. If the SST variability within a bin is small, the RE algorithm will fit very well to the SST but may experience a lower sensitivity (Alerskans et al., 2020).

It was seen in Fig. 2 that the input feature which impacts the output of the XGB model the most is, by far, tb6V. The other input features are at least an order of magnitude less important. For the NN model, tb6V is also the most important feature, however, several other features also contribute significantly to the model performance, such as tb10V, tb18V and tb23V. All of the channels have a relatively large impact on the performance of the NN model and are therefore included, whereas all channels except the 23 and 18 GHz channels are included in the XGB model. Nielsen-Englyst et al. (2021) has previously investigated the importance of different frequency channels using both a physically based and a statistically based retrieval algorithm by including different subsets of the AMSR-E frequency channels (considering the 6–36 GHz frequency range) and validating the resulting SST retrievals against independent drifting buoy observations. Nielsen-Englyst et al. (2021) found that the most important channels for SST retrievals are the 6 GHz channels, which is in agreement with the feature importance analysis for both the NN and XGB models. Following the 6 GHz channels, the 10 and 18 GHz channels were found to be the most important using both the physically and statistically based models (Nielsen-Englyst et al., 2021). The statistical algorithm showed a clear improvement in performance when more channels are included, while the physical algorithm showed less variation among the channel subsets, and it actually performed quite well by only including 6 and 10 GHz. This is similar to the XGB model, which also relies mostly on the 6 and 10 GHz channels (see Fig. 2). The NN model, on the other hand, is more similar to the regression based algorithm in the sense that it performs better when more information is included, as is evident on the more even distribution of importances.

The XGB model was found to perform best with respect to standard deviation but worst with respect to sensitivity. This might be related to a poor generalisation ability of the XGB model, which implies a problem with overfitting. However, as mentioned in Section 3.2, the XGB model is run with early stopping in order to prevent overfitting and no overfitting was observed when analysing the training and generalisation errors. To investigate this issue further, simpler XGB models could be trained and a comparison between performances with respect to standard deviation, bias and sensitivity could be made in order to see if the problem is related to overfitting. The low sensitivity might also be related to the input features used. Even though a feature importance analysis was performed in order to only select the most important input features, the exclusion of some input features might negatively influence the performance of the XGB model. Nielsen-Englyst et al. (2021) found that the 18 GHz channels were important for the both a physically based and a statistically based model. The XGB model includes neither the 18 GHz channels nor the 23 GHz channels, the latter which have been found to be sensitive to atmospheric water content (Nielsen-Englyst et al., 2021). To investigate if the inclusion of some of the excluded features could affect the performance of the model, several new instances of the XGB model could be trained where some of the now-excluded input features are included. The performance of these models could then be compared with the performance of the current XGB model, especially with respect to sensitivity.

The RE and the two ML models are all statistically based retrieval algorithms. However, the RE model used here is constrained to predefined linear relationships (although it can be expanded to include non-linearities), whereas the ML models allow non-linear relationships between input features and retrieved SST. The main advantage of ML models is that they allow approximation of complex functions as they are considered universal approximators (Hornik et al., 1990; Cybenko, 1989; Hornik, 1991). They therefore allow for the learning of new relationships without prior assumptions. This is one of the main advantages of ML-based models in comparison to more traditional regression-based algorithms. On the other hand, one of the disadvantages of ML models is related to the computational cost. The HPO of the models for example, is very computationally heavy, especially if opting for the gridded search. However, not performing an HPO can have an impact on the performance of the model, as there is no optimal model structure that suits all problems (Yang and Shami, 2020). Moreover, the training of the ML models is also more computationally expensive than the training of a linear regression based model such as the RE. Furthermore, the more complex the model, e.g. the larger the architecture and the more input features used, the slower the optimisation and training is. This not only applies to the ML models but also to the RE model. For retrieving SSTs, on the other hand, the ML models are equally as fast as the RE model.

In this study, we focused on the retrieval of PMW SSTs from AMSR-E, however, it is also possible to apply ML models to retrievals of SSTs from other satellite sensors. Initial validation results using AMSR2 show good performances with biases of 0.01 and −0.08 K and standard deviations of 0.34 and 0.44 k for an XGB and an NN model, respectively. The better validation results of AMSR2 compared to AMSR-E is in agreement with those reported in Alerskans et al. (2020). The retrieval of satellite SSTs from PMW observations using ML can also be extended to future satellite missions, such as CIMR, which is currently prepared by the ESA as a part of the Copernicus Expansion Program of the European Union (http://www.cimr.eu/; Donlon, 2020) and to the retrieval of IR satellite SSTs (Sunder et al., 2020).

The uncertainties of the retrieved SSTs have not been considered in this study but it is an important aspect that needs to be addressed in the future as uncertainties are important for a wide variety of applications, such as the use of SSTs within oceanic and atmospheric models (Merchant et al., 2017). Therefore, future work should aim at estimating the uncertainties of the retrieved SSTs for each of the two ML models. Statistical models have previously been used to estimate the uncertainty in SST retrievals, such as a regression based algorithm for the estimation of the uncertainty of the RE SSTs (Alerskans et al., 2020). More recently, Kumar et al. (2021) investigated the use of two ML models for estimating the uncertainty in satellite derived IR SSTs. Good results were obtained, showing the usefulness of ML based algorithms in uncertainty estimates. Another approach could be to train multiple algorithms to obtain an ensemble from which the uncertainties can be estimated. Future work on estimating the uncertainties of the XGB and NN SST retrievals could therefore include an investigation of these approaches.

6. Conclusions

In this study, two types of machine learning (ML) models have been assessed for the retrieval of SSTs using passive microwave (PMW) satellite observations from AMSR-E. The results have been compared with an existing state-of-the-art regression (RE) retrieval algorithm. The ML models considered were the decision tree-based algorithm Extreme Gradient Boosting (XGB) and a multilayer perceptron neural network (NN). The performance of the models was evaluated using independent in situ observations of SST from drifting buoys. The performance of the RE and NN retrieval algorithms with respect to bias and standard deviation is similar, with the NN generally performing slightly better. The XGB model performs significantly better than both the RE and

NN models with respect to standard deviation and has a similar bias. However, the SST sensitivity of both the RE and NN models is significantly higher than that of the XGB model, with the RE model having the highest sensitivity. This demonstrates the importance of including the sensitivity in the validation analysis. It is not yet understood why the XGB model performs well with respect to standard deviation but significantly worse than the other two models with respect to sensitivity. This should be further investigated, especially with respect to overfitting and selected input features.

This is an initial study meant to investigate the possibilities of using ML based algorithms for retrieval of SST from PMW observations. It shows that there is a large potential for the use of ML models but also that further work is needed in order to explore the full potential of ML based retrievals. The NN used here is a very simple form of a neural network and does not represent the full spectrum of neural networks. In order to investigate the use of neural networks for PMW SST retrievals, a study comparing different types of neural networks is needed. Similarly, more work is needed for evaluating the potential of the XGB model and other decision tree based ML models, especially with respect to sensitivity.

The main strength of ML models is that they allow for the approximation of complex functions without prior assumptions. For statistical based algorithms, such as the RE, the relationship between the input and output variables needs to be explicitly specified in the model formulation. For ML models, on the other hand, the model itself will find the best relationship between input and output variables without prior assumptions.

The ML methodology, where the algorithms select the important features based on the information in the observations and the training dataset may also be of great value in complex problems where not all physical or instrumental effects are well determined e.g. in the commissioning phase of new satellites and instruments. This initial study demonstrates that there is a large potential in the use of ML algorithms for the retrieval of SST from PMW observations.

CRediT authorship contribution statement

Emy Alerskans: Conceptualization, Methodology, Software, Investigation, Writing – original draft, Writing – review & editing. **Ann-Sofie P. Zinck:** Conceptualization, Methodology, Software, Investigation, Writing – original draft, Writing - review & editing. **Pia Nielsen-Englyst:** Software, Investigation, Writing – original draft, Writing – review & editing. **Jacob L. Høyer:** Conceptualization, Supervision, Writing – original draft, Writing – review & editing.

Declaration of competing interest

Emy Alerskans reports financial support was provided by European Space Agency. Pia Nielsen-Englyst reports financial support was provided by European Space Agency. Jacob L. Hoeyer reports financial support was provided by European Space Agency.

Data availability

ICODAS version 2.5.1 is available via https://icoads.noaa.gov/products.htmland EN4 version 4.2.0 is available at https://www.metoffice.gov.uk/hadobs/en4/. The resampled L2A data product AMSR-E V12 is produced by Remote Sensing Systems (RSS) and distributed by NASA's National Snow and Ice Data Center (NSIDC). Data are available at https://nsidc.org/data/ae_l2a. The ERA-Interim reanalysis data is available at https://www.ecmwf.int/en/forecasts/datasets/reanalysis-datasets/era-interim and the ERA5 reanalysis data is available via https://www.ecmwf.int/en/forecasts/datasets/reanalysis-datasets/era5.

Acknowledgements

The authors are grateful for computing resources and technical assistance provided by the Danish Center for Climate Computing, a facility built with support of the Danish e-Infrastructure Corporation, Danish Hydrocarbon Research and Technology Centre, VILLUM Foundation, and the Niels Bohr Institute. The PMW algorithm development was partially funded by The European Space Agency Climate Change Initiative for Sea Surface Temperature, grant 4000109848/13/I-NB.

References

Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., Corrado, G.S., Davis, A., Dean, J., Devin, M., Ghemawat, S., Goodfellow, I., Harp, A., Irving, G., Isard, M., Jia, Y., Jozefowicz, R., Kaiser, L., Kudlur, M., Levenberg, J., Mané, D., Monga, R., Moore, S., Murray, D., Olah, C., Schuster, M., Shlens, J., Steiner, B., Sutskever, I., Talwar, K., Tucker, P., Vanhoucke, V., Vasudevan, V., Viégas, F., Vinyals, O., Warden, P., Wattenberg, M., Wicke, M., Yu, Y., Zheng, X., 2015. TensorFlow: Large-scale machine learning on heterogeneous systems. Software available from tensorflow.org, URL: https://www.tensorflow.org/.

Alerskans, E., Høyer, J.L., Gentemann, C.L., Pedersen, L.T., Nielsen-Englyst, P., Donlon, C., 2020. Construction of a climate data record of sea surface temperature from passive microwave measurements. Remote Sens. Environ. 236, 111485.

Ashcroft, P., Wentz, F.J., 2013. AMSR-E/aqua L2A global swath spatially-resampled brightness temperatures, version 3. http://dx.doi.org/10.5067/AMSR-E/AE_L2A.003, Boulder, Colorado USA. NASA National Snow and Ice Data Center Distributed Active Archive Center.

Atlas, R., Hoffman, R.N., Ardizzone, J., Leidner, S.M., Jusem, J.C., Smith, D.K., Gombos, D., 2011. A cross-calibrated, multiplatform ocean surface wind velocity product for meteorological and oceanographic applications. Bull. Am. Meteorol. Soc. 92 (2), 157–174.

Azodi, C.B., Tang, J., Shiu, S.-H., 2020. Opening the black box: Interpretable machine learning for geneticists. Trends Genet. 36 (6), 442–455.

Berrar, D., 2018. Cross-validation. ISBN: 9780128096338, http://dx.doi.org/10.1016/B978-0-12-809633-8.20349-X.

Bojinski, S., Verstraete, M., Peterson, T.C., Richter, C., Simmons, A., Zemp, M., 2014. The concept of essential climate variables in support of climate research, applications, and policy. Bull. Am. Meteorol. Soc. 95 (9), 1431–1443. http://dx.doi.org/10.1175/BAMS-D-13-00047.1, URL: https://journals.ametsoc.org/view/journals/bams/95/9/bams-d-13-00047.1.xml.

Brasnett, B., Colan, D.S., 2016. Assimilating retrievals of sea surface temperature from VIIRS and AMSR2. J. Atmos. Ocean. Technol. 33 (2), 361–375.

Breiman, L., Friedman, J.H., Olshen, R.A., Stone, C.J., 1984. Classification and Regression Trees. Wadsworth International Group, Belmont, CA.

Chang, P., Jelenak, Z., Alsweiss, S., 2015. Algorithm Theoretical Basis Document: GCOM-W1/AMSR2 Day-1 EDR version 1.0.. Technical Report, URL: https://www.star.nesdis.noaa.gov/jpss/documents/ATBD/ATBD_AMSR2_Ocean_EDR_v2.0.pdf.

Chelton, D.B., Wentz, F.J., 2005. Global microwave satellite observations of sea surface temperature for numerical weather prediction and climate research. Bull. Am. Meteorol. Soc. 86 (8), 1097–1116. http://dx.doi.org/10.1175/BAMS-86-8-1097, URL: https://journals.ametsoc.org/view/journals/bams/86/8/bams-86-8-1097.xml.

Chen, T., Guestrin, C., 2016. XGBoost: A scalable tree boosting system. pp. 785–794. http://dx.doi.org/10.1145/2939672.2939785.

Cybenko, G., 1989. Approximation by superpositions of a sigmoidal function. Math. Control Signals Systems 2 (4), 303–314.

Dee, D.P., Uppala, S.M., Simmons, A., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M., Balsamo, G., Bauer, d.P., et al., 2011. The ERA-Interim reanalysis: Configuration and performance of the data assimilation system. Q. J. R. Meteorol. Soc. 137 (656), 553–597.

Donlon, C., 2020. Copernicus imaging microwave radiometer (CIMR) mission requirements document, version 4. European Space Agency, Noordwijk, The Netherlands.

Esposito, F., Malerba, D., Semeraro, G., Kay, J., 1997. A comparative analysis of methods for pruning decision trees. IEEE Trans. Pattern Anal. Mach. Intell. 19 (5), 476–491.

Ferreira, P., Le, D.C., Zincir-Heywood, N., 2019. Exploring feature normalization and temporal information for machine learning based insider threat detection. In: 2019 15th International Conference on Network and Service Management. CNSM, IEEE, pp. 1–7.

Friedman, J.H., 2001. Greedy function approximation: A gradient boosting machine. Ann. Statist. 1189–1232.

Gentemann, C.L., 2014. Three way validation of MODIS and AMSR-E sea surface temperatures. J. Geophys. Res. Oceans 119 (4), 2583–2598.

Gentemann, C.L., Hilburn, K.A., 2015. In situ validation of sea surface temperatures from the GCOM-w 1 AMSR 2 RSS calibrated brightness temperatures. J. Geophys. Res. Oceans 120 (5), 3567–3585.

Gentemann, C.L., Meissner, T., Wentz, F.J., 2009. Accuracy of satellite sea surface temperatures at 7 and 11 GHz. IEEE Trans. Geosci. Remote Sens. 48 (3), 1009–1018.

Gentemann, C.L., Wentz, F.J., Brewer, M., Hilburn, K., Smith, D., 2010. Passive microwave remote sensing of the ocean: An overview. Oceanogr. Space 13–33.

Good, S.A., Martin, M.J., Rayner, N.A., 2013. EN4: Quality controlled ocean temperature and salinity profiles and monthly objective analyses with uncertainty estimates. J. Geophys. Res. Oceans 118 (12), 6704–6716.

Goodfellow, I., Bengio, Y., Courville, A., 2016. Deep Learning. MIT Press, http://www.deeplearningbook.org.

Grimm, K.J., Mazza, G.L., Davoudzadeh, P., 2017. Model selection in finite mixture models: A k-fold cross-validation approach. Struct. Equ. Model. A Multidisciplinary Journal 24 (2), 246–256. http://dx.doi.org/10.1080/10705511.2016.1250638.

Haykin, S., 1999. Neural Networks: A Comprehensive Foundation. Prentice hall.

Hecht-Nielsen, R., 1992. Theory of the backpropagation neural network. In: Neural Networks for Perception. Elsevier, pp. 65–93.

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Mu noz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., et al., 2020. The ERA5 global reanalysis. Q. J. R. Meteorol. Soc. 146 (730), 1999–2049.

Hornik, K., 1991. Approximation capabilities of multilayer feedforward networks. Neural Netw. 4 (2), 251–257.

Hornik, K., Stinchcombe, M., White, H., 1990. Universal approximation of an unknown mapping and its derivatives using multilayer feedforward networks. Neural Netw. 3 (5), 551–560.

Huang, L., Qin, J., Zhou, Y., Zhu, F., Liu, L., Shao, L., 2020. Normalization techniques in training dnns: Methodology, analysis and application. arXiv preprint arXiv:2009.12836.

Just, A.C., De Carli, M.M., Shtein, A., Dorman, M., Lyapustin, A., Kloog, I., 2018. Correcting measurement error in satellite aerosol optical depth with machine learning for modeling PM2. 5 in the Northeastern USA. Remote Sens. 10 (5), 803.

Just, A.C., Liu, Y., Sorek-Hamer, M., Rush, J., Dorman, M., Chatfield, R., Wang, Y., Lyapustin, A., Kloog, I., 2020. Gradient boosting machine learning to improve satellite-derived column water vapor measurement error. Atmos. Meas. Tech. 13 (9), 4669–4681.

Kawanishi, T., Sezai, T., Ito, Y., Imaoka, K., Takeshima, T., Ishido, Y., Shibata, A., Miura, M., Inahata, H., Spencer, R., 2003. The advanced microwave scanning radiometer for the earth observing system (AMSR-E), NASDA's contribution to the EOS for global energy and water cycle studies. IEEE Trans. Geosci. Remote Sens. 41 (2), 184–194. http://dx.doi.org/10.1109/TGRS.2002.808331.

Kilic, L., Prigent, C., Aires, F., Boutin, J., Heygster, G., Tonboe, R.T., Roquet, H., Jimenez, C., Donlon, C., 2018. Expected performances of the copernicus imaging microwave radiometer (CIMR) for an all-weather and high spatial resolution estimation of ocean and sea ice parameters. J. Geophys. Res. Oceans 123 (10), 7564–7580.

Kotsiantis, S.B., Kanellopoulos, D., Pintelas, P.E., 2006. Data preprocessing for supervised leaning. Int. J. Comput. Sci. 1 (2), 111–117.

Kumar, C., Podestá, G., Kilpatrick, K., Minnett, P., 2021. A machine learning approach to estimating the error in satellite sea surface temperature retrievals. Remote Sens. Environ. 255, 112227.

Le Traon, P.-Y., Antoine, D., Bentamy, A., Bonekamp, H., Breivik, L., Chapron, B., Corlett, G., Dibarboure, G., DiGiacomo, P., Donlon, C., et al., 2015. Use of satellite observations for operational oceanography: recent achievements and future prospects. J. Oper. Oceanogr. 8 (sup1), s12–s27.

LeCun, Y., Bengio, Y., Hinton, G., 2015. Deep learning. Nature 521 (7553), 436–444.

Lee, A., Taylor, P., Kalpathy-Cramer, J., Tufail, A., 2017. Machine learning has arrived!. Ophthalmology 124, 1726–1728. http://dx.doi.org/10.1016/j.ophtha.2017.08.046.

Liang, X., Yang, Q., Nerger, L., Losa, S.N., Zhao, B., Zheng, F., Zhang, L., Wu, L., 2017. Assimilating copernicus SST data into a pan-arctic ice–ocean coupled model with a local SEIK filter. J. Atmos. Ocean. Technol. 34 (9), 1985–1999.

Liashchynskyi, P., Liashchynskyi, P., 2019. Grid search, random search, genetic algorithm: A big comparison for NAS. CoRR abs/1912.06059, arXiv:1912.06059.

Liu, H., Li, Q., Bai, Y., Yang, C., Wang, J., Zhou, Q., Hu, S., Shi, T., Liao, X., Wu, G., 2021. Improving satellite retrieval of oceanic particulate organic carbon concentrations using machine learning methods. Remote Sens. Environ. 256, 112316.

Liu, Y., Xia, X., Yao, L., Jing, W., Zhou, C., Huang, W., Li, Y., Yang, J., 2020. Downscaling satellite retrieved soil moisture using regression tree-based machine learning algorithms over Southwest France. Earth Space Sci. 7 (10), e2020EA001267.

Lundberg, S.M., Lee, S.-I., 2017. A unified approach to interpreting model predictions. In: Guyon, I., Luxburg, U.V., Bengio, S., Wallach, H., Fergus, R., Vishwanathan, S., Garnett, R. (Eds.), Advances in Neural Information Processing Systems, Vol. 30. Curran Associates, Inc., pp. 4765–4774, URL: http://papers.nips.cc/paper/7062-a-unified-approach-to-interpreting-model-predictions.pdf.

Maeda, T., Taniguchi, Y., Imaoka, K., 2015. GCOM-W1 AMSR2 level 1R product: Dataset of brightness temperature modified using the antenna pattern matching technique. IEEE Trans. Geosci. Remote Sens. 54 (2), 770–782.

Maeda, T., Tomii, N., Seki, M., Sekiya, K., Shibata, A., 2020. Sea-surface-temperature retrieval at higher spatial resolution in the satellite-Borne microwave radiometer AMSR2 follow-on mission. IEEE Geosci. Remote Sens. Lett. 18 (2), 336–340.

Meissner, T., Wentz, F.J., 2012. The emissivity of the ocean surface between 6 and 90 GHz over a large range of wind speeds and earth incidence angles. IEEE Trans. Geosci. Remote Sens. 50 (8), 3004–3026.

Merchant, C.J., Embury, O., Bulgin, C.E., Block, T., Corlett, G.K., Fiedler, E., Good, S.A., Mittaz, J., Rayner, N.A., Berry, D., et al., 2019. Satellite-based time-series of sea-surface temperature since 1981 for climate applications. Sci. Data 6 (1), 1–18.

Merchant, C., Embury, O., Le Borgne, P., Bellec, B., 2006. Saharan dust in nighttime thermal imagery: Detection and reduction of related biases in retrieved sea surface temperature. Remote Sens. Environ. 104 (1), 15–30.

Merchant, C., Harris, A., Murray, M., Zavody, A., 1999. Toward the elimination of bias in satellite retrievals of sea surface temperature: 1. Theory, modeling and interalgorithm comparison. J. Geophys. Res. Oceans 104 (C10), 23565–23578.

Merchant, C., Harris, A., Roquet, H., Le Borgne, P., 2009. Retrieval characteristics of non-linear sea surface temperature from the advanced very high resolution radiometer. Geophys. Res. Lett. 36 (17).

Merchant, C.J., Paul, F., Popp, T., Ablain, M., Bontemps, S., Defourny, P., Hollmann, R., Lavergne, T., Laeng, A., De Leeuw, G., et al., 2017. Uncertainty information in climate data records from earth observation. Earth Syst. Sci. Data 9 (2), 511–527.

Merchant, C.J., Saux-Picart, S., Waller, J., 2020. Bias correction and covariance parameters for optimal estimation by exploiting matched in-situ references. Remote Sens. Environ. 237, 111590.

Minnett, P., Alvera-Azcárate, A., Chin, T., Corlett, G., Gentemann, C., Karagali, I., Li, X., Marsouin, A., Marullo, S., Maturi, E., et al., 2019. Half a century of satellite remote sensing of sea-surface temperature. Remote Sens. Environ. 233, 111366.

Monzikova, A., Kudryavtsev, V., Reul, N., Chapron, B., 2017. On the upper ocean response to tropical cyclones: Satellite microwave observation. In: 2017 Progress in Electromagnetics Research Symposium-Fall. PIERS-FALL, IEEE, pp. 2437–2444.

Moschos, E., Schwander, O., Stegner, A., Gallinari, P., 2020. Deep-SST-Eddies: A deep learning framework to detect oceanic eddies in sea surface temperature images. In: ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing. ICASSP, IEEE, pp. 4307–4311.

Nielsen, M.A., 2015. Neural Networks and Deep Learning, Vol. 25. Determination press San Francisco, CA.

Nielsen-Englyst, P., Høyer, J.L., Alerskans, E., Pedersen, L.T., Donlon, C., 2021. Impact of channel selection on SST retrievals from passive microwave observations. Remote Sens. Environ. 254, 112252.

Nielsen-Englyst, P., L Høyer, J., Toudal Pedersen, L., L Gentemann, C., Alerskans, E., Block, T., Donlon, C., 2018. Optimal estimation of sea surface temperature from AMSR-E. Remote Sens. 10 (2), 229.

Ning, J., Xu, Q., Wang, T., Zhang, S., 2018. Upper ocean response to super typhoon soudelor revealed by different SST products. In: IGARSS 2018-2018 IEEE International Geoscience and Remote Sensing Symposium. IEEE, pp. 6063–6066.

O'Carroll, A.G., Armstrong, E.M., Beggs, H.M., Bouali, M., Casey, K.S., Corlett, G.K., Dash, P., Donlon, C.J., Gentemann, C.L., Hø yer, J.L., et al., 2019. Observational needs of sea surface temperature. Front. Mar. Sci. 6, 420.

O'Carroll, A.G., Eyre, J.R., Saunders, R.W., 2008. Three-way error analysis between AATSR, AMSR-E, and in situ sea surface temperature observations. J. Atmos. Ocean. Technol. 25 (7), 1197–1207.

Paul, S., Huntemann, M., 2021. Improved machine-learning-based open-water–sea-ice–cloud discrimination over wintertime antarctic sea ice using MODIS thermal-infrared imagery. Cryosphere 15 (3), 1551–1565.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, E., 2011. Scikit-learn: Machine learning in Python. J. Mach. Learn. Res. 12, 2825–2830.

Prigent, C., Aires, F., Bernardo, F., Orlhac, J.-C., Goutoule, J.-M., Roquet, H., Donlon, C., 2013. Analysis of the potential and limitations of microwave radiometry for the retrieval of sea surface temperature: Definition of MICROWAT, A new mission concept. J. Geophys. Res. Oceans 118 (6), 3074–3086.

Prochaska, J.X., Cornillon, P.C., Reiman, D.M., 2021. Deep learning of sea surface temperature patterns to identify ocean extremes. Remote Sens. 13 (4), 744.

Rayner, N.A., Brohan, P., Parker, D., Folland, C.K., Kennedy, J.J., Vanicek, M., Ansell, T., Tett, S., 2006. Improved analyses of changes and uncertainties in sea surface temperature measured in situ since the mid-nineteenth century: The HadSST2 dataset. J. Clim. 19 (3), 446–469.

Ribeiro, R.P., Moniz, N., 2020. Imbalanced regression and extreme value prediction. Mach. Learn. 109 (9), 1803–1835.

Ribeiro, M.T., Singh, S., Guestrin, C., 2016. " Why should i trust you?" Explaining the predictions of any classifier. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. pp. 1135–1144.

Rodriguez-Fernandez, N.J., Aires, F., Richaume, P., Kerr, Y.H., Prigent, C., Kolassa, J., Cabot, F., Jimenez, C., Mahmoodi, A., Drusch, M., 2015. Soil moisture retrieval using neural networks: Application to SMOS. IEEE Trans. Geosci. Remote Sens. 53 (11), 5991–6007.

Sanò, P., Panegrossi, G., Casella, D., Marra, A.C., D'Adderio, L.P., Rysman, J.F., Dietrich, S., 2018. The passive microwave neural network precipitation retrieval (PNPR) algorithm for the CONICAL scanning global microwave imager (GMI) radiometer. Remote Sens. 10 (7), 1122.

Sanò, P., Panegrossi, G., Casella, D., Marra, A.C., Paola, F.D., Dietrich, S., 2016. The new passive microwave neural network precipitation retrieval (PNPR) algorithm for the cross-track scanning ATMS radiometer: Description and verification study over europe and africa using GPM and TRMM spaceborne radars. Atmos. Meas. Tech. 9 (11), 5441–5460.

Saux Picart, S., Tandeo, P., Autret, E., Gausset, B., 2018. Exploring machine learning to correct satellite-derived sea surface temperatures. Remote Sens. 10 (2), 224.

Shapley Ll, S., 1953. A value for n-person games. In: Contributions to the Theory of Games II, Annals of Mathematical Studies, Vol. 28.

Shibata, A., 2006. Features of ocean microwave emission changed by wind at 6 GHz. J. Oceanogr. 62 (3), 321–330.

Shrikumar, A., Greenside, P., Kundaje, A., 2017. Learning important features through propagating activation differences. In: International Conference on Machine Learning. PMLR, pp. 3145–3153.

Shrikumar, A., Greenside, P., Shcherbina, A., Kundaje, A., 2016. Not just a black box: Learning important features through propagating activation differences. arXiv preprint arXiv:1605.01713.

Sismanidis, P., Bechtel, B., Keramitsoglou, I., Göttsche, F., Kiranoudis, C.T., 2021. Satellite-derived quantification of the diurnal and annual dynamics of land surface temperature. Remote Sens. Environ. 265, 112642.

Sunder, S., Ramsankaran, R., Ramakrishnan, B., 2020. Machine learning techniques for regional scale estimation of high-resolution cloud-free daily sea surface temperatures from MODIS data. ISPRS J. Photogramm. Remote Sens. 166, 228–240.

Wentz, F.J., Meissner, T., 2000. Algorithm Theoretical Basis Document(ATBD): AMSR Ocean Algorithm (Version 2). RSS Tech. Proposal 121599A-1, Remote Sensing Systems, Santa Rosa, CA.

Wentz, F.J., Meissner, T., 2007. Supplement 1: Algorithm theoretical basis document for AMSR-E ocean algorithms. 30, NASA: Santa Rosa, CA, USA.

Woodruff, S.D., Worley, S.J., Lubker, S.J., Ji, Z., Eric Freeman, J., Berry, D.I., Brohan, P., Kent, E.C., Reynolds, R.W., Smith, S.R., et al., 2011. ICOADS Release 2.5: Extensions and enhancements to the surface marine meteorological archive. Int. J. Climatol. 31 (7), 951–967.

Xu, K., Zhang, M., Li, J., Du, S.S., Kawarabayashi, K.-i., Jegelka, S., 2020. How neural networks extrapolate: From feedforward to graph neural networks. arXiv preprint arXiv:2009.11848.

Yang, C.-S., Kim, S.-H., Ouchi, K., Back, J.-H., 2015. Generation of high resolution sea surface temperature using multi-satellite data for operational oceanography. Acta Oceanol. Sinica 34 (7), 74–88.

Yang, L., Shami, A., 2020. On hyperparameter optimization of machine learning algorithms: Theory and practice. Neurocomputing 415, 295–316. http://dx.doi.org/10.1016/j.neucom.2020.07.061.

Ying, X., 2019. An overview of overfitting and its solutions. J. Phys. Conf. Ser. 1168, 022022. http://dx.doi.org/10.1088/1742-6596/1168/2/022022.

Young, I., 1999. Seasonal variability of the global ocean wind and wave climate. Int. J. Climatol.: J. Royal Meteorol. Soc. 19 (9), 931–950.

Zhang, C., Liu, C., Zhang, X., Almpanidis, G., 2017. An up-to-date comparison of state-of-the-art classification algorithms. Expert Syst. Appl. 82, 128–150. http://dx.doi.org/10.1016/j.eswa.2017.04.003, URL: https://www.sciencedirect.com/science/article/pii/S0957417417302397.

APPENDIX C

Paper III: Impact of channel selection on SST retrievals from passive microwave observations

This appendix contains a full copy of Paper III:

- **Nielsen-Englyst, Pia**, Jacob L. Høyer, Emy Alerskans, Leif Toudal Pedersen, Craig Donlon. 2021. "Impact of channel selection on SST retrievals from passive microwave observations", *Remote Sensing of Environment*, Volume 254, 112252, ISSN 0034-4257.

Remote Sensing of Environment 254 (2021) 112252

Contents lists available at ScienceDirect

Remote Sensing of Environment

journal homepage: www.elsevier.com/locate/rse

Impact of channel selection on SST retrievals from passive microwave observations

Pia Nielsen-Englyst [a,b,*], Jacob L. Høyer [b], Emy Alerskans [c,b], Leif Toudal Pedersen [a], Craig Donlon [d]

[a] DTU-Space, Technical University of Denmark, DK-2800 Lyngby, Denmark
[b] Danish Meteorological Institute, Lyngbyvej 100, DK-2100 Copenhagen Ø, Denmark
[c] University of Copenhagen, Tagensvej 16, DK-2200 København N, Denmark
[d] European Space Agency/European Space Research and Technology Centre (ESA/ESTEC), 2201 AZ Noordwijk, the Netherlands

ARTICLE INFO

Keywords:
Remote sensing
Passive microwaves
Channel selection
Copernicus imaging microwave radiometer (CIMR)
AMSR-E
Sea surface temperature (SST)
Optimal estimation (OE)

ABSTRACT

Two retrieval algorithms developed as a part of the European Space Agency Climate Change Initiative (ESA-CCI) project are used to assess the effects of withholding observations from selected frequency channels on the retrieved subskin Sea Surface Temperature (SST) from AQUA's Advanced Microwave Scanning Radiometer—Earth Observing System (AMSR-E) and to evaluate a Copernicus Imaging Microwave Radiometer (CIMR) like channel configuration.

The first algorithm is a statistical regression-based retrieval algorithm, while the second is a physically based optimal estimation (OE) algorithm. A database with matching satellite and drifting buoy observations is used to test the performance of each channel configuration using both retrieval algorithms to identify the most optimal channel selection for accurate SST retrievals. The evaluation against in situ observations allows identification of the strengths and weaknesses of the two retrieval algorithms, and demonstrates the importance of using in situ observations to evaluate existing theoretical retrieval uncertainty studies. Overall, the performance increases as expected when more channels are included in the retrieval. In particular, more channels allow a better performance for the range of different observing conditions (e.g. cold waters). The two retrieval algorithms agree that for a three-channel configuration, the 6, 10, 18 GHz (V and H polarization) is better than the 6, 10, 23 GHz configuration (V and H polarization). This is demonstrated for different geographical regions and throughout all seasons. Of the different combinations tested here, it is evident that withholding observations from the 23 and 36 GHz channels from the retrieval has the least impact on the SST performance. Overall, this analysis shows that the CIMR like channel configuration performs very well when compared to an AMSR-E like constellation using both retrieval algorithms.

1. Introduction

Global sea surface temperature (SST) observations are crucial for climate monitoring (e.g. Merchant et al., 2019), numerical weather prediction (NWP; Brasnett and Colan, 2016; Chelton and Wentz, 2005), ocean and coupled models (Liang et al., 2017; Le Traon et al., 2015; Yang et al., 2015) and in understanding and predicting the state of the ocean and the atmosphere (Gentemann et al., 2003; Monzikova et al., 2017; Ning et al., 2018; O'Carroll et al., 2019). Global sampling of SST is only possible through satellite remote sensing, and SST was one of the first ocean variables to be observed from earth observing satellites (see e.g. Minnett et al. (2019) for a detailed overview of the latest progress in satellite SST retrievals). SST has been derived from infrared (IR) satellite observations since 1981, with a typical spatial resolution of 1–4 km (Embury et al., 2012; Gladkova et al., 2016; Merchant et al., 2019; Reynolds et al., 2002). However, the IR observations are biased from aerosols and the temporal/spatial sampling is problematic in regions with persistent cloud cover (such as the Polar and Tropical regions), since IR wavelengths are unable to observe the surface through clouds (Merchant et al., 1999, 2006; Vázquez-Cuervo et al., 2004).

Observations from passive microwave (PMW) sensors are important alternatives to IR observations as PMW measurements are not prevented

* Corresponding author at: DTU-Space, Technical University of Denmark, DK-2800 Lyngby, Denmark.
E-mail address: pne@dmi.dk (P. Nielsen-Englyst).

https://doi.org/10.1016/j.rse.2020.112252
Received 28 April 2020; Received in revised form 26 November 2020; Accepted 9 December 2020
Available online 25 December 2020

by non-precipitating clouds or biased by aerosols (Donlon et al., 2007, 2010; Ulaby et al., 1981; Wentz et al., 2000). The first space borne microwave radiometer capable of measuring SST was the Scanning Multichannel Microwave Radiometer (SMMR) carried on Seasat 1 and Nimbus 7, both launched in 1978 (Lipes, 1982; Milman and Wilheit, 1985). However, SST retrievals from Nimbus-7 SMMR suffered from significant calibration problems, resulting in SST errors as high as 1.12 °C (Milman and Wilheit, 1985) as well as very coarse spatial resolution of approx. 150 km, limiting its usefulness. The high quality PMW SST record began in 1997 after the launch of the Tropical Rainfall Measuring Mission (TRMM) Microwave Imager (TMI; e.g. Kummerow et al., 1998; Wentz, 2015; Wentz et al., 2000) (10 GHz channel), which provided SST observations from 1998 to 2014 between 40°N/S. The TRMM mission was followed by the Global Precipitation Mission (GPM) Microwave Imager (GMI; Bidwell et al., 2005; Draper et al., 2015), which was launched in 2014, and uses a larger antenna compared to TMI resulting in a better spatial resolution. The first global PMW SST data became available in 2002 by AQUA's Advanced Microwave Scanning Radiometer – Earth Observing System (AMSR-E; e.g. Chelton and Wentz, 2005; Kawanishi et al., 2003), using the channels at 6 GHz. The instrument ceased normal operations in October 2011 and was followed by the currently flying Advanced Microwave Radiometer 2 (AMSR2), flown on the Global Change Observing Mission 1st – Water (GCOM-W1), which was launched in 2012 (Imaoka et al., 2010). AMSR2 has a larger antenna compared to AMSR-E, which provides improved spatial resolution and the additional channel at 7 GHz allows for better identification of Radio Frequency Interference (RFI). PMW SST estimates using the 6.9 and 10.7 GHz channels have a typical spatial sampling of about 10 km, a resolution of 50–60 km and an uncertainty of about 0.4 °C (Gentemann, 2014; Gentemann et al., 2010; Nielsen-Englyst et al., 2018; Wentz et al., 2000). The uncertainty is mainly attributed to mismatches in time and space of observations, differences in measurement depth, and errors in satellite estimates and in situ observations.

The current microwave imagers do not capture the subscale to mesoscale variability and are contaminated near (~100 km) coasts and in the marginal ice zone at the 6.9 GHz channels, due to the large field of view (e.g. AMSR2 has a 64 × 32 km field of view at 6.9 GHz). The spatial resolution of SST retrievals is limited by the ratio of the wavelength of measurement to the antenna diameter, the satellite altitude and the incidence angle (Wentz and Meissner, 2000). A larger antenna is therefore required to obtain a high spatial resolution (e.g. for a resolution of ~10 km at 10 GHz an antenna of ~6 m is required assuming a satellite altitude of 830 km. Improving the spatial resolution of the measurements obtained by the 6.9 and 10.7 GHz channels could lead to substantial improvements of PMW SST estimates and their information content in global products and regional analysis systems.

A sustained continuity of PMW imagers, which can provide measurements at higher spatial resolution and with high radiometric fidelity to retrieve SSTs that meet the current operational needs, is a major concern within the SST community (O'Carroll et al., 2019). The Chinese Microwave Radiometer Imager onboard the HaiYang-2B (HY-2B) is a future PMW mission with a 6.9 GHz channel (Zhang et al., 2020), while the FengYun-3 series (FY-3D and FY-3F) also has a 10.65 GHz channel. However, their spatial resolutions are three times lower compared to AMSR2. In addition, the AMSR2 follow-on mission (AMSR3) is officially approved as a project to be installed on the Global Observation SATellite for Greenhouse gases and Water Cycle (GOSAT-GW) to be launched in 2023 by JAXA (Hirabayashi, 2020). AMSR3 will be of almost equivalent capability to that of AMSR2, but with addition of high-frequency channels (166 and 183 GHz) (Hirabayashi, 2020).

A new radiometer, the Copernicus Imaging Microwave Radiometer (CIMR), is currently being studied by the European Space Agency (ESA) for the Copernicus Expansion program of the European Union (http://www.cimr.eu/). CIMR is a polar mission, designed to observe all-weather, high-resolution, high-accuracy, sub-daily observations of SST and sea ice as the primary variables (Donlon, 2019). It is the intention to include the radiometer channels 6.9, 10.7, 18.7 and 36.6 GHz common with AMSR-E and AMSR2, but also the 1.4 GHz channel, providing continuity to the current Soil Moisture and Ocean Salinity (SMOS; Kerr et al., 2010) and the Soil Moisture Active Passive (SMAP; Fore et al., 2016) missions. The anticipated spatial resolution of CIMR is shown in Table 1 and compared to the spatial resolution of AMSR-E. The higher spatial resolution at the lower frequency channels enables retrievals of SST and other surface parameters at a higher resolution and lower uncertainty than possible with the current missions.

Studies have been conducted to assess the expected impact of CIMR observations on retrievals and applications for: sea ice, snow depth and sea surface salinity (Braakmann-Folgmann and Donlon, 2019; Ciani et al., 2019; Scarlat et al., 2020). Theoretical information content studies have been made with the aim of identifying the role of different channel combinations and the expected CIMR performance for SST (Kilic et al., 2018, 2019; Pearson et al., 2018, 2019; Prigent et al., 2013). Kilic et al. (2018) derive a mean global SST retrieval uncertainty of 0.2 °C at a spatial resolution of 15 km using the intended CIMR channel configuration, which is a clear improvement in performance compared to AMSR2. They also show that the estimated CIMR retrieval uncertainty varies strongly with SST (from 0.15 °C for warm SSTs and up to 0.45 °C for cold SSTs) and wind speed (larger uncertainties for higher wind speeds), but is only weakly impacted by the total column water vapor content and total cloud liquid water content. However, a comparison of the theoretical retrieval error standard deviations obtained for AMSR2 by Kilic et al. (2018) (Figs. 5 and 6) and the standard deviations of retrieved versus drifting buoy SSTs in Alerskans et al. (2020) (Figs. 7 and 8) shows significant differences between the theoretical and the observed uncertainties. These differences can arise due to the different retrieval methods (physical versus statistical) or due to limitations in the forward models ability to simulate the real conditions. Kilic et al. (2019) compared some of the ocean forward models used in the theoretical studies and found large differences between the models and significant disagreement with observations, in particular for high wind speeds and cold SSTs.

With the discrepancies among the forward models, it is thus important to apply several types of retrieval algorithms for different channel configurations and to evaluate them against a common in situ reference data set to supplement the existing theoretical studies. The aim of this study is to determine the impact of using different channel selections on retrieved SST and to assess the realistic performances of a physically based optimal estimation (OE) algorithm and a statistical regression (RE) based algorithm against in situ observations. The use of in situ observations can inform us about deficiencies and strengths of different types of retrieval algorithms as well as evaluate the impact of different channel combinations. In addition to the preparation for the CIMR mission, it is also important to investigate the effects of different channel combinations in a realistic scenario, where future ITU frequency allocations may contaminate some of the specific frequency bands used for PMW observations of the ocean. Furthermore, this study can also be used to assess the impact during an instrument failure, where the detector or electronics for one or several channels fail during operations. Finally, the results can aid in the construction of long-term climate data records (CDRs) and harmonization of PMW satellite missions with

Table 1
The frequency channels of the AMSR-E and CIMR mission and their respective spatial resolutions.

Frequency (GHz)		1.4	6.9	10.6	18.7	23.8	36.5	89.0
Spatial Resolution (3 dB footprint size) (km x km)	AMSR-E	–	75 × 43	51 × 29	27 × 16	32 × 18	14 × 8	6 × 4
	CIMR	64 × 36	19 × 11	13 × 7	6 × 4	–	5 × 3	–

different channel configurations.

For the investigations conducted here, a physical and a statistical retrieval algorithm will be applied, where the physically based algorithm is a type of retrieval algorithm similar to the ones used for the information content studies (Kilic et al., 2018, 2019; Pearson et al., 2018; Prigent et al., 2013) and the statistical algorithm resembles the retrievals currently used to generate many existing PMW satellite SST products (Alerskans et al., 2020; Chang et al., 2015; Gentemann et al., 2010; Shibata, 2006; Wentz and Meissner, 2007).

The two different retrieval algorithms used here are developed within the ESA Climate Change Initiative (CCI) for SST (Merchant et al., 2014). The first algorithm is an optimal estimation algorithm (Nielsen-Englyst et al., 2018), which inverts a forward model to retrieve different geophysical parameters including SST from AMSR-E observations. The second retrieval algorithm is a purely statistical retrieval, consisting of a regression model, which retrieves wind speeds and SSTs using AMSR brightness temperatures (Alerskans et al., 2020). In this study, we test different subsets of the channels available in AMSR-E, which means that we do not consider the 1.4 GHz channels, the actual noise equivalent differential temperatures (NEdTs) from CIMR and the higher spatial footprints. However, the results give valuable insight to the relative performance of the different channel combinations, and the differences between the two different types of retrieval algorithms for real conditions using a well-known instrument. This study is therefore an important extension to existing theoretical studies that investigate and prepare for new PMW missions like the CIMR.

The paper is structured such that Section 2 includes a description of the matchup database used for algorithm development and validation. The retrieval algorithms are described in Section 3.1 while Section 3.2 describes the different channel selections and filters, which have been tested. The retrieval results from using different channel selections in both retrieval algorithms are presented in Section 4. Section 5 contains a discussion and Section 6 presents recommendations for future work. Finally, conclusions are given in Section 7.

2. Data

2.1. ESA CCI multisensor matchup dataset (MMD)

This work is built on the ESA CCI Multi-sensor Matchup Dataset (MMD), which is described and used in Nielsen-Englyst et al. (2018) and Alerskans et al. (2020). It includes brightness temperatures from the Level 2A (L2A) swath data product AMSR-E V12 (Ashcroft and Wentz, 2013), with the spatial resampling to the 6.9 GHz resolution (75 × 43 km). The L2A AMSR-E V12 product also includes a satellite scan quality flag, channel quality flags and the incidence- and azimuth angles. The brightness temperatures have been matched with quality controlled in situ measurements from the International Comprehensive Ocean-Atmosphere Dataset (ICOADS) version 2.5.1 (Woodruff et al., 2011), the Met Office Hadley Centre (MOHC) Ensembles dataset version 4.2.0 (EN4; Good et al., 2013) by requiring a maximal geodesic distance of 20 km and a maximal time difference of 4 h. The 20 km was chosen to ensure that the in situ measurement is located within an AMSR-E footprint, while the maximal time difference balances the need for accurate collocated data with the need for a sufficient number of useable matches. The MMD also includes SST, Total Column Water Vapor (TCWV), Total Cloud Liquid Water (TCLW), and Wind Speed (WS) information from the ERA-Interim (ERA-I) reanalysis data (Dee et al., 2011). For sea surface salinity (SSS), we have included data from the GLOBAL-REANALYSIS-PHY-001-030 reanalysis product provided by the Copernicus Marine Environment Monitoring Service (CMEMS; http://marine.copernicus.eu).

The MMD has been filtered for erroneous satellite, in situ and auxiliary data. Erroneous satellite measurements include those contaminated by precipitation and RFI from ground-based sources, geostationary satellites and communication satellites that are identified and flagged in order to obtain accurate SST retrievals (e.g. Gentemann and Hilburn, 2015). Low quality AMSR-E pixels and brightness temperatures outside the accepted range (0−320 K) were also flagged as erroneous. In addition, data were flagged if the difference between the measurements in vertical (V) and horizontal (H) polarization for the 18–36 GHz channels was negative. The standard deviations of the 23 and 36 GHz channels (both polarizations) were calculated over a 21 × 21 subregion surrounding the retrieval pixel and were used to flag low quality data. Low quality in situ data and matchups with an in situ or ERA-I SST outside the accepted range (−2–34 °C) were flagged as erroneous. Similar, matchups with ERA-I wind speeds greater than 20 m s^{-1} were flagged. Further checks were included to account for situations where the SST retrievals could be compromised due to land and sea ice contamination, diurnal warming effects, precipitation and sun glitter contamination, and RFI. Land and sea ice contamination was accounted for by using the AMSR-E land/ocean flag and ERA-I sea ice fraction. To avoid diurnal warming effects, daytime matchups with ERA-I wind speeds less than 4 m s^{-1} were removed. Rain contamination was accounted for by removing data if the brightness temperature of the 18 V channel exceeded 240 K. Contamination from sun glitter was accounted for by removing data with sun glint angles below 25 degrees. Potential contamination due to RFI was accounted for by using Table 2 in Gentemann and Hilburn (2015), as well as the observation location (for ground based RFI) and the reflection longitude and latitude (for geostationary RFI). As a last control, a 3-σ filter was applied on the ERA-I and in situ SST difference. The number of matchups per latitude band is limited in order to ensure a better balance of data across the different latitude bands. For further details on data filtering see Alerskans et al. (2020).

The focus of this study is the year 2010. To obtain an independent validation data set, the matchups from 2010 were divided into four subsets to accommodate the fact that the regression model (see Section 3.1.2) needs to be trained and afterwards validated on independent data. The first three subsets are used exclusively for the training of the regression model whereas the validation subset is used to validate the results from both retrieval algorithms. For a more detailed description on the division into subsets see Alerskans et al. (2020).

The filtered validation subset consists of 1,514,985 matchups from 2010, which are used to compare the performance of the different channel configurations for both retrieval algorithms. Fig. 1 shows the geographical distribution of the validation subset, with the largest number of matchups in North Atlantic Ocean. There is no significant seasonal variation in the matchup distribution. In Section 3.2, different filters are established to be used during the comparison of different channel configurations.

3. Methods

3.1. Retrieval algorithms

3.1.1. Optimal estimation (OE) retrieval algorithm

The OE algorithm used in this study is based on the retrieval algorithm described in Nielsen-Englyst et al. (2018), which is built upon the forward model described in Wentz and Meissner (2000) that relates the observed top of atmosphere brightness temperatures to the relevant

Table 2
The number of matchups (N) left after filtering and the percentage of matchups removed by each filter. Convergence is also required by F1 + F2.

Filter	N	% Removed
No filter	1,514,985	–
OE convergence passed	1,246,425	18
F1: TB RMSE < 0.25 K	694,473	54
F2: TB RMSE < 0.5 K	1,137,765	25
F3: 3RSTD	1,080,336	29

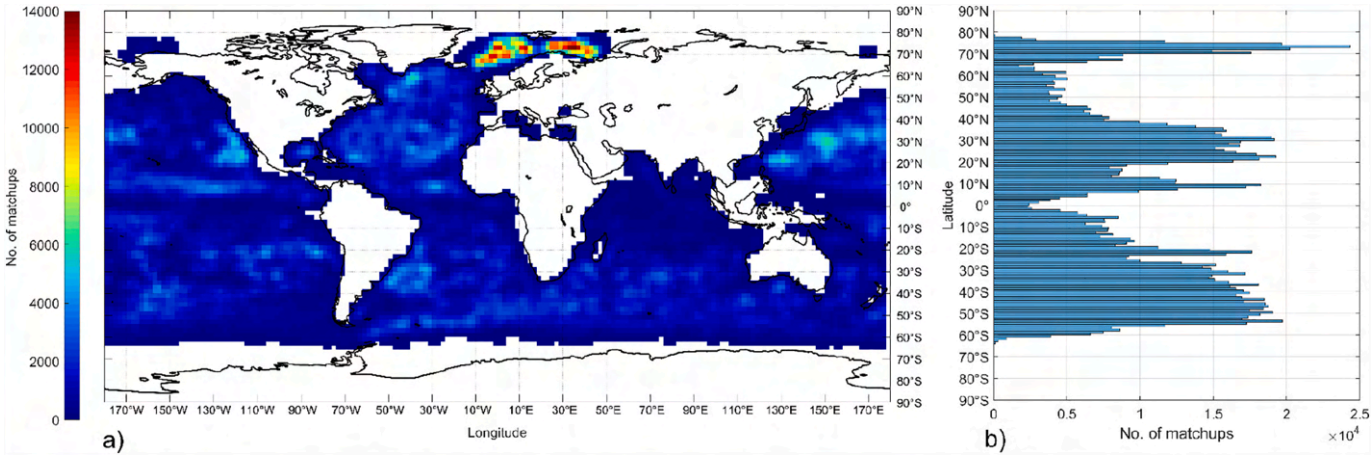


Fig. 1. (a) The geographical distribution of the matchups using a grid size of 5 × 5 degrees and (b) the latitudinal distribution of matchups, for the filtered validation subset.

geophysical parameters as given by the state vector, $\mathbf{x}$ = [WS, TCWV, TCLW, SST]. These are by experience the most important geophysical parameters influencing the amount of microwave radiation received by the satellite antenna at frequencies in the microwave part of the spectrum over open-ocean (Wentz and Meissner, 2000). Following Nielsen-Englyst et al. (2018), we use a regression model to correct for the forward model residuals, which depends on SST and WS from the previous iteration and the ERA-I wind direction relative to the azimuthal look, φ_r. One way to examine the impact of the different channels on the retrieved parameters is to investigate the change in brightness temperature for a perturbation in the individual geophysical parameters (i.e. the Jacobian matrix). The sensitivity of the brightness temperatures to the geophysical parameters varies with both frequency and polarization. The sensitivity of the forward model to SST is shown in Fig. 2a for the AMSR-E frequencies up to 36 GHz. It illustrates that the sensitivity is larger for the vertical polarizations, where 6.9 V GHz provides the highest sensitivity with maximum changes in brightness temperature of 0.5 K/°C. For 6.9 and in particular the 10.7 GHz, the sensitivity decreases for cold SSTs, indicating that SST retrievals are more difficult in cold waters compared to warm waters. The WS influences the sea surface roughness, which affects the surface emissivity and hence the observed brightness temperatures (Hollinger, 1971; Meissner and Wentz, 2002, 2012). Fig. 2b shows the sensitivity to WS (wind induced surface roughness) with larger sensitivities to horizontal than vertical polarization. There is a distinct change in the slopes at WS of about 7 m s^{-1} where foam starts to form, which increases the emissivity for both polarizations. The effect of water vapor is weak at low frequencies, and peaks at 23 GHz (Fig. 2c). Fig. 2d shows the change in brightness temperature over a range of TCLWs, with the main influence at the 36 GHz channels. The SSS has a small effect and the influence is largest at low frequencies (Fig. 2e). These dependencies are similar to the sensitivities found by Prigent et al. (2013) using the fast radiative transfer model RTTOV, but differ from those found by Pearson et al. (2018). For consistency, the sensitivity to the relative wind direction (φ_r) is shown in Fig. 2f, with the largest impact at the vertical polarizations at angles of 100–250°.

The OE technique presented by Rodgers (2000) has been used to retrieve WS, TCWV, TCLW and SST from AMSR-E observations by inverting the forward model. In the inversion, the forward model is constrained by a priori information about the expected mean and covariance of the geophysical parameters. A priori information from ERA-I has been used for SST, WS, TCWV and TCLW. We use a similar setup as in Nielsen-Englyst et al. (2018), but with an increased diagonal covariance matrix element for SST to increase the sensitivity of the retrieved SST to true changes in SST and to ensure the independency from the first guess SST information from ERA-I. The covariance of the a priori state $\mathbf{x_a}$ (the a priori guess of the ocean and atmospheric state $\mathbf{x}$), is therefore given as:

$$\mathbf{S_a} = \begin{bmatrix} e_{WS}^2 & 0 & 0 & 0 \\ 0 & e_{TCWV}^2 & 0 & 0 \\ 0 & 0 & e_{TCLW}^2 & 0 \\ 0 & 0 & 0 & e_{SST}^2 \end{bmatrix} \quad (1)$$

where e_{WS} = 2 m s^{-1}, e_{TCWV} = 0.9 mm, e_{TCLW} = 1 mm and e_{SST} = 4 °C. The uncertainties on WS, TCWV and TCLW are best estimates based on available published validation results (Chelton and Freilich, 2005; Dee et al., 2011; Jakobson et al., 2012; Jiang et al., 2012; Li et al., 2008) and the covariances are assumed zero following Nielsen-Englyst et al. (2018). As in Nielsen-Englyst et al. (2018), we use the 5 lower frequencies: 6.9, 10.7, 18.7, 23.8, 36.5 GHz from AMSR-E for the retrieval. The $\mathbf{S_e}$ is listed in Appendix A and the constant bias correction of the forward model ranges from −0.72 K on 10 GHz H to 0.64 K on 18.7 GHz V.

In the OE retrieval method, the error on the retrieval is given by the retrieval error covariance matrix:

$$\mathbf{S} = \left(\mathbf{S_a^{-1}} + \mathbf{K_i^T S_e^{-1} K_i}\right)^{-1} \quad (2)$$

where $\mathbf{K}$ is the Jacobian of the forward model. The square root of the $\mathbf{S}$ matrix provides the theoretical retrieval error standard deviation. Fig. 3 shows the theoretical retrieval error standard deviation of SST as a function of SST for different WS, TCWV and TCLW conditions using an incidence angle of 55°. The figures show that the theoretical retrieval error is largely dependent on the SST and WS, with largest errors for cold waters and high wind speeds. In comparison, the impact from TCWV and TCLW is very limited. This is in agreement with previous results (Kilic et al., 2018; Prigent et al., 2013).

The OE algorithm also directly provides the sensitivities of the retrieved parameters to the true state through the averaging kernel matrix, which is given by:

$$\mathbf{A_{ij}} = \frac{dx_i}{dx_j^t} \quad (3)$$

where x^t is the true state. Using the validation subset as described in Section 2.1 and a filter requiring the root mean squared difference of simulated versus observed brightness temperatures (TB RMSE) less than 0.5 K, the mean sensitivities of the retrieved parameters to the true state are given as: 0.99, 1.00, 1.00 and 0.99 for WS, TCWV, TCLW and SST. The increased sensitivity to SST, compared to the mean value of 0.5 in

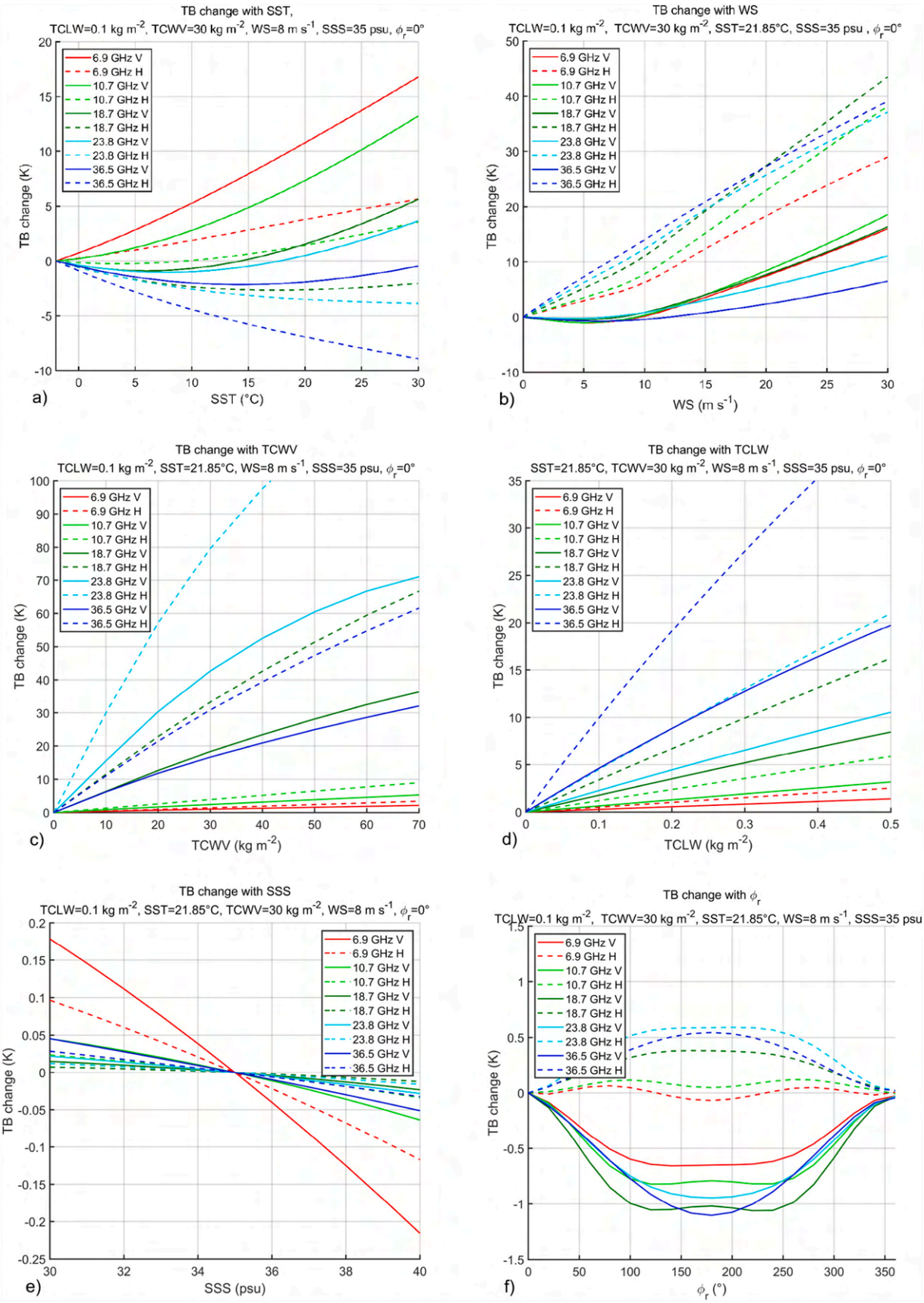


Fig. 2. The change in top-of-atmosphere brightness temperature as a function of a) SST, b) WS, c) TCWV, d) TCLW, e) SSS and f) φ_r as simulated by the forward model for the state defined above the individual plots.

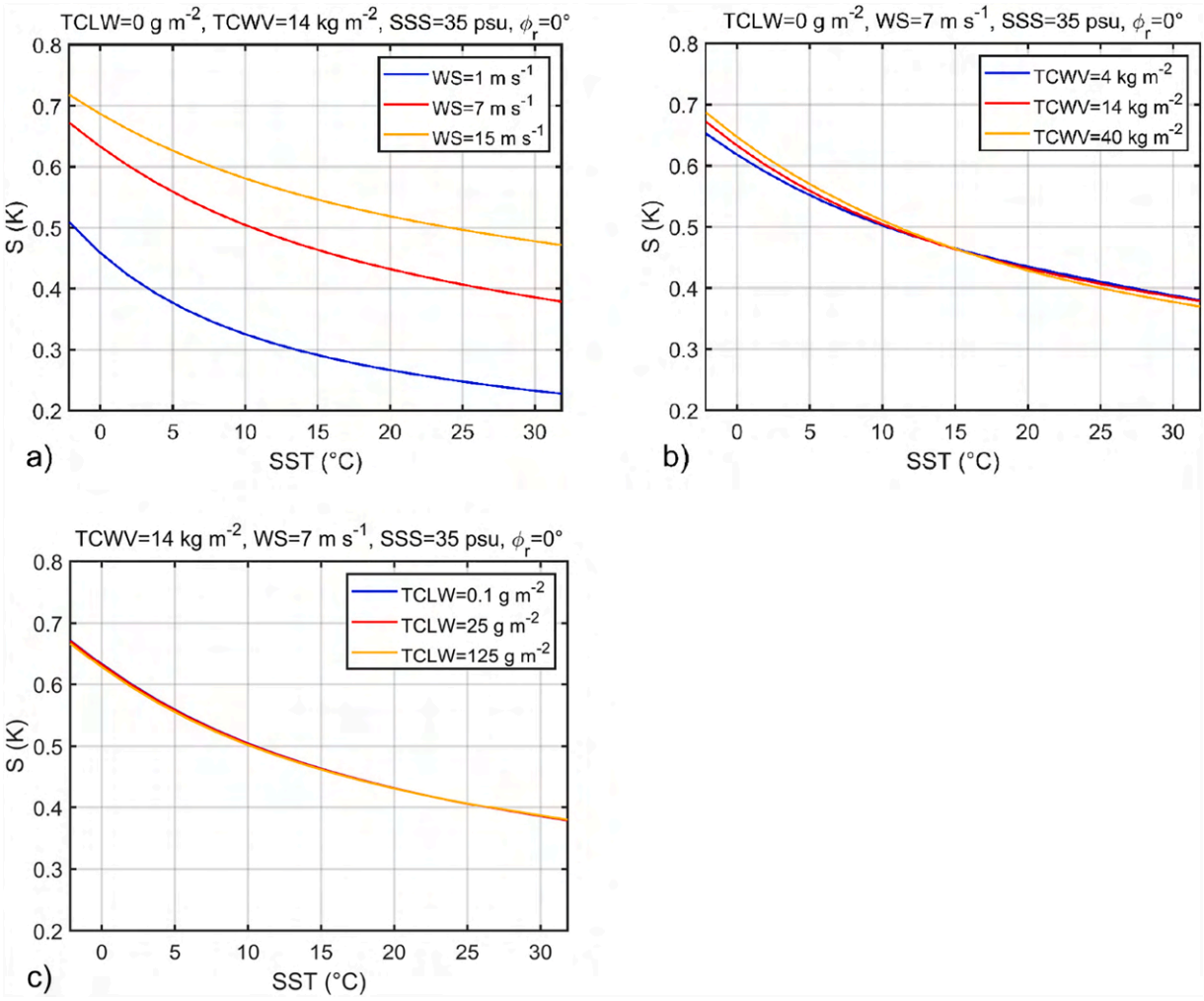


Fig. 3. The simulated retrieval error, S, as a function of SST for different a) WSs, b) TCWVs and c) TCLWs.

Nielsen-Englyst et al. (2018), is a result of the increased uncertainty on the first guess SST from ERA-I. For the same subset (TB RMSE<0.5 K), the standard deviation of retrieved SST versus in situ observed SST is 0.66 °C.

This setup has been used to run the OE algorithm for a number of different channel selections (see Section 3.2). The forward model simulates brightness temperatures for all channels, but only the selected channels (and the corresponding measurement and forward model uncertainties) are included in the inversion, when different channel combinations are tested.

3.1.2. Regression (RE) model

The statistical retrieval algorithm consists of a two-step WS regression model followed by an SST regression model. Both steps in the WS retrieval algorithm use AMSR-E brightness temperatures, where the first step applies a global algorithm and the second step applies a localized algorithm for fixed WS intervals, using the first-step retrieved WSs. The SST retrieval algorithm applies a global algorithm to retrieve SST using AMSR-E brightness temperatures, the retrieved wind speeds from step two of the WS retrieval algorithm and information from ERA-I. The SST retrieval is very similar to the one described in Alerskans et al. (2020), except that the coefficients are derived globally in this study. Alerskans et al. (2020) used a two-stage retrieval algorithm to retrieve SST, where the first-stage coefficients were derived locally for fixed latitude intervals and ascending and descending passes, respectively, and the second-stage coefficients were derived locally for fixed WS and SST intervals. The approach with global coefficients was chosen in this study to ensure a consistent comparison of the different channel selections and to minimize effects from localized algorithm coefficient derivations. The global performance of the retrieved SST compared to drifter SST is 0.62 °C using the frequency range 6–36 GHz and a filter of TB RMSE less than 0.5 K, as derived for the OE retrieval (see Section 3.2). In this study, we are considering the relative differences between the different channel selections and we are therefore not concerned with the slightly larger retrieval errors.

3.2. Selection of channels and subsets

Different channel combinations have been tested by selecting different subsets of the channels available in AMSR-E and a configuration where all channels (except the 89 GHz channels) are included. As the focus of this study is SST retrievals, the main focus will be on the 15 different channel selections that all include the 6 GHz channels, as 6 GHz is the primary frequency for global SST retrievals (Wilheit, 1979). However, it is beneficial to know the effect of excluding the 6 GHz channels, as the GMI instrument could be a potential bridging data set between AMSR2 and future PMW missions (Bidwell et al., 2005; Draper et al., 2015). The importance of the 6 GHz channel is clear when the two channel combinations 6, 10, 18, 23 and 36 GHz and 10, 18, 23 and 36 GHz are compared, where the removal of the 6 GHz channels results in an increase in the uncertainty of the retrieved SST from 0.66 °C to 0.75 °C for the OE algorithm and from 0.62 °C to 0.86 °C using the RE

algorithm and the TB RMSE<0.5 K filter.

In this study, we will refer to the AMSR-E channel selection as the one including: 6, 10, 18, 23 and 36 GHz (without the 89 GHz channels), following Nielsen-Englyst et al. (2018). Alerskans et al. (2020) uses the 89 GHz channels, but the inclusion only resulted in slight improvements in the performance. We will refer to the CIMR channel selection as the one including: 6, 10, 18 and 36 GHz (in contrast to the expected configuration, which also includes the 1.4 GHz frequency). It is assumed that the V and H polarizations are always available (or not available) at the same time. In Section 4.1, the 15 different channel selections have been tested on the same validation subset for both the OE and RE retrieval algorithms. It is important to select a suitable selection filter that enables a comparison of the different retrievals. Here, the subsets have been obtained by applying a common filter based on the matchups for which the OE AMSR-E channel configuration has reached convergence and another filter based on TB RMSE also obtained from the OE AMSR-E channel configuration. Differences in the simulated and observed brightness temperatures can usually be attributed to forward model errors or measurement errors (due to e.g. imperfect calibration or channel contamination). Nielsen-Englyst et al. (2018) showed a close relationship between the TB RMSE and the SST retrieval performance. Table 2 shows the number of matchups removed by applying different filters. Of all matchups, 82% reached convergence as defined in Nielsen-Englyst et al. (2018). The next two filters (F1 and F2) include the convergence criterion as well, but also additional filters with TB RMSE less than 0.25 K and 0.5 K, respectively. The F2 filter removes 25% of the matchups (including the ones that did not reach convergence), while the F1 filter removes 54%. Using the simulated brightness temperatures from the AMSR-E channel configuration for filtering will favor this configuration over the other selections, but ensure that the same subset is used for all channel configurations and that the worst retrievals are excluded. The TB RMSE filters have been used in Section 4.1.

In Section 4.2, the analysis focuses on four of the above-mentioned channel configurations: a) 6, 10, 18 GHz, b) 6, 10, 23 GHz, c) 6, 10, 18, 36 GHz, and d) 6, 10, 18, 23, 36 GHz. These configurations have been chosen because they represent the effect of including the 23 GHz channels through a direct comparison against the 18 GHz channels and through a comparison of the CIMR and AMSR-E channel configurations. Furthermore, selection a) and c) allow us to examine the effects of the 36 GHz channels. For this analysis, another filter (F3) has been applied to ensure that the filtering does not favor any of the four channel configurations, and still retains enough matches to make robust validation statistics. The F3 filter removes matchups if the simulated minus observed brightness temperatures exceeds three times the robust standard deviation (3RSTD) for any of the channels in each of the four channel configurations. This filter removes 29% of the matchups, most of which are the same as those removed by the F2 filter, with 99% of the matchups removed by the F3 filter also removed by the F2 filter. A total of 98% of the matchups removed by the F1 filter are also removed by the F3 filter. Fig. 4 shows the geographical distribution of the removed matchups using the F1 and F3 filter, respectively. Most of removed matchups are located in the areas, where the largest number of matchups is present (see Fig. 1) and the two filters agree on the locations where most matchup are removed.

4. Results

4.1. Performance for all channels

The validation subset (described in Section 2.1) has been used for both retrieval algorithms (RE and OE) to assess the overall impact of using different channel combinations. Fig. 5 shows the robust standard deviation (rstd) of retrieved SST – in situ SST for the OE and RE algorithms, using the 15 different channel combinations, where the F2 and F1 filters have been applied, removing 25% and 54% of the matchups, respectively (see Section 3.2). The channel configurations are ranked from the highest standard deviations to the lowest using the RE model results and the F1 filter to determine the order. This ranking order of the channels will be used for all figures in this section. Adding more channels leads to a better performance of the RE model, with the 6, 10, and 18 GHz channels being the most important. Using the F1 filter improves the RE performance for all channel configurations, with limited changes in the ranking order compared to the F2 filter. The TB RMSE filtering shows improvements in the RE retrieval of comparable magnitude as when using the OE algorithm. This shows the value of the TB RMSE factor as an uncertainty indicator. The OE algorithm shows less variation among the different channel combinations compared to the RE algorithm, and the OE algorithm performs quite well by only including 2 frequencies, where it outperforms the RE algorithm.

The differences and similarities between the RE and OE algorithms have been investigated further by looking at the performance during different weather and surface conditions, using the F2 filter. This filter was chosen to keep sufficient matchups in the different categories and to minimize effects from filtering on the results. Fig. 6a–b show the performance of the retrieved SST for calm or strong winds (calm or rough sea), cold or warm SSTs, and dry or humid conditions for the OE and RE algorithm, respectively. Using 500 bootstrap samples with replacement (same size as N), the average width of the 90% confidence intervals is 0.0030 °C, when considering both algorithms and all the different weather/surface conditions and channel selections. The maximum width of the 90% confidence intervals is 0.0075 °C and calculated for the RE WS $>$ 10 m s^{-1} and the 6, 18 GHz channel selection data subset. Warm, calm waters and humid conditions show a good and similar performance for the different channel selections for both algorithms.

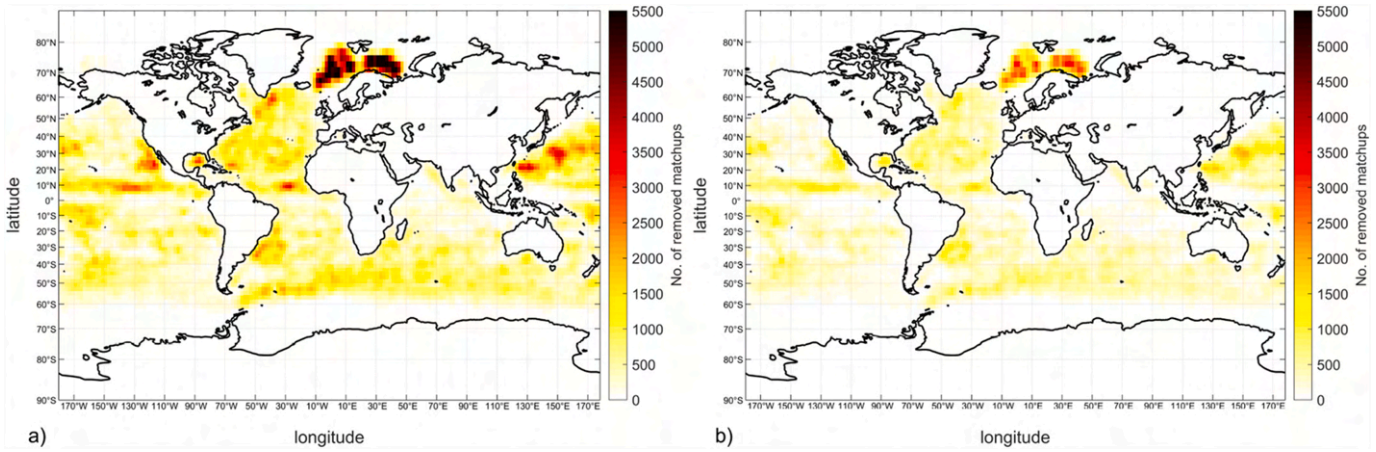


Fig. 4. The number of matchups removed by a) the F1 (TB RMSE<0.25 K) filter and b) the F3 (3RSTD) filter using 5 × 5 degree bins.

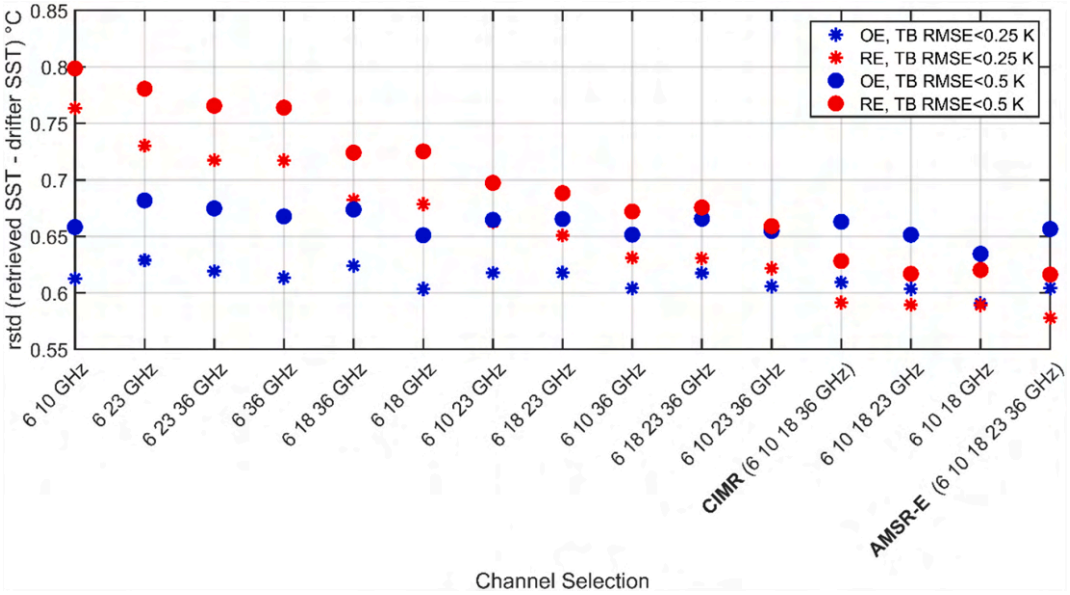


Fig. 5. Robust standard deviation of retrieved SST against drifter SST for different channel selections using the OE and RE algorithm and the filters F1 (TB RMSE<0.25 K) and F2 (TB RMSE<0.5 K), respectively. The channel configurations are ranked from the highest to the lowest standard deviations using F1 filtering and the RE algorithm.

Rough and cold waters provide equally high standard deviations, when compared to the relatively low standard deviations observed for warm and calm waters. Both algorithms show a good performance for high TCWV conditions, but this effect is likely related to the good performance in warm waters, with 96% of the high water vapor matchups located in warm waters (>15 °C). By including more channels, the algorithms tend to retrieve better SSTs in cold and rough waters, reducing the overall spread and thus allowing the algorithms to better represent all observing conditions. One of the key differences between the two algorithms is seen in the variability among the different weather and surface conditions. The OE algorithm shows significant increases in uncertainty for cold and rough waters, but a very good performance in warm and calm waters for all channel configurations. In contrast, the RE model shows limited variability among the different weather and surface conditions, especially when more channels are included. This demonstrates the strength of the statistical RE model, which always maximizes the amount of information in the channels available for retrieval based upon the information in the training data. In few of the RE channel combinations (6, 10 GHz and 6, 36 GHz), the information is limited resulting in inadequate representation of all conditions. These channel selections work for warm waters, but they all show weak performance in cold and rough waters for the RE algorithm. A good performance should not only be judged on the overall performance, since it depends on the number of matchups representing the given conditions, but it should also consider the spread in performance between the different observing conditions.

The overall improvement in SST performance (i.e. decrease in rstd when compared to drifter SST) of including the different frequencies (V and H polarization) can be seen in the first two columns of.

Table 3. These are based on the difference in performance between the algorithm that includes the frequency considered and the same channel configuration excluding the frequency considered. The improvements by adding the 6 GHz channels are thus based on the comparison of the 10, 18, 23, 36 GHz against the AMSR-E configuration, while the statistics on each of the other channels are based on a comparison of 7 pairs of channel configurations (all including the 6 GHz channels). The channels with the largest impact have been marked as bold in the table. The table shows that the OE and the RE algorithm agree that (not surprisingly) the most important channels for SST retrievals are the 6 GHz channels, followed by the 10 and 18 GHz channels.

Fig. 7 shows the sensitivity (Eq. (3)) to SST obtained using the OE algorithm for the different channel configurations during the different weather and surface conditions. The sensitivities to SST are generally high (mean value of 0.98) and the variations are small, but significant. Using 500 bootstrap samples with replacement (same size as N), the average width of the 90% confidence intervals is 2.44e-05 when considering the different weather/surface conditions and channel selections. The maximum width of the 90% confidence intervals is 8.77e-05 and calculated for the WS > 10 m s^{-1} and the 6, 10 GHz channel data subset. The high sensitivities are explained by the increased uncertainty on first guess SST. As more channels are included the sensitivity increases (in particular for cold waters), with the 6, 10 GHz configuration having the lowest sensitivity. This should be viewed in context of the performance of the retrieved SST against in situ SST in Fig. 6a, where increased number of channels improves the performance over the range of different observing conditions. The lowest sensitivity is generally seen for cold SST and high wind speeds, which explains the poor OE performance in these conditions using all channel configurations (Fig. 6a).

The sensitivity to TCLW (not shown) is likewise very high (mean value of 1.00) across the 15 channel selections, due to relative high uncertainty on ERA-I TCLW, which means that the retrieved TCLW is independent of the first guess from ERA-I. The sensitivity to WS and TCWV (also not shown) vary across the different channel selections depending on the information available in the selected channels, with the mean values of 0.93 and 0.44, respectively. The increase in sensitivities with inclusion of different frequencies (V and H polarization) can be seen in the last four columns of.

Table 3. The addition of the 36 GHz channels shows the largest increase in both the WS and TCLW sensitivity, followed by the 18, 23 and 10 GHz channels, while the 6 GHz channels give the smallest increase in sensitivity. For the TCWV sensitivity, the 23 GHz channels show the largest increase, followed by the 36 and 18 GHz channels and with limited impact from the 6 and 10 GHz channels. For the SST sensitivity, the 6 GHz channels show the largest increase, and leave the remaining channels with a small impact on the sensitivity in comparison.

4.2. Geographical dependencies for selected channel configurations

Fig. 8 shows the global performance of retrieved SST when compared

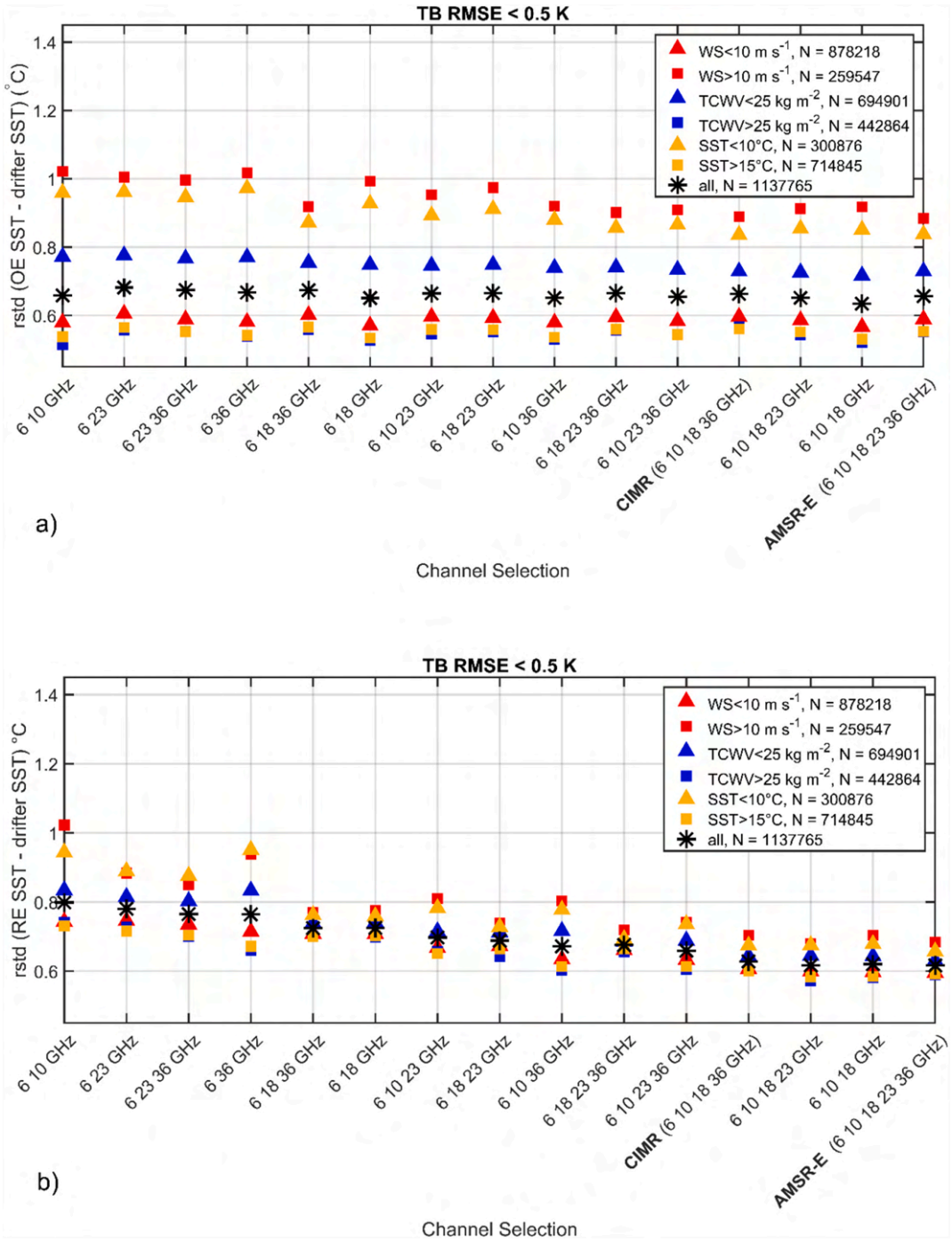


Fig. 6. Robust standard deviation of retrieved SST against drifter SST for different conditions and channel selections using a) the OE algorithm and b) the RE algorithm. The F2 (TB RMSE<0.5 K) filter has been applied for both the OE and RE algorithm and N is the number of matchups in each subset. The channel configurations are ordered from the highest to the lowest standard deviations using "all" conditions/matchups after F1 filtering (TB RMSE<0.25 K) for the RE algorithm.

to drifters for the RE and OE algorithms, respectively, for the AMSR-E channel configuration using the F3 filter. In general, the best performance is seen for low latitudes, while the high latitude and high variability regions show increased standard deviations. This is consistent with the fact that $\partial TB/\partial SST$ is smaller for cold waters compared to warm waters, as is also shown in Fig. 2a. Sea ice and a much smaller Rossby radius (compared to low latitudes), resulting in much smaller scale of eddies and fronts, further complicate SST retrievals in high latitudes. Better spatial resolution (as expected by CIMR) will make it easier to detect these features and therefore improve high latitude retrievals. The large standard deviations in the Sea of Japan are evident for all channel combinations and both types of retrieval algorithms, and are likely effects from RFI at 6 GHz and RFI at 10 and/or 18 GHz, which are not removed by the applied RFI filters (see Section 2.1). Larger standard deviations are seen both with and without the 6 GHz channels, indicating that some matchups are contaminated at 6 GHz, while others are contaminated at 10 GHz and/or 18 GHz. Therefore, more sophisticated RFI filtering may be required, such as being planned for on-board implementation on CIMR. Since, the effect is seen in all channel selections it will not change the conclusions drawn in this study. The OE algorithm shows larger latitudinal differences in performance, with a better performance than the RE algorithm in the region between 25 N

Table 3
Improvements in SST retrievals compared to drifters for the OE and RE algorithms (column 1–2), and the increase in sensitivities (column 3–6), by including information from individual frequencies (both V and H polarization) using the F2 (TB RMSE < 0.5 K) filter. The largest increases in SST performances and sensitivities have been marked with bold.

	OE SST	RE SST	$\Delta A_{1,1}$ (WS)	$\Delta A_{2,2}$ (TCWV)	$\Delta A_{3,3}$ (TCLW)	$\Delta A_{4,4}$ (SST)
6 GHz	**0.2402**	**0.0907**	0.0210	0.0140	0.0000	**0.0145**
10 GHz	0.0876	0.0148	0.0372	0.0913	0.0002	0.0013
18 GHz	0.0811	0.0117	0.0431	0.2762	0.0017	0.0007
23 GHz	0.0309	0.0089	0.0339	**0.5984**	0.0011	0.0007
36 GHz	0.0289	0.0114	**0.0950**	0.4939	**0.0039**	0.0011

and 25 S, but a decrease in the performance at latitudes above 40°. The smaller latitudinal dependence in the RE algorithm indicates that the statistical model finds a way to compensate for the reduced sensitivity in colder waters using information from the other channels and the covariances between the geophysical parameters that have an impact on the observed brightness temperatures.

Fig. 9 shows how the performance of the three channel selections: 6, 10, 18 GHz, 6, 10, 23 GHz, and CIMR differs from the performance using the AMSR-E channel configuration for the OE and RE algorithm, respectively. Using the 6, 10, 18 GHz channels improves the performance in the equatorial region (and mid-latitudes for the OE algorithm) compared to the AMSR-E configuration. However, in the high latitudes (and further into the mid-latitudes for the RE algorithm) the AMSR-E configuration outperforms the 6, 10, 18 GHz configuration. The

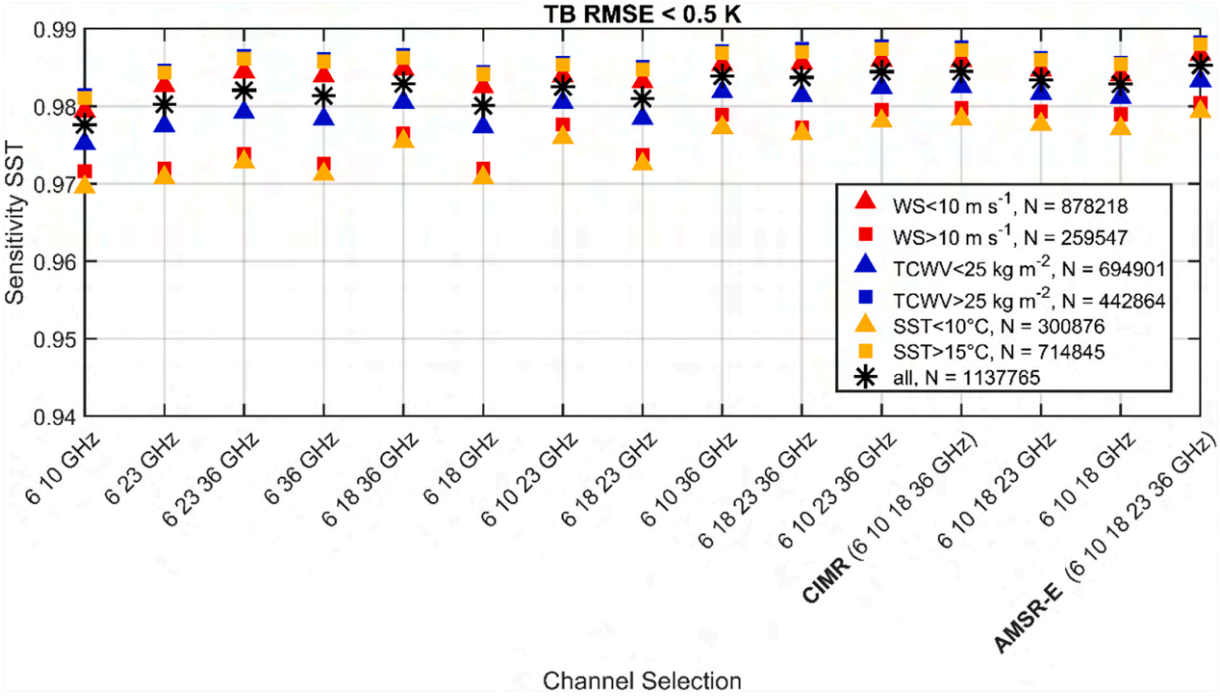


Fig. 7. The sensitivity to SST for the different channel selections and different weather/surface conditions using the same ranking order as in the previous figures and the F2 (TB RMSE < 0.5 K) filtering. N is the number of matchups in each subset.

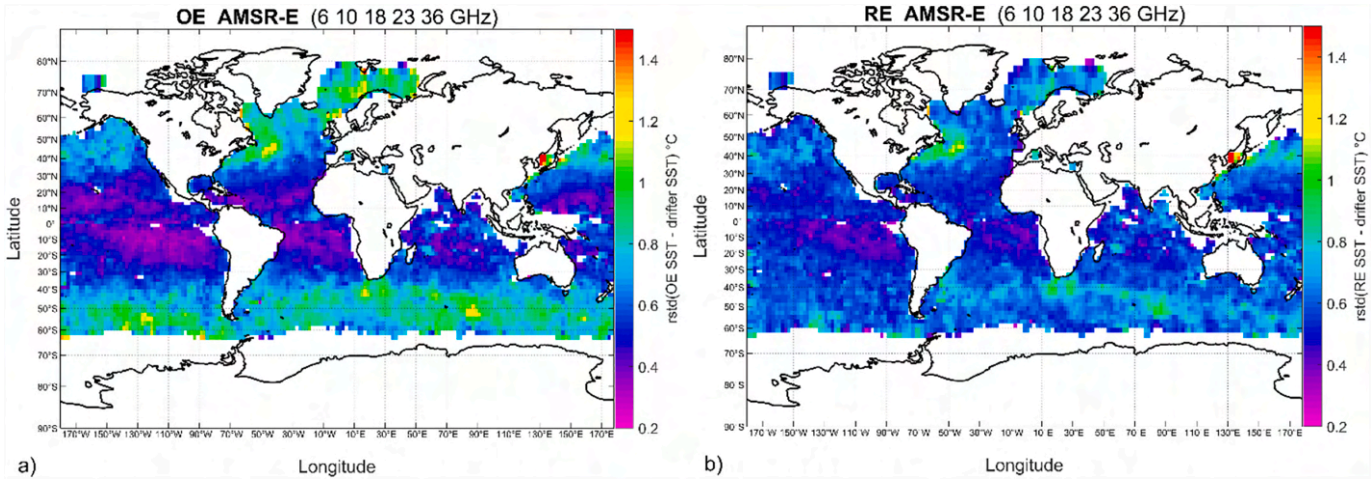


Fig. 8. Retrieved SST performance (robust standard deviations, rstd) against drifter SST using the AMSR-E (6, 10, 18, 23, 36 GHz) channel selection for a) the OE algorithm and b) the RE algorithm, using the F3 (3RSTD) filter. The geographical statistics have been gridded using a grid size of 5 degrees, with a minimum of 50 matchups per grid cell.

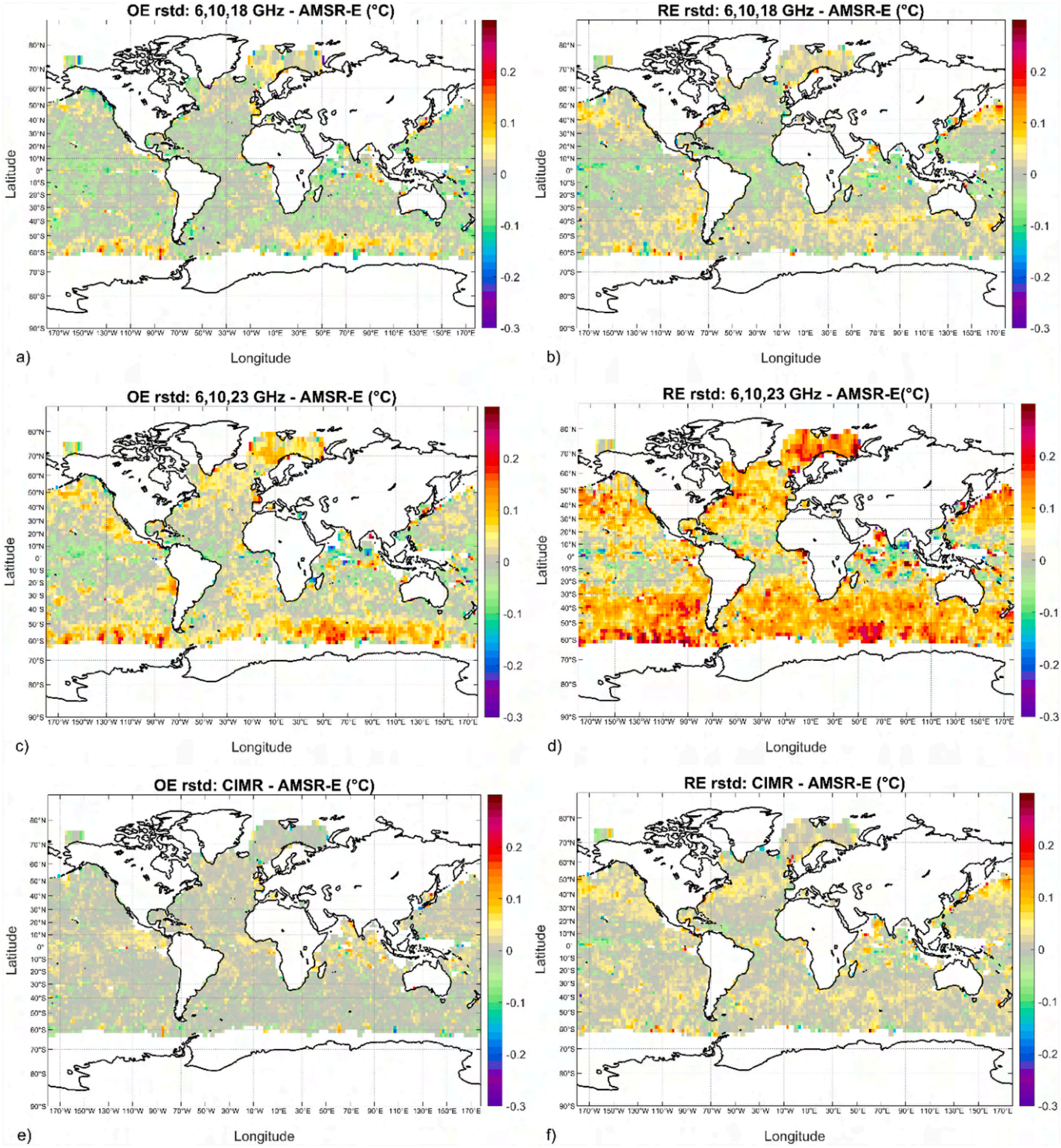


Fig. 9. Difference in robust standard deviations (retrieved minus drifter SST) (rstd) between channel configurations 6, 10, 18 GHz and the AMSR-E channel configuration using a) the OE algorithm, and b) the RE algorithm. Similar, the differences in rstd between 6, 10, 23 GHz and AMSR-E are shown for the OE (c) and RE (d) algorithms and the differences in rstd between CIMR and AMSR-E for the OE (e) and RE (f) algorithms. The F3 (3RSTD) filter has been used in all figures. The geographical statistics have been gridded using a grid size of 5 degrees, with a minimum of 50 matchups per grid cell.

largest reductions in performance (compared to AMSR-E) are seen in the comparison of the 6, 10, 23 GHz configuration for both algorithms, with the most pronounced degradation for the RE algorithm. For both algorithms, the degradations are concentrated in the high latitudes (and mid latitudes for the RE algorithm), while the 6, 10, 23 GHz configuration seems to perform slightly better in the equatorial region. Both the OE and RE algorithms show the best agreement between the CIMR and AMSR-E channel selections, with the smallest differences using the OE algorithm. These findings have been summarized in Table 4 for the nine regions shown in Fig. 10.

Table 4 shows the performance of the four channel configurations using the OE and RE algorithms for each of the nine regions. For all channel selections, the best performance is seen in the equatorial region, with decreasing performance towards the poles. This latitudinal

Table 4
The retrieved versus drifter SST robust standard deviations (°C) within each region using the OE and RE algorithm, for each of the four channel selections using the F3 filter.

Region	RE algorithm				OE algorithm				
	6, 10, 18	6, 10, 23	CIMR	AMSR-E	6, 10, 18	6, 10, 23	CIMR	AMSR-E	N
Arctic	0.70	0.82	0.69	**0.67**	0.92	0.96	0.90	**0.89**	109,493
Subpolar North Atlantic	0.67	0.75	0.68	**0.66**	**0.82**	0.85	**0.82**	**0.82**	43,160
North Atlantic	**0.62**	0.68	0.63	**0.62**	**0.57**	0.60	0.59	0.59	110,870
North Pacific	0.64	0.71	0.65	**0.63**	**0.58**	0.62	0.60	0.60	163,072
Equatorial Region	**0.53**	0.57	0.55	0.55	**0.44**	**0.44**	0.48	0.47	188,331
South Atlantic	**0.61**	0.69	0.62	**0.61**	**0.60**	0.63	0.61	0.61	76,185
South Pacific	0.58	0.65	0.58	**0.57**	**0.53**	0.56	0.55	0.54	116,675
South Indian Ocean	0.63	0.70	0.63	**0.61**	**0.63**	0.65	0.65	0.65	83,632
Southern Ocean	0.68	0.79	0.67	**0.65**	0.83	0.87	**0.82**	0.83	174,069
All regions	**0.62**	0.70	0.63	**0.62**	**0.63**	0.66	0.66	0.66	1,065,487

Bold indicates the lowest robust standard deviation among the channel selections for each region and both algorithms. N is the number of matchups within each region.

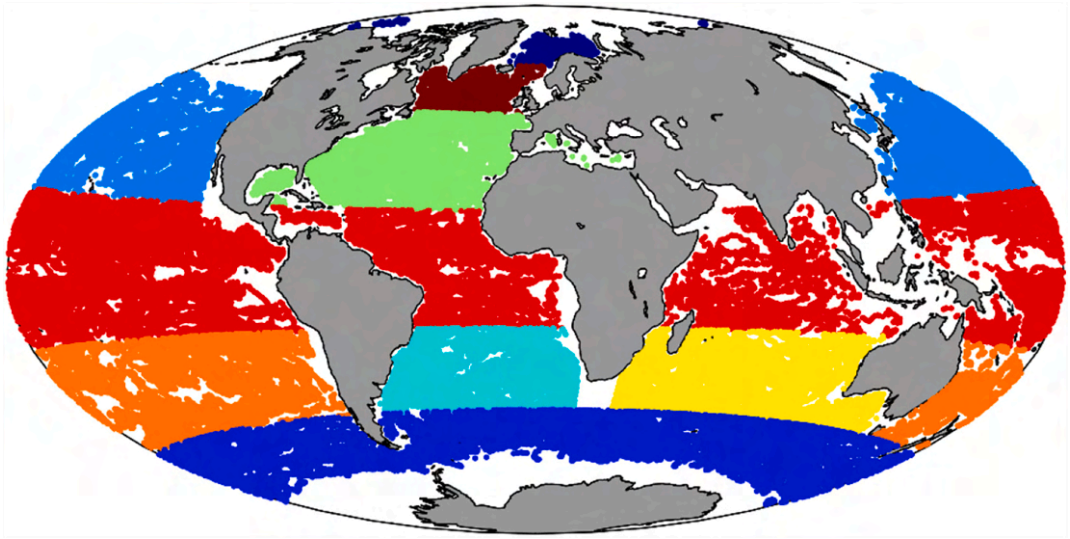


Fig. 10. The matchups in the nine selected regions for which the algorithms have been evaluated.

dependence is more pronounced for the OE algorithm compared to the RE algorithm, in all configurations. The 6, 10, 18 GHz combination shows very good performance in particular for the OE algorithm, where it outperforms the AMSR-E selection in most regions (except for the high latitudes). The AMSR-E and CIMR configurations show very similar performance in all regions.

Fig. 11 shows the seasonal cycle of the robust standard deviations of retrieved minus drifter SST for the same four channel selections using both retrieval algorithms in the nine different regions. The high latitudes show a clear seasonal cycle, with the lowest standard deviations during local summer (warm SST) for both algorithms. The high latitude regions also show the largest differences between the RE and OE performance, which is particularly true during local winter time, where the RE algorithm performs better than the OE algorithm. For most regions, the 6, 10, 23 GHz configuration gives the highest standard deviations for both retrieval algorithms. The RE algorithm shows very similar seasonal performance for the 6, 10, 18 GHz, CIMR and AMSR-E channel configurations for all regions. However, large differences between these configurations and the 6, 10, 23 GHz configuration are observed using the RE algorithm, especially at high latitudes during winter, where significant improvements are seen when replacing the 23 GHz with the 18 GHz channels.

5. Discussion

This study shows that the inclusion of more channels and more information in the retrieval reduces the uncertainty in the SST retrievals. This is clearly seen for the RE algorithm, while the OE algorithm shows less variability in overall performance among the different channel selections. This difference can be explained by the fundamental difference in the two types of retrieval algorithms. The statistical retrieval is based solely on the information available in the selected satellite channel observations. Conversely, the OE types of retrieval algorithm use a priori information to constrain the retrievals. If limited information is available in a given set of channels, the sensitivity will decrease and as a result, the algorithm will put more weight on the first guess. The largest decreases in the wind speed (i.e. sea surface roughness) sensitivity are seen for retrievals in cases, where 36, 23 or 18 GHz are left out from the OE retrieval. Theoretically, the sensitivity to WS is largest at 10 and 18 GHz (Fig. 2b). In the approach used here however, the fact that individual channels are sensitive to other parameters is turned to an advantage by the simultaneous use of all channels. The largest decrease in water vapor sensitivity is seen for retrievals, where the 23 and 36 GHz observations are left out. In particular, the retrieved water vapor will in these cases be dominated by the first guess from ERA-I, which therefore will keep the OE retrievals close to the true state even with limited information from the brightness temperatures. This will be the case in a realistic scenario, as the one considered here, where WS and TCWV information is available from NWP at a similar quality as ERA-I to be used as first guess for the OE algorithm. The OE retrieval algorithm is thus more robust towards different channel combinations, compared to the statistical RE algorithm.

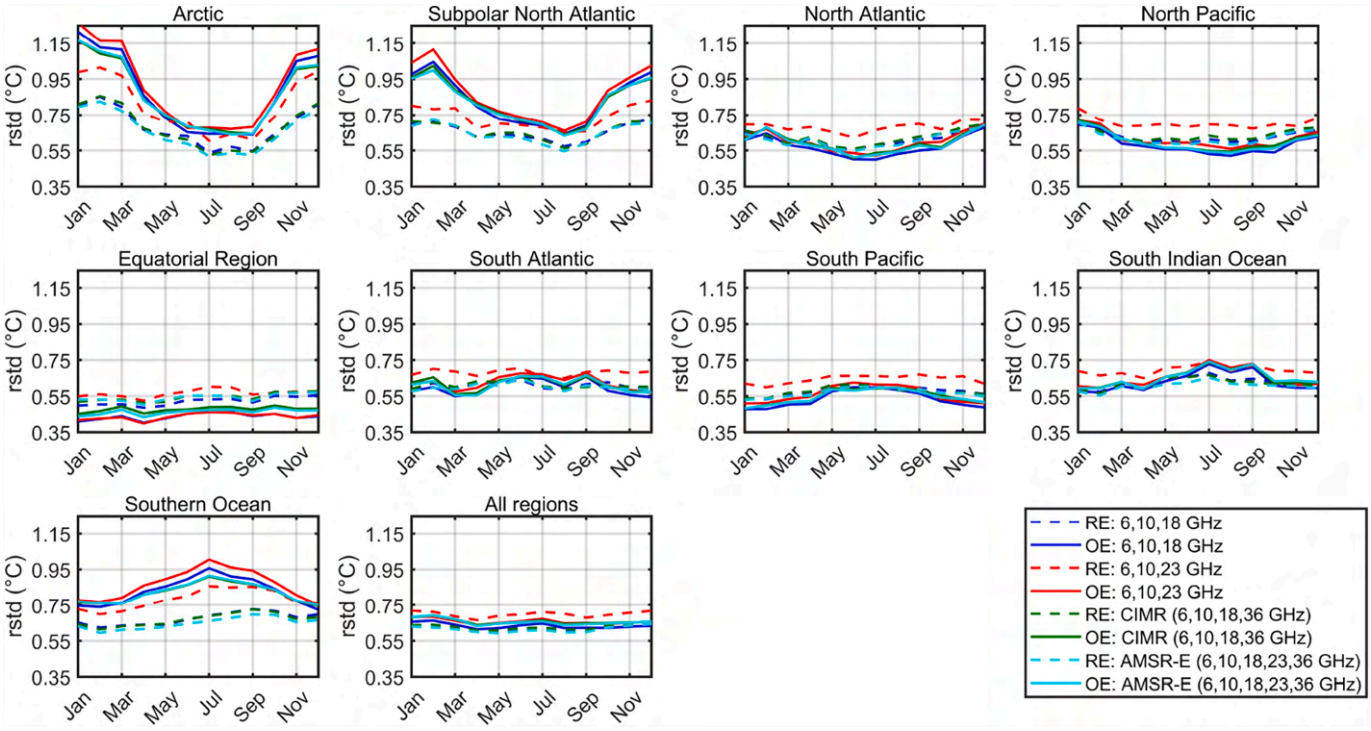


Fig. 11. Retrieved SST performance against drifters for the four different channel selections: 6, 10, 18 GHz, 6, 10, 23 GHz, 6, 10, 18, 36 GHz, and 6, 10, 18, 23, 36 GHz as a function month for the different regions using the F3 filter for both retrieval algorithms.

By including more channels both algorithms show a better representation over the range of the different observing conditions. The performance of the OE algorithm (and the different channel selections) is sensitive to the atmospheric and oceanic conditions with elevated uncertainties for cold and rough waters for all channel selections. The increased uncertainties in cold waters are related to the decreased sensitivity to SST for the low frequency channels (in particular at V polarization) as seen in Fig. 2a. Kilic et al. (2019) found that cold SSTs are a general source of disagreement between simulated and observed brightness temperatures in the emissivity models and they explained this in part by uncertainties in the modelling of the dielectric constants of sea water, but also by inaccuracy in the reanalysis data and high wind speed effects in colder areas. Fig. 2a shows that the sensitivity to SST in cold waters to a large extent originates from the higher frequency channels, with decreasing brightness temperatures for increasing SST. This suggest that the increased uncertainties for cold SST could also be attributed to uncertainty in modelling the higher frequencies due to e.g. imperfections in the atmospheric emissivity model. This effect could possibly to some extent be reduced by including the covariances (off-diagonals in S_a) between the geophysical parameters in the OE algorithm. However, these are currently not well known and it is a complete study to estimate them (see recommendations for future work in Section 6). Using the RE algorithm, the increased uncertainties for cold waters and strong surface roughness is most pronounced for the two-frequency configurations without 18 GHz, while more (>2) frequencies provide enough information to balance the lower sensitivity in cold and rough waters. This indicates that the statistical RE algorithm compensates for the reduced sensitivity of the low frequency channels in colder waters by using information from the higher frequency channels and possibly the covariances between the geophysical parameters influencing the observations. These results demonstrate the importance of using in situ observations and several types of SST retrievals to assess the relative performance of the different channel selections. Information content studies are typically performed using the OE type of algorithms, but different conclusions might be reached using other types of retrieval methods, as indicated here.

Both the OE and RE algorithms agree that in a three frequency selection, the 6, 10, 18 GHz combination is better than 6, 10, 23 GHz, considering the nine different regions in Table 4 and performance throughout the year (Fig. 11). In general, the results for the OE and RE algorithms consistently show that the CIMR channel selection performs very well compared to the AMSR-E configuration. It should be noted that the presented results do not correspond to the expected CIMR performance, as e.g. the CIMR NEdTs will be different from the AMSR-E NEdTs and the CIMR footprints will be significantly smaller than the AMSR-E footprints. Neglecting these observational characteristics for the CIMR instrument, means that the absolute performance of the CIMR retrievals is expected to be significantly better than shown here due to smaller NEdTs and smaller footprints, which better resolve spatial variability at the key frequencies. However, as we focus on the differences between the channel combinations, we consider the AMSR-E NEdTs suitable for the current study. Using the AMSR2 observations instead of AMSR-E observations would very likely resemble the results presented here, as it was shown in Alerskans et al. (2020) that the retrievals from the two instruments showed very similar performance and uncertainty characteristics.

The very good performance of the 6, 10 and 18 GHz combination shown here support the channel selection for the CIMR mission, as the CIMR requirement of the NEdT for the 36 GHz is elevated (Donlon, 2019) and therefore is likely to contribute with limited information to the SST retrievals. Also, note that the CIMR configuration referred to in this paper is without the 1.4 GHz channels (L-band). This could introduce a difference in the wind speed behavior, particularly for high wind speeds (Meissner et al., 2017; Reul et al., 2012, 2016, 2017), and also on the impact of salinity. However, for the winds considered here (<20 m s^{-1}) it was shown in Alerskans et al. (2020) that sufficient information was available in the AMSR-E observations to perform an accurate wind speed retrieval.

To obtain realistic retrieval estimates using the OE algorithm, an accurate forward model is needed to account for the brightness temperature dependencies to the relevant geophysical parameters. The forward model used is based on an earlier version of the RSS (Remote

Sensing Systems) model (Wentz and Meissner, 2000), which showed the least discrepancy with observations in Kilic et al. (2019). The improvements applied to the forward model (Nielsen-Englyst et al., 2018) results in even better agreement with observations and consequently good SST retrievals. The forward model sensitivities presented in this paper (see Fig. 2) indicate a close relationship with Prigent et al. (2013) and Kilic et al. (2018), whereas the use of RTTOV and FASTEM (FAST microwave Emissivity Model) in Pearson et al. (2018) shows significantly different emissivity/brightness temperature dependencies with respect to wind speed and sea surface salinity.

One of the strengths of the OE algorithm is that it offers several possibilities to obtain realistic error estimates and in fact it directly provides a theoretical error estimate for each retrieval. The theoretical uncertainties presented in Section 3.1.1 are in good agreement with those found in Prigent et al. (2013) and Kilic et al. (2018). A validation of the estimated theoretical uncertainties is shown in Fig. 12, where the OE versus drifter SST differences are displayed against the theoretical uncertainties obtained from Eq. 2. The dashed lines represent the ideal uncertainty under the assumptions that drifting buoys have a total uncertainty of 0.2 °C and that the mean global point to footprint sampling uncertainty is 0.15 °C, derived as in Alerskans et al. (2020) taking into account the footprint of the AMSR-E instrument. The figure shows reasonable agreement between the observed uncertainties and the theoretical uncertainties, which is computed for each retrieval in the OE algorithm, but also demonstrates that the theoretical retrieval uncertainties are not representing the true uncertainties for all conditions, demonstrating the need for use of in situ observations for this type of assessment.

6. Future work

This study shows that the statistical retrieval algorithm is more robust over the range of observing conditions, when compared with the physically based OE algorithm that can include information from first guess fields. Future work should therefore aim at improving the OE algorithm. This can be done in several ways. One way is to improve the OE statistics for PMW observations e.g. by including the error covariances for the geophysical variables in the retrieval. The error covariances for the geophysical variables influencing passive microwave observations are currently not well known and it is a complete study to estimate these. One recent study has been conducted to estimate error covariance parameters (and a bias correction) using IR observations (Merchant et al., 2020) and it looks very promising. Repeating the work done by Merchant et al. (2020) should be the focus of future work to improve the OE algorithm. Another way to improve the OE algorithm is to update the forward model to more accurately model the emissivity over the given frequency range and accounting for the different observing conditions. As also recognized in Kilic et al. (2019), current ocean emissivity models have issues particularly in cold and rough waters, and future efforts should thus be devoted to address these issues. The addition of the 1.4 GHz channels in the retrieval to prepare for the CIMR observations is another interesting task that would add information about the salinity and high wind speeds to the channel combinations considered here. Note, however, that the spatial resolution of the 1.4 GHz channels on the CIMR will be significantly coarser than e.g. the 6 GHz observations, which adds a complication to the use of the 1.4 GHz channels.

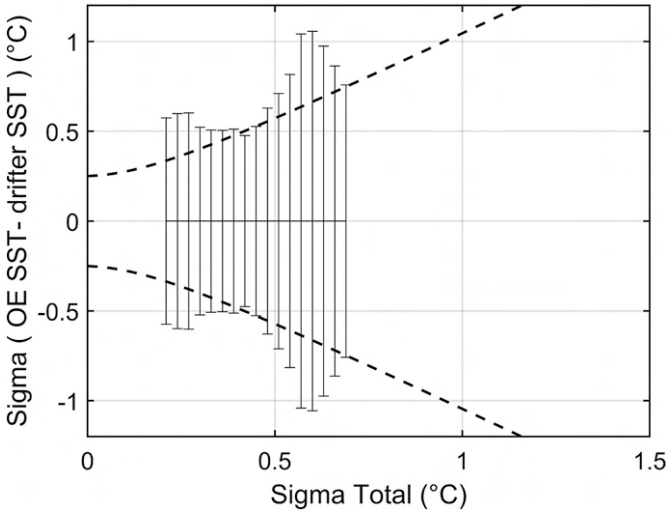


Fig. 12. Theoretical uncertainty validation with respect to drifter SST for the OE algorithm. Dashed lines show the ideal uncertainty model accounting for uncertainties in drifter SST and the sampling error. Solid black lines show one standard deviation of the retrieved minus drifter differences for each 0.03 °C bin.

Both retrieval algorithms agree that the 6, 10, 18 GHz is better than the 6, 10, 23 GHz combination. Although the 23 GHz should be considered the best channel to account for uncertainties due to water vapor variability, these results show that it might add more noise being in the absorption band, while the 18 GHz channels are able to correct for the relative small uncertainties in the 6 GHz channels due to water vapor variability. If 10 GHz is used as the primary frequency for SST retrievals, the 23 GHz may be needed as the 10 GHz channels are affected more by water vapor than the 6 GHz channels. Future studies could extend this study to investigate other channel combinations, including those omitting the lowest frequency(ies) in order to obtain better spatial resolution.

Additional future tasks could also be to examine the impact on the results shown here by using a different forward model or by using ERA5 (Copernicus Climate Change Service (C3S), 2017) instead of ERA-I reanalysis data for training the RE algorithm and for first guess in the OE algorithm, respectively. Irrespective of the different improvements made to the retrievals and the forward models, this study demonstrates the importance of comparing different types of retrieval algorithms (physical versus statistical) using in situ observations before determining the best suitable SST retrieval for a new instrument, such as CIMR.

7. Conclusion

The aim of this study is to analyze the impact of using different channel selections on retrieved sea surface temperature (SST) and to assess the performance of two different types of retrieval algorithms. The first algorithm is a physically based optimal estimation (OE) algorithm (Nielsen-Englyst et al., 2018), which inverts a forward model to retrieve SST, wind speed (WS), total column water vapor (TCWV) and total cloud liquid water (TCLW), while the second algorithm is a statistical regression (RE) based algorithm, which retrieves WS and SST (Alerskans et al., 2020). Here, we use AMSR-E brightness temperatures from 2010 to retrieve SST using different channel combinations in both algorithms, which are then compared to independent in situ SSTs.

This study demonstrated similarities with Kilic et al. (2018) using the OE algorithm, but it also demonstrated the importance of using two different types of retrieval algorithms. Information content studies are typically performed using the OE type of algorithms, and the current forward models have issues reproducing observations in cold waters (Kilic et al., 2019). Conversely, statistical retrievals, such as the RE algorithm, are the often used for generating operational and climate data records of PMW SST. The RE algorithm used here is able to compensate for the decreased sensitivity for cold SSTs, due to the nature of the statistical algorithm. This study also highlights the importance of assessing the retrievals against real reference observations, to ensure that all the different uncertainties contributions are included when drawing conclusions on the impact of using different channels for SST retrievals.

Both algorithms show (as expected) an increase in performance when more channels are included in the retrieval, since more channels

allow a better representation of the different observing conditions (e.g. cold waters). Both algorithms agree that the 6, 10, 18 GHz channel configuration is better than 6, 10, 23 GHz for SST retrievals. Furthermore, it is evident that withholding observations from the 23 and 36 GHz channels from the retrieval has the least impact on the SST performance. The proposed CIMR constellation (not considering the 1.4 GHz channels) has proven to perform very well when compared to an AMSR-E like constellation (excluding the 89 GHz channels) using both the OE and RE algorithm. The actual CIMR performance is expected to be significantly better than shown here due to smaller NEdTs and footprints. The proposed CIMR constellation includes observations at 1.4 GHz (L-band), which will further add information to both SST and WS retrievals (Reul et al., 2012, 2017), but with decreased spatial resolution and with larger influence from sea surface salinity and the ionosphere (i.e. Faraday rotation).

The comparison of the two retrieval algorithms demonstrates strengths and weaknesses related to their use. One of the strengths of the RE algorithm is its ability to compensate for the decreased sensitivity to SST in cold waters as mentioned above. The OE algorithm, on the other hand, shows very good capability during conditions with fewer channels. This can become very useful e.g. in case of instrument failure and we believe that this capability will be useful to develop alternative plan for emergency status of the instrument. Another advantage of the OE algorithm is, that it can be used to retrieve SST from day 1 for all channel combinations, while the RE algorithm requires significant number of matchups for training each combination of channels before it can be used for retrievals. This is important to consider for potential day 1 SST products from e.g. CIMR. All of above needs to be considered before an algorithm is selected for a given application and multiple utilization of the two algorithms is definitely an advantage.

Overall, new insights have been gained into the performance of the SST retrievals for different channel combinations. This is important for designing and evaluating new satellite configurations or for assessing the impact of existing satellite observations in the case of failing channels or RFI contaminated retrievals.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The two SST retrieval algorithms have been developed within the ESA-CCI SST project and this analysis has been carried out as a part of the ESA CIMR contract no. 4000125255/18/NL/AI. ICOADS data are available at https://icoads.noaa.gov/. EN4 version 4.2.0 is available at https://www.metoffice.gov.uk/hadobs/en4/. The resampled L2A data product AMSR-E V12 is produced by Remote Sensing Systems (RSS) and distributed by NASA's National Snow and Ice Data Center (NSIDC). Data are available at https://nsidc.org/data/ae_l2a. The ERA-Interim reanalysis data is available at https://www.ecmwf.int/en/forecasts/datasets/reanalysis-datasets/era-interim.

Appendix A. Appendix

$S_e =$

0.094	0.106	0.023	−0.042	−0.002	−0.119	0.092	0.002	0.081	−0.008
0.106	0.302	−0.067	−0.105	−0.075	−0.220	0.129	0.039	0.135	0.006
0.023	−0.067	0.127	0.048	0.113	−0.011	0.092	−0.071	0.084	−0.059
−0.042	−0.105	0.048	0.090	0.026	0.047	0.005	−0.025	0.000	−0.028
−0.002	−0.075	0.113	0.026	0.243	0.073	0.101	−0.148	0.132	−0.107
−0.119	−0.220	−0.011	0.047	0.073	0.275	−0.154	−0.052	−0.130	−0.011
0.092	0.129	0.092	0.005	0.101	−0.154	0.271	−0.101	0.231	−0.089
0.002	0.039	−0.071	−0.025	−0.148	−0.052	−0.101	0.118	−0.101	0.076
0.081	0.135	0.084	0.000	0.132	−0.130	0.231	−0.101	0.286	−0.120
−0.008	0.006	−0.059	−0.028	−0.107	−0.011	−0.089	0.076	−0.120	0.076

References

Alerskans, E., Høyer, J.L., Gentemann, C.L., Pedersen, L.T., Nielsen-Englyst, P., Donlon, C., 2020. Construction of a climate data record of sea surface temperature from passive microwave measurements. Remote Sens. Environ. 236, 111485. https://doi.org/10.1016/j.rse.2019.111485.

Ashcroft, P., Wentz, F.J., 2013. AMSR-E/Aqua L2A Global Swath Spatially-Resampled Brightness Temperatures (Tb), Version 3. https://doi.org/10.5067/AMSR-E/AE_L2A.003.

Bidwell, S.W., Flaming, G.M., Durning, J.F., Smith, E.A., 2005. The global precipitation measurement (GPM) microwave imager (GMI) instrument: role, performance, and status. In: Proceedings. 2005 IEEE International Geoscience and Remote Sensing Symposium, 2005. IGARSS '05., 1 IEEE, Seoul, Korea, pp. 83–86.

Braakmann-Folgmann, A., Donlon, C., 2019. Estimating snow depth on Arctic Sea ice using satellite microwave radiometry and a neural network. Cryosphere 13 (9), 2421–2438. https://doi.org/10.5194/tc-13-2421-2019.

Brasnett, B., Colan, D.S., 2016. Assimilating retrievals of sea surface temperature from VIIRS and AMSR2. J. Atmos. Ocean. Technol. 33 (2), 361–375. https://doi.org/10.1175/JTECH-D-15-0093.1.

Chang, P.S., Jelenak, Z., Alsweiss, S., 2015. Algorithm Theoretical Basis Document: GCOM-W1/AMSR2 Day-1 EDR version 1.0 [online] Available from: https://www.star.nesdis.noaa.gov/jpss/documents/ATBD/ATBD_AMSR2_Ocean_EDR_v2.0.pdf.

Chelton, D.B., Freilich, M.H., 2005. Scatterometer-based assessment of 10-m wind analyses from the operational ECMWF and NCEP numerical weather prediction models. Mon. Weather Rev. 133 (2), 409–429. https://doi.org/10.1175/MWR-2861.1.

Chelton, D.B., Wentz, F.J., 2005. Global microwave satellite observations of sea surface temperature for numerical weather prediction and climate research. Bull. Am. Meteorol. Soc. 86 (8), 1097–1115. https://doi.org/10.1175/BAMS-86-8-1097.

Ciani, D., Santoleri, R., Liberti, G.L., Prigent, C., Donlon, C., Buongiorno Nardelli, B., 2019. Copernicus imaging microwave radiometer (CIMR) benefits for the Copernicus level 4 sea-surface salinity processing chain. Remote Sens. 11 (15), 1818. https://doi.org/10.3390/rs11151818.

Copernicus Climate Change Service (C3S), 2017. ERA5: Fifth Generation of ECMWF Atmospheric Reanalyses of the Global Climate. Copernicus Climate Change Service Climate Data Store (CDS) [online] Available from: https://cds.climate.copernicus.eu/cdsapp#!/home (Accessed 9 May 2019).

Dee, D.P., Uppala, S.M., Simmons, A.J., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M.A., Balsamo, G., Bauer, P., Bechtold, P., Beljaars, A.C.M., van de Berg, L., Bidlot, J., Bormann, N., Delsol, C., Dragani, R., Fuentes, M., Geer, A.J., Haimberger, L., Healy, S.B., Hersbach, H., Hólm, E.V., Isaksen, L., Kållberg, P., Köhler, M., Matricardi, M., McNally, A.P., Monge-Sanz, B.M., Morcrette, J.J., Park, B.K., Peubey, C., de Rosnay, P., Tavolato, C., Thépaut, J.N., Vitart, F., 2011. The ERA-Interim reanalysis: Configuration and performance of the data assimilation system. Q. J. R. Meteorol. Soc. 137 (656), 553–597. https://doi.org/10.1002/qj.828.

Donlon, C., Rayner, N., Robinson, I., Poulter, D.J.S., Casey, K.S., Vazquez-Cuervo, J., Armstrong, E., Bingham, A., Arino, O., Gentemann, C., May, D., LeBorgne, P., Piollé, J., Barton, I., Beggs, H., Merchant, C.J., Heinz, S., Harris, A., Wick, G., Emery, B., Minnett, P., Evans, R., Llewellyn-Jones, D., Mutlow, C., Reynolds, R.W., Kawamura, H., 2007. The global ocean data assimilation experiment high-resolution sea surface temperature pilot project. Bull. Am. Meteorol. Soc. 88 (8), 1197–1213. https://doi.org/10.1175/BAMS-88-8-1197.

Donlon, C., Casey, K.S., Gentemann, C., Harris, A., 2010. Successes and challenges for the modern sea surface temperature observing system. In: Hall, J., Harrison, D.E., Stammer, D. (Eds.), Proceeding of OceanObs'09: Sustained Ocean Observations and Information for Society, 2. ESA Publication WPP-306, Venice Italy, 21–25 September 2009.

Donlon, C. J.: Copernicus Imaging Microwave Radiometer (CIMR) Mission Requirements Document. Version 3.0., [online] Available from: https://cimr.eu/sites/cimr.met.

no/files/documents/CIMR-MRD-v3.0-20190930-Issued.pdf (Accessed 9 January 2020), 2019.
Draper, D.W., Newell, D.A., Wentz, F.J., Krimchansky, S., Skofronick-Jackson, G.M., 2015. The global precipitation measurement (GPM) microwave imager (GMI): instrument overview and early on-orbit performance. IEEE J. Sel. Top. Appl. Earth Observ. Remote Sens. 8 (7), 3452–3462. https://doi.org/10.1109/JSTARS.2015.2403303.
Embury, O., Merchant, C.J., Corlett, G.K., 2012. A reprocessing for climate of sea surface temperature from the along-track scanning radiometers: initial validation, accounting for skin and diurnal variability effects. Remote Sens. Environ. 116, 62–78. https://doi.org/10.1016/j.rse.2011.02.028.
Fore, A.G., Yueh, S.H., Tang, W., Stiles, B.W., Hayashi, A.K., 2016. Combined active/passive retrievals of ocean vector wind and sea surface salinity with SMAP. IEEE Trans. Geosci. Remote Sens. 54 (12), 7396–7404. https://doi.org/10.1109/TGRS.2016.2601486.
Gentemann, C.L., 2014. Three way validation of MODIS and AMSR-E Sea surface temperatures. J. Geophys. Res. Oceans 119 (4), 2583–2598. https://doi.org/10.1002/2013JC009716.
Gentemann, C.L., Hilburn, K.A., 2015. In situ validation of sea surface temperatures from the GCOM-W1 AMSR2 RSS calibrated brightness temperatures: validation of RSS GCOM-W1 SST. J. Geophys. Res. Oceans 120 (5), 3567–3585. https://doi.org/10.1002/2014JC010574.
Gentemann, C.L., Donlon, C.J., Stuart-Menteth, A., Wentz, F.J., 2003. Diurnal signals in satellite sea surface temperature measurements. Geophys. Res. Lett. 30 (3) https://doi.org/10.1029/2002GL016291.
Gentemann, C.L., Meissner, T., Wentz, F.J., 2010. Accuracy of Satellite Sea surface temperatures at 7 and 11 GHz. IEEE Trans. Geosci. Remote Sens. 48 (3), 1009–1018. https://doi.org/10.1109/TGRS.2009.2030322.
Gladkova, I., Ignatov, A., Shahriar, F., Kihai, Y., Hillger, D., Petrenko, B., 2016. Improved VIIRS and MODIS SST imagery. Remote Sens. 8 (1), 79.
Good, S.A., Martin, M.J., Rayner, N.A., 2013. EN4: quality controlled ocean temperature and salinity profiles and monthly objective analyses with uncertainty estimates: THE EN4 DATA SET. J. Geophys. Res. Oceans 118 (12), 6704–6716. https://doi.org/10.1002/2013JC009067.
Hirabayashi, T., 2020. Contribution of JAXA's Earth Observation Missions to Water Cycle and Climate Studies, Disaster Mitigation, and Operational Applications, EGU General Assembly 2020, Online, 4–8 May 2020, EGU2020–19165. https://doi.org/10.5194/egusphere-egu2020-19165.
Hollinger, J., 1971. Passive microwave measurements of sea surface roughness. IEEE Trans. Geosci. Electron. 9 (3), 165–169. https://doi.org/10.1109/TGE.1971.271489.
Imaoka, K., Kachi, M., Kasahara, M., Ito, N., Nakagawa, K., Oki, T., 2010. Instrument performance and calibration of AMSR-E and AMSR2, international archives of the photogrammetry. Remote Sens. Spat. Inform. Sci. 38 (8), 13–18.
Jakobson, E., Vihma, T., Palo, T., Jakobson, L., Keernik, H., Jaagus, J., 2012. Validation of Atmospheric Reanalyses over the Central Arctic Ocean: tara reanalyses validation. Geophys. Res. Lett. 39 (10) https://doi.org/10.1029/2012GL051591. N/a-N/a.
Jiang, J.H., Su, H., Zhai, C., Perun, V.S., Del Genio, A., Nazarenko, L.S., Donner, L.J., Horowitz, L., Seman, C., Cole, J., Gettelman, A., Ringer, M.A., Rotstayn, L., Jeffrey, S., Wu, T., Brient, F., Dufresne, J.-L., Kawai, H., Koshiro, T., Watanabe, M., LÉcuyer, T.S., Volodin, E.M., Iversen, T., Drange, H., Mesquita, M.D.S., Read, W.G., Waters, J.W., Tian, B., Teixeira, J., Stephens, G.L., 2012. Evaluation of cloud and water vapor simulations in CMIP5 climate models using NASA "A-Train" satellite observations: EVALUATION OF IPCC AR5 MODEL SIMULATIONS. J. Geophys. Res.-Atmos. 117 (D14) https://doi.org/10.1029/2011JD017237.
Kawanishi, T., Sezai, T., Ito, Y., Imaoka, K., Takeshima, T., Ishido, Y., Shibata, A., Miura, M., Inahata, H., Spencer, R.W., 2003. The advanced microwave scanning radiometer for the earth observing system (AMSR-E), NASDA's contribution to the EOS for global energy and water cycle studies. IEEE Trans. Geosci. Remote Sens. 41 (2), 184–194. https://doi.org/10.1109/TGRS.2002.808331.
Kerr, Y.H., Waldteufel, P., Wigneron, J.-P., Delwart, S., Cabot, F., Boutin, J., Escorihuela, M.-J., Font, J., Reul, N., Gruhier, C., Juglea, S.E., Drinkwater, M.R., Hahne, A., Martín-Neira, M., Mecklenburg, S., 2010. The SMOS Mission: new tool for monitoring key elements ofthe global water cycle. Proc. IEEE 98 (5), 666–687. https://doi.org/10.1109/JPROC.2010.2043032.
Kilic, L., Prigent, C., Aires, F., Boutin, J., Heygster, G., Tonboe, R.T., Roquet, H., Jimenez, C., Donlon, C., 2018. Expected performances of the Copernicus imaging microwave radiometer (CIMR) for an all-weather and high spatial resolution estimation of ocean and sea ice parameters. J. Geophys. Res. Oceans 123 (10), 7564–7580. https://doi.org/10.1029/2018JC014408.
Kilic, L., Prigent, C., Boutin, J., Meissner, T., English, S., Yueh, S., 2019. Comparisons of ocean radiative transfer models with SMAP and AMSR2 observations. J. Geophys. Res. Oceans 124 (11), 7683–7699. https://doi.org/10.1029/2019JC015493.
Kummerow, C., Barnes, W., Kozu, T., Shiue, J., Simpson, J., 1998. The tropical rainfall measuring mission (TRMM) sensor package. J. Atmos. Ocean. Technol. 15, 809–817.
Le Traon, P.Y., Antoine, D., Bentamy, A., Bonekamp, H., Breivik, L.A., Chapron, B., Corlett, G., Dibarboure, G., Digiacomo, P., Donlon, C.J., Faugère, Y., Font, J., Girard-Ardhuin, F., Gohin, F., Johannessen, J.A., Kamachi, M., Lagerloef, G., Lambin, J., Larnicol, G., Le Borgne, P., Leuliette, E., Lindstrom, E., Martin, M.J., Maturi, E., Miller, L., Mingsen, L., Morrow, R., Reul, N., Rio, M.H., Roquet, H., Santoleri, R., Wilkin, J., 2015. Use of satellite observations for operational oceanography: recent achievements and future prospects. J. Operat. Oceanogr. 8 (sup 1), 12–27. https://doi.org/10.1080/1755876X.2015.1022050.
Li, J.-L.F., Waliser, D., Woods, C., Teixeira, J., Bacmeister, J., Chern, J., Shen, B.-W., Tompkins, A., Tao, W.-K., Köhler, M., 2008. Comparisons of satellites liquid water estimates to ECMWF and GMAO analyses, 20th century IPCC AR4 climate simulations, and GCM simulations. Geophys. Res. Lett. 35 (19) https://doi.org/10.1029/2008GL035427.
Liang, X., Yang, Q., Nerger, L., Losa, S.N., Zhao, B., Zheng, F., Zhang, L., Wu, L., 2017. Assimilating Copernicus SST data into a pan-Arctic Ice-Ocean coupled model with a local SEIK filter. J. Atmos. Ocean. Technol. 34 (9), 1985–1999. https://doi.org/10.1175/JTECH-D-16-0166.1.
Lipes, R.G., 1982. Description of SEASAT radiometer status and results. J. Geophys. Res. 87 (C5), 3385. https://doi.org/10.1029/JC087iC05p03385.
Meissner, T., Wentz, F., 2002. An updated analysis of the ocean surface wind direction signal in passive microwave brightness temperatures. IEEE Trans. Geosci. Remote Sens. 40 (6), 1230–1240. https://doi.org/10.1109/TGRS.2002.800231.
Meissner, T., Wentz, F.J., 2012. The emissivity of the ocean surface between 6 and 90 GHz over a large range of wind speeds and earth incidence angles. IEEE Trans. Geosci. Remote Sens. 50 (8), 3004–3026. https://doi.org/10.1109/TGRS.2011.2179662.
Meissner, T., Ricciardulli, L., Wentz, F.J., 2017. Capability of the SMAP Mission to measure ocean surface winds in storms. Bull. Am. Meteor. Soc. 98 (8), 1660–1677. https://doi.org/10.1175/BAMS-D-16-0052.1.
Merchant, C.J., Harris, A.R., Murray, M.J., Zavody, A.M., 1999. Toward the elimination of bias in satellite retrievals of skin sea surface temperature 1. Theory, modeling and inter-algorithm comparison. J. Geophys. Res. Oceans 104 (C10), 23565–23578. https://doi.org/10.1029/1999JC900105.
Merchant, C.J., Embury, O., Le Borgne, P., Bellec, B., 2006. Saharan dust in night-time thermal imagery: detection and reduction of related biases in retrieved sea surface temperature. Remote Sens. Environ. 104, 15–30.
Merchant, C.J., Embury, O., Roberts-Jones, J., Fiedler, E., Bulgin, C.E., Corlett, G.K., Good, S., McLaren, A., Rayner, N., Morak-Bozzo, S., Donlon, C., 2014. Sea surface temperature datasets for climate applications from phase 1 of the European Space Agency climate change initiative (SST CCI). Geosci. Data J. 1 (2), 179–191. https://doi.org/10.1002/gdj3.20.
Merchant, C.J., Embury, O., Bulgin, C.E., Block, T., Corlett, G.K., Fiedler, E., Good, S.A., Mittaz, J., Rayner, N.A., Berry, D., Eastwood, S., Taylor, M., Tsushima, Y., Waterfall, A., Wilson, R., Donlon, C., 2019. Satellite-based time-series of sea-surface temperature since 1981 for climate applications. Sci. Data 6 (1), 223. https://doi.org/10.1038/s41597-019-0236-x.
Merchant, C.J., Saux-Picart, S., Waller, J., 2020. Bias correction and covariance parameters for optimal estimation by exploiting matched in-situ references. Remote Sens. Environ. 237, 111590. https://doi.org/10.1016/j.rse.2019.111590.
Milman, A.S., Wilheit, T.T., 1985. Sea surface temperatures from the scanning multichannel microwave radiometer on Nimbus 7. J. Geophys. Res. 90 (C6), 11631. https://doi.org/10.1029/JC090iC06p11631.
Minnett, P.J., Alvera-Azcárate, A., Chin, T.M., Corlett, G.K., Gentemann, C.L., Karagali, I., Li, X., Marsouin, A., Marullo, S., Maturi, E., Santoleri, R., Saux Picart, S., Steele, M., Vazquez-Cuervo, J., 2019. Half a century of satellite remote sensing of sea-surface temperature. Remote Sens. Environ. 233, 111366. https://doi.org/10.1016/j.rse.2019.111366.
Monzikova, A.K., Kudryavtsev, V.N., Reul, N., Chapron, B., 2017. On the upper ocean response to tropical cyclones: Satellite microwave observation. In: Progress in Electromagnetics Research Symposium-Fall (PIERS-FALL). IEEE, pp. 2437–2444.
Nielsen-Englyst, P., Høyer, L.J., Toudal Pedersen, L.L., Gentemann, C., Alerskans, E., Block, T., Donlon, C., 2018. Optimal estimation of sea surface temperature from AMSR-E. Remote Sens. 10 (2), 229. https://doi.org/10.3390/rs10020229.
Ning, J., Qing, X., Wang, T., Zhang, S., 2018. Upper Ocean response to super typhoon soudelor revealed by different SST products. In: IGARSS 2018–2018 IEEE International Geoscience and Remote Sensing Symposium. IEEE.
O'Carroll, A.G., Armstrong, E.M., Beggs, H.M., Bouali, M., Casey, K.S., Corlett, G.K., Dash, P., Donlon, C.J., Gentemann, C.L., Høyer, J.L., Ignatov, A., Kabobah, K., Kachi, M., Kurihara, Y., Karagali, I., Maturi, E., Merchant, C.J., Marullo, S., Minnett, P.J., Pennybacker, M., Ramakrishnan, B., Ramsankaran, R., Santoleri, R., Sunder, S., Saux Picart, S., Vázquez-Cuervo, J., Wimmer, W., 2019. Observational needs of sea surface temperature. Front. Mar. Sci. 6, 420. https://doi.org/10.3389/fmars.2019.00420.
Pearson, K., Merchant, C., Embury, O., Donlon, C., 2018. The role of advanced microwave scanning radiometer 2 channels within an optimal estimation scheme for sea surface temperature. Remote Sens. 10 (2), 90. https://doi.org/10.3390/rs10010090.
Pearson, K., Good, S., Merchant, C.J., Prigent, C., Embury, O., Donlon, C., 2019. Sea surface temperature in global analyses: gains from the Copernicus imaging microwave radiometer. Remote Sens. 11 (20), 2362. https://doi.org/10.3390/rs11202362.
Prigent, C., Aires, F., Bernardo, F., Orlhac, J.-C., Goutoule, J.-M., Roquet, H., Donlon, C., 2013. Analysis of the potential and limitations of microwave radiometry for the retrieval of sea surface temperature: definition of MICROWAT, a new mission concept: sea temperature from microwaves. J. Geophys. Res. Oceans 118 (6), 3074–3086. https://doi.org/10.1002/jgrc.20222.
Reul, N., Tenerelli, J., Chapron, B., Vandemark, D., Quilfen, Y., Kerr, Y., 2012. SMOS satellite L-band radiometer: A new capability for ocean surface remote sensing in hurricanes: SMOS L-BAND radiometer and high winds. J. Geophys. Res. 117 (C2) https://doi.org/10.1029/2011JC007474 n/a-n/a.
Reul, N., Chapron, B., Zabolotskikh, E., Donlon, C., Quilfen, Y., Guimbard, S., Piolle, J.F., 2016. A revised L-band radio-brightness sensitivity to extreme winds under tropical cyclones: the five year SMOS-storm database. Remote Sens. Environ. 180, 274–291. https://doi.org/10.1016/j.rse.2016.03.011.
Reul, N., Chapron, B., Zabolotskikh, E., Donlon, C., Mouche, A., Tenerelli, J., Collard, F., Piolle, J.F., Fore, A., Yueh, S., Cotton, J., Francis, P., Quilfen, Y., Kudryavtsev, V.,

2017. A new generation of tropical cyclone size measurements from space. Bull. Am. Meteor. Soc. 98 (11), 2367–2385. https://doi.org/10.1175/BAMS-D-15-00291.1.

Reynolds, R.W., Rayner, N.A., Smith, T.M., Stokes, D.C., Wang, W., 2002. An improved in situ and satellite SST analysis for climate. J. Clim. 15 (13), 1609–1625. https://doi.org/10.1175/1520-0442(2002)015<1609:AIISAS>2.0.CO;2.

Rodgers, C.D., 2000. Inverse Methods for Atmospheric Sounding - Theory and Practice. World Scientific Publishing Co. Pte. Ltd.

Scarlat, R.C., Spreen, G., Heygster, G., Huntemann, M., Paţilea, C., Pedersen, L.T., Saldo, R., 2020. Sea Ice and atmospheric parameter retrieval from satellite microwave radiometers: synergy of AMSR2 and SMOS compared with the CIMR candidate mission. J. Geophys. Res. Oceans 125 (3). https://doi.org/10.1029/2019JC015749.

Shibata, A., 2006. Features of ocean microwave emission changed by wind at 6 GHz. J. Oceanogr. 62 (3), 321–330. https://doi.org/10.1007/s10872-006-0057-3.

Ulaby, F.T., Moore, R.K., Fung, A.K., 1981. Microwave Remote Sensing: Active and Passive. Volume 1-Microwave Remote Sensing Fundamentals and Radiometry.

Vázquez-Cuervo, J., Armstrong, E.M., Harris, A., 2004. The effect of aerosols and clouds on the retrieval of infrared sea surface temperatures. J. Clim. 17 (20), 3921–3933. https://doi.org/10.1175/1520-0442(2004)017<3921:TEOAAC>2.0.CO;2.

Wentz, F.J., 2015. A 17-Yr climate record of environmental parameters derived from the tropical rainfall measuring Mission (TRMM) microwave imager. J. Clim. 28 (17), 6882–6902. https://doi.org/10.1175/JCLI-D-15-0155.1.

Wentz, F.J., Meissner, T., 2000. AMSR Ocean Algorithm. Algorithm Theoretical Basis Document, Remote Sensing Systems, Santa Rosa, CA.

Wentz, F.J., Meissner, T., 2007. Supplement 1 Algorithm Theoretical Basis Document for AMSR-E Ocean Algorithms. RSS Technical Report 051707. Remote Sensing Systems, Santa Rosa, CA [online] Available from: http://images.remss.com/papers/rsstech/2007_051707_Wentz_AMSR_Ocean_Algorithm_Version_2_Supplement1_ATBD.pdf.

Wentz, F.J., Gentemann, C., Smith, D., Chelton, D.B., 2000. Satellite measurements of sea surface temperature through clouds. Science 288 (5467), 847–850. https://doi.org/10.1126/science.288.5467.847.

Wilheit, T., 1979. A model for the microwave emissivity of the Ocean's surface as a function of wind speed. IEEE Trans. Geosci. Electron. 17 (4), 244–249. https://doi.org/10.1109/TGE.1979.294653.

Woodruff, S.D., Worley, S.J., Lubker, S.J., Ji, Z., Eric Freeman, J., Berry, D.I., Brohan, P., Kent, E.C., Reynolds, R.W., Smith, S.R., Wilkinson, C., 2011. ICOADS Release 2.5: extensions and enhancements to the surface marine meteorological archive. Int. J. Climatol. 31 (7), 951–967. https://doi.org/10.1002/joc.2103.

Yang, C.-S., Kim, S.-H., Ouchi, K., Back, J.-H., 2015. Generation of high resolution sea surface temperature using multi-satellite data for operational oceanography. Acta Oceanol. Sin. 34 (7), 74–88. https://doi.org/10.1007/s13131-015-0694-8.

Zhang, L., Yu, H., Wang, Z., Yin, X., Yang, L., Du, H., Li, B., Wang, Y., Zhou, W., 2020. Evaluation of the initial sea surface temperature from the HY-2B scanning microwave radiometer. IEEE Geosci. Remote Sensing Lett. 1–5. https://doi.org/10.1109/LGRS.2020.2968635.

APPENDIX D

Paper IV: A combined sea and sea-ice surface temperature climate dataset of the Arctic, 1982-2021

This appendix contains a full copy of Paper IV:

- **Nielsen-Englyst, Pia**, Jacob L. Høyer, Wiebke M. Kolbe, Gorm Dybkjær, Thomas Lavergne, Rasmus Tage Tonboe, Sotirios Skarpalezos, Ioanna Karagali. 2023. "A combined sea and sea-ice surface temperature climate dataset of the Arctic, 1982–2021", *Remote Sensing of Environment*, Volume 284, 113331, ISSN 0034-4257.

Remote Sensing of Environment 284 (2023) 113331

Contents lists available at ScienceDirect

Remote Sensing of Environment

journal homepage: www.elsevier.com/locate/rse

A combined sea and sea-ice surface temperature climate dataset of the Arctic, 1982–2021

Pia Nielsen-Englyst [a,b,*], Jacob L. Høyer [b], Wiebke M. Kolbe [a,b], Gorm Dybkjær [b], Thomas Lavergne [c], Rasmus Tage Tonboe [a], Sotirios Skarpalezos [b], Ioanna Karagali [b]

[a] DTU-Space, Technical University of Denmark, DK-2800 Lyngby, Denmark
[b] National Centre for Climate Research, Danish Meteorological Institute, Lyngbyvej 100, DK-2100 Copenhagen Ø, Denmark
[c] Research and Development Department, Norwegian Meteorological Institute, Oslo, Norway

ARTICLE INFO

Editor: Menghua Wang

Keywords:
Arctic
Sea surface temperature
Sea ice surface temperature
Infrared satellite observations
Optimal interpolation
Climate change

ABSTRACT

The surface temperature is one of the main parameters for assessing climate change. Temperature change is most pronounced in the Arctic, and therefore, it is crucial to accurately estimate sea and sea ice surface temperatures in this region. The availability of in situ observations is limited in the Arctic, thus, increasing the need for satellite observations to estimate surface temperatures. We present the first Arctic (>58°N) gap-free climate dataset covering the surface temperatures of the ocean, sea ice and the marginal ice zone from 1 January 1982 to 31 May 2021 based on observations from infrared satellite sensors. The underlying algorithm combines the multi-satellite observations and performs a statistical optimal interpolation to obtain daily gap-free fields, with a spatial resolution of 0.05° in latitude and longitude. In situ observations have been used to derive consistent validation statistics over the ocean and sea ice. Comparison of the derived sea surface temperatures against in situ measurements from drifting buoys, moored buoys and Argo floats shows mean differences of 0.01 °C, 0.04 °C and 0.04 °C and standard deviations of 0.54 °C, 0.56 °C and 0.51 °C, respectively. Over sea ice, the derived ice surface temperatures have been compared with KT-19 measurements from IceBridge flights, showing a mean difference of 1.52 °C and standard deviation of 3.12 °C, and with air temperatures from the North Pole (NP) ice drifting stations as well as ECMWF distributed buoys and CRREL buoys, with mean differences of −2.35 °C, −3.21 °C and −2.87 °C and standard deviations of 3.12 °C, 3.34 °C and 3.36 °C, respectively. The combination of sea and sea-ice surface temperature provides a consistent dataset for climate analysis, which is crucial for studying climate change and trends in the Arctic. The combined sea and sea-ice surface temperature of the Arctic has risen with about 4.5 °C over the period 1982–2021, with a peak warming of around 10 °C in the northeastern Barents Sea.

1. Introduction

The surface temperature is one of the main variables for assessing climate change (Bates and Diaz, 1991; Bojinski et al., 2014; Folland et al., 2001; Kaplan et al., 1998). This is also true for the Arctic, where positive feedback mechanisms, e.g. the ice-albedo feedback, amplify climate change (AMAP, 2021; Comiso and Hall, 2014; Graversen et al., 2008; Hall et al., 2004; IPCC, 2019; Pithan and Mauritsen, 2014; Rantanen et al., 2022; Richter-Menge et al., 2017). In the Arctic Ocean, the surface temperatures play a crucial role for the heat exchange between the ocean and atmosphere, sea ice growth and melt processes (Key et al., 1997; Maykut, 1986) as well as in weather and sea ice forecasts through assimilation into ocean and atmospheric models (Carton et al., 2000; Larsen et al., 2007; Oke et al., 2008; Rasmussen et al., 2018; Rayner, 2003; Song and Yu, 2012). Therefore, it is important to accurately estimate the surface temperature of the Arctic Ocean, which consists of open ocean, sea ice and a marginal ice zone (MIZ), i.e. the transitional region between the open ocean and pack ice. The extreme environment and the poor accessibility (especially in the winter season) make in situ observations challenging and sparse in the Arctic (Centurioni et al., 2019; Donlon et al., 2012). Polar orbiting satellites offer a very good alternative and addition to the in situ observations through their high spatial and temporal coverage of the Arctic.

A large number of global gap-free (optimally interpolated) gridded

* Corresponding author at: DTU-Space, Technical University of Denmark, DK-2800 Lyngby, Denmark.
E-mail address: pne@dmi.dk (P. Nielsen-Englyst).

https://doi.org/10.1016/j.rse.2022.113331
Received 1 April 2022; Received in revised form 14 September 2022; Accepted 26 October 2022
Available online 12 November 2022

sea surface temperature (SST) products (Level 4, L4, analyses) have been developed based on a variety of different satellite observations and sometimes including in situ observations as well (Donlon et al., 2012; Good et al., 2020; Merchant et al., 2014, 2019; Reynolds et al., 2002, 2007; Roberts-Jones et al., 2012). Differences are known to exist among the analyses due to varying input data (both infrared and microwave satellite and in situ), quality control procedures, cloud-masks, land/ice masks, interpolation techniques and related configurations (e.g. correlation scales and grid sizes). Several inter-comparison studies have been performed to understand the consistency and discrepancy of the different L4 SST analyses (Dash et al., 2012; Fiedler et al., 2019; Martin et al., 2012; Okuro et al., 2014; Yang et al., 2021). The L4 SST analyses perform fairly uniformly globally, with accuracies better than 0.5 °C in clear-sky conditions (Martin et al., 2012; Petrenko et al., 2014; Wang et al., 2016), but large uncertainties and significant differences are found at high latitudes in particular in coastal and dynamic regions (Castro et al., 2016; Dash et al., 2012; Vazquez-Cuervo et al., 2022). The extreme atmospheric and oceanographic conditions, sparsely distributed in situ observations, and persistent cloud cover complicate accurate SST satellite retrievals in the Arctic (Donlon et al., 2010; Høyer et al., 2012). Specialized high latitude algorithms have been developed and shown to outperform the global analyses in these regions (Jia and Minnett, 2020; Vincent et al., 2008a, 2008b). Improving the SST data quality in the Arctic has been identified by the SST community as being of high priority for future SST research and developments (O'Carroll et al., 2019).

The presence of both seawater and ice, and the large seasonal and inter-annual fluctuations in the sea ice cover, lead to increased complexity in the SST mapping of the Arctic region. Traditionally, global gridded SST products use independent satellite observed sea ice concentration (SIC) to derive an ice mask. Few products simply do not report SST in ice covered waters, while most products provide a foundation or mixed layer temperature of the sea water just below the ice, which is assumed to be at the freezing point of seawater (−1.8 °C for salinities around 35 psu) for high SICs. Different methods and empirical relationships have been used to generate proxy SSTs from SICs in the MIZ (Banzon et al., 2020), which in some cases are blended with the closest open water SSTs (Rayner, 2003; Reynolds et al., 2007, 2002).

However, the use of under ice SSTs in the MIZ and in sea ice covered regions does not represent the surface temperature as very large temperature gradients can exist through the sea ice and snow (Tonboe et al., 2011). These products cannot be used to estimate the surface heat exchange with the atmosphere, and thus, not appropriate for use as boundary conditions in e.g. atmospheric models or to assess the surface temperature changes in the Arctic Ocean. Combining SST and ice surface temperature (IST) is identified as the most appropriate method for determining the surface temperature of the Arctic (Minnett et al., 2019).

Ice surface temperature retrievals from satellites are based on thermal infrared (IR) data, which implies that satellite IST products represent clear sky temperatures. However, automatic cloud screening over Arctic sea ice is challenging due to the resemblance of ice and cloud top temperatures. Undetected clouds or atmospheric ice/water typically results in larger IST errors than for SST retrievals, including a cold bias (Dybkjær et al., 2012; Hall et al., 2004). A number of satellite-based IST products are available (Hall et al., 2004; Key et al., 1994, 1997, 2013; Maslanik et al., 1997) and some combine Arctic SST and IST (Dybkjær et al., 2012; Vincent et al., 2008b), while few products include land surface temperatures as well (Comiso, 2003; Comiso and Hall, 2014; Dodd et al., 2019). A Near-Real-Time (NRT) gap-free (all-sky) combined SST, IST and MIZT product (L4 SST/IST) has been produced for a few years within the Copernicus Marine Service (CMEMS; doi: https://doi.org/10.48670/moi-00130), however, no L4 SST/IST reanalysis or climate data record (CDR) has previously been generated for the Arctic Ocean, based on IR satellite observations. Significant differences in surface temperatures are observed during all-sky and clear-sky conditions (Nielsen-Englyst et al., 2019; Walsh and Chapman, 1998). Therefore, to study climate trends, it is important with an effective way to fill in the gaps (due to clouds) since inter-annual variations in cloud cover may impact the accuracy of the trends observed when only clear-sky data are used (Liu et al., 2008).

Previously, climate trends have been estimated individually for SST and IST records (e.g. (Bulgin et al., 2020; Comiso, 2003; Comiso and Hall, 2014; Merchant et al., 2019; Wang and Key, 2005). However, this is problematic in the Arctic region due to the large temporal variability in the sea ice cover including the overlying northward migration of the ice edge on decadal timescales, and thus, the resulting climate trends are not easy to interpret (Comiso, 2003). A combined surface temperature dataset of the ocean, sea ice and the MIZ provides a consistent climate indicator, which is important for studying climate trends in the Arctic region.

This paper presents the generation, validation and analysis of the first gap-free, combined sea and sea-ice surface temperature climate dataset, with CDR-like temporal consistency, for the Arctic (>58°N) covering the period from 1 January 1982 to 31 May 2021. It is based on IR satellite products of surface temperatures from the European Space Agency's Climate Change Initiative (ESA CCI), the Copernicus Climate Change Service (C3S) and the Arctic and Antarctic ice Surface Temperatures from thermal Infrared (AASTI). A multiplatform optimal interpolation (OI) scheme has been used to combine these data sources and fill the gaps due to clouds. It has been developed considering the conditions that apply in the Arctic with special attention to the MIZ, where it produces a combination of open water SST and IST. Each surface type (i.e. sea ice, open ocean and MIZ), has its own characteristics, and thus, it is very important with an accurate SIC product to identify the different regions. Because of this, a combination of several SIC products and additional filtering and consistency checks between the SST and SIC fields have been developed to reduce erroneous SIC observations and produce an improved SIC product in particular close to the coasts.

The paper is organized such that Section 2, describes the satellite and in situ data used for construction and validation of the dataset. In Section 3, the methods and processing steps of the SIC and the L4 SST/IST fields are described. The results, validation, and analysis of the L4 SST/IST are provided in Section 4, while discussion and conclusions are given in Section 5 and Section 6, respectively.

2. Data

Multiple sources of satellite observations are used over the sea and sea ice surfaces in the Arctic in order to provide IST and SST information for the L4 SST/IST processing. Fig. 1 shows the temporal coverage of the individual SST/IST satellite products used as input to the Danish Meteorological Institute Optimal Interpolation (DMIOI) L4 processing system and the in situ observations used for validation. Moreover, it shows the temporal coverage of the input SIC products and the auxiliary data used in the generation of the OI SIC product (see Section 3.a). Details about each data source are provided below.

2.1. Satellite data

2.1.1. Sea and sea ice surface temperature

2.1.1.1. SST CCI L2P. The SST CCI version 2.1 data are used for the period 1982–2016, and they are obtained through the ESA CCI project (Merchant et al., 2019). The SST CCI data include observations from the ATSR 1 instrument on board the ERS-1 satellite, ATSR 2 on board the ERS-2 satellite, and the AATSR on board ENVISAT, and the AVHRR on board the NOAA and Metop satellites. The CCI SST retrievals provide the temperature at 20 cm depth at the nearest of 10:30 h or 22.30 h local time to best represent the daily mean (Embury et al., 2012; Merchant et al., 2019).

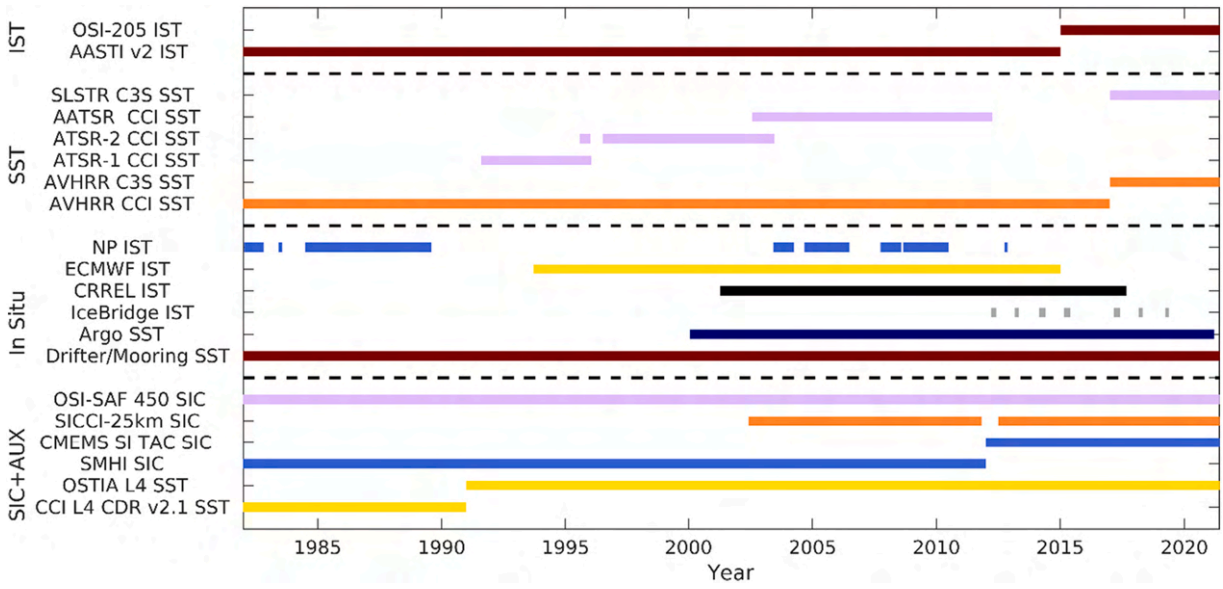


Fig. 1. Temporal coverage of the input satellite IST and SST data, the in situ IST and SST observations as well as the input SIC products and the auxiliary data used for the OI SIC generation.

2.1.1.2. C3S SST. For the period 2017–2021, SST observations from C3S are used. From 2017 to August 31, 2019, the C3S data are provided in satellite swath coordinates, referred to as Level 2 (L2), while the data are provided as gridded single L2 files, referred to as Level 3 uncollated (L3U) files after this period. The C3S data is obtained from Owen Embury (personal communication, 2018) and corresponds to the L3U data files available from https://cds.climate.copernicus.eu/ except for the higher spatial resolution. The C3S data include observations from the SLSTR A/B instruments on board the Sentinel 3 satellites and the AVHRRs on board the NOAA and Metop satellites.

2.1.1.3. AASTI + OSI-205 IST. IST observations are obtained by combining two data sources; a climate dataset and data from a NRT data stream. The Arctic and Antarctic ice Surface Temperatures from thermal Infrared satellite climate dataset version 2 (AASTI v2) covering the period 1982–2014 (Dybkjær et al., 2014, 2022) and data from the operational OSISAF IST product (OSI-205) covering the period from 2015 and onwards (Dybkjær et al., 2018). The AASTI and OSI-205 data are retrieved by identical algorithms from a single sensor type, the AVHRR sensors, where the applied algorithms are calibrated specifically for each sensor. The data format for the two datasets differ, as AASTI uses the Global Area Coverage (GAC) and the OSI-205 uses the Local Area Coverage (LAC), at approximately 4 and 1 km resolution, respectively. In Dybkjær et al. (2018) the retrieval system is described in detail.

A critical step for IST retrievals is the cloud detection procedure. For this purpose, the cloud mask from the EUMETSAT Nowcasting Satellite Application Facility (NWCSAF) Polar Platform System version 2014 (PPS2014) is used (Dybbroe et al., 2005a, 2005b; SMHI, 2014). The PPS has for years been used for automatic cloud masking for AVHRR and other sensors (Dybbroe et al., 2005b). However, cloud mask quality over ice during non-sunlit hours is low due to the similarity of cloud tops and ice surfaces at IR wavelengths.

2.1.2. Sea ice concentration

The sea ice concentration (SIC) field uses the EUMETSAT OSISAF Global SIC CDR v2 product OSI-450 covering the period 1979–2015, and the OSI-430b the Interim CDR extension from 2016 onwards. The product is derived from coarse resolution (30-60 km) passive microwave (PMW) satellite data (SMMR, SSM/I and SSMIS). We hereafter label this whole time series as OSI-450, although the OSI-430b period is involved as well. In addition, we use data from the ESA CCI programme SICCI-25km product, which uses medium-resolution (15–25 km) PMW satellite data from AMSR-E (June 2002 to October 2011) and AMSR2 (July 2012 to May 2017). An extension of the SICCI-25km processing has been performed to provide consistent SIC fields after May 2017 using AMSR2 data. We hereafter label both the SICCI-25km and its extension as SICCI-25km. OSI-450 and SICCI-25km share the same algorithms, processing chains and data format (Lavergne et al., 2019), and continue earlier work by Andersen et al. (2007) and Tonboe et al. (2016). Both OSI-450 and SICCI-25km are presented at 25 km grid spacing. However, the true spatial resolution of OSI-450 is coarser due to the larger footprints of the SMMR, SSM/I and SSMIS channels used by the algorithm. The true spatial resolution of the SICCI-25km product is on the order of its grid spacing: 25 km. As other such SIC CDRs based on PMW satellite data, OSI-450 and SICCI-25km have challenges in coastal regions (Lavergne et al., 2019). For that reason, the SIC fields for the Baltic Sea consist of a SIC product based on ice charts from the Swedish Meteorological and Hydrological Institute (SMHI; 1982–2011) and the CMEMS 1 km SIC fields (2012-present), which include high resolution sea ice information from the Swedish and Finnish ice services. These Baltic Sea SIC products are similar to those used in the regional SST and IST CDR produced for the Baltic region (Høyer and Karagali, 2016).

2.1.3. Auxiliary data

An independent daily SST product has been used to filter the SIC product. SST from ESA SST CCI L4 Analysis CDR v2.1 is used from 1982 to 1990 (Good et al., 2019; Merchant et al., 2019), while the ESA SST CCI Analysis Long Term Product (Merchant et al., 2016) is used for the period 1991 to 2010, and the Operational Sea Surface Temperature and Sea Ice Analysis (OSTIA; Donlon et al., 2012; Good et al., 2020) is used from 2011 and onwards. These auxiliary products are obtained from the CMEMS catalogue (https://marine.copernicus.eu/) and they have a spatial resolution of 0.05° in latitude and longitude. The combined and independent SST product will be referred to as OSTIA/CCI in this paper. The OSTIA/CCI product is only included for filtering the SIC products and not as an input to the L4 SST generation.

2.2. In situ data

2.2.1. Sea surface temperature

For the SST validation, data from drifting buoys, moored buoys and Argo floats are obtained from the Hadley Centre Integrated Ocean

Database (HADIOD; Atkinson et al., 2014) version 1.2.0.0. The Argo and drifter observations are well represented in the Arctic region open waters, while the moorings are concentrated in the North Atlantic and Greenland Sea.

2.2.2. Ice surface temperature

For validation of IST, we use data from 116 drifting buoys obtained from the Meteorological Archival and Retrieval System (MARS) at the European Centre for Medium-Range Weather Forecasts (ECMWF) covering the period 1993–2015 (hereafter referred to as ECMWF buoys). These are supplemented with data from 96 U.S. Army Cold Regions Research Engineering Laboratory (CRREL) mass balance buoys for the period 2001–2017 (Perovich et al., 2016; Richter-Menge et al., 2006) as well as measurements from 14 Russian North Pole (NP) drifting ice stations mainly for the period 1982–1989 but also some for the period 2003–2012 (RU/FSR/HME/AARI and NSIDC, 1993). The temperature observations from NP, ECMWF and CRREL are air temperatures measured at different heights (typically around 2 m above the surface) depending on e.g. snow accumulation, snow drift and snow melt. These temperature observations have been inspected and quality controlled manually for data artefacts. The ECMWF and CRREL observations have previously been used for validation of satellite ISTs and corresponding 2 m air temperature estimates within the European Union's Horizon2020 project EUSTACE (EU Surface Temperatures for All Corners of Earth; Nielsen-Englyst et al., 2021; Rayner et al., 2020).

We also access 117 NASA IceBridge (IAKST1B) flights covering the period 2012–2019, typically conducted during March–May (version 2; Studinger, 2020). The surface temperatures are converted from IR radiation measurements from a Heitronics KT-19 IR Radiation Pyrometer by assuming an emissivity constant of 0.97. The surface temperatures are provided by IceBridge at a spatial resolution of about 15 m, and here they have been averaged for every fifth kilometer to remove the small scale variability, which cannot be represented by the coarser resolution L4 IST product.

3. DMIOI L4 processing system

This section presents the full DMIOI L4 processing system, which integrates individual, single sensor, swath based SST and IST observations to a multi-sensor interpolated (gap-free) field. The processing sequence is outlined in Fig. 2. The OI SIC field (derived and described in Section 3.a) is used as input to identify the different surface types (i.e. ocean, sea ice and the MIZ) for each day during the record. The surface is considered as open water when SIC≤15%, ice covered when SIC>70% and as MIZ when 15 < SIC≤70%. Together with the land mask the SIC is used to construct a dynamic surface mask. This dynamic surface mask is used during the pre-processing of the input L2 + L3 IST/SST to L3 Supercollated (L3S) data, which is described in detail in Section 3.b. The surface mask is also used during the derivation of the error statistics and covariances for each surface type, which are used in the OI method for analysis of the observations (see Section 3.c). In the end, the OI method produces the daily L4 SST/IST and the corresponding uncertainties. The L4 SST/IST generation and post-processing are described in Section 3.d.

3.1. Processing of OI Sea ice concentration

The OI SIC field (used as input for the DMIOI L4 Processing System) is based on different sources of sea ice information, which have been resampled onto the final L4 0.05° regular latitude longitude grid. As stated in Section 2.a.2), the SMHI/CMEMS products (Høyer and Karagali, 2016) are used within the Baltic Sea, while the SICCI-25km and OSI-450 SIC fields are used outside of the Baltic Sea. The SICCI-25km product is used whenever it is available and the OSI-450 product otherwise. The days with missing data are listed in the product user guides, and for these days the SIC field closest in time is used to construct a SIC record for all days from 1982 to May 2021. The SIC field has been extrapolated along the coasts to cover the fjords with SIC values. Low SIC (≤15%) is defined as no ice and the SIC is set to 0% in these grid cells.

Even though a land-spillover correction procedure is already applied in the SIC processing for SICCI-25 and OSI-450, erroneous ice is still common along ice-free coasts since the PMW brightness temperatures from ocean grid cells close to the coast often contain a mixture of the microwave emission from land and ocean (due to the large field of view). This land-spillover effect is more pronounced for the coarser resolution OSI-450 product than the SICCI-25km product (Kern et al., 2022; Lavergne et al., 2019). The resampling of the SIC products to the L4 0.05° grid results in oversampling of the coarse resolution SIC data, which also requires some filtering of the resampled SIC data to comply with the higher resolution SST data. To improve the resampled SIC fields (in particular the coarser resolution OSI-450) and to increase the consistency of the full SIC record, two filters have been used to minimize the effects from erroneous ice along the coasts. Both filters use a spatial 15 × 15 grid point filter around each 0.05° L4 grid cell. The first filter (F1) removes sea ice from the center grid cell if the group of 15 × 15 grid cells contains at least one land and one ocean grid cell. This approach is similar to the NOAA/NSIDC Bootstrap Land-Spillover Correction (Cho et al., 1996; Meier, 2012) for which the center grid cell is replaced by the minimum non-land value within the group of grid cells if at least one of the grid cells is land. The second filter (F2) removes sea ice from the

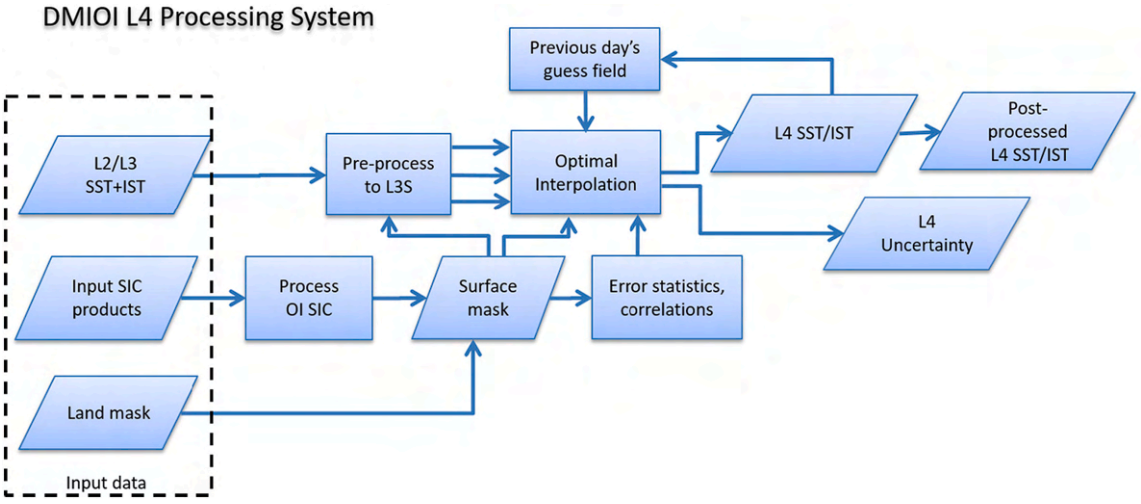


Fig. 2. Schematic diagram illustrating the processing steps of the DMIOI L4 Processing System.

center grid cell if any of the grid cells within the 15 × 15 group is land and the SST (from the OSTIA/CCI product) of the center grid cell at the same time exceeds 3 °C. The rationale for this is that the SST fields in general have fewer coastal issues as they are based on the higher resolution IR observations.

To examine the consistency between the different products, the resampled OSI-450 and SICCI-25km have been compared for 2018 with and without the filters. The Baltic Sea has been excluded in this comparison since the SMHI/CMEMS SIC products have been used here. In general, the largest differences between the resampled OSI-450 and resampled SICCI-25km are found along the coasts, where the resampled OSI-450 has numerous occurrences of sea ice grid cells, which are not seen in the resampled SICCI-25km. This is explained by the larger uncorrected land spill-over effects in OSI-450 compared to SICCI-25km.

Fig. 3 shows the distribution of sea ice as a function of SST (OSTIA/CCI) for OSI-450 and SICCI-25km during July 2018 in the case of no filter and both filters (F1 + F2) applied. July has been chosen as example month since it is one of the months with largest disagreement between OSI-450 and SICCI-25km as well as the month with most cases of coinciding sea ice and SST > 3 °C. As shown in Fig. 3a-b, OSI-450 has many cases of sea ice with SIC up to 60% in warm (6–15 °C) waters (before filtering) which are not seen in SICCI-25km. The OSI-450 and SICCI-25km agree much more after applying the two filters (F1 + F2), but with some remaining differences for warm SSTs (>10 °C).

A final filter (F3) has been applied to remove all sea ice in grid cells with OSTIA/CCI SST exceeding 8 °C. This SST threshold is more relaxed than what has previously been used (Hurrell et al., 2008; Markus and Cavalieri, 2009) and observed during in situ campaigns (Chiodi et al., 2021) to minimize the removal of true ice in regions with large SST gradients.

Before applying any filters, the resampled OSI-450 has more ice grid cells (SIC>15%) than the resampled SICCI-25km ranging from 0.88% more in March to 3.34% more in July. These ranges reduce to 0.84% in March and 2.84% in July after applying all filters (F1 + F2 + F3). The number of grid cells with SIC>15% and SST>3 °C is only a small fraction of all ice grid cells. Without any filters this fraction ranges from 0.018 to 0.639% and 0.002–0.162% for OSI-450 and SICCI-25km, respectively. After applying all the filters, these ranges are reduced to 0.002–0.197% and 0.000–0.140% for OSI-450 and SICCI-25km, respectively.

Fig. 4 provides detail of the number of ice grid cells with SST>3 °C for the entire SIC range (0<SIC≤100%) during July for the resampled OSI-450 and SICCI-25km in the cases where no filter, F1, or F1 + F2 + F3 has been applied. The filters remove limited sea ice in SICCI-25km compared to OSI-450. In OSI-450, F1 primarily removes sea ice with concentrations below 20% while adding F2 leads to removal of sea ice with SICs up to 80%. F3 removes very limited cases in comparison with F1 and F2. After filtering, the distribution of sea ice (with SST> 3°C) in OSI-450 looks more similar to the distribution in SICCI-25km.

The filtering method derived here removes many of the errors due to land spill-over, in particular in OSI-450, but some errors likely remain and likewise some true sea ice may be removed through the filtering. However, overall the filters have proven to be beneficial to reduce spurious ice and to reach higher consistency in the resulting combined and filtered SIC record (referred to as OI SIC). The preprocessed OI SIC

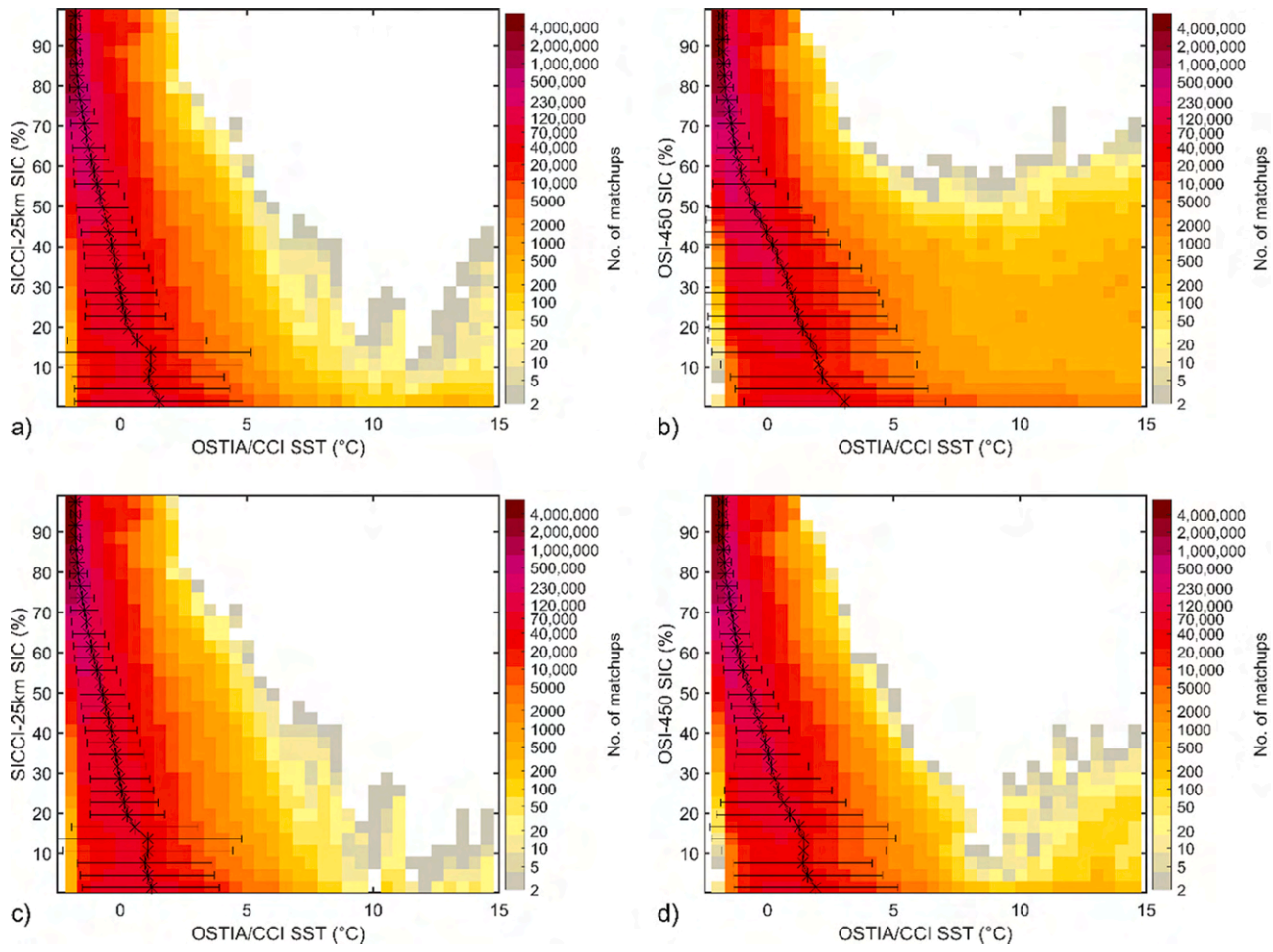


Fig. 3. OSTIA/CCI SST versus SIC from OSI-450 (right) and SICCI-25km (left) during July 2018 without filtering (top) and with both F1 and F2 filters applied (bottom). The average distribution is shown as asterisks with related error-bars showing the standard deviations considering bins with >30 members.

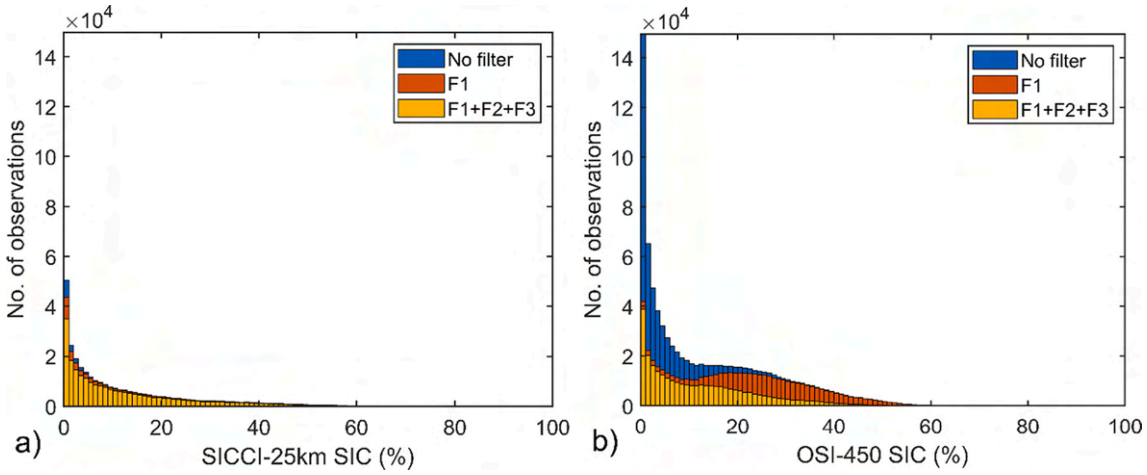


Fig. 4. The number of ice grid cells (SIC>0%) with SST>3 °C during July 2018 for a) SICCI-25km and b) OSI-450 in case where no filter, F1, or F1 + F2 + F3 filters have been applied, respectively.

field is used as input to the DMIOI L4 processing system for the L4 SST/IST generation (see Fig. 2).

3.2. Preprocessing of SST/IST satellite data

The input L2 SST and IST satellite data have been aggregated into L3S data by considering the available data within 24 h from the analysis. The L3S SST fields have been calculated as a noise weighted average of all available ATSR, AVHRR and SLSTR observations. Based on results from validation against in situ observations and the statistics from ingestion in the L4 analysis, we use the following uncertainties for ATSR, AVHRR, SLSTR and AASTI: 0.3 °C, 0.4 °C, 0.4 °C and 1.0 °C. AASTI ISTs are used over the MIZ and sea ice regions.

Only satellite data classified as cloud free by the cloud mask were included. Satellite observations processed using an IST algorithm in areas where the OI SIC field is zero were discarded. A minimum quality flag of 4 was used for all observations except from those from SLSTR, where a quality flag of 5 was required.

3.3. The optimal interpolation (OI) method

The aim of the OI method is to take an irregular distribution of observations, which may have different uncertainties and spatial resolutions, and provide the best possible estimate on a regular grid. This should be done by extracting the maximum information from all available observations and combining these using the proper weight of each observation. OI attempts to accomplish this by minimizing the mean-square interpolation error for a large ensemble of analysis situations. The OI method was first introduced by Gandin (1963) to produce gridded maps of meteorological variables, while oceanic applications was introduced by Bretherton et al. (1976) and has since then been widely used for mapping of SST (e.g. Høyer and She, 2007; Reynolds and Smith, 1994; White, 1995). The OI methodology used in this paper follows the high latitude SST DMI processing scheme (Høyer et al., 2014; Høyer and She, 2007). For the OI estimation, we assume that $\widehat{f}_i$ represents the ith observation of the field and can be written as

$$\widehat{f}_i = f_i + \varepsilon_i \tag{1}$$

where f_i is the true representative value of the field and ε_i is the observational error.

The construction of the OI field employs a first guess field, which in this case is provided as the analysis field from the previous day. The deviation of the estimated field from the first-guess field is calculated as a weighted sum of the nearby observed departures (anomalies) from the first guess:

$$f_0' = \sum_{i=1}^{n} (f_i' + \varepsilon_i) p_i + I_0, \tag{2}$$

where p_i are the weighting factors, I_0 is the interpolation error, n is the number of observations selected for the interpolation and the primes indicate anomalies from the first-guess field. Working with anomalies ensures that the first guess is preserved in regions with limited or no observations. It is assumed that the first guess is unbiased and that the noise on one observation is not correlated with the true value or with the noise of other observations (Bergman, 1978). The optimal weights should be chosen such that ε is minimized and this is obtained by differentiating ε partially with respect to each of the p_j and equate them to zero. By assuming that the variances are homogenous and the covariances are both homogenous and isotropic (i.e. only depending on the distance) the following set of linear equations can be derived:

$$\sum_{j=1}^{n} C_{i,j} p_j + \tau_i^2 p_i = C_{0,i}, \tag{3}$$

where p_i denotes the optimal weights, which need to be determined, $C_{i,j} = \frac{\overline{f_i' f_j'}}{\sigma^2}$ is the correlation function between the individual observations and $C_{0,i} = \frac{\overline{f_0' f_i'}}{\sigma^2}$ is the correlation function between the observations and the estimation point, and $\tau_i^2 = \frac{\sigma_{\varepsilon,i}^2}{\sigma^2}$ with $\sigma_{\varepsilon i}^2$ being the error variance of the ith observation and σ^2 representing the variability. The OI method directly provides the mean-square interpolation error, which can be estimated in all grid points by

$$\varepsilon_{min} = \sigma^2 \left(1 - \sum_{i=1}^{n} C_{0,i} p_i \right). \tag{4}$$

From this, it follows that the mean-square interpolation error never exceeds the variance of the anomalies in the estimation point. Both the weights and the interpolation error depend on the scales of the correlations $C_{i,j}$ and $C_{0,i}$, on the variance, and on the random observational error, $\sigma_{\varepsilon,\, i}$. These are calculated prior to the prediction and assumed to be constant in time.

3.3.1. OI statistical parameters

The analysis field from the previous day is used as first guess and the corresponding first guess error variances are derived from a test run, where the L3 observations were compared against the previous day analysis and the standard deviations were calculated from the

anomalies. For IST, the first guess variance and error covariance have been derived using one year of Metop AVHRR L3 IST observations. For each day, the L4 field of previous day were subtracted from the L3 IST observations. The anomaly time series were then used to calculate the standard deviations for every grid point in order to obtain a spatial two dimensional field of standard deviations (Høyer et al., 2014). The first guess variance and error covariance for SST are the same as those used in Høyer et al. (2014), and Høyer and She (2007). In the MIZ (15% < SIC≤70%), the first guess variance and error covariance are calculated through a weighted linear combination of SST and IST values, using the OI SIC as a weighting factor.

Spatially varying correlation functions in the latitudinal and longitudinal directions have been derived empirically from the observations (i.e. one year (2018) of anomalies) by assuming steady state. For IST, L3 observations from METOP AVHRR were used, while L3 VIIRS observations (e.g. Liang and Ignatov, 2013) were used for SST due to the higher accuracy and higher spatial resolution (~750 m). The correlations are calculated by correlating time-series of anomalies within 10 × 10° bins. Separate functions were derived for open ocean, sea ice and the MIZ, due to the different variabilities in the temperature of the different surface types, but they share the same form:

$$C_{i,j} = exp\left(-\lambda \bullet dist_{i,j}^{\gamma}\right) \quad (7)$$

where $dist_{i,j}^{\gamma}$ is the distance between two observations, and λ and γ are the two empirically determined parameters that vary throughout the domain.

The true correlations will be larger than the observed correlation near zero lag because of the random noise in the observations. The difference in correlation resulting from the random noise (μ) has been estimated by measuring the drop in correlation from the estimation point (correlation = 1) to a small distance (2.5 km) for both IST and SST. Thereby, it is assumed that at distances smaller than 2.5 km from the estimation point, the drop in correlation can be fully attributed to the random noise. For the MIZ, the random noise has been estimated by assuming that the error of MIZT observations is equal to the error on IST observations. The estimated effect from the random noise has been removed before fitting the correlation functions. The fit results in 2D spatial fields of λ and γ for each of the three surface types. Fig. 5 shows the spatial 2D mean observed correlations including the random noise component for SST and IST using all available observations. They are calculated by correlating time-series of anomalies within 10 × 10° bins and ordering the results according to the distance in latitude and longitude.

Fig. 6 shows the best fit mean correlation models for SST, IST and MIZT in the latitudinal direction with the corresponding average values of λ and γ as listed in Table 1. The fitted mean correlations only include observations with >30 pairs of anomalies. The derived *e*-folding scales are shown in Table 1.

3.4. L4 SST/IST generation and post-processing

The derived OI statistical parameters have been used for the generation of the gap-free L4 SST/IST climate dataset. The correlations of all observations within the search radius (100 km) around the estimation point have been estimated and the observations used by the OI have been selected by taking the one with largest correlation within each quarter around the estimation point (if available). This has been repeated until the maximum number of observations (20) is reached or no more observations are available. On average, the selected number of observations used in the OI is 18.6, indicating good coverage. The coverage increases over time (with annual means ranging from 16.3 to 19.8) and reaches a maximum during winter (DJF) with a mean coverage of 19.1 and a minimum during summer (JJA) with a mean coverage of 17.6.

The interpolated anomalies have been added to the first guess field to provide the surface temperatures. Any anomalies exceeding ±9.9 °C

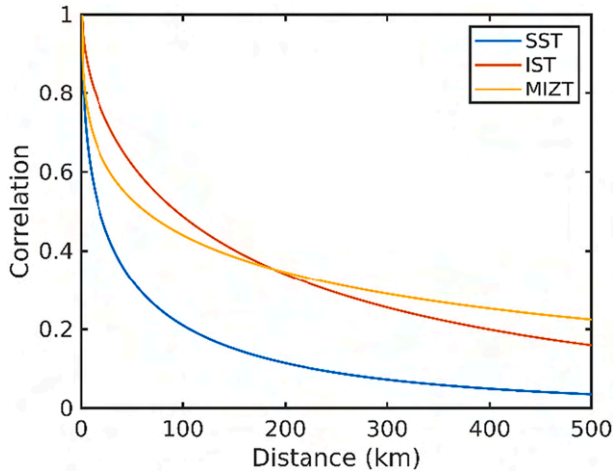


Fig. 6. The best fit mean spatial correlation in the latitudinal direction for SST, IST and MIZT.

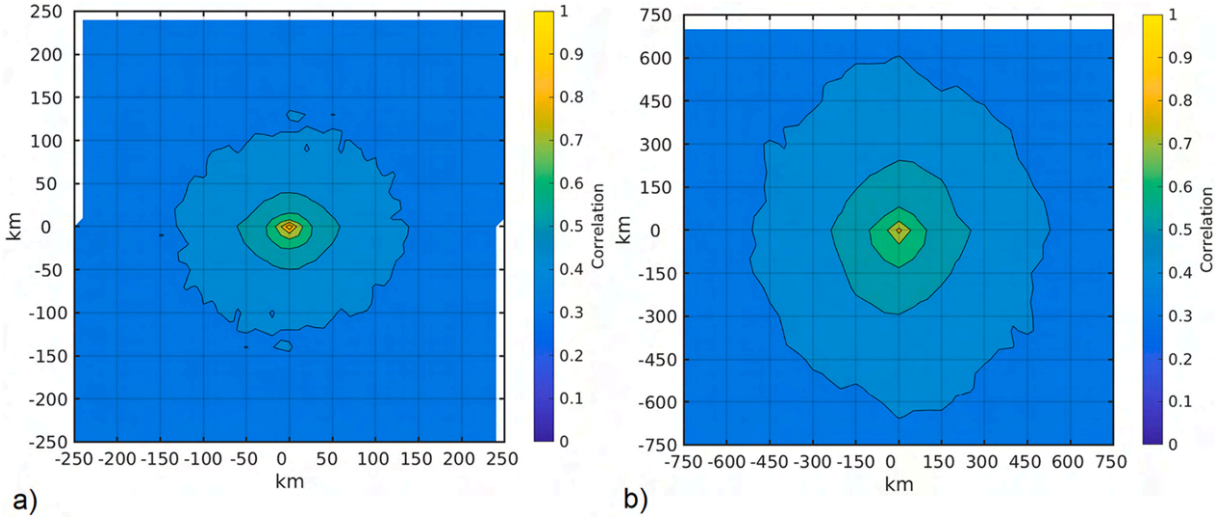


Fig. 5. Spatial mean of observed correlations (all 10 × 10° bins) for SST (left) and IST (right) with the estimated random noise added to the observations.

Table 1
The average covariance parameters, λ, γ, first guess error variance, σ^2, (°C), random noise, μ, and e-folding scale, e^{-1} (km) of each domain.

	SST (SIC≤15%)	IST (SIC>70%)	MIZT (15% < SIC≤70%)
λ	0.18	0.05	0.16
γ	0.47	0.58	0.37
σ^2	0.18	5.15	3.17
μ	0.24	0.29	0.47
e^{-1}	69.5	276.1	344.4

were reset to this value. Surface temperatures outside the accepted range (−60 to +35C°) have been reset to the closest of the two temperature limits.

The input satellite observations are all thermal IR observations which can only be utilized during clear-sky conditions. Here, the gaps due to clouds are filled using OI, but the aggregated and averaged daily surface temperatures are clear-sky averages, which may differ from the all-sky averaged surface temperature. For SST, the difference is small but over ice the clear-sky averaged IST is typically colder than the all-sky averaged IST (referred to as the clear-sky bias). Nielsen-Englyst et al. (2019) found an average cold clear-sky bias of 0.85 °C, using in situ observations over Arctic sea ice and a cloud mask derived from the

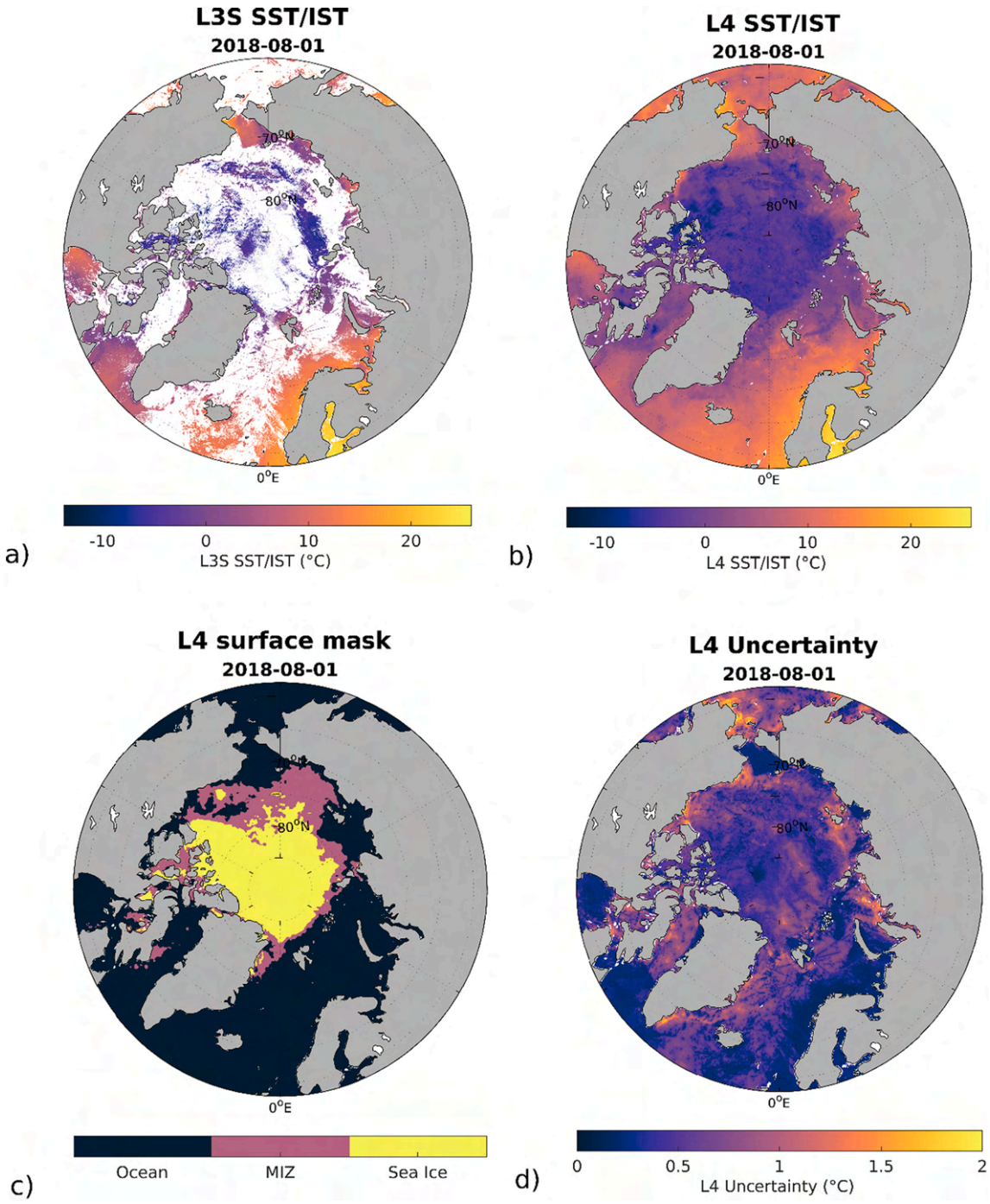


Fig. 7. An example of a) the combined L3S SST/IST field, b) the optimally interpolated L4 SST/IST field, c) the L4 surface mask, and d) the derived L4 uncertainty field for August 1, 2018.

longwave-equivalent cloud cover fraction. To minimize the clear-sky bias over ice in the L4 SST/IST product, we have introduced a constant bias correction over sea ice (SIC>70%) of +0.85 °C, which are linearly scaled towards 0 °C for open ocean using the SIC as a weighting factor.

No clear-sky bias correction has been applied to the SST in open water grid points (SIC≤15%), but a constant bias correction of 0.16 °C has been applied based on the validation against drifting buoys. It has been added to all open water grid points and the value is linearly scaled towards zero for 100% sea ice cover using the SIC as a weighting factor.

4. Results

The DMIOI L4 processing system has been used to generate a daily gap-free combined SST and IST reanalysis covering the Arctic (>58°N) with a spatial resolution of 0.05° for the period 1st of January 1982 to 31st of May 2021. An example of the combined SST and IST L3S field, the L4 SST/IST field, the L4 surface mask, and the L4 uncertainty field is shown in Fig. 7 for August 1, 2018.

4.1. Validation

The L4 SST and IST fields have been validated separately using the independent in situ observations listed in Section 2.b. The OI SIC field is only used as a mask for ocean/ice identification, and it will not be validated. Table 2 shows the overall validation statistics against independent in situ observations for SST and IST. The statistics are divided into the different sources of in situ observations due to differences in type and error characteristics. For both SST and IST, matchups with L4 – in situ temperature differences deviating more than three times the standard deviation (i.e. 3σ) from the mean L4 – in situ temperature difference have been excluded. This has been done for the individual observation types to avoid effects from outliers. The results of the SST and IST validation are discussed separately in the two following sections.

4.1.1. Sea surface temperature

The L4 SST (SIC≤15%) have been validated against drifting buoys, moored buoys, and Argo floats. Matchups with in situ SST or L4 SST colder than −1.8 °C have been excluded from the validation statistics. Table 2 shows that the L4 SST have mean differences <0.05 °C and standard deviations <0.6 °C, and correlations above 0.98 when compared against drifting buoys, moored buoys, and Argo floats. The Argo floats and moored buoys provide independent validation statistics while the drifting buoys have been used to adjust the L4 SST (see Section 3.d). The mean differences are lower than those reported by Castro et al. (2016) where nine different SST analyses and two single sensor satellite products were compared with independent observations from Upper Temperature of the polar Oceans (UpTempO) buoys deployed in the Beaufort Sea in 2012 and 2013. The standard deviations reported here are slightly lower than in Castro et al. (2016) for most cases, but higher than those reported for the most northern UpTempO buoys where SSTs were very uniform and lower than for those observations that coincided with a summer storm over the Arctic Ocean.

Fig. 8 shows the seasonal (3-month) mean L4 SST differences (L4 SST – in situ SST) and standard deviations for the period 1982–2021, when compared to drifting buoys, moored buoys, and Argo Floats. It should be noted that the amount of drifting and moored buoys are much lower in the beginning of the time period compared to the latest years, which makes the validation statistics against drifter and moorings less reliable in the beginning of the record. The Argo floats are only available from May 2001. The statistics are only calculated for 3-month periods with >45 matchups available. Overall, the mean difference and standard deviation are both quite stable over the period of time for all three types of in situ observations. The combined trend in the mean difference is plotted (Fig. 8) and the corresponding overall trend is −0.0002 °C/yr indicating a very stable performance. The trends for the individual in situ types are −0.0001 °C/yr, −0.0019 °C/yr and 0.0027 °C/yr for drifters, moorings and Argo floats, respectively.

Table 2
Overall validation statistics of the L4 SST and IST against in situ observations. The mean difference is given by L4 SST/IST minus in situ SST/IST. Notice that the NP, ECMWF and CRREL do not measure the surface temperature, but the air temperature (T2m). For both the SST and IST validation a 3σ filter has been applied before calculating the statistics. The table shows the mean difference (MD), standard deviation of the difference (std), root mean squared error (RMSE), correlation (corr) and the number of observations (Nobs).

Type	Parameter	MD	std	RMSE	corr	Nobs
Drifting buoys	SST (°C)	0.00	0.54	0.54	0.99	3,062,549
Moored buoys	SST (°C)	0.03	0.56	0.56	0.98	76,052
Argo	SST (°C)	0.03	0.51	0.51	0.99	32,953
NP drifting ice stations (T2m)	IST (°C)	−2.35	3.12	3.91	0.98	7665
Drifting buoys ECMWF (T2m)	IST (°C)	−3.21	3.34	4.63	0.96	55,288
Drifting buoys CRREL (T2m)	IST (°C)	−2.87	3.36	4.42	0.96	22,979
IceBridge KT-19 (IST)	IST (°C)	1.52	3.12	3.48	0.92	36,638

4.1.2. Ice surface temperature

The validation of the L4 IST (SIC>15%) is limited by the sparse number of in situ observations as well as increased in situ uncertainties in the ice covered regions compared to the open ocean. The L4 IST is colder than in situ measurements from NP, ECMWF and CRREL, while the L4 IST is warmer than KT-19 measurements from IceBridge flights. NP, ECMWF and CRREL do not provide the surface temperature, but the air temperatures measured at a varying height, typically of about 2 m (T2m). Nielsen-Englyst et al. (2019) found an average IST-T2m difference of −1.25 °C during all-sky conditions over sea ice. This IST-T2m difference is a real temperature difference between the snow surface and the air above it. This explains part of the temperature differences observed in the validation results against NP, ECMWF and CRREL (presented in Table 2), which is not related to the performance of the L4 IST. In Section 5, this topic is further discussed. Fig. 9 shows the seasonal mean L4 IST difference and standard deviation against NP, ECMWF and CRREL observations, respectively. The statistics are only calculated for the 3-month periods with >45 matchups available. The combined trend in the mean difference is −0.0166 °C/yr, while the individual trends are 0.0047 °C/yr, −0.0391 °C/yr, and 0.0676 °C/yr for NP, ECMWF buoys and CRREL buoys, respectively. A seasonal cycle is present both in the mean difference and the standard deviation. The larger standard deviations during winter are explained mainly by the larger temperature variability during winter, but also by the higher uncertainties in the cloud masking during winter (Nielsen-Englyst et al., 2019). The seasonal variation in the mean difference is corresponding to what was observed in the L3 IST validation and in Nielsen-Englyst et al. (2021).

The L4 IST validation against IceBridge includes 117 IceBridge flights, which provide measured surface temperatures in the period 2012–2019. The IceBridge data are smoothed for every fifth kilometer if >30 observation points were available. Fig. 10 shows an example of the validation against one single IceBridge flight from March 2012. Despite the smoothing of the data, there are still large fluctuations in the IST measured by IceBridge, which are not captured by the L4 IST. These are results of warm leads (open or newly refrozen) in the sea ice and some of the largest ones are actually captured to some extent by the L4 IST. Overall, the figure shows a quite good agreement between IceBridge and the L4 IST in this particular example, with a mean difference of 1.24 °C and a standard deviation of 1.11 °C. Considering all 117 IceBridge flights, and weighting all observations equally, the IST validation shows a mean difference of 1.52 °C, standard deviation of 3.12 °C and correlation of 0.92 (see Table 2) using the 3σ filter on the differences. The

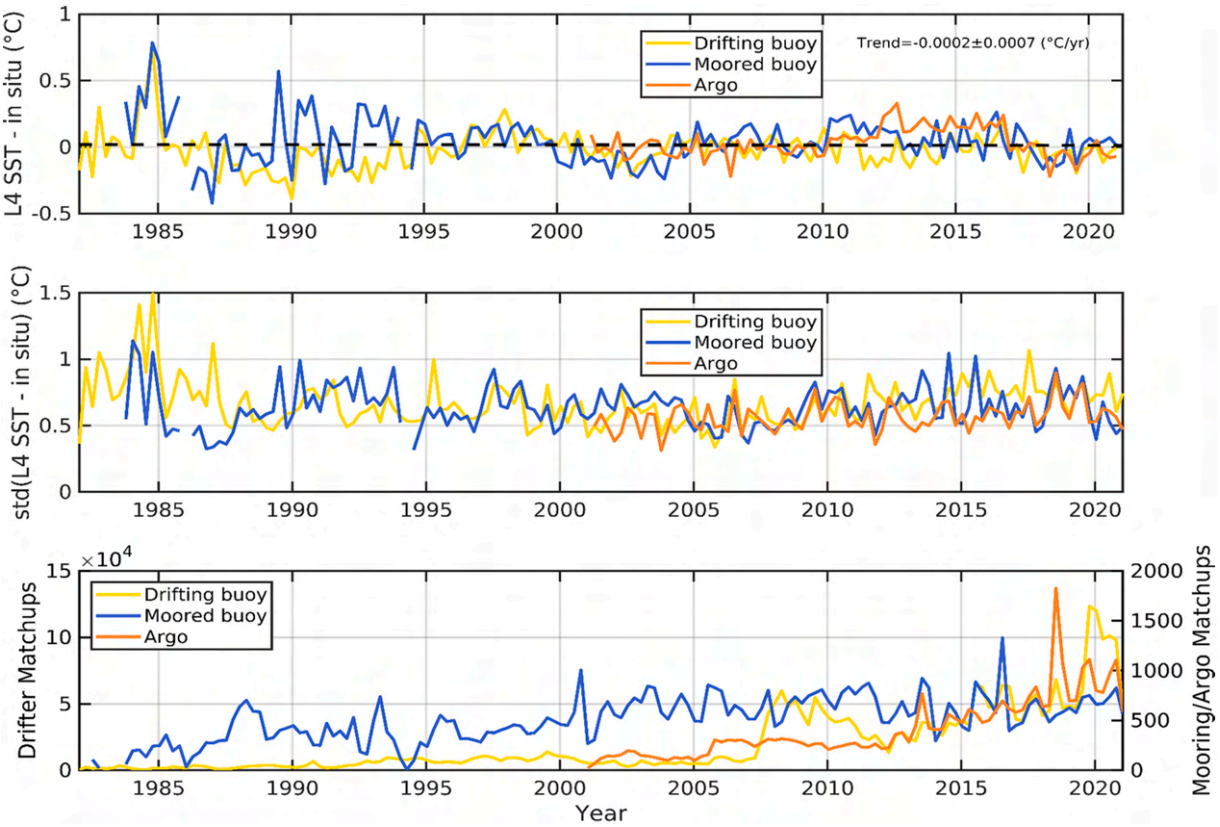


Fig. 8. Seasonal (3-month) mean L4 SST difference (top) and standard deviation (middle) against sea surface temperatures measured by drifting buoys, moored buoys and Argo floats. The bottom plot shows the seasonal mean number of matchups for each in situ source. The linear trend in the mean difference is calculated considering all in situ types. The seasonal difference, standard deviation and trend are only calculated if >45 matchups are available.

IceBridge observations have standard deviations of the 5 km averages ranging from 0.81 °C to 5.41 °C for the different flights and an average value of 2.52 °C. Part of the bias against IceBridge may be explained by the missing cloud screening of IceBridge observations. If clouds are present between the flight and the surface, the radiometer will provide the temperature of the (usually colder) clouds instead of the surface temperature. Furthermore, IceBridge provides "snap-shots" of the surface temperature, while the L4 SST/IST provides daily means, which have been adjusted by using the 24 h average clear-sky bias correction (see Section 3.d).

4.1.3. Uncertainties

Each daily L4 SST/IST is accompanied with an uncertainty estimate, which is a direct output of the OI method (see Eq. 4). The uncertainties for the L4 SST have been validated against drifters. The validation results are shown in Fig. 11, where the actual L4 SST minus drifter SST differences are plotted as a function of the OI derived uncertainty estimates. This uncertainty validation approach is similar to what is done in Alerskans et al. (2020) and Nielsen-Englyst et al. (2018). The dashed lines represent the ideal uncertainty based on the assumption that drifters have a total uncertainty of 0.2 K (Kennedy, 2014). The figure shows a good agreement between the estimated uncertainties and the L4 SST uncertainties for low uncertainties, but overestimated uncertainties for the higher uncertainties. This approach is unfortunately not applicable for the IST validation results due to the lack of good quality in situ reference observations that makes the spatial sampling component very large in comparisons between the L4 IST and the in situ measured near surface air temperatures.

4.2. Climate analysis

The derived L4 SST/IST product is the first gap-free combined SST and IST reanalysis product for the Arctic. The combination of sea ice and ocean surface temperatures in the Arctic provides a consistent climate indicator, which can be used to monitor day-to-day variations as well as long-term climate trends. Fig. 12 shows the average seasonal variation of the daily surface temperature and the sea ice to open water ratio of the Arctic (>58°N). The daily surface temperature reaches a maximum in August with average surface temperature around the melting point and minimum temperatures in February/March with average temperatures around −24 °C. During the minimum in September, only 30% of the L4 SST/IST region is covered with sea ice (and 70% open ocean), while the fraction increases to almost 70% during winter (30% for open ocean). The largest variability in temperature is seen during winter. Part of this is explained by stronger cyclones and anticyclones in contrast to the summer (Serreze et al., 1993), when the variability moreover is limited by the upper temperature limit around the melting point for the ice covered regions. For regions covered by seasonal sea ice, the SSTs usually also stay around the melting point during summer, when the sea ice has retreated(e.g. Timmermans and Labe, 2020), limiting the summer SST variability in these regions as well. The large variability in winter temperatures as well as in the sea ice coverage, shown in Fig. 12, is also explained by the increasing temperatures and decreasing sea ice extent throughout the data record (1982–2021).

Fig. 13 shows the monthly mean surface temperature anomaly for the period 1982–2021 when compared against the monthly mean of the 30-year-long reference period, 1991–2020. The large fluctuations illustrate large monthly variations in sea and ice surface temperatures of the Arctic. The solid black curve represents the yearly mean and the dashed black curve is the linear fit. During the period from 1982 to 2021

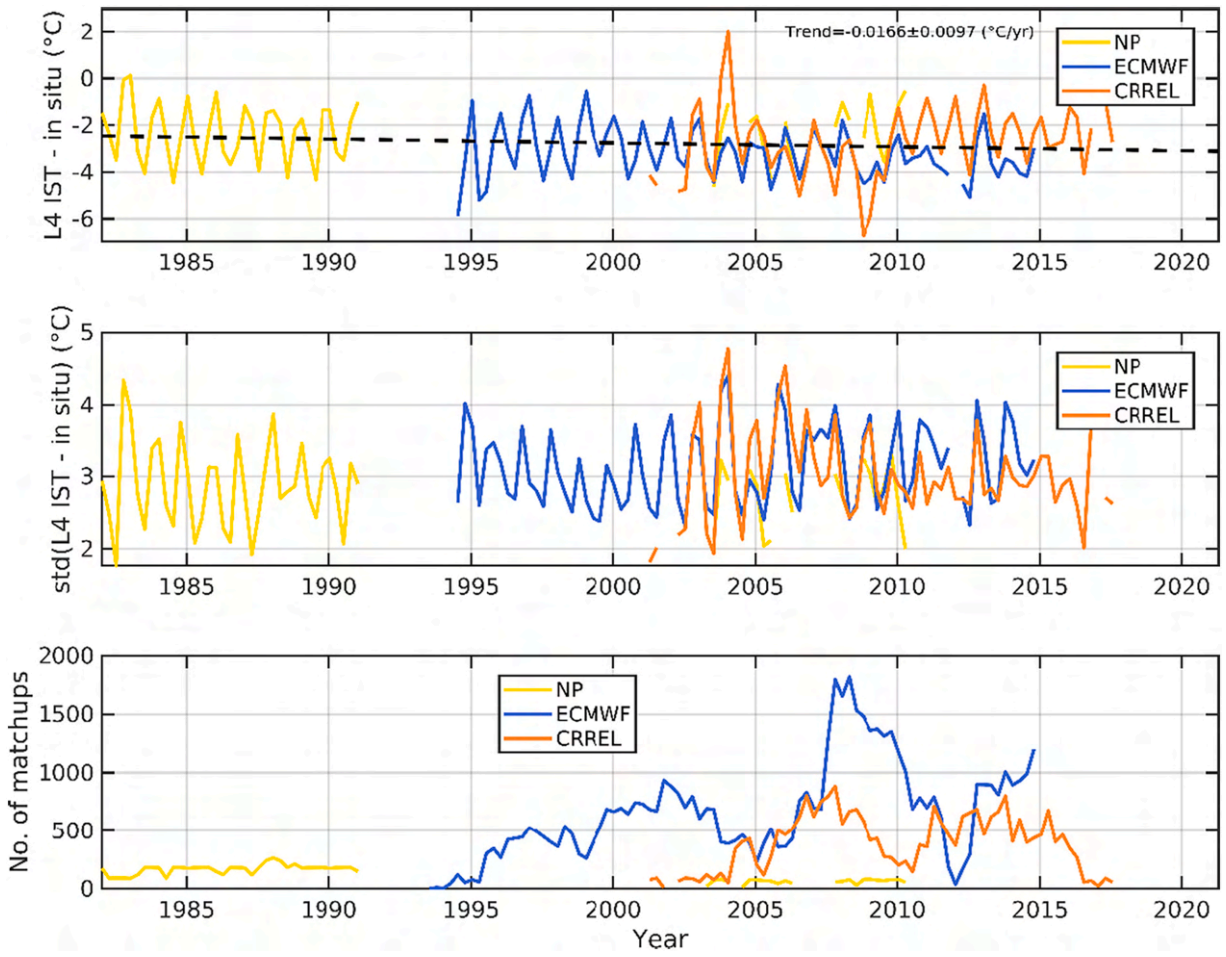


Fig. 9. Seasonal (3-monthly) mean L4 IST difference (top) and standard deviations (middle) against air temperatures from NP, ECMWF, CRREL. The bottom panel shows the seasonal mean number of matchups for each in situ type. The linear trend in the mean difference is calculated considering all in situ types. The seasonal mean difference, standard deviation and trend are only calculated if >45 matchups are available.

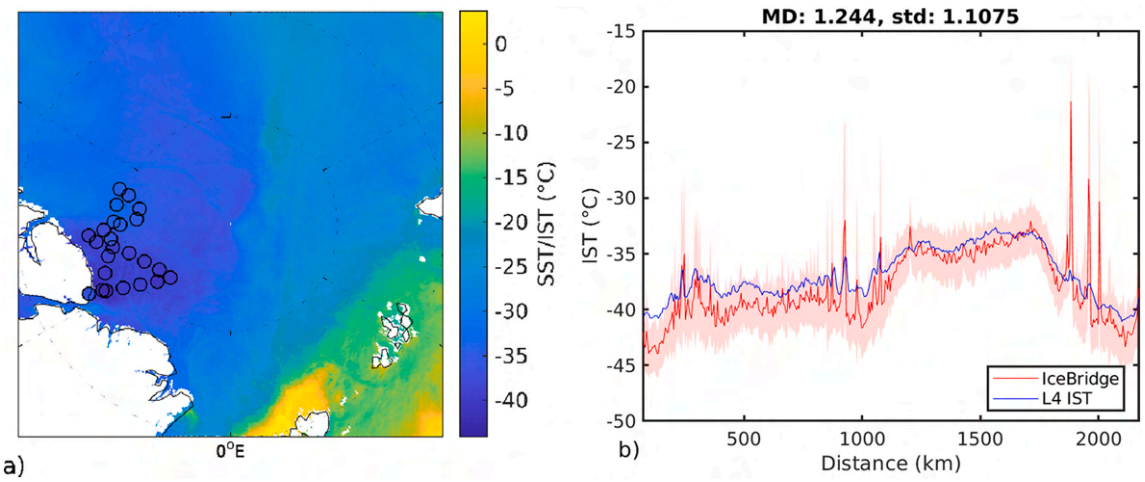


Fig. 10. An example of the validation against one Icebridge flight during March 29, 2012. Figure a) shows the flight path, while b) shows a comparison of the smoothened IceBridge IST and the L4 IST, with a mean temperature difference of 1.24 °C and a standard deviation of 1.11 °C.

the linear trend is 0.114 °C/yr, and the surface temperatures of the Arctic Ocean (>58°N) have thus increased by about 4.5 °C in 39 years. The trend increases with latitude and the total temperature increase to about 5 °C north of 70°N. Fig. 14 shows the corresponding two dimensional trend in the monthly Arctic surface temperatures for the entire period, 1982–2021. On average, the temperature has increased in most regions, with the largest increase in the northeastern Barents Sea, which shows a peak temperature increase of about 10 °C over the 1982–2021 period. The trend pattern is in agreement with the most recent findings based on combined observed and modelled data over the Arctic Ocean

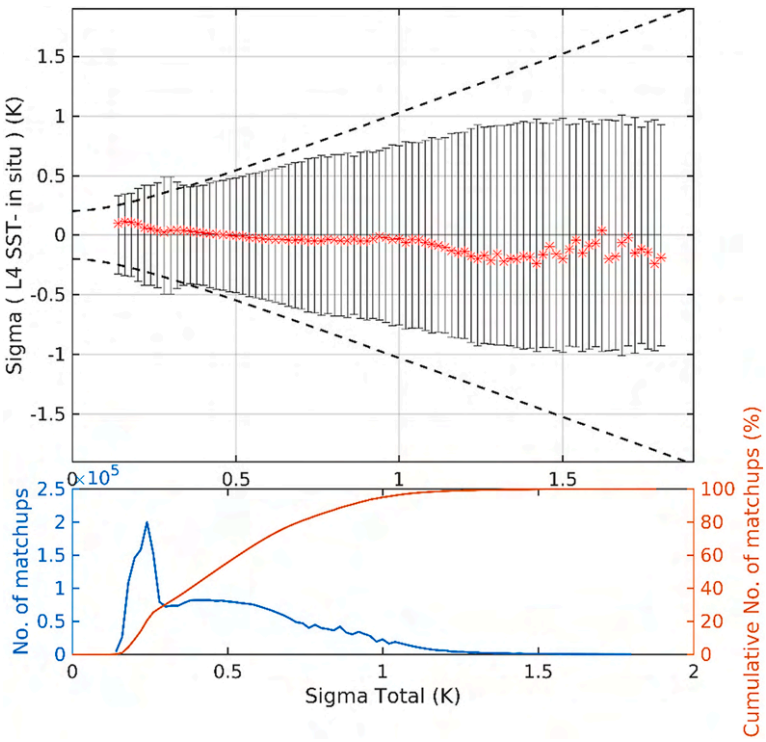


Fig. 11. L4 SST uncertainty validation against independent in situ observations from drifters. The dashed lines show the ideal uncertainty when accounting for uncertainties in the drifter SSTs and the sampling error. The solid black lines show one standard deviation of the L4 SST minus drifter SST differences for each 0.02 K bin and the red asterisks mark the mean difference. The bottom plots show the number of matchups (blue) and the cumulative percentage of matchups for each bin (red). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

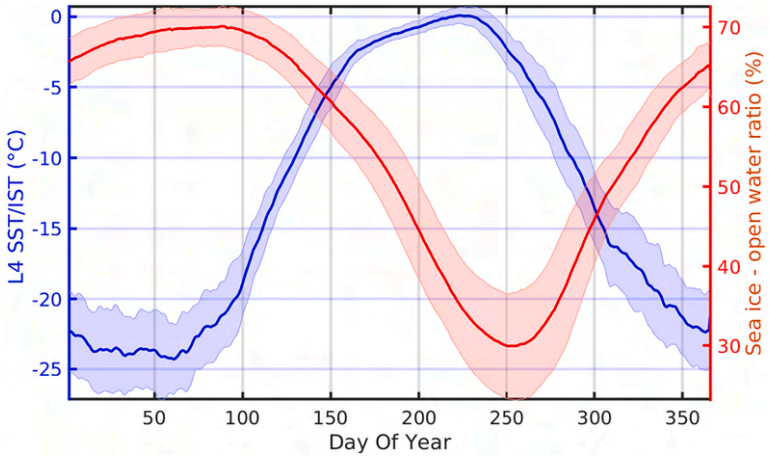


Fig. 12. Seasonal cycle of daily mean L4 SST/IST (blue) and daily mean sea ice – open water ratio in percentage of the total L4 SST/IST coverage (red) considering all years, 1982–2021. The shaded regions show one standard deviation. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

for the period 1979–2019 (AMAP, 2021) and a combination of four observational datasets covering the period 1979–2021 (Rantanen et al., 2022). The largest temperature increases are coinciding with those regions where the number of open water days have increased the most (Tonboe et al., 2016). In general, the surface temperature increases observed in Fig. 14 are slightly larger than the air temperature increases observed in AMAP (2021) and (Rantanen et al., 2022), who find maximum temperature increases of around 6–8 °C in the northeastern Barents Sea. The largest surface temperature trends are found during fall (September–November) and winter (December–February) with average trends of 0.148 °C/yr and 0.142 °C/yr, respectively, leading to an average temperature increase of >5.5 °C from 1982 to 2021 considering these months. During summer (June–August), limited increase in the surface temperature (0.037 °C/yr) is observed. The small trend during summer is mainly explained by the upper constrain of the sea ice surface temperature to the melting point and the fact that the SSTs usually stay around the melting point during summer in regions where the seasonal sea ice has retreated (e.g. Timmermans and Labe, 2020).

5. Discussion

In order for a dataset to be categorized as a CDR it needs to have sufficient length, accuracy and stability to enable study and assessment of long-term climate variability and change (Minnett et al., 2020). A dataset with excellent absolute accuracy is important for understanding climate processes, but not necessary for determining long-term changes or trends as long as the dataset has the required stability (Ohring et al., 2005). Based on the expected magnitude of a climate change signal (per decade), the requirements of an SST CDR are an accuracy of 0.1 °C and a stability (per decade) of 0.04 °C according to Ohring et al. (2005). The

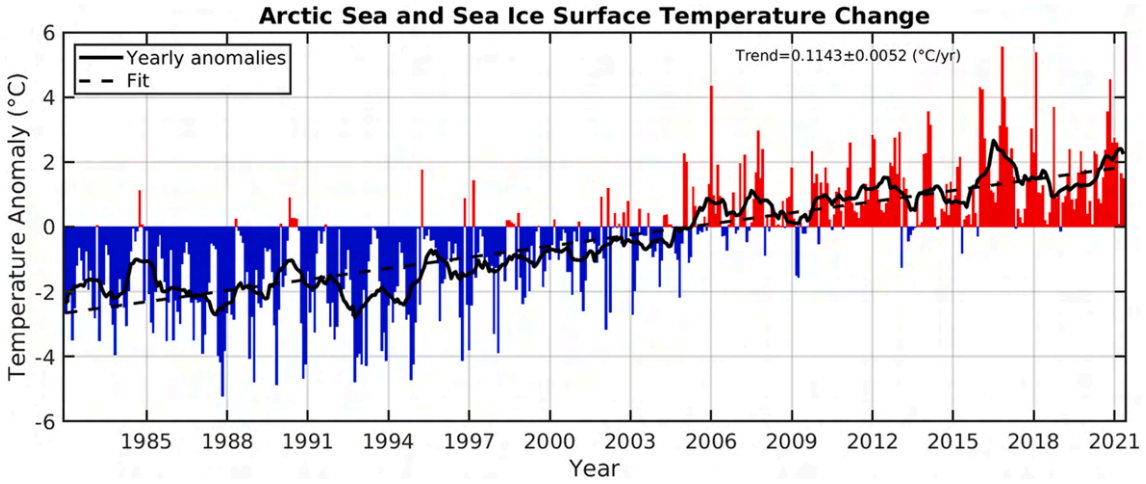


Fig. 13. Monthly mean surface temperature anomalies for the period 1982–2021. The anomalies represent the difference between the monthly mean surface temperature and the monthly mean surface temperature of the reference period, 1991–2020. The yearly mean anomalies are shown as the solid black curve, while the dashed black curve is the linear fit, which has a slope of 0.114 °C/yr.

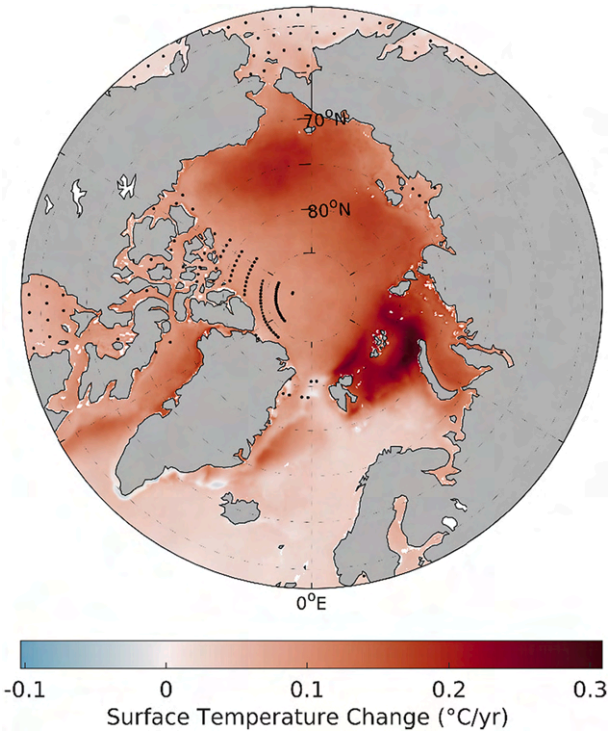


Fig. 14. The average rate of surface temperature change (°C) per year based on the monthly mean L4 SST/IST from January 1, 1982 to May 31, 2021. The black dots indicate regions where trends are not significant.

observed accuracy and stability of the L4 SST meet these SST CDR requirements. For land surface temperature (LST), including land surfaces covered with snow and ice, the CDR requirements are 0.3 °C/decade for the threshold stability and 0.1 °C/decade for the target stability (GCOS, 2016). The L4 IST meets the target requirements when considering the NP validation (which spans the longest time period). Currently, no GCOS requirements have been established for sea ice surface temperature, because this variable is not recognized as an Essential Climate Variable (ECV). Introducing IST as an official GCOS ECV is highly recommended in order to facilitate a common community consensus on IST requirements for climate applications (Lavergne et al., 2022). Moreover, the limited availability of long-term high quality reference in situ IST observations (with known and controlled stability), complicates validation and hence, the extent to which satellite IST products can be tested for meeting the CDR requirements.

In CDRs, quantitative uncertainty information should also be provided with the observations (Merchant et al., 2017). Here, each daily L4 SST/IST is accompanied with an uncertainty estimate derived directly from the OI method. The validation of the L4 SST uncertainty estimates (see Section 4.a.3) shows that the OI method is capable of deriving reliable uncertainty estimates for SST.

The validation of the L4 SST shows differences smaller than 0.05 °C and standard deviations <0.6 °C against drifters, moorings and Argo floats. The L4 IST is colder than in situ measurements from NP, ECMWF and CRREL, and warmer than KT-19 measurements from IceBridge flights with an average difference of 1.52 °C. The L4 IST differences were in the range from −3.21 °C to −2.35 °C and standard deviations of about 3.4 °C against T2m measurements from NP, ECMWF and CRREL. A large fraction of these differences can be attributed to the temperature difference between the in situ air temperatures measured about 2 m height and the actual surface temperature provided by the L4 IST product. Nielsen-Englyst et al. (2021) derived a simple regression model to convert satellite observed ISTs to 2 m air temperatures (T2m) over ice surfaces. Using the sea ice regression model and coefficients from Nielsen-Englyst et al. (2021) to derive T2m from the L4 ISTs, the absolute differences against CRREL, ECMWF and NP reduce to below 0.7 °C, while the standard deviations remain more or less the same. These initial results indicate that it is possible to derive reliable T2m above the sea ice on basis of satellite derived L4 ISTs in order to supplement the sparse in situ air temperature network in the Arctic.

Traditionally, gridded and gap-free satellite SST products provide the temperature of the sea water just below the ice in ice covered regions and this is also the case for the OSTIA/CCI and the NOAA Optimum Interpolation SST (OI SST v2; Reynolds et al., 2007, 2002), which are used in the comparison here. The NOAA OI SST v2 shows an overall warming trend in the Arctic Ocean (>65°N) of 0.036 ± 0.03 °C/yr considering the period from 1982 to 2018 (Carvalho and Wang, 2020), while a trend of 0.015 ± 0.003 °C/yr is found using OSTIA/CCI for the same region and time period. Considering the L4 SST/IST region (>58°N) and period (1982–2021), OSTIA shows an overall warming trend of 0.016 ± 0.001 °C/yr. To enable a comparison against the L4 product derived here, the OSTIA SST has been compared with the L4 SST product for regions that never experience sea ice (considering both products). The two products show very similar spatial trend fields, and the average trend is 0.019 °C/yr for OSTIA and 0.022 °C/yr for the L4 SST product, indicating a good agreement in those regions where both

products always report the temperature of the ocean surface.

In ice covered regions, limited variation is seen in the sub-ice SSTs, and this is not representative for the actual surface conditions. Instead, T2m is normally used in global reanalyses in ice covered regions in combination with SST over the ocean (Dee et al., 2011; Hersbach et al., 2020; Kobayashi et al., 2015; Simmons, 2004). A combination of modelled and observed data for a 49-year long period (1971–2019) showed an average increase in annual near-surface air temperature of 0.063 °C/yr (AMAP, 2021). Marquardt Collow et al. (2020) compared 12 reanalyses of the central Arctic Ocean and found an average trend in SST/T2m of 0.09 ± 0.01 °C/yr for the period 1979–2017 with large spread among the reanalyses (estimated standard deviation of the regression coefficient of 0.2 K). Using the derived T2m estimates from the L4 IST (following Nielsen-Englyst et al., 2021) together with our L4 SST, the combined SST/T2m trend in the Arctic is 0.100 ± 0.005 °C/yr for the region above 58°N and 0.111 ± 0.006 °C/yr above 70°N. This is in good agreement with the average estimate of the 12 reanalyses in the central Arctic Ocean.

Currently, global reanalyses only assimilate in situ measured air temperatures over ice. As a consequence, the global reanalyses usually show large uncertainties and significant differences in the Arctic due to the limited number of in situ observations in this region (Cowtan and Way, 2014; Lenssen et al., 2019; Marquardt Collow et al., 2020; Rapaić et al., 2015). Marquardt Collow et al. (2020) concludes that more in situ and remote sensing observations as well as a better use of existing satellite observations are needed in order to represent the characteristics of the entire Arctic region in both reanalyses and numerical models. Therefore, the L4 SST/IST product as well as a satellite derived T2m dataset based on the L4 IST product have a large potential to supplement existing in situ measurements and improve the current surface air temperature estimates over the Arctic sea ice (Nielsen-Englyst et al., 2021). The L4 SST/IST product has been derived for the Arctic, and the same procedures can be applied to the Antarctic.

Future work should aim at improving the L4 SST/IST product. One possible way of improving the L4 IST could be to include other available IST satellite products such as the MODIS IST data (Hall et al., 2004) or the VIIRS IST product (Key et al., 2013; Liu et al., 2015). More and higher quality in situ observations will also increase the possibilities of improving the satellite derived IST products and also allowing a better parameterization of the constant clear-sky bias, which is believed to have a seasonal variation. In general, there is a need for good quality reference data over sea ice, and in particular observations that also cover the MIZ, to improve retrieval derivation and validation.

In relation to SST, inclusion of other SST products may lead to improvements e.g. those based on PMW observations. The IR observations used as input to the L4 SST/IST product are severely limited by clouds, which are particularly extensive over the Arctic during summer (Intrieri, 2002; Key, 1990). PMW SST observations may provide an important supplement to the IR observations, since they are not prevented by non-precipitating clouds (Chelton and Wentz, 2005; Wentz et al., 2000). A number of different retrieval algorithms have been developed to retrieve SST from PMW observations (Alerskans et al., 2022, 2020; Chang et al., 2015; Meissner and Wentz, 2012; Milman and Wilheit, 1985; Nielsen-Englyst et al., 2018; Shibata, 2006; Wentz and Meissner, 2000). Hence, future work should investigate the possibilities of including PMW SST observations in the L4 SST/IST product derived here. This is particular of interest since the Copernicus Imaging Microwave Radiometer (CIMR), currently being implemented by ESA as a Copernicus Expansion Mission, will provide high-accuracy, high resolution PMW observations of the Polar Regions, which can be used to improve SST and SIC products in the future (Donlon, 2020).

6. Conclusion

This paper presents the first gap-free (L4) surface temperature climate dataset covering the ocean, sea ice and the marginal ice zone (MIZ) of the Arctic (>58°N) for the period from 1982 to May 2021. The dataset is based on thermal infrared observations from A(*A*)TSR, AVHRR and SLSTR, which have been combined using optimal interpolation (OI). Due to differences in the variability over ice, ocean and the MIZ, the OI statistical parameters have been derived separately for each surface type.

The derived L4 sea surface temperature (SST) and sea ice surface temperature (IST) product has been compared with different sources of in situ observations. The validation of the L4 SSTs shows mean differences of 0.01 °C, 0.04 °C and 0.04 °C and standard deviations of 0.54 °C, 0.56 °C and 0.51 °C for drifting buoys, moored buoys and Argo floats, respectively. The L4 ISTs have been compared with KT-19 measurements from 117 IceBridge flights, showing a mean difference of 1.52 °C and standard deviation of 3.12 °C, and with air temperatures (typically measured at about 2 m height) from the North Pole (NP) ice drifting stations as well as ECMWF distributed buoys and CRREL buoys, with mean differences of −2.35 °C, −3.21 °C and –2.87 °C and standard deviations of 3.12 °C, 3.34 °C and 3.36 °C, respectively. The large temperature differences against NP, ECMWF and CRREL are linked to the physical temperature difference between the skin ISTs provided here and the air temperatures measured in situ. The observed stability is −0.0001 °C/yr and 0.0047 °C/yr against drifters (SST) and NP (IST) observations, respectively, indicating a very stable performance throughout the record.

Traditionally, climate surface temperature trends have been estimated individually for SST and IST satellite based records. This is problematic in the Arctic region due to the large variability in the sea ice cover on decadal timescales, and the resulting climate trends are not easy to interpret. A combined surface temperature dataset of the ocean, sea ice and the MIZ provides a consistent climate indicator, which is important for monitoring day-to-day variations as well as climate trends in the Arctic region. The combined sea and sea ice surface temperatures of the Arctic have increased by around 4.5 °C between 1982 and 2021, with a peak warming of around 10 °C in the northeastern Barents Sea.

Data availability statement

The L4 SST/IST data are released through the Copernicus Marine Service, product SEAICE_ARC_PHY_CLIMATE_L4_MY_011_016 (doi: https://doi.org/10.48670/moi-00123) and will be updated regularly. For more information, see https://resources.marine.copernicus.eu/product-detail/SEAICE_ARC_PHY_CLIMATE_L4_MY_011_016/INFORMATION.

CRediT authorship contribution statement

Pia Nielsen-Englyst: Conceptualization, Methodology, Software, Investigation, Writing – original draft, Writing – review & editing. **Jacob L. Høyer:** Conceptualization, Methodology, Software, Investigation, Supervision, Writing – original draft, Writing – review & editing. **Wiebke M. Kolbe:** Methodology, Software, Investigation, Writing – original draft. **Gorm Dybkjær:** Investigation, Software, Writing – original draft. **Thomas Lavergne:** Software, Writing – original draft. **Rasmus Tage Tonboe:** Supervision, Writing – original draft. **Sotirios Skarpalezos:** Software. **Ioanna Karagali:** Project administration, Writing – original draft.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgments

This work is funded by the Copernicus Marine Environment Monitoring Service (CMEMS) and the Danish National Centre for Climate Research (NCKF) at the Danish Meteorological Institute (DMI). The authors would like to thank Leif Toudal Pedersen for his valuable inputs and constructive feedback during this work.

References

Alerskans, E., Høyer, J.L., Gentemann, C.L., Pedersen, L.T., Nielsen-Englyst, P., Donlon, C., 2020. Construction of a climate data record of sea surface temperature from passive microwave measurements. Remote Sens. Environ. 236, 111485 https://doi.org/10.1016/j.rse.2019.111485.

Alerskans, E., Zinck, A.-S.P., Nielsen-Englyst, P., Høyer, J.L., 2022. Exploring machine learning techniques to retrieve sea surface temperatures from passive microwave measurements. Remote Sens. Environ. 281, 113220 https://doi.org/10.1016/j.rse.2022.113220.

AMAP, 2021. Arctic climate change update 2021: Key trends and impacts. In: Summary for Policy-makers. Arctic Monitoring and Assessment Programme (AMAP), Tromsø, Norway.

Andersen, S., Tonboe, R., Kaleschke, L., Heygster, G., Pedersen, L.T., 2007. Intercomparison of passive microwave sea ice concentration retrievals over the high-concentration Arctic Sea ice. J. Geophys. Res. 112, C08004. https://doi.org/10.1029/2006JC003543.

Atkinson, C.P., Rayner, N.A., Kennedy, J.J., Good, S.A., 2014. An integrated database of ocean temperature and salinity observations. J. Geophys. Res. Oceans 119, 7139–7163. https://doi.org/10.1002/2014JC010053.

Banzon, V., Smith, T.M., Steele, M., Huang, B., Zhang, H.-M., 2020. Improved estimation of Proxy Sea surface temperature in the Arctic. J. Atmos. Ocean. Technol. 37, 341–349. https://doi.org/10.1175/JTECH-D-19-0177.1.

Bates, J.J., Diaz, H.F., 1991. Evaluation of multichannel sea surface temperature product quality for climate monitoring: 1982–1988. J. Geophys. Res. 96, 20613. https://doi.org/10.1029/91JC02280.

Bergman, K.H., 1978. Role of observational errors in optimum interpolation analysis. Bull. Am. Meteorol. Soc. 59, 1603–1611. https://doi.org/10.1175/1520-0477-59.12.1603.

Bojinski, S., Verstraete, M., Peterson, T.C., Richter, C., Simmons, A., Zemp, M., 2014. The concept of essential climate variables in support of climate research, applications, and policy. Bull. Am. Meteorol. Soc. 95, 1431–1443. https://doi.org/10.1175/BAMS-D-13-00047.1.

Bretherton, F.P., Davis, R.E., Fandry, C.B., 1976. A technique for objective analysis and design of oceanographic experiments applied to MODE-73. Deep-Sea Res. Oceanogr. Abstr. 23, 559–582. https://doi.org/10.1016/0011-7471(76)90001-2.

Bulgin, C.E., Merchant, C.J., Ferreira, D., 2020. Tendencies, variability and persistence of sea surface temperature anomalies. Sci. Rep. 10, 7986. https://doi.org/10.1038/s41598-020-64785-9.

Carton, J.A., Chepurin, G., Cao, X., Giese, B., 2000. A simple ocean data assimilation analysis of the global upper ocean 1950–95. Part I: method. J. Phys. Oceanogr. 30, 294–309.

Carvalho, K.S., Wang, S., 2020. Sea surface temperature variability in the Arctic Ocean and its marginal seas in a changing climate: patterns and mechanisms. Glob. Planet. Chang. 193, 103265 https://doi.org/10.1016/j.gloplacha.2020.103265.

Castro, S.L., Wick, G.A., Steele, M., 2016. Validation of satellite sea surface temperature analyses in the Beaufort Sea using UpTempO buoys. Remote Sens. Environ. 187, 458–475. https://doi.org/10.1016/j.rse.2016.10.035.

Centurioni, L.R., Turton, J., Lumpkin, R., Braasch, L., Brassington, G., Chao, Y., Charpentier, E., Chen, Z., Corlett, G., Dohan, K., Donlon, C., Gallage, C., Hormann, V., Ignatov, A., Ingleby, B., Jensen, R., Kelly-Gerreyn, B.A., Koszalka, I.M., Lin, X., Lindstrom, E., Maximenko, N., Merchant, C.J., Minnett, P., O'Carroll, A., Paluszkiewicz, T., Poli, P., Poulain, P.-M., Reverdin, G., Sun, X., Swail, V., Thurston, S., Wu, L., Yu, L., Wang, B., Zhang, D., 2019. Global in situ observations of essential climate and ocean variables at the Air-Sea Interface. Front. Mar. Sci. 6, 419. https://doi.org/10.3389/fmars.2019.00419.

Chang, P.S., Jelenak, Z., Alsweiss, S., 2015. Algorithm Theoretical Basis Document: GCOM-W1/AMSR2 Day-1 EDR Version 1.0.

Chelton, D.B., Wentz, F.J., 2005. Global microwave satellite observations of sea surface temperature for numerical weather prediction and climate research. Bull. Am. Meteorol. Soc. 86, 1097–1115. https://doi.org/10.1175/BAMS-86-8-1097.

Chiodi, A.M., Zhang, C., Cokelet, E.D., Yang, Q., Mordy, C.W., Gentemann, C.L., Cross, J. N., Lawrence-Slavas, N., Meinig, C., Steele, M., Harrison, D.E., Stabeno, P.J., Tabisola, H.M., Zhang, D., Burger, E.F., O'Brien, K.M., Wang, M., 2021. Exploring the Pacific Arctic seasonal ice zone with saildrone USVs. Front. Mar. Sci. 8, 640690 https://doi.org/10.3389/fmars.2021.640697.

Cho, K., Sasaki, N., Shimoda, H., Sakata, T., Nishio, F., 1996. Evaluation and improvement of SSM/I sea ice concentration algorithms for the Sea of Okhotsk. J. Remote Sens. Soc. Jpn. 16, 133–144. https://doi.org/10.11440/rssj1981.16.133.

Comiso, J.C., 2003. Warming trends in the Arctic from clear sky satellite observations. J. Clim. https://doi.org/10.1175/1520-0442(2003)016<3498:WTITAF>2.0.CO;2.

Comiso, J.C., Hall, D.K., 2014. Climate trends in the Arctic as observed from space: climate trends in the Arctic as observed from space. WIREs Clim. Change 5, 389–409. https://doi.org/10.1002/wcc.277.

Cowtan, K., Way, R., 2014. Update to "Coverage bias in the HadCRUT4 temperature series and its impact on recent temperature trends". Reconcil. Glob. Temp. Ser. https://doi.org/10.13140/RG.2.1.4334.8564.

Dash, P., Ignatov, A., Martin, M., Donlon, C., Brasnett, B., Reynolds, R.W., Banzon, V., Beggs, H., Cayula, J.-F., Chao, Y., Grumbine, R., Maturi, E., Harris, A., Mittaz, J., Sapper, J., Chin, T.M., Vazquez-Cuervo, J., Armstrong, E.M., Gentemann, C., Cummings, J., Piollé, J.-F., Autret, E., Roberts-Jones, J., Ishizaki, S., Høyer, J.L., Poulter, D., 2012. Group for High Resolution sea Surface Temperature (GHRSST) analysis fields inter-comparisons—Part 2: near real time web-based level 4 SST quality monitor (L4-SQUAM). Deep-Sea Res. II Top. Stud. Oceanogr. 77–80, 31–43. https://doi.org/10.1016/j.dsr2.2012.04.002.

Dee, D.P., Uppala, S.M., Simmons, A.J., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M.A., Balsamo, G., Bauer, P., Bechtold, P., Beljaars, A.C.M., van de Berg, L., Bidlot, J., Bormann, N., Delsol, C., Dragani, R., Fuentes, M., Geer, A.J., Haimberger, L., Healy, S.B., Hersbach, H., Hólm, E.V., Isaksen, L., Kållberg, P., Köhler, M., Matricardi, M., McNally, A.P., Monge-Sanz, B.M., Morcrette, J.-J., Park, B.-K., Peubey, C., de Rosnay, P., Tavolato, C., Thépaut, J.-N., Vitart, F., 2011. The ERA-interim reanalysis: configuration and performance of the data assimilation system. Q. J. R. Meteorol. Soc. 137, 553–597. https://doi.org/10.1002/qj.828.

Dodd, E.M.A., Veal, K.L., Ghent, D.J., Broeke, M.R., Remedios, J.J., 2019. Toward a combined surface temperature data set for the Arctic from the along-track scanning radiometers. J. Geophys. Res. Atmos. 2019JD030262 https://doi.org/10.1029/2019JD030262.

Donlon, C.J., 2020. Copernicus Imaging Microwave Radiometer (CIMR) Mission Requirements Document. Version 4.0.

Donlon, C., Casey, K.S., Gentemann, C., Harris, A., 2010. Successes and challenges for the modern sea surface temperature observing system. In: Proceeding of OceanObs'09: Sustained Ocean Observations and Information for Society.

Donlon, C.J., Martin, M., Stark, J., Roberts-Jones, J., Fiedler, E., Wimmer, W., 2012. The Operational Sea surface temperature and sea ice analysis (OSTIA) system. Remote Sens. Environ. 116, 140–158. https://doi.org/10.1016/j.rse.2010.10.017.

Dybbroe, A., Karlsson, K.-G., Thoss, A., 2005a. NWCSAF AVHRR cloud detection and analysis using dynamic thresholds and radiative transfer modeling. Part I: algorithm description. J. Appl. Meteorol. 44, 39–54. https://doi.org/10.1175/JAM-2188.1.

Dybbroe, A., Karlsson, K.-G., Thoss, A., 2005b. NWCSAF AVHRR cloud detection and analysis using dynamic thresholds and radiative transfer modeling. Part II: tuning and validation. J. Appl. Meteorol. 44, 55–71. https://doi.org/10.1175/JAM-2189.1.

Dybkjær, G., Tonboe, R., Høyer, J.L., 2012. Arctic surface temperatures from metop AVHRR compared to in situ ocean and land data. Ocean Sci. 8, 959–970. https://doi.org/10.5194/os-8-959-2012.

Dybkjær, G., Høyer, J.L., Tonboe, R., Olsen, S.M., 2014. Report on the documentation and description of the new Arctic Ocean dataset combining SST and IST. NACLIM Deliv. D32, 28.

Dybkjær, G., Eastwood, S., Borg, A.L., Høyer, J.L., Tonboe, R., 2018. Algorithm theoretical basis document (ATBD) for the OSI SAF Sea and sea ice surface temperature L2 processing chain. OSI205a and OSI205b.

Dybkjær, G., Eastwood, S., Tonboe, R.T., Nielsen-Englyst, P., Høyer, J.L., Kolbe, W., Jensen, T., Skarpalezos, S., 2022. Arctic and Antarctic Surface Temperatures from AVHRR thermal Infrared satellite sensors 1982-2015; AASTI v2. In prep.

Embury, O., Merchant, C.J., Corlett, G.K., 2012. A reprocessing for climate of sea surface temperature from the along-track scanning radiometers: initial validation, accounting for skin and diurnal variability effects. Remote Sens. Environ. 116, 62–78. https://doi.org/10.1016/j.rse.2011.02.028.

Fiedler, E.K., McLaren, A., Banzon, V., Brasnett, B., Ishizaki, S., Kennedy, J., Rayner, N., Roberts-Jones, J., Corlett, G., Merchant, C.J., Donlon, C., 2019. Intercomparison of long-term sea surface temperature analyses using the GHRSST multi-product ensemble (GMPE) system. Remote Sens. Environ. 222, 18–33. https://doi.org/10.1016/j.rse.2018.12.015.

Folland, C.K., Rayner, N.A., Brown, S.J., Smith, T.M., Shen, S.S.P., Parker, D.E., Macadam, I., Jones, P.D., Jones, R.N., Nicholls, N., Sexton, D.M.H., 2001. Global temperature change and its uncertainties since 1861. Geophys. Res. Lett. 28, 2621–2624. https://doi.org/10.1029/2001GL012877.

Gandin, L.S., 1963. Objective analysis of meteorological field. Gidrometeor. Isdat, Leningrad.

GCOS, 2016. The Global Observing System for Climate: Implementation Needs (No. GCOS-200 (GOOS-214)). World Meteorological Organization (WMO).

Good, S.A., Embury, O., Bulgin, C.E., Mittaz, J., 2019. ESA Sea Surface Temperature Climate Change Initiative (SST_cci): Level 4 Analysis Climate Data Record, Version 2.1. https://doi.org/10.5285/62C0F97B1EAC4E0197A674870AFE1EE6.

Good, S., Fiedler, E., Mao, C., Martin, M.J., Maycock, A., Reid, R., Roberts-Jones, J., Searle, T., Waters, J., While, J., Worsfold, M., 2020. The current configuration of the OSTIA system for operational production of Foundation Sea surface temperature and ice concentration analyses. Remote Sens. 12, 720. https://doi.org/10.3390/rs12040720.

Graversen, R.G., Mauritsen, T., Tjernström, M., Källén, E., Svensson, G., 2008. Vertical structure of recent Arctic warming. Nature 451, 53–56. https://doi.org/10.1038/nature06502.

Hall, D.K., Key, J.R., Case, K.A., Riggs, G.A., Cavalieri, D.J., 2004. Sea ice surface temperature product from MODIS. IEEE Trans. Geosci. Remote Sens. 42, 1076–1087. https://doi.org/10.1109/TGRS.2004.825587.

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., Simmons, A., Soci, C., Abdalla, S., Abellan, X., Balsamo, G., Bechtold, P., Biavati, G., Bidlot, J., Bonavita, M., Chiara, G., Dahlgren, P., Dee, D., Diamantakis, M., Dragani, R., Flemming, J., Forbes, R., Fuentes, M., Geer, A., Haimberger, L., Healy, S., Hogan, R.J., Hólm, E., Janisková, M., Keeley, S., Laloyaux, P., Lopez, P., Lupu, C., Radnoti, G., Rosnay, P.,

Rozum, I., Vamborg, F., Villaume, S., Thépaut, J., 2020. The ERA5 global reanalysis. Q.J.RMeteorol. Soc. 146, 1999–2049. https://doi.org/10.1002/qj.3803.
Høyer, J.L., Karagali, I., 2016. Sea surface temperature climate data record for the North Sea and Baltic Sea. J. Clim. 29, 2529–2541. https://doi.org/10.1175/JCLI-D-15-0663.1.
Høyer, J.L., She, J., 2007. Optimal interpolation of sea surface temperature for the North Sea and Baltic Sea. J. Mar. Syst. 65, 176–189. https://doi.org/10.1016/j.jmarsys.2005.03.008.
Høyer, J.L., Karagali, I., Dybkjær, G., Tonboe, R., 2012. Multi sensor validation and error characteristics of Arctic satellite sea surface temperature observations. Remote Sens. Environ. 121, 335–346. https://doi.org/10.1016/j.rse.2012.01.013.
Høyer, J.L., Le Borgne, P., Eastwood, S., 2014. A bias correction method for Arctic satellite sea surface temperature observations. Remote Sens. Environ. 146, 201–213. https://doi.org/10.1016/j.rse.2013.04.020.
Hurrell, J.W., Hack, J.J., Shea, D., Caron, J.M., Rosinski, J., 2008. A New Sea surface temperature and sea ice boundary dataset for the community atmosphere model. J. Clim. 21, 5145–5153. https://doi.org/10.1175/2008JCLI2292.1.
Intrieri, J.M., 2002. An annual cycle of Arctic surface cloud forcing at SHEBA. J. Geophys. Res. 107, 8039. https://doi.org/10.1029/2000JC000439.
IPCC, 2019. Special Report on the Ocean and Cryosphere in a Changing Climate..
Jia, C., Minnett, P.J., 2020. High latitude sea surface temperatures derived from MODIS infrared measurements. Remote Sens. Environ. 251, 112094 https://doi.org/10.1016/j.rse.2020.112094.
Kaplan, A., Cane, M.A., Kushnir, Y., Clement, A.C., Blumenthal, M.B., Rajagopalan, B., 1998. Analyses of global sea surface temperature 1856–1991. J. Geophys. Res. 103, 18567–18589. https://doi.org/10.1029/97JC01736.
Kennedy, J.J., 2014. A review of uncertainty in in situ measurements and data sets of sea surface temperature: IN SITU SST UNCERTAINTY. Rev. Geophys. 52, 1–32. https://doi.org/10.1002/2013RG000434.
Kern, S., Lavergne, T., Pedersen, L.T., Tonboe, R.T., Bell, L., Meyer, M., Zeigermann, L., 2022. Satellite passive microwave sea-ice concentration data set intercomparison using landsat data. Cryosphere 16, 349–378. https://doi.org/10.5194/tc-16-349-2022.
Key, J., 1990. Cloud cover analysis with Arctic advanced very high resolution radiometer data: 2. Classification with spectral and textural measures. J. Geophys. Res. 95, 7661. https://doi.org/10.1029/JD095iD06p07661.
Key, J., Maslanik, J.A., Papakyriakou, T., Serreze, M.C., Schweiger, A.J., 1994. On the validation of satellite-Derived Sea ice surface temperature. Arctic 47. https://doi.org/10.14430/arctic1298.
Key, J.R., Collins, J.B., Fowler, C., Stone, R.S., 1997. High-latitude surface temperature estimates from thermal satellite data. Remote Sens. Environ. 61, 302–309. https://doi.org/10.1016/S0034-4257(97)89497-7.
Key, J.R., Mahoney, R., Liu, Y., Romanov, P., Tschudi, M., Appel, I., Maslanik, J., Baldwin, D., Wang, X., Meade, P., 2013. Snow and ice products from suomi NPP VIIRS. J. Geophys. Res. Atmos. 118, 12816–12830. https://doi.org/10.1002/2013JD020459.
Kobayashi, S., Ota, Y., Harada, Y., Ebita, A., Moriya, M., Onoda, H., Onogi, K., Kamahori, H., Kobayashi, C., Endo, H., Miyaoka, K., Takahashi, K., 2015. The JRA-55 reanalysis: general specifications and basic characteristics. J. Meteorol. Soc. Jpn. 93, 5–48. https://doi.org/10.2151/jmsj.2015-001.
Larsen, J., Høyer, J.L., She, J., 2007. Validation of a hybrid optimal interpolation and Kalman filter scheme for sea surface temperature assimilation. J. Mar. Syst. 65, 122–133. https://doi.org/10.1016/j.jmarsys.2005.09.013.
Lavergne, T., Sørensen, A.M., Kern, S., Tonboe, R., Notz, D., Aaboe, S., Bell, L., Dybkjær, G., Eastwood, S., Gabarro, C., Heygster, G., Killie, M.A., Brandt Kreiner, M., Lavelle, J., Saldo, R., Sandven, S., Pedersen, L.T., 2019. Version 2 of the EUMETSAT OSI SAF and ESA CCI Sea-ice concentration climate data records. Cryosphere 13, 49–78. https://doi.org/10.5194/tc-13-49-2019.
Lavergne, T., Kern, S., Aaboe, S., Derby, L., Dybkjær, G., Garric, G., Heil, P., Hendricks, S., Holfort, J., Howell, S., Key, J., Lieser, J., Maksym, T., Maslowski, W., Meier, W.N., Munoz-Sabater, J., Nicolas, J., Özsoy, B., Rabe, B., Rack, W., Raphael, M., de Rosnay, P., Smolyanitsky, V., Tietsche, S., Ukita, J., Vichi, V., Wagner, P., Willmes, S., Zhao, X., 2022. A new structure for the sea ice essential climate variables of the global climate observing system. Under review.
Lenssen, N.J.L., Schmidt, G.A., Hansen, J.E., Menne, M.J., Persin, A., Ruedy, R., Zyss, D., 2019. Improvements in the GISTEMP uncertainty model. J. Geophys. Res. Atmos. 124, 6307–6326. https://doi.org/10.1029/2018JD029522.
Liang, X., Ignatov, A., 2013. AVHRR, MODIS, and VIIRS radiometric stability and consistency in SST bands: AVHRR-MODIS-VIIRS consistency. J. Geophys. Res. Oceans 118, 3161–3171. https://doi.org/10.1002/jgrc.20205.
Liu, Y., Key, J.R., Wang, X., 2008. The influence of changes in cloud cover on recent surface temperature trends in the Arctic. J. Clim. 21, 705–715. https://doi.org/10.1175/2007JCLI1681.1.
Liu, Y., Key, J., Tschudi, M., Dworak, R., Mahoney, R., Baldwin, D., 2015. Validation of the suomi NPP VIIRS ice surface temperature environmental data record. Remote Sens. 7, 17258–17271. https://doi.org/10.3390/rs71215880.
Markus, T., Cavalieri, D.J., 2009. The AMSR-E NT2 Sea Ice Concentration Algorithm : Its Basis and Implementation.
Marquardt Collow, A.B., Cullather, R.I., Bosilovich, M.G., 2020. Recent Arctic Ocean surface air temperatures in atmospheric reanalyses and numerical simulations. J. Clim. 33, 4347–4367. https://doi.org/10.1175/JCLI-D-19-0703.1.
Martin, M., Dash, P., Ignatov, A., Banzon, V., Beggs, H., Brasnett, B., Cayula, J.-F., Cummings, J., Donlon, C., Gentemann, C., Grumbine, R., Ishizaki, S., Maturi, E., Reynolds, R.W., Roberts-Jones, J., 2012. Group for High Resolution sea Surface temperature (GHRSST) analysis fields inter-comparisons. Part 1: a GHRSST multi-product ensemble (GMPE). Deep-Sea Res. II Top. Stud. Oceanogr. 77–80, 21–30. https://doi.org/10.1016/j.dsr2.2012.04.013.
Maslanik, J., Fowler, C., Key, J., Scambos, T., Hutchinson, T., Emery, W., 1997. AVHRR-based polar pathfinder products for modeling applications. Ann. Glaciol. 25, 388–392. https://doi.org/10.3189/S0260305500014336.
Maykut, G.A., 1986. The surface heat and mass balance. In: Untersteiner, N. (Ed.), The Geophysics of Sea Ice. Springer, US, Boston, MA, pp. 395–463. https://doi.org/10.1007/978-1-4899-5352-0_6.
Meier, W.N., 2012. Climate algorithm theoretical basis document (C-ATBD). In: Passive Microwave Sea Ice Concentration. National Oceanic and Atmospheric Administration Climate Data Record Program, Asheville, NC.
Meissner, T., Wentz, F.J., 2012. The emissivity of the ocean surface between 6 and 90 GHz over a large range of wind speeds and earth incidence angles. IEEE Trans. Geosci. Remote Sens. 50, 3004–3026. https://doi.org/10.1109/TGRS.2011.2179662.
Merchant, C.J., Embury, O., Roberts-Jones, J., Fiedler, E., Bulgin, C.E., Corlett, G.K., Good, S., McLaren, A., Rayner, N., Morak-Bozzo, S., Donlon, C., 2014. Sea surface temperature datasets for climate applications from phase 1 of the European Space Agency climate change initiative (SST CCI). Geosci. Data J. 1, 179–191. https://doi.org/10.1002/gdj3.20.
Merchant, C.J., Embury, O., Roberts-Jones, J., Fiedler, E., Bulgin, C.E., Corlett, G.K., Good, S., Rayner, N., Donlon, C., 2016. ESA Sea Surface Temperature Climate Change Initiative (ESA SST CCI): Analysis Long Term Product Version 1.1. https://doi.org/10.5285/2262690A-B588-4704-B459-39E05527B59A.
Merchant, C.J., Paul, F., Popp, T., Ablain, M., Bontemps, S., Defourny, P., Hollmann, R., Lavergne, T., Laeng, A., de Leeuw, G., Mittaz, J., Poulsen, C., Povey, A.C., Reuter, M., Sathyendranath, S., Sandven, S., Sofieva, V.F., Wagner, W., 2017. Uncertainty information in climate data records from earth observation. Earth Syst.Sci. Data 9, 511–527. https://doi.org/10.5194/essd-9-511-2017.
Merchant, C.J., Embury, O., Bulgin, C.E., Block, T., Corlett, G.K., Fiedler, E., Good, S.A., Mittaz, J., Rayner, N.A., Berry, D., Eastwood, S., Taylor, M., Tsushima, Y., Waterfall, A., Wilson, R., Donlon, C., 2019. Satellite-based time-series of sea-surface temperature since 1981 for climate applications. Sci Data 6, 223. https://doi.org/10.1038/s41597-019-0236-x.
Milman, A.S., Wilheit, T.T., 1985. Sea surface temperatures from the scanning multichannel microwave radiometer on Nimbus 7. J. Geophys. Res. 90, 11631. https://doi.org/10.1029/JC090iC06p11631.
Minnett, P.J., Kilpatrick, K.A., Podestá, G.P., Evans, R.H., Szczodrak, M.D., Izaguirre, M. A., Williams, E.J., Walsh, S., Reynolds, R.M., Bailey, S.W., Armstrong, E.M., Vazquez-Cuervo, J., 2020. Skin Sea-surface temperature from VIIRS on suomi-NPP—NASA continuity retrievals. Remote Sens. 12, 3369. https://doi.org/10.3390/rs12203369.
Nielsen-Englyst, P., Høyer, L., Pedersen, J.Toudal, Gentemann, L.L., Alerskans, C., Block, E., Donlon, T., 2018. Optimal estimation of sea surface temperature from AMSR-E. Remote Sens. 10, 229. https://doi.org/10.3390/rs10020229.
Nielsen-Englyst, P., Høyer, J.L., Madsen, K.S., Tonboe, R., Dybkjær, G., Alerskans, E., 2019. In situ observed relationships between snow and ice surface skin temperatures and 2 m air temperatures in the Arctic. Cryosphere 13, 1005–1024. https://doi.org/10.5194/tc-13-1005-2019.
Nielsen-Englyst, P., Høyer, J.L., Madsen, K.S., Tonboe, R.T., Dybkjær, G., Skarpalezos, S., 2021. Deriving Arctic 2 m air temperatures over snow and ice from satellite surface temperature measurements. Cryosphere 15, 3035–3057. https://doi.org/10.5194/tc-15-3035-2021.
O'Carroll, A.G., Armstrong, E.M., Beggs, H.M., Bouali, M., Casey, K.S., Corlett, G.K., Dash, P., Donlon, C.J., Gentemann, C.L., Høyer, J.L., Ignatov, A., Kabobah, K., Kachi, M., Kurihara, Y., Karagali, I., Maturi, E., Merchant, C.J., Marullo, S., Minnett, P.J., Pennybacker, M., Ramakrishnan, B., Ramsankaran, R., Santoleri, R., Sunder, S., Saux Picart, S., Vázquez-Cuervo, J., Wimmer, W., 2019. Observational needs of sea surface temperature. Front. Mar. Sci. 6, 420. https://doi.org/10.3389/fmars.2019.00420.
Ohring, G., Wielicki, B., Spencer, R., Emery, B., Datla, R., 2005. Satellite instrument calibration for measuring global climate change: report of a workshop. Bull. Amer. Meteor. Soc. 86, 1303–1314. https://doi.org/10.1175/BAMS-86-9-1303.
Oke, P.R., Brassington, G.B., Griffin, D.A., Schiller, A., 2008. The Bluelink Ocean data assimilation system (BODAS). Ocean Model. 21, 46–70. https://doi.org/10.1016/j.ocemod.2007.11.002.
Okuro, A., Kubota, M., Tomita, H., Hihara, T., 2014. Inter-comparison of various global sea surface temperature products. Int. J. Remote Sens. 35, 5394–5410. https://doi.org/10.1080/01431161.2014.926415.
Petrenko, B., Ignatov, A., Kihai, Y., Stroup, J., Dash, P., 2014. Evaluation and selection of SST regression algorithms for JPSS VIIRS. J. Geophys. Res. Atmos. 119, 4580–4599. https://doi.org/10.1002/2013JD020637.
Pithan, F., Mauritsen, T., 2014. Arctic amplification dominated by temperature feedbacks in contemporary climate models. Nat. Geosci. 7, 181–184. https://doi.org/10.1038/ngeo2071.
Rantanen, M., Karpechko, A.Yu., Lipponen, A., Nordling, K., Hyvärinen, O., Ruosteenoja, K., Vihma, T., Laaksonen, A., 2022. The Arctic has warmed nearly four times faster than the globe since 1979. Commun. Earth Environ. 3 https://doi.org/10.1038/s43247-022-00498-3.
Rapaić, M., Brown, R., Markovic, M., Chaumont, D., 2015. An evaluation of temperature and precipitation surface-based and reanalysis datasets for the Canadian Arctic, 1950–2010. Atmosphere-Ocean 53, 283–303. https://doi.org/10.1080/07055900.2015.1045825.
Rasmussen, T.A.S., Høyer, J.L., Ghent, D., Bulgin, C.E., Dybkjaer, G., Ribergaard, M.H., Nielsen-Englyst, P., Madsen, K.S., 2018. Impact of assimilation of sea-ice surface

temperatures on a coupled ocean and sea-ice model. J. Geophys. Res. Oceans 123, 2440–2460. https://doi.org/10.1002/2017JC013481.
Rayner, N.A., 2003. Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century. J. Geophys. Res. 108, 4407. https://doi.org/10.1029/2002JD002670.
Rayner, N.A., Auchmann, R., Bessembinder, J., Brönnimann, S., Brugnara, Y., Capponi, F., Carrea, L., Dodd, E.M.A., Ghent, D., Good, E., Høyer, J.L., Kennedy, J.J., Kent, E.C., Killick, R.E., van der Linden, P., Lindgren, F., Madsen, K.S., Merchant, C. J., Mitchelson, J.R., Morice, C.P., Nielsen-Englyst, P., Ortiz, P.F., Remedios, J.J., van der Schrier, G., Squintu, A.A., Stephens, A., Thorne, P.W., Tonboe, R.T., Trent, T., Veal, K.L., Waterfall, A.M., Winfield, K., Winn, J., Woolway, R.I., 2020. The EUSTACE project: delivering global, daily information on surface air temperature. Bull. Am. Meteorol. Soc. 101, E1924–E1947. https://doi.org/10.1175/BAMS-D-19-0095.1.
Reynolds, R.W., Smith, T.M., 1994. Improved Global Sea surface temperature analyses using optimum interpolation. J. Clim. 7, 929–948. https://doi.org/10.1175/1520-0442(1994)007<0929:IGSSTA>2.0.CO;2.
Reynolds, R.W., Rayner, N.A., Smith, T.M., Stokes, D.C., Wang, W., 2002. An improved in situ and satellite SST analysis for climate. J. Clim. 15, 1609–1625. https://doi.org/10.1175/1520-0442(2002)015<1609:AIISAS>2.0.CO;2.
Reynolds, R.W., Smith, T.M., Liu, C., Chelton, D.B., Casey, K.S., Schlax, M.G., 2007. Daily high-resolution-blended analyses for sea surface temperature. J. Clim. 20, 5473–5496. https://doi.org/10.1175/2007JCLI1824.1.
Richter-Menge, J.A., Overland, J., Mathis, J.T., Osborne, E., 2017. Arctic Report Card 2017.
Roberts-Jones, J., Fiedler, E.K., Martin, M.J., 2012. Daily, global, high-resolution SST and sea ice reanalysis for 1985–2007 using the OSTIA system. J. Clim. 25, 6215–6232. https://doi.org/10.1175/JCLI-D-11-00648.1.
RU/FSR/HME/AARI, NSIDC, 1993. AARI Russian North Polar Drifting Station Data, from NSIDC. https://doi.org/10.5065/V47S-KW40.
Serreze, M.C., Box, J.E., Barry, R.G., Walsh, J.E., 1993. Characteristics of Arctic synoptic activity, 1952–1989. Meteorl. Atmos. Phys. 51, 147–164. https://doi.org/10.1007/BF01030491.
Shibata, A., 2006. Features of ocean microwave emission changed by wind at 6 GHz. J. Oceanogr. 62, 321–330. https://doi.org/10.1007/s10872-006-0057-3.
Simmons, A.J., 2004. Comparison of trends and low-frequency variability in CRU, ERA-40, and NCEP/NCAR analyses of surface air temperature. J. Geophys. Res. 109, D24115. https://doi.org/10.1029/2004JD005306.
SMHI, 2014. Algorithm Theoretical Basis Document for the Cloud Mask of the NWC/PPS. NWC/CDOP2/PPS/SMHI/SCI/ATBD/1, Issue 1, Rev. 0.
Song, X., Yu, L., 2012. High-latitude contribution to global variability of Air-Sea sensible heat flux. J. Clim. 25, 3515–3531. https://doi.org/10.1175/JCLI-D-11-00028.1.
Studinger, M., 2020. IceBridge KT19 IR Surface Temperature, Version 2. https://doi.org/10.5067/UHE07J35I3NB.
Timmermans, M.-L., Labe, Z., 2020. Arctic Report Card 2020: Sea Surface Temperature. https://doi.org/10.25923/V0FS-M920.
Tonboe, R.T., Dybkjær, G., Høyer, J.L., 2011. Simulations of the snow covered sea ice surface temperature and microwave effective temperature. Tellus A 63, 1028–1037. https://doi.org/10.1111/j.1600-0870.2011.00530.x.
Tonboe, R.T., Eastwood, S., Lavergne, T., Sørensen, A.M., Rathmann, N., Dybkjær, G., Pedersen, L.T., Høyer, J.L., Kern, S., 2016. The EUMETSAT Sea ice concentration climate data record. Cryosphere 10, 2275–2290. https://doi.org/10.5194/tc-10-2275-2016.
Vazquez-Cuervo, J., Castro, S.L., Steele, M., Gentemann, C., Gomez-Valdes, J., Tang, W., 2022. Comparison of GHRSST SST analysis in the Arctic Ocean and alaskan coastal waters using saildrones. Remote Sens. 14, 692. https://doi.org/10.3390/rs14030692.
Vincent, R.F., Marsden, R.F., Minnett, P.J., Buckley, J.R., 2008a. Arctic waters and marginal ice zones: 2. An investigation of arctic atmospheric infrared absorption for advanced very high resolution radiometer sea surface temperature estimates. J. Geophys. Res. 113, C08044. https://doi.org/10.1029/2007JC004354.
Vincent, R.F., Marsden, R.F., Minnett, P.J., Creber, K.A.M., Buckley, J.R., 2008b. Arctic waters and marginal ice zones: a composite Arctic Sea surface temperature algorithm using satellite thermal data. J. Geophys. Res. 113, C04021. https://doi.org/10.1029/2007JC004353.
Walsh, J.E., Chapman, W.L., 1998. Arctic Cloud–Radiation–Temperature associations in observational data and atmospheric reanalyses. J. Clim. 11, 3030–3045. https://doi.org/10.1175/1520-0442(1998)011<3030:ACRTAI>2.0.CO;2.
Wang, X., Key, J.R., 2005. Arctic surface, cloud, and radiation properties based on the AVHRR polar pathfinder dataset. Part II: recent trends. J. Clim. 18, 2575–2593. https://doi.org/10.1175/JCLI3439.1.
Wang, H., Guan, L., Chen, G., 2016. Evaluation of sea surface temperature from FY-3C VIRR data in the Arctic. IEEE Geosci. Remote Sens. Lett. 13, 292–296. https://doi.org/10.1109/LGRS.2015.2511184.
Wentz, F.J., Meissner, T., 2000. AMSR Ocean Algorithm. Algorithm Theoretical Basis Document (No. Version 2). Remote Sensing Systems, Santa Rosa, CA.
Wentz, F.J., Gentemann, C., Smith, D., Chelton, D.B., 2000. Satellite measurements of sea surface temperature through clouds. Science 288, 847–850. https://doi.org/10.1126/science.288.5467.847.
White, W.B., 1995. Design of a global observing system for gyre-scale upper ocean temperature variability. Prog. Oceanogr. 36, 169–217. https://doi.org/10.1016/0079-6611(95)00017-8.

APPENDIX E

Paper V: Impact of microwave observations on Arctic sea surface temperatures

This appendix contains a full copy of Paper V:

- **Nielsen-Englyst, Pia**, Jacob L. Høyer, Ioanna Karagali, Wiebke M. Kolbe, Rasmus T. Tonboe, Leif T. Pedersen. 2023. "Impact of microwave observations on Arctic sea surface temperature". In review –*Remote Sensing of Environment*

Impact of microwave observations on the estimation of Arctic sea surface temperatures

Pia Nielsen-Englyst[a,b,*], Jacob L. Høyer[b], Ioanna Karagali[b], Wiebke M. Kolbe[a,b], Rasmus T. Tonboe[a], Leif T. Pedersen[a]

[a] *DTU-Space, Technical University of Denmark, Lyngby, Denmark*
[b] *National Centre for Climate Research (NCKF), Danish Meteorological Institute (DMI), Copenhagen, Denmark*

Abstract

The frequent and persistent cloud cover in the Arctic limits the extent to which sea surface temperature (SST) can be retrieved from thermal infrared (IR) satellite sensors. Passive microwave (PMW) observations provide highly complementary information to IR, enabling measurements through non-precipitating clouds, although at a coarser spatial resolution. The differences in coverage, accuracy, footprint size, spatial resolution and error characteristics between IR and PMW SSTs require a systematic assessment of how to best combine IR and PMW SST retrievals. This is provided in this study on the basis of the ESA-CCI PMW SST climate data record (CDR) and an existing IR-based gap-free sea and sea ice surface temperature CDR covering the Arctic (>58°N), where cloud cover is a serious limitation to IR sensors. An important step towards a combined IR and PMW SST CDR is to correct for systematic biases in the PMW and IR SST datasets relative to each other. The PMW SSTs show reduced biases against *in situ* SSTs compared to the IR SSTs, but for consistency with time periods when no Arctic PMW SSTs were available, the PMW SSTs have been adjusted to the IR SSTs in this study. This is done using a dynamic bias correction to generate a consistent combined IR and PMW Arctic SST CDR for the period 2002-2017. Including PMW SSTs reduces the standard deviations from 0.54°C, 0.55°C and 0.47°C to 0.47°C, 0.54°C and 0.41°C against drifters, moorings and Argo floats, respectively. The improved performance is seen in almost all regions (including those already covered by IR observations), with the largest improvement in IR data sparse regions. The average theoretical uncertainty reduces by 0.08°C, which is in good agreement with the observed improvement in the standard deviation against drifters. The results are very promising and expected to improve even further in the future with the launch of the Copernicus Imaging Microwave Radiometer (CIMR), which will enable PMW SST retrievals with lower uncertainties and much closer to coasts and sea ice (where the largest uncertainties arise) than what is possible with previous and current PMW radiometers.

Keywords: Arctic, infrared satellite observations, passive microwave satellite observations, sea surface temperature (SST), AMSR-E, AMSR2

*Corresponding author
Email address: pne@dmi.dk (Pia Nielsen-Englyst)

1. Introduction

The sea surface temperature (SST) is an Essential Climate Variable (ECV) used for monitoring, understanding and predicting climate change (Bojinski et al., 2014). The Arctic is warming more rapidly than the global average, due to a number of amplifying feedback mechanisms (e.g. AMAP, 2021; Pithan and Mauritsen, 2014; Meredith et al., 2019; Rantanen et al., 2022), which makes it a very important region to monitor. The extreme environment and the poor accessibility make *in situ* observations challenging and sparse in the Arctic (Centurioni et al., 2019; Donlon et al., 2012).

Satellite observations are an important tool for monitoring the Arctic due to the high spatial and temporal coverage. There are several global satellite-based gap-free (i.e. Level 4, L4) SST products (e.g. Reynolds et al., 2007; Merchant et al., 2019; Donlon et al., 2012), but these usually show large uncertainties and diversity in the Arctic (Dash et al., 2012; Castro et al., 2016; Vazquez-Cuervo et al., 2022), where extreme environmental conditions, limited *in situ* data, persistent cloud cover and a varying length of the sunlit part of the day round the year complicate accurate SST retrievals from satellites (Donlon et al., 2009; Høyer et al., 2012; Minnett et al., 2019). Improving the Arctic SST data has been identified as being of "high priority" for future SST research and developments (O'Carroll et al., 2019). Specialized high latitude algorithms have previously been developed and shown to surpass the global L4 SST products in the Arctic (Jia and Minnett, 2020; Vincent et al., 2008b,a). Moreover, infrared (IR) satellite observations have recently been used to produce the first satellite-based L4 climate data record (CDR) of combined ocean and sea ice surface temperature in the Arctic (>58°N), which enables consistent climate monitoring of the Arctic warming (Nielsen-Englyst et al., 2023).

The IR SST observations are hampered by clouds, and data gaps usually remain after combining different IR SST datasets. Two common methods used to fill in the data gaps are temporal extension (e.g. Reynolds and Smith, 1994) and spatial interpolation techniques (e.g. Reynolds and Smith, 1994; Thiébaux et al., 2003; Donlon et al., 2012; Nielsen-Englyst et al., 2023). In the Arctic, frequent and persistent cloud cover results in long periods without surface coverage from IR sensors, and the cloud-contaminated observations are often difficult to identify. Therefore, the IR SST analyses usually rely heavily on the choice of interpolation technique and cloud masking in the Arctic, resulting in large sampling errors (Liu and Minnett, 2016). SSTs derived from passive microwave (PMW) observations have the potential to fill in large and persistent data gaps in the IR coverage, since PMW observations are less impacted by clouds and aerosols (Donlon et al., 2007, 2009; Ulaby et al., 2014; Wentz and Meissner, 2000) than IR sensors which nevertheless tend to provide higher spatial resolution. SSTs retrieved from IR sensors usually have spatial resolutions of about 1-4 km and uncertainties of 0.2-0.4°C (Donlon et al., 2007; Merchant et al., 2019; Reynolds et al., 2002; Embury et al., 2012), while PMW SSTs have spatial resolutions of about 50 km and uncertainties of 0.4-0.5°C, with the largest uncertainties in high latitudes (Nielsen-Englyst et al., 2018; Alerskans et al., 2020, 2022; Gentemann, 2014; Wentz et al., 2000; Shibata, 2006).

Current and previous PMW sensors do not capture subscale to mesoscale SST features and are influenced

by land near (∼100 km) coasts and sea ice due to the large field of view. Improved spatial resolution of the 6.9 and 10.7 GHz channels could lead to substantial improvements of PMW SST retrievals and their information content in global and regional SST products (O'Carroll et al., 2019). This is one of the primary objectives of the Copernicus Imaging Microwave Radiometer (CIMR) by the European Space Agency (ESA) as a part of the Copernicus Expansion program of the European Union (http://www.cimr.eu/). CIMR will provide high-accuracy, high resolution PMW observations of the Polar Regions, which will enable retrievals of SST and other surface parameters at a higher spatial resolution and lower uncertainty than what is possible with the current PMW missions (Donlon, 2020). In addition to CIMR, the Advanced Microwave Scanning Radiometer 2 (AMSR2) follow-on mission (AMSR3) is currently being prepared by Japan Aerospace Exploration Agency (JAXA) (Kasahara et al., 2020).

The large potential of including high latitude PMW SST retrievals with frequent updates from previous, current and future PMW missions, makes it important to investigate how to best combine IR and PMW SST retrievals in an Arctic analysis. Many global L4 SST analyses already include PMW SST observations e.g. the NOAA Optimum Interpolation (OI) SST V2 (Reynolds et al., 2007; Huang et al., 2021), Operational Sea Surface Temperature and Sea Ice Analysis (OSTIA; Donlon et al., 2012; Good et al., 2020), Canadian Meteorological Center (CMC) SST analysis (Brasnett, 2008), Remote Sensing Systems (REMSS) MW-IR SST product (`http://www.remss.com/measurements/sea-surface-temperature/oisst-description`), and the Multi-scale Ultra-high Resolution (MUR) SST analysis (Chin et al., 2017). Prior systematic efforts have investigated the impact of including satellite SST retrievals from the IR Advanced Very High Resolution Radiometer (AVHRR), the Advanced Microwave Scanning Radiometer – Earth Observing System (AMSR-E) and the Tropical Rainfall Measuring Mission (TRMM) Microwave Imager (TMI) to an existing *in situ* data based global analysis (Reynolds et al., 2004, 2007), and later the impact of further adding the IR Advanced Along Track Scanning Radiometer (AATSR) and PMW TMI satellite SST retrievals (Reynolds et al., 2010). The AMSR-E data was found to have a strong impact in the mid-latitudes particularly in large gradient regions e.g. the Gulf Stream, because of the improved data coverage (Reynolds et al., 2007). This impact was not seen by adding TMI SSTs, because accurate TMI SST retrievals are limited to the tropics (Reynolds et al., 2010). Similarly, Brasnett and Colan (2016) showed clear improvements when assimilating AMSR2 SSTs in the CMC SST analysis. Other studies have put efforts into characterizing the errors of IR and PMW satellite SST products (O'Carroll et al., 2008; Gentemann, 2014; Ricciardulli and Wentz, 2004) and developing bias corrections to facilitate improved merging of the products using moored and drifting buoys as reference (Castro et al., 2008).

This study presents the first systematic assessment of the impact of including PMW SST observations in an Arctic SST analysis. The PMW SSTs are from the ESA Climate Change Initiative (ESA-CCI) PMW SST CDR (Alerskans et al., 2020) based on AMSR-E and AMSR2 observations. As reference, we use the recently generated Arctic (>58°N) L4 combined SST and sea ice surface temperature (IST) CDR, which uses IR-sensors only (Nielsen-Englyst et al., 2023). Several methods of including satellite PMW SSTs have

been tested using one year of data, 2015, in order to identity the best way to combine the SST observations from IR and PMW sensors in the Arctic. One method has been selected and used for including the PMW SST data in the Arctic reanalysis for the entire ESA CCI PMW SST period (2002-2017). The paper is organized as follows. Section 2 briefly describes the IR-based L4 Arctic SST/IST CDR, the PMW SSTs, and the *in situ* observations used for validation. Section 3 provides a description of the different methods tested for including the PMW SSTs in the Arctic L4 SST/IST dataset. Section 4 presents the impact on satellite coverage, validation results and the effective spatial resolution of the different test runs during 2015. Section 5 provides the validation results and uncertainty estimates based on the combined IR and PMW Arctic L4 SST CDR (2002-2017). Section 6 discusses the results and provides suggestions for future work and finally, the conclusions are provided in Section 7.

2. Data

2.1. L4 Arctic SST/IST

This study uses the combined L4 Arctic (>58°N) SST/IST climate dataset (described in Nielsen-Englyst et al., 2023) as baseline. The long term (1982-2021) climate dataset was generated by combining IR observations from A(A)TSR, AVHRR and SLSTR (Sea and Land Surface Temperature Radiometer), and applying a statistical OI method to obtain daily gap-free fields, with a spatial resolution of 0.05° in latitude and longitude. To be included in the OI processing, a minimum quality level (QL) of 4 was required for all observations except from those from SLSTR, where a QL of 5 was required following Nielsen-Englyst et al. (2023). Each daily L4 SST/IST field comes with a theoretical uncertainty estimate, which has been derived directly from the OI method (Nielsen-Englyst et al., 2023). Validation of the OI-derived L4 SST uncertainties against drifting buoy observations showed that the OI method is capable of deriving reliable uncertainty estimates for SST (Nielsen-Englyst et al., 2023).

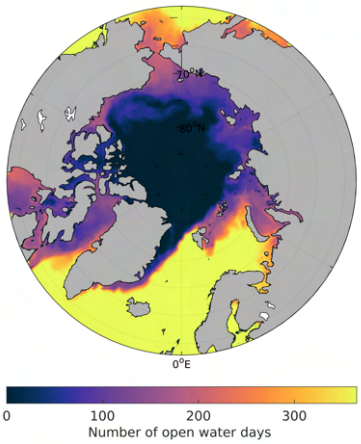


Figure 1: Example of the total number of open water days during one year (2015).

The L4 SST/IST CDR covers both sea and sea-ice surfaces in the Arctic, with open ocean being defined by sea ice concentration (SIC) $\leq$ 15%, the marginal ice zone (MIZ) as 15% <SIC$\leq$ 70%, and ice covered when SIC> 70%. Figure 1 shows an example of the number of open water days during one year (2015), while Figure 2 shows the corresponding seasonal variation in the open water fraction (dark blue + light blue). During winter and spring, only 20% of the (non-land) surface is open water, while the open water coverage increases to almost 60% during September.

Figure 2 also illustrates the relative percentage of SST and IST satellite observations that are available and missing for each day during 2015, which was chosen as a test year. About 75% of the sea ice covered grid cells are covered with observations during winter, while the IST satellite coverage drops to below 10% during summer due to an extensive summer cloud cover. On average only 21.7% of the open ocean is covered with satellite observations. During winter, only ~10% of the open ocean grid cells are covered by observations, while the open ocean satellite coverage reaches a maximum in summer of about 35%. This means that the L4 SST/IST CDR is actually based on a very limited set of satellite observations during long periods of the year. In this study, the focus is on the open ocean regions because of the variable sea ice emissivity, and the fact that thermal microwaves penetrate into the snow-cover on sea ice (Ulaby et al., 1986; Tonboe et al., 2011). The penetration in sea ice means that the IST measured by IR and PMW radiometers is not the same (Lavergne et al., 2022), which complicates a blend of the two (see Section 6).

In 2015, the L4 SST CDR is based only on IR satellite observations from AVHRR sensors (Nielsen-Englyst et al., 2023). Figure 3a shows an example of the SST coverage during one day (i.e. September 1, 2015), with large regions being unobserved. Figure 3b shows the total number of days with IR SST observations in each grid cell during 2015, and it is evident that IR observations of the surface are limited by clouds in many regions for more than half of the year.

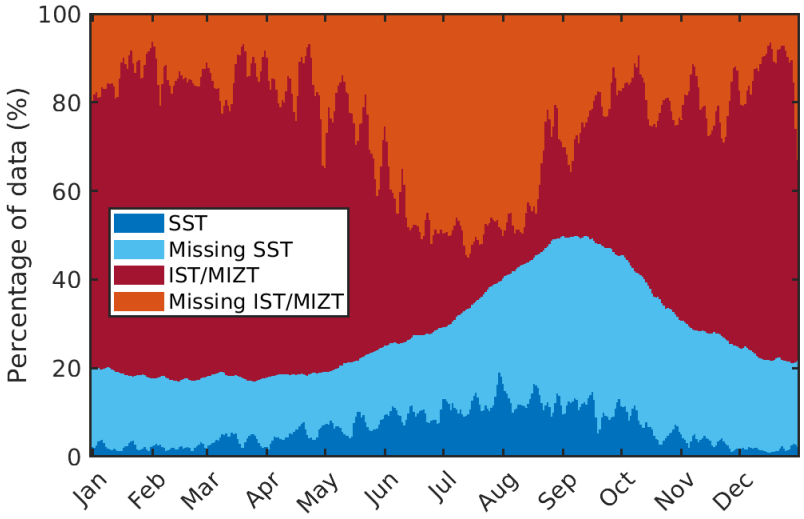


Figure 2: Example of the percentage of satellite data with SST, missing SST, IST/MIZT and missing IST/MIZT for each day during one year (2015).

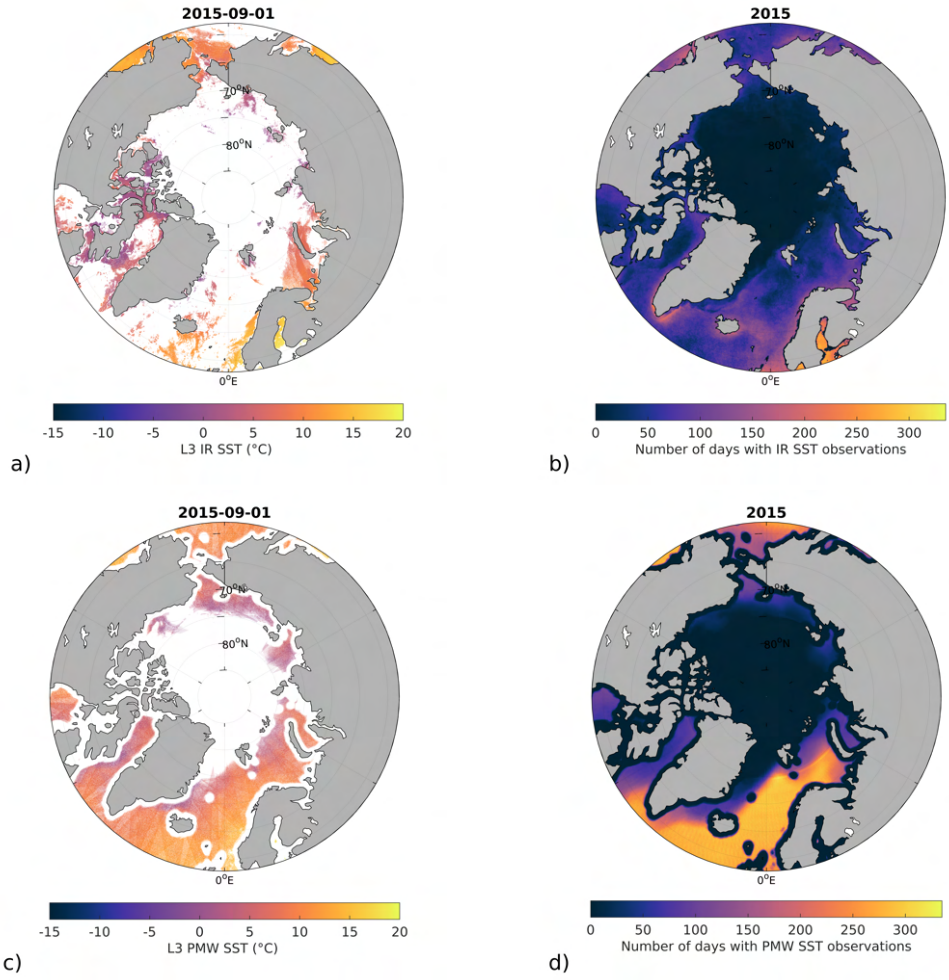


Figure 3: Examples of the L3 SST coverage during one day i.e. September 1 2015 using a) IR and c) PMW observations, and the total number of days with SST observations during 2015 from b) IR and d) PMW sensors.

2.2. PMW data

We use the Level-2 (L2) data from the PMW SST CDR described in Alerskans et al. (2020) and developed within the ESA-CCI SST project (Merchant et al., 2014). The L2 PMW SST CDR is generated using a statistical regression-based retrieval algorithm, which uses observations from AMSR-E and AMSR2 for the period June 2002–October 2017 (Alerskans et al., 2020). The PMW SSTs have been adjusted to best represent the daily mean temperature at 20 cm depth for consistency with the ESA-CCI IR SST retrievals (Embury et al., 2012; Merchant et al., 2019). The resulting L2 PMW SST CDR is provided with a 10 km grid resolution and is available from the Centre for Environmental Data Archival (CEDA) at `http://gws-access.ceda.ac.uk/public/esacci-sst/PMW2.0_release/AMSR/L2P/`.

In this study, only PMW SST observations assigned QL 3-5 are used. For these QLs, no PMW SSTs are retrieved if any sea ice is detected within ±200 km (using the ERA-Interim SIC), or if land is detected within

±100 km (Alerskans et al., 2020). This is done to exclude PMW SST retrievals which may be contaminated by land and sea ice due to the large satellite footprint at low frequencies. Figure 3c shows an example of the coverage from the PMW observations during one day, while Figure 3d shows the total number of days with PMW observations during one year (2015), when included in the OI in a similar way as the IR observations. The large band with no PMW observations along the coasts and the sea ice edge is explained by the fact that only the highest (3-5) QLs are used. In larger distances from coasts and sea ice, the PMW SST observations show superior coverage compared to the IR SST retrievals, which are limited by cloud cover. This illustrates the large potential there may be in combining IR and PMW observations for the SST mapping of the Arctic.

2.3. In situ observations

In situ observations from drifting buoys, moored buoys and Argo floats are used for validation as in Nielsen-Englyst et al. (2023). The *in situ* observations are obtained from the Hadley Centre Integrated Ocean Database v. 1.2.0.0 (HADIOD, Atkinson et al., 2014). The drifters are well represented in the Arctic open ocean region. The Argo floats show good coverage in the North Atlantic and Greenland Sea while the moorings which are concentrated in certain regions of the North Atlantic and southern Greenland Sea. The *in situ* observations are only used for validation and have not been included in the analysis nor used for bias correction of the analysis as was done using the drifters in Nielsen-Englyst et al. (2023).

3. Methods

Satellite IR and PMW radiometers measure top-of-the-atmosphere up-welling thermal emission. However, the IR and PMW satellite SST observations have very different characteristics in terms of spatial resolution and sensitivity to noise sources, which need to be taken into account in order to combine the two data sets properly (Castro et al., 2016). One difference is, as already mentioned, the almost all-weather capability of PMW observations compared to the clear-sky only capability of IR observations. The penetration depth of IR and PMW sensors also differs, with IR measuring the skin SST and PMW sensors measuring the subskin temperature (Donlon et al., 2007; Minnett and Kaiser-Weiss, 2012). However, here both the IR and PMW SST retrievals are adjusted to best represent the daily mean at the same depth (of 20 cm). The IR and PMW observations are also subject to different sources of uncertainty. The Arctic IR SST uncertainties arise mainly due to undetected clouds and insufficient representation of the atmospheric attenuation (e.g. by water vapor) (Castro et al., 2008), while large PMW SST uncertainties usually are related to rain, strong winds (>20 m s^{-1}), sun-glint, radio frequency interference (RFI), sidelobe contamination near (typically within ∼100 km) land and/or sea ice (Gentemann, 2014; Gentemann and Hilburn, 2015). Other differences are related to the different grid and spatial resolutions. The IR observations have a grid resolution of 0.05°, which is close the the actual spatial resolution of the observations. On the other hand, the L2 PMW SST observations are provided with a 10 km grid resolution, while the actual spatial resolution is in the order of 50 km due to the large satellite footprint at low frequencies. Because of the large PMW footprint, PMW SSTs are not retrieved

close to coasts and sea ice (for QL 3-5), and in these areas, the IR observations are thus the only source of SST observations (Alerskans et al., 2020). The different PMW and IR footprint sizes make it important to assess the impact on the effective spatial resolution when blending PMW and IR SSTs.

In this study, different test runs have been designed to assess the best method to include the PMW SST observations in the Arctic L4 SST/IST reanalysis taking into account the differences in the observation characteristics between the IR and PMW observations. All test runs were processed of 2015 using the DMI OI L4 processing system, which is described in detail in Nielsen-Englyst et al. (2023). The DMI OI L4 processing system takes the L2 satellite observations as input and averages these into single sensor daily Level-3 (L3) fields. The L3 fields are afterwards aggregated into L3 super-collocated (L3S) fields by calculating the noise weighted average of the available observations within 24 hours from the analysis. The IR SSTs are assumed to have uncertainties of 0.3°C (ATSR), 0.4°C (AVHRR), 0.4°C (SLSTR) (following Nielsen-Englyst et al., 2023), while the PMW SSTs are assumed to have uncertainties of 0.5°C (Alerskans et al., 2020). The L4 fields are generated using the same OI statistical parameters as in Nielsen-Englyst et al. (2023) for all test runs. Table 1 provides an overview of the different test runs processed for 2015, while a description of each test run is provided below.

3.1. IR only (IR)

Test run IR is identical to the Arctic SST/IST CDR described in Nielsen-Englyst et al. (2023) except that the temporally and spatially constant bias correction (of +0.16°C) against drifters has been excluded here. This IR-only test run is used as reference for the following test runs, which all include PMW observations.

3.2. PMW only (PMW)

Test run PMW excludes all IR SST data and is based only on the PMW SST data. In this case, the PMW observations have been included using the same approach as was done for the IR observations in the first test run. When aggregating to single sensor L3 fields, this leaves gaps in approximately every second grid cell of the L3 PMW field, since the L2 PMW observations are provided at a 10 km grid (in contrast to the L4/L3 0.05 degree grid), and a given satellite observation is only included once in the current DMI OI L4 processing scheme.

3.3. IR and PMW (IR_PMW)

Test run IR_PMW is the first attempt to combine IR and PMW SST observations in the DMI OI L4 processing scheme. The IR and PMW observations have been included as in the first two test runs, leaving gaps in approximately every second grid cell of the L3 PMW field. IR_PMW will be the baseline of the following test runs, which all combine the IR and PMW observations using slightly different approaches.

3.4. IR and averaged PMW (IR_PMW_AVG)

IR_PMW_AVG is tested to see the effect of a more complete PMW SST field. The setup is similar to IR_PMW, but differs in the way the L3 PMW field is aggregated. Instead of only allowing each satellite observation to be included once, each L2 PMW satellite observation is included in all L3 grid cells within a radius of 5 km. This allows observations in all grid cells covered by the PMW footprints of the L2 observations and increases the number of L2 PMW observations available in each L3 grid cell in general. Each L3 PMW field is thus an average of more L2 PMW observations.

3.5. IR and subsampled PMW (IR_PMW_SUB)

Two problems arise in the previous test runs including PMW SST. Firstly, the L3 PMW SST grid cells are substantially over-sampled (with the L2 grid resolution being much higher than the PMW footprint) and each L3 PMW SST grid cell is thereby not independent but noise-correlated with its neighboring grid cells (within the satellite footprint of ∼50 km). Moreover, there is a risk of PMW flooding in the OI scheme, and thus, minimizing the impact of the much less frequent IR SSTs. Test run IR_PMW_SUB investigates the effect of sub-sampling the L2 PMW SST observations. In IR_PMW_SUB the L2 PMW SST observations have been sub-sampled with a step of 4 in the longitude and latitude of the L2 grid before averaging to the L3 PMW fields. This minimizes the dependence between the L3 PMW grid cells, and reduces the amount of L3 PMW grid cells and the risk of L3 PMW SST flooding in the OI scheme.

*3.6. IR and averaged, subsampled and IR-adjusted PMW (IR*_PMW)*

The rationale behind IR*_PMW is to use the averaged and complete L3 PMW fields from IR_PMW_AVG but applying a mask (based on the IR_PMW_SUB L3 PMW fields) to minimize the dependence between the PMW observations and the risk of L3 PMW SST flooding in the OI scheme. Therefore, only those L3 PMW grid cells, which are included in IR_PMW_SUB are kept and included in the L3S and L4 generation. This means that the number of days with observations is the same as in IR_PMW_SUB.

IR*_PMW also includes an inter-sensor bias correction, which has been implemented to correct for systematic biases in the IR and PMW SST data sets relative to each other. This is an important step towards a combined IR and PMW dataset to avoid introducing biases when switching from one sensor to the other/or both. Here, the L3 PMW SST data have been corrected using the L3 IR SST observations as references (the "*" in IR*_PMW denotes the reference field). The bias correction is described in more detail in Section 3.8.

3.7. PMW-adjusted IR and averaged, subsampled PMW (IR_PMW)*

IR_PMW* is similar to IR*_PMW, but instead of correcting the PMW SST observations against IR SST, the IR SST observations have been corrected against the PMW SST observations.

Table 1: Overview of the different test runs. The IR headline refers to IR being included as in Nielsen-Englyst et al. (2023), while the PMW headline refers to PMW being included in a similar way as IR. PMW-AVG specifies that a complete and averaged PMW field is included. In contrast, PMW-SUB corresponds to a sub-sampled PMW field being included. REF indicates if an inter-sensor bias correction has been used and specifies the applied reference sensor. See Section 3.1-3.7 for a detailed description of each test run.

Test run	IR	PMW	PMW-AVG	PMW-SUB	REF
IR	X				
PMW		X			
IR_PMW	X	X			
IR_PMW_AVG	X		X		
IR_PMW_SUB	X			X	
IR*_PMW	X		X	X	IR
IR_PMW*	X		X	X	PMW

3.8. Inter-sensor bias correction

In test run IR*_PMW and IR_PMW*, an inter-sensor bias correction has been implemented to correct for systematic biases in the IR and PMW SST data sets relative to each other. The inter-sensor bias correction method was developed in Høyer et al. (2014), where it was demonstrated to be very efficient in removing biases throughout the year. The bias correction is assumed to be a smooth field, mainly accounting for slowly varying systematic tendencies of each sensor retrieval.

Using IR as reference, the following approach is used to estimate the PMW SST bias correction (subtracted from the L3 PMW observations in IR*_PMW). The IR SST reference field has been averaged onto coarser grid (0.25°) and aggregated using a temporal window of 7 days. A difference field is calculated for each day by subtracting the IR reference field from the corresponding coarse resolution aggregated PMW sensor field. This coarse resolution difference field is afterwards interpolated to high resolution (0.05°) and smoothed over 500 km to reduce small scale noise. The resulting high resolution difference field has been used to bias-correct the L3 PMW SST fields. The bias correction has been subtracted from the L3 PMW observations in IR*_PMW. In IR_PMW*, the bias correction is calculated in a similar way using PMW as reference field, which corresponds to adding the PMW bias correction to the L3 IR SST observations. Figure 4 shows the seasonal spatial variation of the coarse grid, smooth and extrapolated PMW SST bias correction field, while Figure 5 shows the daily mean PMW SST bias correction throughout the year 2015. In all seasons, the bias correction is smallest in the North Atlantic. The average bias correction is 0.31°C, with almost no correction during winter and a maximum during summer of about 0.6°C.

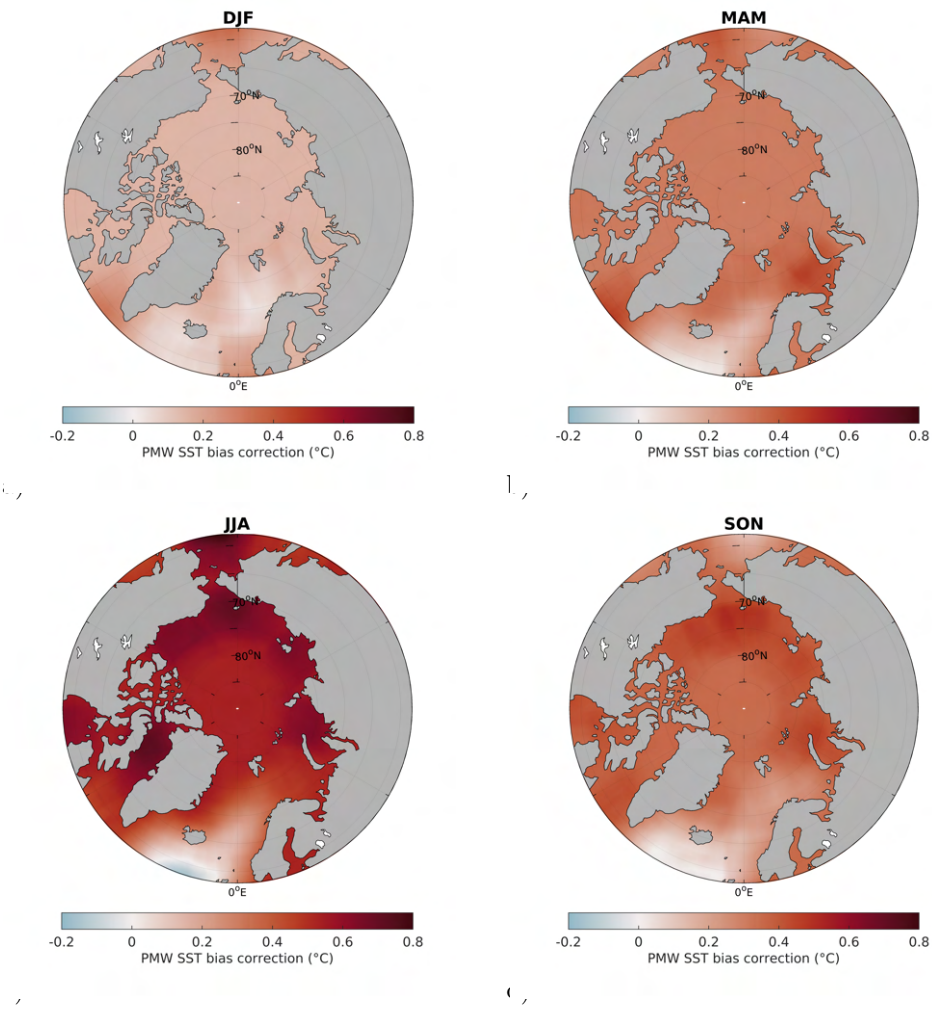


Figure 4: Spatial variation of the mean coarse grid (and extrapolated) PMW bias correction field for the months a) December-February, b) March-May, c) June-August, and d) September-November 2015. The PMW bias correction field is subtracted from the PMW SST observations in IR*_PMW (and added to the IR SST observations in IR_PMW*).

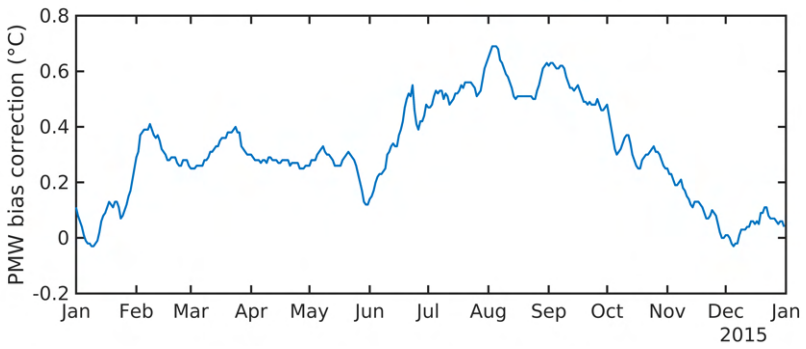


Figure 5: Seasonal variation of the daily mean PMW bias correction field subtracted from the L3 PMW SST observations in IR*_PMW (and added to the IR SST observations in IR_PMW*) during 2015.

4. Test run results

This section investigates the impact on the satellite coverage, validation and the effective spatial resolution by including PMW SST observations in the Arctic L4 SST/IST dataset using one year (2015) of the data and the setup from the test runs described in Section 3.

4.1. Satellite Coverage

The number of daily satellite observations included in the OI processing scheme varies among the test runs. Figure 6 shows the change in the number of daily SST observations compared to the IR reference test run for the individual test runs during 2015. In the case where only PMW observations are used (Figure 6a), no SST observations are available near coasts and sea ice, but more SST observations are available everywhere else compared to test run IR. Combining PMW and IR observations as in IR_PMW (Figure 6b), most regions (away from coasts and sea ice) experience more SST observations compared to test run IR. Performing the L3 PMW averaging by including more L2 PMW observations as in IR_PMW_AVG allows many more days with SST observations compared to the other test runs (Figure 6c). Figure 6d shows the remaining number of days with SST observations after the L3 PMW fields have been sub-sampled to only include the available L3 PMW grid cells from IR_PMW_SUB. The average daily SST coverage for the different test runs are: 21.7% (IR), 33.9% (PMW), 55.6% (IR_PMW), 73.8% (IR_PMW_AVG) and 25.2% (IR_PMW_SUB/IR*_PMW/IR_PMW*) during 2015.

4.2. Validation

The different test runs have been validated against drifting buoy SST observations, which provide the best representation of the Arctic. Table 2 shows the validation results of the L4 SST fields, the aggregated L3 super-collocated (L3S) SST fields, and the single sensor L3 PMW SST fields during 2015. Matchups with drifter SST or L4/L3S/L3 SST below -1.8°C and matchups with L4/L3S/L3 – drifter SST differences deviating more than three times the standard deviation from the mean L4/L3S/L3 – drifter SST difference (referred to as a 3-sigma filter) have been excluded from the validation statistics. The filters are applied to exclude erroneous *in situ* observations and to provide more representative validation statistics (without dominance from outliers). Each filter removes about 2% of the L4 matchups. The varying satellite coverage (as seen in Section 4.1) is reflected in the large variations in the number of L3S and L3 PMW matchups in Table 2.

For test run IR, a significant increase in standard deviation is seen from the L3S to L4 field, which indicates that it is difficult for the OI processing to provide accurate SSTs in the poorly IR observed regions. The L4 IR mean SST difference of -0.14°C is close to the difference (of -0.16°C), which was documented and corrected for in the post-processing of the long-term climate dataset in Nielsen-Englyst et al. (2023). As seen in Figure 5, the PMW observations are generally warmer than the IR observations, and the resulting mean difference against drifters is reduced for all test runs where PMW observations are included in the L4

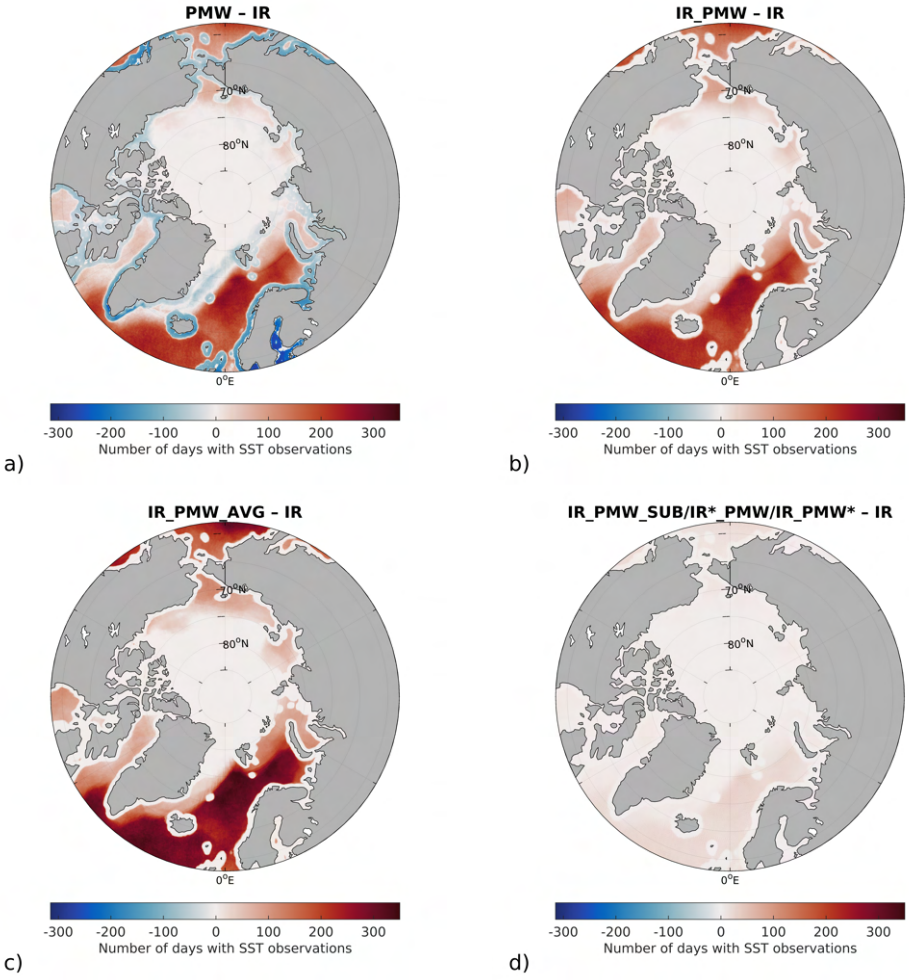


Figure 6: Differences in the total number of daily SST observations compared to test run IR for a) PMW, b) IR_PMW, c) IR_PMW_AVG and d) IR_PMW_SUB/IR*_PMW/IR_PMW* during 2015.

generation (and not referenced against the IR observations as in IR*_PMW). All combinations of IR and PMW observations show reduced L4 standard deviations compared to only using either IR or PMW SST observations.

Table 2 also shows that it is possible to reduce the standard deviations of the L3 PMW observations by including more (of the surrounding) L2 PMW observations in the aggregation of the L3 PMW field. This also allows more L3 grid cells to be assigned with a PMW SST and results in more L3 PMW matchups compared to PMW/IR_PMW. The opposite is the case when the L2 PMW fields are sub-sampled. Here the number of grid cells with a L3 PMW SST is reduced, which results in fewer matchups than for PMW/IR_PMW, and less L2 PMW observations available for averaging the L3 PMW SST fields, resulting in a larger L3 PMW SST standard deviation. If the L3 PMW SST validation subset from IR_PMW_SUB (10,256 matchups) is used for the IR_PMW_AVG L3 PMW SST validation, the standard deviation and bias reduce to 0.44°C and -0.01°C (as also seen for IR*_PMW/IR_PMW*). This clearly shows that including more L2 PMW SST observations

Table 2: Overall performance of the different test runs against *in situ* observations from drifting buoys during 2015. The table shows the mean difference (MD), standard deviation (STD), root mean squared difference (RMSE), and the number of observations (Nobs) for the L4, L3S and L3 PMW SST fields, respectively. A 3-sigma filter has been applied to remove outliers.

	L4				L3S				L3 PMW			
Test run	MD	STD	RMSE	Nobs	MD	STD	RMSE	Nobs	MD	STD	RMSE	Nobs
IR	-0.14	0.64	0.65	193,798	-0.10	0.47	0.48	41,312	-	-	-	-
PMW	-0.03	0.74	0.74	192,140	-0.01	0.57	0.57	97,648	-0.01	0.57	0.57	97,648
IR_PMW	-0.04	0.56	0.57	194,434	-0.03	0.54	0.54	119,014	-0.01	0.57	0.57	97,649
IR_PMW_AVG	-0.03	0.56	0.56	194,457	-0.02	0.47	0.47	153,197	-0.00	0.48	0.48	140,401
IR_PMW_SUB	-0.06	0.56	0.56	194,188	-0.08	0.51	0.51	49,506	0.00	0.63	0.63	10,256
IR*_PMW	-0.18	0.56	0.59	194,159	-0.12	0.47	0.48	49,450	-0.01	0.44	0.44	10,219
IR_PMW*	0.05	0.56	0.56	194,532	0.12	0.48	0.49	49,456	-0.01	0.44	0.44	10,219

in the L3 PMW SST aggregation improves the L3 PMW SST performance substantially.

The improved L3 PMW and L3S standard deviations of IR*_PMW/IR_PMW* (and IR_PMW_AVG) are, however, not reflected in the L4 standard deviations, which do not vary among the test runs including both PMW and IR SST observations. If only the L3S matchups of IR_PMW_AVG (153,197 matchups) are considered in the L4 validation, all combined IR and PMW test runs provide equal same standard deviations of 0.46°C, while test run IR and test run PMW provide standard deviations of 0.53°C and 0.55°C, respectively (not shown). This indicates that if both IR and PMW observations are included, the OI L4 processing is able to provide accurate SSTs for those 153,197 matchups despite the differences in the L3S and L3 PMW fields.

Figures 7a and 7b show the geographical distribution of the standard deviations during 2015 when compared to drifting buoy SST using IR and IR*_PMW SST, respectively, with the latter being similar to IR_PMW* (not shown). In both cases, the largest standard deviations are found along the coasts and in the seasonal ice covered waters. The few or none satellite observations from the IR and PMW sensors in these regions (see Figure 3) make it difficult for the OI method to provide accurate SSTs. Figures 7d and 7e show the mean differences in 2015, when compared to drifting buoy SST for IR*_PMW and IR_PMW*, respectively. Large differences are seen in the mean SST differences for IR*_PMW and IR_PMW*, which are corrected against IR and PMW observations, respectively. IR*_PMW is cold compared to drifters, with an increasing magnitude towards the sea ice edge. In contrast, IR_PMW* shows varying mean differences, with an average difference around zero, but also regions with SSTs warmer than drifters (e.g. the Beaufort and Chukchi Sea). Figure 7c shows the distribution of the L4 matchups (with drifters) available for validation of IR*_PMW during 2015, which is similar to the matchup distribution of the other test runs (not shown). The drifter matchups show good coverage of the open water regions with most matchups in the North Atlantic and southern Greenland Sea. Figure 7f shows the distribution of those IR*_PMW L4 matchups that are

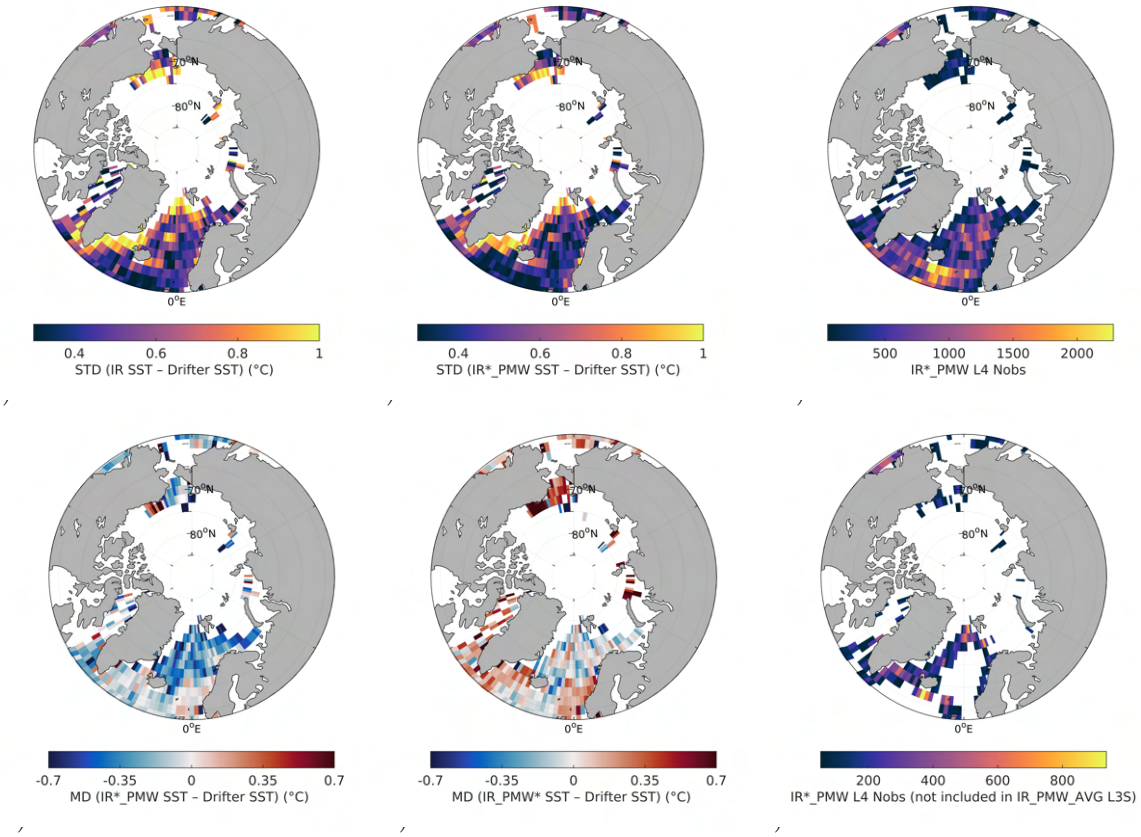
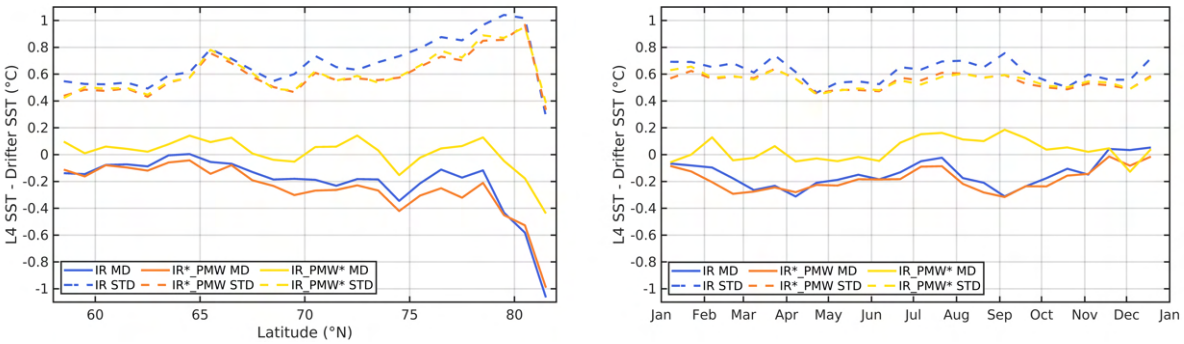


Figure 7: The standard deviation (STD) against drifter SST for a) IR and b) IR*_PMW (similar to IR_PMW*) and the mean differences (MD) between d) IR*_PMW, e) IR_PMW* and drifters. The last column shows the number of L4 matchups (Nobs) for c) IR*_PMW and f) IR*_PMW which are not included in the IR_PMW_AVG L3S matchups. The statistics are calculated for each 2x2 degree grid having more than 50 members during 2015.

not part of the IR_PMW_AVG L3S matchups i.e. those matchups that have no IR or PMW satellite SST observations within the corresponding grid cells. These are concentrated along the coasts and sea ice edge. In these regions, increased standard deviations are observed for both IR and IR*_PMW (Figures 7a and 7b) as well as for the other test runs (not shown). This will be discussed in greater detail in Section 6.

The performances of IR*_PMW and IR_PMW* and IR are shown as a function of latitude (Figure 8a) and time (Figure 8b) for 2015. Generally, IR*_PMW/IR_PMW* show smaller standard deviations, with the largest improvement between 68°N and 80°N compared to test run IR. Test run IR and IR*_PMW show a gradual increase in the absolute mean difference with latitude. This is in contrast to IR_PMW*, which shows a mean difference centered around zero except from northwards of about 80°N, where it is colder than drifters, but to a smaller degree than the IR and IR*_PMW. IR_PMW* also shows a smaller and more stable mean difference as a function of time compared to test run IR and IR*_PMW (Figure 8b). The dynamic bias correction of IR against PMW is thus able to significantly reduce the latitudinal and seasonally discrepancy observed in the IR SSTs.



a) b)

Figure 8: Mean differences (MD) and standard deviations (STD) against drifter SST during 2015 for test run IR, IR*_PMW and IR_PMW* as a function of a) latitude (°N) and b) time, using bin sizes of 1° and 15 days, respectively, and a requirement of minimum 30 matchups per bin.

4.3. Spectral Analysis

Due to the differences in IR and PMW footprint sizes and coverage, it is important to assess the impact on the effective spatial resolution on the L4 product, when ingesting the PMW SSTs into the IR-based L4 SST/IST analysis. Therefore, a spectral analysis has been performed in the two sub-domains shown in Figure 9 for the different L4 SST test runs. Comparisons of the spectral power will indicate added benefit in resolving SST signals or degradation of the effective spatial resolution when including the coarser PMW observations.

The estimation of the power spectrum is performed using the standard FFT method in the zonal direction (Thomson and Emery, 2014) and only open water points are included. The power spectrum per wave number is computed for each latitude band and averaged into bins for the full year for each test run. Figure 10 shows the zonal spectra for each test run (except from IR_PMW* for which there was no visible difference from IR*_PMW) during 2015 for sub-domains D1 (a) and D2 (b). The theoretical -2 and -5/3 (-1.6) curves for the expected decrease of spectral power in the meso- to sub-mesoscale are also included for reference (Vazquez-Cuervo et al., 2022; Castro et al., 2017). Note that the smallest scales are resolved only in the northernmost region of the D2 domain, as the distance between meridionals becomes smaller towards the high latitudes. This means that the results for the smallest scales are based on less data than for the larger scales. Therefore, to avoid noisy signals, only power spectra with wavelengths larger than 6 km are shown.

Figure 10 shows similar power spectra at scales larger than ∼60 km for D1 and scales larger than ∼20 km for D2. The "bump" in spectral power occurring for the PMW and IR_PMW at approximately 12.5 km in D1 (10.5 km in D2) is assumed to be associated with energy contribution from the smaller scales, probably due to the mismatch between the resolution of the PMW data and the L4 grid spacing. PMW and IR_PMW generally have more energy at small scales compared to the other test runs and their spectral slopes are gentler, i.e. slower decrease in power as wavenumbers increase. This is likely explained by the fact that both of these test runs include PMW SST observations (approximately for every tenth kilometer) in regions that

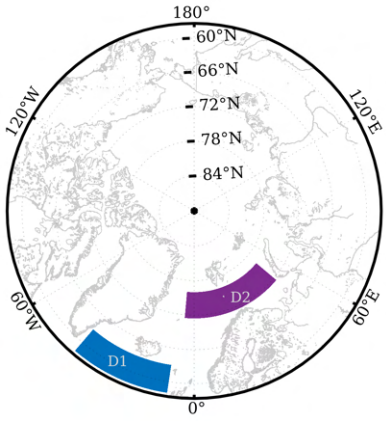


Figure 9: Domain 1 (D1, blue) and 2 (D2, magenta) used for the calculation of the spectral power.

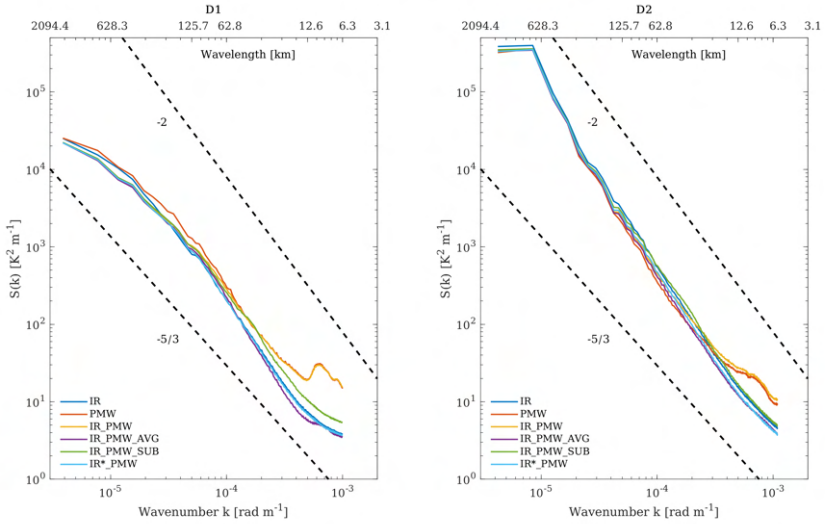


Figure 10: Zonal spectra for the subdomains a) D1 and b) D2 for the different test runs during 2015.

were otherwise under-sampled by IR. This is in contrast to IR_PMW_SUB and IR*_PMW, where the PMW SSTs have been subsampled and the L4 output relies more on the interpolation capability in these regions resulting in a more smooth L4 field. This is manifested as an overall lower spectral power level and a rapid decrease in the power as wavenumbers increase, i.e. steeper spectral slopes. In IR_PMW_AVG, the averaging performed in the aggregation of the L3 PMW SST field also results in a more smooth L4 SST field. Except from PMW and IR_PMW, the test runs are closely aligned with the IR spectra, which is ideal in terms of long-term consistency of a merged IR and PMW SST product.

4.4. Selection for the CDR

Different ways of including PMW SST observations in the L4 Arctic SST/IST reanalysis have been assessed. For all test runs, the inclusion of PMW SSTs reduces the L4 standard deviations against drifters

compared to only using IR SST (and only using PMW SST).

The best L3 PMW validation result is obtained by allowing each L2 PMW observation to be included in all L3 grid cells in a radius of 5 km from the L2 PMW observation. This reduces the noise and provides more robust L3 PMW SST estimates (see IR_PMW_AVG validation in Section 4.2). To reduce the dependence between the L3 PMW observations and the risk of PMW flooding in the OI scheme, this L3 PMW field (from IR_PMW_AVG) has been sub-sampled by only including those grid cells, which were also included in IR_PMW_SUB. This is done in both IR*_PMW and IR_PMW* that differ only in the inter-sensor bias correction, which adjusts the PMW SST to IR SST in IR*_PMW and IR SST to PMW SST in IR_PMW*. The smallest discrepancy against drifters is seen in the case where the IR SSTs are adjusted to PMW SSTs, since the IR SSTs are generally too cold compared to drifters. However, the PMW observations span a much shorter time scale than IR, which limits the extent to which the bias correction against PMW SSTs can be applied. For climate analyses, the PMW SSTs could instead be adjusted to IR SSTs to ensure consistency and avoid jumps when introducing the PMW SSTs.

The spectral analysis revealed similar zonal spectra for all the test runs (and both domains) except from PMW and IR_PMW, which both have more energy in the small-scales compared to the other test runs. This is explained by the fact that both of these test runs include PMW SST observation (approximately for every tenth kilometer) in regions that were otherwise under-sampled by IR. The other test runs are closely aligned with the IR spectra, which is ideal in terms of long-term consistency of a merged IR and PMW SST product.

Based on above analyses, we decided to use the setup from IR*_PMW to generate a blended IR and PMW Arctic L4 SST CDR for the ESA CCI PMW SST CDR period extending from June 2002 to October 2017.

5. CDR Results

This section provides the validation and uncertainty results of the blended Arctic IR and PMW L4 SST CDR for the ESA CCI PMW SST period (2002-2017), hereafter referred to as IR*_PMW (after the test run, which it is based on). The results of the full IR*_PMW SST CDR is compared to the IR-based Arctic SST/IST CDR (described in Nielsen-Englyst et al., 2023) for this period, which will be referred to as IR hereafter. In the PMW sensor gap between AMSR-E and AMSR2 (from October 4, 2011 to July 4, 2012) the IR*_PMW SST CDR is based only on IR observations, and is thus identical to the IR SST CDR for this period.

5.1. Validation

The long-term IR and IR*_PMW runs have been validated against drifting buoys, Argo floats and moorings for the years 2002-2017. The validation statistics are summarized in Table 3 for the L3S and L4 SST products. For both drifters and Argo floats there is a substantial improvement in the L4 standard deviations, which reduce from 0.54°C to 0.47°C and from 0.47°C to 0.41°C, respectively, by including PMW SST observations.

Table 3: Overall performance of IR and IR*_PMW SSTs against drifting buoys, Argo floats and moorings for the years 2002-2017. The table shows the mean difference (MD), standard deviation (STD), root mean squared difference (RMSE), and the number of observations (Nobs). A 3-sigma filter has been applied to remove outliers.

	L4 IR				L4 IR*_PMW				L3S IR				L3S IR*_PMW			
	MD	STD	RMSE	Nobs	MD	STD	RMSE	Nobs	MD	STD	RMSE	Nobs	MD	STD	RMSE	Nobs
Drifters	-0.15	0.54	0.56	1.70e06	-0.15	0.47	0.49	1.70e06	-0.09	0.42	0.43	4.62e05	-0.11	0.42	0.43	5.31e05
Moorings	-0.11	0.55	0.56	39,935	-0.12	0.54	0.56	39,935	-0.05	0.51	0.51	16,649	-0.06	0.52	0.52	17,304
Argo	-0.07	0.47	0.48	19,847	-0.06	0.41	0.41	19,936	0.01	0.38	0.38	5,758	-0.00	0.38	0.38	6,578

Moorings show very limited variation in the L4 performances and this is explained by the fact that the moorings are located only in specific parts of the North Atlantic and the southern Greenland Sea.

The L3S SST validation revealed limited variations in the statistics for IR and IR*_PMW (for all *in situ* types). Argo floats and drifters increase their number of L3S matchups with about 15% when PMW observations are added, while moorings only have 4% more matchups when including PMW observations. These additional matchups are not available for the IR product, and if these matchups are excluded from the L3S IR*_PMW validation, the statistical parameters (provided in Table 3) of L3S IR and L3S IR*_PMW are the exact same (for all *in situ* types).

Figure 11a shows the geographical distribution of standard deviation against drifters for L4 IR*_PMW, while Figure 11b shows the differences in standard deviation between L4 IR*_PMW and L4 IR for drifters during the period 2002-2017. The largest standard deviations are generally found along the coasts and sea ice edge, where few IR and PMW SSTs are available. Compared to the IR run, the inclusion of PMW SSTs provides reduced standard deviations in almost all regions with few exceptions which are likely related to sea ice contamination and residual RFI (Gentemann and Hilburn, 2015).

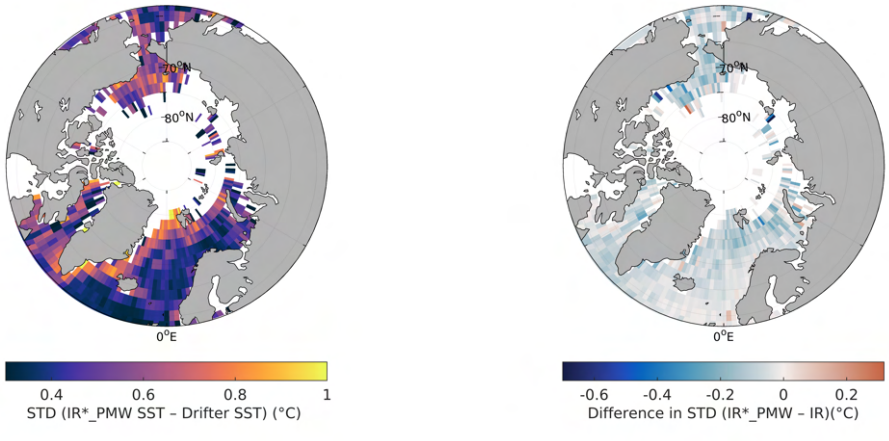


Figure 11: a) Standard deviation (STD) against drifter SST for L4 IR*_PMW and b) the difference in standard deviation between L4 IR*_PMW and IR for drifters, during the period 2002-2017. The statistics are calculated for each 2x2 degree grid having more than 50 members.

Figure 12a shows the annual mean and standard deviation of the L4 SST minus drifter SST for the period 2002-2017. In general, higher standard deviations are seen when no PMW SST observations are included. For both IR and IR*_PMW, the standard deviations are largest in the last part of the period (2011-2017), which is characterized by fewer IR SST observations, since the (A)ATSR and SLSTR are not available (Nielsen-Englyst et al., 2023; Merchant et al., 2019). For this period, we also notice a larger reduction in standard deviation when including PMW SSTs. At the same time, the mean difference differs slightly among the two runs, which was unexpected as the PMW SSTs have been adjusted to the IR SSTs. To investigate this effect in greater detail, Figure 12b shows the validation statistics as a function of the number of days since an IR SST observation was last available in that particular grid point. For both IR and IR*_PMW, the standard deviations against drifters increase almost linearly with the number of days since an IR SST observation was last available. It is also seen that the reduction in standard deviation from including PMW SSTs increases with the number of days since the last IR observation (i.e. the largest improvements are seen in IR data sparse regions).

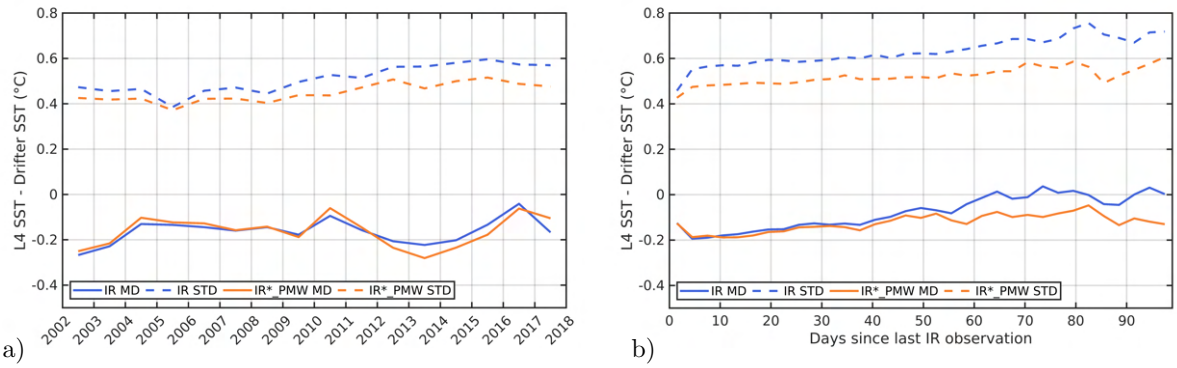


Figure 12: Mean differences (MD) and standard deviations (STD) against drifter SST for IR and IR*_PMW CDR SSTs as a function of a) time and b) days since last IR observation, using bin sizes of one year and 3 days, respectively, and a requirement of minimum 30 matchups per bin.

For those matchups (2%) where IR observations are lacking in more than 60 days, the mean difference (against drifters) differs among the two runs, as the IR SSTs get a warm bias. This is a phenomenon occurring mainly from October through December, when long periods without observations result in the use of first guess (previous day's analysis) many days in a row resulting in the use of erroneously warmer (summer+fall) SSTs. In principle, the OI should account for this by taking neighbor observations into account but what we see is the residual effect from large areas with missing observations for a long time. This is of course undesirable and will be discussed further in Section 6. The inclusion of the more frequent PMW SSTs reduces this effect as seen in Figure 12b. This is another advantage of including PMW observations with frequent updates from past, current and future PMW radiometers.

5.2. L4 Uncertainty

As described in Nielsen-Englyst et al. (2023) each daily L4 SST/IST is assigned with an uncertainty estimate, which is a direct output of the OI method that depends on the data availability, the proximity of the observations and the uncertainty of the observations and the background field. Figure 13a shows the geographical mean L4 SST uncertainty for IR*_PMW for the period 2002-2017. The largest uncertainties are found along the coasts and in the seasonal ice covered regions with maximum uncertainties (of ∼2°C) north of Svalbard. This is in agreement with the increased standard deviations observed against drifters (Figure 11a). Figure 13b shows the reduction in L4 SST uncertainty when including PMW SST observations, with the largest reductions in the Barents Sea, Greenland Sea and the Labrador Sea.

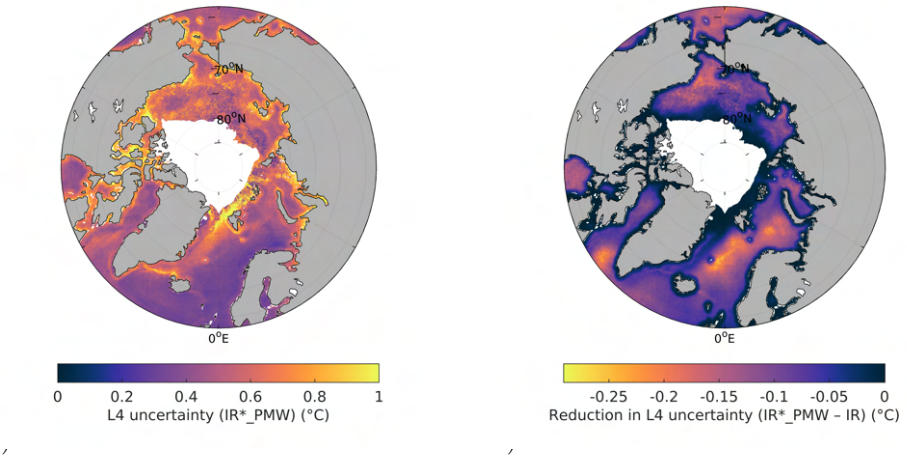


Figure 13: Spatial mean a) L4 SST uncertainty of IR*_PMW and b) L4 SST uncertainty difference between IR*_PMW and IR calculated for grid cells with more than 90 days of open water in the period 2002-2017.

Figure 14 shows the yearly mean L4 SST uncertainty during 2002-2017 for IR and IR*_PMW, respectively. In both cases, the L4 SST uncertainty increases after 2011 as a result of fewer IR satellite observations which is in agreement with the increased standard deviations observed in the end of the record in Figure 12a. At all times, the IR*_PMW provides lower L4 SST uncertainties than the IR run, with the largest improvements in the end of the period, which is also in agreement with Figure 12a. Overall, IR*_PMW provides a reduction in the L4 SST uncertainty of 0.08°C compared to IR, which is comparable to the reduction in standard deviation observed against drifters of 0.07°C (see Table 3). Nielsen-Englyst et al. (2023) also showed good agreement between L4 SST uncertainties and observed uncertainties using drifter SSTs as reference for the full IR record (1982-2021). The reduced L4 SST uncertainty estimates obtained when including PMW SSTs are thus very promising results.

6. Discussion

The inclusion of PMW SST observations provides substantial reductions in the L4 standard deviations compared to only using IR (and only using PMW) SST observations. The improved performance is mainly

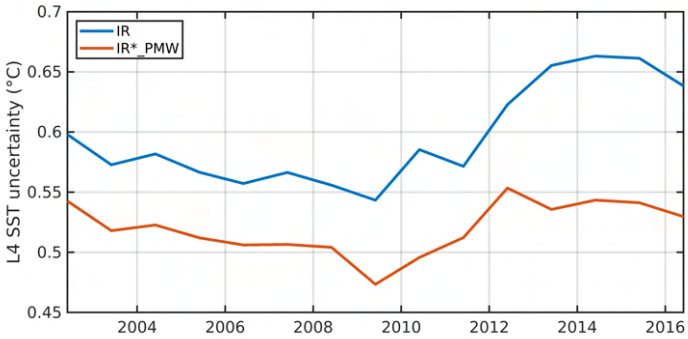


Figure 14: Yearly mean L4 SST uncertainty during 2002-2017 for IR and IR*_PMW.

linked to the superior coverage from PMW observations in the Arctic (Figure 6). However, the independent and highly complementary uncertainty characteristics of PMW and IR observations (see Section 3) are likely also part of the explanation for the observed improvements. The complementary uncertainty characteristics reduce the risk of systematic biases (e.g. as seen for the IR) in a merged product. This is supported by all combinations of IR and PMW SSTs (without introducing an inter-sensor bias correction) having reduced L4 biases compared to the IR test run (Table 2). The colder temperatures observed in the Arctic IR SSTs (compared to drifters) is a well known problem, which is also seen in other SST analyses e.g. OSTIA (Fiedler et al., 2019). To reduce the risk of introducing biases when switching from one sensor to the other/or both, an inter-sensor bias correction is necessary. The smallest L4 mean difference against drifters is seen in the case where the IR SSTs are adjusted to PMW SSTs. However, to be consistent with time periods when no Arctic PMW SSTs are available, a bias correction adjusting the PMW SSTs to IR SST has been implemented here. This introduces a bias in the L4 SST as seen in Table 2. Since, the continuity of PMW imagers have been sustained for the future with AMSR3 and CIMR in the pipeline, future work should focus on using the PMW SST observations to adjust the cold IR SSTs in the Arctic (also extended to periods when no PMW SSTs are available).

In general, the largest standard deviations (Figure 11a) and theoretical uncertainties (Figure 13a) are found along the coasts and sea ice edge for both IR and IR*_PMW as well as for the other test runs (not shown). In these regions, few or no SST observations from the IR and PMW satellite sensors (Figure 3) make it challenging to provide accurate SST estimates. In addition, some of these regions have very large gradients in the surface temperature e.g. along the ice edge (Carvalho and Wang, 2020). The improved spatial resolution (of 15 km) from CIMR will enable PMW SST retrievals much closer to coasts and sea ice in the future. The higher accuracy of 0.3°C (in cold waters) from CIMR will also improve the Arctic SST estimates even further.

The L4 IR standard deviations during 2015 were higher (0.64°C for drifters) compared to those from the full period (0.54°C for drifters). This is mainly explained by the fact that only the AVHRR sensors were available during the period 2011-2017 (Nielsen-Englyst et al., 2023; Merchant et al., 2019). The year, 2015,

was chosen since we expected the largest impact from PMW observations to occur when few IR observations are available, thus making it more feasible to assess the impact and variations among the different test runs. As evident from Figure 12a, the largest reduction in standard deviation by including PMW SSTs was indeed observed during the period 2011-2017. In relation to that, Figure 12b also showed that the reduction in standard deviation is largest for regions rarely observed by IR sensors.

For both IR and IR*_PMW, the reduced satellite coverage during 2011-2017 resulted in increased standard deviations (Figure 12a) and increased L4 uncertainties (Figure 14). In contrast, the overall long-term mean differences do not show any dependence on the satellite coverage in Figure 12a. However, Figure 12b revealed that in fact, there is a coverage dependence when binning the mean differences as a function of the days since the last IR observation in that particular grid point. Both IR and IR*_PMW get warmer as the number of consecutive days without IR observations increases. The issue arises during fall and early winter in unobserved regions and when the surrounding observations are too far away to have any weight in the OI scheme. In that case, the first guess SST (i.e. previous day's analysis in this case) is used many days in succession resulting in the use of erroneous warmer (summer and fall) SSTs.

This is clearly undesirable, in particular in the context of climate monitoring. Most of the existing and widely used L4 global SST analyses also use OI techniques (Castro et al., 2016; Vazquez-Cuervo et al., 2022) and the previous day's analysis as first guess (e.g. NOAA OI SST and REMSS MW-IR SST), and it is thus likely that they also are affected by this artefact. Figure 12b showed that the residual effect was most pronounced when only including IR observations, while the inclusion of the PMW observations reduced the effect. This is another argument of including the more frequent available PMW observations from past, present and future missions. However, despite the inclusion of PMW SSTs, the problem will likely persist in regions very near coasts and sea ice, which cannot be resolved by the current IR and PMW sensors. This should be addressed in future updates of the Arctic reanalysis as well as in the development of new regional and global reanalyses e.g. by applying a seasonal variation to the first guess field.

Future work should also be focused towards improving the validation close to sea ice e.g. by using Saildrone observations (Gentemann et al., 2020; Vazquez-Cuervo et al., 2022; Jia et al., 2022) and potentially improving the surface temperature estimates close to both coasts and the sea ice e.g. in similar ways as done for salinity retrievals in Meissner and Manaster (2021) and Olmedo et al. (2017). Moreover, the capability of using PMW ISTs to supplement the IR ISTs should also be investigated in the future. An increasing number of PMW derived IST products have become available at daily temporal resolution based on the vertically polarized 6.9-GHz channel AMSR-data (e.g. Le Traon et al., 2015; Comiso et al., 2003; Kilic et al., 2019). These approximately represent the physical temperature of the snow/ice interface (Tonboe et al., 2011; Tonboe, 2010; Ulaby et al., 1986), and relating these to IR IST (i.e. the skin surface temperature) is a challenging task considering the large temperature gradients in the snow during winter (e.g. Comiso et al., 2003, 1989).

Finally, future work should aim at updating the OI scheme to include the L2 PMW SST uncertainty

estimates provided with the individual PMW SST retrievals instead of the spatial and temporal constant of 0.5°C. Alerskans et al. (2020) showed good validation results of the L2 PMW SST uncertainties and using these may lead to better L4 SST estimates as well as improved L4 uncertainty estimates.

7. Conclusions

The impact of including passive microwave (PMW) sea surface temperature (SST) observations is investigated using an existing infrared (IR) gap-free (L4) Arctic surface temperature analysis covering the ocean and sea ice northwards of 58°N. The Arctic suffers from frequent and persistent cloud cover, which prevents IR retrievals of SST. Therefore, the almost all-weather PMW sensors have a significant coverage advantage over IR sensors (which nevertheless provide a much better spatial resolution).

This study provides a systematic assessment of how to best combine IR and PMW SST observations in a blended L4 Arctic SST analysis in order to improve existing reanalyses as well as preparing for future PMW missions (such as CIMR). It is found that the addition of PMW SST observations improves the L4 SST validation results against drifting buoy SSTs for all the methods evaluated here. In order to combine IR and PMW, it is important to correct for systematic biases in the PMW and IR SST data sets relative to each other. The PMW SSTs show lower mean differences against drifter SSTs compared to the IR SSTs, but for consistency with time periods when no Arctic PMW SST observations are available, the PMW SSTs have been adjusted to IR SSTs in this study. This has been done in order to generate a blended IR and PMW Arctic SST climate dataset for the ESA-CCI PMW SST data period (2002-2017).

The L4 SST standard deviations decrease from 0.54°C, 0.55°C and 0.47°C to 0.47°C, 0.54°C and 0.41°C against drifters, moorings and Argo floats, respectively, when PMW SST observations are included. As expected, the largest improvements are seen when the IR data is sparse, but improved performance is seen in almost all regions including those already covered by IR observations. The good performance is likely not only due to the superior PMW coverage but also related to the different and complementary uncertainty characteristics of IR and PMW observations. The mean theoretical uncertainty estimate decreases with 0.08°C when including PMW observations, which is in good agreement with the observed reduction in standard deviation against drifters.

The largest theoretical uncertainties and standard deviations against drifters are generally found along the sea ice edge and coasts, which suffer from no or few SST observations (both from IR and PMW sensors). Improved PMW coverage and SST retrievals are expected in the future with the launch of the CIMR mission, which will provide SSTs with a spatial resolution of 15 km and a precision of 0.3°C (in cold waters) (Donlon, 2020). This will allow PMW SST retrievals much closer to the coasts and sea ice compared to what is possible with previous and current radiometers. Therefore, CIMR has a very large potential to improve Arctic SST estimates even further.

Author contributions

Pia Nielsen-Englyst and Jacob L. Høyer conceived and designed the experiments. Pia Nielsen-Englyst, Jacob L. Høyer, Ioanna Karagali and Wiebke M. Kolbe performed the experiments. Pia Nielsen-Englyst, Jacob L Høyer, Rasmus T. Tonboe and Leif T. Pedersen analyzed the data. Pia Nielsen-Englyst wrote the manuscript with contributions from all authors.

Acknowledgement

This work is funded by the Copernicus Marine Environment Monitoring Service (CMEMS) and the Danish National Centre for Climate Research (NCKF) at the Danish Meteorological Institute (DMI). The PMW SST CDR has been developed within the European Space Agency's Climate Change Initiative for Sea Surface Temperature project under grant number 4000109848/13/I-NB.

References

Alerskans E, Høyer JL, Gentemann CL, Pedersen LT, Nielsen-Englyst P, Donlon C. Construction of a climate data record of sea surface temperature from passive microwave measurements. Remote Sensing of Environment 2020;236:111485.

Alerskans E, Zinck ASP, Nielsen-Englyst P, Høyer JL. Exploring machine learning techniques to retrieve sea surface temperatures from passive microwave measurements. Remote Sensing of Environment 2022;281:113220.

AMAP . Arctic climate change update 2021: Key trends and impacts. summary for policy-makers. 2021.

Atkinson CP, Rayner NA, Kennedy JJ, Good SA. An integrated database of ocean temperature and salinity observations. Journal of Geophysical Research: Oceans 2014;119(10):7139–63.

Bojinski S, Verstraete M, Peterson TC, Richter C, Simmons A, Zemp M. The concept of essential climate variables in support of climate research, applications, and policy. Bulletin of the American Meteorological Society 2014;95(9).

Brasnett B. The impact of satellite retrievals in a global sea-surface-temperature analysis. Quarterly Journal of the Royal Meteorological Society 2008;134(636):1745–60.

Brasnett B, Colan DS. Assimilating retrievals of sea surface temperature from viirs and amsr2. Journal of Atmospheric and Oceanic Technology 2016;33(2):361–75.

Carvalho K, Wang S. Sea surface temperature variability in the arctic ocean and its marginal seas in a changing climate: Patterns and mechanisms. Global and Planetary Change 2020;193:103265.

Castro SL, Emery WJ, Wick GA, Tandy Jr W. Submesoscale sea surface temperature variability from uav and satellite measurements. Remote Sensing 2017;9(11):1089.

Castro SL, Wick GA, Jackson DL, Emery WJ. Error characterization of infrared and microwave satellite sea surface temperature products for merging and analysis. Journal of Geophysical Research: Oceans 2008;113(C3).

Castro SL, Wick GA, Steele M. Validation of satellite sea surface temperature analyses in the beaufort sea using uptempo buoys. Remote Sensing of Environment 2016;187:458–75.

Centurioni LR, Turton J, Lumpkin R, Braasch L, Brassington G, Chao Y, Charpentier E, Chen Z, Corlett G, Dohan K, et al. Global in situ observations of essential climate and ocean variables at the air–sea interface. Frontiers in Marine Science 2019;:419.

Chin TM, Vazquez-Cuervo J, Armstrong EM. A multi-scale high-resolution analysis of global sea surface temperature. Remote sensing of environment 2017;200:154–69.

Comiso J, Grenfell T, Bell D, Lange M, Ackley S. Passive microwave in situ observations of winter weddell sea ice. Journal of Geophysical Research: Oceans 1989;94(C8):10891–905.

Comiso JC, Cavalieri DJ, Markus T. Sea ice concentration, ice temperature, and snow depth using amsr-e data. IEEE Transactions on Geoscience and Remote Sensing 2003;41(2):243–52.

Dash P, Ignatov A, Martin M, Donlon C, Brasnett B, Reynolds RW, Banzon V, Beggs H, Cayula JF, Chao Y, et al. Group for high resolution sea surface temperature (ghrsst) analysis fields inter-comparisons—part 2: Near real time web-based level 4 sst quality monitor (l4-squam). Deep Sea Research Part II: Topical Studies in Oceanography 2012;77:31–43.

Donlon C. Copernicus Imaging Microwave Radiometer (CIMR) Mission Requirements Document, version 4. ref. ESA-EOPSM-CIMR-MRD-3236; European Space Agency; Noordwijk, The Netherlands; 2020.

Donlon C, Casey K, Gentemann C, LeBorgne P, Robinson I, Reynolds R, Merchant C, Llewellyn-Jones D, Minnett P, Piolle J, et al. Successes and challenges for the modern sea surface temperature observing system. Community White Paper for OceanObs 2009;9:1–9.

Donlon C, Robinson I, Casey K, Vazquez-Cuervo J, Armstrong E, Arino O, Gentemann C, May D, LeBorgne P, Piollé J, et al. The global ocean data assimilation experiment high-resolution sea surface temperature pilot project. Bulletin of the American Meteorological Society 2007;88(8):1197–214.

Donlon CJ, Martin M, Stark J, Roberts-Jones J, Fiedler E, Wimmer W. The operational sea surface temperature and sea ice analysis (ostia) system. Remote Sensing of Environment 2012;116:140–58.

Embury O, Merchant CJ, Corlett GK. A reprocessing for climate of sea surface temperature from the along-track scanning radiometers: Initial validation, accounting for skin and diurnal variability effects. Remote Sensing of Environment 2012;116:62–78.

Fiedler E, Mao C, Good S, Waters J, Martin M. Improvements to feature resolution in the ostia sea surface temperature analysis using the nemovar assimilation scheme. Quarterly Journal of the Royal Meteorological Society 2019;145(725):3609–25.

Gentemann C, Scott JP, Mazzini PL, Pianca C, Akella S, Minnett PJ, Cornillon P, Fox-Kemper B, Cetinić I, Chin TM, et al. Saildrone: Adaptively sampling the marine environment. Bulletin of the American Meteorological Society 2020;101(6):E744–62.

Gentemann CL. Three way validation of modis and amsr-e sea surface temperatures. Journal of Geophysical Research: Oceans 2014;119(4):2583–98.

Gentemann CL, Hilburn KA. In situ validation of sea surface temperatures from the gcom-w 1 amsr 2 rss calibrated brightness temperatures. Journal of Geophysical Research: Oceans 2015;120(5):3567–85.

Good S, Fiedler E, Mao C, Martin MJ, Maycock A, Reid R, Roberts-Jones J, Searle T, Waters J, While J, et al. The current configuration of the ostia system for operational production of foundation sea surface temperature and ice concentration analyses. Remote Sensing 2020;12(4):720.

Høyer JL, Karagali I, Dybkjær G, Tonboe R. Multi sensor validation and error characteristics of arctic satellite sea surface temperature observations. Remote Sensing of Environment 2012;121:335–46.

Høyer JL, Le Borgne P, Eastwood S. A bias correction method for arctic satellite sea surface temperature observations. Remote Sensing of Environment 2014;146:201–13.

Huang B, Liu C, Banzon V, Freeman E, Graham G, Hankins B, Smith T, Zhang HM. Improvements of the daily optimum interpolation sea surface temperature (doisst) version 2.1. Journal of Climate 2021;34(8):2923–39.

Jia C, Minnett PJ. High latitude sea surface temperatures derived from modis infrared measurements. Remote Sensing of Environment 2020;251:112094.

Jia C, Minnett PJ, Szczodrak M, Izaguirre M. High latitude sea surface skin temperatures derived from saildrone infrared measurements. IEEE Transactions on Geoscience and Remote Sensing 2022;.

Kasahara M, Kachi M, Inaoka K, Fujii H, Kubota T, Shimada R, Kojima Y. Overview and current status of gosat-gw mission and amsr3 instrument. In: Sensors, Systems, and Next-Generation Satellites XXIV. SPIE; volume 11530; 2020. p. 1153007.

Kilic L, Tonboe RT, Prigent C, Heygster G. Estimating the snow depth, the snow–ice interface temperature, and the effective temperature of arctic sea ice using advanced microwave scanning radiometer 2 and ice mass balance buoy data. The Cryosphere 2019;13(4):1283–96.

Lavergne T, Kern S, Aaboe S, Derby L, Dybkjaer G, Garric G, Heil P, Hendricks S, Holfort J, Howell S, et al. A new structure for the sea ice essential climate variables of the global climate observing system. Bulletin of the American Meteorological Society 2022;103(6):E1502–21.

Le Traon PY, Antoine D, Bentamy A, Bonekamp H, Breivik L, Chapron B, Corlett G, Dibarboure G, DiGiacomo P, Donlon C, et al. Use of satellite observations for operational oceanography: recent achievements and future prospects. Journal of Operational Oceanography 2015;8(sup1):s12–27.

Liu Y, Minnett PJ. Sampling errors in satellite-derived infrared sea-surface temperatures. part i: Global and regional modis fields. Remote sensing of environment 2016;177:48–64.

Meissner T, Manaster A. Smap salinity retrievals near the sea-ice edge using multi-channel amsr2 brightness temperatures. Remote sensing 2021;13(24):5120.

Merchant CJ, Embury O, Bulgin CE, Block T, Corlett GK, Fiedler E, Good SA, Mittaz J, Rayner NA, Berry D, et al. Satellite-based time-series of sea-surface temperature since 1981 for climate applications. Scientific data 2019;6(1):1–18.

Merchant CJ, Embury O, Roberts-Jones J, Fiedler E, Bulgin CE, Corlett GK, Good S, McLaren A, Rayner N, Morak-Bozzo S, et al. Sea surface temperature datasets for climate applications from phase 1 of the european space agency climate change initiative (sst cci). Geoscience Data Journal 2014;1(2):179–91.

Meredith M, Sommerkorn M, Cassotta S, Derksen C, Ekaykin A, Hollowed A, Kofinas G, Mackintosh A, Melbourne-Thomas J, Muelbert M, Ottersen G, Schuur E. Polar regions. in: Ipcc special report on the ocean and cryosphere in a changing climate [h.-o. pörtner, d.c. roberts, v. masson-delmotte, p. zhai, m. tignor, e. poloczanska, k. mintenbeck, a. alegría, m. nicolai, a. okem, j. petzold, b. rama, n.m. weyer (eds.)]. in press. 2019.

Minnett P, Alvera-Azcárate A, Chin T, Corlett G, Gentemann C, Karagali I, Li X, Marsouin A, Marullo S, Maturi E, et al. Half a century of satellite remote sensing of sea-surface temperature. Remote Sensing of Environment 2019;233:111366.

Minnett P, Kaiser-Weiss A. Group for high resolution sea-surface temperature discussion document: Near-surface oceanic temperature gradients. 2012.

Nielsen-Englyst P, Høyer JL, Kolbe WM, Dybkjær G, Lavergne T, Tonboe RT, Skarpalezos S, Karagali I. A combined sea and sea-ice surface temperature climate dataset of the arctic, 1982–2021. Remote Sensing of Environment 2023;284:113331.

Nielsen-Englyst P, L Høyer J, Toudal Pedersen L, L Gentemann C, Alerskans E, Block T, Donlon C. Optimal estimation of sea surface temperature from amsr-e. Remote Sensing 2018;10(2):229.

Olmedo E, Martínez J, Turiel A, Ballabrera-Poy J, Portabella M. Debiased non-bayesian retrieval: A novel approach to smos sea surface salinity. Remote Sensing of Environment 2017;193:103–26.

O'Carroll AG, Armstrong EM, Beggs HM, Bouali M, Casey KS, Corlett GK, Dash P, Donlon CJ, Gentemann CL, Høyer JL, et al. Observational needs of sea surface temperature. Frontiers in Marine Science 2019;6:420.

O'Carroll AG, Eyre JR, Saunders RW. Three-way error analysis between aatsr, amsr-e, and in situ sea surface temperature observations. Journal of atmospheric and oceanic technology 2008;25(7):1197–207.

Pithan F, Mauritsen T. Arctic amplification dominated by temperature feedbacks in contemporary climate models. Nature geoscience 2014;7(3):181–4.

Rantanen M, Karpechko AY, Lipponen A, Nordling K, Hyvärinen O, Ruosteenoja K, Vihma T, Laaksonen A. The arctic has warmed nearly four times faster than the globe since 1979. Communications Earth & Environment 2022;3(1):1–10.

Reynolds RW, Gentemann CL, Corlett GK. Evaluation of aatsr and tmi satellite sst data. Journal of Climate 2010;23(1):152 –65.

Reynolds RW, Gentemann CL, Wentz F. Impact of trmm ssts on a climate-scale sst analysis. Journal of Climate 2004;17(15):2938 –52.

Reynolds RW, Rayner NA, Smith TM, Stokes DC, Wang W. An improved in situ and satellite sst analysis for climate. Journal of climate 2002;15(13):1609–25.

Reynolds RW, Smith TM. Improved global sea surface temperature analyses using optimum interpolation. Journal of climate 1994;7(6):929–48.

Reynolds RW, Smith TM, Liu C, Chelton DB, Casey KS, Schlax MG. Daily high-resolution-blended analyses for sea surface temperature. Journal of climate 2007;20(22):5473–96.

Ricciardulli L, Wentz FJ. Uncertainties in sea surface temperature retrievals from space: Comparison of microwave and infrared observations from trmm. Journal of Geophysical Research: Oceans 2004;109(C12).

Shibata A. Features of ocean microwave emission changed by wind at 6 ghz. Journal of oceanography 2006;62:321–30.

Thiébaux J, Rogers E, Wang W, Katz B. A new high-resolution blended real-time global sea surface temperature analysis. Bulletin of the American meteorological Society 2003;84(5):645–56.

Thomson RE, Emery WJ. Data analysis methods in physical oceanography. Newnes, 2014.

Tonboe RT. The simulated sea ice thermal microwave emission at window and sounding frequencies. Tellus A: Dynamic Meteorology and Oceanography 2010;62(3):333–44.

Tonboe RT, Dybkjær G, Høyer JL. Simulations of the snow covered sea ice surface temperature and microwave effective temperature. Tellus A: Dynamic Meteorology and Oceanography 2011;.

Ulaby F, Long D, Blackwell W, Elachi C, Fung A, Ruf C, Sarabandi K, Zebker H, Van Zyl J. Microwave radar and radiometric remote sensing, university of michigan press. Ann Arbor 2014;.

Ulaby FT, Moore RK, Fung AK. Microwave remote sensing: Active and passive. Volume 3- From theory to applications, 1986.

Vazquez-Cuervo J, Castro SL, Steele M, Gentemann C, Gomez-Valdes J, Tang W. Comparison of ghrsst sst analysis in the arctic ocean and alaskan coastal waters using saildrones. Remote Sensing 2022;14(3):692.

Vincent RF, Marsden R, Minnett P, Buckley J. Arctic waters and marginal ice zones: 2. an investigation of arctic atmospheric infrared absorption for advanced very high resolution radiometer sea surface temperature estimates. Journal of Geophysical Research: Oceans 2008a;113(C8).

Vincent RF, Marsden R, Minnett P, Creber K, Buckley J. Arctic waters and marginal ice zones: A composite arctic sea surface temperature algorithm using satellite thermal data. Journal of Geophysical Research: Oceans 2008b;113(C4).

Wentz FJ, Gentemann C, Smith D, Chelton D. Satellite measurements of sea surface temperature through clouds. Science 2000;288(5467):847–50.

Wentz FJ, Meissner T. Algorithm Theoretical Basis Document(ATBD): AMSR Ocean Algorithm (Version 2). RSS Tech. Proposal 121599A-1; Remote Sensing Systems; Santa Rosa, CA; 2000.

APPENDIX F

Paper VI: Deriving Arctic 2 m air temperatures over snow and ice from satellite surface temperature measurements

This appendix contains a full copy of Paper VI:

- **Nielsen-Englyst, Pia**, Jacob L. Høyer, Kristine S. Madsen, Rasmus T. Tonboe, Gorm Dybkjær, Sotirios Skarpalezos. 2021. "Deriving Arctic 2 m air temperatures over snow and ice from satellite surface temperature measurements". *The Cryosphere*, 15(7), 3035-3057.

The Cryosphere, 15, 3035–3057, 2021
https://doi.org/10.5194/tc-15-3035-2021

The Cryosphere Open Access EGU

Deriving Arctic 2 m air temperatures over snow and ice from satellite surface temperature measurements

Pia Nielsen-Englyst[1,2], Jacob L. Høyer[2], Kristine S. Madsen[2], Rasmus T. Tonboe[2], Gorm Dybkjær[2], and Sotirios Skarpalezos[2]

[1]1DTU-Space, Technical University of Denmark, 2800 Kongens Lyngby, Denmark
[2]Research and Development, Danish Meteorological Institute (DMI), 2100 Copenhagen Ø, Denmark

Correspondence: Pia Nielsen-Englyst (pne@dmi.dk)

Received: 31 January 2021 – Discussion started: 3 March 2021
Revised: 2 June 2021 – Accepted: 2 June 2021 – Published: 2 July 2021

Abstract. The Arctic region is responding heavily to climate change, and yet, the air temperature of ice-covered areas in the Arctic is heavily under-sampled when it comes to in situ measurements, resulting in large uncertainties in existing weather and reanalysis products. This paper presents a method for estimating daily mean clear-sky 2 m air temperatures (T2m) in the Arctic from satellite observations of skin temperature, using the Arctic and Antarctic ice Surface Temperatures from thermal Infrared (AASTI) satellite dataset, providing spatially detailed observations of the Arctic. The method is based on a linear regression model, which has been tuned against in situ observations to estimate daily mean T2m based on clear-sky satellite ice surface skin temperatures. The daily satellite-derived T2m product includes estimated uncertainties and covers the Arctic sea ice and the Greenland Ice Sheet during clear skies for the period 2000–2009, provided on a 0.25° regular latitude–longitude grid. Comparisons with independent in situ measured T2m show average biases of 0.30 and 0.35 °C and average root-mean-square errors of 3.47 and 3.20 °C for land ice and sea ice, respectively. The associated uncertainties are verified to be very realistic for both land ice and sea ice, using in situ observations. The reconstruction provides a much better spatial coverage than the sparse in situ observations of T2m in the Arctic and is independent of numerical weather prediction model input. Therefore, it provides an important supplement to simulated air temperatures to be used for assimilation or global surface temperature reconstructions. A comparison of T2m derived from satellite and ERA-Interim/ERA5 estimates shows that the satellite-derived T2m validates similar to or better than ERA-Interim/ERA5 against in situ measurements in the Arctic.

1 Introduction

The Arctic climate is changing rapidly with surface temperatures rising faster than other regions of the world due to Arctic amplification (Graversen et al., 2008; IPCC, 2013; Pithan and Mauritsen, 2014; Richter-Menge et al., 2017), with the maximum warming occurring during late autumn and early winter (Box et al., 2019; Screen and Simmonds, 2010). Meteorological measurements in Greenland show a general warming since the 1780s (Cappelen, 2021; Masson-Delmotte et al., 2012; Hanna et al., 2021; Abermann et al., 2017), with the 2000s being the warmest decade in western and southern Greenland, while the 2010s in parts of eastern Greenland were slightly warmer than the 2000s (Cappelen, 2021).

The Arctic surface air temperature is one of the key climate indicators used to assess regional and global climate changes (Hansen et al., 2010; Pielke et al., 2007), and both model simulations and observations indicate that warming in the global climate is amplified at the northern high latitudes (e.g. Collins et al., 2013; Holland and Bitz, 2003; Overland et al., 2018). Traditionally, near-surface air temperatures have been measured at the height of 1–2 m using automatic weather stations (AWSs) or buoys (Hansen et al., 2010; Jones et al., 2012; Rayner, 2003; World Meteorological Organization, 2014). Extreme temperatures, winds, and the remoteness of the Arctic make in situ observations in the Arctic temporally and spatially sparse (Reeves Eyre and Zeng, 2017). Therefore, it is challenging to achieve climate-quality temperature records for this region.

The key datasets used to assess the Arctic temperature changes are global gridded near-surface air temperature datasets that are derived using in situ observations (Hansen

et al., 2010; IPCC, 2013; Morice et al., 2012; Smith et al., 2008; Vose et al., 2012). These datasets typically have higher uncertainties in the Arctic region due to the limited availability of in situ observations (Cowtan and Way, 2014; Lenssen et al., 2019; Rapaić et al., 2015). In addition, global reanalysis products such as ERA-Interim (ERA-I) and ERA5 (Dee et al., 2011; Hersbach et al., 2020) are frequently used to study the changes in the Arctic and to force ocean and sea ice models. Despite the assimilation of in situ data in the global reanalysis models, significant model differences have been reported for the Arctic (Davy and Outten, 2020; Delhasse et al., 2020; Lindsay et al., 2014; Wesslén et al., 2014) as well as large deviations from observations of T2m over Arctic sea ice (Wang et al., 2019).

Observations from polar-orbiting satellites offer a very good supplement to the in situ observations through high spatial and temporal coverage of the high latitudes and may improve the surface temperature products and the assessment of the Arctic climate changes. Therefore, daily near-surface air temperatures derived from satellite temperature observations have the potential to increase the amount of information in the datasets and improve the quality of the climate records, as recognized in Merchant et al. (2013) and Rayner et al. (2020).

Two fundamental challenges exist when deriving a T2m product from infrared satellite observations. The first challenge is that infrared sensors (in the atmospheric window region of 10–12 µm wavelength) measure the ice surface skin temperature (IST_{skin}), whereas the current global temperature products include the near-surface air temperature as measured continuously by AWSs and buoys. The surface skin temperature may differ considerably from the near-surface air temperature measured by AWSs or buoys. Previous studies have compared satellite-retrieved IST_{skin} and T2m from AWSs located on the Greenland Ice Sheet (GrIS; Dybkjær et al., 2012a; Hall et al., 2008, 2012; Koenig and Hall, 2010; Shuman et al., 2014) and over the Arctic sea ice (Dybkjær et al., 2012) and found temperature differences of which a significant part could be attributed to the temperature difference between T2m and IST_{skin}. Other studies have investigated the relationship between T2m and IST_{skin} over ice using in situ observations (Adolph et al., 2018; Hall et al., 2008, 2004; Hudson and Brandt, 2005; Nielsen-Englyst et al., 2019; Vihma et al., 2008). Nielsen-Englyst et al. (2019) found that on average T2m is 0.65–2.65 °C higher than IST_{skin} with variations depending on the location of the measurement, i.e. over sea ice, seasonal snow cover, and the following zones of the GrIS: lower ablation zone, upper-middle ablation zone, and accumulation zone. The T2m–IST_{skin} difference was found to vary seasonally with the largest differences during the winter (when inversions are most common) and during melting conditions in the summer (where the surface temperature is fixed at the melting point). In Nielsen-Englyst et al. (2019), wind speed and cloud cover were identified as key parameters determining the T2m–IST_{skin} difference.

The second challenge, related to the use of satellite-derived infrared IST_{skin} to derive T2m, is that the availability of IST_{skin} observations is limited to clear-sky conditions while T2m is measured continuously by AWSs and buoys. Previous studies have shown that a satellite-derived, clear-sky, surface temperature record can be significantly colder than an all-sky surface temperature record (Koenig and Hall, 2010; Nielsen-Englyst et al., 2019). To benefit from the good coverage of satellite surface temperature data, above-mentioned challenges should be considered with caution. This work, starting with Nielsen-Englyst et al. (2019), has been initiated to estimate clear-sky T2m from satellite observations (whenever these are available) for the Arctic sea ice and the GrIS in order to provide spatially detailed observations for the areas unobserved by in situ stations and to supplement the in situ observations already available. Here, special attention has been given to the above-mentioned challenges, and the relationships between the near-surface air temperature and the satellite skin measurements have been explored in detail. A regression-based approach has been used to estimate daily T2m using satellite IST_{skin} and a seasonal cycle function as predictors based on the work presented in Høyer et al. (2018). The derived product covers only days with no or limited clouds, when satellite skin temperature observations are available. However, for those days when the satellite-derived T2m product is available, it provides an estimate of the daily averaged all-sky T2m since it has been regressed towards in situ measurements from both clear and cloudy conditions. In order to further facilitate the usage of the derived product in modelling and for monitoring purposes, each satellite-retrieved T2m estimate comes with uncertainties.

Similar efforts have been made to estimate clear-sky near-surface air temperatures (and corresponding uncertainties) over land, ocean, and lakes using satellite observations to cover all surfaces of the Earth (Good, 2015; Good et al., 2017; Høyer et al., 2018). The previous work has mostly been done as a part of the European Union's Horizon2020 project EUSTACE (EU Surface Temperatures for All Corners of Earth, 2015–2019, https://www.eustaceproject.org, last access: 29 June 2021), with the overall aim to produce a globally complete gap-free daily near-surface temperature analysis since 1850. It is outside the scope of this paper to produce a daily continuous gap-free near-surface temperature analysis. However, within EUSTACE this has been done using a statistical model to combine satellite-derived clear-sky near-surface air temperatures and in situ observations and their respective uncertainty estimates (Morice et al., 2019; Rayner et al., 2020). The clear-sky T2m product derived in this paper has been used to generate this daily gap-free EUSTACE T2m product for the GrIS and the Arctic sea ice, while similar clear-sky temperature products have been used over land, ocean, and lakes.

This paper is structured such that Sect. 2 describes the in situ data and the satellite data. Section 3 presents the method used to estimate clear-sky daily T2m and uncertainties. The resulting T2m dataset and its validation are presented in Sect. 4 and discussed in Sect. 5. Conclusions are given in Sect. 6.

2 Data

2.1 In situ data

In situ observations of near-surface air temperatures have been collected from weather stations, expeditions, and campaigns covering ice and snow surfaces to assemble the DMI-EUSTACE database. The database includes quality controlled and uniformly formatted temperature observations covering ice and snow surfaces during the period 2000–2009 (Høyer et al., 2018). For the GrIS we use the Programme for Monitoring of the Greenland Ice Sheet (PROMICE) data provided by the Geological Survey of Denmark and Greenland (GEUS; Fausto and van As, 2019; Ahlstrøm et al., 2008; van As et al., 2011) and the Greenland Climate Network data (GC-Net; Kindig, 2010; Shuman et al., 2001; Steffen and Box, 2001). Only PROMICE data from the middle-upper ablation zone and accumulation zone have been used to ensure that data are only acquired over permanently snow- or ice-covered surfaces. Observations covering seasonal snow have also been used from the Atmospheric Radiation Measurement (ARM) programme from two sites: Atqasuk (ATQ) and Barrow (BAR), at the North Slope of Alaska (Ackerman and Stokes, 2003; Stamnes et al., 1999). Data from Arctic sea ice are primarily retrieved from the meteorological observation archive at the European Centre for Medium-Range Weather Forecasts (ECMWF) MARS data storage facility, providing 196 unique data series from drifting buoys. These sea ice data are supplemented with data from 10 US Army Cold Regions Research Engineering Laboratory (CRREL) mass balance buoys (Perovich et al., 2016; Richter-Menge et al., 2006) and observations from a weather station located 29 m above the sea surface on the research vessel *Polarstern* operated by the Alfred Wegener Institute in the sea-ice-covered parts of the Arctic Ocean (Knust, 2017; König-Langlo et al., 2006a). We also use air temperature measurements obtained from ice buoys deployed in the Fram Strait region within the framework of the Fram Strait Cyclones (FRAMZY) campaigns during the years 2002, 2007, and 2008 as well as air temperatures from the Arctic Climate System Study (ACSYS) campaign in 2003 (Brümmer et al., 2011b, c, 2012b, a). Finally, we use data from two ice buoy campaigns operated by the Meteorological Institute of the University of Hamburg within the framework of the integrated EU research project DAMOCLES (Developing Arctic Modelling and Observing Capabilities for Long-term Environmental Studies; Brümmer et al., 2011a).

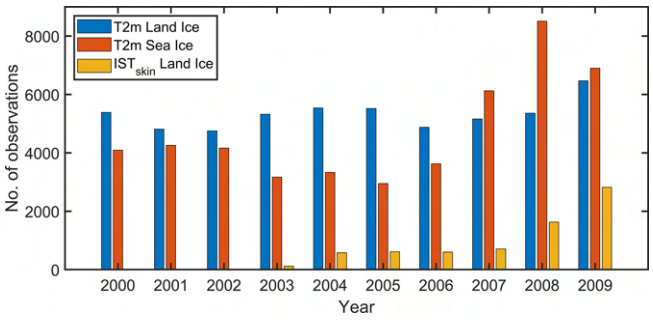


Figure 1. Total number of daily averaged in situ observations of T2m and IST_{skin} over Arctic land ice and sea ice per year covering the period 2000–2009.

The different in situ types measure the air temperature at different heights that furthermore differ over time depending on the amount of snowfall, snow drift, and snowmelt. Here, we will refer to T2m for all observation types regardless of these variations. Nielsen-Englyst et al. (2019) showed small changes ($< 0.22\,°C$) in T2m–IST_{skin} differences when using only observations within the measurement range of 1.90–2.10 m in height compared to using all measurements (ranging in measurement height from 0.3 to 3 m). The observations from *Polarstern* at 29 m height are not included in the derivation of the near-surface air temperature dataset but only used for the validation. The accuracy of the air temperature sensors for all observation sites is approximated to 0.1 °C (Hall et al., 2008; Høyer et al., 2017b). Few data sources provide both skin and air temperatures, e.g. the PROMICE and ARM stations. The PROMICE skin temperatures have been calculated from upwelling longwave radiation, measured by Kipp & Zonen CNR1 or CNR4 radiometers, assuming a surface longwave emissivity of 0.97 (van As, 2011). All in situ data have been screened for spikes and other unrealistic data artefacts by visual inspection. Afterwards, the in situ observations have been averaged to daily temperatures using all available observations. Figure 1 shows the number of daily averaged in situ observations each year (2000–2009) of IST_{skin} and T2m over Arctic land ice and sea ice. The two ARM stations are included as land ice stations in this analysis, and only data from snow-covered periods are used. In total 65 810 observations with daily T2m and 7057 observations with daily IST_{skin} are available over land ice. See Table 1 for more information on the in situ observations used in this study.

2.2 Satellite data

The satellite data used in this study are from the Arctic and Antarctic Ice Surface Temperatures from thermal Infrared satellite sensors (AASTI; Dybkjær et al., 2014, 2018; Høyer et al., 2019) dataset, covering high-latitude seas, sea ice, and ice sheet with clear-sky surface temperatures based on satellite infrared measurements from the CLARA-A1 dataset

Table 1. Overview of in situ observations used in this study, covering the period 2000–2009.

	No. of sites, (AWS, buoys, or ships)	No. of days with observations	Surface type	Observation type	Temperature measurements
ACSYS	7	280	Sea ice	Buoy	T2m
ARM	2	2846	Seasonal snow	AWS	T2m, IST_{skin}
CRREL	10	1031	Sea ice	Buoy	T2m
DAMOCLES	25	2160	Sea ice	Buoy	T2m
ECMWF	196	27 235	Sea ice	Buoy	T2m
FRAMZY	11	251	Sea ice	Buoy	T2m
GC-NET	15	29 133	Land ice	AWS	T2m
POLARSTERN	1	189	Sea ice	Ship	T2m
PROMICE	8	2685	Land ice	AWS	T2m, IST_{skin}

compiled by EUMETSAT's Climate Monitoring, Satellite Application Facility (CM-SAF; Karlsson et al., 2013). The dataset is based on one of the longest existing satellite records from the Advanced Very High Resolution Radiometer (AVHRR) instruments on board a long series of NOAA satellites. AASTI contains swath-based (i.e. Level 2; L2) ice surface skin temperature (IST_{skin_L2}) data processed and error corrected on the original Global Area Coverage (GAC) grid. The first version of the AASTI product, which is used in this study, is available from 2000 to 2009 in the original projection and resolution (L2), i.e. ~ 0.05 arc degree resolution and multiple daily coverage. Since 2000, seven different AVHRR instruments have been orbiting the globe, each 14 times per day, thus providing approximately bi-hourly coverage of the polar regions (Fig. 2). The number of operational satellites increased from two to six from 2000 to 2009. The IST algorithm used to generate the AASTI dataset is based on thermal infrared brightness temperatures of AVHRR channels 4 (centre wavelength at $\sim 11\,\mu m$) and 5 (centre wavelength at $\sim 12\,\mu m$) and the satellite zenith angle. The algorithm is a split window algorithm, working within three temperature domains for each individual satellite (Key et al., 1997). The retrieval calibration of each domain has been done by relating modelled surface temperatures with modelled top-of-atmosphere brightness temperatures, determined by a radiative transfer model (Dybkjær et al., 2014). Cloud masking has been performed using the Polar Platform System (PPS) cloud processing software (Dybbroe et al., 2005a, b).

As discussed in Merchant et al. (2017), satellite-based climate data records should include uncertainty estimates. The AASTI IST_{skin_L2} data come with uncertainties divided into three independent uncertainty components, each with different characteristics: the random uncertainty (μ_{rnd_L2}), a locally systematic uncertainty (μ_{local_L2}), and a large-scale systematic ("global") uncertainty (μ_{glob_L2}). These three components have been chosen since they behave differently when aggregating the observations in time or space (see Sect. 3.2). This uncertainty methodology has been developed within the sea surface temperature (SST) community (Bulgin et al., 2016; Rayner et al., 2015) and will be followed here. The total uncertainty on the IST_{skin_L2}, μ_{total_L2}, is calculated by summing each component in quadrature (i.e. square root of sum of squares). Excluding the cloud mask uncertainty, grid cell systematic uncertainties (μ_{glob_L2}) are set to a fixed value of 0.1 °C to represent systematic uncertainties in the forward models (see e.g. Merchant et al., 1999; Merchant and Le Borgne, 2004). The AASTI IST_{skin_L2} data also come with a quality level (QL) from 1 (bad data) to 5 (best quality), with the addition of level 0 (no data) (GHRSST Science Team, 2010).

Here, we have aggregated the AASTI IST_{skin_L2} observations into 3-hourly and daily gridded Level 3 (L3) averages of IST_{skin_L2} on a fixed 0.25° by 0.25° regular geographical grid. This grid was chosen within the EUSTACE project to ensure a common grid to be used globally. The daily gridded averages (IST_{skin_L3}) are calculated by averaging all available IST_{skin_L2} observations with a quality flag of 4 (good) or 5 (best) for a given date and within the 0.25° bin. This has been done to facilitate the development of the relationship model and to ease the user uptake. The data in the daily aggregated files contain mean surface temperature observations from 00:00 to 24:00 LST, 3-hourly bin averages of surface temperatures, and also the number of observations in the eight time bins during each day. The 3-hourly numbers of observations are used to estimate the satellite sampling throughout the day, and the 3-hourly temperature data are used to gain confidence in the daily cycle estimates (see quality checks below). Figure 3 shows the mean number of observations per day in each of the eight time intervals given in local time for the Arctic region. The variation in coverage throughout the day is a combined effect of the satellite overpassing, performance of the cloud screening algorithm, and the cloud-free conditions during the day. In addition, the fixed 0.25° regular geographical grid results in a decreasing L3 bin area when approaching the North Pole. The maximum satellite coverage is generally seen around 80° N with a min-

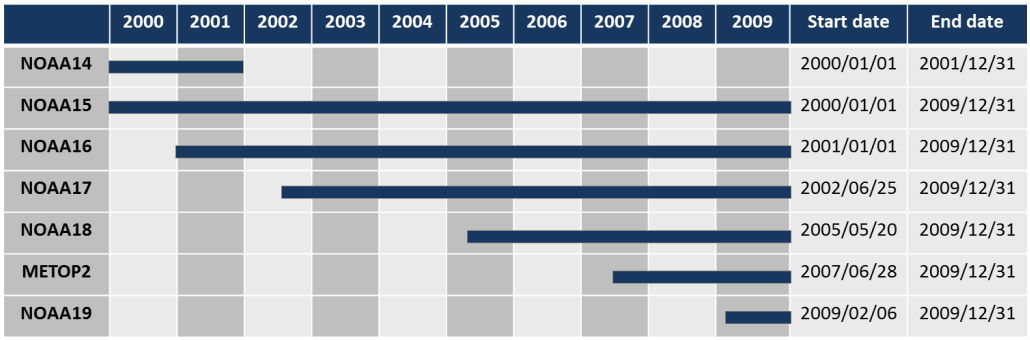


Figure 2. NOAA and Metop satellites carrying the AVHRR sensor, used for AASTI version 1.

imum at the North Pole. Cloud-free conditions over the GrIS are primarily observed around noon and the early afternoon.

In order to best resolve the diurnal cycle with satellite information, we require data during both the night (between 18:00 and 06:00 LST) and the day (between 06:00 and 18:00 LST) in order to calculate IST_{skin_L3}. To identify sea ice, we use an ice mask for which sea ice is characterized by sea ice concentrations above 30 % according to the EUMETSAT OSISAF Global Sea Ice Concentration Climate Data Record (Tonboe et al., 2016). A few more checks have been set up in order to minimize the temporal sampling errors, the effects of undetected clouds and outliers, and inconsistencies between the ice mask and the surface temperatures. Following Høyer et al. (2018), the IST_{skin_L3} is discarded if one of the following criteria is met:

- IST_{skin_L3} exceeds +5 °C, indicating inconsistency between the ice mask and the surface temperatures.
- The standard deviation of satellite IST_{skin_L2} during 1 d exceeds 7.07 °C, corresponding to a sinusoidal daily cycle with a difference between day and night of 20 °C.
- The difference between IST_{skin_L3} and the average of all available 3 h bin averages exceeds 10 °C.
- IST_{skin_L3} is more than 10 °C colder than the corresponding average of up to 24 neighbouring cloud-free observations (in a 5-by-5 grid cell square) with the same surface type.

The criteria above have been derived from analysis and inspection of the satellite data and with considerations to the results presented in Nielsen-Englyst et al. (2019). Inconsistencies between the ice mask and surface temperature typically occur along the coasts and sea ice edge, where the OSISAF product is subject to land-spillover effects causing spurious ice in ice-free areas (Lavergne et al., 2019). Using a surface temperature threshold of 5 °C reduces the land-spillover effects and results in increased consistency between the ice mask and the surface temperatures.

The satellite-derived surface temperature has seasonal differences in daily variability, with the largest standard deviations during the summer in Greenland and during the winter for sea ice, when the freeze-up of sea ice causes higher variability along the sea ice margin (Fig. 4). The main uncertainty components of the IST_{skin_L3} estimates are erroneous cloud screening and the spatial variance of snow and ice surface emissivity, which are not accounted for in the retrieval algorithm. The presence of non-detected clouds will contribute to increased standard deviations and usually a cold IST_{skin_L3} bias, since the cloud tops and other atmospheric constituents are generally colder than the surface (Dybkjær et al., 2012).

2.2.1 Validation

Additional satellite versus in situ differences arise when comparing satellite observations with pointwise ground measurements due to different spatial and temporal characteristics. To assess the magnitude of these effects, the IST_{skin_L3} data have been validated against in situ observations from the PROMICE and ARM stations. Table 2 shows the validation results of daily IST_{skin_L3} against in situ skin temperatures (IST_{skin_insitu}) and in situ 2 m air temperatures ($T2m_{insitu}$). The maximum matchup distance is 14.6 km, and the average distance is 8.1 km, considering the AWSs in Table 2. The topography mask included in the HIRHAM5 regional climate model (see e.g. Langen et al., 2015) has been used to calculate the differences in elevation (Δh) between the in situ stations and corresponding satellite pixels. There is no clear correlation between the large biases and large elevation differences from this table, but the elevation effects are contributing to the spatial sampling error. The spatial and temporal sampling errors contribute to the overall uncertainty, but effects from erroneous cloud screening, algorithm simplifications, and uncertainties in the in situ observations are also included in the results. Previous studies find that erroneous cloud screening (undetected clouds) is one of the main reasons for the cold biases observed when comparing satellite-observed IST with in situ measurements (Hall et al., 2004, 2012; Koenig and Hall, 2010; Østby et al., 2014; Westermann et al., 2012). Another important contribution is the effect of comparing clear-sky satel-

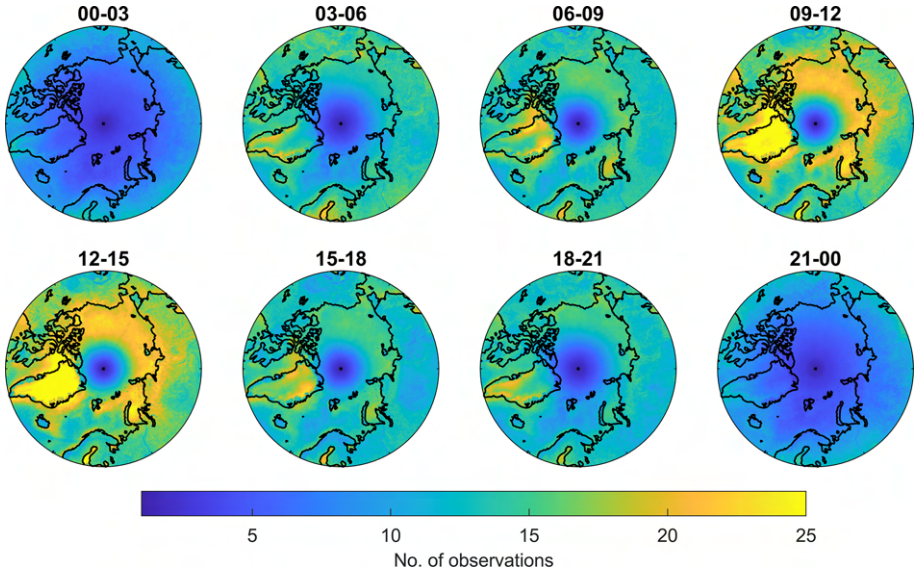


Figure 3. Mean number of observations per day in the L3 bins for each of the eight local solar time intervals, averaged for the period 2000–2009.

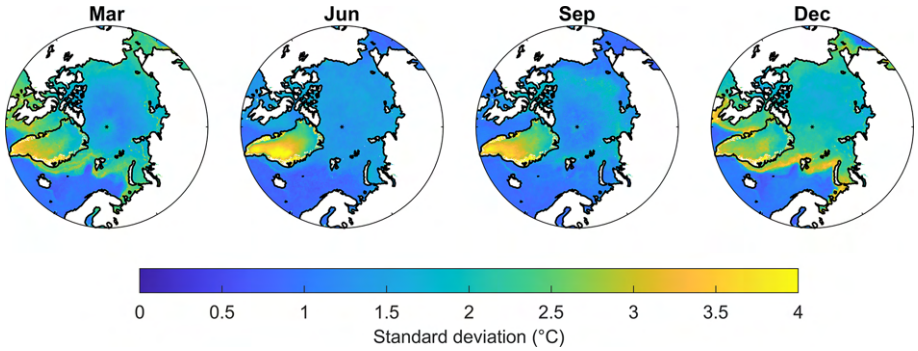


Figure 4. Standard deviations (°C) of daily satellite surface temperature observations for March, June, September, and December of each year averaged for the years 2000–2009.

lite observations with all-sky in situ observations, as discussed in Nielsen-Englyst et al. (2019). In general, IST_{skin_L3} correlates better with $T2m_{insitu}$ than with the IST_{skin_insitu}. Moreover, the IST_{skin_L3}–$T2m_{InSitu}$ difference shows smaller standard deviations than IST_{skin_L3}–IST_{skin_insitu}. However, as expected the biases and root-mean-squared differences (RMSDs) are larger for the IST_{skin_L3}–$T2m_{insitu}$ differences than for the IST_{skin_L3}–IST_{skin_insitu} differences. The reason is that the radiometric surface skin temperature can be significantly different from the surface air temperature measurements (Adolph et al., 2018; Hall et al., 2008; Hudson and Brandt, 2005; Nielsen-Englyst et al., 2019; Vihma et al., 2008). On average, the skin temperature is colder than the air temperature (Nielsen-Englyst et al., 2019), resulting in even more negative biases, when the IST_{skin_L3} is compared to in situ measured T2m, instead of in situ skin temperatures. The generally high correlations are dominated by the synoptic (2–5 d) and seasonal variations, which are pronounced in both IST and T2m.

3 Methods

3.1 Regression model

Nielsen-Englyst et al. (2019) analysed a large number of in situ stations with simultaneous T2m and IST_{skin} observations and showed that empirical relationships exist between T2m and IST_{skin}. However, it was also shown that the relationships varied for different regions. Based upon these results, it was decided to use a simple-regression-based method in this paper to derive the daily mean T2m from the satellite IST_{skin_L3} observations. Separate regression models have been derived for land ice and sea ice.

Table 2. Validation of daily AASTI v.1 Level 3 IST (IST_{skin_L3}) against in situ IST_{skin} (IST_{skin_insitu}) and T2m observations ($T2m_{insitu}$). N: number of matchups; Corr: correlation; SD: standard deviation; RMSD: root-mean-square difference. d is the matchup distance and Δh is the difference in elevation (AWS − satellite).

		$IST_{skin\_L3} - IST_{skin\_insitu}$				$IST_{skin\_L3} - T2m_{insitu}$				d (km)	Δh (m)
Station	N	Corr	Bias	SD	RMSD	Corr	Bias	SD	RMSD		
ARM_ATQ	1235	93.8	−2.47	3.69	4.44	93.7	−3.17	3.69	4.87	10.8	–
ARM_BAR	1594	94.1	−0.73	4.30	4.36	94.6	−1.14	4.02	3.86	6.1	–
PROMICE KAN-M	422	93.9	−3.65	3.37	4.96	94.6	−4.56	3.14	5.53	7.6	15
PROMICE KAN-U	239	93.9	−1.75	3.32	3.75	94.4	−3.39	3.17	4.64	14.6	21
PROMICE KPC-U	488	97.6	−1.31	2.62	2.92	98.2	−3.20	2.27	3.92	5.1	29
PROMICE NUK-U	296	77.7	−4.09	5.00	6.45	84.7	−7.19	4.01	8.23	14.4	64
PROMICE QAS-U	407	83.9	−1.65	4.20	4.51	86.3	−3.70	3.75	5.27	6.5	197
PROMICE SCO-U	403	91.5	−4.60	4.25	6.26	93.7	−7.55	3.75	8.43	4.2	20
PROMICE TAS-U	386	67.5	−1.03	5.43	5.52	79.5	−3.61	4.39	5.68	8.4	214
PROMICE UPE-U	125	88.2	−3.13	3.88	4.97	90.0	−5.49	3.50	6.50	3.0	110
All data	5595	92.9	−2.03	4.24	4.70	93.2	−3.36	4.12	5.32	8.1	83.8

To test different types of regression models, the IST_{skin_L3} data have been matched up with in situ observations for each day (Høyer et al., 2018). This is done by requiring a distance to the nearest in situ site of less than 15 km. The average matchup distance is 8.6 and 7.2 km for land ice and sea ice, respectively, which means that all in situ observations are made within the area of the satellite pixel. The corresponding mean elevation difference is 30 m (while the absolute mean elevation difference is 45 m) and is calculated using the topography mask included in HIRHAM5 (Langen et al., 2015) for the 23 GrIS AWSs. Out of the 23 AWSs, four of them (GC-net JAR1, TAS_U, QAS_U, and UPE_U) have corresponding elevation differences above 100 m. In Sect. 4.3, the effect of these AWSs has been estimated and discussed. All in situ observations, described in Sect. 2.1., have been matched with IST_{skin_L3} data, resulting in a total number of daily matchups of 65 810 from 275 different observation sites (see Table 1). These have been divided into two subsets: one for training and one for validation of the different regression models for land ice and sea ice, respectively. This has been done while ensuring similar coverage of training and validation data over the two domains, which is shown in Fig. 5. The result is that 40 % (13 792 matchups) are used for testing the regression models (and generating the regression coefficients), and the remaining 60 % (20 872 matchups) are left for validation of the regression models over land ice. Over sea ice 48 % (15 035 matchups) are used for testing, and 52 % (16 111 matchups) are left for validation.

The regression model is based on multiple linear regression analysis using least squares (Menke, 1989). The multiple linear regression analysis equations can be written in matrix form,

$$\boldsymbol{d}^{\mathrm{obs}} = \mathbf{G}\boldsymbol{m} + \boldsymbol{e}, \quad (1)$$

$$\boldsymbol{d}^{\mathrm{pre}} = \mathbf{G}\boldsymbol{m}, \quad (2)$$

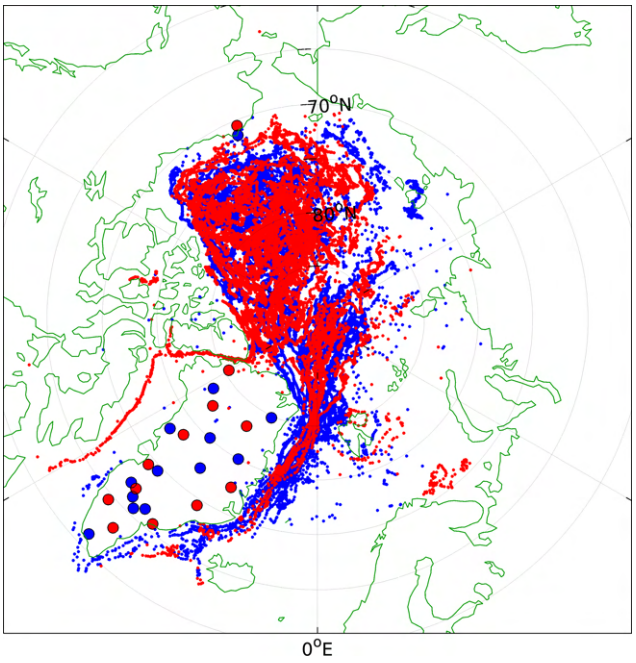


Figure 5. Positions of matchups on sea ice and land ice (red: training; blue: validation).

where $\boldsymbol{d}^{\mathrm{obs}}$ and $\boldsymbol{d}^{\mathrm{pre}}$ are vectors containing the observed and modelled in situ air temperatures, respectively, $\mathbf{G}$ is a matrix containing the various predictors, $\boldsymbol{m}$ is a vector containing regression coefficients, and $\boldsymbol{e}$ is the fitting error.

The regression coefficients are found using damped least squares (Menke, 1989). The least-squares method is used since the problem is generally over-determined, and the damping is added to limit effects of noisy data. The regression coefficients are thus given as

$$\mathbf{G}^{-\mathrm{g}} = \left(\mathbf{G}^T\mathbf{G} + \varepsilon^2\mathbf{I}\right)^{-1}\mathbf{G}^T, \quad (3)$$

$$\boldsymbol{m} = \mathbf{G}^{-\mathrm{g}}\boldsymbol{d}^{\mathrm{obs}}, \quad (4)$$

where $\mathbf{G}^{-\mathrm{g}}$ is called the generalized inverse, ε is a damping factor, and $\mathbf{I}$ is an identity matrix (with ones in the diagonal and zeros elsewhere). The superscript operator T denotes transposing and -1 denotes inversion. We have tested a range of damping factors to assess the relation to the error coefficients. A damping factor of 0.2 was chosen to avoid overfitting noise in the data, while keeping the error coefficients low.

The choice of predictors is based on current knowledge of the parameters that influence the relationship between $\mathrm{IST_{skin}}$ and $\mathrm{T2m_{insitu}}$ (Adolph et al., 2018; Hall et al., 2008; Hudson and Brandt, 2005; Nielsen-Englyst et al., 2019; Vihma and Pirazzini, 2005), limited by the available satellite data. Nielsen-Englyst et al. (2019) showed that the T2m–Tskin difference varies over the season with the smallest differences during the spring, autumn, and summer in non-melting conditions. For that reason, we have also tested the effect of including a seasonal cycle as predictor. A total of five regression models with different predictors have been tested (Høyer et al., 2018).

$$\hat{\mathrm{IST}}_{\mathrm{skin}}: \quad \mathrm{T2m_{sat}} = \alpha_0 + \alpha_1 \mathrm{IST_{skin\_L3}} \quad (5)$$

$$\hat{\mathrm{IST}}_{\mathrm{skin}}\mathrm{SWd}: \quad \mathrm{T2m_{sat}} = \alpha_0 + \alpha_1 \mathrm{IST_{skin\_L3}} + \alpha_2 \mathrm{SWd} \quad (6)$$

$$\hat{\mathrm{IST}}_{\mathrm{skin}}\mathrm{WS}: \quad \mathrm{T2m_{sat}} = \alpha_0 + \alpha_1 \mathrm{IST_{skin\_L3}} + \alpha_2 \mathrm{WS} \quad (7)$$

$$\hat{\mathrm{IST}}_{\mathrm{skin}}\mathrm{Lat}: \quad \mathrm{T2m_{sat}} = \alpha_0 + \alpha_1 \mathrm{IST_{skin\_L3}} + \alpha_2 \mathrm{Lat} \quad (8)$$

$$\hat{\mathrm{IST}}_{\mathrm{skin}}\mathrm{Season}: \quad \mathrm{T2m_{sat}} = \alpha_0 + \alpha_1 \mathrm{IST_{skin\_L3}} + \alpha_2 \cos((t \cdot 2\pi)/(1\,\mathrm{yr})) + \alpha_3 \sin((t \cdot 2\pi)/(1\,\mathrm{yr})) \quad (9)$$

The regression model in Eq. (8) is limited to an offset and a scaling of $\mathrm{IST_{skin\_L3}}$, where the latter term accounts for the synoptic and seasonal variations, which are the dominating factors in both the IST and T2m variability. This part is thus included in all regression models tested. The other regression models also have a third predictor, which is included to examine how to best represent the residual variations in the T2m–IST difference. The model in Eq. (9) uses theoretical top-of-atmosphere shortwave radiation, Eq. (10) uses the wind forcing (from ERA-I and ERA5, respectively), Eq. (11) uses latitude variation, and Eq. (12) uses a seasonal variation. In the regression model in Eq. (12), the seasonal variation is assumed to be the shape of a cosine function, $A \cdot \cos((t \cdot 2\pi)/(1\,\mathrm{yr}) - \varphi)$, where A is the amplitude, φ is the phase and t is time. Since $\cos(x_1 - x_2) = \cos(x_1)\cos(x_2) + \sin(x_1)\sin(x_2)$, the seasonal cycle can be rewritten to the form in Eq. (12) with $A = \sqrt{\alpha_2^2 + \alpha_3^2}$ and $\varphi = \arctan(\alpha_3/\alpha_2)$.

Table 3. Statistics on the relation between observed and modelled temperatures for the training data. N: number of matchups used for testing; Corr: correlation; RMSD: root-mean-square difference. Since, the training data are used for the regression, the bias is zero, and thus the standard deviation equals RMSD.

		N	Corr (%)	RMSD (°C)
Land ice	$\hat{\mathrm{IST}}_{\mathrm{skin}}$	13 792	95.7	3.51
	$\hat{\mathrm{IST}}_{\mathrm{skin}}\mathrm{SWd}$	13 792	96.2	3.28
	$\hat{\mathrm{IST}}_{\mathrm{skin}}\mathrm{WS_{ERA\text{-}I}}$	13 792	95.8	3.47
	$\hat{\mathrm{IST}}_{\mathrm{skin}}\mathrm{WS_{ERA5}}$	13 792	95.9	3.42
	$\hat{\mathrm{IST}}_{\mathrm{skin}}\mathrm{Lat}$	13 792	95.8	3.48
	$\hat{\mathrm{IST}}_{\mathrm{skin}}\mathrm{Season}$	13 792	96.3	3.28
Sea ice	$\hat{\mathrm{IST}}_{\mathrm{skin}}$	15 035	96.0	3.32
	$\hat{\mathrm{IST}}_{\mathrm{skin}}\mathrm{SWd}$	15 035	96.0	3.32
	$\hat{\mathrm{IST}}_{\mathrm{skin}}\mathrm{WS_{ERA\text{-}I}}$	15 035	96.0	3.32
	$\hat{\mathrm{IST}}_{\mathrm{skin}}\mathrm{WS_{ERA5}}$	15 035	96.0	3.32
	$\hat{\mathrm{IST}}_{\mathrm{skin}}\mathrm{Lat}$	15 035	96.1	3.28
	$\hat{\mathrm{IST}}_{\mathrm{skin}}\mathrm{Season}$	15 035	96.2	3.25

The training data have been used to calculate the regression coefficients for each regression model covering the land ice and sea ice. The performance of each regression model has been investigated using the training data, and the results are shown in Table 3. The best performance is found by using the regression model where $\mathrm{T2m_{sat}}$ is predicted from $\mathrm{IST_{skin\_L3}}$ combined with a seasonal variation ($\hat{\mathrm{IST}}_{\mathrm{skin}}\mathrm{Season}$). This model predicts $\mathrm{T2m_{sat}}$ better compared to the other regression models, with correlations above 96 % and RMSD values of 3.25–3.28 °C against training data for both surface types (Table 3). In the following, we will use the regression model given in Eq. (12) with the seasonal term included and with separate regression coefficients for land ice and sea ice (see Table 4). The phase corresponds to a maximum on the 19 January and 12 February for land ice and sea ice, respectively. This is in agreement with Nielsen-Englyst et al. (2019), who found the strongest clear-sky inversion during the winter months (December–February) for all sites included in the analysis except from the ones located in the lower ablation zone (not included here), where pronounced surface melt takes place for long periods of time.

3.2 Uncertainty estimates for $\mathrm{T2m_{sat}}$

Uncertainty estimates on the derived $\mathrm{T2m_{sat}}$ are crucial to facilitate the usage of the dataset in modelling and for monitoring purposes. The uncertainty estimates of the satellite-derived $\mathrm{T2m_{sat}}$ data follow the approach in Bulgin et al. (2016) and Rayner et al. (2015), which has also been used for the AASTI data. The uncertainty on a single $\mathrm{T2m_{sat}}$ estimate is divided into random, locally correlated, and systematic uncertainty components, with the total uncertainty

Table 4. Model regression coefficients for $\hat{\mathrm{IST}}_{\mathrm{skin}}$Season.

	Offset, α_0 (°C)	$\mathrm{IST}_{\mathrm{skin\_L3}}$ factor, α_1	Amplitude, A	Phase, φ
Land ice	4.20	1.06	2.26	−0.33
Sea ice	1.46	0.89	1.83	−0.75

$\mu_{\mathrm{total\_T2m}}$ given as the square root of the sum of the three squared components:

$$\mu_{\mathrm{total\_T2m}} = \sqrt{\mu^2_{\mathrm{rnd\_T2m}} + \mu^2_{\mathrm{local\_T2m}} + \mu^2_{\mathrm{glob\_T2m}}}.$$

The random uncertainty component for the $\mathrm{T2m}_{\mathrm{sat}}$ belonging to a particular grid cell at a particular point in time is found by propagating the AASTI $\mathrm{IST}_{\mathrm{skin\_L3}}$ random uncertainty through the regression model:

$$\mu_{\mathrm{rnd\_T2m}} = \sqrt{\left(\alpha_1 \mu_{\mathrm{rnd\_L3}}\right)^2},$$

with $\mu_{\mathrm{rnd\_L3}}$ given as the aggregated $\mu_{\mathrm{rnd\_L2}}$:

$$\mu_{\mathrm{rnd\_L3}} = \frac{\mu_{\mathrm{rnd\_L2}}}{\sqrt{N}},$$

where N is the number of observations for each bin in the aggregation from L2 to L3. The $\sqrt{N}$ reduction applies because the random uncertainty of each L2 data point that goes into the L3 calculation is by definition independent from the other.

The L3 global uncertainty component does not average out in any aggregation and is thus transferred directly from the L2 uncertainty estimate and multiplied by α_1 to make up $\mu_{\mathrm{glob\_T2m}}$:

$$\mu_{\mathrm{glob\_T2m}} = \alpha_1 \mu_{\mathrm{glob\_L3}} = \alpha_1 \cdot 0.1\,^{\circ}\mathrm{C}.$$

The $\mu_{\mathrm{local\_T2m}}$ contains the local uncertainty component of L2, a sampling error $\mu_{\mathrm{lsamp\_L3}}$ related to sampling errors in space and time due to the aggregation, a relationship error, cloud mask uncertainty, etc. When aggregating from L2 to daily L3, additional sources of uncertainty enter through the gridding process as $\mathrm{IST}_{\mathrm{skin\_L3}}$ can only be retrieved for clear-sky pixels. This introduces a temporal and spatial sampling uncertainty. If all our satellite observations were obtained during all-sky conditions, we assume that the high polar temporal coverage is such that the temporal sampling uncertainty in the L3 files can be set to zero. However, this is not the case, and using only clear-sky observations generally leads to a clear-sky bias in averaged $\mathrm{IST}_{\mathrm{skin}}$ satellite observations when compared to in situ observations (Hall et al., 2012; Nielsen-Englyst et al., 2019; Rasmussen et al., 2018). The relationship error represents the standard deviation of the residuals calculated at in situ stations, where both skin and air temperatures are available, i.e. $\mathrm{T2m}_{\mathrm{sat}}$–$\mathrm{T2m}_{\mathrm{insitu}}$. Estimating all the different components that make up the $\mu_{\mathrm{local\_T2m}}$ is a very challenging task and is out of the scope of this paper. Instead, we estimate the $\mu_{\mathrm{local\_T2m}}$ component using a simple regression model fitted to the satellite-derived T2m and in situ T2m differences. Separate models have been chosen for the land ice and sea ice, due to the differences in the error characteristics. The variables to include in the uncertainty regression models have been chosen from a careful examination of the matchup dataset. For land ice and sea ice the most relevant variables were the $\mathrm{IST}_{\mathrm{skin\_L3}}$ itself and the number of 3 h time bins with observations in the L3, N_{bins}.

For land ice the regression model for $\mu_{\mathrm{local\_T2m}}$ is given as follows:

$$\mu_{\mathrm{local\_T2m\_landice}} = \beta_0 + \beta_1 \mathrm{IST}_{\mathrm{skin\_L3}} + \beta_2 N_{\mathrm{bins}}, \quad (10)$$

while the regression model for sea ice is given as

$$\mu_{\mathrm{local\_T2m\_seaice}} = \gamma_0 + \gamma_1 \mathrm{IST}_{\mathrm{skin\_L3}} + \gamma_2 \mathrm{IST}^2_{\mathrm{skin_{L3}}} + \gamma_3 N_{\mathrm{bins}}. \quad (11)$$

The coefficients have been determined by fitting to the $\mathrm{T2m}_{\mathrm{sat}}$–$\mathrm{T2m}_{\mathrm{insitu}}$ standard deviations calculated for the training data with $\mathrm{IST}_{\mathrm{skin\_L3}}$ bin intervals of 2 °C and a N_{bins} interval of 1. The $\mu_{\mathrm{rnd\_T2m}}$ and $\mu_{\mathrm{glob\_T2m}}$ components have been removed from the standard deviations in each bin as well as an assumed in situ uncertainty of 0.1 °C and an average sampling uncertainty of 0.5 °C (Høyer et al., 2017a; Reeves Eyre and Zeng, 2017) before fitting the regression models. The optimal regression coefficients for each domain are listed in Table 5.

4 Results

In Sect. 3.1, we selected the best (Eq. 12) of the five different algorithms and used it together with the derived coefficients (Tables 3 and 4) to retrieve T2m from satellite surface temperature estimates. The derived dataset consists of daily estimates of near-surface air temperature on a 0.25° regular latitude–longitude grid, during the period 2000–2009 (Høyer et al., 2018; Kennedy et al., 2019). Days with clouds and few clear-sky observations (as explained in Sect. 2.2) are not included in the dataset. However, for those days when the satellite-derived T2m product is available, it provides an estimate of the daily averaged all-sky T2m (see Sect. 5). Each temperature estimate is associated with three components of uncertainty on the 0.25° daily scale: a random uncertainty, a synoptic-scale correlated uncertainty, and a globally correlated uncertainty excluding uncertainties related to

Table 5. Uncertainty model regression coefficients.

Land ice	$\beta_0 = 3.82\,^{\circ}\mathrm{C}$	$\beta_1 = -0.24$	$\beta_2 = -0.03$	
Sea ice	$\gamma_0 = 2.01\,^{\circ}\mathrm{C}$	$\gamma_1 = -0.06$	$\gamma_2 = -0.12$	$\gamma_3 = -0.001$

the masking of clouds. The three types of uncertainties are also gathered in a total uncertainty estimate (see Sect. 3.2). The land ice temperatures have been calculated for grid cells categorized as ice sheet by the ETOPO1 global relief model (Amante and Eakins, 2009), averaged to the 0.25° grid. Sea ice temperatures have been calculated for grid cells with sea ice concentrations above 30 %, according to OSISAF (Tonboe et al., 2016).

4.1 Validation of $T2m_{sat}$

The derived $T2m_{sat}$ product has been validated against independent in situ data (i.e. the validation subset described in Sect. 3.1). Figure 6 shows an example of the daily near-surface air temperature coverage (from 1 January 2008). Circles are in situ T2m measurements from coincidence-independent AWSs and buoys, and there seems to be quite good agreement between these and $T2m_{sat}$ during this specific day. The overall model performance, when compared to all independent AWS and buoy observations, is summarized in Table 6. The satellite-derived air temperatures are about 0.3 °C warmer than measured in situ air temperature for both land ice and sea ice. For the GrIS, the bias is partly explained by topographic effects (see Sect. 4.3). The correlations are above 95 % for both surface types, and the RMSD is 3.47 and 3.20 °C for land and sea ice, respectively. Note that the uncertainty of the in situ data is also included in these RMSD values.

Figure 7 shows the average seasonal variation in bias and standard deviation for land ice and sea ice, respectively. For both land ice and sea ice, there is a seasonal dependency in standard deviation, with the largest values during the winter and smallest values during the summer. This is likely explained by a better cloud screening performance during sunlit periods (Karlsson and Dybbroe, 2010) and by the smaller natural thermal variability that is observed during summer conditions. Similar seasonality in performance is seen in five reanalysis products (including ERA-I/ERA5) for the GrIS (Zhang et al., 2021). As shown in Fig. 7, the average seasonal variation in bias is largest over sea ice, with the largest values in March and August. However, this seasonal tendency in bias over sea ice is only reflected at the beginning of the time period (i.e. 2000–2004). This can be seen in Fig. 8, which shows the seasonal averaged independent validation statistics for the entire period for land ice and sea ice. The figure also shows a quite stable performance over the time period for both land ice and sea ice.

As more satellite observations have become available over the time period, increased coverage of the surface temperature is expected over time. Figure 9 shows the average number of filled 3 h bins per day for the GrIS and Arctic sea ice for 2000–2009. Both surface types show an increase in filled 3 h bins over time, with large seasonal variations. In most years, sea ice has 1–1.5 filled bins per day more during winter than summer, due to a more extensive cloud cover over sea ice during summer (Curry et al., 1996; Beesley and Moritz, 1999). The GrIS typically has fewer filled bins per day during the winter and summer than spring and autumn, which is also explained by differences in cloud coverage (Griggs and Bamber, 2008). Note that the increase in the average number of filled 3 h bins from 2000 to 2009 is not reflected in the performance of the T2m product (Fig. 8).

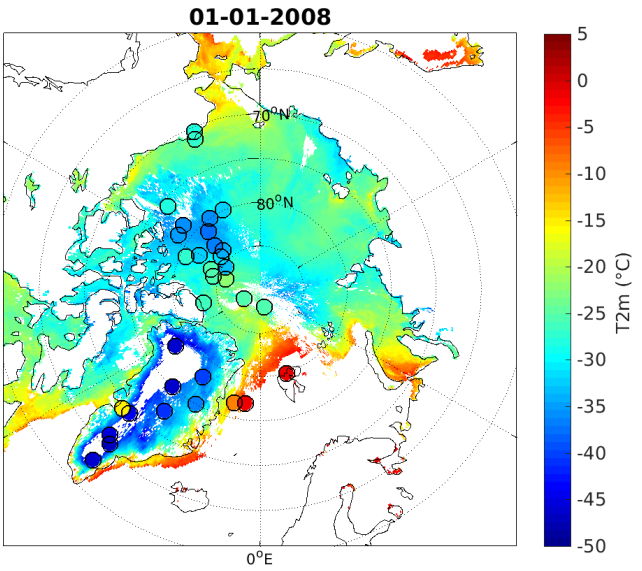


Figure 6. Daily mean 2 m air temperature over land ice and sea ice from 1 January 2008. Circles show in situ measurements.

Figure 10 shows $T2m_{sat}$–$T2m_{insitu}$ differences plotted as a function of AASTI L3 skin temperature for land ice and sea ice. Over land ice, the standard deviation decreases as a function of IST_{skin_L3}, while the bias is around zero for IST_{skin_L3} between −45 and −10 °C, positive for higher temperatures and negative for lower temperatures. For sea ice, the maximum standard deviation is found at skin temperatures of about −20 °C, with smaller standard deviations for higher and lower IST_{skin_L3}. Positive biases are found for very cold skin temperatures ($< -25\,^{\circ}\mathrm{C}$) and for temperatures around the melting point ($> -4\,^{\circ}\mathrm{C}$), while the intermediate temperatures have a slightly negative bias. This effect is included

Table 6. Statistics on the relation between satellite-derived and in situ measured temperatures for comparison with independent validation data. N: number of matchups used for validation; Corr: correlation; bias: $T2m_{sat}$–$T2m_{insitu}$ difference; SD: standard deviation; RMSD: root-mean-square difference.

	N	Corr (%)	Bias (°C)	SD (°C)	RMSD (°C)
Land ice	20 872	95.5	0.30	3.45	3.47
Sea ice	16 111	96.5	0.35	3.18	3.20

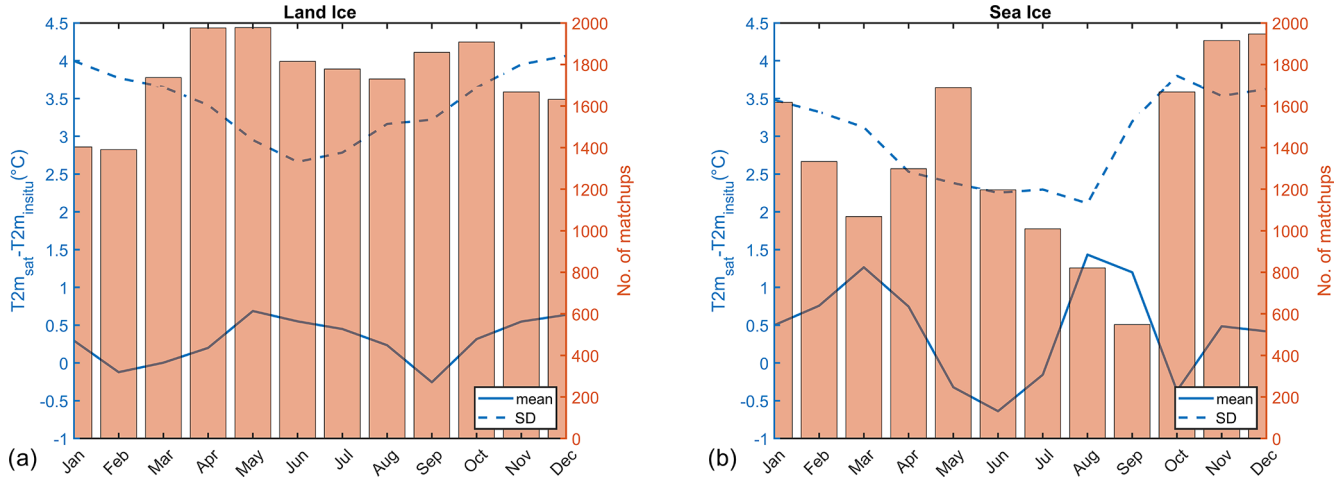


Figure 7. Estimated T2m minus observed T2m averaged for each month for **(a)** land ice and **(b)** sea ice. The dashed lines are standard deviations while the solid lines are biases. The bars show the average number of matchups for each month.

in the uncertainty estimates as presented in Sect. 3.2, which include IST_{skin_L3} as a predictor for both land ice and sea ice.

Figure 11 shows the validation results of the estimated uncertainties, where the $T2m_{sat}$–$T2m_{insitu}$ difference is plotted against the theoretical total uncertainties as obtained in Sect. 3.2 for land ice and sea ice. The dashed lines represent the ideal uncertainty with the assumptions that the in situ observations have an uncertainty of 0.1 °C and that the sampling uncertainty is 0.5 °C. The estimated uncertainties show good agreement with the observed uncertainties when the error bars follow the dashed line, which is the case here for both land ice and sea ice.

4.2 Comparison with reanalyses

The performance of $T2m_{sat}$ has been compared to the performance of T2m from ECMWF's reanalysis ERA-I ($T2m_{ERA\text{-}I}$; Dee et al., 2011) and the replacement reanalysis ERA5 ($T2m_{ERA5}$; Hersbach et al., 2020). Table 7 shows the performance of $T2m_{ERA\text{-}I}$ and $T2m_{ERA5}$ against the independent in situ T2m observations, which should be compared with the performance of the regression-derived $T2m_{sat}$ as shown in Table 6. The comparison may not be truly independent as a number of stations and buoys have been assimilated into the ERA-I and ERA5 data products (Dee et al., 2011; Hersbach et al., 2020), which would favour the reanalysis products in the comparison. Yet, the bias is significantly lower for $T2m_{sat}$ than for both $T2m_{ERA\text{-}I}$ and $T2m_{ERA\text{-}5}$, while the other validation parameters are similar, with slightly better correlation and standard deviation but slightly worse RMSD results for $T2m_{ERA}$. Previous studies have also found that ERA-I suffers from a consistent warm bias in the Arctic (Lüpkes et al., 2010; Jakobson et al., 2012; Vihma et al., 2002; Batrak and Müller, 2019; Simmons and Poli, 2014), and recent studies suggest that the warm bias still exists in ERA5 over sea ice (Wang et al., 2019; Graham et al., 2019). Similarly, recent studies found no significant improvements in 2 m temperatures over the GrIS for ERA5 compared to ERA-I (Delhasse et al., 2020; Zhang et al., 2021). Note, however, that the NCEP-CFSR, which is based on a coupled atmosphere–sea ice–ocean model, has shown better performance than ERA-I for near-surface atmospheric variables over sea ice (Jakobson et al., 2012).

Figure 12 shows the RMSD between in situ measured T2m and $T2m_{ERA\text{-}I}$ as well as $T2m_{ERA5}$ and $T2m_{sat}$ for the individual validation sites and both surface types. Due to the large number of buoys, these have been validated for each data source with all observations weighted equally. The last bars refer to the RMSD obtained by validating all validation sites in one long time series weighting all daily observations equally. The total $T2m_{sat}$ agrees better with in situ observations for both surface types compared to both ERA-I and ERA5. For most land ice stations, the $T2m_{sat}$ outper-

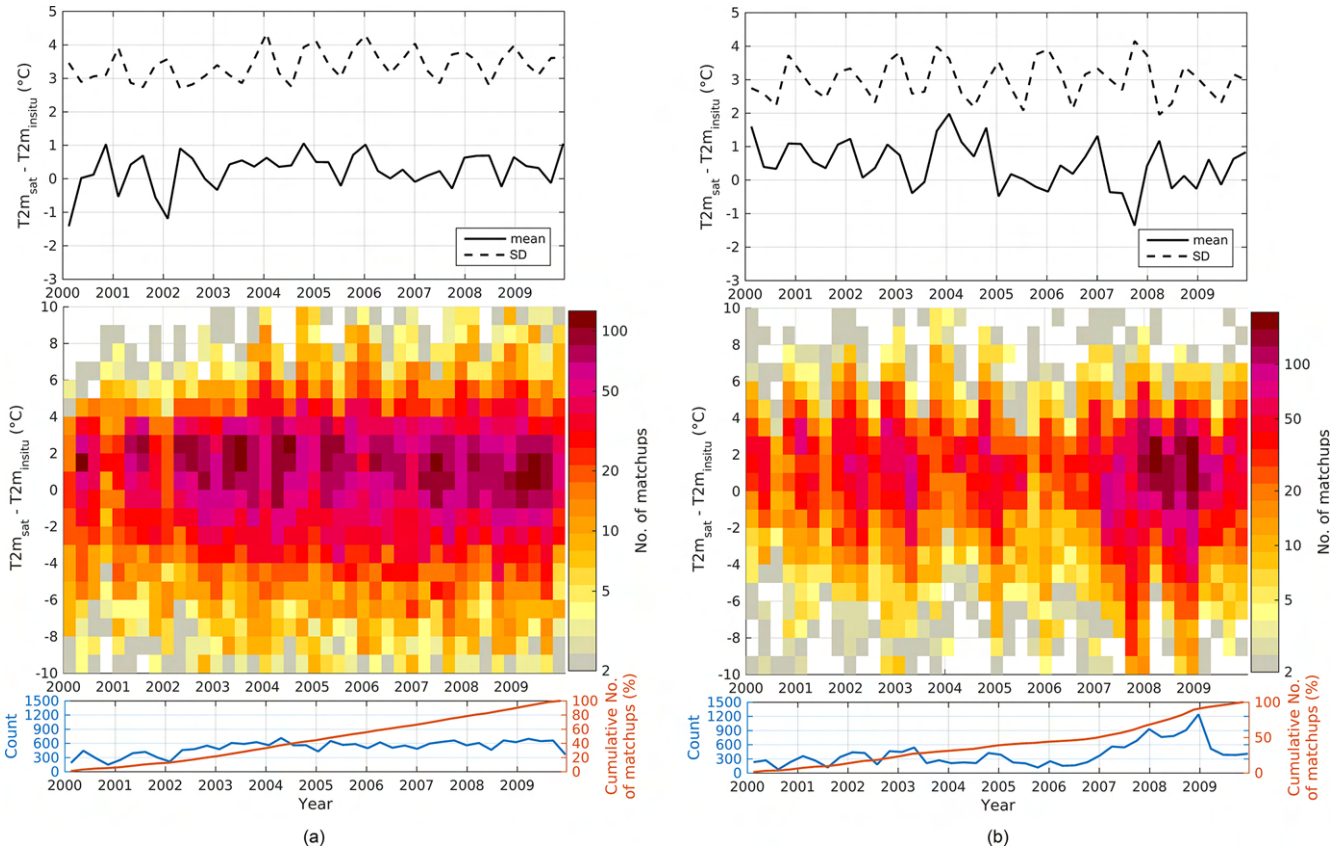


Figure 8. Estimated T2m minus observed T2m (bin size of 1 °C) for the full time period (bin size of 90 d) for **(a)** land ice and **(b)** sea ice. The dashed lines are standard deviations while the solid lines are bias in the upper figures. The surface plots in the middle figures show the number of matchups in each bin, while the bottom plots show the number of matchups (blue) and the cumulative percentage of matchups (red) in each time bin.

Table 7. Statistics on the relation between ERA-I/ERA5 and in situ measured temperatures for independent test data. N: number of matchups used for validation; Corr: correlation; bias: $T2m_{ERA}$–$T2m_{insitu}$ difference; SD: standard deviation; RMSD: root-mean-square difference.

	N		Corr (%)	Bias (°C)	SD (°C)	RMSD (°C)
Land ice	20 872	ERA-I	96.4	3.41	3.18	4.66
		ERA5	97.1	2.03	3.08	3.69
Sea ice	16 111	ERA-I	96.9	1.14	3.02	3.22
		ERA5	95.7	2.19	3.67	4.27

forms ERA-I and ERA5. One exception is the ARM station (BAR), where a bias of 2.49 °C gives rise to a relatively large RMSD for $T2m_{sat}$. This is likely explained by physical differences between the seasonal snow-covered sites and the GrIS sites, which are not fully captured by the regression model. ERA5 is significantly better than ERA-I over the GrIS, but ERA5 performs worse than both ERA-I and $T2m_{sat}$ over sea ice. Over sea ice, $T2m_{ERA\text{-}I}$ agrees better with in situ observations from the ECMWF data stream and *Polarstern*. However, these may be assimilated into both ERA-I and ERA5. The validation against *Polarstern* is relatively good even though the temperature measurements are made at 29 m height. This is likely because the data are mainly from the summer, when the vertical temperature gradients in the boundary layer are mostly small, and the performance of the cloud screening algorithm reaches its maximum. The independent in situ observations by ACSYS, CRREL, DAMOCLES, and FRAMZY are better reproduced by the satellite-derived T2m. The errors in the $T2m_{ERA\text{-}I}$/$T2m_{ERA5}$ and $T2m_{sat}$ datasets are expected to be independent and uncorrelated. For that reason, a combination of either $T2m_{ERA\text{-}I}$ or $T2m_{ERA5}$ and $T2m_{sat}$ can lead to an improved T2m estimate.

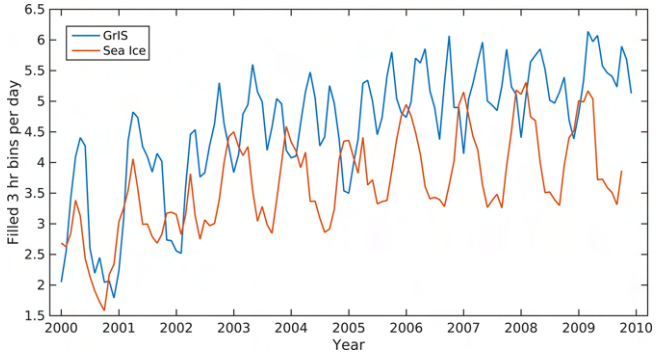


Figure 9. Average number of filled 3 h bins per day for the Greenland Ice Sheet and the Arctic sea ice.

4.3 Topographic effects

The effects from topography over the GrIS have been assessed by introducing a new matchup dataset that ensures that the elevation difference between satellite and in situ observations is less than 100 m over the GrIS. Excluding those AWSs (4 out of 23) with a larger elevation difference than 100 m results in a reduction of the training dataset of 2935 matchups (i.e. from GC-net_JAR1 and PROMICE TAS_U) and a reduction in the validation dataset of 560 matchups (i.e. from PROMICE QAS_U and UPE_U). The performance of the satellite-derived T2m improves the bias in particular, which decreases to 0.07 °C, while the standard deviation decreases to 3.41 °C over land ice. ERA-I and ERA5 show limited changes in performance, with slightly increased biases of 3.48 and 2.07 °C and standard deviations of 3.14 and 3.08 °C, respectively, when introducing the new matchup dataset over land ice. A similar good performance of the regression model is found when the two AWSs in the validation subset are kept. Despite the increased performance of the regression model, we have included all observations in the training of the model to ensure a robust and spatial representative solution.

4.4 Analysis of $T2m_{sat}$

The monthly mean $T2m_{sat}$ is shown in Fig. 13 for March, June, September and December averaged over the period 2000–2009. The interior and northern part of the GrIS is typically colder than other parts of the Arctic in all months, while the warmest regions are found along the sea ice marginal ice zone and the ablation zone of the GrIS. Limited spatial variability is seen over the Arctic sea ice during summer.

Figure 14 shows the monthly mean near-surface air temperature estimates averaged over the GrIS for the period 2000–2009. The GrIS records a distinct annual cycle in near-surface air temperature, with the maximum temperatures of around −4 °C during July and minimum temperatures of about −28 °C during winter. The range in monthly mean air temperature is in agreement with those reported by van As et al. (2011) at a number of PROMICE AWSs. The temporal variability is largest during winter due to a larger cloud radiative effect (compared to near-zero during summer) and a larger meridional temperature gradient resulting in a more vigorous atmospheric circulation in winter (Serreze et al., 1993). In addition, the temporal variability is lower during summer due to the fact that when the surface begins to melt, the sensible heat is used for melting and hence reducing surface air temperature variability (Steffen, 1995).

As illustrated in Fig. 15, $T2m_{sat}$ provides increasing coverage over the period 2000–2003 and quite stable coverage for the years 2003–2009. The average daily coverage is 84 % and 67 % for land ice and sea ice, respectively, for the stable 2003–2009 period and the 0.25° grid. When considering a 1° grid resolution, these numbers increase to 94 % and 81 %, respectively. Over land ice, the maximum coverage is during the spring and autumn, while the sea ice coverage has a clear drop in coverage during the summer due to increased cloud cover (Curry et al., 1996; Beesley and Moritz, 1999).

5 Discussion

Due to the limited number of in situ observations in the Arctic, and especially over sea ice, gathering in situ observations for testing and validating the regression models is not a simple task. The lack of observations that represent all conditions and regions in the Arctic and the resulting matching threshold of 15 km combined with the large topographical variations over the GrIS increase the uncertainty in the pixel-to-point comparison, thereby complicating the derivation and validation of the regression models. Despite this, the validation against independent in situ observations and the comparison with ERA-I and ERA5 demonstrate the value of the $T2m_{sat}$ product in the Arctic.

Five regression models were tested, and the best regression model predicts $T2m_{sat}$ from daily satellite IST_{skin_L3} combined with a seasonal variation. The performance of the $T2m_{sat}$ product did not improve much when the wind speed information from ERA-I or ERA5 (Table 3) was included despite the fact that previous studies have shown a strong dependency of wind speed for both land ice and sea ice (Adolph et al., 2018; Hudson and Brandt, 2005; Miller et al., 2013; Nielsen-Englyst et al., 2019). This was unexpected, at least for sea ice. The reason is likely that the quality of the wind speed fields is not adequate for use in the relationship model. In particular, accurately representing katabatic winds in numerical weather prediction (NWP) models is a challenging task due to the high resolution needed in the vertical direction (Grisogono et al., 2007; Steeneveld, 2014; Weng and Taylor, 2003; Zilitinkevich et al., 2006). Furthermore, the representation of surface roughness and the processes of snow–surface coupling, radiation, and turbulent mixing are hampered by limited resolution, while the relative importance of the processes varies with wind speed (Sterk et al., 2013). More accurate information on the wind speed

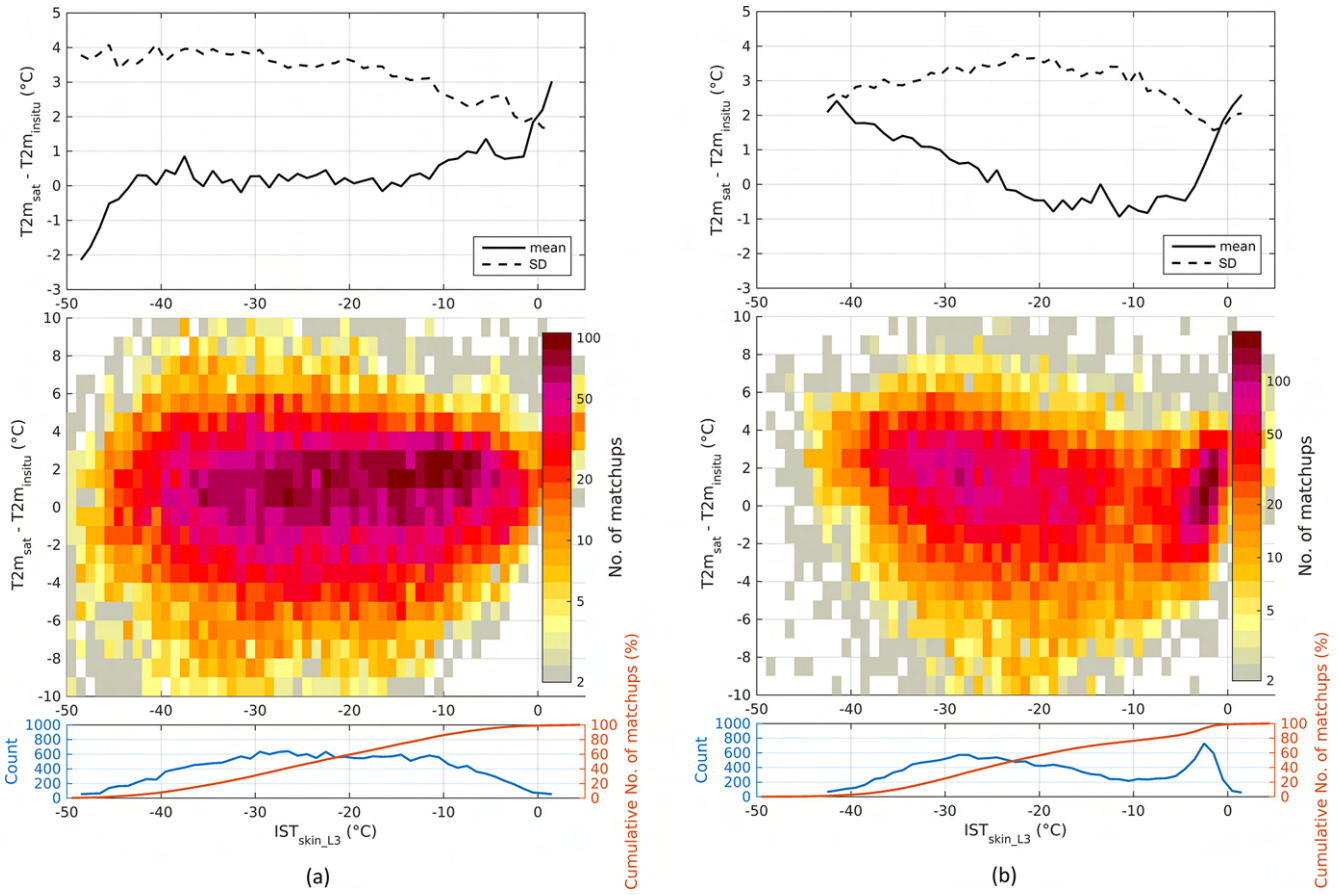


Figure 10. Estimated T2m minus observed T2m (bin size of 1 °C) as a function of binned (bin size of 1 °C) satellite IST_{skin_L3} for **(a)** land ice and **(b)** sea ice. The dashed lines are standard deviations while the solid lines are bias in the upper figure. The surface plots in the middle figures show the number of matchups in each bin while the bottom plots show the number of matchups (blue) and the cumulative percentage of matchups (red) in each IST_{skin_L3} bin.

is expected to improve the performance of the regression model, which includes wind speed as a predictor. In particular, the higher-resolution NWP output may be very beneficial in the regions of the GrIS where the local topography interacts with the wind through katabatic effects (DuVivier and Cassano, 2013; Oltmanns et al., 2015; Renfrew, 2004). Regional high-resolution reanalysis products are currently being developed within the Copernicus Arctic regional Reanalysis service C3S project (https://climate.copernicus.eu/copernicus-arctic-regional-reanalysis-service, last access: 29 June 2021). It is likely that such products will provide winds that can be used within a relationship model.

Since infrared satellites cannot measure the surface temperature during cloudy conditions, a cold clear-sky bias is often observed in infrared satellite IST_{skin_L3} averages compared to all-sky temperature averages (see, e.g. Table 2; Hall et al., 2008; Koenig and Hall, 2010). When using satellite IST_{skin_L3} observations, it is thus important to be aware of the clear-sky bias, which moreover varies with different temporal averaging windows (Nielsen-Englyst et al., 2019). Here, when using an empirical statistical method, which is trained against daily averaged in situ T2m (obtained in both clear-sky and cloudy conditions), the conversion from IST_{skin_L3} to $T2m_{sat}$ removes the systematic IST_{skin_L3} clear-sky bias effects that may be present in the satellite data. As a result, we obtain a $T2m_{sat}$ estimate which performs similarly or better than the IST_{skin_L3}, when validated against in situ observations. For the IST_{skin_L3}, the temporal sampling errors resulting from clouds have been minimized through a number of requirements. For short-lasting (< 24 h) cloudy conditions, the division into 3 h bin averages and the requirement of filled 3 h bins during both the night (between 18:00 and 06:00 LST) and day (between 06:00 and 18:00 LST) ensure that the diurnal cycle is best resolved despite the gaps with clouds. For long-lasting (≥ 24 h) cloudy conditions, IST_{skin_L3} is not available, and we do not retrieve $T2m_{sat}$ for these days.

The $T2m_{sat}$ product derived here provides increasing coverage over the period 2000–2003 and stable coverage for 2003–2009. The coverage varies with the season, with the minimum coverage over sea ice in the period from July to

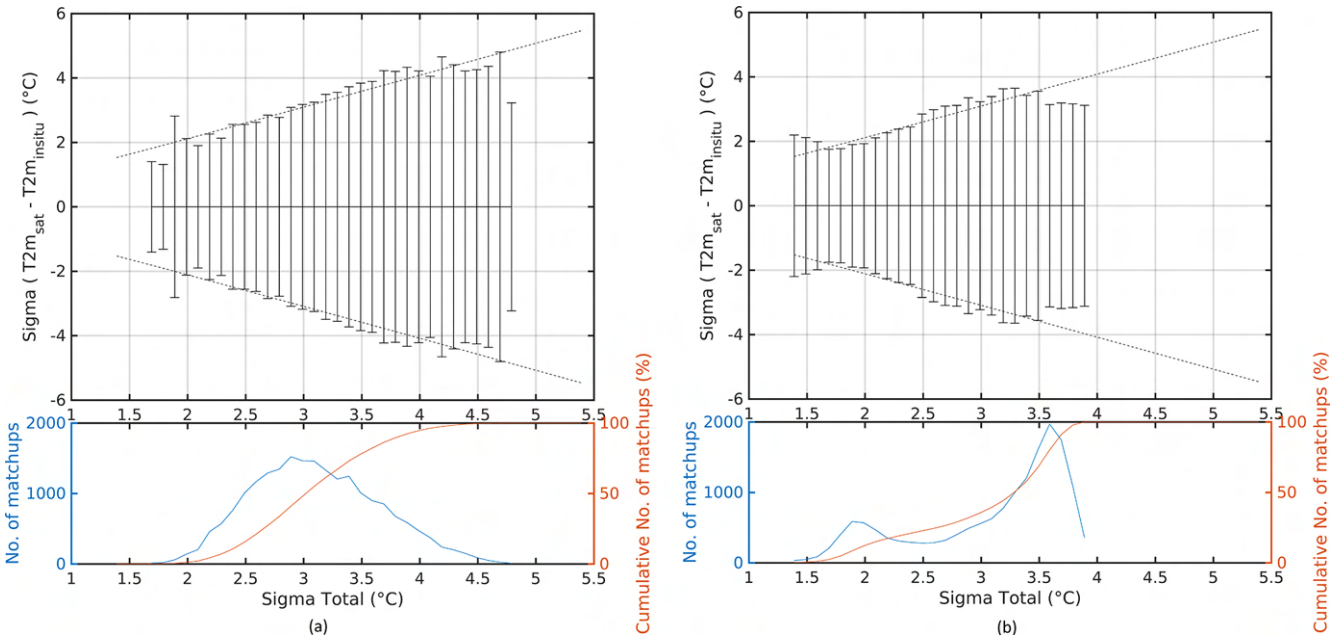


Figure 11. Satellite-estimated T2m uncertainty validation with respect to independent in situ T2m for **(a)** land ice and **(b)** sea ice. Dashed lines show the modelled uncertainty accounting for uncertainties in the in situ T2m and the sampling error. Solid black lines show 1 standard deviation of the estimated minus in situ differences for each 0.1 °C bin. The bottom plots show the number of matchups (blue) and the cumulative percentage of matchups for each bin (red).

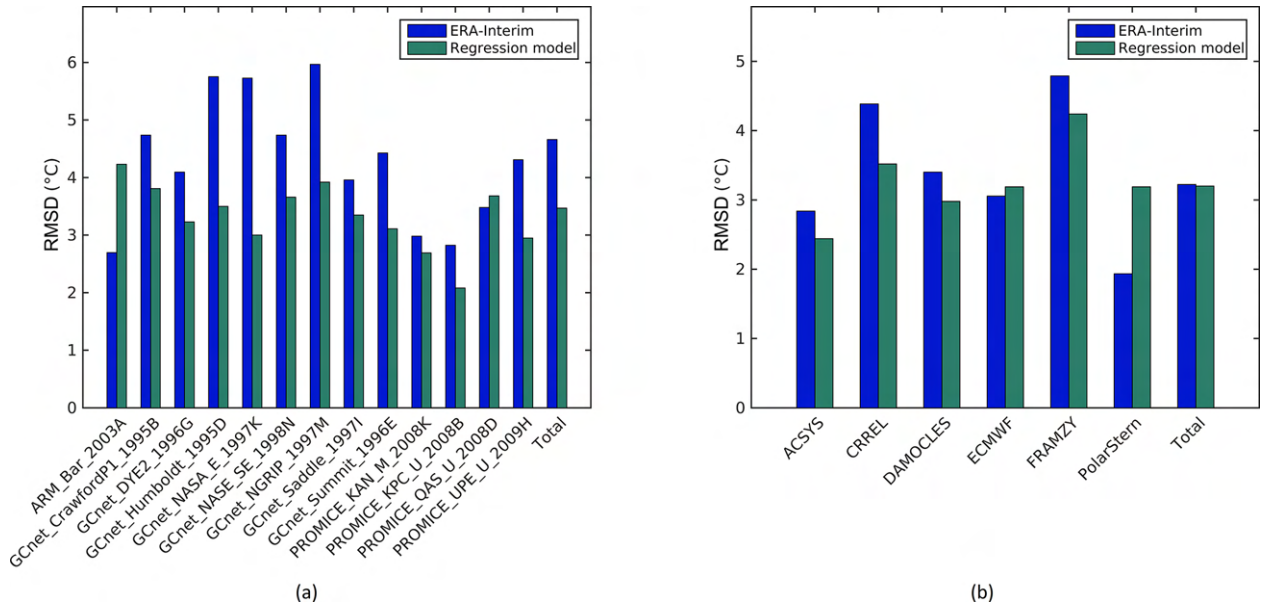


Figure 12. Root-mean-square differences (RMSDs) calculated for the **(a)** land ice sites and **(b)** sea ice sites using T2m from ERA-Interim, ERA5, and the regression model, respectively. Only buoys with more than 200 observations are included. The last two bars listed as "total" are the RMSD obtained by using all validation data.

September due to extensive cloud cover over the Arctic sea ice during summer (Curry et al., 1996; Beesley and Moritz, 1999). Nevertheless, the average daily coverage is 84 % and 67 % for land ice and sea ice, respectively, for the stable 2003–2009 period. The high percentages in coverage demonstrate that the gaps due to cloudy days are limited (except for over sea ice in the summer) and that the dataset contains a significant amount of information on the all-sky daily T2m even though it is based on clear-sky satellite observations.

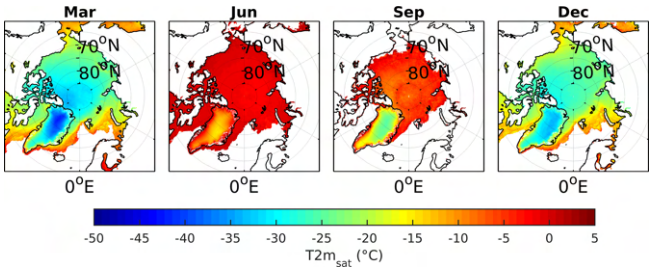


Figure 13. Monthly mean $T2m_{sat}$ during March, June, September, and December, averaged for the period 2000–2009.

Atmospheric models using data assimilation or statistical techniques may be applied to fill in the gaps due to clouds. This has already been done in the EUSTACE project by using an advanced statistical model to combine in situ observed and clear-sky satellite-derived T2m estimates (over land, lakes, ocean, and ice), including uncertainty estimates, into a global and gap-free daily analysis of surface air temperatures from 1850 to 2015 (Morice et al., 2019; Rayner et al., 2020). The $T2m_{sat}$ product derived in this paper is used as input to the EUSTACE surface air temperature analysis for the GrIS and the Arctic sea ice.

The $T2m_{sat}$ dataset developed here only covers Arctic, but the AASTI satellite dataset also covers the Antarctica. This implies that similar statistical methods can be derived for the Antarctic ice sheet and sea ice. Preliminary investigations indicate that a T2m product can be derived for the Antarctic ice sheet with similar performance to GrIS, whereas the Southern Ocean sea ice is challenging due to very few in situ observations (Morice et al., 2012). For both southern regions, more in situ observations are needed to repeat the work performed for the Arctic and to determine a reliable statistical model. This product can also be extended to seasonal snow and ice, but it requires a dynamic surface mask and the derivation of the regression model to be repeated. However, similar efforts have already been made within EUSTACE to cover seasonal snow (Good, 2015; Morice et al., 2019; Rayner et al., 2020).

The AASTI version builds on the Clara version 1 dataset from the CM-SAF. A version 2 of the dataset is now available (Karlsson et al., 2017), which facilitates the production of an AASTI version 2 dataset that covers the period 1982 up to present. With consistency in the retrieval algorithm and datasets, it will be possible to use the relationship model to produce a satellite-based climate data record of T2m from 1982 to today.

Including other available satellite products such as MODIS IST observations (Hall et al., 2004) or the (A)ATSR dataset (Ghent et al., 2017) may improve the quality of the $T2m_{sat}$ product. However, adding new data requires detailed knowledge of the characteristics of the dataset such as sampling frequency and uncertainty of the IST observations. In addition, determination of the relationship model is needed again. At the same time, adding more satellite overpasses to the daily estimates may not reduce the uncertainty of the products. This is evident when comparing Figs. 7 and 8 where the variation in the number of satellite observations during the record (Fig. 8) is not reflected in a similar variation in the performance of the product (Fig. 7). The uncertainty in the beginning of the record is comparable to the uncertainty at the end of the record, despite an almost doubling of the observed 3-hourly averages throughout the day.

6 Conclusions

The surface air temperature is one of the key indicators for Arctic climate change, and it can easily be compared with climate change indicators from other regions. This study introduces a methodology for using satellite skin temperatures for estimating air temperatures to compensate for the lack of in situ measurements and as a supplement to reanalysis products in the Arctic. Daily near-surface air temperatures (T2m) have been estimated based on daily clear-sky satellite Level 3 (L3) observations of ice surface skin temperatures (IST_{skin_L3}), using the Arctic and Antarctic ice Surface Temperatures from thermal Infrared satellite sensors (AASTI) reanalysis. A regression-based method has been used and tuned against in situ observed T2m using IST_{skin_L3} observations covering both Arctic sea ice and the Greenland Ice Sheet (GrIS). In general, there is a good correlation between T2m and IST_{skin_L3} due to the seasonal cycle in both IST and T2m. Different models have been tested to examine how to best capture the variability in the T2m–IST difference. The highest correlation and lowest RMSDs were found using a model where $T2m_{sat}$ is predicted from daily satellite IST_{skin_L3} combined with a seasonal variation, assumed to have the shape of an annual harmonic. This model has been used to derive daily T2m on a 0.25° regular latitude–longitude grid from the clear-sky AASTI IST_{skin_L3} over the Arctic during the time period 2000–2009 (Kennedy et al., 2019), using different regression coefficients for land ice and sea ice. Days with clouds or limited clear-sky observations have been excluded from the analysis. Considering a 1° regular latitude–longitude grid, the average daily coverage of the $T2m_{sat}$ product is 94 % over the GrIS and 81 % for sea ice for the years 2003–2009. The days when the $T2m_{sat}$ is available, the T2m estimate can be considered a daily averaged all-sky T2m, since it has been tuned against all-sky in situ observations.

The estimated $T2m_{sat}$ data show average biases of 0.30 and 0.35 °C and average root-mean-square errors of 3.47 and 3.20 °C for land ice and sea ice, respectively, when validated against independent in situ observations. All daily $T2m_{sat}$ estimates include a total uncertainty estimate divided into a random, locally systematic, and large-scale systematic uncertainty component. The total uncertainty of $T2m_{sat}$ shows good validation results when validated against independent in situ observations. A comparison with two of ECMWF's

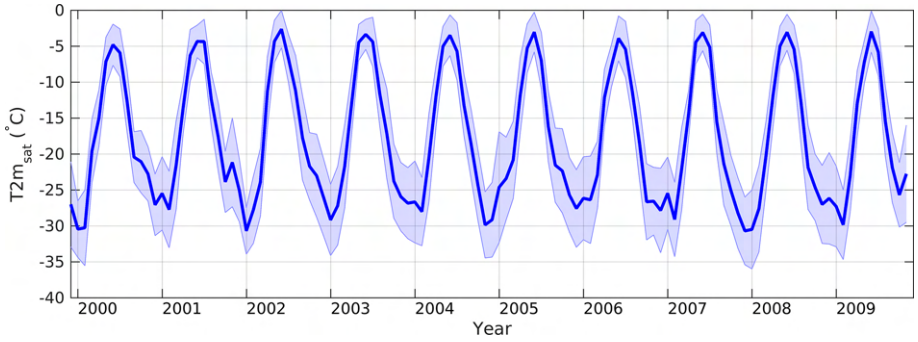


Figure 14. Monthly mean $T2m_{sat}$ for the Greenland Ice Sheet. The shading represents the variability.

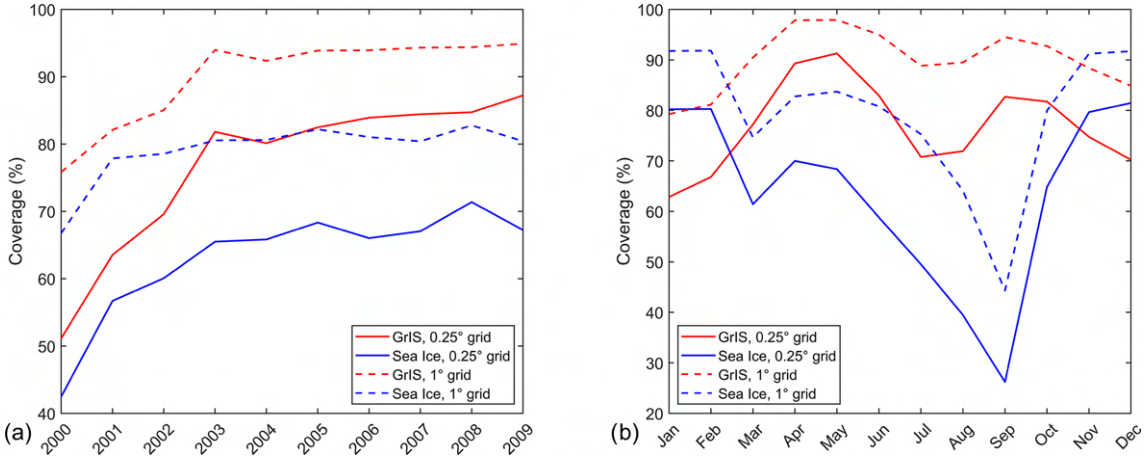


Figure 15. $T2m_{sat}$ coverage averaged for **(a)** each year and **(b)** each month for the GrIS and sea ice, and using a grid resolution of 0.25° and 1°, respectively.

reanalyses (i.e. ERA-I and ERA5) shows that $T2m_{sat}$ validates similarly or better than both of these even though the reanalyses actively assimilate available in situ observations. The $T2m_{sat}$ product is independent of the quality of the NWP forecasts, and thus it represents an important supplement to the model-based T2m. The errors in NWP products (e.g. $T2m_{ERA\text{-}I}$ or $T2m_{ERA5}$) and the errors in the product derived here ($T2m_{sat}$) are expected to be independent and uncorrelated, and a combination of a NWP product and the $T2m_{sat}$ data can therefore lead to an even better T2m estimate. The regression models presented here both work on satellite observations that are available from reprocessed records but open up for a near-real-time estimation of T2m from satellites. The results obtained for the ice-covered areas show that there is a large potential for using satellite-observed surface temperatures to estimate near-surface air temperatures. These estimates are not supposed to replace the already existing air temperature measurements or reanalyses, but rather to supplement these in particular in areas where no in situ observations are currently available.

Data availability. The derived surface air temperatures from satellite surface skin temperatures over ice can be downloaded from https://doi.org/10.5285/f883e197594f4fbaae6edebafb3fddb3 (Kennedy et al., 2019). The PROMICE data can be accessed through http://www.promice.dk (last access: 16 November 2018, https://doi.org/10.22008/promice/data/aws, Fausto and van As, 2019). The ARM data are available at https://www.archive.arm.gov/discovery/#v/results/s/s::co (last access: 21 December 2018, https://doi.org/10.5439/1025220, ARM Archive, 2018). GC-Net data can be found through https://doi.org/10.5067/6S7UHUH2K5RI (Greenland Climate Network (GC-Net) Radiation for Arctic System Reanalysis, Version 1., 2016). Data from CRREL mass balance buoys are available from http://imb-crrel-dartmouth.org (The CRREL-Dartmouth Mass Balance Buoy Program, 2016), while *Polarstern* data can be downloaded at https://dship.awi.de/Polarstern.html (last access: 24 November 2016, https://doi.org/10.1594/PANGAEA.761654, König-Langlo et al., 2006b). FRAMZY data are available from https://doi.org/10.1594/WDCC/UNI_HH_MI_FRAMZY2002 (Brümmer et al., 2012b), https://doi.org/10.1594/WDCC/UNI_HH_MI_FRAMZY2007 (Brümmer et al., 2011b), and https://doi.org/10.1594/WDCC/UNI_HH_MI_FRAMZY2008 (Brümmer et al., 2011c), while ACSYS data are found here: https://doi.org/10.1594/WDCC/UNI_HH_MI_ACSYS2003.

Damocles data can be found here: https://doi.org/10.1594/wdcc/uni_HH_MI_DAMOCLES2007 (Brümmer et al., 2011a). The traditional buoy and ship data obtained from ECMWF are distributed through the World Meteorological Organization's (WMO) Global Telecommunication System (GTS) and available for members at the ECMWF Meteorological Archival and Retrieval System (MARS). Finally, the AASTI IST_{skin_L2} data are available from https://doi.org/10.5285/60b820fa10804fca9c3f1ddfa5ef42a1 (Høyer et al., 2019).

Author contributions. PNE, KSM, and GD compiled and quality-checked the in situ data. PNE, JLH, and KSM designed and developed the regression model and estimated uncertainties. GD, JLH, and RT developed the AASTI IST_{skin_L2} data. SS did the ERA5 matchup. PNE prepared the manuscript with contributions from all authors.

Competing interests. The authors declare that they have no conflict of interest.

Disclaimer. Publisher's note: Copernicus Publications remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Acknowledgements. This study was carried out as a part of the European Union Surface Temperatures for All Corners of Earth (EUSTACE), which is financed by the European Union's Horizon 2020. The authors would also like to thank the data providers.

Financial support. This research has been supported by the Horizon 2020 (EUSTACE (grant no. 640171)).

Review statement. This paper was edited by Chris Derksen and reviewed by Emma Dodd, Christopher J. Merchant, and Timo Vihma.

References

Abermann, J., Hansen, B., Lund, M., Wacker, S., Karami, M., and Cappelen, J.: Hotspots and key periods of Greenland climate change during the past six decades, Ambio, 46, 3–11, https://doi.org/10.1007/s13280-016-0861-y, 2017.

Ackerman, T. P. and Stokes, G. M.: The Atmospheric Radiation Measurement Program, Phys. Today, 56, 38–44, https://doi.org/10.1063/1.1554135, 2003.

Adolph, A. C., Albert, M. R., and Hall, D. K.: Near-surface temperature inversion during summer at Summit, Greenland, and its relation to MODIS-derived surface temperatures, The Cryosphere, 12, 907–920, https://doi.org/10.5194/tc-12-907-2018, 2018.

Ahlstrøm, A., van As, D., Citterio, M., Andersen, S., Fausto, R., Andersen, M., Forsberg, R., Stenseng, L., Lintz Christensen, E., and Kristensen, S. S.: A new Programme for Monitoring the Mass Loss of the Greenland Ice Sheet, Geol. Surv. Den. Greenl., 15, 61–64, 2008.

Amante, C. and Eakins, B. W.: ETOPO1 Global Relief Model converted to PanMap layer format, PANGAEA, https://doi.org/10.1594/PANGAEA.769615, 2009.

Atmospheric Radiation Measurement (ARM) Archive: ARM-standard Meteorological Instrumentation at Surface, https://doi.org/10.5439/1025220, 2018.

Batrak, Y. and Müller, M.: On the warm bias in atmospheric reanalyses induced by the missing snow over Arctic sea-ice, Nat. Commun., 10, 4170, https://doi.org/10.1038/s41467-019-11975-3, 2019.

Beesley, J. A. and Moritz, R. E.: Toward an explanation of the annual cycle of cloudiness over the Arctic Ocean, J. Climate, 12, 395–415, 1999.

Box, J. E., Colgan, W. T., Christensen, T. R., Schmidt, N. M., Lund, M., Parmentier, F.-J. W., Brown, R., Bhatt, U. S., Euskirchen, E. S., Romanovsky, V. E., Walsh, J. E., Overland, J. E., Wang, M., Corell, R. W., Meier, W. N., Wouters, B., Mernild, S., Mård, J., Pawlak, J., and Olsen, M. S.: Key indicators of Arctic climate change: 1971–2017, Environ. Res. Lett., 14, 045010, https://doi.org/10.1088/1748-9326/aafc1b, 2019.

Brümmer, B., Müller, G., Haller, M., Kriegsmann, A., Offermann, M., and Wetzel, C.: DAMOCLES 2007–2008 – Hamburg Arctic Ocean Buoy Drift Experiment: meteorological measurements of 16 autonomous drifting ice buoys, World Data Center for Climate (WDCC) at DKRZ, https://doi.org/10.1594/wdcc/uni_HH_MI_DAMOCLES2007, 2011a.

Brümmer, B., Müller, G., Lammert-Stockschläder, A., Jahnke-Bornemann, A., and Wetzel, C.: FRAMZY 2007 – Third Field Experiment on Fram Strait Cyclones and their Impact on Sea Ice: meteorological measurements of the research aircraft Falcon, 16 autonomous ice buoys and 13 autonomous water buoys, World Data Center for Climate (WDCC) at DKRZ, https://doi.org/10.1594/WDCC/UNI_HH_MI_FRAMZY2007, 2011b.

Brümmer, B., Müller, G., and Wetzel, C.: FRAMZY 2008 – Fourth Field Experiment on Fram Strait Cyclones and their Impact on Sea Ice: meteorological measurements of 7 autonomous ice buoys, World Data Center for Climate (WDCC) at DKRZ, https://doi.org/10.1594/WDCC/UNI_HH_MI_FRAMZY2008, 2011c.

Brümmer, B., Launiainen, J., Müller, G., Kirchgaessner, A., and Wetzel, C.: ACSYS 2003 – Arctic Atmospheric Boundary Layer and Sea Ice Interaction Study north of Spitsbergen: meteorological measurements of the research aircraft Falcon, 11 autonomous ice buoys and radiosoundings at the research vessels Aranda and Polarstern, World Data Center for Climate (WDCC) at DKRZ, https://doi.org/10.1594/WDCC/UNI_HH_MI_ACSYS2003, 2012a.

Brümmer, B., Launiainen, J., Müller, G., and Wetzel, C.: FRAMZY 2002 – Second Field Experiment on Fram Strait Cyclones and their Impact on Sea Ice: meteorological measurements of the research aircraft Falcon, 15 autonomous ice buoys and radiosoundings at the research vessel Aranda, World Data Center for Climate (WDCC) at DKRZ,

https://doi.org/10.1594/WDCC/UNI_HH_MI_FRAMZY2002, 2012b.

Bulgin, C. E., Embury, O., and Merchant, C. J.: Sampling uncertainty in gridded sea surface temperature products and Advanced Very High Resolution Radiometer (AVHRR) Global Area Coverage (GAC) data, Remote Sens. Environ., 177, 287–294, https://doi.org/10.1016/j.rse.2016.02.021, 2016.

Cappelen, J. (Ed.): Greenland – DMI Historical Climate Data Collection 1768-2020., DMI Rep. 21–02, Copenhagen, Denmark, Danish Meteorological Institute, Copenhagen, Denmark, 2021.

Collins, M., Knutti, R., Arblaster, J., Dufresne, J.-L., Fichefet, T., Friedlingstein, P., Gao, X., Gutowski, W. J., Johns, T., Krinner, G., Shongwe, M., Tebaldi, C., Weaver, A. J., and Wehner, M.: Long-term Climate Change: Projections, Commitments and Irreversibility, in: Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change, edited by: Stocker, T. F., Qin, D., Plattner, G.-K., Tignor, M., Allen, S. K., Boschung, J., Nauels, A., Xia, Y., Bex, V., and Midgley, P. M., Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 1029–1136, https://doi.org/10.1017/CBO9781107415324.024, 2013.

Cowtan, K. and Way, R.: Update to "Coverage bias in the HadCRUT4 temperature series and its impact on recent temperature trends", Reconciling global temperature series, https://doi.org/10.13140/RG.2.1.4334.8564, 2014.

Curry, J. A., Schramm, J. L., Rossow, W. B., and Randall, D.: Overview of Arctic Cloud and Radiation Characteristics, J. Climate, 9, 1731–1764, https://doi.org/10.1175/1520-0442(1996)009<1731:OOACAR>2.0.CO;2, 1996.

Davy, R. and Outten, S.: The Arctic Surface Climate in CMIP6: Status and Developments since CMIP5, J. Climate, 33, 8047–8068, https://doi.org/10.1175/JCLI-D-19-0990.1, 2020.

Dee, D. P., Uppala, S. M., Simmons, A. J., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M. A., Balsamo, G., Bauer, P., Bechtold, P., Beljaars, A. C. M., van de Berg, L., Bidlot, J., Bormann, N., Delsol, C., Dragani, R., Fuentes, M., Geer, A. J., Haimberger, L., Healy, S. B., Hersbach, H., Hólm, E. V., Isaksen, L., Kållberg, P., Köhler, M., Matricardi, M., McNally, A. P., Monge-Sanz, B. M., Morcrette, J.-J., Park, B.-K., Peubey, C., de Rosnay, P., Tavolato, C., Thépaut, J.-N., and Vitart, F.: The ERA-Interim reanalysis: configuration and performance of the data assimilation system, Q. J. Roy. Meteor. Soc., 137, 553–597, https://doi.org/10.1002/qj.828, 2011.

Delhasse, A., Kittel, C., Amory, C., Hofer, S., van As, D., S. Fausto, R., and Fettweis, X.: Brief communication: Evaluation of the near-surface climate in ERA5 over the Greenland Ice Sheet, The Cryosphere, 14, 957–965, https://doi.org/10.5194/tc-14-957-2020, 2020.

DuVivier, A. K. and Cassano, J. J.: Evaluation of WRF Model Resolution on Simulated Mesoscale Winds and Surface Fluxes near Greenland, Mon. Weather Rev., 141, 941–963, https://doi.org/10.1175/MWR-D-12-00091.1, 2013.

Dybbroe, A., Karlsson, K.-G., and Thoss, A.: NWCSAF AVHRR Cloud Detection and Analysis Using Dynamic Thresholds and Radiative Transfer Modeling. Part I: Algorithm Description, J. Appl. Meteorol., 44, 39–54, https://doi.org/10.1175/JAM-2188.1, 2005a.

Dybbroe, A., Karlsson, K.-G., and Thoss, A.: NWCSAF AVHRR Cloud Detection and Analysis Using Dynamic Thresholds and Radiative Transfer Modeling. Part II: Tuning and Validation, J. Appl. Meteorol., 44, 55–71, https://doi.org/10.1175/JAM-2189.1, 2005b.

Dybkjær, G., Tonboe, R., and Høyer, J. L.: Arctic surface temperatures from Metop AVHRR compared to in situ ocean and land data, Ocean Sci., 8, 959–970, https://doi.org/10.5194/os-8-959-2012, 2012.

Dybkjær, G., Høyer, J. L., Tonboe, R., and Olsen, S. M.: Report on the documentation and description of the new Arctic Ocean dataset combining SST and IST, NACLIM Deliverable, D32.28, 2014.

Dybkjær, G., Eastwood, S., Borg, A. L., Høyer, J. L., and Tonboe, R.: Algorithm theoretical basis document (ATBD) for the OSI SAF Sea and Sea Ice Surface Temperature L2 processing chain, OSI205a and b, http://osisaf.met.no/docs/osisaf_cdop2_ss2_pum_ice-conc_v1p4.pdf, last access: 19 February 2018.

Fausto, R. S. and van As, D.: Programme for monitoring of the Greenland ice sheet (PROMICE): Automatic weather station data, Version: v03, Dataset published via Geological Survey of Denmark and Greenland, https://doi.org/10.22008/promice/data/aw, 2019.

Ghent, D. J., Corlett, G. K., Göttsche, F.-M., and Remedios, J. J.: Global Land Surface Temperature From the Along-Track Scanning Radiometers: Global LST from the ATSRs, J. Geophys. Res.-Atmos., 122, 12167–12193, https://doi.org/10.1002/2017JD027161, 2017.

GHRSST Science Team: The Recommended GHRSST Data Specification (GDS) 2.0, document revision 4, available from the GHRSST International Project Office, 2011, 123 pp., 2010.

Good, E.: Daily minimum and maximum surface air temperatures from geostationary satellite data: Satellite min and max air temperatures, J. Geophys. Res.-Atmos., 120, 2306–2324, https://doi.org/10.1002/2014JD022438, 2015.

Good, E. J., Ghent, D. J., Bulgin, C. E., and Remedios, J. J.: A spatiotemporal analysis of the relationship between near-surface air temperature and satellite land surface temperatures using 17 years of data from the ATSR series, J. Geophys. Res.-Atmos., 122, 9185–9210, https://doi.org/10.1002/2017JD026880, 2017.

Graham, R. M., Cohen, L., Ritzhaupt, N., Segger, B., Graversen, R. G., Rinke, A., Walden, V. P., Granskog, M. A., and Hudson, S. R.: Evaluation of Six Atmospheric Reanalyses over Arctic Sea Ice from Winter to Early Summer, J. Climate, 32, 4121–4143, https://doi.org/10.1175/JCLI-D-18-0643.1, 2019.

Graversen, R. G., Mauritsen, T., Tjernström, M., Källén, E., and Svensson, G.: Vertical structure of recent Arctic warming, Nature, 451, 53–56, https://doi.org/10.1038/nature06502, 2008.

Griggs, J. A. and Bamber, J. L.: Assessment of Cloud Cover Characteristics in Satellite Datasets and Reanalysis Products for Greenland, J. Climate, 21, 1837–1849, https://doi.org/10.1175/2007JCLI1570.1, 2008.

Grisogono, B., Kraljević, L., and Jeričević, A.: The low-level katabatic jet height versus Monin-Obukhov height, Q. J. Roy. Meteor. Soc., 133, 2133–2136, https://doi.org/10.1002/qj.190, 2007.

Hall, D., Box, J., Casey, K., Hook, S., Shuman, C., and Steffen, K.: Comparison of satellite-derived and in-situ observations of ice and snow surface temperatures over

Greenland, Remote Sens. Environ., 112, 3739–3749, https://doi.org/10.1016/j.rse.2008.05.007, 2008.

Hall, D. K., Key, J. R., Case, K. A., Riggs, G. A., and Cavalieri, D. J.: Sea ice surface temperature product from MODIS, IEEE T. Geosci. Remote, 42, 1076–1087, https://doi.org/10.1109/TGRS.2004.825587, 2004.

Hall, D. K., Comiso, J. C., DiGirolamo, N. E., Shuman, C. A., Key, J. R., and Koenig, L. S.: A Satellite-Derived Climate-Quality Data Record of the Clear-Sky Surface Temperature of the Greenland Ice Sheet, J. Climate, 25, 4785–4798, https://doi.org/10.1175/JCLI-D-11-00365.1, 2012.

Hanna, E., Cappelen, J., Fettweis, X., Mernild, S. H., Mote, T. L., Mottram, R., Steffen, K., Ballinger, T. J., and Hall, R. J.: Greenland surface air temperature changes from 1981 to 2019 and implications for ice-sheet melt and mass-balance change, Int. J. Climatol., 41, E1336–E1352, https://doi.org/10.1002/joc.6771, 2021.

Hansen, J., Ruedy, R., Sato, M., and Lo, K.: Global surface temperature change, Rev. Geophys., 48, RG4004, https://doi.org/10.1029/2010RG000345, 2010.

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., Nicolas, J., Peubey, C., Radu, R., Schepers, D., Simmons, A., Soci, C., Abdalla, S., Abellan, X., Balsamo, G., Bechtold, P., Biavati, G., Bidlot, J., Bonavita, M., Chiara, G., Dahlgren, P., Dee, D., Diamantakis, M., Dragani, R., Flemming, J., Forbes, R., Fuentes, M., Geer, A., Haimberger, L., Healy, S., Hogan, R. J., Hólm, E., Janisková, M., Keeley, S., Laloyaux, P., Lopez, P., Lupu, C., Radnoti, G., Rosnay, P., Rozum, I., Vamborg, F., Villaume, S., and Thépaut, J.: The ERA5 global reanalysis, Q. J. Roy. Meteor. Soc., 146, 1999–2049, https://doi.org/10.1002/qj.3803, 2020.

Holland, M. M. and Bitz, C. M.: Polar amplification of climate change in coupled models, Clim. Dynam., 21, 221–232, https://doi.org/10.1007/s00382-003-0332-6, 2003.

Høyer, J. L., Alerskans, E., Nielsen-Englyst, P., Thejll, P., Dybkjær, G., and Tonboe, R.: Detailed investigation of the uncertainty budget for non-recoverable IST observations and their SI traceability (ESA Tech. Rep. FRM4STS OP-70), available at: http://www.frm4sts.org/wp-content/uploads/sites/3/2018/08/OP-70-FRM4STS_option3_report_v1-signed.pdf (last access: 29 June 2021), 2017a.

Høyer, J. L., Lang, A. M., Tonboe, R., Eastwood, S., Wimmer, W., and Dybkjær, G.: Towards field intercomparison experiment (FICE) for ice surface temperature (ESA Tech. Rep. FRM4STS OP-40), available at: http://www.frm4sts.org/wp-content/uploads/sites/3/2017/12/OFE-OP-40-TR-5-V1-Iss-1-Ver-1-Signed.pdf (last access: 29 June 2021), 2017b.

Høyer, J. L., Good, E., Nielsen-Englyst, P., Madsen, K. S., Woolway, I., and Kennedy, J.: Report on the relationship between satellite surface skin temperature and surface air temperature observations for oceans, land, sea ice, ice sheets, and lakes, available at: https://www.eustaceproject.org/eustace/static/media/uploads/d1.5_revised.pdf (last access: 29 June 2021), 2018.

Høyer, J. L., Dybkjær, G., Eastwood, S., and Madsen, K. S.: EUSTACE/AASTI: Global clear-sky ice surface temperature data from the AVHRR series on the satellite swath with estimates of uncertainty components, v1.1, 2000–2009, Centre for Environmental Data Analysis, https://doi.org/10.5285/60b820fa10804fca9c3f1ddfa5ef42a1, 2019.

Hudson, S. R. and Brandt, R. E.: A Look at the Surface-Based Temperature Inversion on the Antarctic Plateau, J. Climate, 18, 1673–1696, https://doi.org/10.1175/JCLI3360.1, 2005.

IPCC: Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change, edited by: Stocker, T. F., Qin, D., Plattner, G.-K., Tignor, M., Allen, S. K., Boschung, J., Nauels, A., Xia, Y., Bex V., and Midgley, P. M., Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 1535 pp., https://doi.org/10.1017/CBO9781107415324, 2013.

Jakobson, E., Vihma, T., Palo, T., Jakobson, L., Keernik, H., and Jaagus, J.: Validation of atmospheric reanalyses over the central Arctic Ocean, Geophys. Res. Lett., 39, L10802, https://doi.org/10.1029/2012GL051591, 2012.

Jones, P. D., Lister, D. H., Osborn, T. J., Harpham, C., Salmon, M., and Morice, C. P.: Hemispheric and large-scale land-surface air temperature variations: An extensive revision and an update to 2010:, J. Geophys. Res.-Atmos., 117, D05127, https://doi.org/10.1029/2011JD017139, 2012.

Karlsson, K.-G. and Dybbroe, A.: Evaluation of Arctic cloud products from the EUMETSAT Climate Monitoring Satellite Application Facility based on CALIPSO-CALIOP observations, Atmos. Chem. Phys., 10, 1789–1807, https://doi.org/10.5194/acp-10-1789-2010, 2010.

Karlsson, K.-G., Riihelä, A., Müller, R., Meirink, J. F., Sedlar, J., Stengel, M., Lockhoff, M., Trentmann, J., Kaspar, F., Hollmann, R., and Wolters, E.: CLARA-A1: a cloud, albedo, and radiation dataset from 28 yr of global AVHRR data, Atmos. Chem. Phys., 13, 5351–5367, https://doi.org/10.5194/acp-13-5351-2013, 2013.

Karlsson, K.-G., Anttila, K., Trentmann, J., Stengel, M., Fokke Meirink, J., Devasthale, A., Hanschmann, T., Kothe, S., Jääskeläinen, E., Sedlar, J., Benas, N., van Zadelhoff, G.-J., Schlundt, C., Stein, D., Finkensieper, S., Håkansson, N., and Hollmann, R.: CLARA-A2: the second edition of the CM SAF cloud and radiation data record from 34 years of global AVHRR data, Atmos. Chem. Phys., 17, 5809–5828, https://doi.org/10.5194/acp-17-5809-2017, 2017.

Kennedy, J. J., Capponi, F., Ghent, D., Good, E. J., Høyer, J. L., Kent, E. C., Madsen, K. S., Mitchelson, J. R., Nielsen-Englyst, P., and Tonboe, R. T.: EUSTACE: Globally gridded clear-sky daily air temperature estimates from satellites with uncertainty estimates for land, ocean and ice, 1995–2016, Centre for Environmental Data Analysis, https://doi.org/10.5285/f883e197594f4fbaae6edebafb3fddb3, 2019.

Key, J. R., Collins, J. B., Fowler, C., and Stone, R. S.: High-latitude surface temperature estimates from thermal satellite data, Remote Sens. Environ., 61, 302–309, https://doi.org/10.1016/S0034-4257(97)89497-7, 1997.

Kindig, D.: Greenland Climate Network (GC-Net) Radiation for Arctic System Reanalysis, Version 1. Boulder, Colorado USA, NASA National Snow and Ice Data Center Distributed Active Archive Center, https://doi.org/10.5067/6S7UHUH2K5RI, 2010.

Knust, R.: Polar Research and Supply Vessel POLARSTERN operated by the Alfred-Wegener-Institute, Journal of Large-Scale Research Facilities JLSRF, 3, A119, https://doi.org/10.17815/jlsrf-3-163, 2017.

Koenig, L. S. and Hall, D. K.: Comparison of satellite, thermochron and air temperatures at Summit, Greenland, during the winter of 2008/09, J. Glaciol., 56, 735–741, https://doi.org/10.3189/002214310793146269, 2010.

König-Langlo, G., Loose, B., and Bräuer, B.: 25 Years of Polarstern Meteorology, WDC-MARE Rep., 4 (CD-ROM), 1–137, 2006a.

König-Langlo, G., Loose, B., and Bräuer, B.: 25 Years of Polarstern Meteorology, World Data Center for Marine Environmental Sciences, PANGAEA, https://doi.org/10.1594/PANGAEA.761654, 2006b.

Langen, P. L., Mottram, R. H., Christensen, J. H., Boberg, F., Rodehacke, C. B., Stendel, M., van As, D., Ahlstrøm, A. P., Mortensen, J., Rysgaard, S., Petersen, D., Svendsen, K. H., Aðalgeirsdóttir, G., and Cappelen, J.: Quantifying Energy and Mass Fluxes Controlling Godthåbsfjord Freshwater Input in a 5 km Simulation (1991–2012)., J. Climate, 28, 3694–3713, https://doi.org/10.1175/JCLI-D-14-00271.1, 2015.

Lavergne, T., Sørensen, A. M., Kern, S., Tonboe, R., Notz, D., Aaboe, S., Bell, L., Dybkjær, G., Eastwood, S., Gabarro, C., Heygster, G., Killie, M. A., Brandt Kreiner, M., Lavelle, J., Saldo, R., Sandven, S., and Pedersen, L. T.: Version 2 of the EUMETSAT OSI SAF and ESA CCI sea-ice concentration climate data records, The Cryosphere, 13, 49–78, https://doi.org/10.5194/tc-13-49-2019, 2019.

Lenssen, N. J. L., Schmidt, G. A., Hansen, J. E., Menne, M. J., Persin, A., Ruedy, R., and Zyss, D.: Improvements in the GISTEMP Uncertainty Model, J. Geophys. Res.-Atmos., 124, 6307–6326, https://doi.org/10.1029/2018JD029522, 2019.

Lindsay, R., Wensnahan, M., Schweiger, A., and Zhang, J.: Evaluation of Seven Different Atmospheric Reanalysis Products in the Arctic, J. Climate, 27, 2588–2606, https://doi.org/10.1175/JCLI-D-13-00014.1, 2014.

Lüpkes, C., Vihma, T., Jakobson, E., König-Langlo, G., and Tetzlaff, A.: Meteorological observations from ship cruises during summer to the central Arctic: A comparison with reanalysis data, Geophys. Res. Lett., 37, L09810, https://doi.org/10.1029/2010GL042724, 2010.

Masson-Delmotte, V., Swingedouw, D., Landais, A., Seidenkrantz, M.-S., Gauthier, E., Bichet, V., Massa, C., Perren, B., Jomelli, V., Adalgeirsdottir, G., Hesselbjerg Christensen, J., Arneborg, J., Bhatt, U., Walker, D. A., Elberling, B., Gillet-Chaulet, F., Ritz, C., Gallée, H., van den Broeke, M., Fettweis, X., de Vernal, A., and Vinther, B.: Greenland climate change: from the past to the future: Greenland climate change, WiRes. Clim. Change, 3, 427–449, https://doi.org/10.1002/wcc.186, 2012.

Menke, W.: Geophysical Data Analysis: Discrete Inverse Theory, Elsevier, New York, 1989.

Merchant, C. J. and Le Borgne, P.: Retrieval of Sea Surface Temperature from Space, Based on Modeling of Infrared Radiative Transfer: Capabilities and Limitations, J. Atmos. Ocean. Tech., 21, 1734–1746, https://doi.org/10.1175/JTECH1667.1, 2004.

Merchant, C. J., Harris, A. R., Murray, M. J., and Závody, A. M.: Toward the elimination of bias in satellite retrievals of sea surface temperature: 1. Theory, modeling and interalgorithm comparison, J. Geophys. Res.-Oceans, 104, 23565–23578, https://doi.org/10.1029/1999JC900105, 1999.

Merchant, C. J., Matthiesen, S., Rayner, N. A., Remedios, J. J., Jones, P. D., Olesen, F., Trewin, B., Thorne, P. W., Auchmann, R., Corlett, G. K., Guillevic, P. C., and Hulley, G. C.: The surface temperatures of Earth: steps towards integrated understanding of variability and change, Geosci. Instrum. Method. Data Syst., 2, 305–321, https://doi.org/10.5194/gi-2-305-2013, 2013.

Merchant, C. J., Paul, F., Popp, T., Ablain, M., Bontemps, S., Defourny, P., Hollmann, R., Lavergne, T., Laeng, A., de Leeuw, G., Mittaz, J., Poulsen, C., Povey, A. C., Reuter, M., Sathyendranath, S., Sandven, S., Sofieva, V. F., and Wagner, W.: Uncertainty information in climate data records from Earth observation, Earth Syst. Sci. Data, 9, 511–527, https://doi.org/10.5194/essd-9-511-2017, 2017.

Miller, N. B., Turner, D. D., Bennartz, R., Shupe, M. D., Kulie, M. S., Cadeddu, M. P., and Walden, V. P.: Surface-based inversions above central Greenland, J. Geophys. Res.-Atmos., 118, 495–506, https://doi.org/10.1029/2012JD018867, 2013.

Morice, C. P., Kennedy, J. J., Rayner, N. A., and Jones, P. D.: Quantifying uncertainties in global and regional temperature change using an ensemble of observational estimates: The HadCRUT4 data set: The HADCRUT4 data set, J. Geophys. Res.-Atmos., 117, D08101, https://doi.org/10.1029/2011JD017187, 2012.

Morice, C. P., Capponi, F., Kennedy, J. J., Killick, R. E., Lindgren, F., Mitchelson, J. R., Rayner, N. A., and Winn, J. P.: EUSTACE: Global daily air temperature combining surface and satellite data, with uncertainty estimates, for 1850–2015, v1.0, Centre for Environmental Data Analysis, Centre for Environmental Data Analysis, https://doi.org/10.5285/468abcf18372425791a31d15a41348d9, 2019.

Nielsen-Englyst, P., Høyer, J. L., Madsen, K. S., Tonboe, R., Dybkjær, G., and Alerskans, E.: In situ observed relationships between snow and ice surface skin temperatures and 2 m air temperatures in the Arctic, The Cryosphere, 13, 1005–1024, https://doi.org/10.5194/tc-13-1005-2019, 2019.

Oltmanns, M., Straneo, F., Seo, H., and Moore, G. W. K.: The Role of Wave Dynamics and Small-Scale Topography for Downslope Wind Events in Southeast Greenland, J. Atmos. Sci., 72, 2786–2805, https://doi.org/10.1175/JAS-D-14-0257.1, 2015.

Østby, T. I., Schuler, T. V., and Westermann, S.: Severe cloud contamination of MODIS Land Surface Temperatures over an Arctic ice cap, Svalbard, Remote Sens. Environ., 142, 95–102, https://doi.org/10.1016/j.rse.2013.11.005, 2014.

Overland, J., Dunlea, E., Box, J. E., Corell, R., Forsius, M., Kattsov, V., Olsen, M. S., Pawlak, J., Reiersen, L.-O., and Wang, M.: The urgency of Arctic change, Polar Sci., 21, 6–13, https://doi.org/10.1016/j.polar.2018.11.008, 2018.

Perovich, D. K., Richter-Menge, J. A., and Polashenski, C. M.: Observing and understanding climate change: Monitoring the mass balance, motion, and thickness of Arctic sea ice, The CRREL-Dartmouth Mass Balance Buoy Program, CRREL-Dartmouth, available at: http://imb-crrel-dartmouth.org, last access: 24 November 2016.

Pielke, R. A., Davey, C. A., Niyogi, D., Fall, S., Steinweg-Woods, J., Hubbard, K., Lin, X., Cai, M., Lim, Y.-K., Li, H., Nielsen-Gammon, J., Gallo, K., Hale, R., Mahmood, R., Foster, S., McNider, R. T., and Blanken, P.: Unresolved is-

sues with the assessment of multidecadal global land surface temperature trends, J. Geophys. Res., 112, D24S08, https://doi.org/10.1029/2006JD008229, 2007.

Pithan, F. and Mauritsen, T.: Arctic amplification dominated by temperature feedbacks in contemporary climate models, Nat. Geosci., 7, 181–184, https://doi.org/10.1038/ngeo2071, 2014.

Rapaić, M., Brown, R., Markovic, M., and Chaumont, D.: An Evaluation of Temperature and Precipitation Surface-Based and Reanalysis Datasets for the Canadian Arctic, 1950–2010, Atmos. Ocean, 53, 283–303, https://doi.org/10.1080/07055900.2015.1045825, 2015.

Rasmussen, T. A. S., Høyer, J. L., Ghent, D., Bulgin, C. E., Dybkjær, G., Ribergaard, M. H., Nielsen-Englyst, P., and Madsen, K. S.: Impact of Assimilation of Sea-Ice Surface Temperatures on a Coupled Ocean and Sea-Ice Model, J. Geophys. Res.-Oceans, 123, 2440–2460, https://doi.org/10.1002/2017JC013481, 2018.

Rayner, N., Good, S., and Block, T.: SST CCI Product User Guide, Project Document, SST_CCI-PUG-UKMO-201, available at: https://climate.esa.int/media/documents/SST_CCI-PUG-UKMO-201-Issue_1-signed.pdf (last access: 29 June 2021), 2015.

Rayner, N. A.: Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century, J. Geophys. Res., 108, 4407, https://doi.org/10.1029/2002JD002670, 2003.

Rayner, N. A., Auchmann, R., Bessembinder, J., Brönnimann, S., Brugnara, Y., Capponi, F., Carrea, L., Dodd, E. M. A., Ghent, D., Good, E., Høyer, J. L., Kennedy, J. J., Kent, E. C., Killick, R. E., van der Linden, P., Lindgren, F., Madsen, K. S., Merchant, C. J., Mitchelson, J. R., Morice, C. P., Nielsen-Englyst, P., Ortiz, P. F., Remedios, J. J., van der Schrier, G., Squintu, A. A., Stephens, A., Thorne, P. W., Tonboe, R. T., Trent, T., Veal, K. L., Waterfall, A. M., Winfield, K., Winn, J., and Woolway, R. I.: The EUSTACE Project: Delivering Global, Daily Information on Surface Air Temperature, B. Am. Meteorol. Soc., 101, E1924–E1947, https://doi.org/10.1175/BAMS-D-19-0095.1, 2020.

Reeves Eyre, J. E. J. and Zeng, X.: Evaluation of Greenland near surface air temperature datasets, The Cryosphere, 11, 1591–1605, https://doi.org/10.5194/tc-11-1591-2017, 2017.

Renfrew, I. A.: The dynamics of idealized katabatic flow over a moderate slope and ice shelf, Q. J. Roy. Meteor. Soc., 130, 1023–1045, https://doi.org/10.1256/qj.03.24, 2004.

Richter-Menge, J. A., Perovich, D. K., Elder, B. C., Claffey, K., Rigor, I., and Ortmeyer, M.: Ice mass-balance buoys: a tool for measuring and attributing changes in the thickness of the Arctic sea-ice cover, Ann. Glaciol., 44, 205–210, https://doi.org/10.3189/172756406781811727, 2006.

Richter-Menge, J. A., Overland, J., Mathis, J. T., and Osborne, E. (Eds.): Arctic Report Card 2017, 2017.

Screen, J. A. and Simmonds, I.: Increasing fall-winter energy loss from the Arctic Ocean and its role in Arctic temperature amplification, Geophys. Res. Lett., 37, L16707, https://doi.org/10.1029/2010GL044136, 2010.

Serreze, M. C., Box, J. E., Barry, R. G., and Walsh, J. E.: Characteristics of Arctic synoptic activity, 1952–1989, Meteorol. Atmos. Phys., 51, 147–164, https://doi.org/10.1007/BF01030491, 1993.

Shuman, C. A., Steffen, K., Box, J. E., and Stearns, C. R.: A Dozen Years of Temperature Observations at the Summit: Central Greenland Automatic Weather Stations 1987–99, J. Appl. Meteorol., 40, 741–752, https://doi.org/10.1175/1520-0450(2001)040<0741:ADYOTO>2.0.CO;2, 2001.

Shuman, C. A., Hall, D. K., DiGirolamo, N. E., Mefford, T. K., and Schnaubelt, M. J.: Comparison of Near-Surface Air Temperatures and MODIS Ice-Surface Temperatures at Summit, Greenland (2008–13), J. Appl. Meteorol. Clim., 53, 2171–2180, https://doi.org/10.1175/JAMC-D-14-0023.1, 2014.

Simmons, A. J. and Poli, P.: Arctic warming in ERA-Interim and other analyses: Arctic Warming in ERA-Interim and Other Analyses, Q. J. Roy. Meteor. Soc., 141, 1147–1162, https://doi.org/10.1002/qj.2422, 2014.

Smith, T. M., Reynolds, R. W., Peterson, T. C., and Lawrimore, J.: Improvements to NOAA's Historical Merged Land–Ocean Surface Temperature Analysis (1880–2006), J. Climate, 21, 2283–2296, https://doi.org/10.1175/2007JCLI2100.1, 2008.

Stamnes, K., Ellingson, R. G., Curry, J. A., Walsh, J. E., and Zak, B. D.: Review of Science Issues, Deployment Strategy, and Status for the ARM North Slope of Alaska-Adjacent Arctic Ocean Climate Research Site, J. Climate, 12, 46–63, https://doi.org/10.1175/1520-0442(1999)012<0046:ROSIDS>2.0.CO;2, 1999.

Steeneveld, G.-J.: Current challenges in understanding and forecasting stable boundary layers over land and ice, Frontiers in Environmental Science, 2, 41, https://doi.org/10.3389/fenvs.2014.00041, 2014.

Steffen, K.: Surface energy exchange at the equilibrium line on the Greenland ice sheet during onset of melt, Ann. Glaciol., 21, 13–18, https://doi.org/10.3189/S0260305500015536, 1995.

Steffen, K. and Box, J.: Surface climatology of the Greenland Ice Sheet: Greenland Climate Network 1995–1999, J. Geophys. Res.-Atmos., 106, 33951–33964, https://doi.org/10.1029/2001JD900161, 2001.

Sterk, H. A. M., Steeneveld, G. J., and Holtslag, A. A. M.: The role of snow-surface coupling, radiation, and turbulent mixing in modeling a stable boundary layer over Arctic sea ice, J. Geophys. Res.-Atmos., 118, 1199–1217, https://doi.org/10.1002/jgrd.50158, 2013.

The CRREL-Dartmouth Mass Balance Buoy Program: CRREL-Dartmouth, available at: http://imb-crrel-dartmouth.org, last access: 24 November 2016.

Tonboe, R. T., Eastwood, S., Lavergne, T., Sørensen, A. M., Rathmann, N., Dybkjær, G., Pedersen, L. T., Høyer, J. L., and Kern, S.: The EUMETSAT sea ice concentration climate data record, The Cryosphere, 10, 2275–2290, https://doi.org/10.5194/tc-10-2275-2016, 2016.

van As, D.: Warming, glacier melt and surface energy budget from weather station observations in the Melville Bay region of northwest Greenland, J. Glaciol., 57, 208–220, https://doi.org/10.3189/002214311796405898, 2011.

van As, D., Fausto, R. S., Ahlstrøm, A., Andersen, S., Citterio, M., Edelvang, K., Graversen, P., Machguth, H., Nick, F., Nielsen, S., and Weidick, A.: Programme for Monitoring of the Greenland Ice Sheet (PROMICE): first temperature and ablation record., Geol. Surv. Den. Greenl., 23, 73–76, 2011.

Vihma, T. and Pirazzini, R.: On the Factors Controlling the Snow Surface and 2 m Air Temperatures Over the Arctic Sea Ice in Winter, Bound.-Lay. Meteorol., 117, 73–90, https://doi.org/10.1007/s10546-004-5938-7, 2005.

Vihma, T., Uotila, J., Cheng, B., and Launiainen, J.: Surface heat budget over the Weddell Sea: Buoy results and model comparisons, J. Geophys. Res., 107, 3013, https://doi.org/10.1029/2000JC000372, 2002.

Vihma, T., Jaagus, J., Jakobson, E., and Palo, T.: Meteorological conditions in the Arctic Ocean in spring and summer 2007 as recorded on the drifting ice station Tara, Geophys. Res. Lett., 35, L18706, https://doi.org/10.1029/2008GL034681, 2008.

Vose, R. S., Arndt, D., Banzon, V. F., Easterling, D. R., Gleason, B., Huang, B., Kearns, E., Lawrimore, J. H., Menne, M. J., Peterson, T. C., Reynolds, R. W., Smith, T. M., Williams, C. N., and Wuertz, D. B.: NOAA's Merged Land–Ocean Surface Temperature Analysis, B. Am. Meteorol. Soc., 93, 1677–1685, https://doi.org/10.1175/BAMS-D-11-00241.1, 2012.

Wang, C., Graham, R. M., Wang, K., Gerland, S., and Granskog, M. A.: Comparison of ERA5 and ERA-Interim near-surface air temperature, snowfall and precipitation over Arctic sea ice: effects on sea ice thermodynamics and evolution, The Cryosphere, 13, 1661–1679, https://doi.org/10.5194/tc-13-1661-2019, 2019.

Weng, W. and Taylor, P. A.: On Modelling the One-Dimensional Atmospheric Boundary Layer, Bound.-Lay. Meteorol., 107, 371–400, https://doi.org/10.1023/A:1022126511654, 2003.

Wesslén, C., Tjernström, M., Bromwich, D. H., de Boer, G., Ekman, A. M. L., Bai, L.-S., and Wang, S.-H.: The Arctic summer atmosphere: an evaluation of reanalyses using ASCOS data, Atmos. Chem. Phys., 14, 2605–2624, https://doi.org/10.5194/acp-14-2605-2014, 2014.

Westermann, S., Langer, M., and Boike, J.: Systematic bias of average winter-time land surface temperatures inferred from MODIS at a site on Svalbard, Norway, Remote Sens. Environ., 118, 162–167, https://doi.org/10.1016/j.rse.2011.10.025, 2012.

World Meteorological Organization: World Meteorological Organization (2014) Guide to Meteorological Instruments and Methods of Observation, WMO-No.8, 1128 pp., preprint, available at: http://hdl.handle.net/11329/365 (last access: 29 June 2021), 2014.

Zhang, W., Wang, Y., Smeets, P. C. J. P., Reijmer, C. H., Huai, B., Wang, J., and Sun, W.: Estimating near-surface climatology of multi-reanalyses over the Greenland Ice Sheet, Atmos. Res., 259, 105676, https://doi.org/10.1016/j.atmosres.2021.105676, 2021.

Zilitinkevich, S., Savijärvi, H., Baklanov, A., Grisogono, B., and Myrberg, K.: Forthcoming Meetings on Planetary Boundary-layer Theory, Modelling and Applications, Bound.-Lay. Meteorol., 119, 591–593, https://doi.org/10.1007/s10546-006-9069-1, 2006.

APPENDIX G

Paper VII: Concerns on benchmarking climate models in the Arctic

This appendix contains a full copy of Paper VII:

- Tian Tian, Shuting Yang, Jacob L. Høyer, **Pia Nielsen-Englyst**, Suman Singha. 2023. "Concerns on benchmarking climate models in the Arctic." *Preprint submitted to Nature Geoscience - Correspondence*

Concerns on benchmarking climate models in the Arctic

Tian Tian[1*], Shuting Yang[1], Jacob L. Høyer[1], Pia Nielsen-Englyst[1,2], Suman Singha[1]

[1*]National Center for Climate Research, Danish Meteorological Institute, Copenhagen, 2100, Denmark.
[2]DTU-Space, Technical University of Denmark, Lyngby, 2800, Denmark.

*Corresponding author(s). E-mail(s): tian@dmi.dk;
Contributing authors: sy@dmi.dk; jlh@dmi.dk; pne@dmi.dk; ssi@dmi.dk;

Abstract

This letter presents a reevaluation of earlier CMIP6 Arctic assessments in IPCC 2021, focusing on the ongoing issue of climate models tending to represent a colder Arctic. We show that this long-standing bias is originated from the comparison with the commonly used ERA5 reanalysis dataset, which has a warm bias over Arctic sea ice. In contrast, our findings show that the CMIP6 models perform fairly well when compared to a recently developed high-resolution satellite-derived 2m air temperature dataset in the Arctic. These findings have significant implications for future model evaluation, climate change assessment, and the upcoming CMIP7.

Keywords: Arctic, sea ice, 2m air temperature, ERA5, climate models, CMIP6

Several recent studies including the IPCC AR6 show that climate models generally simulate a too cold Arctic, also known as a longstanding bias in the past phases of the Coupled Model Intercomparison Project (CMIP) (IPCC, 2021; Bock et al, 2020). This bias, however, is often derived from comparisons with global reanalyses rather than direct observational data, because the latter are typically available as anomalies relative to a reference period (Simmons et al, 2017; Benestad et al, 2019). Nevertheless, recent studies (Batrak and Müller, 2019; Zampieri et al, 2023) found a large warm bias of 5-10°C in the sea ice surface temperature (IST) in nearly all known reanalyses. When using these global atmospheric reanalysis products, e.g., the

widely used ERA5(Hersbach et al, 2020; IPCC, 2021; Bock et al, 2020), for model validation, this bias may lead to inaccurate conclusions, such as presuming that the climate models have a "cold temperature" bias.

Because the Arctic, particularly over sea ice, is data sparse, global reanalyses in the Arctic are only weakly constrained by observations and heavily rely on model formulation with simplified physical processes tied to the radiation budget, such as missing representation of snow on sea ice and constant thickness for sea ice, resulting in considerable uncertainty (Simmons et al, 2017; Benestad et al, 2019; Batrak and Müller, 2019; AMAP, 2021). Accurately assessing the Arctic climate state is crucial, as many essential elements in the Arctic (e.g., sea ice and permafrost melting points, ecosystems, and their possible tipping points, etc.) respond to specific temperature thresholds. We recently developed a new high spatial resolution observational dataset of 2m air temperature (T2m), converted from satellite derived ISTs Nielsen-Englyst et al (2023) using an existing model (Nielsen-Englyst et al, 2021) (hereafter referred to as $T_{2\mathrm{m}}^{(\mathrm{SAT})}$), which enables benchmarking climate models in the Arctic. When compared to various in-situ T2m observations (Nielsen-Englyst et al, 2023), the $T_{2\mathrm{m}}^{(\mathrm{SAT})}$ mean differences range from -0.45°C to 0.65°C, significantly smaller than those of $T_{2\mathrm{m}}^{(\mathrm{ERA5})}$ ranging from 1.73 to 3.73°C.

Fig. 1a shows the $T_{2\mathrm{m}}^{(\mathrm{SAT})}$ climatology over the regions with Arctic sea ice concentrations (SIC) above 30%. Using $T_{2\mathrm{m}}^{(\mathrm{SAT})}$ as reference, ERA5 exhibits a wide spread warm bias of more than 2°C (Fig. 1b) for areas where the SIC is typically above 70% (compassed by the red line and hereafter denoted as SIC70). The bias is markedly larger in wintertime when it may reach 6-10°C, consistent with (Batrak and Müller, 2019). In contrast, the latest generation of CMIP (CMIP6) historical ensemble of 47 models performs remarkably well in this region, with a mean difference of only 0.5°C or less (Fig. 1c), falling within the range of observational uncertainties. Outside of SIC70 (i.e., the marginal ice zone), the largest biases are primarily associated with discrepancies in sea ice cover for both ERA5 and CMIP6 (Fig. 1b-c). Particularly for CMIP6, it is attributed to large regional disparities among models (Davy and Outten, 2020; Shen et al, 2021) and only the subset of models with SIC above 30% have been accounted.

Figure 1 also shows time evolutions of the annual T2m differences of ERA5 and CMIP6 ensemble mean (Fig. 1d-e) with respect to $T_{2\mathrm{m}}^{(\mathrm{SAT})}$ (Fig. 1f) area averaged over SIC70. From 1982 to 2020, $\Delta T_{2\mathrm{m}}^{(\mathrm{ERA5})}$ remains positive by more than 1.41°C, whereas $\Delta T_{2\mathrm{m}}^{(\mathrm{CMIP6})}$ is modest and varies around zero. This contradicts previous conclusions that the Arctic is too cold in CMIP6 historical simulations when compared to ERA5 (IPCC, 2021; Davy and Outten, 2020). Here we emphasize the limitations of using global reanalysis such as ERA5 to evaluate climate models over Arctic sea ice.

Despite the prevailing view that climate models systematically underestimate Arctic warming and amplification compared to ERA5 (Rantanen et al, 2022), CMIP6 (0.61 °C/decade) performs slightly better than ERA5 (0.56 °C/decade) in capturing the warming trend over SIC70, aligned with the observed trends of surface warming (0.79 °C/decade) and sea ice loss (Fig. 1g). Overall, the differences between $T_{2\mathrm{m}}^{(\mathrm{CMIP6})}$ and $T_{2\mathrm{m}}^{(\mathrm{ERA5})}$ (Fig. 1b,c) are similar to the previous studies regarding the longstanding

cold bias over the Arctic region in climate models (IPCC, 2021; Davy and Outten, 2020). However, using $T_{2\mathrm{m}}^{(\mathrm{SAT})}$ as a benchmark, the cold bias in $T_{2\mathrm{m}}^{(\mathrm{CMIP6})}$ is only evidenced in the marginal ice zone. These results highlight the need for improving the relevant processes in assimilation and simulations, as well as including observations as new benchmark fields for evaluating climate models for the Arctic region.

Abbreviations. IPCC: the Intergovernmental Panel on Climate Change; AR6: the Sixth Assessment Report; CMIP: the Coupled Model Intercomparison Project; CMIP6: the Sixth generation of CMIP; IST: sea-Ice Surface Temperature; ERA5: the fifth generation ECMWF reanalysis for the global climate; ECMWF: the European Centre for Medium-Range Weather Forecasts; T2m: 2m air temperature; SIC: sea ice concentration (%); SIC70: areas with SICge70%; NCKF: National Center for Climate Research.

Acknowledgments. The authors were supported by the Danish National Center for Climate Research (NCKF). SY is also partly funded by the EU Horizon Europe research and innovation project OptimESM under grant agreement No101081193. JLH and PNE are also funded by the Copernicus Marine Services.

Author contributions. Conception of the work: TT, SY, JLH. Acquisition of data: TT, PNE and SSI. TT designed analysis, produced figures and wrote the manuscript. PNE performed evaluation of T2m with in-situ observations. All authors read, improved and approved the final manuscript.

Availability of data and materials. The satellite-derived IST and sea ice data used for the T2m calculation are available from https://doi.org/10.48670/moi-00123 (ref. 6). ERA-5 reanalysis data are freely available at the Copernicus Climate Change Service Climate Data Store (https://cds.climate.copernicus.eu/). CMIP6 multi-model model means are obtained using the first member of each model that are available freely and publicly from the Earth System Grid Federation (ESGF, https://esgf.llnl.gov).

Declarations

Competing interests. The authors declare no competing interests.

References

AMAP (2021) Arctic Climate Change Update 2021: Key Trends and Impacts. Summary for Policy-Makers. Arctic Monitoring and Assessment Programme (AMAP), Tromsø, Norway, 16 pp

Batrak Y, Müller M (2019) On the warm bias in atmospheric reanalyses induced by the missing snow over arctic sea-ice. Nat Commun 10(1):4170. https://doi.org/https://doi.org/10.1038/s41467-019-11975-3

Benestad R, Erlandsen H, Mezghani A, et al (2019) Geographical distribution of thermometers gives the appearance of lower historical global warming 46(13):7654–7662. https://doi.org/https://doi.org/10.1029/2019GL083474

Bock L, Lauer A, Schlund M, et al (2020) Quantifying progress across different cmip phases with the esmvaltool 125(21):e2019JD032321. https://doi.org/https://doi.org/10.1029/2019JD032321

Davy R, Outten S (2020) The arctic surface climate in cmip6: status and developments since cmip5 33(18):8047–8068. https://doi.org/https://doi.org/10.1175/JCLI-D-19-0990.1

Hersbach H, Bell B, Berrisford P, et al (2020) The era5 global reanalysis 146(730):1999–2049. https://doi.org/https://doi.org/10.1002/qj.3803

IPCC (2021) Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 2391 pp

Nielsen-Englyst P, Høyer JL, Madsen KS, et al (2021) Deriving arctic 2 m air temperatures over snow and ice from satellite surface temperature measurements. Cryosphere 15(7):3035–3057. https://doi.org/https://doi.org/10.5194/tc-15-3035-2021

Nielsen-Englyst P, Høyer JL, Kolbe WM, et al (2023) A combined sea and sea-ice surface temperature climate dataset of the arctic, 1982–2021. Remote Sens of Environ 284:113331. https://doi.org/https://doi.org/10.1016/j.rse.2022.113331

Rantanen M, Karpechko AY, Lipponen A, et al (2022) The arctic has warmed nearly four times faster than the globe since 1979. Commun Earth Environ 3(1):168. https://doi.org/https://doi.org/10.1038/s43247-022-00498-3

Shen Z, Duan A, Li D, et al (2021) Assessment and ranking of climate models in Arctic Sea ice cover simulation: From CMIP5 to CMIP6 34(9):3609–3627. https://doi.org/https://doi.org/10.1175/JCLI-D-20-0294.1

Simmons A, Berrisford P, Dee D, et al (2017) A reassessment of temperature variations and trends from global reanalyses and monthly surface climatological datasets 143(702):101–119. https://doi.org/https://doi.org/10.1002/qj.2949

Zampieri L, Arduini G, Holland M, et al (2023) A machine learning correction model of the winter clear-sky temperature bias over the arctic sea ice in atmospheric reanalyses 151(6):1443 – 1458. https://doi.org/https://doi.org/10.1175/MWR-D-22-0130.1

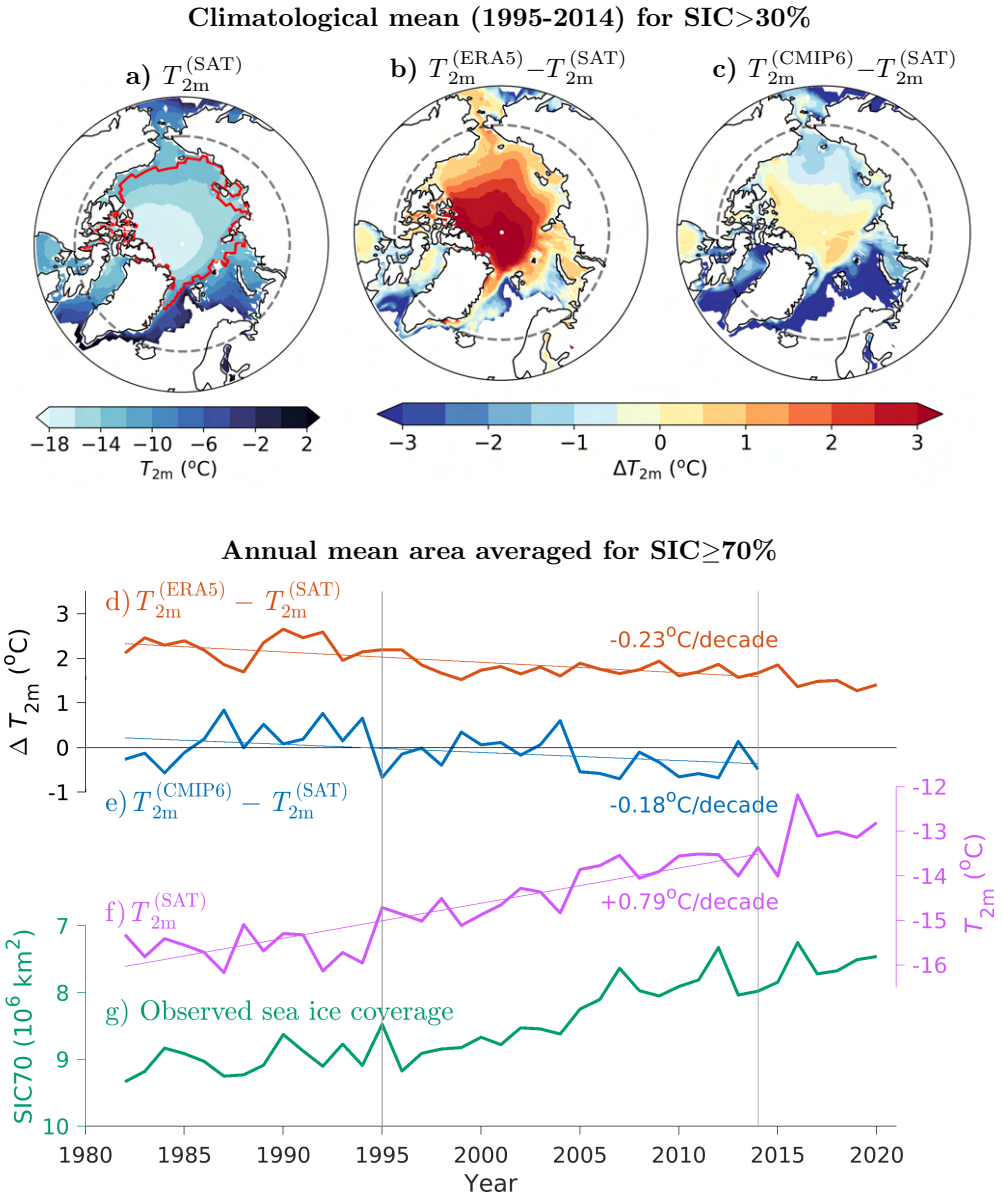


Fig. 1 (a) The 20 years climatology of satellite derived $T_{2m}^{(SAT)}$ over sea ice. The climatological mean difference of T2m from (b) ERA5 and (c) CMIP6 ensemble means versus $T_{2m}^{(SAT)}$. All T2m data are collected for the region with SIC>30% (>58°N) using the observed SIC in Nielsen-Englyst et al (2023). The CMIP6 ensemble mean is calculated from T2m with SIC>30% for each of 47 models to exclude grid points over open sea. In (a-c) the map is bounded at 58°N with the dashed line marking 66.5°N. The red line in (a) indicates the observed SIC≥70% (SIC70) averaged over 1995-2014. Time series of annual mean observed sea ice extent for SIC70 on an inverted y-axis (g), $T_{2m}^{(SAT)}$ (f), and its difference from the ERA5 (d)/CMIP6 model mean (e) averaged for the same area and the respective linear trends (°C/decade) for 1980-2014.

APPENDIX H

Unpublished report: Exploring machine learning to retrieve sea surface temperature from satellite observations

This appendix contains a revised version of a short, unpublished report submitted for the DTU PhD Course *02910 Computational Data Analysis*, taken as part of the PhD degree:

- **Pia Nielsen-Englyst**. 2021. "Exploring machine learning to retrieve sea surface temperature from satellite observations." *Unpublished*

Exploring machine learning to retrieve sea surface temperature (SST) from satellite observations

Abstract

Accurate sea surface temperatures are crucial for weather, ocean, and climate models as well as for monitoring of climate change. This study evaluates different machine learning (ML) techniques for their capability to predict sea surface temperatures from passive microwave (PMW) satellite observed brightness temperatures. Each of the algorithms also use some instrument and environmental information as input. Overall, the performances are good, with the neural network outperforming three different ensemble learning algorithms. The neural network further outperforms the current state-of-art retrieval algorithms based on inversion of a physical model and regression, respectively. This indicates that there is a great potential for using ML techniques to retrieve SST from PMW observations.

Introduction

The sea surface temperature (SST) is an essential climate variable used for climate monitoring, understanding of air–sea interactions, and numerical weather prediction (NWP). Two kinds of retrieval algorithms have generally been used to retrieve SST from PMW satellite observations: statistical algorithms developed by comparisons of satellite observed brightness temperatures and collocated in situ and model observations (most common, see e.g. Alerskans et al., 2020; Wentz and Meissner, 2007), and radiative transfer model (RTM) based algorithms, which use an RTM to simulate the top of atmosphere brightness temperatures. This requires instrument information (azimuth/earth incidence angles, frequency and polarisation) and environmental information (SST, sea surface salinity (SSS), wind speed/direction, water vapour density, liquid water density, pressure, and atmospheric profiles of temperature). Optimal estimation (OE) theory has previously been used to invert a RTM to retrieve SST (Nielsen-Englyst et al., 2018) using a priori information about the ocean and atmosphere (and corresponding uncertainties) to constrain the retrievals. The OE types of algorithms clearly tie the physics of the observations to the geophysical retrieval algorithm, which allows for indication of measurement errors (Minnett et al., 2019). This also means that the performance of OE algorithms is constrained by the accuracy of the RTM as well as the representativeness of the observation and prior error covariances (Merchant et al., 2020). Moreover, measurement errors require ad-hoc corrections to the geophysical retrievals in the OE type of algorithms (Meissner and Wentz, 2012; Nielsen-Englyst et al., 2018). In contrast, statistically based algorithms may account for some of the measurement errors through the coefficient derivation process, but they are limited by the established statistical relationships between the variables.

Machine learning (ML) may improve or supplement existing retrieval algorithms through their higher flexibility and capability of recognizing meaningful patterns and structures in complex problems (Lee et al., 2017). Until recently, the use of ML techniques has been very limited within the field of SST retrievals. However, there has been an increasing amount of research applying ML techniques to

specific parts of retrieval algorithms such as for cloud detection (Paul and Huntemann, 2021), bias correction (Saux Picart et al., 2018), error estimation (Kumar et al., 2021) and identification of eddies (Moschos et al., 2020). Latest, Alerskans et al. (2021) investigated the ability of retrieving PMW SST using of a neural network (the TensorFlow interface; (Haykin, 1999; Nielsen, 2015; Abadi et al., 2016) and the tree-based algorithm Extreme Gradient Boosting (XGBoost; Chen and Guestrin, 2016). The results look promising and we wish to build on this work to further explore the possibilities of using ML techniques for SST retrievals.

In this study, four different ML algorithms have been tested for their ability to retrieve SST from PMW observations: bootstrap aggregation (bagging), random forests and least-squares boosting and a neural network. The first tree algorithms are all different types of ensemble learning, which combine the results of many regression trees in different ways. The neural network has been re-implemented in MATLAB using the optimized hyperparameters from Alerskans et al. (2021), but leaving out two of the input variables. The different algorithms have been compared and assessed in terms of their ability to retrieve SST using matchups with *in situ* observations. Furthermore, the performances have been compared with a simple regression tree and the two benchmark algorithms from Alerskans et al. (2021).

There is a generally good understanding of the geophysical parameters impacting the brightness temperatures observed by the satellites and the relationship with the SST. Therefore, this study has not put any efforts into evaluating the feature importance but included all the inputs known to be of importance. However, it is also important to know the impact of missing input information e.g. in the case of instrument failure on satellites and in this relation it could be of interest to perform an analysis of feature importance in the future.

Data

For this study we use a matchup data base (MMD) consisting of collocated brightness temperatures from AMSR-E and *in situ* observed surface temperatures from drifting buoys. Each matchup also includes instrument information as well a number of geophysical variables (from numerical weather prediction (NWP) models) known to affect the observed brightness temperatures (Nielsen-Englyst et al., 2018). Table 1A (Appendix) shows the number of variables included in the MMD. A total of 5,105,663 matchups covering the years 2002-2011. The MMD has been quality controlled and filtered to remove erroneous observations and to ensure a good and representative coverage of the global ocean. The data has been divided randomly into three subsets consisting of 50%, 25% and 25% used for training, validation and testing, respectively for all algorithms below. The division into subsets is performed by the Neural Network, but is used for the other algorithms as well.

Methods

Bootstrap aggregation (bagging)

Bagging uses many bootstrap samples (with replacement) of the training data to grow individual regression trees for each bootstrap sample and afterwards averaging (aggregating) the predictions. This allow us to average out the noise and reducing the variance compared to growing just one regression tree, and the bias of the bagged tree is the same as that of the individual trees. In this case,

we use 100 bootstrap samples and a minimum leaf size of 5 (1, 3, 5, and 10 were tested manually), which is the default leaf size for regression and close to optimal for the predictive power of an ensemble.

When drawing each bootstrap sample 37% of the observations will on average be omitted. These omitted observations, called “out-of-bag” (OOB) observations, can be used for reliable estimates of the retrieval performance and the feature importance. However, here we have used the test data set for a consistent comparison with the other algorithms in Table 1 and Figure 1-2.

Random Forest

Random Forest (RF) is similar to Bagging, but every tree in the ensemble randomly selects only a subset of predictors for each decision split, while Bagging is provided with the full range of predictors each time. Due to the random feature selection, the trees are more independent of each other compared to bagging, which often results in better performance (due to better variance-bias trade-offs). The number of variables to be sampled randomly at each split must be set and this is the most important parameter to define. Figure A2 shows the results from the `'OptimizeHyperparameters'` Bayesian optimization software in MATLAB. The optimal number of variables to be sampled is found to be 14, but the variations are relatively small for the values in the range 7-20. Therefore, to save time, the results presented below uses a value of 8, which is the default value in MATLAB (and run at the very beginning as benchmark). We use 100 trees and maximum number of splits equal to one minus the number of matchups in the validation subset, while the minimum leaf size is set to 5 as for the bagging algorithm.

Least-squares boosting (LSBoost)

Least-squares gradient boosting (LSBoost) is also an ensemble learning algorithm, but it differs from RF in the way that each tree is trained. While RF uses randomness and grow each regression tree independently, LSBoost grows one tree at a time and adds a new tree to its structure by fitting it to the difference between the observed response and the aggregated prediction of all the pre-trained trees. LSBoost fits to minimize mean-squared error and it can be used with shrinkage by providing the learning rate (LearnRate) parameter. It is very important to select a sufficient learning rate. If the learning rate is too small, the ensemble of trees has to be very large to reach a sufficient performance, while higher learning rates may take too large steps and overshoot the minima. In this case, the LearnRate, MaxNumSplits and B have been chosen by using the `'OptimizeHyperparameters'` random search and a smaller subset of the validation data (~12,764). The final search ranges for each of the hyperparameters were: 0.05:0.8, 1:7, and 300:800, with the optimized values being 0.44, 3, and 705, respectively. This structure is a lot different from the one used in the XGBoost algorithm, which uses 103 trees, a maximum number of splits of 22, a minimum leaf size of 3, and a learning rate of 0.085. More over the XGBoost algorithm have used fractions of subsampling, subsampling by tree, and subsampling by level of 0.58, 0.70, and 0.63, respectively.

Neural Network

Following Alerskans et al. (2021), we use a neural network with two hidden layers with 10 and 15 hidden neurons, respectively. The data are randomly divided into three subsets. The training data are presented to the neural network during training, where the network is adjusted according to its error. The testing data have no effect on the training process, and provide a completely independent measure of network performance. The network is trained with Levenberg-Marquardt backpropagation algorithm (trainlm.m). The validation data are used to measure network generalization, and to stop the training process when the generalization stops improving (and before overfitting). Early stopping is applied, which means that the training automatically stops when generalization stops improving, as indicated by an increase in the mean squared error of the validation data. In this case, the training stopped when the validation error failed to decrease for six iterations, which occurred at iteration 926. Once, the neural network has fit the data, it forms a generalization of the input-output relationship which can be used to generate outputs for inputs it was not trained on. Table 1 and Figure 1-2 show the performance against the independent test data. The standard deviations for the training and validation data sets agree within two digits. If the performance against the training data had been significantly better this could indicate overfitting and a solution would be to reduce the number of hidden neurons. In this case, the performance is equally good and there is no reason to test different numbers of neurons.

Results

The retrieved SSTs have been compared with independent *in situ* SST using the test data (shown in Table 1). The table also shows the results of using a single regression tree for comparison. Furthermore, the performance of the two benchmark algorithms in Alerskans et al. (2021) have been included for comparison using the same validation subset. As expected all algorithms improve the retrievals over a single regression tree. The best performance is obtained using the XGBoost algorithm followed by the two neural networks, while RF and LSBoost show a comparable performance, except from the sensitivity (AK; explained below). Table 2A (Appendix) shows the mean validations statistics of the Arctic >58°N.

*Table 1. Validation results against independent drifter observations of SST from test data (N=1,276,416) of the different retrieval algorithms. The table shows the mean error (ME, i.e. the bias), mean squared error (MSE), mean absolute error (MAE), standard deviation (SD), robust SD (RSD), the coefficient of determination (R^2) and the sensitivity (AK). The last two rows show the performance of the two algorithms used in Alerskans et al. (2021) using the same test subset. All differences are calculated as retrieved minus in situ SST. *The sensitivity is calculated from a different subset consisting of 1,021,133 matchups which includes simulated brightness temperatures.*

	ME	MAE	MSE	SD	RSD	R^2	AK*
Regression Tree	0.001	0.57	0.60	0.77	0.66	0.993	0.80
Bagging	0.002	0.48	0.42	0.65	0.56	0.995	0.81
Random Forest	0.001	0.45	0.37	0.61	0.53	0.996	0.79
LSBoost	0.001	0.45	0.37	0.61	0.54	0.995	0.86
Neural Network	0.001	0.33	0.20	0.45	0.40	0.998	0.92
XGBoost	0.007	0.24	0.13	0.36	0.26	0.998	0.78
NN Tensorflow	0.012	0.37	0.25	0.50	0.44	0.997	0.88

The geographical distribution of the ME and SD have been compared for the different algorithms in Figure 1 and 2, respectively. All algorithms have a decreasing performance with latitude and largest SD in the high variability regions such as the Gulf Stream. For the Bagging and RF, the high variability regions and the Southern Ocean also suffer from high biases. RF, LSBoost and NN all improve the SD in the Southern and Equatorial Region compared to Bagging. LSboost and NN also improves the bias compared to Bagging and RF. Figure 2A (appendix) shows the ME and SD as a function of drifter SST, wind speed, total column water vapor and total cloud liquid water for each of the ML models. The NN shows superior, and very stable, performance compared to the other models.

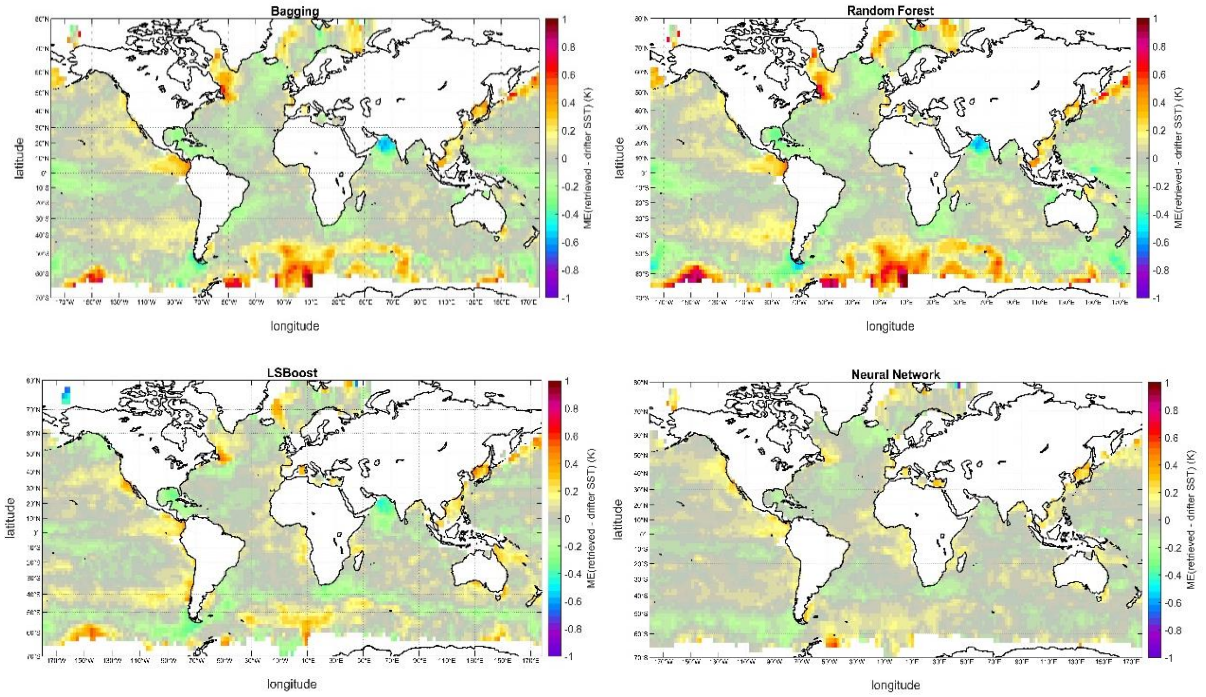


Figure 1. Mean retrieved SST minus drifter SST (ME) for each of the four algorithms.

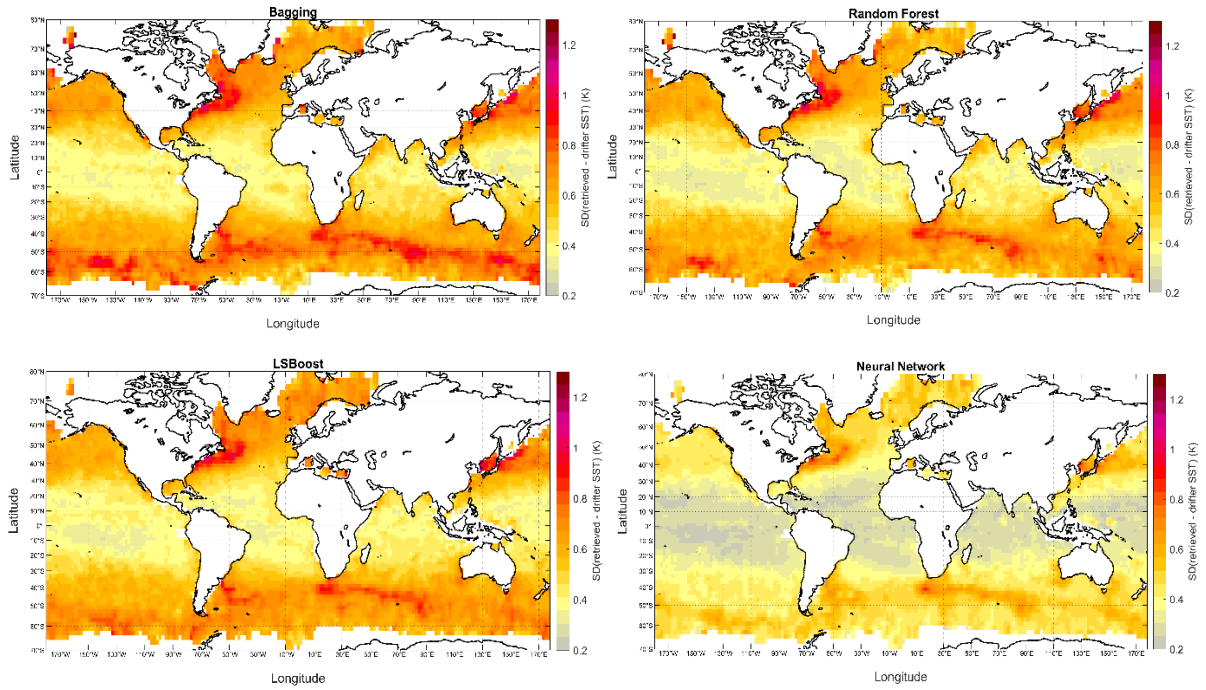


Figure 2. Mean standard deviations (SD) of the retrieved minus drifter SST difference for each of the four algorithms.

The sensitivity (AK) is a very useful estimate of the algorithms ability to capture true changes in SST. It is defined as the change in retrieved SST per unit change in the true SST i.e. a value of 1 would mean an optimal relation. To estimate the sensitivity, we use a forward model to simulate brightness temperatures for three different scenarios: one using reference in situ SST, and two using in situ SST with 1°C added and subtracted, respectively. The simulated brightness temperatures are then provided to the ML retrieval algorithms instead of observed ones to estimate the change in retrieved SST per change in "true" SST i.e. the change in simulated brightness temperature. For this, a different matchup data set is used consisting of 1,021,133 matchups. The average sensitivity results can be seen in Table 1, while Figure 3 show the geographical distributions. The sensitivities are high for all algorithms and highest for the NN developed in this study. However, the high sensitivity for LSBoost should also be noted. All algorithms show a general tendency of higher sensitivities in warm waters and decreasing sensitivities with higher latitudes, but with slightly lower sensitivities at equator. This is in agreement with previous findings (Nielsen-Englyst et al., 2018).

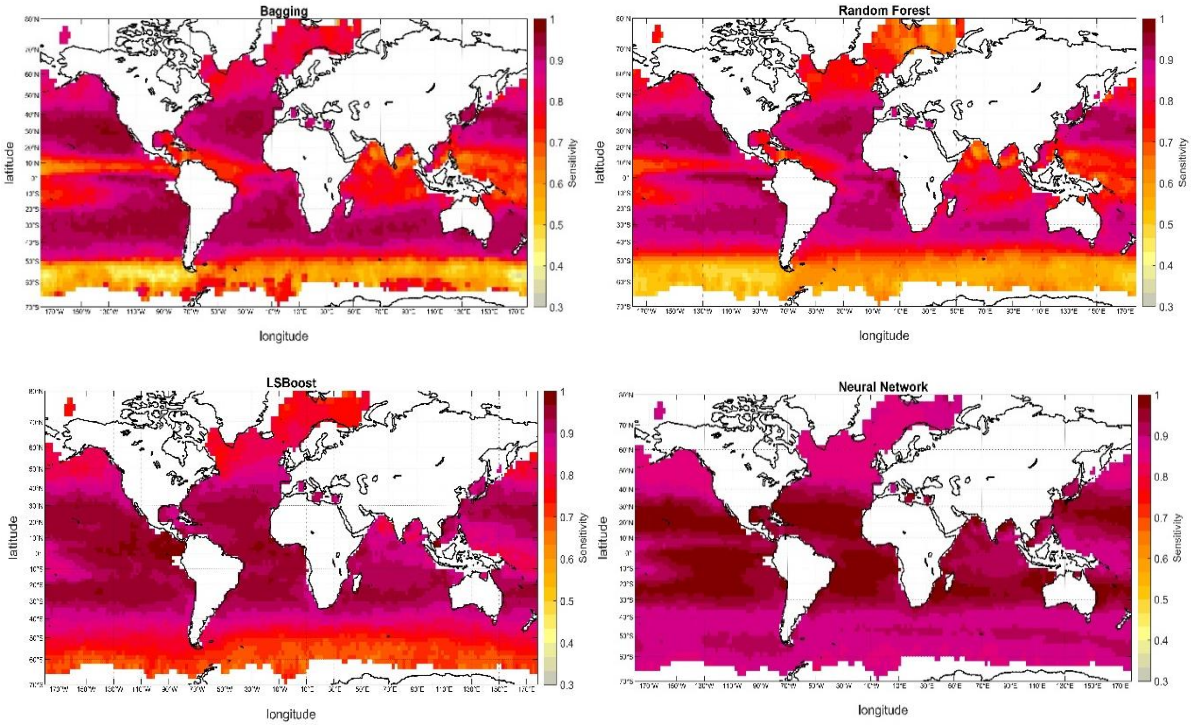


Figure 3. Mean SST sensitivity for each of the four algorithms.

Discussion

Four different ML algorithms have been assessed for the ability to retrieve SST using passive microwave observations. Alerskans et al. (2021) has previously used two different ML algorithms (XGBoost and a neural network) to retrieve SST and these have been used as benchmark.

The performance from a single regression tree has been included and all algorithms improve as expected compared to that. The better performance of bagging and RF is a result of a variance reduction since they provide a bias equal to the bias of the individual trees. Due to the random feature selection in RF, the trees are more independent of each other compared to bagging, and this results in a better performance (due to better variance-bias trade-offs). It was expected that LSboost would outperform RF, but with the hyperparameters used here, this is not the case when considering the global statistics

in Table 1. However, LSBoost better represent the extreme cases better (Figure 2A) and provides a better (higher) sensitivity. More efforts can be put into the tuning of the hyperparameters of the LSBoost algorithm. The tuning is computational expensive with the amount of data, which is needed in order to get a good coverage of the global ocean. It was observed that a better performance could be reached with increasing trees (up to 10,000 trees was tested) and reducing learn rate. This is not surprising, but would take a long time to run on a standard computer. Future work should be aimed at finding a good approach for parameter tuning of the LSBoost algorithm. Furthermore, different fractions of subsampling, subsampling by tree, and subsampling by level could be tested. The RF training and testing can be repeated for the optimal number of variables to be sampled at each split, and this may also improve the retrievals (but likely only slightly).

In terms of SD, the XGBoost outperforms all the other algorithms including the two NNs and the LSboost algorithm used in this study. The XGBoost and LSboost algorithm differ a lot in their structure. While XGBoost grows 103 large trees with a maximum depth of 22, LSboost grows more trees (705), but very shallow trees with a maximum depth of 3. Growing large/complex trees may result in overfitting (which would also result in a low sensitivity) and this could be the case for the XGBoost algorithm. Another explanation could be that the test data is not completely independent, i.e. information has leaked in some way e.g. through the normalization, which is done using the scikit-learn quantiletransformation method. The NN set up in this study outperforms the NN (using the TensorFlow interface) in Alerskans et al. (2021). One difference between the two NNs is the number of input variables. In this study, two input variables were left out as these are thought to have limited impact on the SST. This can likely explain the difference in performance between the two NNs.

Overall, the results look very promising and they indicate that ML techniques can be used to retrieve SST at a similar or even better accuracy than existing retrieval algorithms (Nielsen-Englyst et al., 2018; Alerskans et al., 2020). However, more work is needed in order to fine-tune the algorithms in particular the LSBoost algorithm, which suffered from limited time (and computational) resources.

In this study, no attempts have been made to estimate uncertainties. However, it is good practice to provide uncertainties with SST retrievals and when inverting physical models each retrieval is accompanied with an uncertainty estimate directly through the optimal estimation method. This could be another focus of future work and it could be tackled in different ways. One approach could be to train multiple algorithms to obtain an uncertainty estimate. Multiple algorithms could also be trained to ensure that an algorithm with good generalization is used. Another approach could be to train a new uncertainty algorithm using the same input parameters and the target being the retrieved SST minus *in situ* SST instead. A third, and very interesting approach, is to model the a posteriori probability distribution of the output parameter (here SST) as a mixture of Gaussians and let the various parameters of the mixture model be given the outputs of a conventional neural network in a similar way as was done in Meier et al. (2007). Providing uncertainty estimates should be the focus of future work to further enhance ML retrievals of SST.

References

Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., Corrado, G. S., Davis, A., Dean, J., Devin, M., Ghemawat, S., Goodfellow, I., Harp, A., Irving, G., Isard, M., Jia, Y., Jozefowicz, R., Kaiser, L., Kudlur, M., Levenberg, J., Mane, D., Monga, R., Moore, S., Murray, D., Olah, C., Schuster, M., Shlens, J., Steiner, B., Sutskever, I., Talwar, K., Tucker, P., Vanhoucke, V., Vasudevan, V., Viegas, F., Vinyals, O., Warden, P., Wattenberg, M., Wicke, M., Yu, Y., and Zheng, X.: TensorFlow: Large-Scale Machine Learning on Heterogeneous Distributed Systems, arXiv:1603.04467 [cs], 2016.

Alerskans, E., Høyer, J. L., Gentemann, C. L., Pedersen, L. T., Nielsen-Englyst, P., and Donlon, C.: Construction of a climate data record of sea surface temperature from passive microwave measurements, Remote Sensing of Environment, 236, 111485, https://doi.org/10.1016/j.rse.2019.111485, 2020.

Alerskans, E., Zinck, A.-S., Nielsen-Englyst, P., and Høyer, J. L.: Retrieval of sea surface temperatures from passive microwave measurements, In prep., 2021.

Chen, T. and Guestrin, C.: XGBoost: A Scalable Tree Boosting System, in: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '16: The 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco California USA, 785–794, https://doi.org/10.1145/2939672.2939785, 2016.

Haykin, S.: Neural networks: a comprehensive foundation., Prentice hall, 1999.

Kumar, C., Podestá, G., Kilpatrick, K., and Minnett, P.: A machine learning approach to estimating the error in satellite sea surface temperature retrievals, Remote Sensing of Environment, 255, 112227, https://doi.org/10.1016/j.rse.2020.112227, 2021.

Lee, A., Taylor, P., Kalpathy-Cramer, J., and Tufail, A.: Machine Learning Has Arrived!, Ophthalmology, 124, 1726–1728, https://doi.org/10.1016/j.ophtha.2017.08.046, 2017.

Meier, U., Curtis, A., and Trampert, J.: Global crustal thickness from neural network inversion of surface wave data, Geophysical Journal International, 169, 706–722, https://doi.org/10.1111/j.1365-246X.2007.03373.x, 2007.

Meissner, T. and Wentz, F. J.: The Emissivity of the Ocean Surface Between 6 and 90 GHz Over a Large Range of Wind Speeds and Earth Incidence Angles, IEEE Transactions on Geoscience and Remote Sensing, 50, 3004–3026, https://doi.org/10.1109/TGRS.2011.2179662, 2012.

Merchant, C. J., Saux-Picart, S., and Waller, J.: Bias correction and covariance parameters for optimal estimation by exploiting matched in-situ references, Remote Sensing of Environment, 237, 111590, https://doi.org/10.1016/j.rse.2019.111590, 2020.

Minnett, P. J., Alvera-Azcárate, A., Chin, T. M., Corlett, G. K., Gentemann, C. L., Karagali, I., Li, X., Marsouin, A., Marullo, S., Maturi, E., Santoleri, R., Saux Picart, S., Steele, M., and Vazquez-Cuervo, J.: Half a century of satellite remote sensing of sea-surface temperature, Remote Sensing of Environment, 233, 111366, https://doi.org/10.1016/j.rse.2019.111366, 2019.

Moschos, E., Schwander, O., Stegner, A., and Gallinari, P.: Deep-SST-Eddies: A Deep Learning Framework to Detect Oceanic Eddies in Sea Surface Temperature Images, in: ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Barcelona, Spain, 4307–4311, https://doi.org/10.1109/ICASSP40776.2020.9053909, 2020.

Nielsen, M. A.: Neural networks and deep learning., San Francisco, CA: Determination press, 2015.

Nielsen-Englyst, P., L. Høyer, J., Toudal Pedersen, L., L. Gentemann, C., Alerskans, E., Block, T., and Donlon, C.: Optimal Estimation of Sea Surface Temperature from AMSR-E, Remote Sensing, 10, 229, https://doi.org/10.3390/rs10020229, 2018.

Paul, S. and Huntemann, M.: Improved machine-learning-based open-water–sea-ice–cloud discrimination over wintertime Antarctic sea ice using MODIS thermal-infrared imagery, The Cryosphere, 15, 1551–1565, https://doi.org/10.5194/tc-15-1551-2021, 2021.

Saux Picart, S., Tandeo, P., Autret, E., and Gausset, B.: Exploring Machine Learning to Correct Satellite-Derived Sea Surface Temperatures, Remote Sensing, 10, 224, https://doi.org/10.3390/rs10020224, 2018.

Wentz, F. J. and Meissner, T.: AMSR_E Ocean Algorithms, Remote Sensing Systems, Santa Rosa, CA, 2007.

Appendix

Inputs	Satellite orbit	'amsr_orbit'
	Latitude	'amsr_lat'
	Longitude	'amsr_lon'
	Solar Zenith Angle	'amsr_solza'
	Satellite Zenith Angle	'amsr_satza'
	Satellite Azimuth Angle	'amsr_sataz'
	Relative direction	'relative_dir'
	NWP wind speed	'nwp_wind_speed'
	NWP total column water vapor	'nwp_tcwv'
	NWP cloud liquid water content	'nwp_clwt'
	Brightness temperature 6V	'tb6V'
	Brightness temperature 6H	'tb6H'
	Brightness temperature 10V	'tb10V'
	Brightness temperature 10H	'tb10H'
	Brightness temperature 18V	'tb18V'
	Brightness temperature 18H	'tb18H'
	Brightness temperature 23V	'tb23V'
	Brightness temperature 23H	'tb23H'
	Brightness temperature 36V	'tb36V'
	Brightness temperature 36H	'tb36H'
	Brightness temperature 89V	'tb89V'
	Brightness temperature 89H	'tb89H'
Output	*In situ* sea surface temperature	'insitu_sst'

Table 1A. List of input features/predictors and the output SST from drifting buoys.

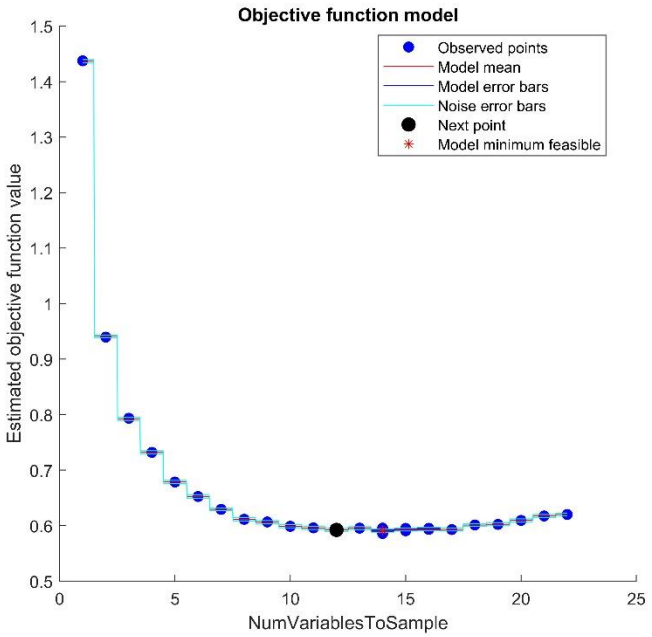


Figure 1A. Estimated objective function value for different numbers of variables to be sampled for the RF algorithm.

	ME	MAE	MSE	SD	RSD	R^2	AK*
Regression Tree	0.026	0.65	0.77	0.88	0.77	0.915	0.75
Bagging	0.024	0.55	0.54	0.74	0.66	0.940	0.75
Random Forest	0.017	0.51	0.45	0.67	0.61	0.950	0.68
LSBoost	0.008	0.57	0.55	0.74	0.69	0.939	0.78
Neural Network	-0.010	0.42	0.29	0.54	0.52	0.968	0.86
XGBoost	0.010	0.23	0.13	0.36	0.22	0.985	0.65
NN Tensorflow	-0.014	0.48	0.40	0.63	0.58	0.956	0.82

Table 2A. Performances of the different retrievals for the matchups located in the Arctic (>58°N). Number of matchups are 106,417.

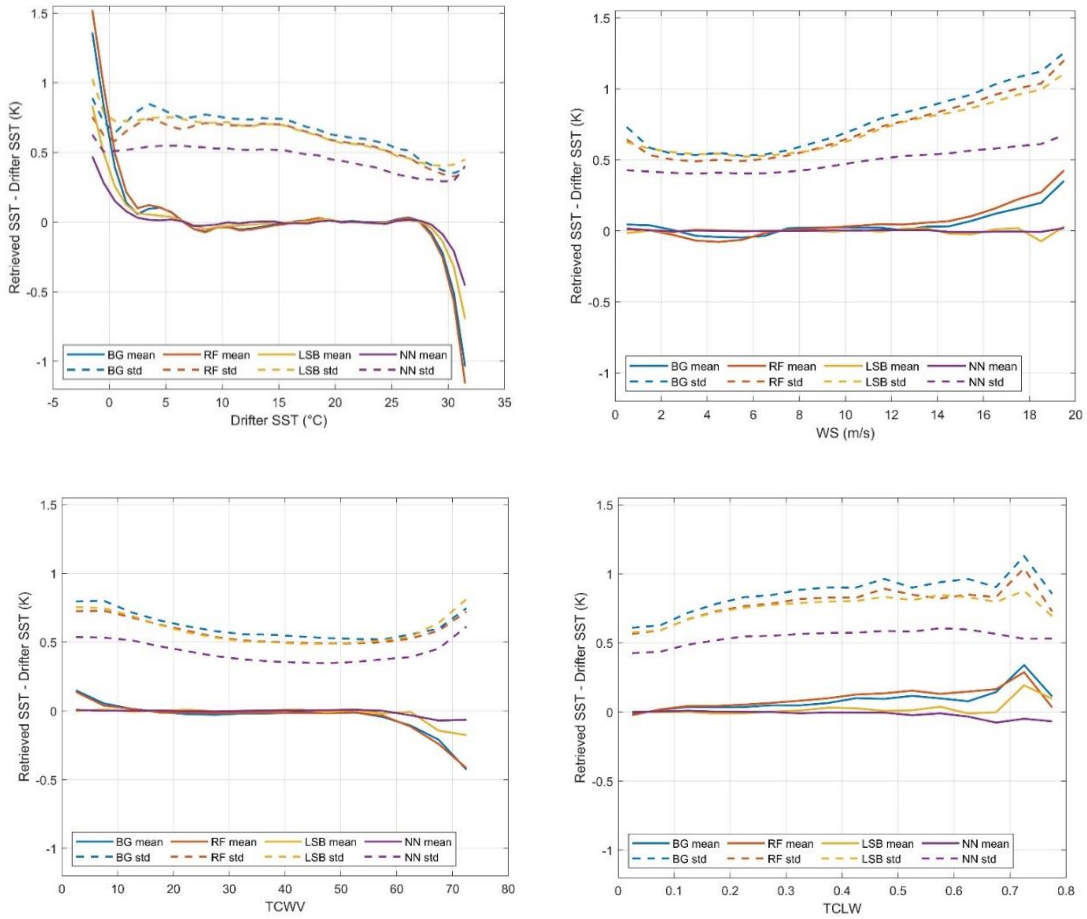


Figure 2A: Performance of retrieved SST compared to drifter SST as a function of a) drifter SST, b) wind speed, c) total column water vapour and c) total cloud liquid water.

Technical
University of
Denmark

Ørsteds Plads, Building 348
2800 Kgs. Lyngby
Tlf. 4525 1700

www.space.dtu.dk