Calibration of a distributed hydrology and land surface model using energy flux measurements

Larsen, Morten Andreas Dahl; Refsgaard, Jens Christian; Jensen, Karsten H.; Butts, Michael B.; Stisen, Simon; Mollerup, Mikkel

Published in: Agricultural and Forest Meteorology

Link to article, DOI: 10.1016/j.agrformet.2015.11.012

Publication date: 2016

Document Version
Peer reviewed version

Link back to DTU Orbit

Abstract: In this study we develop and test a calibration approach on a spatially distributed groundwater-surface water catchment model (MIKE SHE) coupled to a land surface model component with particular focus on the water and energy fluxes. The model is calibrated against time series of eddy flux measurements from three sites of different land surface type (agriculture, forest and meadow) and river discharge data from the 2500 km² Skjern River catchment in Denmark. The approach includes initial calibrations of three one-dimensional models representing the three land surface types using the flux measurements for calibration. This step provides initial values for the subsequent modelling and calibration at catchment scale. To test the validity of the approach, two additional catchment scale distributed simulations were performed with no calibration and only calibration of the one-dimensional models, respectively. In addition, a subsequent validation period was simulated. A mean energy closure imbalance of 20% was seen for the three sites. For the distributed simulations, the energy imbalance was accounted for by two energy balance closure hypotheses ascribing the error to either energy fluxes or net radiation. In general, the distributed calibration approach improved model results substantially compared to using default values (no calibration) or calibration of the one-dimensional models only. For the distributed model simulations, the assumption regarding the energy balance closure had a substantial impact on the parameter sensitivities and on the simulated discharge and energy balance. During calibration, the simulation with corrected energy fluxes showed better performance on discharge than the simulation with corrected net radiation whereas the reverse was true for the validation period. Regarding energy fluxes, the simulation with corrected net radiation was superior in both the calibration and validation period.
Dear Editor

With great pleasure we hereby submit our research paper “Calibration of a distributed hydrology and land surface model using energy flux measurements” for consideration in the Journal of Agricultural and Forest Meteorology.

Traditionally, distributed hydrology models often rely on potential evapotranspiration to estimate actual evapotranspiration rather than solving the energy equation. SVATs and LSMs on the other hand do solve the energy equation and are often calibrated in 1D or uses remote sensing data in distributed setups.

With the present paper we present a novel approach to calibrating a combined spatially distributed hydrology-land surface model using energy flux tower observations in combination with discharge reaching good results.

The present paper is not a resubmission and has no concurrent submissions to other journals.

Sincerely, Morten A D Larsen, corresponding author, madla@dtu.dk/+4525119895
Ms. Ref. No.: AGRFORMAT-D-15-00325
Title: Calibration of a distributed hydrology and land surface model using energy flux measurements
Agricultural and Forest Meteorology

I have now completed the review of the above manuscript. I apologize for the unusually long review time. We invited 14 different reviewers to assess your work. I have decided to make a decision based on one review and my own assessment of your manuscript.

In my opinion, the manuscript is extremely well written and addresses an important topic. I appreciate the objective and rigorous calibration approach and the testing of the models at the eddy covariance sites and catchment scales. This manuscript was a pleasure to read. I have a few minor comments/questions:

Dear editor/co-editor and reviewers

Thank you for considering our manuscript, the overall positive critique, the large effort in finding reviewers and the reviews itself which address important points even in the light of the suggested “minor review”. Below you find a comment by comment reply to the points raised by the two reviewers.

Best regards,
Morten A. D. Larsen

1. I wondered how much of the ag land is tile drained and how this might influence the model performance compared to the non ag lands?

   Being a part of the DK-model nationwide MIKE SHE setup a global (100% coverage) drainage is used since the dense network of smaller streams and ditches cannot be resolved by the 500 m resolution. The differences between land surface types should therefore be minimal. The drainage is described in:


   At the scale in question the setup is capable of reproducing observed physical variables such as discharge, groundwater heads and soil moisture due to rigorous calibration and refinement during the last 12-15 years. We do not have information on the exact share of drained land over the agricultural areas within the catchment, but refer to technical papers on the performance of specific output variables:


2. Line 193. you mention simulated irrigation, but it is not clear how much water this adds to the system or how you determined the simulated amount. Further, how much uncertainty does this add to your analysis?

   The irrigation is calculated in the MIKE SHE model code on the basis of knowledge on well locations, filter depths, known annual abstraction, a demand area and agricultural land around each well by a demand function and a simulated water deficit in the root zone as described in:

To be able to add irrigation to the precipitation file, these simulations have however been performed prior to the simulations were results are extracted for this manuscript. We agree that this is currently not clear in the manuscript and have revised accordingly. Good suggestion.

3. You do a nice job assessing the impact of lack of energy balance closure and showing the sensitivity to such assumptions. I recommend you include in your discussion some of Alan Barr’s work.

Energy balance closure at the BERMS flux towers in relation to the water balance of the White Gull Creek watershed 1999-2009
By: Barr, A. G.; van der Kamp, G.; Black, T. A.; et al.
AGRICULTURAL AND FOREST METEOROLOGY Volume: 153 Special Issue: SI Pages: 3-13 Published: FEB 15 2012

Thank you very much. We have been aware that some would might think the paper is on the heavy side with regards to the number of references which is partly caused by the literature study on parameter values. Other than the included studies on parameter values, the number of studies on energy balance closure is of course vast and the number of references to include among these is a subjective matter. We are glad to be made aware of this study and definitely agree that it should be included due the catchment scale, energy balance considerations, the long evaluation period and streamflow evaluation. This is done.

It would be helpful to me if you could return the reviews annotated, in a different colour or font, to indicate how you have responded to each of the comments.

I am not sure if you mean in the response letter or in a manuscript version but I do both to make sure: In the present document I answer each comment separately and point out where I have applied the corrections and I also attach a manuscript version with corrections highlighted.
Reviewers' comments:

Reviewer #1: Summary of Paper:
The objective of this work is to develop a method for parameter calibration of a spatially distributed hydrologic model using both energy fluxes and catchment runoff. The study used the MIKE SHE hydrologic model and the SWET land surface model. The authors conducted a parameter sensitivity test (Monte-Carlo) and presented four stepwise calibration schemes.

Main impression:
The approach used in this paper was novel, complex, and adds great understanding to the parameterization of distributed hydrologic models. Their results do show an overall improvement in the simulated energy balance post-calibration. I would suggest very minor edits/additions prior to publication.

As also stated above – we would like to thank for the effort in reviewing and the positive judgement of the paper.

Specific Comments and Suggestions:
1. L22: "...the three land surface types and using the flux measurements for calibration." I am not sure what is suggested here? The above text is exactly equal to the manuscript and can therefore not be seen as a suggestion for a change. After reading the sentence again I can see that the “and” could be removed which could have been the intent of the suggestion. This is done.

2. L29: Are you referring to the calibration approach of the distributed systems? Yes – this is specified.

3. L46: Explain acronyms before use.
We agree fully but these are model names which was not clear. This is therefore clarified.

That sentence was indeed uneasy. It was corrected in relation to including a reference as suggested by the editor:

5. L130: Referring the outlier and replacement data in Line 130 - Can the authors expand on how much data was considered an outlier and removed in the results section? Or was any data removed from the calibration/validation periods? This is important since both periods are only 1 year.
We agree that this could potentially infer with the quality of the calibration results and could be addressed in quantitative terms in addition to using the reference (Ringgard et al. 2012). For the calibration period the LE outliers made up 0.16%, 0.74%, 0.13% of the data material for Agriculture, forest and meadow respectively and for H the corresponding numbers were 0.05%, 0.8% and 0.15%. In essence a relative small portion of the data. We have added the outlier-replaced percentage between LE and H to the text (weighted average between stations).
6. **Section 4.2** - By this section I already forgot about what HYP1, 2, 3, 4 represented as far as the calibration scheme. Adding a short sentence reminding the reader about the calibration approaches would be helpful.

Good suggestion. This is added.

7. **Figure 4**. Agriculture - Root shape parameter - I'm a little baffled as to how the input scenarios affect the parameter sensitivity so much as to designate the parameter highly sensitive for HYP1 and HYP4 (100% sensitivity), and moderately sensitive (<40%) for the other two scenarios. This is the same site, the inputs should not be so different.

We agree that from an immediate understanding the difference should not be in this range. More conditions could however attribute to this:

- The sensitivities are relative and therefore the change in relative root shape (A_root) sensitivity can be caused by changes in sensitivities for other parameters – in this case Rst_min.
- The processes are highly non-linear causing this shift when using either corrected (HYP1 and HYP4) or uncorrected fluxes (HYP2 and HYP3).
- The root shape parameter controls the distribution of root water uptake from either the upper or lower part of the root zone. Thus, it appears that this parameter, as opposed to root depth, is the most effective when adjusting for a simulated reproduction of actual evapotranspiration.

8. **Figure 8**. Add x and y labels

We have only written labels once, since they are equal for all three figures. We have however changed the layout slightly to more clearly reflect this.

For further assistance, please visit our customer support site at [http://help.elsevier.com/app/answers/list/p/7923](http://help.elsevier.com/app/answers/list/p/7923). Here you can search for solutions on a range of topics. You will also find our 24/7 support contact details should you need any further assistance from one of our customer support representatives.
Highlights

- A new hydrology model calibration approach solving the energy equation is proposed.
- Initial 1D calibration is performed on three representative flux tower sites.
- LE+H+G were 20% below Rn requiring two hypotheses on the erroneous component.
- The approach was skilful – validated against runs with fewer calibration steps.
- The hypothesis ascribing Rn error was superior over the LE+H+G error hypothesis.
Calibration of a distributed hydrology and land surface model using energy flux measurements

Morten A. D. Larsen*, Jens C. Refsgaard, Karsten H. Jensen, Michael B. Butts, Simon Stisen and Mikkel Mollerup.

1 - Technical University of Denmark, Department of Management Engineering, Frederiksborgvej 399, 4000 Roskilde, Denmark
2 - Geological Survey of Denmark and Greenland, Øster Voldgade 10, 1350 Copenhagen, Denmark.
3 - University of Copenhagen, Department of Geosciences and Natural Resource Management, Øster Voldgade 10, 1350 Copenhagen, Denmark.
4 - DHI, Agern Alle 5, DK 2970, Hørsholm, Denmark
5 - The Danish Road Directorate, Guldalderen 12, 2640 Hedehusene.

* - Corresponding author: madla@dtu.dk/004525119895

Keywords: hydrology/land surface modelling; calibration; water and energy fluxes; evapotranspiration; energy closure imbalance

Abstract

In this study we develop and test a calibration approach on a spatially distributed groundwater-surface water catchment model (MIKE SHE) coupled to a land surface model component with particular focus on the water and energy fluxes. The model is calibrated against time series of eddy flux measurements from three sites of different land surface type (agriculture, forest and meadow) and river discharge data from the 2500 km² Skjern River catchment in Denmark. The approach includes initial calibrations of three one-dimensional models representing the three land surface types using the flux measurements for calibration. This step provides initial
values for the subsequent modelling and calibration at catchment scale. To test the validity of the approach, two additional catchment scale distributed simulations were performed with no calibration and only calibration of the one-dimensional models, respectively. In addition, a subsequent validation period was simulated. A mean energy closure imbalance of 20% was seen for the three sites. For the distributed simulations, the energy imbalance was accounted for by two energy balance closure hypotheses ascribing the error to either energy fluxes or net radiation. In general, the distributed calibration approach improved model results substantially compared to using default values (no calibration) or calibration of the one-dimensional models only. For the distributed model simulations, the assumption regarding the energy balance closure had a substantial impact on the parameter sensitivities and on the simulated discharge and energy balance. During calibration, the simulation with corrected energy fluxes showed better performance on discharge than the simulation with corrected net radiation whereas the reverse was true for the validation period. Regarding energy fluxes, the simulation with corrected net radiation was superior in both the calibration and validation period.

2 Introduction

Water and energy fluxes between land surface and atmosphere are important components of atmospheric and hydrological processes. These fluxes can be quantified by the use of land surface models (LSM) or soil-vegetation-atmosphere-transfer models (SVAT). The calculation of e.g. evapotranspiration in SVATs and LSMS is based on solving the energy and radiation equations often on a sub-daily basis and they therefore differ from the less physically stringent schemes often used in many traditional hydrological models which are based on potential evapotranspiration. LSMS, originating from atmospheric sciences, include spatially distributed, often large scale descriptions of land surface processes. Examples include the Noah model (Rosero et al. 2010) and the CLM model.
LSMs are typically coupled with or forced by atmospheric models and have recently been included in fully coupled climate-hydrology models (Maxwell et al. 2011, Shrestra et al. 2014). SVATs, originating from soil and hydrological sciences, are one-dimensional descriptions typically used for small-scale descriptions linked with soil water flow models (Mauser and Schädlich 1998, Ridler et al. 2012). When included in spatially distributed hydrological models they possess the potential for providing improved evapotranspiration descriptions and enable hydrological catchment models to better utilise remote sensing data to force and constrain hydrological models (Stisen et al. 2011a). SVATs have also recently been included in fully coupled climate-hydrology models (Butts et al. 2014, Larsen et al. 2014). LSMs and SVATs linked to spatially distributed hydrological models are basically similar, and hence we shall in the following refer to both of them as SVAT models.

Assessment of parameter values is critical for the use of SVAT models and an essential challenge is related the vast number of parameters often seen in this type of models. Franks et al. (1997, 1999), Beven and Franks (1999) and Gupta et al. (1999) all highlight the high uncertainty in the predictive capabilities of multi-parameter SVATs due to equifinality. Yet; Franks et al. (1997, 1999) still show good results in terms of reproducing point site flux measurements from the FIFE area in Kansas, USA, and in the Amazon area in Brazil. The added value of a multi-criteria approach as opposed to a single criterion method is confirmed by Gupta et al. (1999) and Demarty et al. (2004). Pollacco et al. (2013) apply an objective function weighing algorithm based on the uncertainties related to remote sensing based surface soil moisture and evapotranspiration calibration variables. Currently no explicit guidelines have been developed on calibrating complex and distributed SVAT and hydrology models.

Parameter estimation for hydrological models is traditionally performed by use of calibration where parameter values are modified to obtain best possible fit between model
simulations and observed target data. While calibration was previously often performed manually by a trial-and-error approach, parameter optimisation by inverse modelling is now the method of choice (Gupta et al. 1998, Madsen 2003, Moore and Doherty 2005). An example is the study by Sun et al. (2013) where inverse calibration based on Monte Carlo-Bayesian techniques was used for calibrating a model both against energy fluxes at point scale and runoff at catchment scale (4.9 km²). In Ingwersen et al. (2011) inverse calibration was used to simulate the water and energy budget for a winter wheat stand at plot scale. Similarly Ridler et al. (2012) utilized inverse techniques for calibrating the combined MIKE SHE/SWET model to simulate energy fluxes at point scale in Mali.

A particular problem related to calibration of SVAT models is that observations of water and energy fluxes are usually not available from operational monitoring networks but only from a few research stations and often for short periods (Wilson et al. 2002, Franssen et al. 2010, Leuning et al. 2012). In addition, energy flux data are known to often have problems with energy balance closure which severely hampers parameter optimisation by inverse modelling, as a SVAT model per definition assumes a closed energy balance (Twine et al. 2000, Choi et al. 2009) and a failure to meet this demand can result in significant biases in long term climate model simulations (Grimmond et al. 2010). Also, the lack of measured energy balance closure will yield an erroneous parameterization when the fluxes are used for the calibration of a hydrological model. Therefore certain assumptions need to be made to account for the lack of closure. To accommodate this, Beven (2006) suggested creating artificial hypotheses to provide closure.

Catchment water balances are linked to energy balances, because the latent energy/evapotranspiration appears as a key element in both balances. The observed catchment runoff and the catchment water balance assessed by hydrological models hence include important information also on the energy balance. On a catchment scale (603 km²) Barr et al. (2012) used
distributed flux tower measurements of evapotranspiration against measured precipitation and discharge for a residual analysis on water balance closure concluding a 15% lack of energy flux closure compared to the measured net energy. For catchment scale (205 km²) model calibration, Li et al. (2011) used the CLM4 model, modified to include a runoff scheme, to evaluate both runoff and energy fluxes. Similarly, operating on a regional to continental scale Maurer et al. (2002) simulated energy flux components while the model was manually calibrated only against runoff.

The objectives of the present study is to develop and test a methodology for calibrating and assessing parameters of a SVAT model linked to a spatially distributed hydrological model by using observations of both energy fluxes and catchment runoff. A comprehensive literature study was carried out to obtain feasible initial values and range of variation for parameters for the considered land surface types. The impact of energy imbalance is of particular emphasis and we analyse to which extent inclusion of discharge observations in the calibration process will improve the model performance and robustness.

3 Methodology

3.1 Study area and data

The Skjern catchment (2500 km²) is located in the western part of the Jutland peninsula, Figure 1. The catchment is dominated by sandy soils generated by glacial outwash plains from the last glacial period Weichsel and intersected by older till deposits from the previous glacial period Saalian (Greve et al. 2007). The topography reaches 130 m above sea level in the eastern part of the catchment and the Skjern River flow into Ringkøbing fjord at sea level to the west. The yearly average precipitation for the catchment is 940 mm for the period 2000-2009 based on direct measurements. When corrected for undercatch using standard monthly correction factors (Allerup et al. 1998) the average precipitation amounts to 1130 mm. In the same period the mean annual
temperature is 9.3 °C and the mean monthly temperatures range between 2.1 and 17.3 °C. Inside the catchment flux towers are placed at the three predominant surface types; agriculture (61%), meadow/grass (24%) and forest (13%), Figure 1. At all sites measurements of short-, long-wave and net radiation components; latent (LE), sensible (H) and soil heat fluxes (G); soil water content; precipitation; air temperature; wind speed; and water table levels have been carried out since late 2008. Measurements of radiation and energy fluxes are based on standard methods. Radiation components are measured using a NR01 Hukseflux radiometer (www.hukseflux.com), LI-COR eddy covariance equipment is used for measuring LE fluxes, Gill sonic anemometers for measuring H fluxes, and Hukseflux plates for measuring G fluxes.

The energy flux data used in the study have undergone quality control as part of the processing (Ringgard, 2012) (Step 1.3, Figure 2). Inaccurate observations caused by e.g. low turbulence condition were replaced by data representing similar conditions. Replacement of data was thus for periods with low energy fluxes and therefore this source of uncertainty is expected to be of minor significance. Individual data points clearly outside the expected range at the time of day and season were considered as outliers and removed (equal to 0.2% on average between LE, H and the three stations weighted relative the areal share). For two periods July 21-August 16 and August 24-October 28, 2009, no flux measurements were available from the agricultural site and data were replaced from the forest site. As these periods are mostly placed in the spin-up period (see below) the calibration results are not expected to be significantly affected.

3.2 Modelling system and setup

This study uses the spatially distributed MIKE SHE hydrological modelling system capable of including all key hydrological processes such as ET, channel flow, overland flow, unsaturated flow, saturated flow as well as irrigation and drainage (Graham and Butts 2005). The land surface
model SWET component (Overgaard, 2005) is used in the analysis. SWET is based on the Shuttleworth-Wallace model (Shuttleworth and Wallace 1985). It considers vegetation and energy balance processes in a two-layer system and is extended to include energy fluxes from ponded water and interception storage. Models are constructed for both the three local measurement sites (1D representation) and for the entire catchment (Figure 1).

The 1D models simulate energy fluxes and vertical unsaturated flow based on Richards’ equation. The SWET model is driven by 30 min climatic observations of precipitation, air temperature, wind speed, net radiation, surface air pressure and relative humidity. Measured water table elevations are specified as lower boundary condition. As opposed to the fully distributed model application overland flow, river flow and groundwater flow are not considered. Relevant to both the 1D and distributed analyses, the SWET land-surface module does not consider snow accumulation and melting. Initial values for root depth and vegetation height applicable for the 1D simulations of the agricultural and meadow sites were based on estimates from model simulations for relevant soil type and management conditions using the vegetation model Daisy (Styczen et al. 2004a). Since the footprint of the flux tower at the agricultural site was influenced by several crop types, the vegetation height and root depth were based on average simulations for winter- and spring cereals. The observational data for leaf area index (LAI), another important crop parameter, was inconsistent, and the seasonal variation of this parameter was therefore derived by combining the observed seasonal trends and the simulation results by the Daisy model. The initial values for soil parameters for each site were derived from the HYPRES pedotransfer function (Wösten et al. 1999). Specifically, the Van Genuchten parameters $\alpha$ and $n$ (van Genuchten, 1980) as well as saturated hydraulic conductivity, residual- and saturated water content were estimated from soil texture, organic matter and bulk density. The soil texture data were retrieved from grid values from a distributed soil map of 250 m resolution in the A horizon (0-30 cm) and 500 m resolution in B and
C horizons (30-80 cm and below 80 cm) (Styczen et al. 2004b, Greve et al. 2007, Iversen et al. 2011). For the sensitivity and optimization analyses the soil parameter bounds were defined by calculating the 90% confidence intervals of variation ranges from Meyer et al. (1997) for soils of similar texture as for the three field sites. These soil types include sand, loamy sand and sandy loam for the agriculture and forest site roughly corresponding to JB1 soil (coarse sand) in the Danish soil classification system (Greve et al. 2007, www.djfgeodata.dk). Due to the lack of agreement on soil type at the meadow site between site specific soil retention data and available soil maps the 90% confidence intervals for the meadow site is based on a wider range of soil types from loamy sand to clay loam (Meyer et al. 1997). The discrepancy of the data is most likely due to highly heterogeneous soils caused by shifting alluvial deposits from the nearby Skjern River. In the parameterization the seasonal patterns of vegetation characteristics as induced by both climate and management practice were kept constant whereas the respective amplitudes were calibrated by a single factor for each parameter. The annual sequence of these ratios was found by balancing model simulations by Daisy, relevant literature values (Table 1) and occasional on-site measurements of LAI and vegetation height. Similarly, the relation between the parameter values in the three vertical soil type horizons were kept constant implying that the individual parameters for the A, B and C horizon were shifted by the same factor for each grid cell. By introducing factors between parameters in space and time the number of parameters to be optimized is reduced.

The MIKE SHE model for catchment scale is based on the Danish national water resources model (DK-model) (Stisen et al. 2012, Højberg et al. 2013). As opposed to the DK-model the present model includes the SWET land-surface model and the parameterization of this model is partly based on the work by Stisen et al. (2011a). The DK model has a 500 m resolution and includes a detailed river network. A maximum time step of 1 hour is used for overland, unsaturated and saturated flow. The MIKE 11 open channel river model component utilizes 3616 cross sections
in 100 river branches and uses a 30 min time step. For precipitation exceeding 2 mm per time step
an automatic reduction in time step takes place. The input for the catchment scale SWET model is
hourly values of the same climatic variables as for the 1D models. These values are obtained by thin
plate spline interpolation of climate station data (Stisen et al. 2011a) to a resulting grid in 2 km
resolution for all variables except precipitation used in 500 m resolution. Precipitation input is
based on interpolation of daily rain gauge data using kriging and dynamically corrected for
undercatch. Hereafter, simulated irrigation is added to the precipitation input file. The simulated
irrigation is calculated from the root zone deficit and a demand function based on well locations,
filter depths, annual abstractions and demand area information (Stisen et al. 2011b). The
agricultural crop distribution was developed from statistical data, where the various crops were
grouped in four classes with different growth characteristics: spring sown cereals, winter sown
cereals, grass/clover and maize. The crop classes were distributed without georeferencing due to
lack of data. Forest vegetation parameters were based on relevant literature and observations. A
total of 36 parameters were analysed in the sensitivity analysis. As the calibration is targeted
towards the energy fluxes no sensitivity and calibration analyses are performed on the saturated
zone parameterization which is already well calibrated as a part of the DK-model (Stisen et al.

To generate proper initial conditions for soil moisture in the 1D simulations a spin-up period
of 5-10 months prior to the calibration period was used. The spin-up periods were decided by data
availability as measurements started in April 2009 for the agriculture and meadow sites and
December 2008 for the forest site. The distributed model involves a considerable increase in
computation time compared to the 1D setups and a spin-up period of 3 months was therefore used.
Groundwater quasi steady-state was reached by looping simulations in 3-year runs each time using
the resulting groundwater heads as initial conditions for the subsequent simulation. Three loops were required to reach quasi steady-state.

3.3 Calibration approach

We propose a calibration and validation approach involving three steps as illustrated in Figure 2: (1) calibration of 1D models representing the flux measurement sites, (2) use of parameter values from step 1 as initial values in the calibration of the distributed catchment model, and (3) validation of the distributed parameterization on independent data. Additionally we test the calibration results by comparing to model results based on 1D calibration only and with no calibration respectively.

Both the 1D and the distributed simulations were calibrated using inverse modelling adopting initial and range of parameter values from literature, observations, databases and simulation results from the model code Daisy (Hansen et al. 1990, 1991) as described above and as listed in Table 1 and 2. The initial input parameter values as well as their ranges were specified according to relevance: If available, either observations or Daisy model simulations based on site specific conditions were used. When using literature values conditions similar to the sites in terms of vegetation species, climate and soil type were used. The calibration period covers a one year period from October 1, 2009 to September 30, 2010 while the validation period is from May 1, 2011 to April 30, 2012.

The sensitivity and auto-calibration analyses were performed using the AUTOCAL software included in the MIKE SHE package (Madsen 2003). At the point scale three components were included in the objective function: latent heat flux LE, sensible heat flux H, and averaged soil water content based on measurements at three depths (2.5, 22.5 and 52.5 cm below the surface). The objective function was built from the root mean square errors (RMSE) between observations and
 simulations. The RMSEs for the three individual components were normalized, weighted equally and summed to arrive at an aggregated objective function. The normalization was carried out using the common distance scale which has proven to be robust in accounting for differences in magnitude of each component of the objective function (Madsen 2003, Butts et al. 2004, Mertens et al. 2004).

At catchment scale the optimization was performed against two overall components of equal weight: (1) Discharge observations at three discharge stations aggregated with equal weight (Figure 1) and (2) latent and sensible energy fluxes at the three sites weighted according to the relative areal share of each land surface type (agriculture 62%, forest 13% and grass/meadow 25%).

The sensitivity analysis (Step 1.2 and 2.2 in Figure 2) for both point and catchment scales were based on the AUTOCAL local sensitivity analysis procedure accounting for parameter sensitivities at the location in parameter space defined by the initial parameter values. Since the sensitivity analysis was local and therefore unable to account for numerous combinations in the parameter space, the sensitivity analyses were repeated using randomly sampled parameter sets within the parameter bounds. The sensitivity analysis was based on a backward difference approximation method around the parameter value using a 2% perturbation fraction. Also, covariance matrices were calculated to test for parameter correlation.

For both scales the ten most sensitive parameters were identified and subsequently used in the auto-calibration process similar to Blasone et al. (2007). Only parameters with relative sensitivity coefficients above 1%, in relation to the most sensitive parameter, were included in the optimization (Hill 1998). For the low sensitivity parameters that were not included in the auto-calibration of the catchment model their values were derived from the point scale calibrations. The remaining distributed parameters were based on relevant literature (Table 2).
The point scale parameter optimization was based on the global Shuffled Complex Evolution method (Duan et al. 1993) which has proven to be robust for comprehensive parameter estimation problems (Butts et al. 2004, Mertens et al. 2005, Blasone et al. 2007). A maximum time step of 6 minutes was used for the 1D models, which for the calibration period corresponded to a computation time of around 2-3 minutes per model run. The model convergence criteria were defined as max. 1% change in objective function after three iterations loops and this criterion was usually met after 200-250 runs. For obtaining a more efficient optimization of the catchment scale model, parallel model runs were carried out using the global Population Simplex Evolution optimization method (DHI 2010). The convergence criterion was typically reached after 250-400 runs. To assess the accuracy of the parameterization and modelling approach as such, an independent validation analysis was performed subsequent to the auto-calibration simulations. The validation was performed for a one-year period and assessed on discharge, LE and H as for the calibration analysis (Step 3.1, Figure 2).

3.4 Handling of Energy Balance Flux problems

Energy balance closure requires that the net radiation (difference between incoming and outgoing long- and short-wave radiation) equals the sum of latent heat (LE), sensible heat (H), and soil heat (G) fluxes. All terms are highly sensitive to the surface characteristics and furthermore they are subject to different diurnal and seasonal variations. The albedo of bare soil is dependent on texture and moisture content while the albedo for vegetation generally decreases with vegetation height and stand complexity. Changes in albedo will affect the amount of available energy, and soil moisture influences how this energy is distributed between LE and H heat fluxes. Independent measurements of the components of the energy balance equation rarely provide closure, which has

The impact of the energy balance closure problem on the calibration results was addressed by creating four hypotheses each based on different assumptions regarding measurement errors (Step 1.4 and 2.4, Figure 2). All four hypotheses were used for the 1D runs whereas only the first two were used in the distributed runs to reduce computation time.

- **HYP1:** The measured energy fluxes (LE, H and G) are scaled with a single factor (increase) such that the Bowen-ratio is maintained and the sum of the energy fluxes equals measured net radiation over the one-year calibration period;

- **HYP2:** The measured energy fluxes are maintained whereas the net radiation is scaled with a factor (reduction) to match the sum of measured energy fluxes over the one-year calibration period;

- **HYP3:** The measured energy fluxes are unaltered whereas the sum of these is used as net radiation on a daily basis;

- **HYP4:** The measured energy fluxes (LE, H and G) are scaled with a single factor (increase) such that the Bowen-ratio is maintained and the sum of the energy fluxes equals measured net radiation on a daily basis;

For the first two scenarios the entire dataset is multiplied by a single scaling factor to provide energy closure for the one year calibration period. Closure on a daily basis may not be met, which is the case for the last two scenarios. Of the investigated scenarios more energy is available in HYP1 and HYP4 based on the measured net radiation (yearly mean values of 55 w/m², 68 w/m² and 55 w/m² for agriculture, forest and meadow respectively) whereas HYP2 and HYP3 are based on reduced values (yearly mean values of 39 w/m², 48 w/m² and 42 w/m² for the same sites).
Only the HYP1 and HYP2 hypotheses were used for the distributed simulations since they both utilize measured net radiation, scaled and unscaled, and hereby the daily temporal pattern is likely to be more representative than if based on the sum of the three fluxes. In contrast to the point scale where net radiation measurements were available, global radiation was used in the catchment scale analysis. To fulfil the same assumptions on energy balance closure as used at point scale the albedo was adjusted to obtain energy balance closure over the entire one year calibration period. This was done based on the flux measurements for the three general land use types agriculture, forest and meadow. As a result HYP1 is based on measured albedos whereas HYP2 is based on estimated albedos that are higher than measured. All three distributed simulation cases shown in Figure 2 were performed for both HYP1 and HYP2 as well as for the calibration and validation periods for a total of 12 simulations.

4 Results

4.1 Energy balance closure

The measurements used in this study show lack of energy balance closure. The sum of LE, H and G over the one-year calibration period accounted for 71%, 71% and 76% of the measured net radiation at the agricultural, forest and meadow/grass sites respectively (Figure 3). The corresponding measured Bowen-ratios in the same period were 0.20, 0.17 and 0.18 for the three land surface types. Radiation and particularly data for energy fluxes are generally sparse in Denmark. Therefore, in Figure 3, measured net and global radiation as well as the sum of energy fluxes are compared to results from (1) the corresponding grid cell in HIRHAM regional climate model simulations (Larsen et al. 2013) using ERA-Interim reanalysis data as boundary conditions and (2) observation data from the Foulum research station (grass surface) located app. 50 km north of the agriculture and forest sites. HIRHAM simulation data are also used to compare LE and H.
fluxes since no other flux data were available. The simulations of net radiation by HIRHAM and the
observed data from Foulum are much closer to measured LE, H and G flux sum at all sites for both
years although perhaps less distinct at the forest site (Figure 3). For global radiation there is a close
match between all data series, especially between the three study sites and Foulum. For LE fluxes
good agreement is seen between HIRHAM simulations and observations for the forest and meadow
sites and less favourable for the agricultural site. For the H fluxes larger discrepancies are seen
particularly for the forest site.

4.2 Simulation results

The results of the sensitivity analyses for both point and catchment scales are shown in
Figure 4. Similar parameter sensitivity results are obtained for the agricultural and forest sites with
nine and ten parameters, respectively, having a relative sensitivity coefficient above 10%. The
parameter sensitivities for hypotheses HYP2 and HYP3 (with unaltered energy fluxes) show similar
patterns with the minimal stomata resistance (Rst_min) as the most sensitive parameter. In contrast,
for hypotheses HYP1 and HYP4 (with corrected energy fluxes) the root shape factor is the most
sensitive parameter. For the meadow site only the three parameters Rst_min, leaf area index (LAI)
and vegetation height have relative sensitivity coefficients above 10% and the influence of the
energy balance hypothesis is limited given the large drop in sensitivity between Rst_min and the
second most sensitive parameter LAI.

At catchment scale saturated hydraulic conductivity (K_sat) for the unsaturated zone is the
most sensitive parameter for both energy balance hypotheses, whereas the sensitivities of the
remaining parameters are smaller and also subject to larger variation between the two hypotheses
(Figure 4). The forest energy fluxes were found to be poorly reproduced by the calibrated forest leaf
width values which were therefore manually calibrated prior to auto-calibration.
The simulated discharge based on the parameter values obtained by the full auto-calibration approach for the two distributed energy balance hypotheses are shown in Figure 5 for both the calibration and validation periods. For both simulations and periods there is a tendency of underestimation of baseflow and mostly pronounced for HYP2. For peak flow the HYP2 hypothesis exhibits more dynamic behaviour with more rapid responses. In summary, the HYP1 simulation shows better discharge performance statistics for the calibration period whereas the opposite pattern is seen for the validation period (Figure 6). Also, the validation confirms a reasonable discharge calibration although the HYP1 results are moderately poorer for two out of three discharge stations.

Figure 7 shows simulated LE and H energy fluxes on an hourly basis for a 10 day summer period in the calibration period. The results from the two energy hypothesis runs are compared to their relevant fluxes: HYP1 to scaled and HYP2 to unscaled fluxes, corresponding to the assumptions of the energy balance closure. For both the calibration and validation periods, there is a tendency for better energy flux statistics for HYP2 compared to HYP1 (Figure 6). Overall the agriculture and grass surfaces are better reproduced compared to forest. Especially observations of H in the forest are rather poorly simulated during high global radiation (Figure 7); however, part of this discrepancy is likely caused by errors in the observation data (Sonnenborg et al. 2013). Simulations for night time conditions compare favourably to observations of both energy fluxes and for both scenarios. As shown in Figure 6 the simulations for the validation period shows comparable or better performance statistics for both fluxes as compared to the calibration period.

In the auto-calibration, simulations are compared to the same observed discharge data for both HYP1 and HYP2 while the energy fluxes and albedos were adjusted to obtain energy balance closure. This is reflected by differences in water balance components for the two hypotheses HYP1 and HYP2 both on catchment and local scales (Table 3). With a lower albedo in the HYP1 scenario higher ET and therefore lower recharge and discharge are simulated at catchment scale. For the
agriculture grid cell the slightly higher ET for the HYP2 hypothesis can be ascribed to the lower
K_sat value. The significant difference in drainage is likely a function of higher groundwater
elevations in the HYP2 hypothesis. For forest and meadow the difference in albedo is reflected in
higher ET for HYP1.

The comparison of the performances of the full calibration and the scenarios with calibration
of the 1D models only and no calibration respectively is shown in figure 6. For discharge the
simulations with no calibration using default values is comparable to the results of the full
calibration whereas the 1D calibration is generally poorer. The most significant improvement in the
calibration approach is however seen for the LE and H energy fluxes where the mean root mean
square error (RMSE) and mean absolute error (MAE) statistics for the three sites is substantially
improved. The distributed simulations based on the values from the 1D calibration transferred to
catchment scale are generally of lower quality compared to the case where no calibration is carried
out. This, as opposed to using initial values using no calibration at all (default values), is because
neither input source includes parameters related to the distributed scale (e.g. related to discharge
and overland flow). Also, the extreme 1D mean RMSE statistics for H is due to the forest site which
was manually calibrated prior to the autocalibration (see above) and it does therefore not affect the
full 1D+2D calibration.

5 Discussion

We have presented an approach for a stepwise calibration of a distributed combined land
surface/hydrology model using observations of energy fluxes and discharge as calibration target. In
general the method proved effective. The validation period confirmed the performance statistics
from the calibration period and the simulations using a reduced calibration approach or no
calibration were generally less accurate.
5.1 Energy and water balance

As expected a key finding was that the choice of energy balance hypothesis has a major impact on the resulting parameterization as well as on the energy and water balance results. The lack of energy balance closure for flux measurements is therefore a great challenge when using such measurements for calibrating distributed hydrological models. Energy balance closure is required when flux measurements are used for inverse modelling and we investigated two hypotheses where the error was attributed to either the measured energy fluxes (HYP1) or the measured net radiation (HYP2).

Either of these two approaches may be applicable. Even though the eddy covariance method is generally regarded as the best practical method for measuring energy fluxes, measurement difficulties are well documented (Foken et al. 2006, Sun et al. 2008, Franssen et al. 2010, Ringgard, 2012, Stoy et al. 2013). A consistent deficit in measured LE fluxes have been suggested as an explanation in the lack of energy balance closure (Sun et al. 2008) while on the other hand Foken et al. (2006) suggested that low frequency turbulence structures cause errors in the flux measurements. Also, measurement errors in H fluxes have been suggested, see (Mauder et al. 2008) and Sonnenborg et al. (2013). The energy flux statistics (Figure 6) were better for the HYP2 hypothesis possibly indicating that larger measurement errors are attributed to the energy fluxes compared to radiation components. An additional contributing factor is the difference in the footprint of the flux measurements reflecting conditions upstream of the wind direction and the support scale of the radiation sensor.

Contrary to indications of flux measurement biases, estimates of ET based on eddy covariance and soil moisture balances of the root zone compared well and surprisingly the highest ET was seen for the eddy covariance method over a seven week period in May-June 2009 (Schelde
et al. 2011). Similarly, net radiation measurements from Foulum research station match the sum of LE, H and G fluxes better than for the agriculture, forest and meadow sites used in this study (Figure 3). Advection has also been suggested as a missing component in the energy balance suggesting that the measured energy fluxes may indeed be more reliable than anticipated (Leuning et al. 2012). Likewise, large scale eddies due to land surface heterogeneity have recently been suggested to make up a part of the energy imbalance (Stoy et al. 2013).

Snow processes are currently not considered the SWET land-surface model and this impacts the calibration results especially with regard to discharge. Even though the calibration period is relatively cold, little precipitation occurs on days with temperatures below $0 \, ^\circ\text{C}$. Nevertheless, the snow melt related discharge peaks in Feb-Mar 2010 are clearly not reproduced (Figure 5).

5.2 Sensitivity, calibration and validation

The distinct differences between parameter sensitivities at the agriculture and forest sites compared to the meadow site are related to the water available for ET (Figure 8). At the agricultural and forest sites, complete soil saturation is rarely or never reached and there is a clear tendency for higher ET with lower soil moisture as this occurs during the summer months with a higher level of available solar energy. Soil parameters affecting available soil water are therefore crucial for determining the partitioning of energy fluxes at these sites. For the meadow site, located next to the river bank, saturated conditions are often reached (41% of the year for HYP2) and soil moisture levels are generally high. As a result the relation between root zone water and ET is less distinct. The high water availability at the meadow site therefore makes vegetation parameters substantially more important compared to soil parameters.

The choice of energy balance closure hypothesis, and therefore the energy available, is highly influential on which parameters are most sensitive especially for the agriculture and forest
sites but also to some extent for the meadow site. This is reflected in the sensitivities for the HYP1 and HYP4 hypotheses on one side and the HYP2 and HYP3 hypotheses on the other. This is consistent with the HYP1/HYP4 hypotheses being based on measured net radiation (averaged to 54, 67 and 56 W/m² in the calibration period for agriculture, forest and meadow) while the HYP2/HYP3 hypotheses are based on downscaled net radiation (40, 47 and 40 W/m² for the same sites).

The results of the sensitivity analysis and auto-calibration procedure naturally reflect the interplay between a vast number of parameters, assumptions regarding the energy balance terms, assumed relations between parameters for the individual soil layers, and the specific conditions that apply for the period subject to the analysis. These assumptions are required for reasons such as energy balance closure, calibration simplicity, data availability and focus of the study in question. For example, the parameter Rst_min is assumed constant in the present study whereas Ingwersen et al. (2011) obtained better simulations of LE and H based on the Noah LSM model using monthly values of Rst_min. Hou et al. (2012) related simulations of LE and H to parameter uncertainty in a study using the Community Land Model (CLM4) and documented substantial variations in the simulations of both fluxes (over 100 W/m²) in the summer months depending on the parameterization. The rather poor agreement between simulations and observations of H fluxes was likely due to observation errors. Separate auto-calibration runs were performed assessing forest heat fluxes based on forest parameterization alone with limited improvement and this is in line with the results by Sonnenborg et al. (2013).

5.3 Land use

Standard seasonal crop characteristics were used in this study and the actual crop rotation and management conditions at a given site were therefore not considered. This could lead to
discrepancies between the vegetation parameterization at the agricultural site particular around the
time of harvest and at the grass site due to irregular and undocumented cattle grazing. Possibly
more significant is the effect of the spatial distribution of crops derived from available county
average data, which may not reflect the actual distribution. This is the case both in terms of the
overall distribution, type and location of crops within the catchment. But it also affects the
evaluation of energy flux data from the agricultural site as the actual footprint may differ from the
simulated crop as also reported by Göckede et al. (2008).

The distributed MIKE SHE parameterization was performed by transferring the calibrated
parameters from the point scale setups, representing agriculture, meadow and forest, to the relevant
land surfaces within the catchment for subsequent sensitivity and auto-calibration analyses. For
agriculture, this procedure includes the tradeoff between using the calibrated parameter values
reflecting site specific conditions at the measurement site and using crop and soil specific
parameters from the Daisy model. However, the former was used since parameters were calibrated
under conditions of energy balance closure but also because of the statistical spatial distribution of
the Daisy modelled crop characteristics.

6 – Conclusions

In the present study we present a novel approach to calibrating a combined and distributed
hydrology/land-surface model. The approach first involves calibration of the land-surface model
component against measurements of energy fluxes from flux towers. Secondly, the distributed
hydrological model component, including unsaturated and saturated zone components, is calibrated
using the point scale parameterization as initial conditions and using stream flow and energy flux
measurements as the calibration targets. The method employs a meticulous literature inspection for
realistic initial and range of parameter values. Previous calibration efforts of SVAT and LSM
models with a hydrological focus have often used satellite data having certain measurement deficiencies, relied on previously point scale based parameterizations or used traditional hydrological variables such as discharge and hydraulic head. Modelling the highly interrelated and dynamical processes and variables in the entire water and energy budget from groundwater to lower atmosphere is challenging. Our results show that the proposed approach significantly improves the simulated energy balance components as compared to using default values or a simpler 1D calibration. The simulations were able to reproduce features across water and energy budgets taking into considerations measurement errors and lack of energy balance closure.

The calibration of the one dimensional models provided good insight to data quality and parameter sensitivity and served as a useful initial parameterization of the catchment scale simulations. Further, the sensitivities of parameters in the point scale analysis were highly affected by the surface type and more specifically the water availability as well as the choice of energy balance hypothesis. Similarly, the catchment model simulations showed that parameter sensitivities are largely determined by the choice of energy balance hypothesis, yet saturated hydraulic conductivity was the most sensitive parameter for both energy balance hypotheses. Despite likely errors in measured forest sensible heat fluxes, our results suggest that the measured energy fluxes are reliable and that the energy imbalance is likely caused by advection and large scale eddies rather than erroneous net radiation measurements.

7 - Acknowledgments
The present study was funded by a grant from the Danish Strategic Research Council for the project HYdrological Modelling for Assessing Climate Change Impacts at differeNT Scales (HYACINTS– www.hyacints.dk) under contract no: DSF-EnMi 2104-07-0008. The study is made possible
through a collaboration with the HOBE project (www.hobecenter.dk) funded by the Villum Kann Rasmussen Foundation. The data used in the study can be accessed through the first author.

8 – References


Figure 1. Skjern catchment (Denmark). Locations of the flux tower and discharge sites and land use map (land use map from Jensen and Illangasekare (2011)).

Figure 2. Flow chart of the calibration process.

Figure 3. Measured radiation and energy flux components at the three Skjern catchment sites for the years 2009 and 2010, simulated components by the HIRHAM regional climate model, and measurements for year 2009 from an agricultural research station in Foulum located approx. 50 km north of the agriculture and forest sites. Yearly average values for each radiation and energy flux component is given in the legend. Measurements from the agricultural station in 2009 are bracketed due to extensive gap filling in this period.

Figure 4. Results of the sensitivity analyses for the individual sites and the distributed setup. For the distributed setup only the 20 most sensitive parameters are shown. The y-axis represents relative sensitivity in relation to the most sensitive parameter.
Figure 5. Observed and simulated daily and accumulated discharge at three gauging stations based on two energy balance hypothesis setups for the calibration and validation periods.

Figure 6. Catchment scale simulation statistics. Root mean square error (RMSE) and mean absolute error (MAE) statistics for discharge, latent heat flux (LE) and sensible heat flux (H) for the two energy balance hypotheses and for both the calibration and validation period (RMSE only).

Figure 7. Example of model simulated LE and H energy fluxes for the distributed case and for the two energy balance hypotheses in the calibration period. Un-scaled and scaled observations are also shown. The data represent a summer period June 1-11, 2010 with cloudless conditions from June 1-5 and varying global radiation from June 6-11. Lines of the same colour are comparable.

Figure 8. Daily values for plant transpiration and soil evaporation as a function of the average root zone water content at the three sites in the calibration period for HYP2 and the corresponding soil field capacity (pF 2) and saturation levels (dotted lines).
Table 1. Point scale parameterization.


* Seasonal variation included. If brackets only some of the sites include seasonality (A: agriculture, M: meadow). With seasonality, the values are from mid-summer.

Parameter only included only in the sensitivity analysis.

Initial parameter value based on ° observations, °° simulated values by Daisy model and °°° Van Genuchten parameter derived from pedotransfer function. Parameter values are considered for the A, B and C horizons and the parameter ranges are derived from Meyer et al (1997).
<table>
<thead>
<tr>
<th>Parameter</th>
<th>Agriculture</th>
<th>Forest</th>
<th>Meadow</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Initial value</td>
<td>Range</td>
<td>Calibrated values (hypothesis range)</td>
</tr>
<tr>
<td>Leaf area index* (m²/m³)</td>
<td>4.6 (SIM, 9.29, 37.27, 515)</td>
<td>2.56-6.64</td>
<td>2.69-3.35 (O.19, 1.5, 29, 37.20, 3.7, 8, 16, 21.2, 23.0)</td>
</tr>
<tr>
<td>Root depth*(A, M) (m)</td>
<td>0.78 (SIM, 29)</td>
<td>0.2-0.8</td>
<td>-</td>
</tr>
<tr>
<td>Minimum stomatal resistance (Rs)</td>
<td>90 (9.28, 29.37, 23.27, 54)</td>
<td>50-200</td>
<td>100-194</td>
</tr>
<tr>
<td>Extinction coefficient</td>
<td>0.33 (23.17, 18.15, 12)</td>
<td>0.4-0.8</td>
<td>0.47-0.66</td>
</tr>
<tr>
<td>Leaf width (mm)</td>
<td>0.01 (9.29)</td>
<td>0.005-0.05</td>
<td>0.015-0.26</td>
</tr>
<tr>
<td>Vegetation height*(A, M) (m)</td>
<td>0.78 (SIM)</td>
<td>0.51-1.03</td>
<td>0.55-0.98</td>
</tr>
<tr>
<td>Roughness length*(A, M) (m)</td>
<td>0.03 (10, 29.31)</td>
<td>0.007-0.012</td>
<td>-</td>
</tr>
<tr>
<td>Intercept storm coefficient (mm)</td>
<td>0.05 (29.31, 32.33, 34)</td>
<td>0.01-0.2</td>
<td>0.02-0.13</td>
</tr>
<tr>
<td>Root shape factor</td>
<td>1 (31, 32, 33)</td>
<td>0.2-0.4</td>
<td>0.24-3.66</td>
</tr>
<tr>
<td>Detention storage (mm)</td>
<td>10 (34, 38)</td>
<td>5-100</td>
<td>-</td>
</tr>
<tr>
<td>Initial water content (m³/m³)</td>
<td>0.2</td>
<td>0.1-0.4</td>
<td>-</td>
</tr>
<tr>
<td>Van Genuchten e (cm³/m³)</td>
<td>0.0620, 0.0820, 0.085</td>
<td>0.008</td>
<td>0.11-0.12</td>
</tr>
<tr>
<td>Van Genuchten n</td>
<td>1.445/1.58/1.72</td>
<td>1.25-1.88</td>
<td>-</td>
</tr>
<tr>
<td>Saturated hydraulic conductivity (m s⁻¹)</td>
<td>3.9E-7</td>
<td>7.</td>
<td>3.9E-7</td>
</tr>
<tr>
<td>Residual water</td>
<td>1.0E-3</td>
<td>0.028</td>
<td>-</td>
</tr>
</tbody>
</table>

Comment [madla1]: We noticed that, by mistake, leaf width results had been copied to here.
<table>
<thead>
<tr>
<th>Parameter</th>
<th>Initial Value</th>
<th>Parameter Range</th>
<th>Parameter Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>Saturated water content (m³/m³)</td>
<td>0.40/0.38/0.35</td>
<td>0.25-0.54</td>
<td>0.25-0.54</td>
</tr>
<tr>
<td></td>
<td>(VG)</td>
<td></td>
<td>(VG)</td>
</tr>
<tr>
<td>Gr. water table* (m) (m below surface)</td>
<td>5 (O)</td>
<td>2-10</td>
<td>5-15</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.25 (O)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.5 (O)</td>
</tr>
</tbody>
</table>

* Seasonal variation included. If brackets only some of the sites include seasonality (A; agriculture, M; meadow). With seasonality, the values are from mid-summer.

Table 2. Catchment scale parameterization.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Distribution type</th>
<th>Range</th>
<th>Hyp1 and Hyp2</th>
<th>Hyp1 Initial value</th>
<th>Hyp1 Calibrated value</th>
<th>Hyp2 Initial value</th>
<th>Hyp2 Calibrated value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Leaf area index (m²/m²)</td>
<td>Agriculture</td>
<td>2.56-6.64</td>
<td>2.97</td>
<td>5.66</td>
<td>2.69</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Root depth (A,M) (m)</td>
<td>0.2-0.8</td>
<td>0.78</td>
<td>0.78</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>minimum stomatal resistance (Rst_min) (s/m)</td>
<td>50-200</td>
<td>136</td>
<td>194</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Extinction coefficient</td>
<td>0.4-0.85</td>
<td>0.66</td>
<td>0.47</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Leaf width (m)</td>
<td>0.005-0.3</td>
<td>0.05</td>
<td>0.25</td>
<td>0.23</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vegetation height (A,M) (m)</td>
<td>0.51-1.03</td>
<td>0.8</td>
<td>0.98</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Roughness length (A,M) (m)</td>
<td>0.05-0.13</td>
<td>0.08 (1)</td>
<td>0.099 (1)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Leaf area index (m²/m²)</td>
<td>Forest</td>
<td>3.6-10.2</td>
<td>3.84</td>
<td>7.41</td>
<td>3.15</td>
<td>3.12</td>
<td></td>
</tr>
<tr>
<td>Root depth (A,M) (m)</td>
<td>1.2</td>
<td>1.95</td>
<td>1.34</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>minimum stomatal resistance (Rst_min) (s/m)</td>
<td>65-258</td>
<td>188</td>
<td>193</td>
<td>237</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Extinction coefficient</td>
<td>0.15-0.52</td>
<td>0.46</td>
<td>0.46</td>
<td>0.22</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Leaf width (m)</td>
<td>0.0005-2</td>
<td>0.04</td>
<td>1</td>
<td>0.85</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vegetation height (A,M) (m)</td>
<td>10-22.5</td>
<td>19.8</td>
<td>11.1</td>
<td>10.6</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Roughness length (A,M) (m)</td>
<td>1.2 25 (1)</td>
<td>0.2</td>
<td>0.11</td>
<td>0.04</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Leaf area index (m²/m²)</td>
<td>Meadow</td>
<td>2.56-6.66</td>
<td>2.74</td>
<td>3.95</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Root depth (A,M) (m)</td>
<td>0.4-0.8</td>
<td>0.6</td>
<td>0.6</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>minimum stomatal resistance (Rst_min) (s/m)</td>
<td>55-250</td>
<td>247</td>
<td>238</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Extinction coefficient</td>
<td>0.25-0.75</td>
<td>0.26</td>
<td>0.45</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Leaf width (m)</td>
<td>0.005-0.3</td>
<td>0.02</td>
<td>0.01</td>
<td>0.0085</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vegetation height (A,M) (m)</td>
<td>0.11-0.49</td>
<td>0.27</td>
<td>0.15</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Roughness length (A,M) (m)</td>
<td>0.011-0.049 (1)</td>
<td>0.007</td>
<td>0.004</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Interception coefficient (mm)</td>
<td>Global</td>
<td>0.01-0.2</td>
<td>0.13</td>
<td>0.07</td>
<td>0.148</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Root shape factor</td>
<td>0.2-4</td>
<td>0.89</td>
<td>0.46</td>
<td>3.66</td>
<td>0.39</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Detention storage</td>
<td>5-100</td>
<td>40</td>
<td>40</td>
<td>95</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Night cloudiness</td>
<td>0.01-0.99</td>
<td>0.5</td>
<td>0.0136</td>
<td>0.5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Drainage time constant (s)</td>
<td>6.50E-08-5.5E-7 (4)</td>
<td>1.02E-07</td>
<td>1.02E-07</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Leakage coefficient (m/s)</td>
<td>1.38E-7-1.38E-5 (2.3)</td>
<td>1.38E-06</td>
<td>1.38E-06</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Overland Manning number (m²/s)</td>
<td>2.15 (2.3)</td>
<td>7.5</td>
<td>7.5</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>River Manning number (m²/s)</td>
<td>14.75-31.25 (3)</td>
<td>25</td>
<td>25</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Soil heat capacity reduction</td>
<td>0.01-0.99</td>
<td>0.5</td>
<td>0.5</td>
<td>0.13</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Soil heat conductivity reduction</td>
<td>0.01-0.99</td>
<td>0.5</td>
<td>0.5</td>
<td>0.039</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Van Genuchten α (cm⁻¹)</td>
<td>By soil type (values for top layer)</td>
<td>0.009-0.128</td>
<td>0.0129</td>
<td>0.093</td>
<td>0.126</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Van Genuchten n</td>
<td>1.2459-1.8842</td>
<td>1.27</td>
<td>1.3</td>
<td>1.51</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Saturated hydraulic conductivity (K_sat) (m/s)</td>
<td>3.9E-7-1.86E-4</td>
<td>1.40E-05</td>
<td>4.00E-07</td>
<td>2.02E-6</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Residual water content (m³/m³)</td>
<td>0.0-0.0278</td>
<td>0</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Saturated water content (m³/m³)</td>
<td>0.247-0.543</td>
<td>0.53</td>
<td>0.373</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 3. Catchment scale simulation water balance. Water balance components for the total catchment as well as values from the grid cells of the three flux tower sites extracted from the distributed catchment. The total catchment albedos are area weighted values and the observed ET values are in brackets.

<table>
<thead>
<tr>
<th></th>
<th>Total catchment</th>
<th>Agri.</th>
<th>Forest</th>
<th>Meadow</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>HYP1</td>
<td>HYP2</td>
<td>HYP1</td>
<td>HYP2</td>
</tr>
<tr>
<td>Albedo</td>
<td>0.18</td>
<td>0.35</td>
<td>0.19</td>
<td>0.32</td>
</tr>
<tr>
<td>Precipitation (mm)</td>
<td>1094</td>
<td>1094</td>
<td>1119</td>
<td>1119</td>
</tr>
<tr>
<td>Evapotranspiration (mm)</td>
<td>481</td>
<td>454</td>
<td>424 (407)</td>
<td>434 (407)</td>
</tr>
<tr>
<td>Canopy Evap. (mm)</td>
<td>192</td>
<td>133</td>
<td>165</td>
<td>83</td>
</tr>
<tr>
<td>Baseflow to river (mm)</td>
<td>513</td>
<td>579</td>
<td>708</td>
<td>780</td>
</tr>
<tr>
<td>Discharge (mm)</td>
<td>412</td>
<td>433</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Overland flow (mm)</td>
<td>31</td>
<td>33</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Pumping (mm)</td>
<td>55</td>
<td>50</td>
<td>104</td>
<td>104</td>
</tr>
<tr>
<td>Drain to river (mm)</td>
<td>207</td>
<td>237</td>
<td>30</td>
<td>164</td>
</tr>
</tbody>
</table>
Calibration of a distributed hydrology and land surface model using energy flux measurements

Morten A. D. Larsen1*, Jens C. Refsgaard2, Karsten H. Jensen3, Michael B. Butts4, Simon Stisen2 and Mikkel Mollerup5.

1 - Technical University of Denmark, Department of Management Engineering, Frederiksborgvej 399, 4000 Roskilde, Denmark
2 - Geological Survey of Denmark and Greenland, Øster Voldgade 10, 1350 Copenhagen, Denmark.
3 - University of Copenhagen, Department of Geosciences and Natural Resource Management, Øster Voldgade 10, 1350 Copenhagen, Denmark.
4 - DHI, Agern Alle 5, DK 2970, Hørsholm, Denmark
5 - The Danish Road Directorate, Guldalderen 12, 2640 Hedehusene.

* - Corresponding author: madla@dtu.dk/004525119895

Keywords: hydrology/land surface modelling; calibration; water and energy fluxes; evapotranspiration; energy closure imbalance

1 Abstract

In this study we develop and test a calibration approach on a spatially distributed groundwater-surface water catchment model (MIKE SHE) coupled to a land surface model component with particular focus on the water and energy fluxes. The model is calibrated against time series of eddy flux measurements from three sites of different land surface type (agriculture, forest and meadow) and river discharge data from the 2500 km² Skjern River catchment in Denmark. The approach includes initial calibrations of three one-dimensional models representing the three land surface types and using the flux measurements for calibration. This step provides
initial values for the subsequent modelling and calibration at catchment scale. To test the validity of
the approach, two additional catchment scale distributed simulations were performed with no
calibration and only calibration of the one-dimensional models, respectively. In addition, a
subsequent validation period was simulated. A mean energy closure imbalance of 20% was seen for
the three sites. For the distributed simulations, the energy imbalance was accounted for by two
ergy balance closure hypotheses ascribing the error to either energy fluxes or net radiation. In
general, the distributed calibration approach improved model results substantially compared to
using default values (no calibration) or calibration of the one-dimensional models only. For the
distributed model simulations, the assumption regarding the energy balance closure had a
substantial impact on the parameter sensitivities and on the simulated discharge and energy balance.
During calibration, the simulation with corrected energy fluxes showed better performance on
discharge than the simulation with corrected net radiation whereas the reverse was true for the
validation period. Regarding energy fluxes, the simulation with corrected net radiation was superior
in both the calibration and validation period.

2 Introduction

Water and energy fluxes between land surface and atmosphere are important components of
atmospheric and hydrological processes. These fluxes can be quantified by the use of land surface
models (LSM) or soil-vegetation-atmosphere-transfer models (SVAT). The calculation of e.g.
evapotranspiration in SVATs and LSMS is based on solving the energy and radiation equations often
on a sub-daily basis and they therefore differ from the less physically stringent schemes often used
in many traditional hydrological models which are based on potential evapotranspiration. LSMSs,
originating from atmospheric sciences, include spatially distributed, often large scale descriptions of
land surface processes. Examples include the Noah model (Rosero et al. 2010) and the CLM model
(Lawrence et al. 2011). LSMs are typically coupled with or forced by atmospheric models and have recently been included in fully coupled climate-hydrology models (Maxwell et al. 2011, Shrestra et al. 2014). SVATs, originating from soil and hydrological sciences, are one-dimensional descriptions typically used for small-scale descriptions linked with soil water flow models (Mauser and Schädlich 1998, Ridler et al. 2012). When included in spatially distributed hydrological models they possess the potential for providing improved evapotranspiration descriptions and enable hydrological catchment models to better utilise remote sensing data to force and constrain hydrological models (Stisen et al. 2011a). SVATs have also recently been included in fully coupled climate-hydrology models (Butts et al. 2014, Larsen et al. 2014). LSMs and SVATs linked to spatially distributed hydrological models are basically similar, and hence we shall in the following refer to both of them as SVAT models.

Assessment of parameter values is critical for the use of SVAT models and an essential challenge is related the vast number of parameters often seen in this type of models. Franks et al. (1997, 1999), Beven and Franks (1999) and Gupta et al. (1999) all highlight the high uncertainty in the predictive capabilities of multi-parameter SVATs due to equifinality. Yet; Franks et al. (1997, 1999) still show good results in terms of reproducing point site flux measurements from the FIFE area in Kansas, USA, and in the Amazon area in Brazil. The added value of a multi-criteria approach as opposed to a single criterion method is confirmed by Gupta et al. (1999) and Demarty et al. (2004). Pollacco et al. (2013) apply an objective function weighing algorithm based on the uncertainties related to remote sensing based surface soil moisture and evapotranspiration calibration variables. Currently no explicit guidelines have been developed on calibrating complex and distributed SVAT and hydrology models.

Parameter estimation for hydrological models is traditionally performed by use of calibration where parameter values are modified to obtain best possible fit between model
simulations and observed target data. While calibration was previously often performed manually by a trial-and-error approach, parameter optimisation by inverse modelling is now the method of choice (Gupta et al. 1998, Madsen 2003, Moore and Doherty 2005). An example is the study by Sun et al. (2013) where inverse calibration based on Monte Carlo-Bayesian techniques was used for calibrating a model both against energy fluxes at point scale and runoff at catchment scale (4.9 km$^2$). In Ingwersen et al. (2011) inverse calibration was used to simulate the water and energy budget for a winter wheat stand at plot scale. Similarly Ridler et al. (2012) utilized inverse techniques for calibrating the combined MIKE SHE/SWET model to simulate energy fluxes at point scale in Mali.

A particular problem related to calibration of SVAT models is that observations of water and energy fluxes are usually not available from operational monitoring networks but only from a few research stations and often for short periods (Wilson et al. 2002, Franssen et al. 2010, Leuning et al. 2012). In addition, energy flux data are known to often have problems with energy balance closure which severely hampers parameter optimisation by inverse modelling, as a SVAT model per definition assumes a closed energy balance (Twine et al. 2000, Choi et al. 2009) and a failure to meet this demand can result in significant biases in long term climate model simulations (Grimmond et al. 2010). Also, the lack of measured energy balance closure will yield an erroneous parameterization when the fluxes are used for the calibration of a hydrological model. Therefore certain assumptions need to be made to account for the lack of closure. To accommodate this, Beven (2006) suggested creating artificial hypotheses to provide closure.

Catchment water balances are linked to energy balances, because the latent energy/evapotranspiration appears as a key element in both balances. The observed catchment runoff and the catchment water balance assessed by hydrological models hence include important information also on the energy balance. On a catchment scale (603 km$^2$) Barr et al. (2012) used
distributed flux tower measurements of evapotranspiration against measured precipitation and discharge for a residual analysis on water balance closure concluding a 15% lack of energy flux closure compared to the measured net energy. For catchment scale (205 km$^2$) model calibration, Li et al. (2011) used both runoff and energy fluxes for model calibration. The CLM4 model was modified to include a runoff scheme and calibration for a 205 km$^2$ catchment was performed manually against, to evaluate both runoff and at the same time considering energy fluxes. Similarly, operating on a regional to continental scale Maurer et al. (2002) simulated energy flux components while the model was manually calibrated only against runoff.

The objectives of the present study is to develop and test a methodology for calibrating and assessing parameters of a SVAT model linked to a spatially distributed hydrological model by using observations of both energy fluxes and catchment runoff. A comprehensive literature study was carried out to obtain feasible initial values and range of variation for parameters for the considered land surface types. The impact of energy imbalance is of particular emphasis and we analyse to which extent inclusion of discharge observations in the calibration process will improve the model performance and robustness.

3 Methodology

3.1 Study area and data

The Skjern catchment (2500 km$^2$) is located in the western part of the Jutland peninsula, Figure 1. The catchment is dominated by sandy soils generated by glacial outwash plains from the last glacial period Weichsel and intersected by older till deposits from the previous glacial period Saalian (Greve et al. 2007). The topography reaches 130 m above sea level in the eastern part of the catchment and the Skjern River flow into Ringkøbing fjord at sea level to the west. The yearly average precipitation for the catchment is 940 mm for the period 2000-2009 based on direct
measurements. When corrected for undercatch using standard monthly correction factors (Allerup et al. 1998) the average precipitation amounts to 1130 mm. In the same period the mean annual temperature is 9.3 °C and the mean monthly temperatures range between 2.1 and 17.3 °C. Inside the catchment flux towers are placed at the three predominant surface types; agriculture (61%), meadow/grass (24%) and forest (13%), Figure 1. At all sites measurements of short-, long-wave and net radiation components; latent (LE), sensible (H) and soil heat fluxes (G); soil water content; precipitation; air temperature; wind speed; and water table levels have been carried out since late 2008. Measurements of radiation and energy fluxes are based on standard methods. Radiation components are measured using a NR01 Hukseflux radiometer (www.hukseflux.com), LI-COR eddy covariance equipment is used for measuring LE fluxes, Gill sonic anemometers for measuring H fluxes, and Hukseflux plates for measuring G fluxes.

The energy flux data used in the study have undergone quality control as part of the processing (Ringgard, 2012) (Step 1.3, Figure 2). Inaccurate observations caused by e.g. low turbulence condition were replaced by data representing similar conditions. Replacement of data was thus for periods with low energy fluxes and therefore this source of uncertainty is expected to be of minor significance. Individual data points clearly outside the expected range at the time of day and season were considered as outliers and removed. (equal to 0.2% on average between LE, H and the three stations weighted relative the areal share). For two periods July 21-August 16 and August 24-October 28, 2009, no flux measurements were available from the agricultural site and data were replaced from the forest site. As these periods are mostly placed in the spin-up period (see below) the calibration results are not expected to be significantly affected.

3.2 Modelling system and setup
This study uses the spatially distributed MIKE SHE hydrological modelling system capable of including all key hydrological processes such as ET, channel flow, overland flow, unsaturated flow, saturated flow as well as irrigation and drainage (Graham and Butts 2005). The land surface model SWET component (Overgaard, 2005) is used in the analysis. SWET is based on the Shuttleworth-Wallace model (Shuttleworth and Wallace 1985). It considers vegetation and energy balance processes in a two-layer system and is extended to include energy fluxes from ponded water and interception storage. Models are constructed for both the three local measurement sites (1D representation) and for the entire catchment (Figure 1).

The 1D models simulate energy fluxes and vertical unsaturated flow based on Richards’ equation. The SWET model is driven by 30 min climatic observations of precipitation, air temperature, wind speed, net radiation, surface air pressure and relative humidity. Measured water table elevations are specified as lower boundary condition. As opposed to the fully distributed model application overland flow, river flow and groundwater flow are not considered. Relevant to both the 1D and distributed analyses, the SWET land-surface module does not consider snow accumulation and melting. Initial values for root depth and vegetation height applicable for the 1D simulations of the agricultural and meadow sites were based on estimates from model simulations for relevant soil type and management conditions using the vegetation model Daisy (Styczen et al. 2004a). Since the footprint of the flux tower at the agricultural site was influenced by several crop types, the vegetation height and root depth were based on average simulations for winter- and spring cereals. The observational data for leaf area index (LAI), another important crop parameter, was inconsistent, and the seasonal variation of this parameter was therefore derived by combining the observed seasonal trends and the simulation results by the Daisy model. The initial values for soil parameters for each site were derived from the HYPRES pedotransfer function (Wösten et al. 1999). Specifically, the Van Genuchten parameters $\alpha$ and $n$ (van Genuchten, 1980) as well as
saturated hydraulic conductivity, residual- and saturated water content were estimated from soil
texture, organic matter and bulk density. The soil texture data were retrieved from grid values from
a distributed soil map of 250 m resolution in the A horizon (0-30 cm) and 500 m resolution in B and
C horizons (30-80 cm and below 80 cm) (Styczen et al. 2004b, Greve et al. 2007, Iversen et al.
2011). For the sensitivity and optimization analyses the soil parameter bounds were defined by
calculating the 90% confidence intervals of variation ranges from Meyer et al. (1997) for soils of
similar texture as for the three field sites. These soil types include sand, loamy sand and sandy loam
for the agriculture and forest site roughly corresponding to JB1 soil (coarse sand) in the Danish soil
classification system (Greve et al. 2007, www.djfgeodata.dk). Due to the lack of agreement on soil
type at the meadow site between site specific soil retention data and available soil maps the 90%
confidence intervals for the meadow site is based on a wider range of soil types from loamy sand to
clay loam (Meyer et al. 1997). The discrepancy of the data is most likely due to highly
heterogeneous soils caused by shifting alluvial deposits from the nearby Skjern River. In the
parameterization the seasonal patterns of vegetation characteristics as induced by both climate and
management practice were kept constant whereas the respective amplitudes were calibrated by a
single factor for each parameter. The annual sequence of these ratios was found by balancing model
simulations by Daisy, relevant literature values (Table 1) and occasional on-site measurements of
LAI and vegetation height. Similarly, the relation between the parameter values in the three vertical
soil type horizons were kept constant implying that the individual parameters for the A, B and C
horizon were shifted by the same factor for each grid cell. By introducing factors between
parameters in space and time the number of parameters to be optimized is reduced.

The MIKE SHE model for catchment scale is based on the Danish national water resources
model (DK-model) (Stisen et al. 2012, Højberg et al. 2013). As opposed to the DK-model the
present model includes the SWET land-surface model and the parameterization of this model is
partly based on the work by Stisen et al. (2011a). The DK model has a 500 m resolution and includes a detailed river network. A maximum time step of 1 hour is used for overland, unsaturated and saturated flow. The MIKE 11 open channel river model component utilizes 3616 cross sections in 100 river branches and uses a 30 min time step. For precipitation exceeding 2 mm per time step an automatic reduction in time step takes place. The input for the catchment scale SWET model is hourly values of the same climatic variables as for the 1D models. These values are obtained by thin plate spline interpolation of climate station data (Stisen et al. 2011a) to a resulting grid in 2 km resolution for all variables except precipitation used in 500 m resolution. Precipitation input is based on interpolation of daily rain gauge data using kriging and dynamically corrected for undercatch. Simulated irrigation is subsequently added to precipitation. Hereafter, simulated irrigation is added to the precipitation input file. The simulated irrigation is calculated from the root zone deficit and a demand function based on well locations, filter depths, annual abstractions and demand area information (Stisen et al. 2011b). The agricultural crop distribution was developed from statistical data, where the various crops were grouped in four classes with different growth characteristics: spring sown cereals, winter sown cereals, grass/clover and maize. The crop classes were distributed without georeferencing due to lack of data. Forest vegetation parameters were based on relevant literature and observations. A total of 36 parameters were analysed in the sensitivity analysis. As the calibration is targeted towards the energy fluxes no sensitivity and calibration analyses are performed on the saturated zone parameterization which is already well calibrated as a part of the DK-model (Stisen et al. 2012, Højberg et al. 2013).

To generate proper initial conditions for soil moisture in the 1D simulations a spin-up period of 5-10 months prior to the calibration period was used. The spin-up periods were decided by data availability as measurements started in April 2009 for the agriculture and meadow sites and December 2008 for the forest site. The distributed model involves a considerable increase in
computation time compared to the 1D setups and a spin-up period of 3 months was therefore used.

Groundwater quasi steady-state was reached by looping simulations in 3-year runs each time using the resulting groundwater heads as initial conditions for the subsequent simulation. Three loops were required to reach quasi steady-state.

3.3 Calibration approach

We propose a calibration and validation approach involving three steps as illustrated in Figure 2: (1) calibration of 1D models representing the flux measurement sites, (2) use of parameter values from step 1 as initial values in the calibration of the distributed catchment model, and (3) validation of the distributed parameterization on independent data. Additionally we test the calibration results by comparing to model results based on 1D calibration only and with no calibration respectively.

Both the 1D and the distributed simulations were calibrated using inverse modelling adopting initial and range of parameter values from literature, observations, databases and simulation results from the model code Daisy (Hansen et al. 1990, 1991) as described above and as listed in Table 1 and 2. The initial input parameter values as well as their ranges were specified according to relevance: If available, either observations or Daisy model simulations based on site specific conditions were used. When using literature values conditions similar to the sites in terms of vegetation species, climate and soil type were used. The calibration period covers a one year period from October 1, 2009 to September 30, 2010 while the validation period is from May 1, 2011 to April 30, 2012.

The sensitivity and auto-calibration analyses were performed using the AUTOCAL software included in the MIKE SHE package (Madsen 2003). At the point scale three components were included in the objective function: latent heat flux LE, sensible heat flux H, and averaged soil water
content based on measurements at three depths (2.5, 22.5 and 52.5 cm below the surface). The objective function was built from the root mean square errors (RMSE) between observations and simulations. The RMSEs for the three individual components were normalized, weighted equally and summed to arrive at an aggregated objective function. The normalization was carried out using the common distance scale which has proven to be robust in accounting for differences in magnitude of each component of the objective function (Madsen 2003, Butts et al. 2004, Mertens et al. 2004).

At catchment scale the optimization was performed against two overall components of equal weight: (1) Discharge observations at three discharge stations aggregated with equal weight (Figure 1) and (2) latent and sensible energy fluxes at the three sites weighted according to the relative areal share of each land surface type (agriculture 62%, forest 13% and grass/meadow 25%).

The sensitivity analysis (Step 1.2 and 2.2 in Figure 2) for both point and catchment scales were based on the AUTOCAL local sensitivity analysis procedure accounting for parameter sensitivities at the location in parameter space defined by the initial parameter values. Since the sensitivity analysis was local and therefore unable to account for numerous combinations in the parameter space, the sensitivity analyses were repeated using randomly sampled parameter sets within the parameter bounds. The sensitivity analysis was based on a backward difference approximation method around the parameter value using a 2% perturbation fraction. Also, covariance matrices were calculated to test for parameter correlation.

For both scales the ten most sensitive parameters were identified and subsequently used in the auto-calibration process similar to Blasone et al. (2007). Only parameters with relative sensitivity coefficients above 1%, in relation to the most sensitive parameter, were included in the optimization (Hill 1998). For the low sensitivity parameters that were not included in the auto-
calibration of the catchment model their values were derived from the point scale calibrations. The remaining distributed parameters were based on relevant literature (Table 2).

The point scale parameter optimization was based on the global Shuffled Complex Evolution method (Duan et al. 1993) which has proven to be robust for comprehensive parameter estimation problems (Butts et al. 2004, Mertens et al. 2005, Blasone et al. 2007). A maximum time step of 6 minutes was used for the 1D models, which for the calibration period corresponded to a computation time of around 2-3 minutes per model run. The model convergence criteria were defined as max. 1% change in objective function after three iterations loops and this criterion was usually met after 200-250 runs. For obtaining a more efficient optimization of the catchment scale model, parallel model runs were carried out using the global Population Simplex Evolution optimization method (DHI 2010). The convergence criterion was typically reached after 250-400 runs.

To assess the accuracy of the parameterization and modelling approach as such, an independent validation analysis was performed subsequent to the auto-calibration simulations. The validation was performed for a one-year period and assessed on discharge, LE and H as for the calibration analysis (Step 3.1, Figure 2).

3.4 Handling of Energy Balance Flux problems

Energy balance closure requires that the net radiation (difference between incoming and outgoing long- and short-wave radiation) equals the sum of latent heat (LE), sensible heat (H), and soil heat (G) fluxes. All terms are highly sensitive to the surface characteristics and furthermore they are subject to different diurnal and seasonal variations. The albedo of bare soil is dependent on texture and moisture content while the albedo for vegetation generally decreases with vegetation height and stand complexity. Changes in albedo will affect the amount of available energy, and soil
moisture influences how this energy is distributed between LE and H heat fluxes. Independent measurements of the components of the energy balance equation rarely provide closure, which has been documented in numerous studies (Wilson et al. 2002, Franssen et al. 2010, Leuning et al. 2012).

The impact of the energy balance closure problem on the calibration results was addressed by creating four hypotheses each based on different assumptions regarding measurement errors (Step 1.4 and 2.4, Figure 2). All four hypotheses were used for the 1D runs whereas only the first two were used in the distributed runs to reduce computation time.

- HYP1: The measured energy fluxes (LE, H and G) are scaled with a single factor (increase) such that the Bowen-ratio is maintained and the sum of the energy fluxes equals measured net radiation over the one-year calibration period;

- HYP2: The measured energy fluxes are maintained whereas the net radiation is scaled with a factor (reduction) to match the sum of measured energy fluxes over the one-year calibration period;

- HYP3: The measured energy fluxes are unaltered whereas the sum of these is used as net radiation on a daily basis;

- HYP4: The measured energy fluxes (LE, H and G) are scaled with a single factor (increase) such that the Bowen-ratio is maintained and the sum of the energy fluxes equals measured net radiation on a daily basis;

For the first two scenarios the entire dataset is multiplied by a single scaling factor to provide energy closure for the one year calibration period. Closure on a daily basis may not be met, which is the case for the last two scenarios. Of the investigated scenarios more energy is available in HYP1 and HYP4 based on the measured net radiation (yearly mean values of 55 w/m², 68 w/m²
and 55 w/m$^2$ for agriculture, forest and meadow respectively) whereas HYP2 and HYP3 are based on reduced values (yearly mean values of 39 w/m$^2$, 48 w/m$^2$ and 42 w/m$^2$ for the same sites).

Only the HYP1 and HYP2 hypotheses were used for the distributed simulations since they both utilize measured net radiation, scaled and unscaled, and hereby the daily temporal pattern is likely to be more representative than if based on the sum of the three fluxes. In contrast to the point scale where net radiation measurements were available, global radiation was used in the catchment scale analysis. To fulfil the same assumptions on energy balance closure as used at point scale the albedo was adjusted to obtain energy balance closure over the entire one year calibration period. This was done based on the flux measurements for the three general land use types agriculture, forest and meadow. As a result HYP1 is based on measured albedos whereas HYP2 is based on estimated albedos that are higher than measured. All three distributed simulation cases shown in Figure 2 were performed for both HYP1 and HYP2 as well as for the calibration and validation periods for a total of 12 simulations.

4 Results
4.1 Energy balance closure

The measurements used in this study show lack of energy balance closure. The sum of LE, H and G over the one-year calibration period accounted for 71%, 71% and 76% of the measured net radiation at the agricultural, forest and meadow/grass sites respectively (Figure 3). The corresponding measured Bowen-ratios in the same period were 0.20, 0.17 and 0.18 for the three land surface types. Radiation and particularly data for energy fluxes are generally sparse in Denmark. Therefore, in Figure 3, measured net and global radiation as well as the sum of energy fluxes are compared to results from (1) the corresponding grid cell in HIRHAM regional climate model simulations (Larsen et al. 2013) using ERA-Interim reanalysis data as boundary conditions.
and (2) observation data from the Foulum research station (grass surface) located app. 50 km north of the agriculture and forest sites. HIRHAM simulation data are also used to compare LE and H fluxes since no other flux data were available. The simulations of net radiation by HIRHAM and the observed data from Foulum are much closer to measured LE, H and G flux sum at all sites for both years although perhaps less distinct at the forest site (Figure 3). For global radiation there is a close match between all data series, especially between the three study sites and Foulum. For LE fluxes good agreement is seen between HIRHAM simulations and observations for the forest and meadow sites and less favourable for the agricultural site. For the H fluxes larger discrepancies are seen particularly for the forest site.

4.2 Simulation results

The results of the sensitivity analyses for both point and catchment scales are shown in Figure 4. Similar parameter sensitivity results are obtained for the agricultural and forest sites with nine and ten parameters, respectively, having a relative sensitivity coefficient above 10%. The parameter sensitivities for hypotheses HYP2 and HYP3 (with unaltered energy fluxes) show similar patterns with the minimal stomata resistance (Rst_min) as the most sensitive parameter. In contrast, for hypotheses HYP1 and HYP4 (with corrected energy fluxes) the root shape factor is the most sensitive parameter. For the meadow site only the three parameters Rst_min, leaf area index (LAI) and vegetation height have relative sensitivity coefficients above 10% and the influence of the energy balance hypothesis is limited given the large drop in sensitivity between Rst_min and the second most sensitive parameter LAI.

At catchment scale saturated hydraulic conductivity (K_sat) for the unsaturated zone is the most sensitive parameter for both energy balance hypotheses, whereas the sensitivities of the remaining parameters are smaller and also subject to larger variation between the two hypotheses.
The forest energy fluxes were found to be poorly reproduced by the calibrated forest leaf width values which were therefore manually calibrated prior to auto-calibration.

The simulated discharge based on the parameter values obtained by the full auto-calibration approach for the two distributed energy balance hypotheses are shown in Figure 5 for both the calibration and validation periods. For both simulations and periods there is a tendency of underestimation of baseflow and mostly pronounced for HYP2. For peak flow the HYP2 hypothesis exhibits more dynamic behaviour with more rapid responses. In summary, the HYP1 simulation shows better discharge performance statistics for the calibration period whereas the opposite pattern is seen for the validation period (Figure 6). Also, the validation confirms a reasonable discharge calibration although the HYP1 results are moderately poorer for two out of three discharge stations.

Figure 7 shows simulated LE and H energy fluxes on an hourly basis for a 10 day summer period in the calibration period. The results from the two energy hypothesis runs are compared to their relevant fluxes: HYP1 to scaled and HYP2 to unscaled fluxes, corresponding to the assumptions of the energy balance closure. For both the calibration and validation periods, there is a tendency for better energy flux statistics for HYP2 compared to HYP1 (Figure 6). Overall the agriculture and grass surfaces are better reproduced compared to forest. Especially observations of H in the forest are rather poorly simulated during high global radiation (Figure 7); however, part of this discrepancy is likely caused by errors in the observation data (Sonnenborg et al. 2013). Simulations for night time conditions compare favourably to observations of both energy fluxes and for both scenarios. As shown in Figure 6 the simulations for the validation period shows comparable or better performance statistics for both fluxes as compared to the calibration period.

In the auto-calibration, simulations are compared to the same observed discharge data for both HYP1 and HYP2 while the energy fluxes and albedos were adjusted to obtain energy balance closure. This is reflected by differences in water balance components for the two hypotheses HYP1
and HYP2 both on catchment and local scales (Table 3). With a lower albedo in the HYP1 scenario higher ET and therefore lower recharge and discharge are simulated at catchment scale. For the agriculture grid cell the slightly higher ET for the HYP2 hypothesis can be ascribed to the lower K_sat value. The significant difference in drainage is likely a function of higher groundwater elevations in the HYP2 hypothesis. For forest and meadow the difference in albedo is reflected in higher ET for HYP1.

The comparison of the performances of the full calibration and the scenarios with calibration of the 1D models only and no calibration respectively is shown in figure 6. For discharge the simulations with no calibration using default values is comparable to the results of the full calibration whereas the 1D calibration is generally poorer. The most significant improvement in the calibration approach is however seen for the LE and H energy fluxes where the mean root mean square error (RMSE) and mean absolute error (MAE) statistics for the three sites is substantially improved. The distributed simulations based on the values from the 1D calibration transferred to catchment scale are generally of lower quality compared to the case where no calibration is carried out. This, as opposed to using initial values using no calibration at all (default values), is because neither input source includes parameters related to the distributed scale (e.g. related to discharge and overland flow). Also, the extreme 1D mean RMSE statistics for H is due to the forest site which was manually calibrated prior to the autocalibration (see above) and it does therefore not affect the full 1D+2D calibration.

5 Discussion

We have presented an approach for a stepwise calibration of a distributed combined land surface/hydrology model using observations of energy fluxes and discharge as calibration target. In general the method proved effective. The validation period confirmed the performance statistics
from the calibration period and the simulations using a reduced calibration approach or no calibration were generally less accurate.

5.1 Energy and water balance

As expected a key finding was that the choice of energy balance hypothesis has a major impact on the resulting parameterization as well as on the energy and water balance results. The lack of energy balance closure for flux measurements is therefore a great challenge when using such measurements for calibrating distributed hydrological models. Energy balance closure is required when flux measurements are used for inverse modelling and we investigated two hypotheses where the error was attributed to either the measured energy fluxes (HYP1) or the measured net radiation (HYP2).

Either of these two approaches may be applicable. Even though the eddy covariance method is generally regarded as the best practical method for measuring energy fluxes, measurement difficulties are well documented (Foken et al. 2006, Sun et al. 2008, Franssen et al. 2010, Ringgard, 2012, Stoy et al. 2013). A consistent deficit in measured LE fluxes have been suggested as an explanation in the lack of energy balance closure (Sun et al. 2008) while on the other hand Foken et al. (2006) suggested that low frequency turbulence structures cause errors in the flux measurements. Also, measurement errors in H fluxes have been suggested, see (Mauder et al. 2008) and Sonnenborg et al. (2013). The energy flux statistics (Figure 6) were better for the HYP2 hypothesis possibly indicating that larger measurement errors are attributed to the energy fluxes compared to radiation components. An additional contributing factor is the difference in the footprint of the flux measurements reflecting conditions upstream of the wind direction and the support scale of the radiation sensor.
Contrary to indications of flux measurement biases, estimates of ET based on eddy covariance and soil moisture balances of the root zone compared well and surprisingly the highest ET was seen for the eddy covariance method over a seven week period in May-June 2009 (Schelde et al. 2011). Similarly, net radiation measurements from Foulum research station match the sum of LE, H and G fluxes better than for the agriculture, forest and meadow sites used in this study (Figure 3). Advection has also been suggested as a missing component in the energy balance suggesting that the measured energy fluxes may indeed be more reliable than anticipated (Leuning et al. 2012). Likewise, large scale eddies due to land surface heterogeneity have recently been suggested to make up a part of the energy imbalance (Stoy et al. 2013).

Snow processes are currently not considered the SWET land-surface model and this impacts the calibration results especially with regard to discharge. Even though the calibration period is relatively cold, little precipitation occurs on days with temperatures below 0 °C. Nevertheless, the snow melt related discharge peaks in Feb-Mar 2010 are clearly not reproduced (Figure 5).

5.2 Sensitivity, calibration and validation

The distinct differences between parameter sensitivities at the agriculture and forest sites compared to the meadow site are related to the water available for ET (Figure 8). At the agricultural and forest sites, complete soil saturation is rarely or never reached and there is a clear tendency for higher ET with lower soil moisture as this occurs during the summer months with a higher level of available solar energy. Soil parameters affecting available soil water are therefore crucial for determining the partitioning of energy fluxes at these sites. For the meadow site, located next to the river bank, saturated conditions are often reached (41% of the year for HYP2) and soil moisture levels are generally high. As a result the relation between root zone water and ET is less distinct.
The high water availability at the meadow site therefore makes vegetation parameters substantially more important compared to soil parameters.

The choice of energy balance closure hypothesis, and therefore the energy available, is highly influential on which parameters are most sensitive especially for the agriculture and forest sites but also to some extent for the meadow site. This is reflected in the sensitivities for the HYP1 and HYP4 hypotheses on one side and the HYP2 and HYP3 hypotheses on the other. This is consistent with the HYP1/HYP4 hypotheses being based on measured net radiation (averaged to 54, 67 and 56 W/m² in the calibration period for agriculture, forest and meadow) while the HYP2/HYP3 hypotheses are based on downscaled net radiation (40, 47 and 40 W/m² for the same sites).

The results of the sensitivity analysis and auto-calibration procedure naturally reflect the interplay between a vast number of parameters, assumptions regarding the energy balance terms, assumed relations between parameters for the individual soil layers, and the specific conditions that apply for the period subject to the analysis. These assumptions are required for reasons such as energy balance closure, calibration simplicity, data availability and focus of the study in question. For example, the parameter Rst_min is assumed constant in the present study whereas Ingwersen et al. (2011) obtained better simulations of LE and H based on the Noah LSM model using monthly values of Rst_min. Hou et al. (2012) related simulations of LE and H to parameter uncertainty in a study using the Community Land Model (CLM4) and documented substantial variations in the simulations of both fluxes (over 100 W/m²) in the summer months depending on the parameterization. The rather poor agreement between simulations and observations of H fluxes was likely due to observation errors. Separate auto-calibration runs were performed assessing forest heat fluxes based on forest parameterization alone with limited improvement and this is in line with the results by Sonnenborg et al. (2013).
5.3 Land use

Standard seasonal crop characteristics were used in this study and the actual crop rotation and management conditions at a given site were therefore not considered. This could lead to discrepancies between the vegetation parameterization at the agricultural site particular around the time of harvest and at the grass site due to irregular and undocumented cattle grazing. Possibly more significant is the effect of the spatial distribution of crops derived from available county average data, which may not reflect the actual distribution. This is the case both in terms of the overall distribution, type and location of crops within the catchment. But it also affects the evaluation of energy flux data from the agricultural site as the actual footprint may differ from the simulated crop as also reported by Göckede et al. (2008).

The distributed MIKE SHE parameterization was performed by transferring the calibrated parameters from the point scale setups, representing agriculture, meadow and forest, to the relevant land surfaces within the catchment for subsequent sensitivity and auto-calibration analyses. For agriculture, this procedure includes the tradeoff between using the calibrated parameter values reflecting site specific conditions at the measurement site and using crop and soil specific parameters from the Daisy model. However, the former was used since parameters were calibrated under conditions of energy balance closure but also because of the statistical spatial distribution of the Daisy modelled crop characteristics.

6 – Conclusions

In the present study we present a novel approach to calibrating a combined and distributed hydrology/land-surface model. The approach first involves calibration of the land-surface model component against measurements of energy fluxes from flux towers. Secondly, the distributed
hydrological model component, including unsaturated and saturated zone components, is calibrated using the point scale parameterization as initial conditions and using stream flow and energy flux measurements as the calibration targets. The method employs a meticulous literature inspection for realistic initial and range of parameter values. Previous calibration efforts of SVAT and LSM models with a hydrological focus have often used satellite data having certain measurement deficiencies, relied on previously point scale based parameterizations or used traditional hydrological variables such as discharge and hydraulic head. Modelling the highly interrelated and dynamical processes and variables in the entire water and energy budget from groundwater to lower atmosphere is challenging. Our results show that the proposed approach significantly improves the simulated energy balance components as compared to using default values or a simpler 1D calibration. The simulations were able to reproduce features across water and energy budgets taking into considerations measurement errors and lack of energy balance closure.

The calibration of the one dimensional models provided good insight to data quality and parameter sensitivity and served as a useful initial parameterization of the catchment scale simulations. Further, the sensitivities of parameters in the point scale analysis were highly affected by the surface type and more specifically the water availability as well as the choice of energy balance hypothesis. Similarly, the catchment model simulations showed that parameter sensitivities are largely determined by the choice of energy balance hypothesis, yet saturated hydraulic conductivity was the most sensitive parameter for both energy balance hypotheses. Despite likely errors in measured forest sensible heat fluxes, our results suggest that the measured energy fluxes are reliable and that the energy imbalance is likely caused by advection and large scale eddies rather than erroneous net radiation measurements.

7 - Acknowledgments
The present study was funded by a grant from the Danish Strategic Research Council for the project HYdrological Modelling for Assessing Climate Change Impacts at different Scales (HYACINTS—www.hyacints.dk) under contract no: DSF-EnMi 2104-07-0008. The study is made possible through a collaboration with the HOBE project (www.hobecenter.dk) funded by the Villum Kann Rasmussen Foundation. The data used in the study can be accessed through the first author.

8 – References


Complete calibration

1D+2D calibration

Step 1  1D Calibration

- Step 1.3 Observation data processing
- Step 1.4 Modification of observation data to obtain energy balance closure

1D calibration only

Step 1.1 Initial + range definitions

Step 1.2 Sensitivity analysis

Step 1.5 Parameter optimisation

Step 1.6 Simulation: 1D Output

Validity tests

Step 2  2D Calibration

- Step 2.3 Observation data processing
- Step 2.4 Modification of observation data to obtain energy balance closure

No calibration

Step 2.1 Initial + range definitions from 1D simulation output

Step 2.2 Sensitivity analysis

Step 2.5 Parameter optimisation

Step 2.6 Simulation: 2D Output

Step 3  Output validation

- Step 3.1 - Independent validation