**Oceanographic variability shapes the spawning distribution of blue whiting (*Micromesistius poutassou*)**

**ANNA K. MIESNER\*, MARK R. PAYNE***\* Corresponding Author (E-mail: amie@aqua.dtu.dk)*

*Centre for Ocean Life, National Institute of Aquatic Resources (DTU Aqua), Technical University of Denmark, Kemitorvet B202, 2800 Kongens Lyngby*

**ACKNOWLEDGMENTS**We wish to thank Sophie Pitois (Cefas) and David Johns (SAHFOS) for providing the CPR data. Moreover we wish to thank Georg H. Engelhard (Cefas) and a second, anonymous, reviewer whose comments at the review stage greatly improved the manuscript. The order of authors is determined following the “First-last-author-emphasis” norm.

**FUNDING**The research leading to these results has received funding from the European Community’s Seventh Framework Programme (FP7 2007-2013) under grant agreement No 308299 (NACLIM) and the European Union’s Horizon 2020 research and innovation programme under grant agreement No 727852 (Blue-Action). Analysis of the CPR fish larval samples from 1979 to 2005 was funded by the United Kingdom Department of Environment, Fisheries and Rural Affairs (Defra) through project MF1101. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

# ABSTRACT

The spawning distribution of blue whiting (*Micromesistius poutassou*) has varied considerably between years but quantitative understanding of the processes driving this change is lacking. Using 55 years of larval-observations from the wide-ranging Continuous Plankton Recorder (CPR) survey, we show that changes in the spawning distribution of blue whiting are associated with variations in the marine environment and particularly salinity. We first corroborated previously reported associations between variations in the spawning distribution and environmental regimes in the spawning region based on space-time interpolation models. We then applied species distribution models to quantify the linkage between the environment and the distribution of blue whiting larvae and verified these model results against independent fisheries and scientific survey data. Models incorporating salinity in the spawning region gave the best agreement with data, with observations of larvae in the CPR being limited to a window of salinities between35.3 and 35.5. Changes in the area of suitable spawning habitat (estimated here to be up to 2.5 times) can therefore be understood as arising from the spread of saline subtropical water masses throughout the spawning region due to a weak North Atlantic subpolar gyre. We postulate that blue whiting actively select optimum oceanographic conditions to deliver their eggs to enhance their offspring’s likelihood of survival and thereby their fitness. The knowledge derived here, together with the high predictability of salinity at depth in the North-East Atlantic, can potentially form the basis for forecasting the spawning distribution of this species.

**Key words**

spawning habitat, species distribution model, climate variability, continuous plankton recorder, North Atlantic subpolar gyre, blue whiting, *Micromesistius poutassou*

# INTRODUCTION

Blue whiting (*Micromesistius poutassou* (Risso,1827)) is a commercially important gadoid found throughout the North-East Atlantic. This species migrates annually between its feeding grounds in the Norwegian Sea to its spawning region west of Great Britain and Ireland (Fig. 1; Bailey, 1982; ICES, 2016a; ICES, 2016b). While there are indications of limited spawning activity outside of the main spawning area (Bainbridge and Cooper, 1973; Bailey, 1982), most spawning takes place from March to April along the European Continental Shelf edge and over banks to the west of Great Britain and Ireland (Bailey, 1982; Pointin and Payne, 2014).

The spatial distribution of blue whiting spawning, however, has been shown to vary substantially between years. Scientific surveys have shown that in some years the distribution extends all the way from Rockall Trough to Rockall Plateau and Hatton Bank (ICES, 2007), while in other years it is compacted close to the Continental Shelf edge (ICES, 2015). These shifts in distribution are typically attributed to the variability of the marine environment: already in 1909, Schmidt noted that blue whiting were sensitive to temperatures and salinities during spawning (Schmidt, 1909). A century later, Hátún *et al.* (2009b) suggested that under fresher and colder conditions in the spawning region, blue whiting mainly spawn along the European Continental Shelf edge west of Ireland, in particular on Porcupine Bank, while during more saline and warmer conditions, spawning expands further westward across Rockall Trough onto Rockall Plateau and shifts northward along the European Continental Shelf (Fig. 1). Unfortunately, the current understanding of these dynamics is largely qualitative in nature: it is not currently possible, for example, to make quantitative predictions of the spawning distribution. Such knowledge, however, could be potentially of great value in the scientific monitoring of the stock, its assessment and management, and even in optimising the performance of the fishery.

A key line of evidence supporting the current understanding stems from the Continuous Plankton Recorder (CPR) survey. The CPR survey is one of the most wide-ranging long term monitoring programmes of marine organisms in the world (Batten *et al.*, 2003) and has been influential in the history of the blue whiting fishery. The first records of blue whiting larvae in CPR samples were reported in the early 1950s (Henderson, 1953). Further CPR data revealed a great abundance and broad spatial distribution of blue whiting larvae in the North-East Atlantic, causing Raitt (1968) to conclude that the spawning stock would be sufficiently large for commercial exploitation and resulting in fisheries biologists becoming interested in the species (Polonsky, 1968; Bainbridge and Cooper, 1973). Since the subsequent development of the fishery in the late 1970s, CPR larval data has also been used to resolve the population structure of blue whiting (Pointin and Payne, 2014). In addition, Hátún *et al.* (2009a) used a spatial subset of CPR data from 1951 to 1970 along a narrow axis across the Rockall Trough to support their hypothesised link between the marine environment and blue whiting spawning.

Since much of this work was performed, the use of species distribution models (SDMs) to characterise the linkage between environmental variables and the distribution of organisms has become commonplace (Guisan and Zimmermann, 2000; Elith and Leathwick, 2009). SDMs correlate species observations with environmental variables to characterize a set of environmental conditions (the “niche”) where the species can occur. Once the relationship between distribution and the environment is appropriately parameterized, spatially resolved fields of environmental variables can then be used with the SDM to produce predictions of the potential habitat and/or geographical distributions of the species(Araújo and Guisan, 2006; Kearney and Porter, 2009; Araújo and Peterson, 2012). A strength of SDMs is their ability to estimate spatial distributions and habitat from pointwise observations, a particularly valuable attribute that can add value to data that is unevenly distributed in time and/or space (e.g. CPR data). Furthermore, SDMs can be used to both predict and project environmental conditions beyond the conditions where they were developed, which is useful when considering the responses of an organism to climate change or climate variability.

In this work, we use SDMs to elucidate the mechanisms that are important for regulating blue whiting’s spawning distribution. Our analysis is based on the unique and underutilised data set of blue whiting larvae obtained from the CPR survey. Building on previous work (e.g. Hátún *et al.*, 2009b), we use an expanded set of blue whiting CPR observations with an additional 35 years of data together with modern SDM modelling techniques to develop a quantitative model of the distribution. We first investigate the spawning distribution of blue whiting in time and space with the aim of developing a better picture of the modes of variability seen in the distribution and confirming previous results. Then we examine the role played by the marine environment in determining the spawning distribution of blue whiting relative to other processes, such as migration dynamics or bathymetric features. Finally, we validate our model results by comparing them to completely independent observations from fisheries and scientific surveys of the spawning distribution.

# MATERIALS AND METHODS

### Analysis strategy

Our analysis of the dynamics of blue whiting spawning distribution, and the factors controlling them, is based primarily on data obtained from the Continuous Plankton Recorder (CPR). We first applied space-time interpolation models (STI) to estimate average spawning distributions from this data, and how the distribution relates to marine climatic regimes. Species distribution models (SDM) were then developed to allow quantitative characterisation of the key environmental processes driving variability in the spawning distribution. The factors identified as important by the SDM were examined in detail using a standard suite of model checking tools. Finally, the CPR-based results were verified against independent data sources from fisheries and scientific surveys.

### Continuous Plankton Recorder (CPR) data

Blue whiting larval data from the Continuous Plankton Recorder (CPR) survey (Bainbridge and Cooper, 1973; Reid *et al.*, 2003) from 1951 to 2005 were obtained from the Sir Alister Hardy Foundation for Ocean Science (SAHFOS), Plymouth, UK. The CPR is a plankton sampling device that is towed behind vessels of opportunity at 7 to 10 m depth, allowing for continuous sampling of the upper mixed water column across major commercial shipping routes (Reid *et al.*, 2003; Richardson *et al.*, 2006). Water enters the device through a small (12.7 mm) square opening in the front of the recorder, and is subsequently filtered through a 270 µm silk screen that is continuously replaced and preserved in formalin (Richardson *et al.*, 2006). On shore, the silk is divided into samples corresponding to 10 nautical miles of towing and analysed under a microscope by taxonomists (Richardson *et al.*, 2006). The position in space and time of CPR measurements are characterized by the spatial and temporal mid-points (Richardson *et al.*, 2006).

For the purposes of this analysis, and following previous work (Pointin and Payne, 2014) the CPR data was geographically restricted to the region where blue whiting larvae were most commonly observed (Fig. 2a), and to the period between February and June, covering more than 99% of all larval-presences (Fig. 2b; Pointin and Payne 2014). The CPR data consisted of 34 422 CPR observations over 54 years (1951 - 2005) including 1 122 presences of blue whiting larvae between February to June (Fig. 2b).

### Physical and oceanographic data

The UK Met Office Hadley Centre’s EN4 ocean analysis product (version EN4.1.1) was used as the primary source of oceanographic data in the analysis (Good *et al.*, 2013). EN4 is based solely on observational data, and provides monthly, quality controlled objective analyses with a 1 degree spatial resolution and associated estimates of uncertainty.

A key feature of this data set, in the context of this analysis, is that it is also depth resolved, consisting of estimates of conditions at 42 vertical levels: blue whiting adults and larvae utilise a range of vertical habitats, however, and the appropriate choice of depth layer requires careful consideration. Spawning blue whiting adults are typically observed at depth ranges from 300 m to 500 or 600 m (ICES, 1996; ICES, 2014), with the majority of eggs and non-feeding larvae reported at depths between 300 - 400 m (Hillgruber and Kloppmann, 1999); 300 - 600 m (Ådlandsvik *et al.*, 2001) or 250 - 600 m (eggs: 250 - 450 and larvae ≤ 2.5mm length: 300 - 600; according to Coombs *et al.*, 1981). After spawning, the larvae require around 20 - 25 days to complete the ascent from the depth at which they are spawned to the sea surface (Ådlandsvik et al., 2001), where they can first be detected in the CPR. The majority of larvae found in the CPR samples are smaller than 6 mm in length (Bainbridge and Cooper, 1973): from an average growth rate of > 0.3 mm per day (Bailey and Heath, 2001) and hatching length of ≤ 2.5 mm (Coombs *et al.*, 1981; Ådlandsvik *et al.*, 2001), it is very likely that larvae captured by the CPR were spawned in the three weeks prior to capture.

This understanding can then be used as the basis for extracting relevant environmental data from EN4. Temperature (T) and salinity (S) variables representative of the spawning conditions experienced by adults and eggs (hereafter referred to as TSPAWN and SSPAWN) were extracted from the EN4 data set for the spatial location of each CPR observation (both presence and absence) over the depth bounds between 252 and 596 m (inclusive) and averaged vertically. However, the environmental data used were that one calendar month prior to the timing of the CPR observation to allow for the lag time between spawning and observation of larvae in the CPR (within the restrictions imposed by the monthly temporal resolution of EN4). While drift of eggs and larvae during this one month may blur the relationship between the actual environmental conditions in which they were spawned and those used in the model, the error introduced is not expected to be worse than that due to the coarse scale (1 degree) of the EN4 product and will be dampened by the strong spatial correlation present in these fields. Furthermore, preliminary analyses showed that the T and S at spawning depth within the month of the CPR observation were highly correlated with TSPAWN and SSPAWN (correlation coefficient = 0.98) indicative of the slow dynamics of these sub-surface water masses. This approach to extracting environmental conditions therefore appears appropriate. Finally, in cases where the water depth at the location of a CPR observation was shallower than 252 m, T and S closest to the sea floor were extracted.

As a proxy for the environmental conditions experienced by blue whiting larvae at the time of capture in the CPR, sea surface temperature (SST) and salinity (SSS) were extracted from the EN4 data set, corresponding to the average T and S between the sea surface and 10 m depth (Good *et al.*, 2013). In this case, the data extracted from EN4 corresponded directly to the month of the CPR observation (i.e. with no time lag).

Oceanographic data for each CPR haul was complemented with geographic-based variables. The water depth corresponding to each CPR observation was extracted by means of bilinear interpolation from NOAAs ETOPO1 product (Amante and Eakins, 2009). The associated slope (in degrees) was computed according to Horn (1981) based on 8 neighbouring cells located adjacent to each grid point using the terrain() function of the raster package (version 2.5.8; Hijmans, 2016) in R (version 3.3.2; R Core Team, 2016).

It is hypothesised that blue whiting larvae perform diel vertical migrations by ascending to the surface waters during dusk and dawn (Hillgruber and Kloppmann, 2000), which could affect the CPR’s capture efficiency as pointed out by Pointin and Payne (2014). To quantify light conditions in the upper water column at the time of capture, the angle of the sun measured from the horizon upwards at the time and point of observation was calculated using the solarpos() function in the maptools package (version 0.8.39; Bivand and Lewin-Koh 2015) in R.

### Marine environment of the study region

We calculated a time series indicative of the state of the marine environment in the spawning region of blue whiting from 1951 to 2016 (red box, Fig. 2a) by area-weighted averages of T and S over blue whiting’s spawning depth (~250 - 600m) in regions with water depths ≥ 600 m for each year during the main spawning period of blue whiting (March - May). We then calculated the 33rd and 67th percentile of these values over the 65 year period and partitioned the time-series into three environmental regimes to represent the broad forms of variability present in the system (Fig. 3).

### Space-time and species distribution modelling

#### Basic model structure

CPR data is provided as abundance categories (Richardson *et al.*, 2006) with the majority (ca. 60 %) of the reported blue whiting abundance data comprising of one, two or three larvae. Initial explorations examined this data using a spatial model similar to Pointin and Payne (2014) with an ordered categorical response variable (Wood *et al.*, 2015): however, this analysis suggested that there was little further information in the abundance categories and the data is therefore treated as presence-absence for the remainder of the study.   
Observations of blue whiting larvae were modelled using a Generalized Additive Model (GAM) with a binomial distribution and logit link structure: P(Xi = PRESENT|πi) ~ Bernoulli(πi) ; with logit(πi) = f() (1)  
where *Xi* is the presence/absence of observation *i* and πi is the probability (P) that blue whiting larvae are observed. The explanatory variables, regardless of whether they are spatial or environmental, are indicated by the function *f()*. All models were fitted using the mgcv()package (version 1.8.16) in R with a “gamma” parameter of 1.4 to avoid overfitting (Wood, 2006). Light conditions in the upper water column, as represented by the elevation of the sun, **, was included as a cubic spline smoother in all models.

Space-time interpolation (STI)   
In order to allow a simple characterisation of the spawning distribution in relation to the state of the marine environment, a model of the CPR data over space (latitude and longitude) and time (i.e. the day of the year, DOY) was employed, termed the “space-time interpolation” model (STI), which is similar to the model of Pointin and Payne (2014). A three-dimensional tensor-product smoother (Wood, 2006), with latitude, longitude and day-of-year as the dimensions was employed to characterise the space-time variability. In order to investigate whether the previously reported shifts in the distribution of spawning blue whiting in response to the climatic regime in the spawning region is also apparent in the larval CPR data, separate space-time models were also fitted for individual climatic regimes, and compared to a single baseline model using all data.

Species Distribution Models (SDMs) Species distribution models (SDMs) were used to examine the processes driving the spatial and temporal distribution of spawning. Due to the relatively large number of potential variables that could be considered, the development of the SDM took place in two steps – first, the effect of constant geographical features was considered in a model set termed geographical model set (GEO) and the “best” model identified. Environmental variables were then added to this model to develop the full SDM.

As blue whiting are typically found between 250 and 600 m, their distribution can be expected to be shaped, at least in part, by oceanic bathymetry. There is little evidence to support spawning in shallow waters and we therefore considered a depth term in our models to account for this effect. Similarly, spawning blue whiting are commonly associated with the Continental Shelf edge (ICES, 2016a) and we therefore considered the bottom-slope as an additional term in the model.

A particular challenge in developing SDMs for this species stems from the large migrations that blue whiting undertake. While the feeding (and overwintering) grounds for this species are mainly in the Norwegian Sea, spawning occurs to the west of Great Britain and Ireland (Bailey, 1982; ICES, 2016b). After spawning, the species returns quickly to the Norwegian Sea feeding grounds in time to take advantage of the productivity in this region in late-spring and summer. The need to return to the feeding grounds therefore can be expected to introduce an additional consideration regarding where to spawn: migrating far from the feeding grounds in search of optimal spawning habitat incurs a penalty due to both the extra migration distance required and the risk of missing part or all of the feeding season. Such behavioural considerations are potentially problematic for the standard species distribution modelling approach, however, which assumes that the environmental niche is the only factor determining distribution (e.g. Wiens *et al.*, 2009), rather than, in this case, additional constraints imposed by migration and life-history considerations. Here we have circumvented this problem by incorporating the potential costs of migration-distance explicitly into our species distributions models. We represent this process with a two-dimensional interaction term (tensor-product smoother) between latitude and day-of-year, as the cost and feasibility of returning to the feeding grounds will clearly depend on both distance and time: spawning further south, for example, could be a feasible proposition if it were to happen earlier in the year, allowing sufficient time to return. The form of this term, however, is left unspecified and is fitted as part of the modelling procedure. In addition, this term also allows for changes in the timing of spawning with latitude to be incorporated, in line with previous results (Pointin and Payne, 2014).

In many fish species, spawning fish as well as eggs and larvae are particularly sensitive to ambient environmental conditions since their tolerable ranges are more restrictive than those of other life-history stages (Pörtner and Peck, 2010; and references therein). In particular, salinity has also been proposed as being critical via its effect on water density and therefore the buoyancy of marine fish eggs, including those of blue whiting (Sundby and Kristiansen, 2015) and has shown to be important for blue whiting larvae (Ådlandsvik *et al.*, 2001). An ensemble of different temperature and salinity structures was therefore considered as explanatory variables and incorporated into the model structure based on this *a priori* reasoning (Anderson, 2008).

### Model validation and assessment

Model validation metrics To assess a model’s goodness of fit, several standard measures were employed. In generalized linear and generalised additive models the “explained deviance” (Dev. Expl.) is an analogue to the coefficient of determination (R2) and was used here as an overall indicator of model quality. The Akaike Information Criterion (AIC), which measures the trade-off between model fit and model complexity (Burnham *et al.*, 2011), was used as the primary tool for model selection: in a set of models, the model with the smallest AIC has the smallest information loss and is therefore defined as the “best” (most parsimonious) model within the set (Anderson, 2008; Burnham *et al.*, 2011). The difference in AIC relative to the smallest AIC value in the model set (∆AIC), provides an easy way to compare and rank models (Anderson, 2008; Burnham *et al.*, 2011): the “best” model within a set has by definition ∆AIC ≡ 0. However, evaluating models based on the same data set used for calibration has been proposed to be a poor approach (Guisan and Zimmermann, 2000). Therefore, for the purpose of evaluation, 5-fold cross validation (CV) was employed, where the models were trained with 80% of the data and predictions subsequently made and validated against for the remaining 20%. Data partitioning was performed based on years, with every 5th year being incorporated into the same fold. The prediction-based model validation metrics were derived for each of the five iterations and their mean calculated.

The ability of the models to discriminate between the presence and absence of larvae were derived from a confusion matrix, summarizing the four possible outcomes between modelled presence/absence predictions and the validation data set (Table 1). The threshold used for translating the predicted probability of blue whiting larval-occurrence into presences and absences was chosen for each model so that the total number of presences in the prediction data set was equal to the number of presences in the observed dataset, in accordance with Freeman and Moisen (2008). The positive predictive value (PPV) is the proportion of hauls where the presence of larvae is predicted and was in fact a larval-presence:

(2)  
where TP (true positive) and FP (false positive) are the elements of the confusion matrix (Table 1). The negative predictive value (NPV) is the proportion of sites where no larvae are predicted that are actually an absence:

(3)  
where TN are the true negatives and FN the false negatives. The true skill statistic (TSS) is the average of the net prediction success rate for presences and absences (Liu *et al.*, 2011) and independent of prevalence (Allouche *et al.*, 2006) and calculated by:   
 TSS= sensitivity + specificity -1, (4)  
where sensitivity is the probability that the model correctly predicts a presence at a site  
 (5)  
and specificity is probability that a known absence site is correctly predicted  
 (6)  
Accordingly, when the TSS is 1 it indicates that the model accuracy is perfect, while a TSS of zero is associated with a purely random model. The best performing models within the set were also validated using standard model diagnostic tools appropriate to generalised additive models e.g. the use of simulation-based quantile-quantile plots (Augustin *et al.*, 2012).   
The relative importance of each variable was evaluated using permutation importance. Data for each explanatory variable was randomly reassigned to a different CPR observation from that which it was originally associated with: in this way, any potential relationship between larval-presence/absence and the environmental variable is broken, while the statistical properties of the data remain unchanged. The model was then refitted and the predictive performance in terms of TSS and PPV calculated, as described before. The process was repeated for each variable and the difference between the original model and the TSS and PPV for each of the randomised models was calculated. A strong reduction in the performance metrics indicates a greater sensitivity of the model to the variable that has been randomized.

Model visualisation Estimated larval distributions were visualised by applying models based on the full CPR data set (i.e. not based on the cross-validation data sets) on a regular grid in space and time. The data used for model predictions was extracted by means of bilinear interpolation onto a regularly spaced grid of 0.25 x 0.25 degree resolution within the study region (Fig. 2a). For spatial maps, and due to the monthly resolution of the EN4 data, the 15th of each month was used as the day-of-year (DOY), while solar elevation angle (**) was fixed to 0°, representing the time of sunrise or sunset. To show the progression of spawning over time with respect to latitude, predictions were made for each DOY from February to June. Afterwards, the predicted larval-observation probabilities were averaged over longitude for each DOY.

Comparison to fisheries & survey data The SDM results were also validated by comparison with entirely independent data sources from fisheries and scientific surveys. This comparison took place by comparing the relative distributions of each of these data sources with respect to the key environmental variables identified by the SDMs to check for agreement between data sets: observations from fisheries and scientific surveys were matched-up with the corresponding environmental data derived from the EN4 data. Comparisons of spatial distribution were also made between outputs from the best SDM and the individual data sources.

Spatially and temporally resolved catch statistics of adult blue whiting fished between 1977 - 2012 have been prepared by the North East Atlantic Fisheries Commission (NEAFC, 2013). This data is available gridded onto 0.5° latitude x 1° longitude pixels with monthly resolution: data from March, the peak timing of spawning, was used in this analysis.

Acoustic biomass survey data of blue whiting spawning aggregations from 1981 to 2013 was also available. Data prior to 2004 originate from Norwegian surveys on the spawning grounds of blue whiting, while data from 2004 onwards stem from the International Blue Whiting Spawning Stock (IBWSS) Survey that is carried out annually for two weeks from late March to early April (ICES, 2016b). The acoustic survey records data continuously along its cruise track and provides biomass estimated of blue whiting per 0.5° latitude x 1° longitude rectangle.

# RESULTS

### Marine climate in blue whiting’s spawning region

The mean temperature and salinity in the spawning area of blue whiting has varied greatly between years. However a distinct pattern is noticeable: the marine climate was typically either both saline and warm or to the contrary fresh (i.e. of low-salinity) and cold (Fig. 3, correlation between temperature and salinity (r) = 0.82). Years where salinity and temperature both exceeded the 67th percentile (S > 35.40 and T > 9.87 °C) were defined as belonging to the “saline and warm” regime (17 years, indicated by the red horizontal bars in Fig. 3), while years with mean salinity and temperature below the 33rd percentile (S < 35.38 and T < 9.64 °C) were defined as “fresh and cold” regime (17, indicated by the blue horizontal bars in Fig. 3).

The spatial distribution of temperature and salinity within the study region varies between regimes (Fig. 4). During the saline and warm regime water masses of subtropical origin with high salinity and temperature spread along Rockall Trough towards Rockall Plateau , with mean salinities during March at blue whiting’s spawning depth reaching ≥ 35.35 and temperatures ≥ 10 °C within Rockall Plateau region (Fig. 4a). Conversely, during the fresh and cold regime such high salinities are constrained within Rockall Trough and onto Porcupine Bank with more waters of North Atlantic origin pushing in from the west (Fig. 4b) due to the influence of a strong subpolar gyre (Hátún *et al.*, 2005).

### Space-time interpolation of CPR data

Simple interpolation of the CPR data, using a statistical model with a space-time smoother, revealed substantial differences in distribution associated with the different marine climate regimes (Table 2). A large decrease in AIC was seen when adding regime as a conditional term (STI2), indicating an improved model over that which is not conditional on regime (STI1, Table 2), although the predictive skill of STI2 was marginally worse, indicated by slightly lower TSS, PPV and NPV.

Comparison of the larval-observation probability during the two most extreme regimes revealed clear differences in distribution (Figure 5). A map of the interpolated larval distribution for April (the peak of larval-presence probability, corresponding to March spawning) highlights a greater westward and northward larval distribution during saline and warm conditions, peaking at the Continental Shelf edge with high probabilities extending from the shelf edge through the Rockall Trough and along the north-eastern edge of the Rockall Plateau (Fig. 5a): the peak at 45 °N is likely to be erroneous, since GAMs perform worse on the edges as opposed to the centre of the modelling domain (Wood, 2006). In contrast, during the fresh and cold regime, the larvae are more concentrated within Rockall Trough and along the shelf edge (Fig. 5b). The probability difference plot (Fig. 5c) highlights regions that are most dissimilar between the regimes: during the saline and warm regime the probability of larval-presence is higher in the vicinity of Rockall Plateau, in particular towards its west and north up to 63 ºN, in contrast to the fresh and cold regime, when larvae are encountered with a higher probability along the shelf edge, on Porcupine Bank and extending into Rockall Trough. Also of note is the absence of larvae in the Porcupine Seabight during warm and saline years.

These differences are also reflected in the temporal changes in the distribution of larvae between the regimes. While a northward progression of spawning over time was apparent during both regimes, the onset and latitudinal distribution of larval appearance differed (Fig. 5d,f). During the fresh and cold regime, significant amounts of larvae are first seen in mid- to late-March at latitudes of around 50 °N, associated with of Porcupine Seabight spawning area (Fig. 5e). Conversely, larval appearance generally commences later at higher latitudes during the saline and warm regime when the Porcupine Seabight spawning area is not used (Fig. 5d), as highlighted by the probability difference plot (Fig. 5f). Furthermore, while the peak of larval occurrence probably occurs at approximately the same latitude and day-of-year in both regimes, larval-observation probabilities are generally higher during the warm and saline regime, peak probabilities (p ≥ 0.225) persist over a longer period and are found at higher latitudes compared to the fresh and cold regime.   
While we have focused primarily on the extremes of the distribution as a way to understand the modes of variability, it is important to note that the intermediate regime also occurs approximately one-third of the time. Comparison of the distributions of spawning in this regime in time and space reveal that it is, as the name suggests, intermediate between the two extremes, with a more northerly distribution and reduced (although not eliminated) spawning in Porcupine Seabight. Westward extension, however, is limited, and the distribution does not extend much beyond the Rockall trough. (Figure S2 and S3).

### Species distribution modelling

Model selection Species distribution modelling proceeded in two steps, first considering the optimal structure for geographical predictors, and then further considering environmental predictors on top of the best geographical model. Including depth into the geographical model (GEO2, Table 3) enhanced model skill compared to the starting model (GEO1, Table 3). Improvements were seen in terms of AIC, explained deviance, and enhanced predictive ability compared to the baseline model (Table 3). However, the addition of seabed slope did not increase model skill further, either alone (GEO3) or in combination with depth (GEO4, Table 3). For further analysis, the simpler geographical model structure GEO2 was as chosen as a refined baseline model for incorporating environmental variability.

The addition of environmental variables to the geographical baseline model further improved the fit and slightly increased the predictive performance (Table 4). Extending the baseline model to incorporate inter-annual variation of T and S reduced the AIC up to nearly 150 units, indicating a great improvement in model parsimony (Table 4): models with a difference in AIC above about 15 to 20 are judged as being strongly different (Anderson, 2008), meaning that this is strong evidence that inter-annual variation in environmental conditions plays a key role in shaping the spawning distribution of blue whiting.

However, the choice of the best environmental model is not immediately clear. The best three models (SDMs 3, 6, 7) are within 1 AIC unit of each other, and the fourth (SDM 8) is 8 AIC units different. Furthermore, while the predictive skill of these models varies, the differences are generally minor: any of these four models could potentially be chosen as the “best model” (for the sake of simplicity of interpretation we have chosen not to employ an ensemble approach or model averaging). Upon closer examination, we can see that the best of these models, SDM 7, gets good results using two-dimensional interactions between surface variables (SSS and SST) as it’s explanatory variables, even though neither of these terms show much skill individually (i.e. SDMs 1 and 2). Examination of the form of this smoother indicates dependencies that are hard to reconcile biologically and it is furthermore difficult to understand why surface variables would be determining for a species that lives primarily at depth. Conversely, all of the three remaining top models (SDMs 3, 6, 8) include spawning salinity (SSPAWN) as a predictor. The addition of spawning temperature to the model (i.e. from SDM 3 to 6) only makes for a very minor improvement, and a more-complex two-dimensional interaction between spawning salinity and temperature (SDM 8) has less support than salinity on its own (SDM 3). The biological interpretability of these models is an important feature and we therefore choose the simplest of them, SDM 3 (spawning salinity) as the “best” model for further analysis. However, it is important to bear in mind that this is a subjective choice, and that other models show comparable explanatory and predictive skill.

SDM visualisation Examining the individual terms in the SDM gives insight into the processes that it views as being important in shaping the distribution of blue whiting larvae. The 2D tensor-product smoother on latitude and day-of-year shows a clear relationship between the two variables, with peak larval-observation probabilities occurring later at higher latitudes (Fig. 6a). There are two centres of high larval-observation probability, one during mid-March at around 50 °N latitude (approximately the Porcupine Seabight area) and a greater region during mid-April to mid-May, reaching a peak observation probability between 57 °N and 59 °N (Fig. 6a). The smoother of the solar elevation angle (**) took positive values from -20° to 40° and assigned the greatest probability of observing larvae at around 10°, corresponding to the period shortly before sunset and/or after sunrise (Fig. 6b). The probability of observing larvae decreased and effectively vanishes in waters shallower than 300m, while is constant above this value (Fig. 6c). The larval-observation probability with respect to SSPAWN showed a clear window of high observation probability where the smooth took positive values between 35.3 and 35.5 (Fig. 6d), later referred to as the optimum salinity window.

Permutation importance experiments can be used to gain insight into the relative importance of each variable in the model fit. Permutation importance was calculated here as the relative loss in TSS upon scrambling of the variable(s) in question. Latitude and DOY, and in particular DOY, appeared to be the most important of these variables followed by depth (Fig. 7).

Model-based estimates of larval-presence probability were visualised for both regimes during April, the month with observed peak larval-presence, corresponding to March spawning (Fig. 2b). During the saline and warm regime, the main region of predicted high larval-observation probability extends from Rockall Plateau over Rockall Trough up to the European Continental Shelf between latitudes of 55 and 59 °N, in particular along Rockall Plateau’s eastern slope and within Rockall Trough (Fig. 8a). During fresh and cold conditions the main region of larval-observation probability is more contracted, spanning across Rockall Trough between 55 and 58 °N, with peak probabilities of observation along the Continental Shelf at 57 °N (Fig. 8b). Overall, the projected area of high observation probability (i.e. greater than 0.45) is 2.5 times larger during the saline and warm regime (147 000 km2), compared to the fresh and cold regime (59 000 km2).

A map of the differences in distribution between the two extreme regimes clearly highlights these changes. Areas of high larval-observation probability have the greatest west- and northward extent during saline and warm conditions and are especially high on Rockall Plateau, while during fresher and colder conditions they are more constricted towards the Continental Shelf and found slightly further south (Fig. 8c). In particular, more larvae are observed on Porcupine Bank and Seabight during fresher and colder conditions while they are virtually absent from Porcupine Seabight during the saline and warm regime (Fig. 8c).

Validation against independent data   
We found good agreement between the SDM’s response to salinities based on CPR observations of blue whiting larvae, and that inferred for adult fish, based on observations from fisheries and scientific surveys. The majority of adult blue whiting were encountered between salinities of 35.3 and 35.5 (Fig. 9). This result is in good agreement with the smooth function of SSPAWN of the SDM, which predicts the highest larval-observation probability at this salinity range. The low abundance of blue whiting at salinities < 35.3 is, however, not in reflected the SDM and may arise due to the capture of fish that are migrating to or from their spawning location.

Good spatial agreement is also seen between the estimated distribution of larvae from the SDM and independent observations from fisheries and scientific surveys targeting spawning adults (Fig. 10). In particular, the expanded, westward distribution of larvae during saline and warm conditions in 2007, and the contracted distribution during fresh and cold conditions in 1993 resemble observations from independent fisheries and scientific surveys targeting spawning adults well (Fig. 10). The optimum salinity window encompasses the majority of observations and delineates the maximum westwards extent specifically in 2007 (Fig. 10). However, observations of adult blue whiting reach further north and south than predicted by the SDM or expected from the optimum salinity window and observed catches from 1993 fail to match the predicted peak occurrence (Fig. 10): again, these may be due to fish on their way to or from the spawning grounds.

# DISCUSSION & CONCLUSION

This study provides a quantitative insight into the processes determining the spatial distribution of blue whiting spawning for the first time. By combining the wide ranging observations of blue whiting presence and absence from the CPR with species distribution modelling tools and ocean observation products, it is possible to characterise the processes driving the observed distributional shifts. Here we place these conclusions in a broader context.

### The marine environment impacts the spawning distribution

Firstly, the distributional shifts reported here agree with previously published work. The north- and westward expansion of the region with high probability of observing blue whiting larvae in the CPR (i.e. the potential spawning region) during saline and warm conditions, and its contraction towards the Continental Shelf during fresher and colder conditions (e.g. Fig. 5) has been observed previously (Bailey, 1970; Bainbridge and Cooper, 1973). Distributional changes of larvae (Hátún *et al.*, 2009b) and spawning adults (Hátún *et al.*, 2009a) have previously been linked to variations in the North Atlantic subpolar gyre (SPG). However, in view of the fact that the SPG is a basin-scale process, while individual fish respond to their local environment, our analysis offers a more nuanced perspective. Furthermore, our use of a statistical model and the focus on the presence-absence aspect of the CPR data circumvents many of the problems hidden in previous analyses such as the information content associated with CPR abundance values used by Hátún *et al.* (2009b) and biases resulting from the CPR’s opportunistic sampling design, it’s small sampling volume (Batten *et al.*, 2003), and the uneven and temporally varying distribution of blue whiting larvae within the water column (Hillgruber and Kloppmann, 1999; Ådlandsvik *et al.*, 2001; Pointin and Payne, 2014).

Our results confirm that blue whiting spawning distribution appears to be following inter-annual variations in the marine environment during spawning. The highest probability of observing larvae in the CPR was found betweensalinities of35.3 to 35.5: this possibly resembles the optimum salinity window for spawning in blue whiting, since independent observations from scientific and fisheries surveys targeting spawning adults also encountered the majority of adult blue whiting at this salinity range (Fig.10 and 11). Moreover, these results are in rough agreement with ichthyoplankton surveys that demonstrated peak spawning above salinities of 35.3 (Schmidt, 1909) and between 35.2 to 35.4 (Bailey and Heath, 2001). Our results therefore suggest that spawning is constrained by fresh and cold water masses in the spawning region associated with a strong SPG.

The apparent selection for this salinity range by spawning blue whiting may have a number of explanations. Blue whiting eggs are positively buoyant initially and their density increases during egg development enabling them to maintain a stable bathypelagic distribution (Ådlandsvik *et al.*, 2001). Subsequently, blue whiting larvae ascend passively through the water column towards the surface where they feed (Ådlandsvik *et al.*, 2001). Since variations in water density affect the buoyancy of eggs and larvae and thus their vertical distribution, changes in salinity can alter the ascent of larvae from their spawning depth towards the food-rich surface waters, which is critical for their survival (Ådlandsvik *et al.*, 2001).

Additionally, changes in the marine environment also affect the plankton community and thus the food conditions for blue whiting. Hátún *et al.* (2009a) have found that during a weak SPG, when the spawning region is influenced by saline, subtropical water masses, the zooplankton community switches from a dominance of *Calanus finmarchicus* to a greater abundance of smaller copepod species such as *Pseudocalanus, Acartia* and *Oithona.* These species represent the main food items of blue whiting larvae (Bailey, 1982) and it is therefore possible that shifts in the spawning distribution may in fact be a result of spawning adults choosing regions to spawn where there is abundant and suitable prey for the larvae, thereby maximising larval survival probabilities. The strong relationship between the potential spawning region and the ambient salinities found here may therefore arise in part as a response to the correlation between salinity and the planktonic community in this region. Alternatively, spawning blue whiting may be using salinity (which they most likely can sense) as a proxy for the planktonic community to be encountered by the larvae (which they most likely cannot sense). Based on this analysis alone it is not possible to determine which, if any, of these mechanisms is causative in nature, and further research is required to clarify this issue.

An interesting, although subtle, result from this analysis concerns the presence of blue whiting larvae in the Porcupine Seabight region. In a previous paper (Pointin and Payne, 2014) it has been suggested that these observations may represent a southern spawning population. However, the results gathered here show that there is essentially no spawning in this region during more saline and warmer regions (Fig. 5) (although larvae are present in this region in both the fresher and colder regime and the intermediate regime). The absence of larvae during this regime can be explained by the distribution of regions of suitable salinity (Fig. 4 and 10) and is a result that is replicated in the species distribution model (Fig. 8). The consequences of this result for the interpretation of a southern spawning component are unclear but suggest that more analysis and particularly more observations in this region are required.

A second important question raised by these results is how the shifts in spawning distribution relate to the large changes in recruitment seen in this stock. It has been proposed, for example, that the overlap between the distribution of blue whiting larvae and the distribution of mackerel in this area is a key factor determining recruitment success (Payne *et al.*, 2012). The hypothesis arises from the observation that mackerel distributions are typically restricted to close to the Continental Shelf edge: in years where the blue whiting distribution expands far to the west, beyond Rockall Plateau, the larvae spawned in these regions would be free from predation by mackerel. There is some empirical evidence to support this hypothesis: large year classes of blue whiting in the mid-late 1990s are associated with the collapse of the SPG and therefore an expanded spawning distribution. However, understanding the processes driving recruitment is notoriously tricky and much more work is required to resolve this issue: nevertheless, the model and the results developed here can be expected to make a valuable contribution to understanding the recruitment dynamics of this stock in the future.

### Limitations

The approach employed here to estimate the spatial distribution of blue whiting and the processes driving it has a number of limitations. In particular, our approach is limited by processes that are not included in the modelling approach, which can affect how the potential niche in environmental space is realized as a distribution in geographical space.

First of all, it is unlikely that all factors determining the niche of blue whiting has been captured by this approach. Due to the complexity of nature, one can never find all factors that determine a species niche (Wiens *et al.*, 2009), and even if one could, some relevant processes are difficult to observe and quantify, such as biotic interactions (i.e. competition and predation) or dispersal limitation (Elith and Leathwick, 2009; Colwell *et al.*, 2009). For example, our study cannot rule out that changes in the spawning distribution of blue whiting are caused by secondary processes resulting from changes in the marine environment that we have not characterised in our model, such as variations in the phyto- and zooplankton composition and abundance and thus in the food conditions for blue whiting (Payne *et al.*, 2012; Hátún *et al.*, 2009a). Migration behaviour and other aspects of the life-history (such as the need to return to the feeding grounds) are also particularly problematic (although we have attempted to compensate for them in our model structure). These and other processes can prevent the species from being in full equilibrium with the current climate, thereby violating a critical assumption of SDMs (Guisan and Zimmermann, 2000; Araújo and Peterson, 2012).

Furthermore, larval-presences are assumed to be recorded within physiologically suitable (environmental) conditions: in practice however, larvae might have drifted away from their initial spawning location into regions beyond suitable environmental conditions for spawning.Nonetheless, it is possible that blue whiting (instinctively or consciously) chose a region to spawn, where eggs and larvae are likely to be retained in a suitable environment or conversely where they drift within suitable water masses, as already hypothesised by Bailey (1982).

Another requirement of SDMs is that the environmental variables have appropriate temporal and spatial scales. The rather coarse spatial and temporal resolution of the temperature and salinity data might have overestimated potentially suitable spawning areas (i.e. the realised niche) and thus inaccurately amplified the area of high larval-observation probability. An extension of this work would take into account multiple ocean analysis and reanalysis products, with the goal of assessing the relative importance of this source of uncertainty.

Observations of adult blue whiting during spawning from both scientific surveys and fisheries data, however, also struggle to characterize their spawning habitat properly. The species distribution model developed here is based on larval-observations, and is therefore a very direct proxy for spawning distribution. However, while data obtained from fisheries and scientific surveys provide a snapshot of the adult distribution, it is ambiguous whether observed individuals are spawning or migrating. The lack of distinction between these two processes creates uncertainty in the correct geographic representation of the spawning adults, and also explains the partial mismatch in places between the modelled suitable habitat (from CPR observations) and observations from scientific surveys/fisheries (e.g. Fig. 10). Nevertheless, the general agreement between the model and the adult observations is satisfying, and increases our confidence in the results.

### Outlook

In this study, the observed changes of the spawning distribution of blue whiting have been clearly linked to inter-annual variations in the marine environment. The North Atlantic subpolar gyre region is one of the most predictable marine regions worldwide (Matei *et al.*, 2012; Meehl *et al.*, 2014) and therefore opens the door to forecasting this distribution. In particular, the switch from one marine climatic regime to another could be a key starting point: modelling studies have shown the ability to retrospectively predict the mid-1990s contraction of the SPG (Wouters *et al.*, 2013; Msadek *et al.*, 2014) and the associated increase in the upper 500 m heat content up to five years in advance (Robson *et al.*, 2017). Consequently, the onset of warmer and more saline conditions in the spawning region of blue whiting and an expanded spawning distribution of this species could potentially be predicted. The high predictive potential of the marine environment in the spawning region of blue whiting, coupled with the persistence of salinity at depth, might therefore enable us to forecast the extent of blue whiting’s spawning distribution at time scales relevant for the monitoring and management of this stock.

# REFERENCES

Ådlandsvik, B., Coombs, S. H., Sundby, S. and Temple, G. (2001). Buoyancy and vertical distribution of eggs and larvae of blue whiting (*Micromesistius poutassou*): observations and modelling. *Fisheries Research* 50**,** 59–72. doi:10.1016/S0165-7836(00)00242-3

Allouche, O., Tsoar, A. and Kadmon, R. (2006). Assessing the accuracy of species distribution models: Prevalence, kappa and the true skill statistic (TSS). *Journal of Applied Ecology* 43**,** 1223–1232. doi:10.1111/j.1365-2664.2006.01214.x

Amante, C. and Eakins, B. W. (2009). *ETOPO1 1 Arc-Minute Global Relief Model: Procedures, Data Sources and Analysis. NOAA Technical Memorandum NESDIS NGDC 24*.

Anderson, D. R. (2008). *Model based inference in the life sciences: A primer on evidence*. New York: Springer Science & Business Media.

Araújo, M. B. and Guisan, A. (2006). Five (or so) challenges for species distribution modelling. *Journal of Biogeography* 33**,** 1677–1688. doi:10.1111/j.1365-2699.2006.01584.x

Araújo, M. B. and Peterson, A. T. (2012). Uses and misuses of bioclimatic envelope modeling. *Ecology* 93**,** 1527–1539. doi:10.1890/11-1930.1

Augustin, N. H., Sauleau, E. A. and Wood, S. N. (2012). On quantile quantile plots for generalized linear models. *Computational Statistics and Data Analysis* 56**,** 2404–2409. doi:10.1016/j.csda.2012.01.026

Bailey, M. C. and Heath, M. R. (2001). Spatial variability in the growth rate of blue whiting (*Micromesistius poutassou*) larvae at the shelf edge west of the UK. *Fisheries Research* 50**,** 73–87. doi:10.1016/S0165-7836(00)00243-5

Bailey, R. S. (1970). Blue Whiting Stocks in the North-East Atlantic. *Scottish Fisheries Bulletin* 33**,** 4–8.

Bailey, R. S. (1982). The population biology of blue whiting in the North Atlantic. *Advances in Marine Biology* 19**,** 257–355.

Bainbridge, V. and Cooper, G. (1973). The distribution and abundance of the larvae of the blue whiting, Micromestistius poutassou (Risso), in the north-east Atlantic, 1948-1970. *Bull. Mar. Ecol.* 8**,** 99–114.

Batten, S. D., Clarke, R., Flinkman, J., Hays, G., *et al.* (2003). CPR sampling: the technical background, materials and methods, and issues of consistency and comparability. *Progress in Oceanography* 58**,** 193–215. doi:10.1016/j.pocean.2003.08.004

Bivand, R. S. and Lewin-Koh, N. (2015). maptools: Tools for Reading and Handling Spatial Objects. URL: CRAN.R-project.org/package=maptools

Burnham, K. P., Anderson, D. R. and Huyvaert, K. P. (2011). AIC model selection and multimodel inference in behavioral ecology: Some background, observations, and comparisons. *Behavioral Ecology and Sociobiology* 65**,** 23–35. doi:10.1007/s00265-010-1029-6

Colwell, R. K., Rangel, T. F., Grinnell, J. and Elton, C. (2009). Hutchinson’s duality : The once and future niche. *Proceedings of the National Academy of Sciences* 106**,** 1–8.

Coombs, S. H., Pipe, R. K. and Mitchell, C. E. (1981). The Vertical Distribution of Eggs and Larvae of Blue Whiting (*Micromesistius poutassou*) and Mackerel (*Scomber Scombrus*) in the Eastern North Atlantic and North Sea. *Rapports et Procès- Verbeaux des Réunions. Conseil International pour l’Exploration de la Mer* 178**,** 188–195.

Elith, J. and Leathwick, J. R. (2009). Species Distribution Models: Ecological Explanation and Prediction Across Space and Time. *Annu Rev Ecol Syst* 40**,** 415–436. doi:10.1146/annurev.ecolsys.l

Freeman, E. A. and Moisen, G. G. (2008). A comparison of the performance of threshold criteria for binary classification in terms of predicted prevalence and kappa. *Ecological Modelling* 217**,** 48–58. doi:10.1016/j.ecolmodel.2008.05.015

Good, S. A., Martin, M. J. and 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**,** 6704–6716. doi:10.1002/2013JC009067

Guisan, A. and Zimmermann, N. E. (2000). Predictive habitat distribution models in ecology. *Ecological Modelling* 135**,** 147–186. doi:10.1016/S0304-3800(00)00354-9

Hátún, H., Payne, M. R., Beaugrand, G., Reid, P. C., *et al.* (2009a). Large bio-geographical shifts in the north-eastern Atlantic Ocean: From the subpolar gyre, via plankton, to blue whiting and pilot whales. *Progress in Oceanography* 80**,** 149–162. doi:10.1016/j.pocean.2009.03.001

Hátún, H., Payne, M. R. and Jacobsen, J. A. (2009b). The North Atlantic subpolar gyre regulates the spawning distribution of blue whiting (*Micromesistius poutassou*). *Canadian Journal of Fisheries and Aquatic Sciences* 66**,** 759–770. doi:10.1139/F09-037

Hátún, H., Sandø, A. B., Drange, H., Hansen, B., *et al.* (2005). Circulation Influence of the Atlantic Subpolar Gyre on the Thermohaline Circulation. *Science* 309**,** 1841–1844. doi:10.1126/science.1116723

Henderson, G. T. D. (1953). Continuous Plankton Records: The young fish and eggs, 1932-39 and 1946-49. *Hull Bulletins of Marine Ecology* 3**,** 215–252.

Hijmans, R. J. (2016). raster: Geographic Data Analysis and Modeling. URL: http://cran.r-project.org/package=raster

Hillgruber, N. and Kloppmann, M. (1999). Distribution and feeding of blue whiting *Micromesistius poutassou* larvae in relation to different water masses in the Porcupine Bank area, west of Ireland. *Marine Ecology Progress Series* 187**,** 213–225. doi:10.3354/meps187213

Hillgruber, N. and Kloppmann, M. (2000). Vertical distribution and feeding of larval blue whiting in turbulent waters above Porcupine Bank. *Journal of Fish Biology* 57**,** 1290–1311. doi:10.1111/j.1095-8649.2000.tb00488.x

Horn, B. K. P. (1981). Hill Shading and the Reflectance Map. *Proceedings of the IEEE* 69**,** 14–47.

ICES (1996). *Report Of the Joint Norwegian-Russian Acoustic Survey on Blue Blue Whiting during Spring 1996. ICES CM H:12*.

ICES (2007). *Report of the Planning Group on Northeast Atlantic Pelagic Ecosystem Surveys (PGNAPES). ICES CM 2007/RMC:07*.

ICES (2015). *Report of the Working Group on Widely Distributed Stocks (WGWIDE), 25 August - 31 August 2015, Pasaia, Spain. ICES CM 2015/ACOM: 5*.

ICES (2016a). *Report of the Working Group on Widely Distributed Stocks (WGWIDE). ICES CM 2016/ACOM:16*.

ICES (2016b). *Stock Annex for Blue Whiting (Subareas I-IX, XI I and XIV). ICES Stock Annex 2016.*

ICES (2014). Widely distributed and migratory stocks. Blue Whiting (Subareas I – IX , XII , and XIV). ICES (International Council for the Exploration of the Sea). In *ICES Advice 2014 Book 9*. pp. 1–14.

Kearney, M. and Porter, W. (2009). Mechanistic niche modelling: Combining physiological and spatial data to predict species’ ranges. *Ecology Letters* 12**,** 334–350. doi:10.1111/j.1461-0248.2008.01277.x

Liu, C., White, M. and Newell, G. (2011). Measuring and comparing the accuracy of species distribution models with presence-absence data. *Ecography* 34**,** 232–243. doi:10.1111/j.1600-0587.2010.06354.x

Matei, D., Pohlmann, H., Jungclaus, J., Müller, W., *et al.* (2012). Two Tales of Initializing Decadal Climate Prediction Experiments with the ECHAM5 / MPI-OM Model. *Journal of Climate* 25**,** 8502–8523. doi:10.1175/JCLI-D-11-00633.1

Meehl, G. A., Goddard, L., Boer, G., Burgman, R., *et al.* (2014). Decadal Climate Prediction: An Update from the Trenches. *Bulletin of the American Meteorological Society* 95**,** 243–267. doi:10.1175/BAMS-D-12-00241.1

Msadek, R., Delworth, T. L., Rosati, A., Anderson, W., *et al.* (2014). Predicting a decadal shift in North Atlantic climate variability using the GFDL forecast system. *Journal of Climate* 27**,** 6472–6496. doi:10.1175/JCLI-D-13-00476.1

NEAFC (2013). *Report from the NEAFC Working Group on Collating Information on the Distribution of All Life Stages of Blue Whiting in the North-East Atlantic and the Distribution of Catches from the Stock, London, 26 – 28 November 2013*.

Payne, M. R., Egan, A., Fässler, S. M. M., Hátún, H., *et al.* (2012). The rise and fall of the NE Atlantic blue whiting (*Micromesistius poutassou*). *Marine Biology Research* 8**,** 475–487. doi:10.1080/17451000.2011.639778

Pointin, F. and Payne, M. R. (2014). A Resolution to the Blue Whiting (Micromesistius poutassou) Population Paradox? *PloS ONE* 9**,** 1–11. doi:10.1371/journal.pone.0106237

Polonsky, A. S. (1968). Materials on the biology of Poutassou. In *Rapports et Procès- verbeaux des Réunions. Conseil International pour l’Exploration de la Mer*. pp. 105–108.

Pörtner, H. O. and Peck, M. A. (2010). Climate change effects on fishes and fisheries: Towards a cause-and-effect understanding. *Journal of Fish Biology* 77**,** 1745–1779. doi:10.1111/j.1095-8649.2010.02783.x

R Core Team (2016). R: A language and environment for statistical computing. Foundation for Statistical Computing, Vienna, Austria. URL: https://www.R-project.org/

Raitt, D. F. S. (1968). The biology and commercial potential of blue whiting in the north-west Atlantic. In *Rapports et Procès- Verbeaux des Réunions. Conseil International pour l’Exploration de la Mer*. pp. 108–115.

Reid, P. C., Colebrook, J. M., Matthews, J. B. L., Aiken, J., *et al.* (2003). The Continuous Plankton Recorder: concepts and history, from Plankton Indicators to undulating recorders. *Progress in Oceanography* 58**,** 117–173. doi:10.1016/j.pocean.2003.08.002

Richardson, A. J., Walne, A. W., John, A. W. G., Jonas, T. D., *et al.* (2006). Using continuous plankton recorder data. *Progress in Oceanography* 68**,** 27–74. doi:10.1016/j.pocean.2005.09.011

Robson, J., Polo, I., Hodson, D. L. R., Stevens, D. P., *et al.* (2017). Decadal prediction of the North Atlantic subpolar gyre in the HiGEM high-resolution climate model. *Climate Dynamics* 0**,** 1–17. doi:10.1007/s00382-017-3649-2

Schmidt, S. (1909). The distribution of the pelagic fry and the spawning regions of the gadoids in the North Atlantic from Iceland to Spain. *Rapp. P.-V. Reun. Cons. Int. Explor. Mer* 10**,** 1–229.

Sundby, S. and Kristiansen, T. (2015). The principles of buoyancy in Marine Fish Eggs and their vertical distributions across the World Oceans. *PLoS ONE* 10**,** 1–23. doi:10.1371/journal.pone.0138821

Wiens, J. A., Stralberg, D., Jongsomjit, D., Howell, C. A., *et al.* (2009). Niches, models, and climate change: assessing the assumptions and uncertainties. *Proceedings of the National Academy of Sciences of the United States of America* 106**,** 19729–19736. doi:10.1073/pnas.0901639106

Wood, S. N. (2006). Generalized Additive Models: An Introduction with R. *Chapman and Hall/CRC Press* 410.

Wood, S. N., Pya, N. and Säfken, B. (2015). Smoothing parameter and model selection for general smooth models. *arXiv* . doi:10.1080/01621459.2016.1180986

Wouters, B., Hazeleger, W., Drijfhout, S., Oldenborgh, G. J. Van, *et al.* (2013). Multiyear predictability of the North Atlantic subpolar gyre. *Geophysical Research Letters* 40**,** 3080–3084. doi:10.1002/grl.50585

**TABLES**

Table 1. Confusion matrix used to evaluate the predictive accuracy of presence/absence models.  
TP (true positives): correctly predicted presences; FP (false positives): erroneously predicted presences; FN (false negatives): erroneously predicted absences; TN (true negatives): correctly predicted absences.

|  |  |  |  |
| --- | --- | --- | --- |
|  |  | **Validation data set** | |
|  |  | Presence | Absence |
| **Model** | Presence | TP | FP |
|  | Absence | FN | TN |

Table 2. Model fitting result of the space-time interpolation.   
Model formulation (*f()* in equation 1) is expressed as an equation where each term is implemented in the model using a either a 1D spline smoother (single terms) or a 2D tensor-product smoother (multiplied terms) and |regime indicates the term is conditional on the oceanographic regime (3 regimes: fresh & cold, saline & warm, and neutral regime). Lat: Latitude; Lon: Longitude; ** : solar elevation angle; DOY: day of the year; DevExpl: Explained Deviance; AIC: Akaike Information Criteria; ∆AIC: difference in AIC relative to the smallest AIC value within the model set; TSS: True Skill Statistic; PPV = Positive predictive value; NPV = Negative predictive value.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Model** | **Formulation, f()** | **DevExpl** | **AIC** | **∆AIC** | **TSS** | **PPV** | **NPV** |
| STI1 | Lon x Lat x DOY + ** | 0.419 | 5977 | 105 | 0.356 | 0.374 | 0.858 |
| STI2 | Lon x Lat x DOY |regime + ** | 0.451 | 5861 | 0 | 0.342 | 0.36 | 0.849 |

Table 3. Model fitting result of the geographic models.   
Model formulation, *f()*, where abbreviations are defined in Table 2. Models GEO2 – GEO4 use the baseline structure of GEO1, plus addition modifications of their own.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Model** | **Formulation, f()** | **DevExpl** | **AIC** | **∆AIC** | **TSS** | **PPV** | **NPV** |
| GEO1 | Lat x DOY + ** | 0.251 | 7465 | 1160 | 0.222 | 0.243 | 0.825 |
| GEO2 | GEO1 + Depth | 0.369 | 6305 | 0 | 0.331 | 0.35 | 0.849 |
| GEO3 | GEO1 + Slope | 0.318 | 6812 | 507 | 0.311 | 0.33 | 0.847 |
| GEO4 | GEO1 + Depth + Slope | 0.369 | 6305 | 0 | 0.332 | 0.351 | 0.853 |

Table 4. Model fitting results for models incorporating environment variables.  
Model formulation, *f()*, where abbreviations are defined in Table 2. *GEO2* is the best fitting geographical model (Table 3): the details of this model are duplicated here from Table 3 for reference.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Model** | **Formulation, f()** | **DevExpl** | **AIC** | **∆AIC** | **TSS** | **PPV** | **NPV** |
| GEO2 | Lat x DOY + ** + Depth | 0.369 | 6305 | 146 | 0.331 | 0.350 | 0.849 |
| SDM1 | GEO2 + SSS | 0.369 | 6304 | 145 | 0.332 | 0.350 | 0.849 |
| SDM2 | GEO2 + SST | 0.380 | 6213 | 54 | 0.344 | 0.362 | 0.851 |
| SDM3 | GEO2 + SSPAWN | 0.385 | 6160 | 1 | 0.356 | 0.373 | 0.850 |
| SDM4 | GEO2 + TSPAWN | 0.376 | 6243 | 84 | 0.337 | 0.355 | 0.849 |
| SDM5 | GEO2 + SSS + SST | 0.382 | 6195 | 36 | 0.337 | 0.356 | 0.852 |
| SDM6 | GEO2 + SSPAWN + TSPAWN | 0.385 | 6159 | 0 | 0.351 | 0.368 | 0.851 |
| SDM7 | GEO2 + SSS x SST | 0.385 | 6159 | 0 | 0.348 | 0.366 | 0.856 |
| SDM8 | GEO2 + SSPAWN x TSPAWN | 0.384 | 6167 | 8 | 0.343 | 0.361 | 0.847 |

# FIGURES

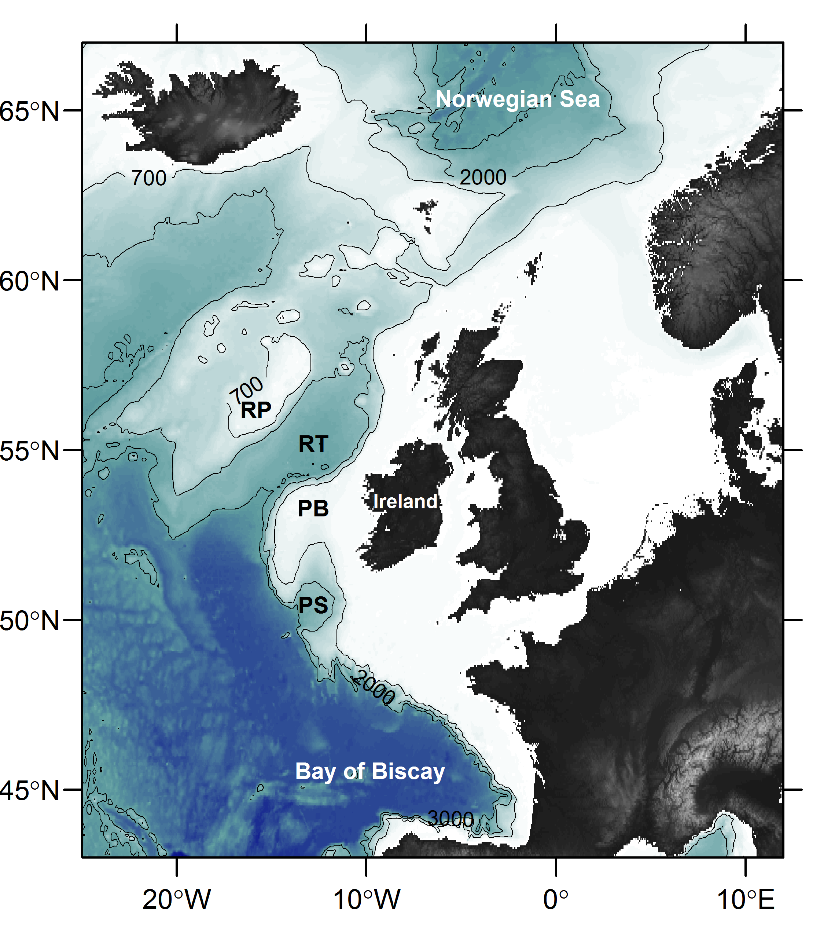


Figure 1. Bathymetric relief map of the study region. Contour lines indicate the water depth in meters (showing the isobaths of 700, 2000 and 3000 m). Geographic features mentioned in the text are labelled Rockall Plateau (RP), Rockall Trough (RT) Porcupine Bank (PB) and Porcupine Seabight (PS).

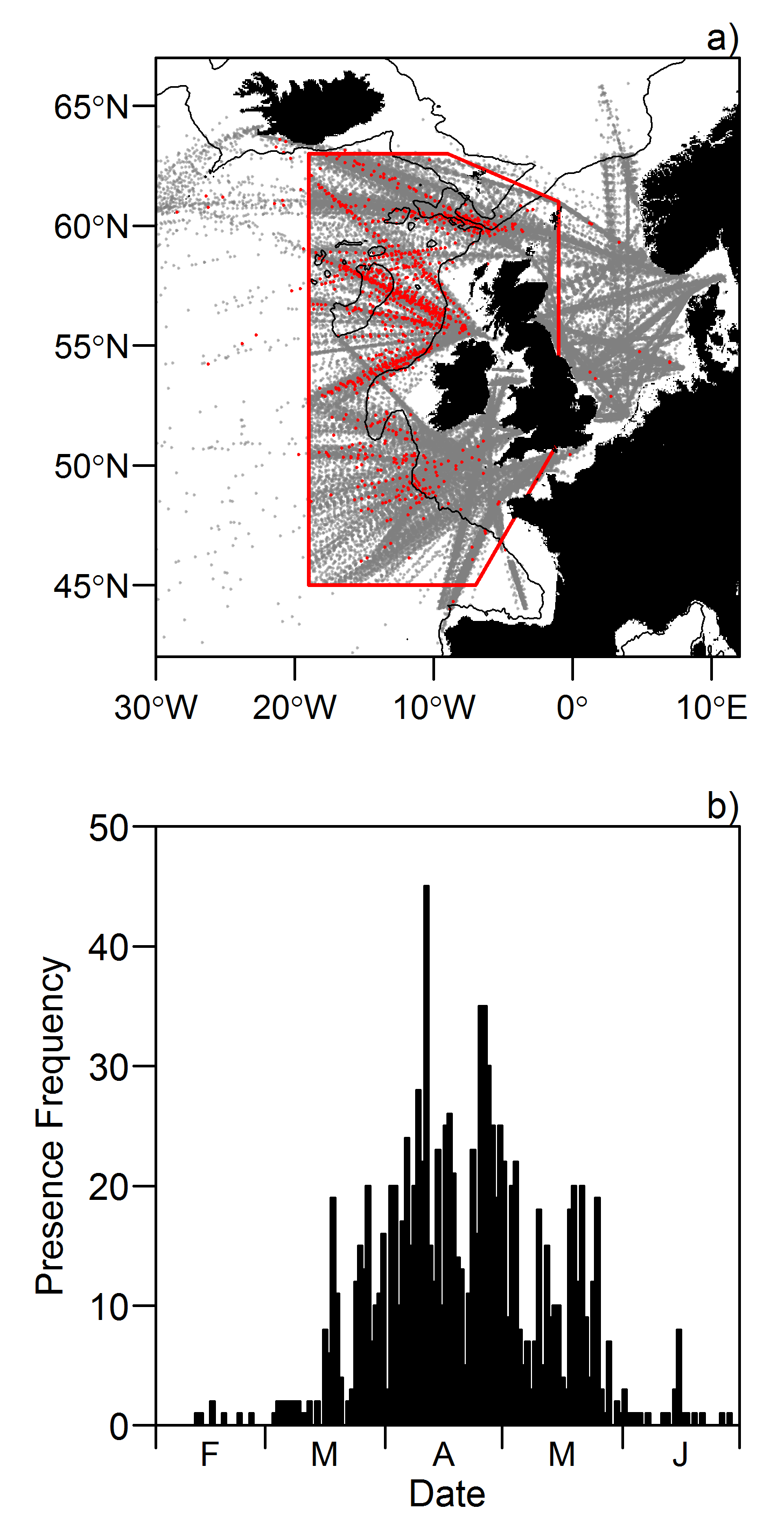


Figure 2. a) Spatial distribution of CPR samples from 1951 to 2005. Grey points indicate CPR sampling stations where blue whiting larvae were absent. Red points indicate CPR sampling stations where blue whiting larvae were present. The red polygon delineates the study region used in further model-based analysis. The black lines indicate the 700 m isobath. b) Temporal distribution of larval-presences (counts) within the study region shown in panel a), as a function of day in the year from February (F) to June (J). Each bar corresponds to a day.



Figure 3. Mean spring (March - May) temperature a) and salinity b) averaged over blue whiting’s spawning depth (250 to 600 m) within the study region (red box in Fig. 2a). Dashed horizontal lines indicate the median spring temperature and salinity within the study region from 1951 to 2016 (= 9.74 °C, = 35.39) and dotted lines the 33rd and 67th percentiles. Red colour indicates warmer/more saline conditions (T *>*, S > ) and blue colder/fresher conditions (T < , S < ). Horizontal bars indicate years defined by one of the two regimes: the saline and warm regime (red bar) with T and S above the 67th percentile, and the fresh and cold regime (blue bar) with T and S below the 33rd percentile.

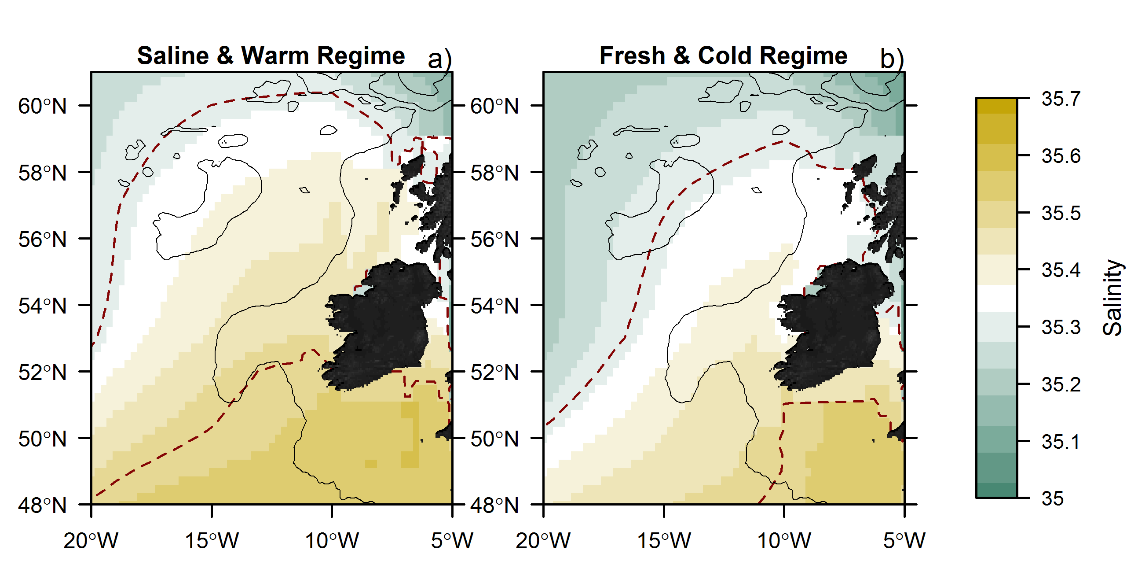


Figure 4. Salinity averaged over blue whiting’s spawning depth (250 to 600 m) during the saline and warm a) and fresh and cold b) regimes in March. The black solid lines indicate the 700 m isobaths and the red dashed lines the isoline of 35.3 and 35.5 psu averaged over blue whiting’s spawning depth, indicative of the optimum salinity range for spawning in blue whiting (Fig. 6d). In regions where water depth is less than 250m, bottom salinity is plotted.



**Figure 5.** Larval-observation probabilites, a proxy for blue whiting spawning distribution, based on space-time interpolation (STI2, Table 2), mapped for April (a-c) and as latitude-time plot from February to June (d-f), during the saline and warm a & d), and fresh and cold b & e) regimes (1951 - 2005); Figure c & f) depict the probability difference of larval-presence between the two regimes, i.e. the difference in larval observation probability between the saline & warm regime (a/d) and fresh & cold regime (b/e); with red colours corresponding to higher larval-observation probabilities during the saline and warm regime compared to the fresh and cold regime, while blue colours indicate the opposite, i.e. higher larval-observation probabilities during the fresh and cold regime. The black lines in a-c indicate the 700 m isobaths.

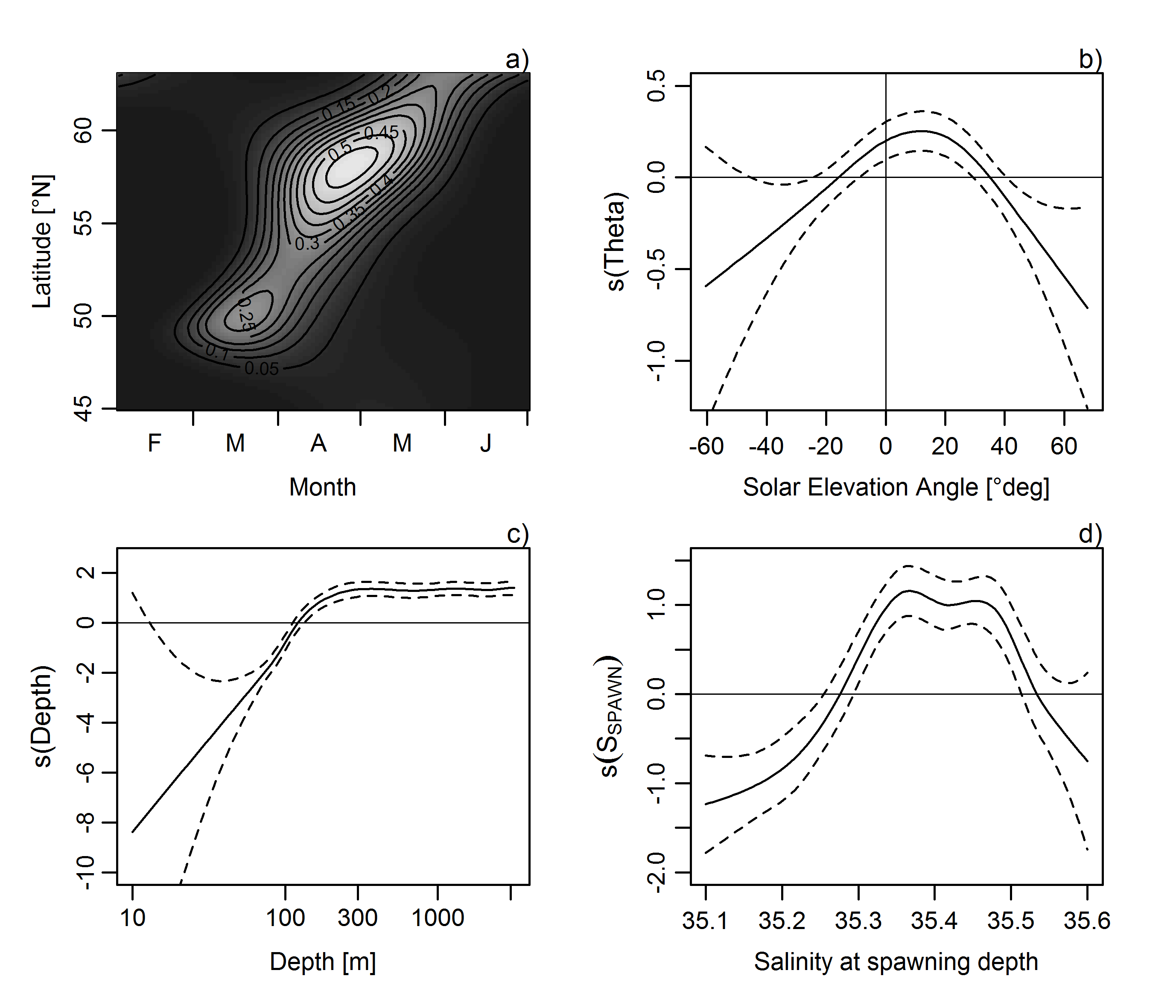


Figure 6. SDM components. a) Contour plot of the larval-observation probability as a function of latitude [in °N] and the day of the year (DOY, here shown as month). b-d) Smooth functions of the SDM with b) the solar elevation angle (** in °degrees), c) the depth [m], and d) the salinity at spawning depth during the spawning time of blue whiting (SSPAWN). In Figures b-d) the solid non-linear lines indicates the estimate of the smooth s() and the dashed lines indicate the 95% confidence interval. The vertical line in b) indicates sunrise/sunset, with positive values indicating that the sun is above the horizon, i.e. that there is light.

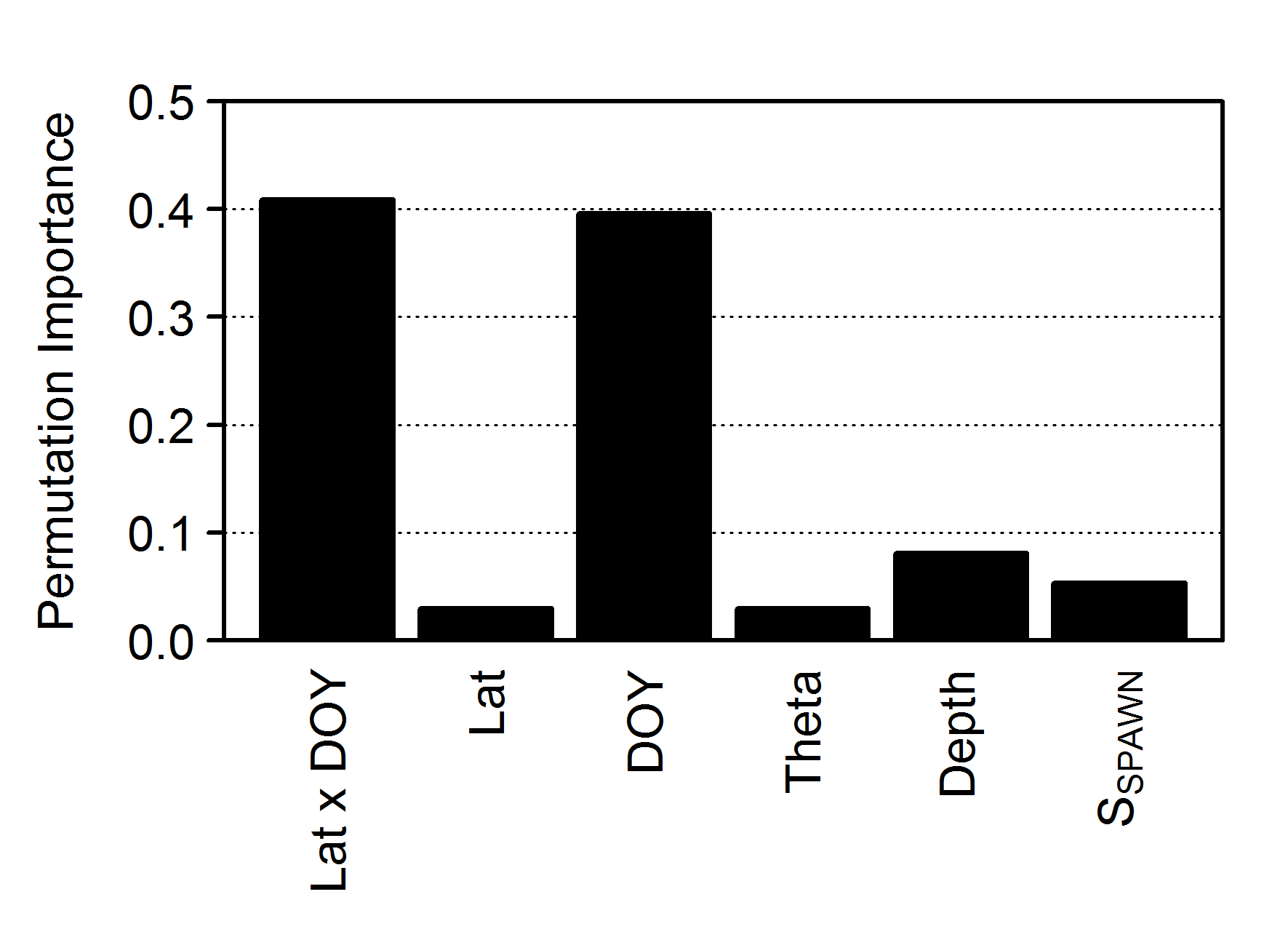


Figure 7. Permutation importance of variables. The difference in true skill statistics (TSS) between the original model and models with randomized explanatory variables is used as an indicator of the importance of each explanatory variable in the model fit: Latitude (Lat), day-of-year (DOY), solar elevation angle (Theta), log-transformed depth (Depth) and the salinity at spawning depth during the spawning time of blue whiting (SSPAWN). The “x” indicates that both Latitude and DOY variables within the 2D smoother were randomized at the same time. The permutation importance in terms of positive predictive prevalence (PPV) was very similar to TSS and is therefore not shown.

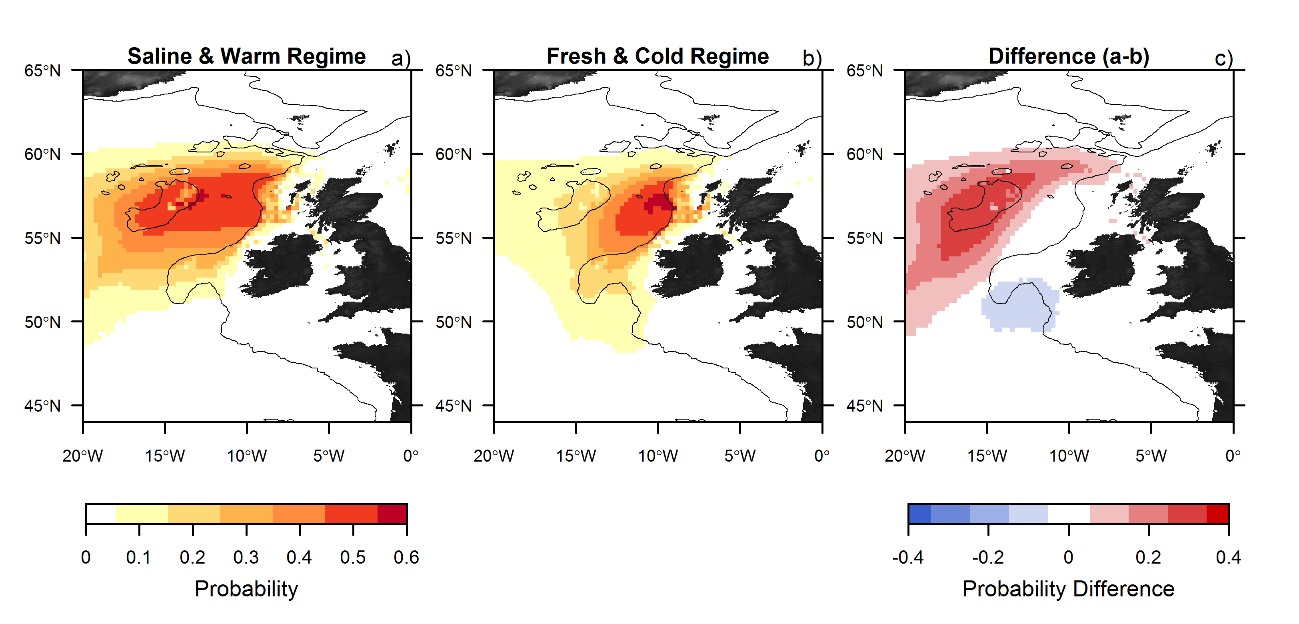


Figure 8. Map of CPR larval-presence probabilities (a proxy for blue whiting spawning distributions) estimated by the SDM in April during a) the more saline and warmer and b) and fresher and colder regime and c) the probability difference between the two regimes. The probability difference was calculated by subtracting the CPR larval-presence probability encountered during the saline & warm regime by that during fresh & cold regime. Accordingly, positive values (red) correspond to higher probability of larval observation during a saline & warm regime compared to the fresher & colder regime. The black lines indicate the 700 m isobaths.

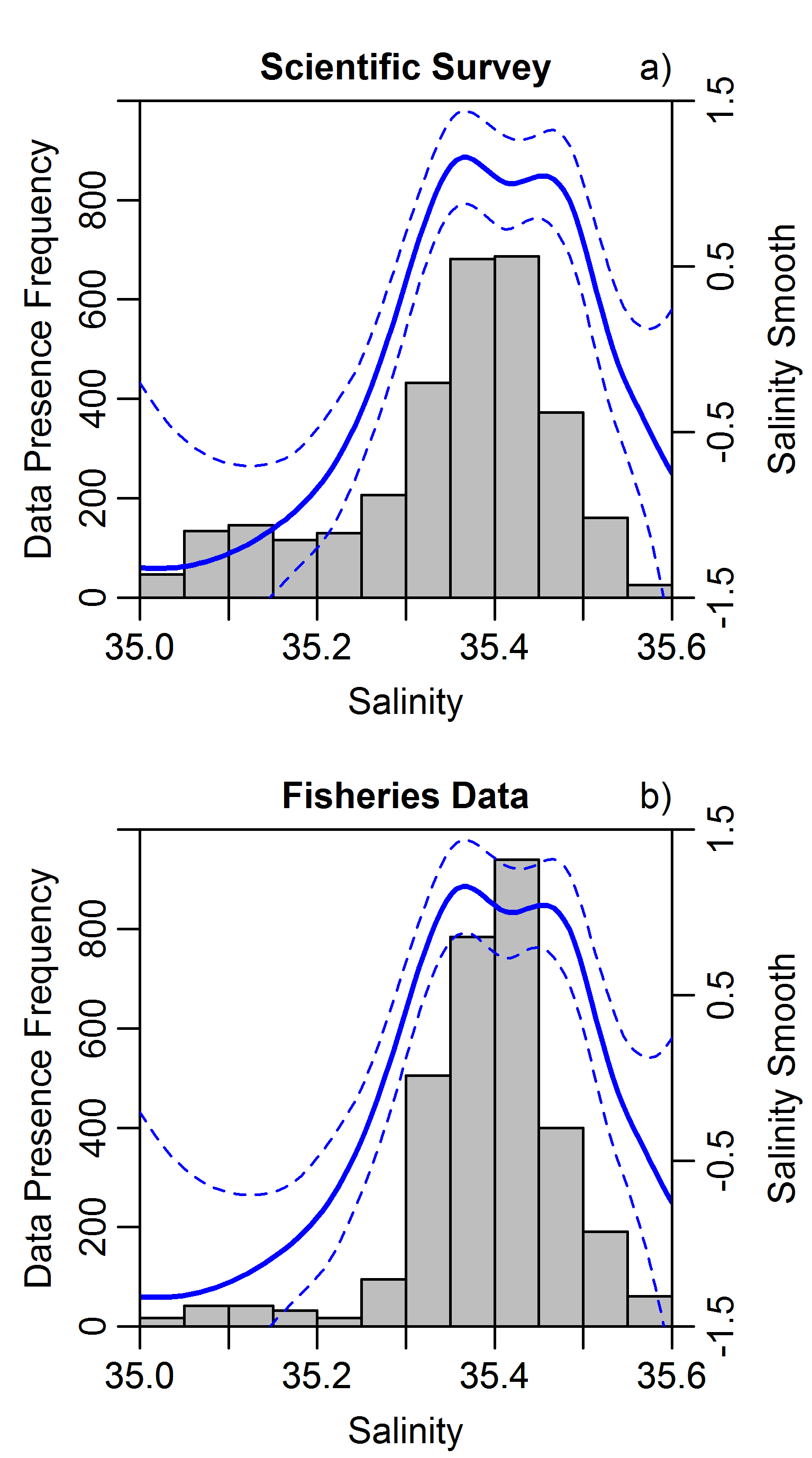


Figure 9. Presence frequency of spawning blue whiting (number of pixels with presences) observed in a) scientific surveys and b) caught in fisheries during March compared to the salinity at which these observations were made (bars). The solid (blue) line indicates the modelled smooth function of blue whiting larval-presence obtained from the SDM (Fig. 6d), with dashed lines indicating the 95% confidence interval.



Figure 10. Distribution of blue whiting in March 2007 during the saline and warm regime (left) and in March 1993 during the fresh and cold regime (right) obtained from the fisheries catch data (top, triangles) and the scientific surveys (bottom, circles). The blue isolines indicate the window of optimum salinity for spawning of blue whiting (35.3 < S < 35.0), averaged over blue whiting’s spawning depth (250 to 600 m) in March. The black lines indicate the 700 m isobaths and the grey shading the bathymetry. Colours correspond to the larval presence probability predicted to be observed in the CPR in the following month of the year in question (April).

# SUPPORTING INFORMATION

The geographical model was adapted to incorporate abundance as an ordered categorical variable by including the family argument “ocat” within the mgcv package including 7 categories in R (version 3.3.2; R Core Team, 2016). The first six categories corresponded to the abundance classes in which zooplankton and fish larvae are classified within the CPR data, consisting of larval abundances of : 0, 1, 2, 3, 4-11, 12-25 (Richardson *et al.*, 2006), while all abundance classes containing more than 25 larvae were grouped into one abundance class, since these consisted of less than 5% of the entire presence data**.** A generalized additive model of the abundance categories of blue whiting larvae with ordered categorical (ocat) family was set up (dev.expl. = 45.9 %) applying the same model formulation as STI 1 (Table 2).

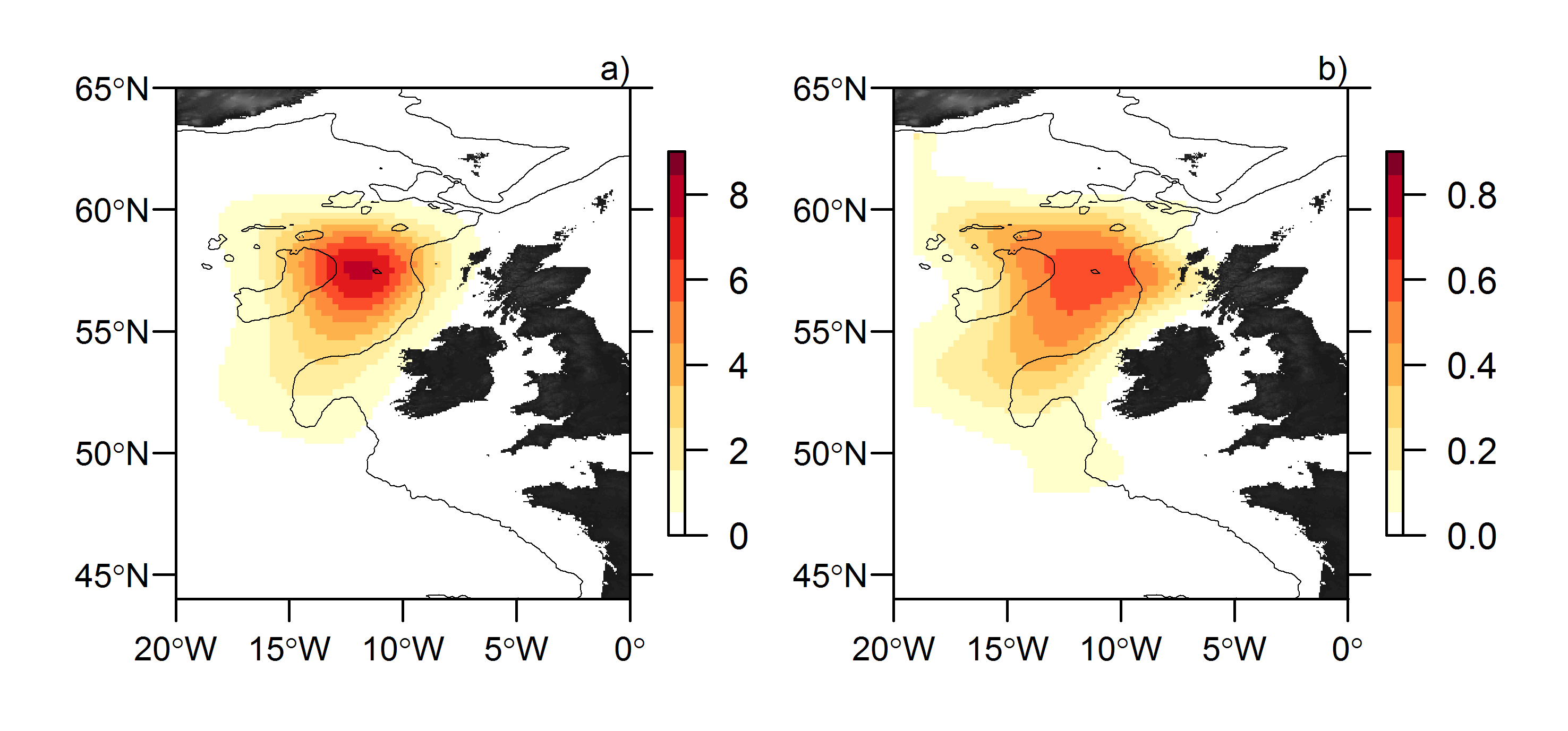
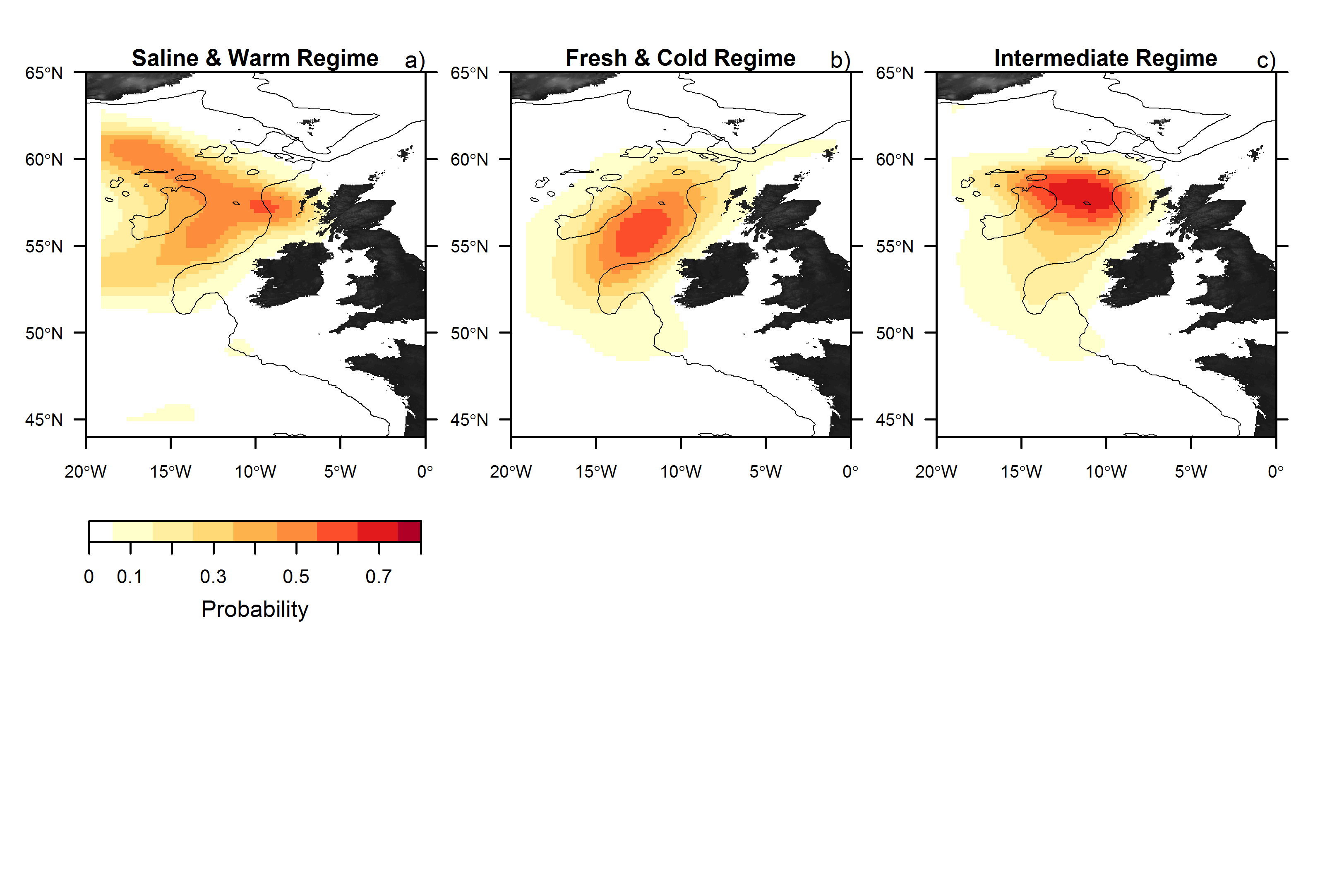


Figure S1. Map of the overall expected abundance [number] a) and presence probability b) of blue whiting larvae in April as modelled by STI 1 (Table 2). The expected abundance is the product of the predicted probability and the accepted value in each pixel summed up for each abundance class (Richardson *et al.*, 2006).



**Figure S2.** Map of predicted larval-presence probabilities as a proxy for blue whiting spawning distributions produced by the SDM during a) the more saline and warmer and b) and fresher and colder and the intermediate c) regime in April (1951-2016). The black lines indicate the 700 m isobaths.

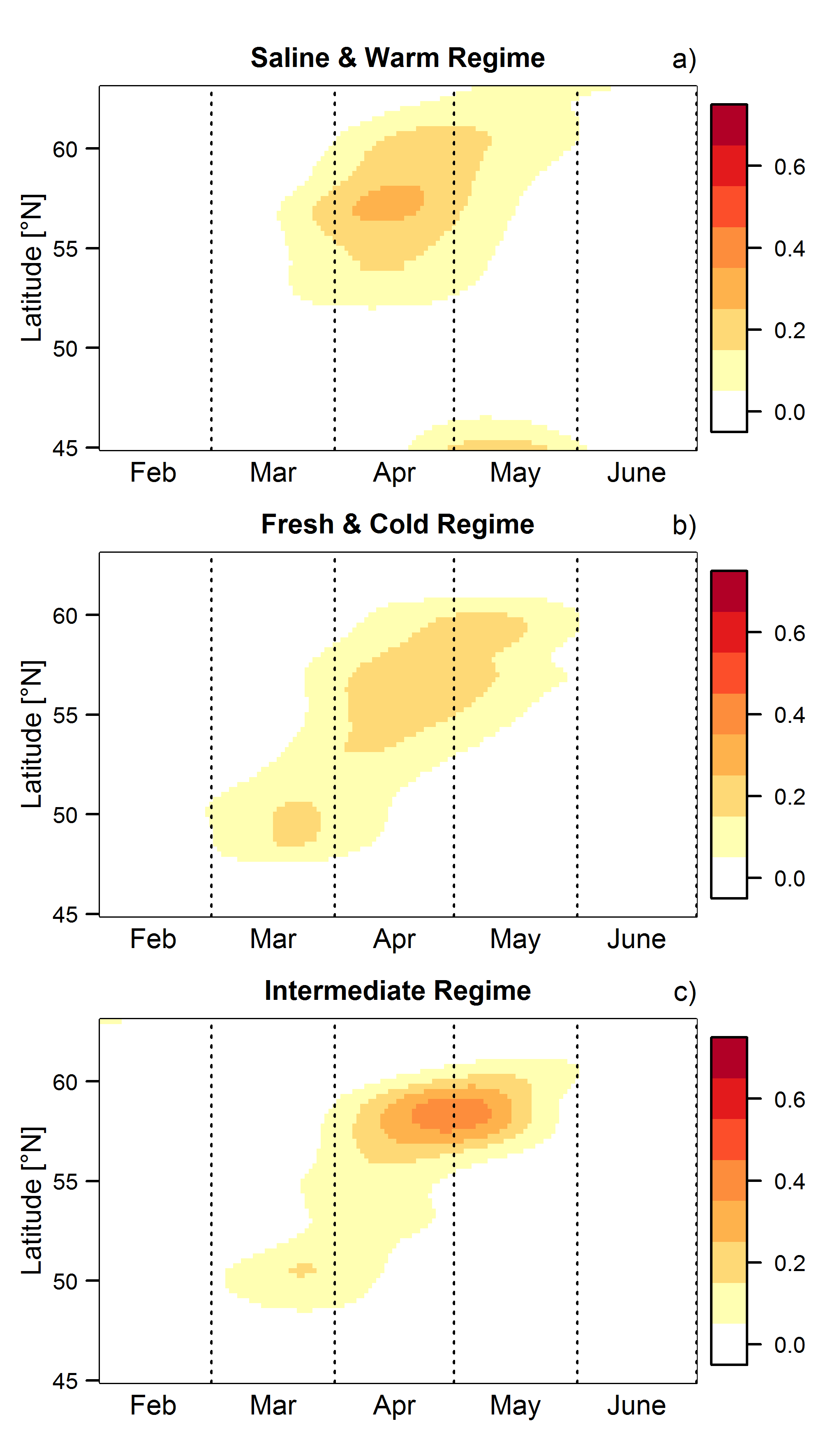


Figure S3. Latitude-Time plot of larval-observation probability during the saline and warm a), fresh and cold b) and intermediate regime c) of the space-time interpolation (STI2, Table 2).

*On the following pages:*

**Figure S4.** Annual distribution of blue whiting observed by the North East Atlantic Fisheries Commission (blue triangles) and the International Blue Whiting Spawning Stock Survey (red circles). The size of the symbol increases with the amount of blue whiting caught/observed. The black lines indicate the 700 m isobaths. The green area indicates the window of optimum salinity for spawning of blue whiting (35.3 < S < 35.0), averaged over blue whiting’s spawning depth (250 to 600 m) for March.

