Interactive comment on “Spatial pattern evaluation of a calibrated national hydrological model – a remote sensing based diagnostic approach” by Gorka Mendiguren et al.

Correspondence to Gorka Mendiguren (gmg@geus.dk)

Response to anonymous Referee #1

Reply: The authors would like to thank the reviewer for his/her detailed and elaborated review of the manuscript. The comments and suggestions are very much taken into thorough consideration as we believe they will improve the reading and add a significant contribution to the manuscript increasing the scientific quality. We are very pleased to read that he/she considers the manuscript appropriate for publication after major revision in Hydrology and Earth System Sciences (HESS). We hope that the changes conducted in the revised version of the manuscript will be well received by the reviewer and that he/she will regard the publication as fit for submission in Hydrology and Earth System Sciences.

1 General comments

The manuscript is dealing with the topic of spatial patterns in distributed hydrological modeling. The authors present a study in which they derived a remote-sensing based ET dataset which they use to improve the spatial patterns of the MIKE-SHE national model of Denmark (DK model). These improvements are achieved by adjusting the parameterizations and input data of the existing DK model. They conclude, that spatial patterns of remote sensing data are a valuable information to inform hydrologic models about spatial patterns, whereas these models are usually calibrated on integral signals, such as streamflow.

The topic fits the scope of HESS and is of scientific interest. The chosen methods seems to be appropriate but some clarification in the methods section is still missing. The authors introduce novelty to the field of applying remotely sensed data for hydrological modeling by consideration of bias insensitive pattern matching techniques. The manuscript lacks, here and there, the soundness of the applied evaluations using scientific methods. A lot of evaluation of the spatial patterns in figures 5 to 7 is done on visual basis without proper numerical/scientific quantification. Some features, e.g., the often mentioned “clear” distinction between model region 5 and 6 are hard to observe for the reader if even existing. Further, the adopted DK model is never evaluated regarding streamflow or groundwater levels which is the main purpose/application of this model. Another criticism is the absence of a proper discussion of the findings of this study, there is little referencing to any other study such as Mu et al. (2007, 2011) for remotely sensing based ET estimates. With exception of section 2.2 the manuscript is well written and good to understand. It could improve by better organization of the sections. I would swap sections 2.3 and 2.4 because sections 2.2 and 2.4 belong together in my opinion. Further, I suggest to fully reorganize and rewrite section 2.2 since it is hard to follow and the storyline is missing in there. Concluding, I suggest to accept the manuscript for publication in HESS after major revision.

Reply:

We agree with the reviewer that the inclusion of statistics of model performance regarding streamflow and groundwater head will add significantly to the manuscript. Initially we decided that the comparison between the performance of the original and modified DK-model would not be fair, since the original model has been calibrated against streamflow and head, whereas the modified has not (single model run time of around 40 hours makes re-calibration a huge task). However we have decided to include a full section comparing the streamflow statistics, water balance errors and groundwater head errors with the original and modified version of the DK model in the manuscript.

Regarding the numerical quantification of the spatial patterns similarities, this is something our group has worked extensively on (Koch et al., 2015) and we currently have another manuscript under review that addresses this issue. However, for this particular study the detailed numerical quantification of the pattern similarities were believed to be a bit irrelevant because the differences are so large between the TSEB and DK-model and therefore a visual interpretation was thought as more appropriate. However we did quantify the correlation coefficients in Figure 10. In a revision we will also add correlations between TSEB AET and different variables (fig 9) to provide a more numerical quantification (Page 12, Lines 29-30).

Regarding the description of the differences between model domains, we agree that this has not been written very clearly and this was improved during revision. But the point was that the lack of spatial consistency in model parametrization results in simulated pattern artifacts. This is still the case and is now exemplified better.

Regarding Mu et al., 2007 we are aware of that product, but chose to produce a satellite based AET dataset based on land surface temperature (LST), which is not included in the Mu model. Their model is primarily driven by LAI and NDVI. This choice is justified by the fact that in our study the LST/TSEB and NDVI/LAI patterns are sometimes very different, especially during the senescence phase. We added an elaborated description of the TSEB-method and discussed how it differs from other regional AET methods and argue why we chose the TSEB.

Koch, J., K. H. Jensen, and S. Stisen (2015), Toward a true spatial model evaluation in distributed hydrological modeling: Kappa statistics, Fuzzy theory, and EOF-analysis benchmarked by the human perception and evaluated against a modeling case study, Water Resour. Res., 51, 1225–1246, doi:10.1002/2014WR016607.

2 Specific comments

Introduction: The introduction is well written and gives a appropriate overview on the topic and shows the novelty of this study compared to existing research.

Methods: In general section 2.2 should be reorganized and rewritten because it is difficult to follow (I am missing the storyline here) and hard to understand what all the variables and equations are needed for. I think a major thing missing here is the presentation of the TSEB equation to assess which variables are needed in order to estimate ET. This will make clearer why you estimate LAI and vegetation height among others.

- please include TSEB equations

Reply: We agree that the method section should be reorganized, and this was done by clearly splitting the TSEB model and the hydrological model in two sections with the appropriate subsections on methodology, remote sensing derived inputs and calibration/validation setup.

The TSEB method cannot be summarized into a single equation, but we understand that it becomes difficult for the reader to follow the TSEB model without some extra explanation. Therefore following the reviewer recommendation we have extended the model description, included several equations that hopefully helps the reader to understand how TSEB works.

However, the model implies several steps and some theory is involved as well so the reader is encouraged to read the article from Norman et al. 1995 to get a complete idea of how the model works in detail.

The text will be modified and the description of the model gets in the manuscript will get this (Page 4 Line 4 – Page 5 Line 9:

"2.1 TSEB setup

2.1.1 TSEB theory

The Two Source Energy Balance Model (TSEB) proposed by Norman et al. (1995) is used to retrieve mean monthly maps of ET across Denmark. In our study we have incorporated the code which is provided by the pyTSEB package (https://github.com/hectornieto/pyTSEB last accessed 30/01/2017). The applied model is a two layer model that treats soil and vegetation separately and estimates fluxes on the basis of LST and air temperature (T_{Air}) among other input variables. As presented in Norman et al. (1995), and presented here in a very simplified explanation. The model is based on the energy balance equation:

$$LE[Wm^{-2}] = R_n - H - G \quad (1)$$

Where H is the sensible heat flux, G is the ground heat flux, R_n is the net radiation and LE is the latent heat flux all in Wm^{-2}.

The sensible heat flux (H) is calculated as:

$$H[Wm^{-2}] = H_C + H_S = \rho C_P \left[\frac{T_C - T_A}{R_A} + \frac{T_S - T_A}{R_A + R_S}\right] \quad (2)$$

Where H_C and H_S are the sensible heat flux for the canopy and soil respectively in Wm^{-2}, T_C and T_S are the canopy and soil temperatures (K), ρC_P is the volumetric heat capacity of air ($Jm^{-2}s^{-1}$), R_S is the resistance to heat (s m^{-1}) flow in the boundary layer above the soil surface and R_A is the aerodynamic resistance (s m^{-1}) expressed as:

$$R_A[sm^{-1}] = \frac{\left[ln\left(\frac{z_U - d}{z_M}\right) - \Psi_M\right]\left[ln\left(\frac{z_T - d}{z_M}\right) - \Psi_H\right]}{0.16U}, \quad (3)$$

where z_U and z_T are the height of the wind speed (U) and air temperature in meters, Ψ_M and Ψ_H are the adiabatic correction factors for momentum and heat, d is the displacement height ($d \approx 0.65\ h_c$, and h_c is the height of the canopy (m)), z_M is the displacement height for momentum ($z_M \approx\ h_c/8$).

The model starts with an iterative process in which finds the T_C which satisfies the energy balance equation. The divergence of net radiation in the canopy (ΔR_n) is used to partition sensible and latent heat fluxes using the Priestley-Taylor approximation (Priestley and Taylor, 1972) for the green part of the canopy. The transpiration is given by the next eq.

$$LE_C[Wm^{-2}] = 1.3 f_g \frac{S}{S+\gamma} \Delta R_n, \qquad (4)$$

Where f_g represents the fraction of LAI that is green, S is the slope of the saturation vapor versus temperature curve and γ is the psychrometric constant and where ΔR_n is calculated as:

$$\Delta R_n = R_n - R_n exp(0.9\ ln(1 - f_c) \qquad (5)$$

And where f_c is calculated as:

$$f_c[-] = 1 - exp\left(\frac{-0.5LAI}{cos\ \theta}\right) \qquad (6)$$

Where θ is the viewing angle.

The model iterates until the energy balance equations are satisfied for soil and canopy. The readers are referred to Norman et al. (1995) to find a fully description of the model for a more detailed explanation of how the iteration process is carried out."

Norman, J. M., Kustas, W. P., and Humes, K. S.: Source approach for estimating soil and vegetation energy fluxes in observations of directional radiometric surface temperature, Agricultural and Forest Meteorology, 77, 263-293, http://dx.doi.org/10.1016/0168-1923(95)02265-Y, 1995.

Priestley, C. H. B., and Taylor, R. J.: On the Assessment of Surface Heat Flux and Evaporation Using Large-Scale Parameters, Monthly Weather Review, 100, 81-92, 10.1175/1520-0493(1972)100<0081:OTAOSH>2.3.CO;2, 1972.

• P4L24: Is the LAI estimate sensitive to its source satellite? So is there any difference in LAI data originating from TERRA compared to AQUA?

Reply: I think you mean the LST instead of LAI (P4L24). TERRA and AQUA satellites have identical characteristics regarding spectral wavelengths of the sensor they carry. Both sensors operate together with different overpass times and therefore the difference in LST is mostly due to the overpass time, not due to sensor.

Regarding the estimation of LAI the MODIS product that we used combines information from both satellites, TERRA and AQUA.

• P4L30: What does BRDF mean?

Reply: we apologize for the typo. We will include in the text what BRDF means: Bidirectional Reflectance Distribution Function. (P5 L20)

• Eq. 1: Please state the wavelengths for B1 and B2

Reply: Wavelengths have been included in the text. B1=645.5nm and B2= 856.5 nm. (P5 L24)

• clarify for what purpose LAI, albedo, VH and others are needed, I think the TSEB equation will help a lot for that

Reply: We have included a more detailed explanation of the TSEB that hopefully helps to clarify this question. We added a brief description in the text indicating how the albedo is used to calculate the net radiation.

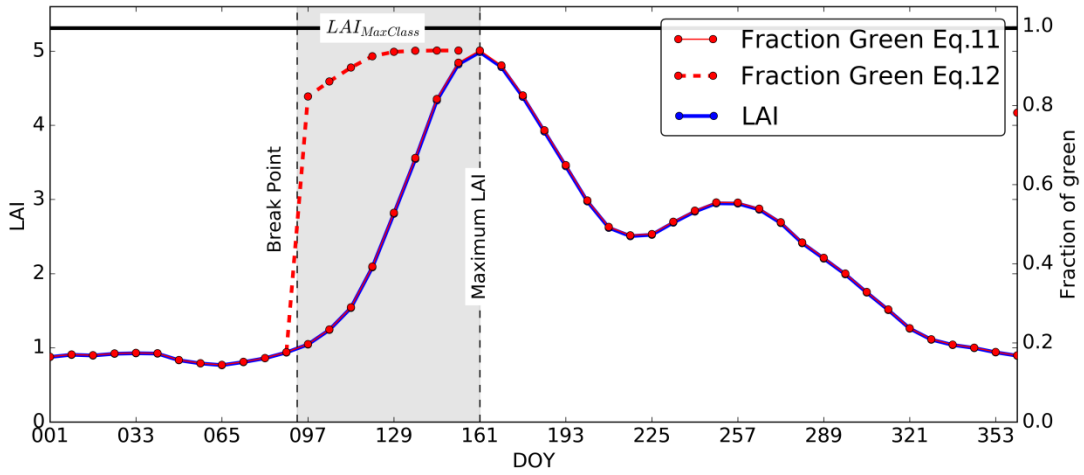
Albedo was used in the study to calculate the net radiation.

• Eq. 6: please explain LAIMaxClass

Reply: LAIMaxClass, is the maximum LAI value for a given land use class. This value is used to scale any given LAI value for a particular grid and day relatively to the land use class it belongs to**).** We will include the explanation of the LAI MaxClass in the manuscript. (P5 L30)

• Fig. 2: probably add the growing phase as a gray box, I think the red line should be dotted outside the growing phase as it was estimated with Eq. 5, merge both legends to one, caption: probably show pixel in map (Fig. 1) - row 100 and column 84 definitely means nothing to anybody, add: LAI corresponds to the left ordinate and Fg to the right one.

Reply: We have followed the reviewer recommendation and added a grey box in the figure. The figure looks now like this.



Caption in the text has been modified following the suggestions and now looks like this:

"Figure 2. Diagram of Fraction of Green (Fg) calculation based on the leaf area index (LAI). LAI corresponds to the left ordinate and Fg to the right one. Shadow highlights the region in time were the Eq.11 is replaced by Eq.12. Data presented corresponds to an agricultural pixel from the dataset."

• Are the data interpolated around the breaking point or is a jump appearing in Fg? How reasonable is that?

Reply: The data is not interpolated in the breaking point. By the approach we implemented in the study, we aimed at detecting the beginning of the growing season by finding the point in which the LAI presented an increase of 20% compared to the pre-growing season low. This approach is taken from Cong et al., 2012. Once the date where that happened (the green-up date) was found we assume that most of the vegetation can be considered and green vegetation and therefore we considered that a rapid change in Fg, reaching values close to 0.9, were meaningful. From there the Fg values follow the evolution of the LAI and decreases as soon as the senescence of the vegetation starts and LAI decreases.

• P6L5-9: How is that approach justified? Do you have any evidence with observations or references in literature?

Reply: The exact approach is not found in the literature; however the 20% increase in vegetation index as green-up date is taken from Cong et al. (2012). The Fg equation itself is a simplified form of the vegetation index based method by Gutman and Ignatov (1998), where we exclude the subtraction of bare soil LAI in the equation. Guzinski et al. (2013), documented the shortcoming of Vegetation index based methods regarding estimating Fg during the greening phase, where they tend to highly underestimate Fg in contrast to the senescence phase, where VI-based methods corresponds well to field observations. We agree with the reviewer that the explanation of what we conducted to retrieve the Fg was very difficult to understand and read (in addition we also had put in a wrong equation). We have completely rewritten this section hoping to make it clearer for the reader. In principle we divide

the Fg estimation into two periods, one is the greening phase and one is the senescence and winter phase. This will allow us to get a more realistic evolution in Fg, with high values throughout the greening phase and gradually changing Fg as a function of LAI during the rest of the year. The new text says:

Page 6 Lines 9-21

"….Fraction of Green vegetation was derived from LAI following the next equation:

$$Fg_i = \frac{LAI_i}{LAI_{MaxClass}} \qquad \textit{(EQ 11)}$$

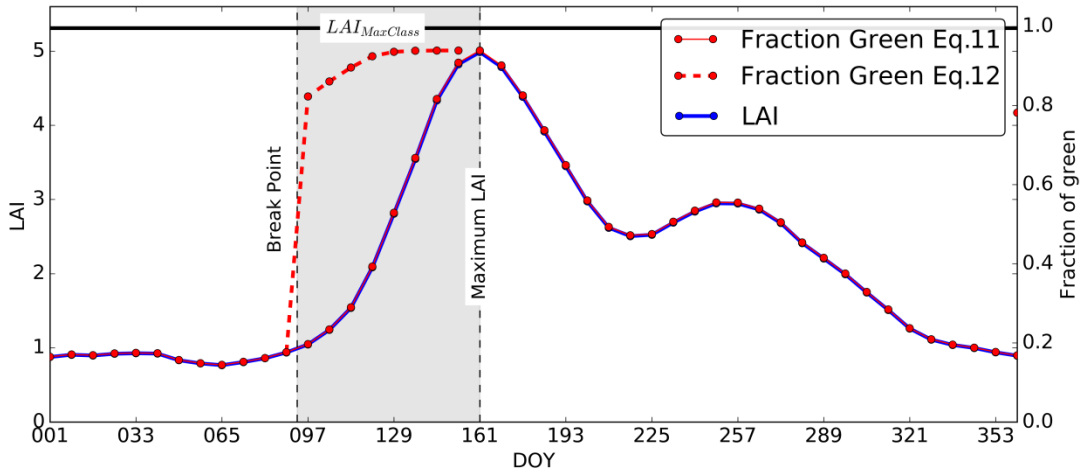
Where Fg_i indicates the Fraction of green for a certain pixel i, LAI_i indicates the LAI value for a pixel i and $LAI_{MaxClass}$ is the maximum LAI value for a specific land cover type. The approach is a simplified form of the vegetation index based method by Gutman and Ignatov (1998). This equation was applied to needle leaf forest land cover type.

For the other land cover types (crops, grasslands, deciduous forest etc.) equation 11 was modified adding another term. These land cover types show a stronger seasonality and a clear distinction between a greening phase and a senescence phase. In order to represent the strong difference in fraction of green vegetation between the period before and after senescence we introduced a different equation for the period between crop emergence and senescence, where we assigned higher values of Fg to non-needle leaf forest land covers, Figure 2. For these types of vegetation Fg will be allowed to increase rapidly just after crop emergence by substituting EQ 11 by EQ 12.

$$Fg_i = \frac{LAI_{i,max}}{LAI_{MaxClass}} \cdot (1 - e^{(-2 \cdot LAI_i)}) \qquad \textit{(EQ 12)}$$

Where $LAI_{i,max}$ indicates the Maximum LAI value for a pixel i.

This substitution is only conducted during part of the phenological year, more specifically for the period starting at the green-up date, corresponding to the point defined by an increase in 20% in LAI compared to the winter low ($LAI_{i,Min}$)(Cong et al. 2012) and continuing until the time at which LAI reaches its maximum ($LAI_{i,Max}$) (see next Figure)." This approach will mediate the shortcomings of the vegetation index based methods, which has been shown to underestimate fraction of green during the greening phase while corresponding well to field observations during senescence (Guzinski et al. 2013).



We believe these assumptions fits well with reality, since a given LAI value before and after senescence can have quite different Fg values. During the growing season most of the plant remains green, which is quite well represented with the modification including the exponential term in the equation. After the point, where vegetation has reached its maximum seasonal development (we assume at maximum LAI), senescence starts and more non- photosynthetically active regions start to appear in the plant, what is translated in lower Fg values.

Cong, N., Piao, S., Chen, A., Wang, X., Lin, X., Chen, S., Han, S., Zhou, G., and Zhang, X.: Spring vegetation green-up date in China inferred from SPOT NDVI data: A multiple model analysis, Agricultural and Forest Meteorology, 165, 104-113, http://dx.doi.org/10.1016/j.agrformet.2012.06.009, 2012.

Gutman, G., and Ignatov, A.: The derivation of the green vegetation fraction from NOAA/AVHRR data for use in numerical weather prediction models, International Journal of Remote Sensing, 19, 1533-1543, 10.1080/014311698215333, 1998.

Guzinski, R., Anderson, M. C., Kustas, W. P., Nieto, H., and Sandholt, I.: Using a thermal-based two source energy balance model with time-differencing to estimate surface energy fluxes with day-night MODIS observations, Hydrology and Earth System Sciences, 17, 2809-2825, 2013.

• Eq. 8: I think you mean EF = ET/Rn

Reply: Yes we meant that, we apologize for the mistake. It has been also changed in the text. (P7 L9)

• P6L30-32: I do not understand this sentence

Reply: We have modified the sentence as it was confusing. In the manuscript it was written as:

"The observed values used during TSEB evaluation are the Bowen ratio (Bowen, 1926) corrected values and the associated uncertainty estimate is the span between all error in the closure problem being assigned to the latent heat and no error in the latent heat."

And now is written as:

Page 7 Lines 26-28

"The evaluation of the TSEB was conducted using as reference the data of the EC systems from the 3 different land cover types that were corrected for the energy closure using the Bowen ratio approach (Bowen, 1926). The uncertainty of the energy closure issue spam between all closure errors being assigned to sensible heat for the first limit and for the second limit all error being assigned to latent heat."

Bowen, I. S.: The ratio of heat losses by conduction and by evaporation from any water surface, Physical Review, 27, 779-787, 10.1103/PhysRev.27.779, 1926.

• Eq. 9 and 10: Why was the original RD approach based on LAI adopted to NDVI. LAI is available as seen on previous page. Could you please elaborate a bit on that?

Reply: We apologize for the confusion, Koch et al. 2017 mentions describing seasonal root depth evolution by scaling either LAI or NDVI, but actually applies the NDVI based scaling for Danish agricultural conditions. We have followed that implementation.

We noticed that this was not very well explained in the text and we have rewritten that part. Please, check now Page 8 Lines 20-Page 9 Line 11.

• Eq. 9: Please explain NDVIi the same as NDVImax

Reply: We have rewritten all this section in the new manuscript and explained the terms.

Page 8 Lines 20-Page 9 Line 11.

• P7L7: LAI in meters?

Reply: Sorry, only root depth in meters, LAI in [m2/m2]

• How is RDmax estimated?

Reply: The RDmax values are fixed in time but varying in space. For the very sandy soils in Western Denmark, the effective maximum routing depth is known to be lower than for the regions with lower sand content. This is described on page 25 of Refsgaard et al. (2011) (http://vandmodel.dk/xpdf/77-2011_vandbalance.pdf) and in (Jensen and Jensen 1999). The RDmax values are scaled as a function of clay content so that the mean values of RDmax across the entire country match the average calibrated root

depths of the original DK-model. This is done to ensure the water balances of the original and modified DK-model does not deviate too much, while avoiding a time consuming (several months of simulation time) recalibration of the modified DK-model. It also has to be acknowledged that the Root depth in the DK-model is considered an effective parameter, that also accounts for variations in vegetation type, phenology and soil types, because it is the main control in the ET estimation.

Jensen and Jensen (1999) Textbook - Soilphysics and agricultural meteorology (In Danish)

• I don't understand what "matching the original DK model" for RD and KC means. I thought the aim is to make them variable. How did you achieve to make them matching, by parameter calibration? Please elaborate a bit more on that.

Reply: We agree with the reviewer that it becomes a little bit confusing. In all the study we have used the original configuration of the DK-Model as reference. To derive the LAI from NDVI, RD, etc… The aim of doing this is to keep the spatial average values of the parameters across the country as similar as possible and focus only on changing/improving the spatial distribution of them without affecting so much the mean model inputs and therefore avoiding a new model calibration. See also the reply above.

• P8L3-6: At P6L14-16: you state an actual value comparison is not anticipated. Here you are calibrating your TSEB model with eddy covariance data. Why? Please elaborate more on that.

Reply: We decided to adjust a few of the vegetation related parameters of the TSEB model based on land cover specific eddy covariance data from 3 different towers representing different vegetation types. Even though we are not utilizing the absolute values of the TSEB (only the pattern) the patterns are representing the differences in space and therefore we believe it is valuable compare and adjust the TSEB to different land covers where flux tower data were available.

We will elaborate on this in a revised manuscript, which will also include a clearer explanation of the sensitivity analysis and calibration of the TSEB.

• P8L1-2: Wouldn't a variance based sensitivity method better fit the purpose of identifying the parameters which have to be used for model calibration instead of the derivative based approach applied herein? Probably provide some details about the chosen sensitivity approach.

Reply: Yes a variance based sensitivity method would be better for the purpose of identifying parameters which have to be used for model calibration. However, our sensitivity analysis was not really designed for that purpose. We use the simple sensitivity analysis to illustrate that the TSEB is mainly sensitive to the remotely sensed variables and the climate forcing which are not subject to calibration, and less sensitive to the vegetation parameters. Subsequently we adjust a few of the TSEB vegetation parameters to get a better discrimination between land cover classes, but the selection of parameters is based more on subjective choices than the sensitivity analysis. We will explain this better in a revised manuscript and also argue why we select the parameters we do for calibration.

• P9L5,L8: Please make a distinction between the terms parameter and variable, the reader gets confused otherwise.

Reply: Sorry you are right, changed to**:**

"The results show that the most sensitive variable for the estimation of AET is LST. Interpreting the sensitivity values for each group individually stress that, for the remote sensing input, parameters that are directly related to LST such as emissivity of vegetation (EmissV) and soil (EmissS) are characterized by a high sensitivity as well. The next group, forcing data, exhibited high sensitivity for all variables, except for wind speed. Overall Air temperature (TempAir) is the most sensitive forcing variable."

• P9L8/Fig 3: better: TempA = Ta

Reply: Thanks for the suggestion. The legend has been modified in the figure.

• Fig. 3: LAIAgri max , LAIForest max , LAIMeadow max do not appear.

Reply: When the sensitivity analysis was conducted, LAI was not evaluated specifically. Instead we evaluated the sensitivity of the fraction of green vegetation fg. Later when calibrating the model we utilized the LAImax for each land cover class to adjust the fg, which is estimated based on equation 5(11) and 6(12). (we unfortunately found a typo in eq5, which also includes LAImaxclass) and linearly proportional to LAImaxclass We realize that this part of the manuscript is weakly explained and we will improve it in the revision.

• P9L14 & P8L5-6: Why did you select only those 4 parameters out of 10. For PT the others seems to be more sensitive then the forest PT, for example.

Reply: The choice of parameters was subjective. We chose not to calibrate PTmeadow and PTagri, because there is not really any physical reason that it should deviate from the original value of 1.28. In contrast literature (Komatsu, 2005) suggests that PTforest is generally lower than 1.28, and therefore we decided to calibrate that value. Canopy height was considered better parameterized with realistic values for both forest and seasonally varying crop height for agriculture, therefore we preferred not to adjust the canopy height. Leaf width was not included due to low sensitivity.

Komatsu, H.: Forest categorization according to dry-canopy evaporation rates in the growing season: comparison of the Priestley–Taylor coefficient values from various observation sites, Hydrological Processes, 19, 3873-3896, 10.1002/hyp.5987, 2005.

• Please justify the assumption to add the residual energy to LE. I only know approaches using corrections based on the Bowen ratio or adding the residual energy to SH.

Reply: We used the standard Bowen ratio corrected data from the EC systems for our calibration of TSEB. The purpose of adding the residual energy to LE or SH in the figure was mainly to illustrate the size of the energy balance closure issue and to illustrate that the TSEB estimates were generally within the limits.

• Fig. 4: Thanks for including error bars to the plot. I think it is misleading showing only the error bars of the observation. Could you also show error bars on the simulation, e.g., emerging from different parameter sets?

Reply: The idea of the figure is to show that values of ET that we estimated from TSEB are within the uncertainty of the reference data. The red lines in the figure represent the range of the energy balance closure problems, they do not include the entire uncertainty range, since uncertainty in the measurements themselves are not included. It would be great to have an estimate of the uncertainty of the TSEB, but even if we included uncertainty arising from the vegetation parameters that would only cover a fraction of the true uncertainty. Given that the TSEB is mainly sensitive to the LST, albedo and climate forcing, the uncertainty of those constitutes a much larger uncertainty. Therefore uncertainty bars based on parametrization alone would be misleading. We will elaborate on this in the revised manuscript.

• P9L20: Could you please mentioned the spatial resolutions of EC and RS data? • The results section is missing in general a discussion with other studies. E.g., estimating ET from MODIS data comparing to Mu et al. (2007, 2011).

Reply: The resolution of the remote sensing dataset is at 1 km.

EC spatial resolutions can vary. The resolution depends on the heights at which the measurements are taken. Another factor that affects the footprint size is the surface roughness, and the last is the thermal stability therefore suggesting a footprint size is difficult. We know that at all stations the instruments are located at best possible locations and heights to be representative of the area and capturing the fluxes in area of tens of meters to hundreds of meters, but under some conditions might be affected by fluxes from areas nearby but not much as the EC captures most of the eddies from the area nearby the station.

• P10L14-15: Is it reasonable to observe lower ET for forest areas? Wouldn't canopy interception increase ET only after precipitation events?

Reply: We agree with the comment from the reviewer. However, we are only evaluating non-cloudy days, meaning that there is no rainfall and thereby no interception in our evaluation. When we compared the data from the EC of the different sites, we noticed that the ET of croplands was higher than those obtained in the forest areas, for these specific cloud free days especially during the peak of the crop growing season (May-July). It has to be remembered that the days we are evaluating are not representative of all conditions but limited to cloud free conditions. Most probably the AET maps for all weather conditions would look different.

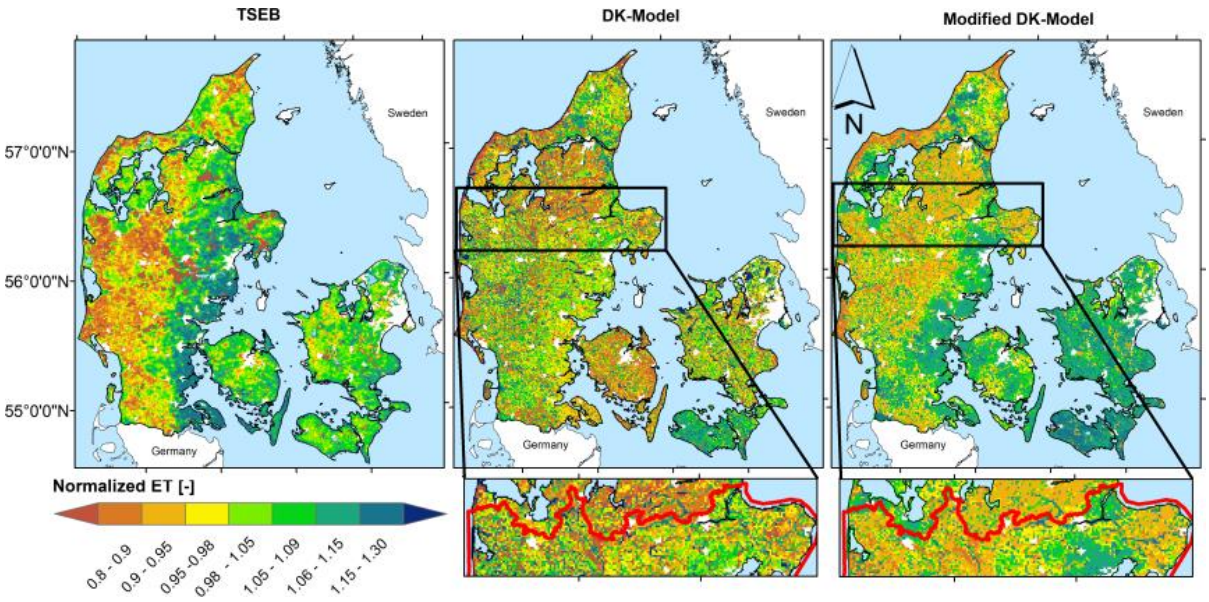
• P10L17: are causing differences in area 2 in the model domain

Reply: We changed the text in the manuscript.

• P10L19, Fig 6, P11L5, and others: I am very sorry but I cannot observe the pronounced difference between zone 5 and 6. Could you provide some more information on that, e.g., zoomed plot numerical analysis? At the provided plots I do not see this features.

Reply: We agree that it is difficult to observe the difference in the contrast between domain 5 and 6. This is partly as a consequence of using the same color ramp and color stretch for all the maps in the figure. We have changed the text to focus on the differences between domains 1,2 and 3 in the DK-model.

Figure 8 below, which might be a better example of the differences between domain 5 and 6.



P10L10: reformulate: extracted

Reply: I think you mean P10L20. We have reformulated to:

Figures 5 and 6 indicate that there is very little resemblance between the spatial patterns of the TSEB ET and the DK-model simulations on the national scale.

- P10L24: ... does not necessarily lead to reasonable ET ...

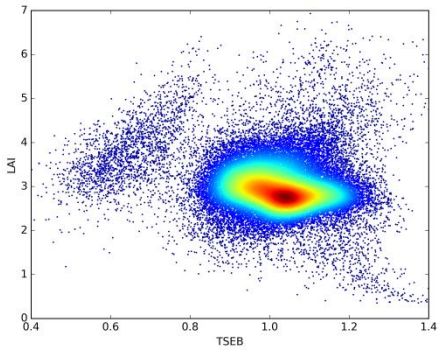
Reply: Thanks for the suggestion. Text has been modified.

- Fig. 9 and P11L9-15: Could you provide numerical evidence to the explanatory variables of the spatial patterns of ET. I can see the E-W gradient in clay content and ET but the others are not observable. Consider rewriting or deleting some of your conclusions since they are not supported by your data. Possibilities to get evidence: scatterplots or SPEARMAN rank correlations.

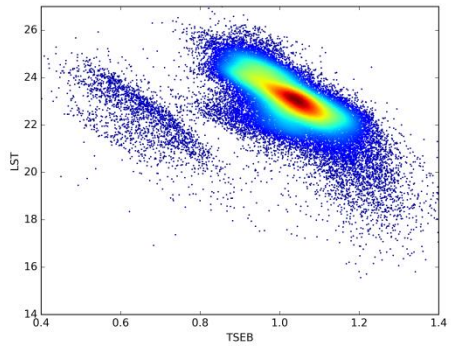
Reply: We agree that the visual interpretation should be backed by some quantification; we will add correlation coefficients between variables and TSEB AET to the maps in figure 9.

We do not consider necessary incorporating the scatter plots in the manuscript as there are already a large number of figures and maps in it. However, we will show them here:

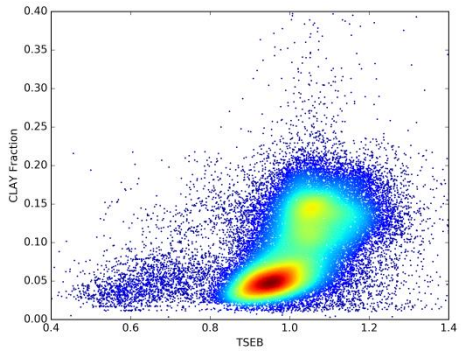
Density scatter plot TSEB- LAI (r =-0.15)



Density scatter plot TSEB- LST (r =-0.50)



Density scatter plot TSEB- Clay Fraction (r = 0.44)



• I miss the comparison of the model performance in streamflow and groundwater table between the original and modified DK model. I understand that the spatial representativeness of the modified DK has improved compared to the original one. But shouldn't be made sure that the water balance is still sufficiently represented by assessing the streamflow and groundwater tables since that is the major purpose of the model? Therefore, the model performance shouldn't deteriorate significantly if evaluated with those variables

Reply: We agree with the reviewer that some information on the performance of the model should be provided. We will include in the manuscript a new section that evaluates the performance of the Original and the modified version of the DK-Model with respect to streamflow and groundwater head. We first included the results regarding the discharge. (See next figure)

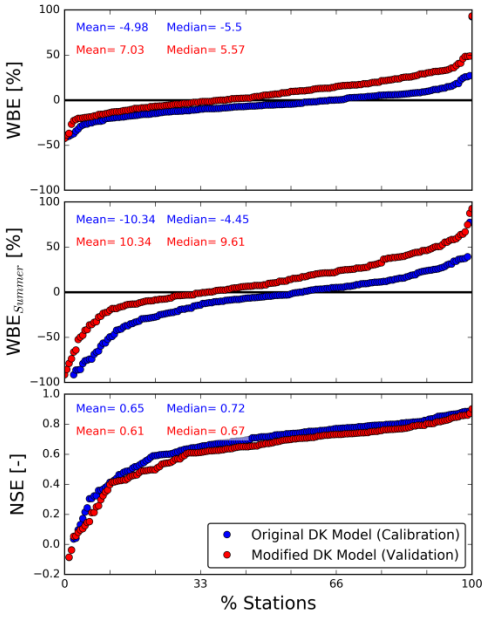


Figure 12. Model performance statistics showing the results of the original and modified versions of the DK model. Stations have been ranked by performance and presented in the x axis as a percentage. Figure shows the Nash-Sutcliffe Efficiency (NSE), water balance error (WBE), and water balance error only for the summer period (WBESummer)

Following the suggestion of the reviewer we have also included the results of the water heads (See next figure)

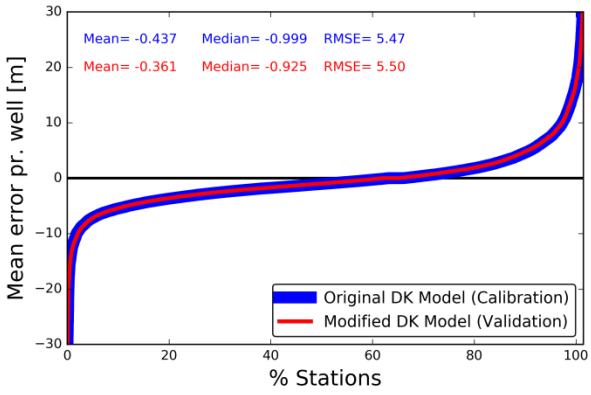


Figure 13. The figure shows the error and statistics in the ground water heads estimated from the original DK model and the modified version of the DK model. Stations have been ranked by performance and presented in the x axis as a percentage.

In both cases the performance statistics have decreased, but not in a very significant way. We expect the statistics of the modified version of the DK model to be similar to those of the original model after a new model calibration that includes also a spatial performance metric in the objective functions, however, at this moment that task is not yet feasible as it requires a new framework to carry on the calibration focusing on both, the spatial pattern performance and the temporal discharge performance.We are starting now to develop the framework and data preparation, but is still unknown when it will be finalized as it also requires a large computing time once all the model inputs and datasets are ready. It is important to highlight that the performance of the original model is a calibration performance whereas the performance of the modified model is a validation.

We will include a new section in the text were the results of the modifications are included.

3.3 Stream discharge and groundwater head DK model performance

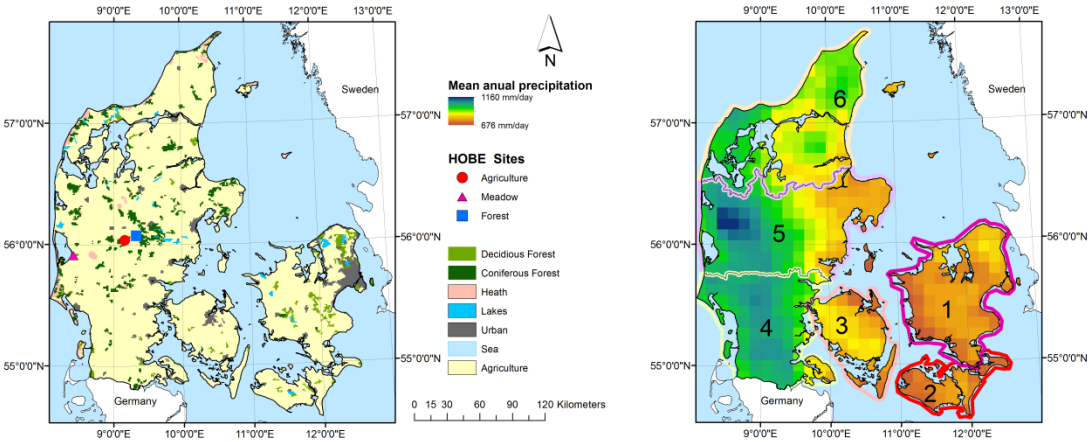
Besides comparing the spatial patterns of the original and modified DK-model, the stream discharge and groundwater head performance is also compared. In this comparison it is important to acknowledge that the original DK- model has been calibrated against these variables, whereas the evaluation of the modified DK-model has to be considered as a validation. Results showing the annual water balance error (WBE) and summer (Jun-Jul-Aug) water balance as well as NSE (Nash-Sutcliffe Efficiency) for 181 discharge stations are presented in Figure 12. The first noticeable thing that can be concluded is that the average water balance error changes from a slight overestimation to a moderate underestimation (Median WBE changes from -5.5 % to 5.5% for the original and modified models respectively). Regarding, the summer water balance which is expected to be influenced the most by the model modifications; the picture is similar although the performance get worse with a larger positive bias. The NSE showed a decrease in performance, from NSE= 0.72 in the original DK model to NSE=0.67 in the modified version.

Ground water heads were also evaluated for 25,365 wells across the country and results are shown in figure 13. The results in this case are very similar between the original version and the modified one. Statistics showed a RMSE of 5.5 m in both cases, with the RMSE being dominated by relatively few very large errors while 78 % of the wells have absolute errors below 5 m. The similarity in simulated groundwater heads between the two model versions indicates that the changes in evapotranspiration patterns have little effect. However, it has to be considered that the simulated groundwater head is controlled by mainly hydraulic conductivity (which does not change between the two versions) and annual recharge upstream of the point of comparison. Since the changes in evapotranspiration patterns mainly effects the summer period, where recharge is low, the effect on annual recharge is limited. In addition, the changes in evapotranspiration patterns will redistribute recharge patterns, but the combined effect of that at some deeper well filter location will be a mixed signal causing limited changes in groundwater head.

The results of this comparison are promising considering that the model was not re-calibrated with the new inputs. In the future, the model will be recalibrated including a spatial metric as an objective function during the calibration, and it is believed that especially the model bias on discharge can be minimized.

3 Technical corrections

• Fig. 1: excluded in figure) - parentheses missing, consider using different symbols for Agri and Meadow because they are hard to distinguish.



Corrected

• P2L10: rational behind developing. Corrected.

• P2L21-25: because you do not provide exhaustive list of references for each application example i suggest to use 'e.g.,' in front of the references. Corrected.

• P3L7: Figure 1 presents the herein used study domain. Text added to caption of figure 1. Corrected.

P4L27: I would put the LAI sentence to the previous paragraph and start the new paragraph with: "The study focuses". Corrected.

• P5L6: delete successfully after Boegh et al. Corrected.

• P5L7: this study instead of the study - you should check that in the entire manuscript

Reply: We checked and corrected in the manuscript.

• P5L7: similar approach was applied where ... - please delete "was applied" later in the sentence Corrected.

• P5LL25:please do not introduce abbreviation like 10U which are never used in the manuscript Corrected.

• P6L5: To identify the different periods, first, the dates ... Corrected.

• Is LAIi.max the same as LAIMax in Eq. 4 and Eq. 9? check consistency

Reply: Equation 9 was unfortunately miswritten, the correct equation is in the revised manuscript

$$Fg_i = \frac{LAI_i}{LAI_{MaxClass}}$$

Regarding EQ4: LAI should be LAI_i (LAI of current cell and timestep) while LAI max should be LAI_{iMAX} is the maximum LAI for the current grid cell. This is different from the LAImaxclass in EQ 11 and 12, which corresponds to the maximum LAI for the entire land cover class.

• P6L6: breakpoint Fig. 2 not figure 3 Corrected.

• P6L6: better: breakpoint Fig. 2, i.e., the onset of the growing season

Corrected. We have rephrased this section.

• P6L8: Eq. 6 instead of 5

Corrected

• you are switching from Eq. to equation and Fig. to figure in the entire manuscript check consistency • probably check for figure and equation referencing in the entire manuscript

Corrected. We have checked the consistency in the manuscript for EQ and Fig.

• Eq. 7 & 8: netRad and net radiation - consistency

Corrected

• I would suggest to use formula symbols like Rn instead of words like netRad

Corrected

• P6L20: The resulting maps ...

Corrected

• P6L21: in just climatological maps

Corrected

• P6L26: latent heat (LE) or evapotranspiration measurements are

Corrected

• P6L29: which is usual instead of not unusual

Corrected

• P7L7: RDi = RDi

Corrected

• P10L11: Fig. 5 instead of Fig. 56

Corrected

• P10L11: pattern identified the TSEB

Reply: I think you meant P10L20. Corrected

• P12L22: the meso.. instead of The meso..

Corrected

Bowen, I. S.: The ratio of heat losses by conduction and by evaporation from any water surface, Physical Review, 27, 779-787, 10.1103/PhysRev.27.779, 1926.
Cong, N., Piao, S., Chen, A., Wang, X., Lin, X., Chen, S., Han, S., Zhou, G., and Zhang, X.: Spring vegetation green-up date in China inferred from SPOT NDVI data: A multiple model analysis, Agricultural and Forest Meteorology, 165, 104-113, http://dx.doi.org/10.1016/j.agrformet.2012.06.009, 2012.
Gutman, G., and Ignatov, A.: The derivation of the green vegetation fraction from NOAA/AVHRR data for use in numerical weather prediction models, International Journal of Remote Sensing, 19, 1533-1543, 10.1080/014311698215333, 1998.
Guzinski, R., Anderson, M. C., Kustas, W. P., Nieto, H., and Sandholt, I.: Using a thermal-based two source energy balance model with time-differencing to estimate surface energy fluxes with day-night MODIS observations, Hydrology and Earth System Sciences, 17, 2809-2825, 2013.
Komatsu, H.: Forest categorization according to dry-canopy evaporation rates in the growing season: comparison of the Priestley–Taylor coefficient values from various observation sites, Hydrological Processes, 19, 3873-3896, 10.1002/hyp.5987, 2005.
Norman, J. M., Kustas, W. P., and Humes, K. S.: Source approach for estimating soil and vegetation energy fluxes in observations of directional radiometric surface temperature, Agricultural and Forest Meteorology, 77, 263-293, http://dx.doi.org/10.1016/0168-1923(95)02265-Y, 1995.
Priestley, C. H. B., and Taylor, R. J.: On the Assessment of Surface Heat Flux and Evaporation Using Large-Scale Parameters, Monthly Weather Review, 100, 81-92, 10.1175/1520-0493(1972)100<0081:OTAOSH>2.3.CO;2, 1972.
Refsgaard, J. C., Stisen, S., Højberg, A. L., Olsen, M., Henriksen, H. J., Børgesen, C. D., Vejen, F., Kern-Hansen, C., and Blicher-Mathiesen, G.: DANMARKS OG GRØNLANDS GEOLOGISKE UNDERSØGELSE RAPPORT 2011/77, Geological Survey of Danmark and Greenland (GEUS), 2011.

Interactive comment on “Spatial patternevaluation of a calibrated national hydrological model – a remote sensing based diagnostic approach” by Gorka Mendiguren et al.

Correspondence to Gorka Mendiguren (gmg@geus.dk)

Response to anonymous Referee #2

Reply: The authors would like to thank the reviewer for his/her detailed and elaborated review of the manuscript. The comments and suggestions are very much taken into thorough consideration as we believe they will improve the reading and add a significant contribution to the manuscript increasing the scientific quality. We are very pleased to read that he/she considers the manuscript appropriate for publication after major revision in Hydrology and Earth System Sciences (HESS). We hope that the changes conducted in the revised version of the manuscript will be well received by the reviewer and that he/she will regard the publication as fit for submission in Hydrology and Earth System Sciences.

Major comments:

Mendiguren et al. studied the importance for a hydrological model simulation to reproduce similar spatial patterns as those of remote sensing data. Their modified version of the DK-model provides a simulated evaporation result that has more similar spatial features found in the remote sensing based ET. Generally, I read the paper with great interest. The paper fits very well within the stated scope of journal. I consider that the evaluation of spatial patterns of model simulation result is still quite novel and rarely done in common hydrological model practices. However, there are still some major issues to be addressed before this manuscript is being accepted for publication.

The authors present an improved model in which remote sensing derived data were used for parameterizing vegetation parameters. They claim that the improved model provides better results as it has more similar spatial patterns as those of remote sensing data. However, it seems that the benchmarking of simulation results is limited to the evaporation flux only (especially its spatial pattern). I suggest performing more evaluation and comparison to the ‘original’ and ’improved’ model simulation results, particularly to their discharge and groundwater head results.

Reply: The other anonymous reviewer addressed a similar comment. In the revised version of the manuscript we have included a new section in which we present the results of the original DK model and the modified DK version**.**

In addition, I fell that the presentation, writing and structure of the paper must be upgraded.

Often, there is no clear gap/interval between paragraphs. There are paragraphs and sentences do not flow with their previous ones. These make the paper difficult to read and understand in quite a lot of places. Moreover, in the Introduction section, I hardly find any sentences related to the actual or the main objective of the study (which is to evaluate spatial patterns of a model simulation result?). I also think that the presentation and structuring of the Methods section must be improved. Furthermore, I recommend to have separated sections of Results and Discussion. Please also see some suggestions in the following minor comments.

Reply: We agree with the reviewer that the main objective can be better introduced, and we strive to do so in a revised manuscript. However, we have thought a lot about splitting the results and discussion and decided that it will be better to keep them together. We feel that the interpretation and discussion is better communicated along with the presentation of results. We hope the reviewer can accept this.

Minor comments:

Page 1, lines 13-15: "The main hypothesis of the study is . . . ." I suggest rephrasing this sentence. Moreover, I could not find this hypothesis in the Introduction section.

Reply: We agree with the reviewer that the sentence is confusing and therefore have removed it in the new version of the manuscript.

Regarding the second part of the comment, we highlighted and clarified what the objectives of the study are including also the hypothesis of the study.

Page 1, lines 26-28: Using your modified version of the DK-model, did you get any improvements on your discharge and groundwater head simulation results?

Reply: No, we did not get any improvement on the results of discharge and ground water head simulations (these results will be added, see reply to reviewer 1). Generally we do not expect the results to improve for discharge and heads when using the modified DK-model without re-calibration. We recently submitted another manuscript where we include a spatial metric in the calibration of a hydrological model in a smaller sub catchment. In this study the results indicated that very similar performance metrics for discharge can be achieved when the spatial component is included (indicating limited tradeoffs), but we do not expect the results to improve, but with a closer spatial pattern to the observed using remote sensing.

Page 2, lines 21-24: I suggest including some references about the applications of using satellite data for assessing groundwater as well (e.g. Rodell et al., 2009, http://dx.doi.org/10.1038/nature08238; Sutanudjaja et al., 2013, https://doi.org/10.1016/j.rse.2013.07.022; Richey et al., 2015, http://dx.doi.org/10.1002/2015WR017349). I guess that they are relevant for your study as you use the DK-model that simulates groundwater head dynamics. I am also curious how the improvement that you introduced (based on remote sensing data) affects groundwater head simulation.

Reply: we have included the references in the new version of the manuscript. Included in Page 2 Line 18.

Page 3, lines 1-3: Could you please elaborate more with what you meant by the "eminent risk" here? Some references will be helpful.

Reply: What we mean is that when a hydrological model is calibrated individually for each model grid, with independent parameter values adjusted to fit the land surface temperature at the grid level, there is a great risk of producing a parameter field that represents a physically unrealistic spatial distribution, even though it minimizes the error in each grid, because the model parameters are used to also compensate for pixel level uncertainties in the satellite product. Thus this would overestimate the

credibility of the remote sensing data. Instead we promote the idea that the model should be calibrated using relatively few global parameters linked to transfer functions which are linked to spatial distributions of basin characteristics (e.g. DEM, soil texture, vegetation, etc.). This approach aims at identifying which parts of the model parametrization generates these differences in the spatial patterns

Page 3, lines 14-16: Could you please elaborate more with what you meant by the "diagnostic approach"?

Reply: By a diagnostic approach we mean that we seek to identify which parts of the model parameterization causes the differences (errors) in spatial patterns between the satellite based estimates and models and between the two models. We will rephrase that part to make it clearer to:

"…The model evaluation is based on a diagnostic approach inspired by the study of Schuurmans et al. (2003) who utilized satellite estimates to identify conceptual model errors in a small sub basin of the MetaSWAP model in the Netherlands. This approach aims at identifying which parts of the model parametrization generates these differences in the spatial patterns."

Page 3, lines 26-27: Neglecting biases/differences in the absolute values is a quite brave assumption. Could you please provide some justification behind it? Why?

Reply: We believe that the greatest value of the satellite based AET estimates is related to the spatial pattern information and not to the absolute values, in this study. Our hydrological model evaluation scheme is based on the assumption that spatial patterns are best observed by satellite measurements whereas the overall water balance is better observed by the aggregated stream discharge measurements. Therefore we chose to only evaluate the spatial pattern of the model against the TSEB monthly maps while the water balance is evaluated against stream discharge. A limitation of the TSEB data regarding water balances is that we only observe AET on specific cloud free days and therefore our evaluation of the spatial patterns is limited to the same days, which will not guarantee a reasonable evaluation of the overall water balance.

In addition, there are different sources for the net energy data behind the hydrological model and TSEB. In the case of the DK model the potential evapotranspiration is used as provided by the Danish Meteorological Institute (DMI) whilst for the TSEB the net radiation data was obtained from the ECMWF dataset. The differences between these two introduce a small bias in AET that is not the focus of our study.

Page 3, line 30: This should be "Sections 2.2 and 2.3".

Reply: We apologize for the mistake. This has been corrected and updated in the new version.

Section 3: Please consider to reorganize the structure (sub-sections) of Section 2. I found that it is quite difficult to understand and follow the sequence of each step of your methods.

Reply: We fully agree with the reviewer that the structure of the manuscript in its actual form is difficult to follow, and we apologize for the extra effort of the reviewer to follow the study.

We have decided to reorganize the structure following the reviewer recommendation. In the new manuscript we have divided the methods in three defined groups, one for the TSEB and a second one for the hydrological model and a third for the EOF, following a more logical sequence. This will also allow a more detailed theoretical description of TSEB as requested by the reviewers.

The scheme for the new methods section in manuscript looks as follows:

2.- Methods

2.1.- TSEB

2.1.1 TSEB theory

2.1.2. Derived remote sensing inputs

2.1.3 Sensitivity analysis and TSEB calibration

2.2.- Hydrological model

2.2.1.- Remote sensing derived hydrological model input data

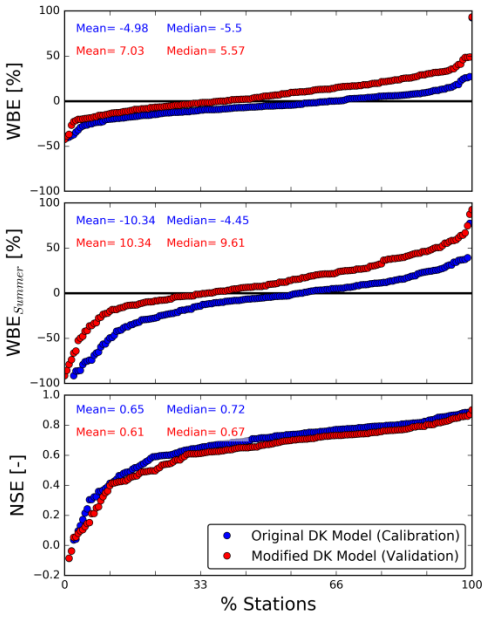
2.3.- Spatial patterns analysis: Empirical orthogonal functions

Page 4, lines 10-11: You have discharge and groundwater head measurement data. Could you please evaluate the results of your improved model to these data?

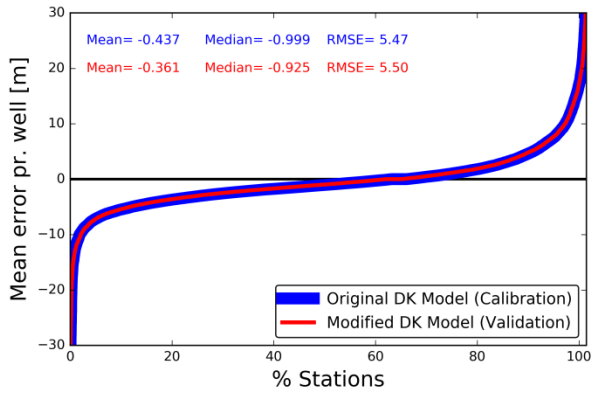
Reply: Reviewer 1 has pointed to this and we agree that this information should be included in the manuscript. To reduce the size of the response, please read the response give to reviewer 1 in the last comment before the technical corrections.

We include the figures for guidance.

Water balance error:



Ground Water heads:



Please, check section 3.3 in the new version of the manuscript.

Page 4, line 19: "TSEB can successfully be applied with a single LST observation, . . . " This sentence is not clear for me. What do you mean by "single LST observation"?'

Reply: We agree that it is a bit confusing the sentence. The TSEB separates the fluxes for soil and canopy by first calculating the temperature of the soil and the canopy from a measurement that contains both. When two observations from the same area are measured simultaneously this can be obtained easily as we have a system that is composed of two equations with two unknowns.

$T_{Rad}(\theta) = [f(\theta)T_C^n + (1 - f(\theta))T_S^n)]^{1/n}$, where θ is the observation angle, TRad is the radiometric temperature, Ts is the soil temperature, Tc is the canopy temperature and n is the power of temperature that is usually 4.

On the other hand, when a single observation is obtained it is necessary to iterate until the temperatures of soil and canopy provide a good solution to the energy balance equation.

Equation 1: Please put the reference to this equation. Moreover, it will be very helpful if you include the unit or the dimension for every variable. Example: NDVI [-]; LAI [-]; VH [unit: m]

Reply: NDVI is referenced (Rouse et al. 1973) few lines up in the text but we agree that the other formulas need to present the units and the references; therefore we have included in the new manuscript the formulas with the units of the variables.

Page 5, line 1: ". . . MODIS band number." Please be more specific with these "band numbers".

Reply: we have included in the manuscript the wavelength each of the bands corresponds to. In the case of the study the wavelengths are B1 (645.5 nm) and B2 (856.5 nm).

Equations 2 and 3: Please put the reference to these equations (e.g. as you introduced the reference for Equation 4).

Reply: Equation 2 has the reference from Boegh et al. (2004) two lines before the formula, and equation is the same as Equation 2 in which the coefficients have been calculated.

Page 5, line 10: Please include the dimension/unit for the parameters alpha and beta. I guess they are dimensionless.

Reply: LAI is measure in m2/m2, alfa is m2/m2 and betta is dimensionless. We have modified equation in the manuscript with the units.

Page 5, line 18: This sentence does not flow well with the ones before it.

Reply: Sorry for the mistake. We have moved it to a more appropriate location, right after the equation.

Page 5, line 21: Could you please elaborate more with what you meant by the "highest quality pixels"?

Reply: The MODIS MCD43B3 provides information on albedo and the quality flags are contained in MCD43B2 product. When we use the term highest quality pixels we mean that only those pixels in which the Quality Flag is good were used to generate the maps.

We have included a better explanation in the manuscript. Now the manuscript says:

"…across different years using only the pixel where the quality flag indicated the albedo was categorized as good quality of the pixel."

Page 5, line 27: I miss the explanation why you have to calculate "Fraction of Green"? For what purpose?

Reply: Reviewer 1 also stated the necessity to give an explanation on this. The reason why fraction of green needs to be calculated is due to the equation the TSEB is using to calculate the fluxes.

The model uses the next equation to calculate the latent heat from the green canopy:

$$\mathrm{LE_C} = 1.3\mathrm{f_g}\frac{\mathrm{S}}{\mathrm{S}+\gamma}\Delta\mathrm{R_n}$$

Where $\mathrm{f_g}$represents the fraction of LAI that is green, S is the slope of the saturation vapor versus temperature curve and γ is the psychrometric constant.

In the new revised version of the manuscript we have included a more elaborated description of the model that we hope helps to understand why the different inputs are generated. Please, find the new explanation of the model in reviewers 1 first specific comment.

Page 5, line 27 to Page 6, line 9: Please rewrite this part. I hardly follow it. And can you please justify this assumption to the reality?

Reply: We agree with the reviewer that the explanation of what we conducted to retrieve the Fg was very difficult to understand and read (in addition we also had put in a wrong equation). We have completely rewritten this section hoping to make it clearer for the reader. The new text says:

"….Fraction of Green vegetation was derived from LAI following the next equation:

$$Fg_i = \frac{LAI_i}{LAI_{MaxClass}} \qquad \textit{(EQ 11)}$$

Where Fg_i indicates the Fraction of green for a certain pixel i, LAI_i indicates the LAI value for a pixel i and $LAI_{MaxClass}$ is the maximum LAI value for a specific land cover type. The approach is a simplified form of the vegetation index based method by Gutman and Ignatov (1998). This equation was applied to needle leaf forest land cover type.

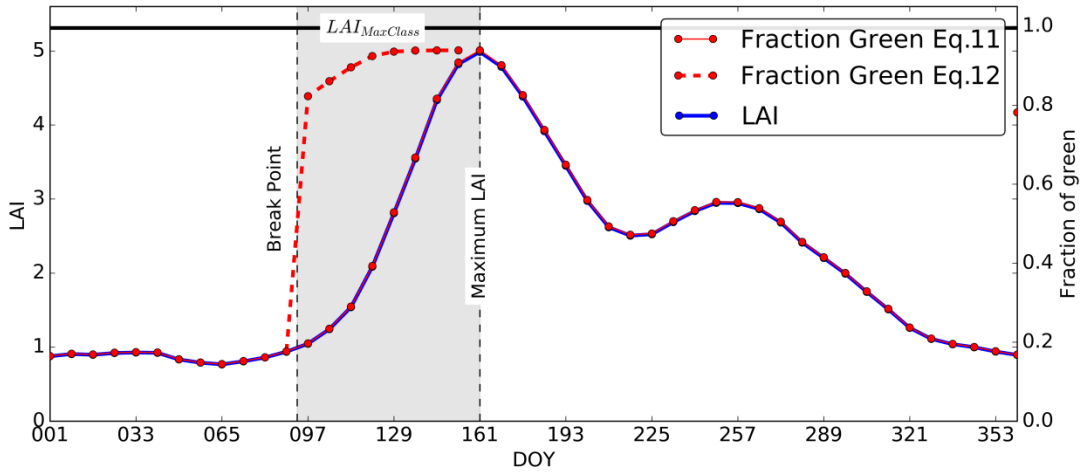
For the other land cover types (crops, grasslands, deciduous forest etc.) equation 11 was modified adding another term. These land cover types show a stronger seasonality and a clear distinction between a greening phase and a senescence phase. In order to represent the strong difference in fraction of green vegetation between the period before and after senescence we introduced a different equation for the period between crop emergence and senescence, where we assigned higher values of Fg to non-needle leaf forest land covers, Figure 2. For these types of vegetation Fg will be allowed to increase rapidly just after crop emergence by substituting EQ 11 by EQ 12.

$$Fg_i = \frac{LAI_{i,max}}{LAI_{MaxClass}} \cdot (1 - e^{(-2 \cdot LAI_i)}) \qquad \textit{(EQ 12)}$$

Where $LAI_{i,max}$ indicates the Maximum LAI value for a pixel i.

This substitution is only conducted during part of the phenological year, more specifically for the period starting at the green-up date, corresponding to the point defined by an increase in 20% in LAI compared to the winter low ($LAI_{i,Min}$)(Cong et al. 2012) and continuing until the time at which LAI reaches its maximum ($LAI_{i,Max}$) (see next Figure). This approach will mediate the shortcomings of the vegetation

index based methods, which has been shown to underestimate fraction of green during the greening phase while corresponding well to field observations during senescence (Guzinski et al. 2013).”



We believe this assumption fits well with reality, since a given LAI value before and after senescence can have quite different Fg values. During the growing season most of the plant remains green, which is quite well represented with the modification including the exponential term in the equation. After the point, where vegetation has reached its maximum seasonal development (we assume at maximum LAI), senescence starts and more non- photosynthetically active regions start to appear in the plant, what is translated in lower Fg values.

Page 6, line 10: This sentence does not flow well with the ones before it. Equations 7 and 8: What is the difference between “Net Radiation” and “netRad”. If they are the same, please be consistent with your variable names. Please also include the dimension/unit.

Reply: the reviewer is right here. We have homogenized it and include the units (W/m^2)

Page 6, lines 22-23: I hardly understand this sentence.

Reply: What we mean is that the monthly TSEB maps are calculated based on cloud free pixels on days with a high overall fraction of cloud free pixels, therefore they are not representative of all weather conditions. Moreover, when we calculate monthly mean maps from the hydrological model we only average the exact same grids and days as used to make the TSEB maps.

Page 6, lines 24-32: “Data from three eddy covariance (EC) flux towers is used as a reference to perform a sensitivity analysis and calibration of some of the vegetation parameters of the TSEB.“ I guess that the methods for sensitivity analysis and calibration are given in the Section 2.4? Please consider to reorganize Section 2 so that all your method steps are presented in a logical sequence.

Reply: we agree with the reviewer in this point and new restructure of the manuscript has been conducted in the new version of the manuscript.

Section 2.3: It makes more sense to put this Section 2.3 directly after Section 2.1.

Reply: We agree with the reviewer and the manuscript sections have been rearranged in a more logical way.

Page 7, lines 8-9, Equation 9: Why did you have to substitute LAI with NDVI?

Reply: In the first stages of the study we used the approach based on LAI but we found some problems when utilizing that equation to convert to root depth. In this study the LAI was developed based on an empirical relationship based on an exponential equation on NDVI and LAI. When converting LAI to root depth that resulted in very abrupt temporal transitions in root depths during the year. Therefore we change the root depth calculation to be a function of NDVI in order to keep the simple linear formulation in eq. 15/16 in the revised version.

Equations 11 and 12: Please put some references.

Reply: We cannot provide any reference for equation 11 in the old manuscript as it was created by us during the study. We noticed that the way the equations are presented is not the best as some values appear like fix coefficients and might create confusion. We have rewritten the eq. 16 and eq. 17 in the new manuscript (P9L1-L16) more clearly. References for equations that are not from this study are now also included in the new text.

Page 8, line 1: ". . . perturbation with respect to a change in model performance." What is your model performance objective function? RMSE? NSE? KGE?

Reply: Mean Error was selected as the objective function. The way it was conducted was finding the parameters where it was obtained minimum error between the monthly mean values of ET for each month and the 3 different sites.

Page 8, line 5: How did you choose these four parameters?

Reply: The choice of parameters was subjective. We chose not to calibrate PTmeadow and PTagri, because there is not really any physical reason that it should deviate from the original value of 1.28. In contrast literature suggests that PTforest is generally lower than 1.28, and therefore we decided to calibrate that value. Canopy height was considered better parameterized with realistic values for both forest and seasonally varying crop height for agriculture, therefore we preferred not to adjust the canopy height. Leaf width was not included due to low sensitivity. See also reply to reviewer 1 regarding the use of the sensitivity analysis.

Equation 13: What is the best value of SEOF? 0? Please clarify.

Reply: This is a good question. Ideally the best score should be 0 which will indicate that there is a matching in the spatial pattern. Regarding the upper limit, it will depend on the spatial variability of the data; high loading values are found for cases with a distinct variability and vice versa. This makes it difficult to put the score into context or to give a physical relevancy. Also the comparison between models or catchments may be limited.

On the other hand, as 0 is the most similar spatial pattern can be used as an objective function and therefore can be used in calibration.

Section 3: I recommend to have separated sections for Results and Discussion.

Reply: we consider is more appropriate the way is presented in the actual form as it allows the reader to follow the story line of the manuscript. Please also see our reply to the general comments.

Section 3.1: Sensitivity analysis: Can you please explain more about how you perturb your model input data? What is their range for the maximum and minimum values of each variable?

Reply: The sensitivity analysis is very simple, we perturb each variable or parameter by a fixed percentage compared to the initial and evaluates the change in objective function relative to teh change in variable/parameter. For temperatures these are changes based on their ^{o}C values.

Page 8, line 14: How did you choose these four parameters? Based on your sensitivity analysis (Fig. 3)? How?

Reply: We have addressed this question previously.

The choice of parameters was subjective. We chose not to calibrate PTmeadow and PTagri, because there is not really any physical reason that it should deviate from the original value of 1.28. In contrast literature suggests that PTforest is generally lower than 1.28, and therefore we decided to calibrate that value. Canopy height was considered better parameterized with realistic values for both forest and seasonally varying crop height for agriculture, therefore we preferred not to adjust the canopy height. Leaf width was not included due to low sensitivity.

Section 3.1: TSEB calibration: How realistic is your calibrated TSEB result map (including its spatial pattern)? Did you compare it to other studies (e.g. to MODIS, GLEAM, etc.)?

Reply: The reviewer makes a good point here and we really consider that is a task that should be addressed in a future study. We did not conduct any comparison against other models i.e MODIS, GLEAM etc… As was mentioned in the study we could use information from three different eddy covariance towers to evaluate the estimates from TSEB.

Regarding the representativeness of the spatial patterns we believe they are representative as LST is highly correlated with ET. Moreover, as shown in the sensitivity analysis the LST drives the TSEB and as shown in figure the correlation between LST map and TSEB is quite high. We believe that LST is a key indicator of evaporative state at the surface and that LST based ET algorithms are more appropriate

than purely vegetation based ones eg. MODIS (Mu et al 2007). Other models such as GLEAM are very much fusion of models and remote sensing data (e.g. including a soil moisture model and plant water stress model) and as such becomes difficult to regard as an observation. Even though TSEB is also a model it is very much driven by the key observations of LST and vegetation, making it more observation based.

Figures 5, 6 and 7: Please use the same and consistent legend (color and values) for all figures so that they can easily be compared.

Reply: we agree with the reviewer that the legend should be homogenized. The spatial patterns in Figures 5,6 and 7 do not represent themselves well on a common color scale. Therefore we have now normalized the maps from figures 5,6 and 7 with their own mean to focus only on the spatial pattern, as was conducted in figure 8. However that will mean that the seasonal component of the ET will disappear. It's open for discussion which presentation is better, real values (mm/day) or normalized values on a common color scale. The results and discussion were adjusted to the new maps.

Page 9, lines 25-28: "The main aim of TSEB . . . ." It seems that the sentences do not flow with their previous ones.

Reply: We included a break point in the text.

Figure 8: How did you normalize all three maps? What was the motivation to normalize these three maps?

Reply: We noticed that we did not explain in the manuscript how the maps were normalized. The way they were normalized was by dividing the mean map by the mean value of the mean map itself.

We have included the explanation in the caption of the figure to clarify it and in the text

Figure 10: Please improve the caption for Fig. 10. What does the color mean here?

Reply: In these types of plots that are called density scatter plots the warmer the color the higher number of points (higher frequency). Therefore those points that are bluish indicate low frequency of points whilst the red colors indicate high frequency of points.

Page 10, line 16: MIKE-SHE = DK-model?

Reply: Yes, sorry for the confusion here. The text has been checked for this confusion and replaced were necessary.

Page 11, lines 1-15: Please rephrase the sentences in these lines. This seems a very important finding in your result. I guess that this result appears very dependent on your choice to use Equation 11 (Page 7, line 19). I am wondering how you derive this equation, particularly their factor 12 and constant 0.2? Can we implement this equation for other study areas, e.g. to other climate regions (e.g. tropical areas).

Reply: The reviewer makes a good point here. We do not believe that this equation will provide good results in other areas, mostly because in our case was specifically calibrated for the study. Basically the equation is composed of two terms, the maximum RD and a factor that scales it and distributes it especially. For the agricultural land covers we used a more elaborated approach than the one used in forest areas for example. For the first term of the equation we established an empirical relationship between the maximum RD and clay fraction for 7 different soil types based on the data from the original DK model setup. The linear fitting gave as result the constants shown in the equation.

Page 12, line 27 to Page 13, line 16: I suggest starting a new subsection for this part.

Reply: We agree here with the reviewer that a subsection will be helpful. We have included it in the revised manuscript(P14L13).

Others:

Figure 1: For the figure on the right, what do the numbers (1 to 6) stand for?

Reply: The numbers state for the DK model subdomains. The DK model runs in 7 different domains (the one corresponding not Bornholm is not included in the study here), therefore the 6 numbers and lines that divide the country aim at showing the domain numbers and the location of each domain.

Figure 2: Please indicate the pixel (row 100, column 84) in Fig. 1.

Reply: We decided to make it more general and just specify the type of land cover that is represented as we also believe that does not add any additional information to the manuscript.

Spatial pattern evaluation of a calibrated national hydrological model – a remote sensing based diagnostic approach

~~G.~~Gorka Mendiguren [1], ~~J.~~Julian Koch [1], ~~S.~~Simon Stisen [1]

[1]Department of hydrology, Geological Survey of Denmark and Greenland, Copenhagen, Denmark.

Correspondence to: Gorka Mendiguren (gmg@geus.dk)

Abstract. Distributed hydrological models are traditionally evaluated against discharge stations, emphasizing the temporal and neglecting the spatial component of a model. The present study widens the traditional paradigm by highlighting spatial patterns of evapotranspiration (ET), a key variable at the land-atmosphere interface, obtained from two different approaches at the national scale of Denmark. The first approach is based on a national water resources model (DK-~~model~~Model), using the MIKE-SHE model code, and the second approach utilizes a two source energy balance model (TSEB) driven mainly by satellite remote sensing data.

~~The main hypothesis of the study is that while both approaches are essentially estimates, the spatial patterns of the remote sensing based approach are explicitly driven by observed land surface temperature and therefore represent the most direct spatial pattern information of ET; enabling its use for distributed hydrological model evaluation.~~ Ideally the hydrological model simulation and remote sensing based approach should present similar spatial patterns and driving mechanism of ET. However, the spatial comparison showed that the differences are significant and indicating insufficient spatial pattern performance of the hydrological model.

The differences in spatial patterns can partly be explained by the fact that the hydrological model is configured to run in 6 domains that are calibrated independently from each other, as it is often the case for large scale multi-basin calibrations. Furthermore, the model incorporates predefined temporal dynamics of ~~Leaf Area Index~~leaf area index (LAI), root depth (RD) and ~~Crop~~crop coefficient (Kc) for each land cover type. This zonal approach of model parametrization ignores the spatio-temporal complexity of the natural system. To overcome this limitation, ~~the~~this study features a modified version of the DK-Model in which LAI, RD, and KC are empirically derived using remote sensing data and detailed soil property maps in order to generate a higher degree of spatio-temporal variability and spatial consistency between the 6 domains. The effects of these changes are analyzed by using the empirical orthogonal functions (EOF) analysis to evaluate spatial patterns. The EOF-analysis shows that including remote sensing derived LAI, RD and KC in the distributed hydrological model adds spatial features found in the spatial pattern of remote sensing based ET.

1 Introduction

The application of spatially distributed hydrological models has become common practice for a wide range of water resources assessments. Such models are valuable tools to manage terrestrial water resources and to provide insights into the overall water balance as well as the internal distribution of multiple hydrological states and fluxes. The spatial predictability of these models is however severely hampered by the general lack of suitable spatial pattern oriented model evaluation frameworks since evaluation remains focused on spatially aggregated objective functions such as discharge. As stated by Conradt et al. (2013) "In conjunction with distributed hydrological modelling spatial calibration usually means individual multi-site calibration". The neglect of a specific focus on spatial patterns in model evaluation is a paradox in the light of an increasing acknowledgement of the role of patterns in the functioning of hydrological systems (Vereecken et al., 2016). Moreover it is against the ~~rational of~~rationale behind developing and applying distributed models (Freeze and Harlan, 1969;Refsgaard, 1997). If the spatial variability of a hydrological system is not of importance to the modeler it seems not worth the effort to apply a distributed model, since numerous studies indicate that equal model fidelity can be achieved with a lumped approach when evaluated solely at the catchment outlet (Stisen et al., 2011;Vansteenkiste et al., 2014)

The concept of spatial pattern comparisons in catchment hydrology was pioneered by Grayson and Blöschl (2000), who developed the theoretical framework and terminology. Since then significant progress has been made in areas such as model code development (Clark et al., 2015;Maxwell and Kollet, 2008;Samaniego et al., 2010), remote sensing (Lettenmaier et al., 2015) and data assimilation (Zhang et al., 2016;Ridler et al., 2014). Nevertheless, explicit spatial pattern evaluation of distributed hydrological models remains a rarity.

In order to perform qualified assessments of simulated spatial patterns reliable observations are a prerequisite. For this purpose satellite remote sensing comes into play as an independent data source with the required spatial resolution and coverage for many catchment scale applications. Satellite imagery has been used for estimation of numerous states and fluxes of interest to hydrological modelling, such as snow cover (Immerzeel et al., 2009), ~~storage change (Chen et al., 2016)~~ground water storage change (Chen et al., 2016;Rodell et al., 2009;Sutanudjaja et al., 2013;Richey et al., 2015), soil moisture (SM) (Wanders et al., 2014), vegetation water content (Mendiguren et al., 2015), land surface temperature (LST) (Corbari et al., 2015) ~~and~~or actual evapotranspiration (ET) (Guzinski et al., 2015). The conversions of the remotely sensed signal to hydrological variables is far from trivial, and usually require in situ measurements and observations for model evaluation. However, in spite of their overall uncertainty; satellite based estimates contain valuable pattern information (Mascaro et al., 2015). More precisely, we propose to utilize remote sensing data solely for the purpose of pattern validation, using bias insensitive metrics, and leave the task to validate water balance closure to the more trusted discharge observations. Several studies have explored the obvious potential in utilizing satellite estimates for hydrological model evaluation and calibration. Some utilize time series of a basin average observation from remote sensing to guide model calibration (Rajib et al., 2016;Rientjes et al., 2013) and therefore rely heavily on the accuracy of the remote sensing estimate while neglecting the spatial pattern information. Others utilize the satellite based estimates to perform a pixel to pixel calibration of the model

Formatted: Danish
Formatted: Danish
Field Code Changed
Formatted: Danish
Field Code Changed
Formatted: Danish
Formatted: Danish
Formatted: Danish
Field Code Changed
Field Code Changed
Formatted: Danish
Formatted: Danish
Field Code Changed
Formatted: Danish
Formatted: Danish
Field Code Changed
Formatted: Danish
Formatted: Danish
Field Code Changed
Formatted: Danish

(Corbari and Mancini, 2014). The latter approach might explore the full information content of the observations. However, there is ~~an eminent~~a risk of a highly parameterized solution problem and possibly unrealistic spatial parameter distributions as each grid cell is parameterized independently. Such a highly parametrized approach will be weak in light of the uncertainty of the remote sensing estimates at the pixel level. Instead we advocate approaches that seek to utilize the general pattern information of remote sensing data with less focus on specific pixels/grids and the general bias.

Relatively few studies utilize the actual pattern of remote sensing estimates in distributed hydrological model evaluation. Interesting examples are (Li et al., 2009) who used remote sensing derived patterns of actual evapotranspiration, and Hendricks Franssen et al. (2008) who utilized satellite based recharge patterns to constrain the calibration of groundwater models. Immerzeel and Droogers (2008) included actual evapotranspiration estimates in the calibration of the Krishna basin in southern India and obtained a good correlation across 115 sub basins while Githui et al. (2012) successfully applied a multi-objective calibration by combining river discharge and remotely sensed actual evapotranspiration of 59 sub basins in Victoria, Australia. Other pattern oriented model evaluations without calibration were conducted by Bertoldi et al. (2010), Wang et al. (2009) ~~and (Koch et al., 2016)~~ and Koch et al. (2016); the latter applied different spatial performance metrics in the evaluation of three land surface models over the continental United States based on remotely sensed land surface temperature maps.

The aim of this study is to develop a remote sensing based actual evapotranspiration (AET) dataset suitable for validation and calibration of the spatial patterns of the national hydrological model of Denmark (DK-~~model). The model evaluation will be~~Model). The idea is to thoroughly investigate the observed differences in spatial patterns between observations and simulations in order to understanding the underlying processes that generate patterns in actual evapotranspiration. The model evaluation is based on a diagnostic approach inspired by the study of Schuurmans et al. (2003) who utilized satellite estimates to identify conceptual model errors in a small sub basin of the MetaSWAP model in the Netherlands. ~~The idea is to thoroughly investigate the observed differences in spatial patterns between observations and simulations in order to understand the underlying processes that generate patterns (Conradt et al., 2013). Subsequently, the~~This approach aims at identifying which parts of the model parametrization generates these differences in the spatial patterns. Later, in an attempt to increase the similarity between the observed and simulated pattern of ET some new inputs and parameter distribution schemes are generated and included in a modified version of the DK model. The response of these modifications in the modified DK model are later evaluated against both spatial patterns of evapotranspiration, stream discharge and ground water heads. Results show that newly gained insights can guide the development of a new parameterization scheme and calibration framework that can facilitate an improvement in the spatial model performance.

Formatted: Danish
Formatted: Danish
Formatted: Danish
Formatted: Danish
Field Code Changed
Formatted: English (U.K.)

2 Methods

In this study two approaches are undertaken to estimate spatial patterns of evapotranspiration (ET) at national scale (Denmark) (Fig. 1); the first is based on a ~~distributed hydrological model (MIKE SHE) (described in section 2.1) and a~~

~~second approach based on a~~ Two Source Energy Balance (TSEB) driven by remote sensing data (~~Section 2.~~section 2.1); and the second is based on a distributed hydrological model (DK-Model) (described in section 2.2). We acknowledge that both approaches are subject to uncertainties; however, the aim of this study is to evaluate the dominating spatial pattern across Denmark and to gain insights into which processes and variables generate these patterns. The evaluation will focus on the spatial pattern itself by neglecting differences in the absolute values of evapotranspiration.

In the last step of ~~the~~this study the current version of the DK-Model is modified by replacing the original root depth (RD), crop coefficient (Kc) and leaf area index (LAI) based on lookup tables and a land cover map by remote sensing based data which features detailed spatio-temporal information~~. A more detailed description is presented in (Section 2.2 and 2.3)~~

2.1 ~~Hydrological model~~

~~The National Water Resources Model (DK-model) (Højberg et al., 2013;Stisen et al., 2012) was developed at the Geological Survey of Denmark and Greenland in 1996 and updated several times (Henriksen et al., 2003;Højberg et al., 2013). The model is constructed within the hydrological model system MIKE SHE (Abbott et al., 1986).~~ ~~The model works at 500 m resolution and due to computational efficiency and differences in geology the DK-model was divided into 7 different domains that cover the entire country; however in this study only 6 of the 7 domains were selected covering approximately 98% of the country with and extension of 42.087 km^2. Domains used are presented in Fig. 1.~~

~~The model is based on a full 3-D finite difference groundwater module that is connected to a simplified two layer unsaturated zone module (Yan and Smith, 1994). Furthermore, the model was previously calibrated using 191 discharge stations and approximately 17.500 data entries of ground water head (GWH) (Stisen et al., 2012). The DK-Model has been extensively used in different applications with different objectives; assessment of climatic change (Karlsson et al., 2016) , water resources management (Henriksen et al., 2008), large scale nitrogen modelling (Windolf et al., 2011;Hansen et al., 2009;van der Keur et al., 2008) highlighting the importance of the spatial component of the model and its reliability.~~

~~2.2~~ TSEB setup ~~and remotely sensed derived inputs~~

2.1.1 TSEB theory

The Two Source Energy Balance Model (TSEB) proposed by Norman et al. (1995) is used to retrieve mean monthly maps of ET across Denmark. Although several global remote sensing based data sets of actual evapotranspiration, such as MODIS ET (Mu et al., 2007) and GLEAM (Miralles et al., 2011), are available, the TSEB model was selected as the most appropriate remote sensing algorithm for the current study. The main reason is that TSEB is mainly driven by land surface temperature, which is a key indicator of evaporative state at the surface (Kalma et al., 2008). This makes LST based ET algorithms more appropriate than purely vegetation based algorithms such as (Mu et al., 2007). Other models, such as

Formatted: English (U.S.)
Field Code Changed
Field Code Changed
Field Code Changed
Formatted: English (U.S.)
Formatted: English (U.S.)
Formatted: English (U.S.)

GLEAM (also not including LST), can be considered a fusion of models and remote sensing data (e.g. including a soil moisture model and plant water stress model) and as such become difficult to regard as an observation.

In our study we have incorporated the code which is provided by the pyTSEB package (https://github.com/hectornieto/pyTSEB last accessed 30/01/2017). The applied model is a two layer model that treats soil and vegetation separately and estimates fluxes on the basis of LST and air temperature (T_{Air}) among other input variables. ~~TSEB can successfully be applied with a single LST observation, if the viewing geometry of the observation is available (VZA). The method uses an iterative process of canopy temperature (Tc) and soil temperature (Ts) to find the temperatures that satisfy the energy balance (Norman et al., 1995). Several inputs to TSEB are directly obtained from the LST product~~As presented in Norman et al. (1995), and presented here in a very simplified way, the model is based on the energy balance equation:

$$LE[Wm^{-2}] = R_n - H - G \quad (1)$$

Where H is the sensible heat flux, G is the ground heat flux, R_n is the net radiation and LE is the latent heat flux all in Wm^{-2}. The sensible heat flux (H) is calculated as:

$$H[Wm^{-2}] = H_C + H_S = \rho C_P \left[\frac{T_C - T_A}{R_A} + \frac{T_S - T_A}{R_A + R_S}\right] \quad (2)$$

Where H_C and H_S are the sensible heat flux for the canopy and soil respectively, T_C and T_S are the canopy and soil temperatures (K), ρC_P is the volumetric heat capacity of air ($Jm^{-2}s^{-1}$), R_S is the resistance to heat ($s\ m^{-1}$) flow in the boundary layer above the soil surface and R_A is the aerodynamic resistance ($s\ m^{-1}$) expressed as:

$$R_A[sm^{-1}] = \frac{\left[ln\left(\frac{z_U - d}{z_M}\right) - \Psi_M\right]\left[ln\left(\frac{z_T - d}{z_M}\right) - \Psi_H\right]}{0.16U}, \quad (3)$$

where z_U and z_T are the height of the wind speed (U) and air temperature in meters, Ψ_M and Ψ_H are the adiabatic correction factors for momentum and heat, d is the displacement height ($d \approx 0.65\ h_c$, and h_c is the height of the canopy (m)), z_M is the displacement height for momentum ($z_M \approx h_c/8$).

The model starts with an iterative process in which it finds the T_C which satisfies the energy balance equation. The divergence of net radiation in the canopy (ΔR_n) is used to partition sensible and latent heat fluxes using the Priestley-Taylor approximation (Priestley and Taylor, 1972) for the green part of the canopy. The transpiration is given by the next eq.

$$LE_C[Wm^{-2}] = 1.3 f_g \frac{S}{S+\gamma} \Delta R_n, \quad (4)$$

Where f_g represents the fraction of LAI that is green, S is the slope of the saturation vapour versus temperature curve and γ is the psychrometric constant and where ΔR_n is calculated as:

$$\Delta R_n[Wm^{-2}] = R_n - R_n \exp(0.9 \ln(1 - f_c)) \quad (5)$$

And where f_c is calculated as:

$$f_c[-] = 1 - \exp\left(\frac{-0.5LAI}{\cos\theta}\right) \quad (6)$$

Where θ is the viewing angle.

The model iterates until the energy balance equations are satisfied for soil and canopy. The readers are referred to Norman et al. (1995) to find a fully description of the model for a more detailed explanation of how the iteration process is carried out.

2.1.2 Derived remote sensing inputs

Several inputs to TSEB are directly obtained from the Moderate Resolution Imaging Spectroradiometer (MODIS) sensor all at 1 km spatial resolution; day time LST and day time VZA obtained from MOD11A1 and MYD11A1 products ~~flown~~on board of on TERRA and AQUA satellites respectively. The decision of whether to use LST from TERRA or AQUA is based on the percentage of high quality pixels available covering Denmark in each scene. The quality flags included in the products ~~is~~are used to select only those pixels with the best observation possible, no cloud present, no cloud shadow etc. In addition only satellite observations obtained between11:00 to 13:00 local solar time are utilized to ensure minimum effect of acquisition time.

~~LAI is derived using an empirical relationship with the Normalized Difference Vegetation Index (NDVI) (Rouse et al., 1973). The~~This study focuses on the growing season from April to September. From a water resources perspective the spatial patterns of ET are regarded as irrelevant for the remaining months of the year due to energy limited conditions and very low potential evapotranspiration. First the Nadir BRDF (Bidirectional Reflectance Distribution Function) Adjusted Reflectance (NBAR) from the MCD43B4 product for that time period was used to calculate the NDVI ~~using the following equation Eq. 1:~~(Rouse et al., 1973) using the following equation (Eq. 7):

~~$$NDVI = \frac{B_2 - B_1}{B_2 + B_1}, \quad (1)$$~~

$$NDVI[-] = \frac{B_2 - B_1}{B_2 + B_1}, \quad (7)$$

Field Code Changed

where ~~Bx indicates~~B1 and B2 is the reflectance from bands 1 and 2 from MODIS ~~band number.~~(645.5 nm and 856.5 .nm respectively).

Later, the Savitzky-Golay filter (Savitzky and Golay, 1964) available in the TIMESAT code (Jönsson and Eklundh, 2002;Jönsson and Eklundh, 2004) is selected to smooth the NDVI time series as it preserves the maximum and minimum values of the original dataset and guarantees consistency of time series. In situ measurements of LAI are usually expensive and time consuming, therefore reference LAI was based on the tables used for the Danish National Water Resources model (Stisen et al., 2012) which are based on previous works of Refsgaard et al. (2011). Boegh et al. (2004) ~~successfully~~ derived LAI variability in Denmark using NDVI with an exponential function. In ~~the~~this study a similar approach where coefficients are adjusted to match the input LAI data ~~from~~used in the National Water Resources model ~~was applied~~ resulting in Eq. ~~2~~8:

~~$$LAI = \alpha \cdot e^{\beta \cdot NDVI} \quad (2)$$~~

$$LAI\left[\frac{m^2}{m^2}\right] = \alpha \cdot e^{\beta \cdot NDVI} \quad (8)$$

Where α and β are specific parameters for ~~the~~this study case~~.~~

~~Later, coefficients α~~ and ~~β were~~that is adjusted to ~~maximized~~maximize the fit between the ~~estimated LAI~~calculated NDVI and the reference LAI from the DK model, leading to Eq. ~~3~~9:

~~$LAI = 0.0633 \cdot e^{5.524 \cdot NDVI}$ (3)~~

$$LAI\left[\frac{m^2}{m^2}\right] = 0.0633 \cdot e^{5.524 \cdot NDVI} \quad (9)$$

Another parameter that is needed to run the TSEB is the Vegetation Height (VH~~) was~~). This is derived assuming a simple linear regression as in Stisen et al. (2011) following Eq. ~~4~~10 :

$$VH\ [m] = 0.1 \cdot \text{Height}_{\text{Max}} + 0.9 \cdot \text{Height}_{\text{Max}} \cdot \left(\frac{LAI}{LAI_{Max}} \frac{LAI_i}{LAI_{i,Max}}\right)$$

~~(4)~~ (10)

Where $\text{Height}_{\text{Max}}$ is a value that changes for each land use class~~.~~ , $LAI_{i,Max}$ indicates the maximum LAI value for a particular pixel. This relationship ~~was~~is applied on all land use classes except forest, which ~~was~~is set to a year round constant VH of 15m.

~~LAI_{Max} indicates the maximum LAI value for each class.~~

The input albedo data was obtained from the MODIS 8-day MCD43B3 product ~~and in~~using only good quality pixel according to the quality flag (MCD43B2). In order to further reduce noise, mean albedo maps are generated by creating 46 mean maps (at 8 day intervals) using all the scenes available for each 8-day ~~of the year (DOY)~~interval across different years ~~using only the highest quality pixels.~~

, The albedo mean maps for each 8 ~~DOY~~day period are later used ~~as a climatology for the different years processed~~to calculate the net radiation in TSEB.

Climate forcing data to run the TSEB are obtained from the ERA-Interim reanalysis data set (Berrisford et al., 2011;Dee et al., 2011) provided by the European Centre for Medium-Range Weather Forecast (ECMWF). We have incorporated 10 m horizontal wind speed ~~(10U),~~ 2 m air temperature ~~(2T),~~ surface solar radiation ~~(SSRD)~~ and longwave incoming radiation ~~(SLRD).~~

Fraction of Green vegetation was derived from LAI ~~by two separate equations for the period defined as 1) beginning of growing season to peak of growing season (the greening phase) and 2) the rest of the year. The equation used outside the greening phase (and for needle leaf forest all year) was defined as~~following Eq.~~5~~ 11:

~~$Fg_i = \frac{LAI_i - LAI_{i,min}}{LAI_{i,min}} \cdot 100$ (5~~

$$[-] = \frac{LAI_i}{LAI_{MaxClass}} \quad (11)$$

~~Where Fg~~ ~~indicates the Fraction of green for a certain pixel *i*, $LAI_{,min}$ indicates the minimum LAI value for a pixel i in the time series.~~

~~For the greening phase (start to peak of growing season) a different equation was used to describe the very quick increase in fraction of green vegetation that occurs right when the very green plants emerge in the spring. The equation used in that interval is Eq. 6:~~

Where Fg_i indicates the Fraction of green for a certain pixel i , LAI_i indicates the LAI value for a pixel i and $LAI_{MaxClass}$ is the maximum LAI value for an specific land cover type. The approach is a simplified form of the vegetation index based method by Gutman and Ignatov (1998). This equation was applied to needle leaf forest land cover type.

For the other land cover types (crops, grasslands, deciduous forest etc.) equation 11 was modified adding another term. These land cover types show a stronger seasonality and a clear distinction between a greening phase and a senescence phase. In order to represent the strong difference in fraction of green vegetation between the period before and after senescence we introduced a different equation for the period between crop emergence and senescence, where we assigned higher values of Fg to non-needle leaf forest land covers, Fig. 2. For these types of vegetation Fg will be allowed to increase rapidly just after crop emergence by substituting Eq. 11 by Eq. 12.

$$Fg_i = \frac{LAI_{i,max}}{LAI_{MaxClass}}[-] = \frac{LAI_{i,max}}{LAI_{MaxClass}} \cdot (1 - e^{(-2 \cdot LAI_i)}) \quad (12)$$

First the dates corresponding to the maximum LAI (Fig. 2) and minimum LAI are identified. Later the date where the LAI increases at least 20% above the minimum is identified as the breakpoint (Fig. 3). The break point is considered to be the start of the growing season and therefore the time when the Fg starts increasing rapidly. Between the breakpoint and maximum LAI equation 5 was used to force an abrupt change in the Fg value and make it rapidly increase to high values before decreasing after reaching a peak in LAI representing the onset of the senescent phase.Where $LAI_{i,max}$ indicates the maximum LAI value for a pixel i.

This substitution is only conducted during part of the phenological year, more specifically for the period starting at the green-up date, corresponding to the point defined by an increase in 20% in LAI compared to the winter low ($LAI_{i,Min}$) (Cong et al., 2012) and continuing until the time at which LAI reaches its maximum ($LAI_{i,max}$) (Fig. 2). This approach will mediate the shortcomings of the vegetation index based methods, which has been shown to underestimate fraction of green during the greening phase while corresponding well to field observations during senescence (Guzinski et al., 2013).

Formatted: Font color: Text 1

Finally, instantaneous estimates of ET are converted to daily ET values based on the assumption of a constant value of the evaporative fraction throughout the day (Sugita and Brutsaert, 1991;Brutsaert and Sugita, 1992) following Eq. 13 :

$$dET = [Wm^{-2}] = EF \cdot Rn \quad (13)$$

where dET represents the daily ET and EF is the Evaporative Fraction and Rn represents the daily Net Radiation.

And net radiation and where EF is calculated as in Eq.14:

$$EF = [-] = ET/Rn \quad (14)$$

Where ET is the actual evapotranspiration and ~~netRad~~Rn is the Net Radiation at the same acquisition time. The assumption of constant EF over the course of a day is often not completely true~~.~~ and also affects the estimates of daily ET (Gentine et al., 2007) but in the current application is not crucial since only the spatial pattern of the remote sensing estimates is utilized.

Later, the daily maps are aggregated to monthly mean maps by using only those days in which the national coverage exceeded 50%. The final monthly mean maps ~~consist of~~comprises all available ET estimates for a given month across all years (2001-2014) resulting in just ~~6~~six climatological maps (April-Sept~~.~~)

). This temporal aggregation process is conducted in the same way with simulations from the DK-Model~~.~~ considering only the same cloud free pixels/grids as from the TSEB estimate and ensuring the comparability of the maps.

2.1.3 Optimization of vegetation parameters

Data from three eddy covariance (EC) flux towers is used as a reference to perform a sensitivity analysis and calibration of some of the vegetation parameters of the TSEB. A detailed description of the instrumentation and data processing of each of the flux sites can be found in Ringgaard et al. (2011). Latent heat (LE) or evapotranspiration measurements ~~is~~are obtained at a frequency of 30 min and the mean value of the observations from 11:00 to 13:00 were used as reference in the calibration of TSEB for the cloud free days where the TSEB estimates are available. The observed eddy covariance data are subject to energy balance closure problems, typically in the order of 20-25%, which is ~~not unusual~~usual for this type of measurements (Hendricks Franssen et al., 2008).

The ~~observed values used during TSEB~~ evaluation ~~are~~of the TSEB was conducted using as reference the data of the EC systems from the 3 different land cover types that were corrected for the energy closure using the Bowen ratio approach (Bowen, 1926) ~~corrected values and the~~. The associated uncertainty ~~estimate~~ is the ~~span~~spam between all ~~error in the~~ closure ~~problem~~error being assigned to ~~the latent~~either sensible heat ~~and no error in the~~or to latent heat.

~~2.3 Remote sensing derived hydrological input data~~

In order to identify the input and parameters that assert the main control of the TSEB model a sensitivity analysis~~Part of the current study is to identify model inadequacies and test possible directions of model parametrization improvements. In an attempt to improve the initial DK-Model, spatially and temporally distributed root depth (RD) maps are generated using remote sensing data using a similar approach to Andersen et al. (2002) and Koch et al. (2017) where RD is calculated by Eq. 9:~~

~~$$RD_i = RD_{max} \frac{LAI_i}{LAI_{max}}, \quad (9)$$~~

~~where RDi is RD in pixel *i*, LAI_i is the LAI in that cell, and LAI_{max} and RD_{max} indicate the maximum values at *i*. This equation was modified and LAI was substituted by NDVI, leading to Eq. 10~~

~~$$RD_i = RD_{max} \frac{NDVI_i}{NDVI_{max}} \quad (10)$$~~

and used for forested land cover types, poorly vegetated areas and urban areas. RD can be considered an effective parameter in the DK-model which partly compensates for the lack of a specific vegetation (LAI) component in the evapotranspiration calculations. For instance, eq. 10 equips forest cells with a temporal varying RD which is contrary to our physical understanding, but it compensates for mixed land-use cells and undergrowth. RD of croplands and grasslands, denoted as RD_{Agri}, was estimated by implementing some modifications to equation 10. RD_{max} is known to vary depending on the soil type with shallower roots for sandy soils (Refsgaard et al., 2011) and the opposite for clay soils, therefore an empirical relation between the clay fraction (CF) in the soil and RD_{max} is established for the different soil types. This equation is then included as a substitute of RD_{max}. In order to allow RD to reach zero, the second term in equation 11 is normalized by including $NDVI_{min}$ in the Eq. 11,

$$RD_{(agri),i} = [(12 \cdot CF_i) + 0.2] \cdot \frac{NDVI_i - NDVI_{min}}{NDVI_{max} - NDVI_{min}} \quad (11)$$

Where CF_i indicates the clay fraction at pixel *i*, $NDVI_i$ indicates the NDVI value of pixel *i*, and $NDVI_{Min}$ and $NDVI_{Max}$ represent the minimum and maximum values of NDVI for the same pixels.

The constants of equation 10 are derived by matching the root depths of the original DK-model.

In addition, the crop coefficient (Kc), which is a correction factor for the reference evapotranspiration (ET_{ref}) accounting for the difference between a given crop or land surface and the reference crop on which the ET_{ref} is based. Kc is derived from remotely sensed LAI using the following Eq. 12:

$$KC = 0.95 + 0.2 * (1 - e^{(-0.7*LAI)}) \quad (12)$$

Again the constants of equation 11 are derived to match the Kc-values of the original DK-model.

2.4 Optimization of vegetation parameters

A sensitivity analysis of the TSEB model is conducted with the help of PEST (http://www.pesthomepage.org/Home.php), a model-independent parameter estimation and uncertainty analysis tool. PEST evaluates the sensitivity of each parameter by perturbing the value of each of the parameters one at a time and subsequently analyse the response of the performed perturbation with respect to a change in model performance. At last, the sensitivities are normalized using the most sensitive parameter as reference.

In order to optimize theLater an optimization of selected vegetation parameters, the is performed. The objective function is set to reduce the differences in mean monthly ET estimates at three EC measurements sites that represent the three main land cover types in DK (Agriculture, Forest and Meadow).

Only 4 parameters are calibrated using PEST. These parameters are: Priestley Taylor parameter for forested areas (PT_{Forest}), and the LAI max Class$LAI_{MaxClass}$ for the three land covers used to estimate the Fg ($LAI_{Max}^{Agriculture}$, LAI_{Max}^{Forest}, LAI_{Max}^{Meadow}).

~~2.5~~2.2 Hydrological model

The National Water Resources Model (DK-Model) (Højberg et al., 2013;Stisen et al., 2012) was developed at the Geological Survey of Denmark and Greenland in 1996 and updated several times (Henriksen et al., 2003;Højberg et al., 2013). The model is constructed within the hydrological model system MIKE-SHE (Abbott et al., 1986). The model works at 500 m resolution and due to computational efficiency and differences in geology the DK-Model was divided into 7 different domains that cover the entire country; however in this study only 6 of the 7 domains were selected covering approximately 98% of the country with and extension of 42.087 km^2. Domains used are presented in Fig. 1. The model is based on a full 3-D finite difference groundwater module that is connected to a simplified two–layer unsaturated zone module (Yan and Smith, 1994). Furthermore, the model was previously calibrated using 191 discharge stations and approximately 17.500 data entries of ground water head (GWH) (Stisen et al., 2012). The DK-Model has been extensively used in different applications with different objectives; assessment of climatic change (Karlsson et al., 2016) , water resources management (Henriksen et al., 2008), large scale nitrogen modelling (Windolf et al., 2011;Hansen et al., 2009;van der Keur et al., 2008) highlighting the importance of the spatial component of the model and its reliability.

Field Code Changed

Field Code Changed

Field Code Changed

2.2.1Remote sensing derived hydrological model input data

Part of the current study is to identify model inadequacies and test possible directions of model parametrization improvements. In an attempt to improve the initial DK-Model, spatially and temporally distributed root depth (RD) maps are generated using remote sensing data using a vegetation index based approach (Koch et al. (2017)) where RD is calculated by Eq. 15:

$$RD_i[m] = RD_{max} \frac{NDVI_i}{NDVI_{max}} \quad (15)$$

where RDi is root depth in pixel *i*, $NDVI_i$ is the NDVI in that cell, and $NDVI_{max}$ is the and RD_{max} indicate the maximum values at *i* in meters. This equation was used for forested land cover types, poorly vegetated areas and urban areas. RD can be considered an effective parameter in the DK-Model which partly compensates for the lack of a specific vegetation component in the evapotranspiration calculations, and therefore variability in LAI and phenology. For instance, Eq. 15 equips forest cells with a temporal varying RD which is contrary to our physical understanding, but it compensates for mixed land-use cells and undergrowth. RD of croplands and grasslands, denoted as RD_{Agri}, was estimated by implementing some modifications to Eq. 15. For Danish agricultural land covers, the effective maximum root depth (RD_{max}) is known to be lower for the very sandy soils in western Denmark (Refsgaard et al., 2011;Breuning Madsen and Platou, 1983). This dependency of maximum root depth on soil type is accounted for by a linear relation between the clay fraction (CF) in the soil and RD_{max}. This relation is then included as a substitute of RD_{max}. In order to allow RD to reach zero for croplands, the second term in Eq. 15 is normalized by including $NDVI_{min}$ in the Eq. 16,

$$RD_{(agri)i}[m] = [(\alpha_{RD} \cdot CF_i) + \beta_{RD}] \cdot \frac{NDVI_i - NDVI_{min}}{NDVI_{max} - NDVI_{min}} \quad (16)$$

Where CF_i indicates the clay fraction at pixel *i*, $NDVI_i$ indicates the NDVI value of pixel *i*, and $NDVI_{Min}$ and $NDVI_{Max}$ represent the minimum and maximum values of NDVI for the same pixels.

The constants α_{RD} and β_{RD} are considered calibration parameters that should be tuned to the best overall water balance and spatial pattern in AET. For the initial run of the modified DK-Model, values of α_{RD} = 12 and β_{RD} = 0.2 are assigned. These values are derived by matching the average root depth across all grids obtained through equation 16 with the corresponding average root depth of the original DK-Model.

In addition, the crop coefficient (Kc), which is a correction factor for the reference evapotranspiration (ET_{ref}) accounting for the difference between a given crop or land surface and the reference crop on which the ET_{ref} is based. Kc is derived from remotely sensed LAI using the approach presented in Allen et al. (1998) and used by Stisen et al. (2008):

$$Kc[-] = Kc_{c,min} + \left(Kc_{c,max} - Kc_{c,min}\right) \cdot \left(1 - e^{(-0.7 \cdot LAI)}\right) = 0.95 + 0.2 * (1 - e^{(-0.7 * LAI)}) \quad (17)$$

Where the Kc_{min} and Kc_{max} are set to 0.95 and 1.15 respctively.

2.3 Spatial pattern analysis; Empirical Orthogonal Functions (EOF)

The Empirical-Orthogonal-Functions (EOF) analysis is a statistical technique commonly used to evaluate large spatio-temporal datasets of hydrological states and fluxes (Mascaro et al., 2015;Perry and Niemann, 2007;Graf et al., 2014). EOF decomposes the variability of a spatio-temporal dataset in two main components. First, a set of orthogonal spatial patterns (EOFs) which are time invariant and define statistically significant patterns of covariation. Second, a set of loadings that are time variant and specifying the importance of each EOF over time. Graf et al. (2014);Perry and Niemann (2007) described briefly the mathematical background of the EOF analysis. Most commonly, the EOF analysis is applied on either observational or modelled datasets, but recent applications stressed it's usability as a tool for spatial validation of distributed hydrological models at catchment scale (Fang et al., 2015;Koch et al., 2016). Koch et al. (2015) suggested performing a joint EOF analysis on an integral data matrix that contains both, observed and simulated data. In this way, the resulting EOF maps honour the spatio-temporal variability of both datasets and the weighted difference between the loadings at specific times can be utilized to derive a quantitative pattern similarity score. The weighting is necessary, because the amount of covariation that lies in each EOF differs, where the first EOF is oriented in the direction of maximum covariation. Therefore, the EOF based similarity score (SEOF) between an observed and a predicted ET map at time x can be formulated as ~~(~~Eq. ~~13):~~18:

$$S_{EOF}^{x} = \sum_{i=1}^{n} w_i \left|\left(load_i^{simx} - load_i^{obsx}\right)\right| \quad (~~13~~18)$$

where w_i, the covariation contribution of the i'th EOF, is multiplied with the absolute difference between the simulated loading ($load^{sim}$) and the observed loading ($load^{obs}$) of the i'th EOF at time x.

The EOF analysis applied in this study evaluated the differences in spatial patterns between the ~~MIKE SHE~~DK model outputs in the original configuration and a modified version where three inputs (RD, LAI and Kc) of the model were replaced by those derived from remote sensing data.

3 Results and discussion

The results and discussion are presented in two sections; the first focuses on the sensitivity analysis and parameter optimization of the TSEB model and the second features the spatial pattern evaluation of the DK-~~model~~Model using the maps obtained from the TSEB model.

3.1 Sensitivity analysis and TSEB calibration

The normalized sensitivity values of the 23 incorporated variables and parameters are illustrated in Fig. 3. The results are presented in three groups depending on the group they belong to: remote sensing data, forcing data and vegetation parameters.

The results show that the most sensitive ~~parameter~~variable for the estimation of AET is LST. Interpreting the sensitivity values for each group individually stress that, for the remote sensing input, parameters that are directly related to LST such as emissivity of vegetation (EmissV) and soil (EmissS) are characterized by a high sensitivity as well. The next group, forcing data, exhibited high sensitivity ~~in~~for all ~~parameters~~variables, except for wind speed. Overall Air temperature (TempAir) is the most sensitive forcing variable. These results indicate that the algorithm is largely controlled by the LST, LAI and climate forcing data. This finding is considered ideal, since the actual parameters of the algorithm do not dominate the final spatial pattern. In general, the remote sensing and forcing data inputs can be considered observations which are not subject to calibration.

~~The least sensitive values are found among the vegetation parameters and some parameters in this group are calibrated to optimize the match to each of the three land cover types. Only minor adjustments of the estimated ET are made by calibrating the PT_{Forest}, $LAI_{Max}^{Agriculture}$, LAI_{Max}^{Forest}, LAI_{Max}^{Meadow}.~~

The sensitivity analysis was utilized for illustrating the main controlling variables of the TSEB algorithm, not for selection of calibration parameters. Subjectively, four vegetation related parameters were selected for calibration to optimize the match to each of the three land cover types. First, the PT_{Forest} parameter was selected because the Priestly-taylor coefficient of forests is believed to be below the standard value of 1.28 assigned for agriculture and meadow (Komatsu, 2005). Secondly, the $LAI_{Maxclass}$values ($LAI_{Max}^{Agriculture}$, LAI_{Max}^{Forest}, LAI_{Max}^{Meadow}.) which control the fraction of green vegetation (Fg) through Eq. 11 and 12 are selected for calibration.

The results of monthly ET estimates are presented for all three sites in Fig. 4. The bars on the observed values indicate the uncertainty associated with the energy balance closure issues where the upper bound of the uncertainty bar represent the situation in which the residual energy is assigned to the latent heat (LE), whilst the lower bound represent the opposite situation in which all the residual energy is assigned to the sensible heat (H).

Generally the estimated ET values agree well with the EC-measurements especially considering the uncertainties associated with energy balance closure and the spatial scale mismatch between the EC footprint and the remote sensing estimations. In order to minimize the effect of scale issues the EC values of ET at the three sites are compared to the average ET of the

Formatted: Heading 2

surrounding pixels estimated by TSEB. For this comparison, pixels that are considered as purely representative of the specific land cover type and therefore not contaminated by other land cover types are used. The selection of the pixels is carried out manually with the help of a high resolution image of the study area.

The comparison is meant as an illustration of the ability of the TSEB to describe the general annual variation and differentiate between land covers. The main aim of the TSEB application is to get robust national maps of growing season ET and the results show agreement on both, the seasonal variation and absolute levels of ET. On the other hand the separation between land covers is somewhat harder to evaluate because all three sites exhibit a similar level of ET.

ET in the forested areas remains mostly constant during the growing season with a tendency to increase at the end. Agricultural areas on the other hand presented much higher variability with a rapid increase at the beginning of the growing season (May-June) and a decrease at the end (August-September). The Wetland shows a similar shape as the forest but with slightly higher ET values, and presents a big increase in the month of August that is not capture in the TSEB.

Mean monthly maps of ET are generated from daily TSEB estimations across all years to ensure consistent spatial patterns for robust spatial model evaluation with the aim to evaluate and improve the model performance. Such an improvement can be facilitated through optimal parametrization, and we therefore focus on the consistent spatial patterns rather than the temporal dynamics of ET variability. This is also reflected in the way the TSEB ET estimates are evaluated.

Results indicate that the TSEB ET estimates are within the measurement uncertainty of the EC at the three stations. The only pronounced disagreement is observed in the wetland during the month of August. Ringgaard et al. (2011) showed how the water level of the Skjern River raised during that time of the year and therefore increasing the values of ET in the EC measurements which is located at the bank of the river.

3.2 Spatial patterns

The mean monthly maps of cloud free ET generated with the TSEB model and DK-Model are presented in Fig. 5 and Fig. 6. ~~The TSEB ET (Fig. 56~~For a better visualization of the spatial patterns the maps were normalized in this case by dividing each map by the mean value of the map itself. The TSEB ET (Fig. 5) exhibits a clear difference between Eastern and Western Denmark with lower ET values in the sandy Western Denmark especially in the peak of the growing season (May-June). The clear E-W pattern ~~identifies~~identified by the ~~RSEB~~TSEB model is remarkable considering that it is opposite the general precipitation gradient (Fig 1). This highlights the strong influence of soil properties on the ET pattern across Denmark. Another feature is that forest areas have lower ET for the selected cloud free days where canopy interception is not included.

Regarding the results from ~~MIKE SHE~~the DK model (Fig. 6) it can be observed that the E-W trend is not noticeable in the maps and the difference between forest and agriculture is less distinguishable. Moreover the effects of the zonal calibration are causing ~~the area of modelling~~ differences in model domain 2 (Fig. 1) ~~to~~which have much higher ET in comparison to the other domains, especially in May and June. ~~In June the pronounced difference between parametrization in domains 4 and 5 is also clearly evident.~~ From ~~figures~~Figs. 5 and 6 it can be extracted that there is almost no resemblance between the spatial

Formatted: Heading 2

patterns ~~of~~identified in the TSEB ET and the DK-~~model~~Model simulations on the national scale. This seems substantial since the model has been calibrated extensively. However, the applied discharge based calibration is dominated by the winter peak runoff, which conveys little information with respect to the spatial patterns of summer ET. This finding actually highlights the need for spatial pattern evaluation of distributed hydrological models since traditional discharge and groundwater head calibration does not ~~secure~~necessarily lead to reasonable ET patterns.

In a first attempt to improve the simulated spatial patterns of the DK-~~model~~Model, new parameterizations of ~~root depth (~~RD~~),~~ LAI and ~~crop coefficient (~~Kc~~)~~ are prepared based on fully distributed remote sensing and soil data as explained in section 2.~~3~~2.1. These contain a higher degree of spatio-temporal detail than the original model input based on predefined tables ~~(Refsgaard et al., 2011)~~from Refsgaard et al. (2011) and should reflect distributions that are more realistic and spatially consistent. Fig. 7 shows the modified DK-Model mean maps of ET. The patterns of these maps are more similar to those observed with the TSEB in which the E-W pattern is quite evident, although this pattern seems to be exaggerated.

Fig. 8 shows the spatial differences in growing season average ET between the DK-Model in its original (r = 0.07) configuration and the modified version (r = 0.33) based on remote sensing input. It is important to highlight at this point that the modified DK-Model is not recalibrated with the new inputs as this goes beyond the scope of this study. A recalibration may modify the water balance in comparison to the original setup. ~~-~~However the performed modifications show some relevant features; the most noticeable visual improvement is the much larger gradient in the East-West pattern obtained in the modified DK-~~model~~Model, which emphasizes the more distinct resemblance to the pattern, estimated with TSEB~~. Secondly~~ which also translated in an improvement in the ~~abrupt changes between~~Pearson coefficient from 0.07 to 0.33. Visually this improvement can be attributed to a more clear East-West pattern and smoother transition in the ~~different model~~ values of domains ~~of~~1,2 and 3 in the ~~DK-Model are removed due to a more spatially consistent parameterization. This is especially~~modified version compared to the ~~case for model domain 2 and the boundary between domain 5 and 6.~~ original DK model.

To analyse the driving mechanism behind the "observed" TSEB patterns and the simulated DK-~~model~~Model patterns the clay fraction used as input to the root depth calculation of the modified DK-Model, the observed average LST input to the TSEB model and the growing season average LAI are illustrated in Fig. ~~9.~~ 9. The similarities between LST and clay fraction maps with TSEB are quite evident (r= -0.50 and r=0.44 respectively) whilst the similarity with LAI is low (r = -0.15).

These maps reveal interesting findings; first the presence of the East-West pattern in the clay fraction map coincides visually quite well with the TSEB model mean outputs (Fig. 8) in spite of the fact that no soil information has been included in the TSEB ET estimation. This indicates that the general perception, of lower ET for the sandy soils in the West due to soil moisture stress in the summer period, is captured well by the TSEB ET. The East-West pattern is not captured by the DK-~~model~~Model simulation even though the model is based on soil type information on field capacity and wilting point. On the other hand the modified DK-~~model~~Model captures much more of the East-West pattern because the clay fraction information is utilized to stretch the root depth distribution. Moreover, the TSEB pattern is mainly controlled by the LST input combined with a fine scale variability introduced by the LAI patterns (Fig. 9).

Fig. 10 underlines that the changes in the original setup of the DK-Model and the modified version are large when compared using scatter plots using the mean normalized map of TSEB as reference. Even the dispersion in the scatter plots is large, the results reveal an improvement in the Pearson correlation coefficient (from r= 0.07 to r=0.33) and also the points move closer to the 1:1 line.

The EOF analysis (Fig. 11) extracts the spatio-temporal similarities and dissimilarities between the two different DK-Model configurations. The analysis is based on monthly mean maps generated using the daily simulations. Only the first three EOFs which, in combination, explain 71% of the total variance are presented. The first EOF captures 45.2% variance and the EOF loadings present very small differences and are equipped with positive sign throughout the entire period. Hence it can be interpreted that EOF 1 addresses the major similarities between the two model configurations. The EOF1 map captured the component of the ET pattern which is driven by the soil properties, as it relates nicely with the mapped clay content in Fig. ~~10~~9. The loadings of the second EOF in combination with its map add 15.7% to the explained variance. The apparent disagreement in values and sign between the loadings stressed that EOF2 captures the major dissimilarities between the two model configurations. The EOF2 pattern resembles the one found in EOF1, however it was characterized by less contrast and overall, it represents the added spatial detail of RD which is defined as a function of clay content and vegetation. The evident E-W trend is strongest in the first three months of the growing season and afterwards the loadings drop to close to zero for the modified DK-~~model~~Model. The third EOF explains around 10% of variance and further records dissimilarities between the two models. Examining the loadings stresses that the modified DK-~~model~~Model plays a minor role in EOF3, as loadings are close to zero. However the first three months of the original DK-~~model~~Model seems well represented and the map underlines the granularity of the original setup, which is strongly driven by the discrete land-use map. Also, the model boundary between area 5 and 6 appears in EOF3, which was caused by the zonal calibration of RD in the original DK-~~model~~Model. The overall similarity scores derived by the EOF analysis presented the maximum value for a pattern comparison in June (≈0.11) and the minimum corresponded to a day in April (≈0.02).

The results highlight a soil properties driven spatial pattern which is expected due to larger water holding capacity in clay dominated areas. This relationship is clearly evident in the TSEB data, although soil data do not drive the TSEB algorithm, but this information is embedded in the LST as LST can be used to map soil textures (Wang et al., 2015). In contrast the original DK-~~model~~Model includes soil type information, but clearly the soil parametrization does not have sufficient effect on the simulated patterns of ET. The spatial patterns can probably be improved through calibration, by increasing the contrast in soil parametrizations or by modifying the model formulations on the soil stress function. In the current study, a new root depth parametrization is applied where the spatio-temporal variation in the effective root depth is estimated based on a combination of the clay fraction map and remotely sensed NDVI time series. The simulated ET maps resulting from the new RS-based DK-~~model~~Model (including root depth Kc and LAI) are clearly much more similar to the TSEB ~~estimates,~~(Fig. 10) although significant differences still occur in some regions. In order to achieve a better performance the transfer-function of the root depth and Kc parametrizations have to be calibrated against the spatial patterns of the TSEB (in combination with discharge and groundwater head). Unfortunately, re-calibration of the National DK-~~model~~Model goes

beyond the scope of the current study, since a single model run of the entire DK-~~model~~Model requires around 40 hours (wall-clock time), but re-calibration will be part of future improvements of the model. This will ensure both spatially consistent parameterizations by utilizing transfer functions inspired by the parameterization scheme of ~~The~~the mesoscale Hydrologic Model (mHM) (Samaniego et al., 2010) and an optimal trade-off between discharge, groundwater head and spatial patterns of ET.

The applied EOF analysis identifies the spatio-temporal similarities and dissimilarities between the two DK-Model configurations. It allows pointing out driving mechanisms behind the simulated spatial patterns, such as the effect of the effective RD in the modified DK-Model in EOF2 or the sharp boundary of simulated ET caused by the zonal calibration of the original DK-Model (EOF3). These findings strengthened the EOF analysis as a suitable tool to meaningfully compare spatial patterns and to diagnose spatial model deficiencies. Recently, the proposed approach was applied by Koch et al. (2017) and Ruiz-Pérez et al. (2016) in a spatial sensitivity analysis and in a spatial pattern oriented model calibration, respectively. In the future the EOF analysis will be considered as a metric to re-calibrate the DK-Model with focus on spatial patterns of ET.

~~Two~~3.2.1 Key differences between the models

The two different approaches to retrieve ET are compared based on the ~~hypothesis~~idea that both models, TSEB and the DK-~~model~~Model should present similar spatial pattern of ET. Results showed that the differences in the outputs were noticeable. These differences can be divided in three groups.

- Differences due to model setup: The DK-~~model~~Model is an aggregate of 6 domains. Each of these domains is calibrated individually, which leads to inconsistent spatial distributions of hydrological properties across domains. This increases the accuracy and performance of the model when evaluated only using discharge stations and ground water heads, but ignores the spatial component as it is an aggregated evaluation. On the contrary, when the ET maps are obtained with TSEB these problems are not present as all the study area is treated the same way and with the same parameterization.
- Spatial differences due to the land cover parameterizations are also important and are clearly evident in the case of the TSEB maps when comparing forest and agriculture areas. In this study three different EC datasets are used to calibrate the vegetation parameters and these sites are assumed to be representative of each land cover at a national scale. In some cases this assumption might not be adequate as soil type/ forest type and other variables might affect the plant response to ET. The TSEB was adjusted to show this pattern which might in some cases be overestimated and therefore enhancing the contrast of the TSEB between two land cover types.
- Differences due to the models: Estimations of ET in the DK-Model are mainly driven by precipitation, root depth and soil properties and represent a residual in the water balance equation. On the other hand the TSEB relies mostly on forcing data and LST to estimate ET as a residual in the energy balance equation and does not take into account any soil information or rainfall. The similar features found between the mean annual maps of TSEB and clay fraction may indicate that the LST, and thereby TSEB, is sensitive to some soil properties. On the other hand these similarities

between LST and soil property patterns can also be explained by the fact that areas with sandy soils and low clay fraction are coincident with areas with lower agricultural production and higher risk of summer drought and vegetation under soil water stress.

43.3 Stream discharge and groundwater head performance

Besides comparing the spatial patterns of the original and modified DK-Model, the stream discharge and groundwater head performance is also compared. In this comparison it is important to acknowledge that the original DK- model has been calibrated against these variables, whereas the evaluation of the modified DK-Model has to be considered as a validation. Results showing the annual water balance error (WBE) and summer (Jun-Jul-Aug) water balance as well as NSE (Nash-Sutcliffe Efficiency) for 181 discharge stations are presented in Fig. 12. The first noticeable thing that can be concluded is that the average water balance error changes from a slight overestimation to a moderate underestimation (Median WBE changes from -5.5 % to 5.5% for the original and modified models respectively). Regarding, the summer water balance which is expected to be influenced the most by the model modifications; the picture is similar although the performance get worse with a larger positive bias. The NSE showed a decrease in performance, from NSE= 0.72 in the original DK model to NSE=0.67 in the modified version.

Ground water heads were also evaluated for 25,365 wells across the country and results are shown in Fig. 13. The results in this case are very similar between the original version and the modified one. Statistics showed a RMSE of 5.5 m in both cases, with the RMSE being dominated by relatively few very large errors while 78 % of the wells have absolute errors below 5 m. The similarity in simulated groundwater heads between the two model versions indicates that the changes in evapotranspiration patterns have little effect. However, it has to be considered that the simulated groundwater head is controlled by mainly hydraulic conductivity (which does not change between the two versions) and annual recharge upstream of the point of comparison. Since the changes in evapotranspiration patterns mainly effects the summer period, where recharge is low, the effect on annual recharge is limited. In addition, the changes in evapotranspiration patterns will redistribute recharge patterns, but the combined effect of that at some deeper well filter location will be a mixed signal causing limited changes in groundwater head.

The results of this comparison are promising considering that the model was not re-calibrated with the new inputs. In the future, the model will be recalibrated including a spatial metric as an objective function during the calibration, and it is believed that especially the model bias on discharge can be minimized.

Conclusions

In this study the potential of remote sensing outputs to evaluate spatial patterns of hydrological models has been shown. The information derived from remote sensing data can be used as a diagnostic tool for revealing model structural insufficiencies and inconsistencies. Additionally, remote sensing derived variables can be used in hydrological models and hence adding

spatial information that is finally translated to the outputs of the models. The use of spatial metrics is beneficial to identify spatial model deficiencies. Furthermore such metrics are required for a spatial pattern oriented model calibration in order to meaningfully compare the changes in the spatial patterns.

Hydrological model evaluations have traditionally focussed on the temporal dynamics of the outputs and not so much on the spatial component. This study shows that more attention must be given to the spatial pattern evaluation as traditional calibration does not ensure a realistic spatially representation. If the spatial component of the model is neglected, the use of distributed hydrological models is not always meaningful and therefore the use of more simple models could be more appropriate.

Model re-calibration should focus on a combination of improved parameter regionalization including clay fraction and other derived variables using remote sensing data, spatially countrywide consistency in parametrization and inclusion of dedicated spatial pattern oriented objective functions in combination with discharge and groundwater head observations.

This study was conducted over an energy limited region and over a specific time of the year where ET pays a more important role in the water cycle. Extending ~~the~~this study to other areas with different ecosystems that combine energy and water limited ecosystems will provide a wider overview on the factors controlling the ET spatial patterns.

5 Acknowledgments

The present work has been carried out under the SPACE (SPAtial Calibration and Evaluation in distributed hydrological modelling using satellite remote sensing data) project; The SPACE project is funded by the Villum Foundation (http://villumfonden.dk/) through their Young Investigator Programme.

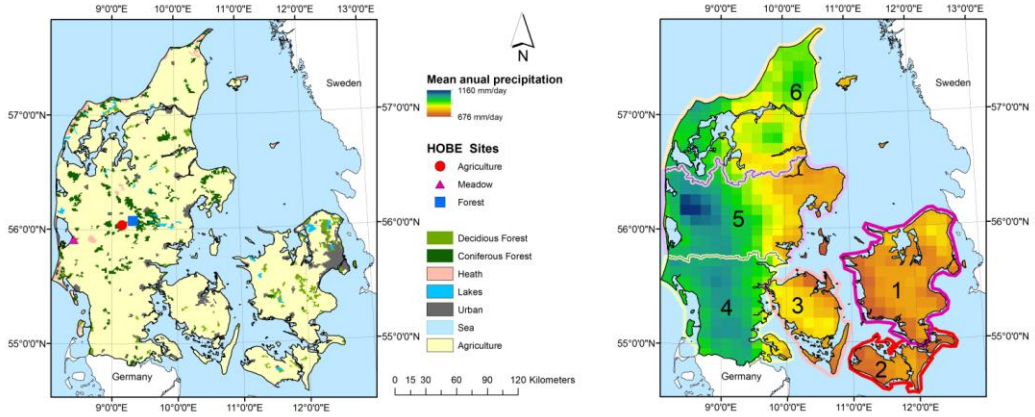


Figure 1. Presents the study domain. Left map: **Land cover map and model domains of the National Water Research Model of Denmark (Bornholm Island excluded in the figure.). Right map: National Water Research Model of Denmark domains and mean annual precipitation.**

Formatted: Caption, Line spacing: single

Formatted: Font: Not Bold

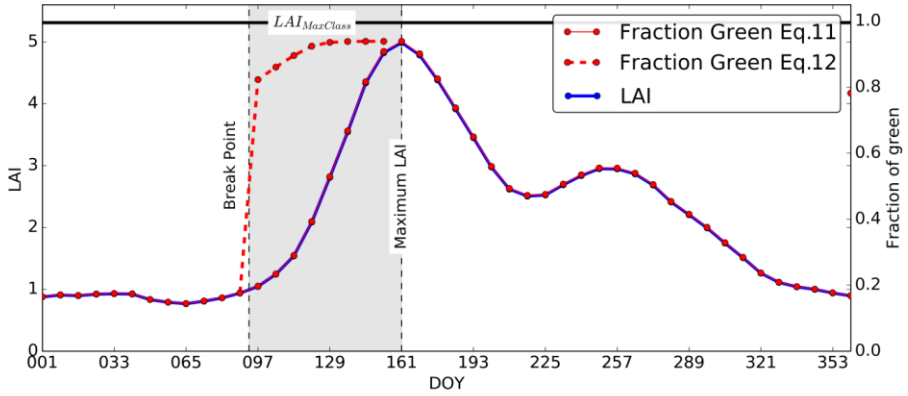


Figure 2. Diagram of Fraction of Green (Fg) calculation based on the leaf area index (LAI). LAI corresponds to the left ordinate and Fg to the right one. Shadow highlights the region in time were the Eq.11 is replaced by Eq.12. Data presented corresponds to an ~~Agricultural~~agricultural pixel ~~(row 100, coloum 84)~~ from the dataset.

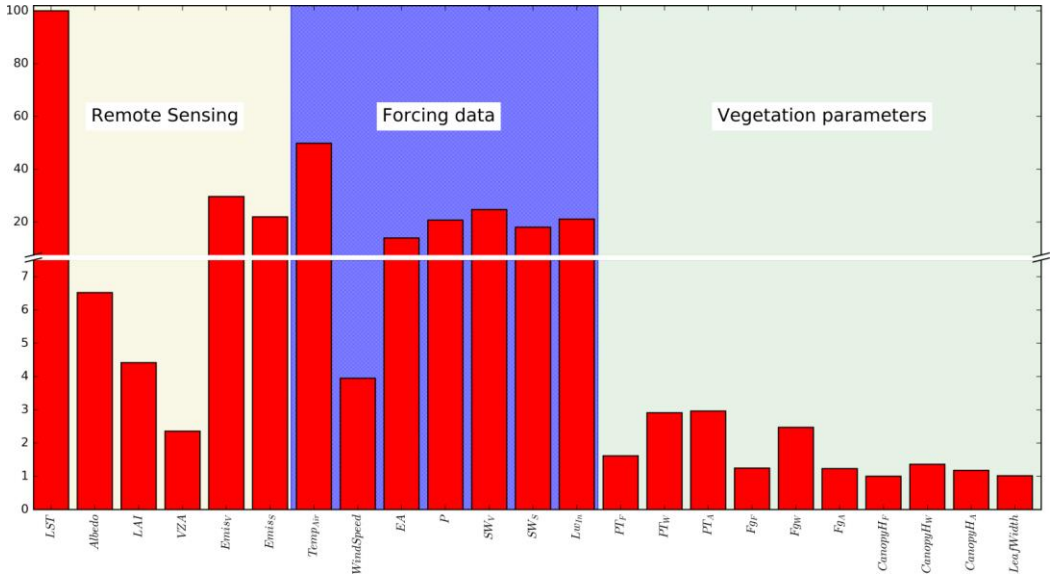


Figure 3. Sensitivity of 28 TSEB model inputs obtained with PEST. Results are normalized using the most sensitive as reference. Acronims used: LST (Land Surface Temperature), LAI (Leaf Area Index), VZA (View Zenithal Angle), $Emiss_V$ (Emissivity of Vegetation), $Emiss_S$ (Emissivity of Soil), ~~$Temp_{Air}$~~ T_A (Air Temperature), EA (Water Vapor pressure above canopy), P (Atmospheric pressure), SW_V (Short wave incoming radiation for vegetation), SW_S(Short wave incoming radiation for soil), LW_{In} (Long wave incoming radiation), PT_F (Pristley Taylor parameter for Forest), PT_W (Pristley Taylor parameter for meadow), PT_A (Pristley Taylor parameter for Agriculture), Fg_F (Fraction of green vegetation for forest), Fg_W (Fraction of green vegetation for meador), Fg_A (Fraction of green vegetation for Agriculture), $CanopyH_F$ (Canopy height for forest), $CanopyH_W$ (Canopy height for meadow), $CanopyH_A$ (Canopy height for Agriculture).

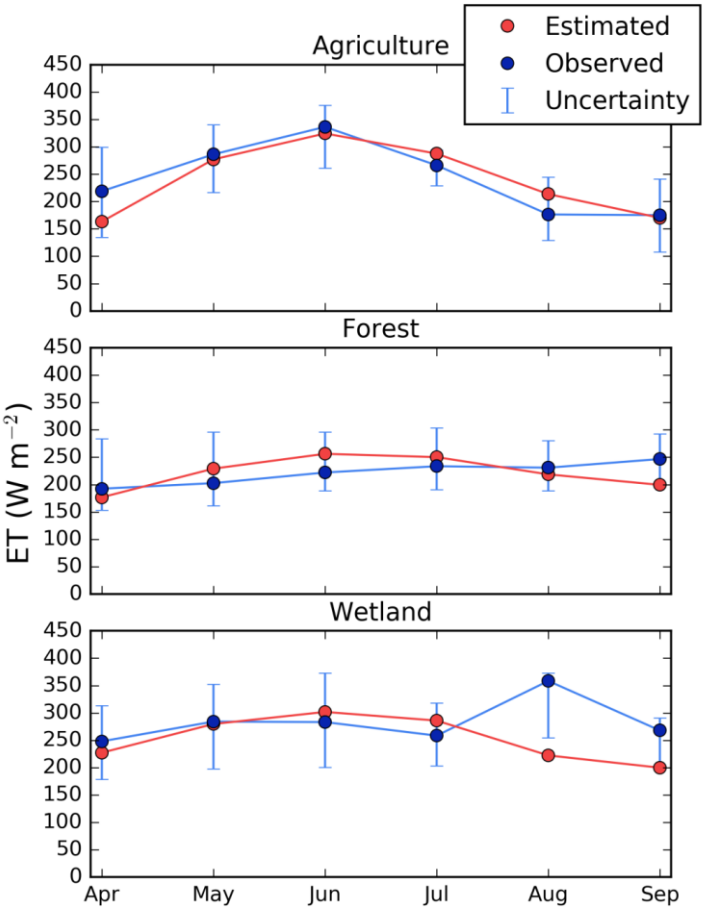


Figure 4. Comparison of TSEB ET estimates in different land cover types. Uncertainty bars limits represent two situations, the upper in which all the residual is assigned to the latent heat and the lower one in which the residual is assigned to the sensible heat flux.

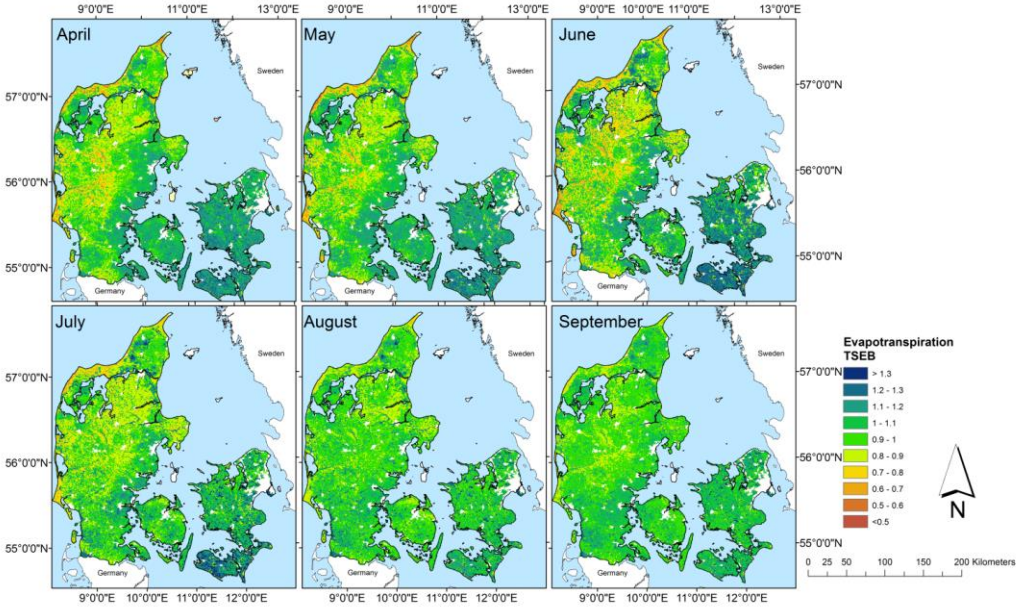


Figure 5. Mean normalized monthly TSEB ET maps [-] in which urban areas have been masked out.

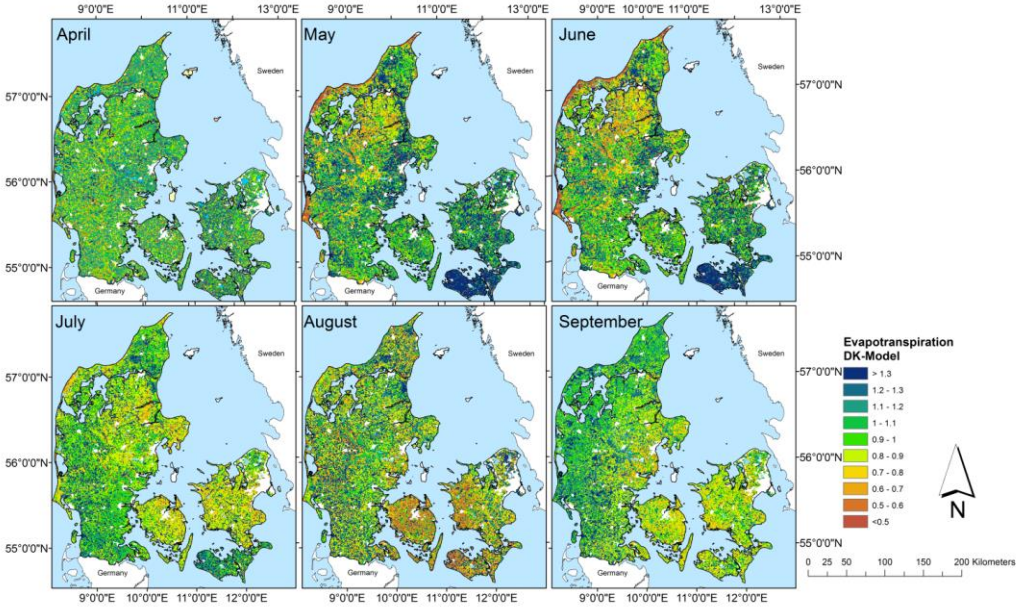


Figure 6. Mean normalized monthly DK-Model ET maps [-] in which urban areas have been masked out.

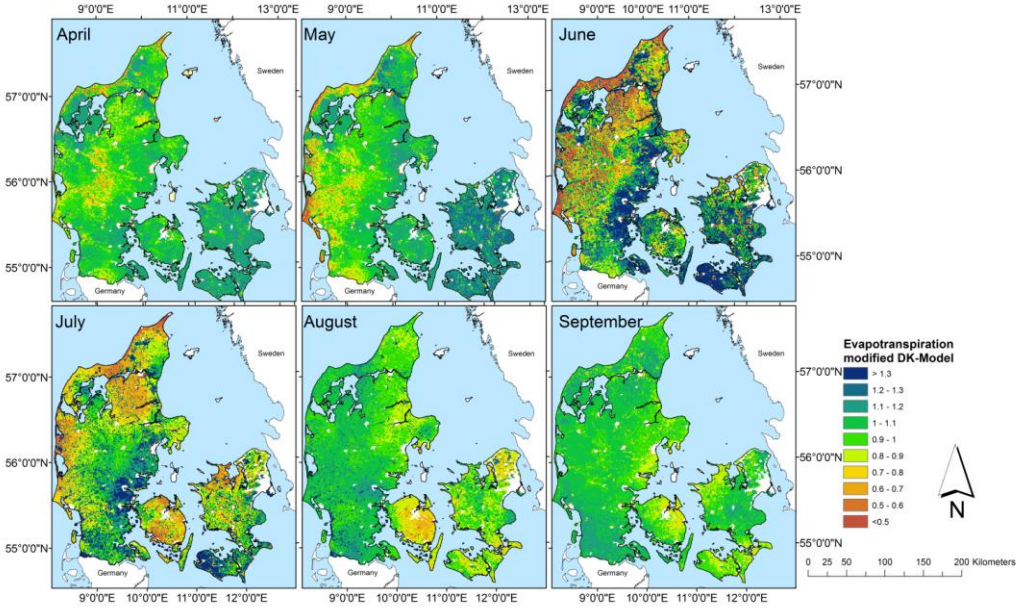


Figure 7. ~~Modified~~Mean normalized monthly modified DK-Model ET mean maps [-] in which urban areas have been masked out.

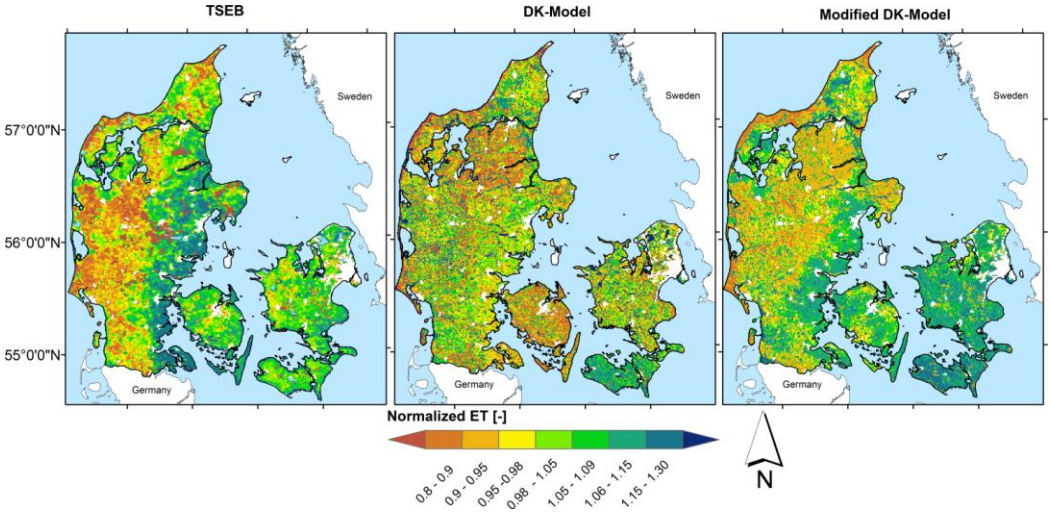


Figure 8. Normalized growing season maps for the TSEB , DK-Model and the modified version of the DK-Model.Normalization was conducted dividing the mean map by the mean value of the map.

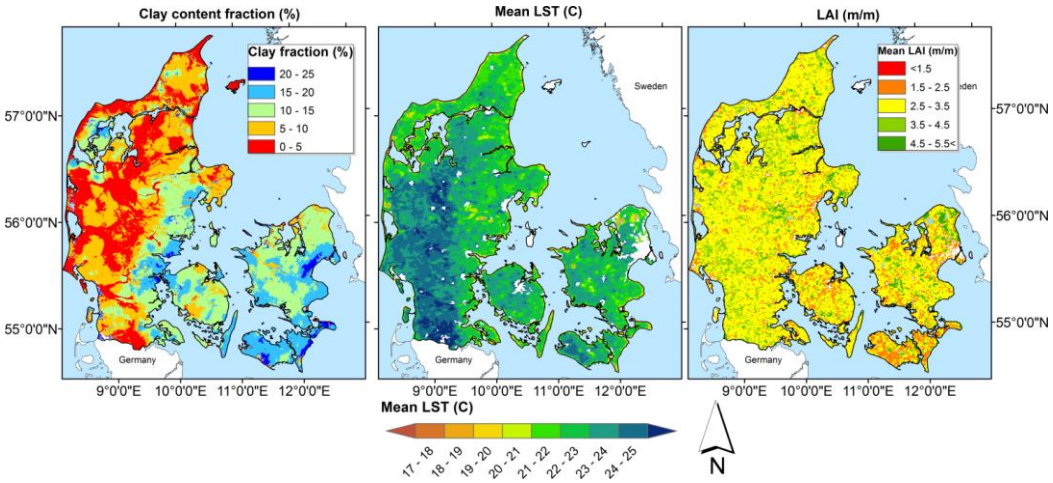


Figure 9. Maps of three different parameters used in this study. Left map shows the clay fraction distribution. Center map displays the mean values of LST during the growing season and right map displays the mean values of LAI during thre growing season and used in the TSEB.

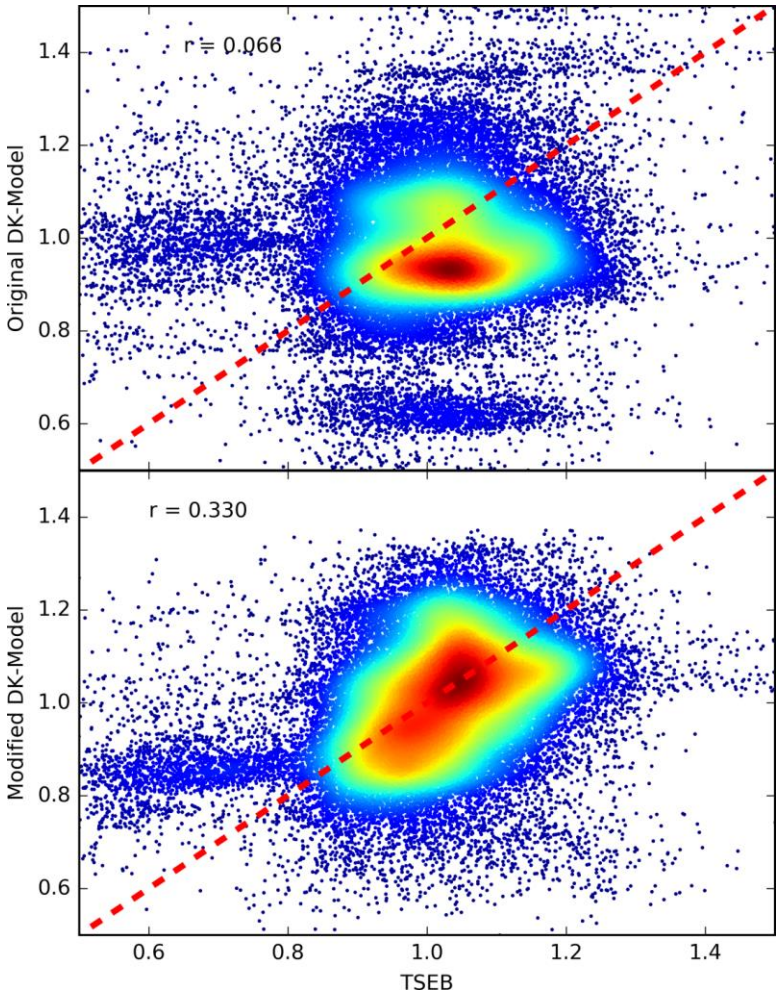


Figure 10. Scatter plots of normalized ET showing the comparison of the TSEB against the two DK-~~Molde~~Model configurations. Upper scatter plot compares the TSEB against the original DK-Model, and lower compares against the modified DK-Model version.

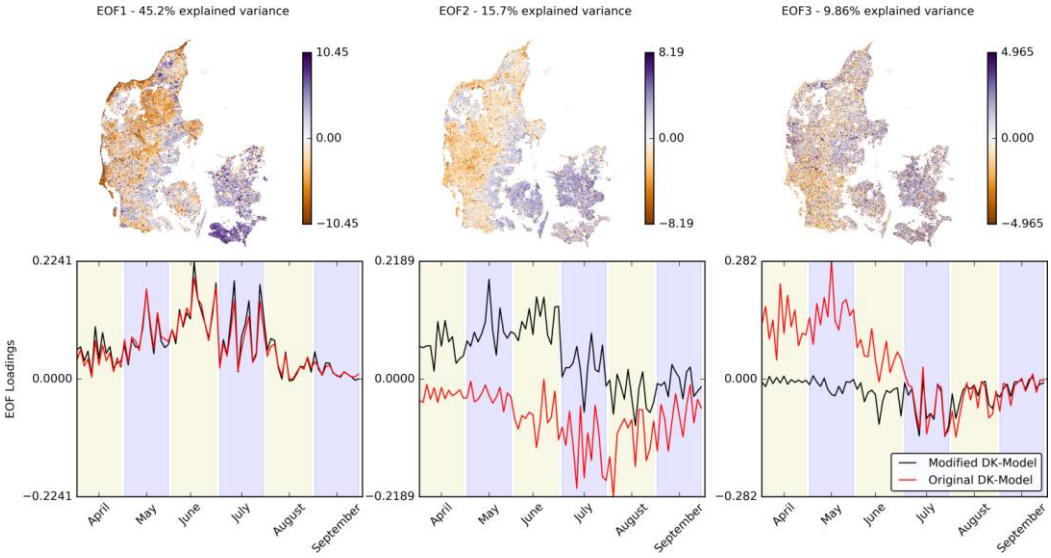


Figure 11. Maps of the first three EOFs comparing the original setup of the ~~MIKE SHE~~DK model model and the modified version of it.

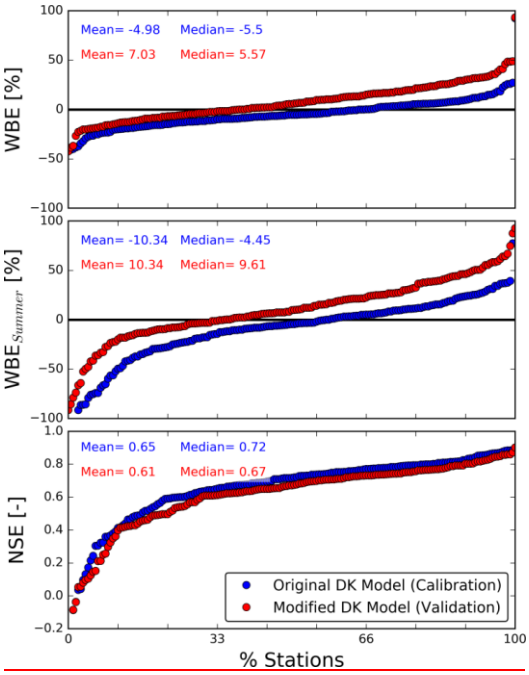


Figure 12. Model performance statistics showing the ~~model.~~results of the original and modified versions of the DK model. Stations have been ranked by performance and presented in the horizontal axis as a percentage. Figure shows the Nash-Sutcliffe Efficiency (NSE), water balance error (WBE), and water balance error only for the summer period (WBE_{Summer})

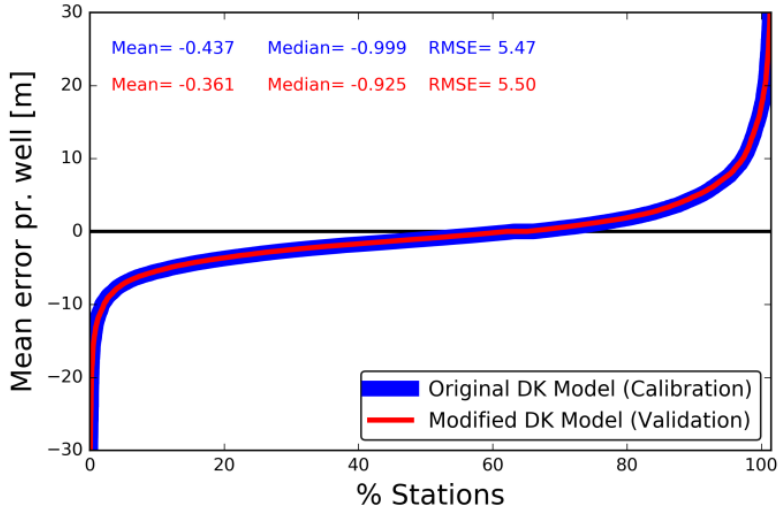


Figure 13. The figure shows the error and statistics in the ground water heads estimated from the original DK model and the modified version of the DK model. Stations have been ranked by performance and presented in the horizontal axis as a percentage.

Formatted: Normal

References

Abbott, M. B., Bathurst, J. C., Cunge, J. A., O'Connell, P. E., and Rasmussen, J.: An introduction to the European Hydrological System — Systeme Hydrologique Europeen, "SHE", 1: History and philosophy of a physically-based, distributed modelling system, Journal of Hydrology, 87, 45-59, http://dx.doi.org/10.1016/0022-1694(86)90114-9, 1986.

~~Andersen, J., Dybkjaer, G., Jensen, K. H., Refsgaard, J. C., and Rasmussen, K.: Use of remotely sensed precipitation and leaf area index in a distributed hydrological model, Journal of Hydrology, 264, 34-50, http://dx.doi.org/10.1016/S0022-1694(02)00046-X, 2002.~~

Allen, R. G., Pereira, L. S., Raes, D., and Smith, M.: Crop evapotranspiration-Guidelines for computing crop water requirements-FAO Irrigation and drainage paper 56, FAO, Rome, 300, D05109, 1998.

Berrisford, P., Dee, D. P., Poli, P., Brugge, R., Fielding, K., Fuentes, M., Kållberg, P. W., Kobayashi, S., Uppala, S., and Simmons, A.: The ERA-Interim archive Version 2.0, in: ERA Report Series, ECMWF, Shinfield Park, Reading, 23, 2011.

Bertoldi, G., Notarnicola, C., Leitinger, G., Endrizzi, S., Zebisch, M., Della Chiesa, S., and Tappeiner, U.: Topographical and ecohydrological controls on land surface temperature in an alpine catchment, Ecohydrology, 3, 189-204, 10.1002/eco.129, 2010.

Boegh, E., Thorsen, M., Butts, M. B., Hansen, S., Christiansen, J. S., Abrahamsen, P., Hasager, C. B., Jensen, N. O., van der Keur, P., Refsgaard, J. C., Schelde, K., Soegaard, H., and Thomsen, A.: Incorporating remote sensing data in physically based distributed agro-hydrological modelling, Journal of Hydrology, 287, 279-299, http://dx.doi.org/10.1016/j.jhydrol.2003.10.018, 2004.

Bowen, I. S.: The ratio of heat losses by conduction and by evaporation from any water surface, Physical Review, 27, 779-787, 10.1103/PhysRev.27.779, 1926.

Breuning Madsen, H., and Platou, S. W.: Land Use Planning in Denmark, The Use of Soil Physical Data in Irrigation Planning, 14, 267-276, 1983.

Brutsaert, W., and Sugita, M.: Application of self-preservation in the diurnal evolution of the surface energy budget to determine daily evaporation, Journal of Geophysical Research: Atmospheres, 97, 18377-18382, 10.1029/92JD00255, 1992.

Chen, J., Famigliett, J. S., Scanlon, B. R., and Rodell, M.: Groundwater Storage Changes: Present Status from GRACE Observations, Surveys in Geophysics, 37, 397-417, 10.1007/s10712-015-9332-4, 2016.

Clark, M. P., Nijssen, B., Lundquist, J. D., Kavetski, D., Rupp, D. E., Woods, R. A., Freer, J. E., Gutmann, E. D., Wood, A. W., Brekke, L. D., Arnold, J. R., Gochis, D. J., and Rasmussen, R. M.: A unified approach for process-based hydrologic modeling: 1. Modeling concept, Water Resources Research, 51, 2498-2514, 10.1002/2015WR017198, 2015.

Cong, N., Piao, S., Chen, A., Wang, X., Lin, X., Chen, S., Han, S., Zhou, G., and Zhang, X.: Spring vegetation green-up date in China inferred from SPOT NDVI data: A multiple model analysis, Agricultural and Forest Meteorology, 165, 104-113, http://dx.doi.org/10.1016/j.agrformet.2012.06.009, 2012.

Conradt, T., Wechsung, F., and Bronstert, A.: Three perceptions of the evapotranspiration landscape: Comparing spatial patterns from a distributed hydrological model, remotely sensed surface temperatures, and sub-basin water balances, Hydrology and Earth System Sciences, 17, 2947-2966, 10.5194/hess-17-2947-2013, 2013.

Corbari, C., and Mancini, M.: Calibration and validation of a distributed energy-water balance model using satellite data of land surface temperature and ground discharge measurements, Journal of Hydrometeorology, 15, 376-392, 10.1175/JHM-D-12-0173.1, 2014.

Corbari, C., Mancini, M., Li, J., and Su, Z.: Can satellite land surface temperature data be used similarly to river discharge measurements for distributed hydrological model calibration?, Hydrological Sciences Journal, 60, 202-217, 10.1080/02626667.2013.866709, 2015.

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

Fang, Z., Bogena, H., Kollet, S., Koch, J., and Vereecken, H.: Spatio-temporal validation of long-term 3D hydrological simulations of a forested catchment using empirical orthogonal functions and wavelet coherence analysis, Journal of Hydrology, 529, Part 3, 1754-1767, http://dx.doi.org/10.1016/j.jhydrol.2015.08.011, 2015.

Freeze, R. A., and Harlan, R. L.: Blueprint for a physically-based, digitally-simulated hydrologic response model, Journal of Hydrology, 9, 237-258, 1969.

Gentine, P., Entekhabi, D., Chehbouni, A., Boulet, G., and Duchemin, B.: Analysis of evaporative fraction diurnal behaviour, Agricultural and Forest Meteorology, 143, 13-29, http://dx.doi.org/10.1016/j.agrformet.2006.11.002, 2007.

Githui, F., Selle, B., and Thayalakumaran, T.: Recharge estimation using remotely sensed evapotranspiration in an irrigated catchment in southeast Australia, Hydrological Processes, 26, 1379-1389, 10.1002/hyp.8274, 2012.

Formatted: German (Germany)

Formatted: German (Germany)

Formatted: German (Germany)

Formatted: German (Germany)

Formatted: German (Germany)

Formatted: German (Germany)

Formatted: German (Germany)

Formatted: German (Germany)

Graf, A., Bogena, H. R., Drüe, C., Hardelauf, H., Pütz, T., Heinemann, G., and Vereecken, H.: Spatiotemporal relations between water budget components and soil water content in a forested tributary catchment, Water Resources Research, 50, 4837-4857, 10.1002/2013WR014516, 2014.

Grayson, R. B., and Blöschl, G.: Spatial modelling of catchment dynamics, in: Spatial Patterns in Catchment Hydrology: Observations and Modelling, edited by: Grayson, R. B., Blöschl, G. (Eds.), Cambridge University Press, 51–81, 2000.

Gutman, G., and Ignatov, A.: The derivation of the green vegetation fraction from NOAA/AVHRR data for use in numerical weather prediction models, International Journal of Remote Sensing, 19, 1533-1543, 10.1080/014311698215333, 1998.

Guzinski, R., Anderson, M. C., Kustas, W. P., Nieto, H., and Sandholt, I.: Using a thermal-based two source energy balance model with time-differencing to estimate surface energy fluxes with day-night MODIS observations, Hydrology and Earth System Sciences, 17, 2809-2825, 2013.

Guzinski, R., Nieto, H., Stisen, S., and Fensholt, R.: Inter-comparison of energy balance and hydrological models for land surface energy flux estimation over a whole river catchment, Hydrol. Earth Syst. Sci., 19, 2017-2036, 10.5194/hess-19-2017-2015, 2015.

Hansen, J. R., Refsgaard, J. C., Ernstsen, V., Hansen, S., Styczen, M., and Poulsen, R. N.: An integrated and physically based nitrogen cycle catchment model, Hydrology Research, 40, 347-363, 10.2166/nh.2009.035, 2009.

Hendricks Franssen, H. J., Brunner, P., Makobo, P., and Kinzelbach, W.: Equally likely inverse solutions to a groundwater flow problem including pattern information from remote sensing images, Water Resources Research, 44, 10.1029/2007WR006097, 2008.

Henriksen, H. J., Troldborg, L., Nyegaard, P., Sonnenborg, T. O., Refsgaard, J. C., and Madsen, B.: Methodology for construction, calibration and validation of a national hydrological model for Denmark, Journal of Hydrology, 280, 52-71, http://dx.doi.org/10.1016/S0022-1694(03)00186-0, 2003.

Henriksen, H. J., Troldborg, L., Højberg, A. L., and Refsgaard, J. C.: Assessment of exploitable groundwater resources of Denmark by use of ensemble resource indicators and a numerical groundwater–surface water model, Journal of Hydrology, 348, 224-240, http://dx.doi.org/10.1016/j.jhydrol.2007.09.056, 2008.

Højberg, A. L., Troldborg, L., Stisen, S., Christensen, B. B. S., and Henriksen, H. J.: Stakeholder driven update and improvement of a national water resources model, Environmental Modelling & Software, 40, 202-213, http://dx.doi.org/10.1016/j.envsoft.2012.09.010, 2013.

Immerzeel, W. W., and Droogers, P.: Calibration of a distributed hydrological model based on satellite evapotranspiration, Journal of Hydrology, 349, 411-424, 10.1016/j.jhydrol.2007.11.017, 2008.

Immerzeel, W. W., Droogers, P., de Jong, S. M., and Bierkens, M. F. P.: Large-scale monitoring of snow cover and runoff simulation in Himalayan river basins using remote sensing, Remote Sensing of Environment, 113, 40-49, http://dx.doi.org/10.1016/j.rse.2008.08.010, 2009.

Jönsson, P., and Eklundh, L.: Seasonality extraction by function fitting to time-series of satellite sensor data, IEEE Transactions on Geoscience and Remote Sensing, 40, 1824-1832, 10.1109/TGRS.2002.802519, 2002.

Jönsson, P., and Eklundh, L.: TIMESAT—a program for analyzing time-series of satellite sensor data, Computers & Geosciences, 30, 833-845, http://dx.doi.org/10.1016/j.cageo.2004.05.006, 2004.

Kalma, J. D., McVicar, T. R., and McCabe, M. F.: Estimating Land Surface Evaporation: A Review of Methods Using Remotely Sensed Surface Temperature Data, Surveys in Geophysics, 29, 421-469, 10.1007/s10712-008-9037-z, 2008.

Karlsson, I. B., Sonnenborg, T. O., Refsgaard, J. C., Trolle, D., Børgesen, C. D., Olesen, J. E., Jeppesen, E., and Jensen, K. H.: Combined effects of climate models, hydrological model structures and land use scenarios on hydrological impacts of climate change, Journal of Hydrology, 535, 301-317, http://dx.doi.org/10.1016/j.jhydrol.2016.01.069, 2016.

Koch, J., Jensen, K. H., and Stisen, S.: Toward a true spatial model evaluation in distributed hydrological modeling: Kappa statistics, Fuzzy theory, and EOF-analysis benchmarked by the human perception and evaluated against a modeling case study, Water Resources Research, 51, 1225-1246, 10.1002/2014WR016607, 2015.

Koch, J., Siemann, A., Stisen, S., and Sheffield, J.: Spatial validation of large-scale land surface models against monthly land surface temperature patterns using innovative performance metrics, Journal of Geophysical Research: Atmospheres, 121, 5430-5452, 10.1002/2015JD024482, 2016.

Koch, J., Mendiguren, G., Mariethoz, G., and Stisen, S.: Spatial Sensitivity Analysis of Simulated Land Surface Patterns in a Catchment Model Using a Set of Innovative Spatial Performance Metrics, Journal of Hydrometeorology, 18, 1121-1142, 10.1175/jhm-d-16-0148.1, 2017.

Komatsu, H.: Forest categorization according to dry-canopy evaporation rates in the growing season: comparison of the Priestley–Taylor coefficient values from various observation sites, Hydrological Processes, 19, 3873-3896, 10.1002/hyp.5987, 2005.

Lettenmaier, D. P., Alsdorf, D., Dozier, J., Huffman, G. J., Pan, M., and Wood, E. F.: Inroads of remote sensing into hydrologic science during the WRR era, Water Resources Research, 51, 7309-7342, 10.1002/2015WR017616, 2015.

Li, H. T., Brunner, P., Kinzelbach, W., Li, W. P., and Dong, X. G.: Calibration of a groundwater model using pattern information from remote sensing data, Journal of Hydrology, 377, 120-130, http://dx.doi.org/10.1016/j.jhydrol.2009.08.012, 2009.

Formatted: German (Germany)

Formatted: German (Germany)

Formatted: German (Germany)

Formatted: German (Germany)

Formatted: German (Germany)

Formatted: German (Germany)

Formatted: German (Germany)

Formatted: German (Germany)

Formatted: German (Germany)

Mascaro, G., Vivoni, E. R., and Méndez-Barroso, L. A.: Hyperresolution hydrologic modeling in a regional watershed and its interpretation using empirical orthogonal functions, Advances in Water Resources, 83, 190-206, http://dx.doi.org/10.1016/j.advwatres.2015.05.023, 2015.
Maxwell, R. M., and Kollet, S. J.: Interdependence of groundwater dynamics and land-energy feedbacks under climate change, Nature Geosci, 1, 665-669, 2008.
Mendiguren, G., Pilar Martín, M., Nieto, H., Pacheco-Labrador, J., and Jurdao, S.: Seasonal variation in grass water content estimated from proximal sensing and MODIS time series in a Mediterranean Fluxnet site, Biogeosciences, 12, 5523-5535, 10.5194/bg-12-5523-2015, 2015.
Miralles, D. G., Holmes, T. R. H., De Jeu, R. A. M., Gash, J. H., Meesters, A. G. C. A., and Dolman, A. J.: Global land-surface evaporation estimated from satellite-based observations, Hydrol. Earth Syst. Sci., 15, 453-469, 10.5194/hess-15-453-2011, 2011.
Mu, Q., Heinsch, F. A., Zhao, M., and Running, S. W.: Development of a global evapotranspiration algorithm based on MODIS and global meteorology data, Remote Sensing of Environment, 111, 519-536, http://dx.doi.org/10.1016/j.rse.2007.04.015, 2007.
Norman, J. M., Kustas, W. P., and Humes, K. S.: Source approach for estimating soil and vegetation energy fluxes in observations of directional radiometric surface temperature, Agricultural and Forest Meteorology, 77, 263-293, http://dx.doi.org/10.1016/0168-1923(95)02265-Y, 1995.
Perry, M. A., and Niemann, J. D.: Analysis and estimation of soil moisture at the catchment scale using EOFs, Journal of Hydrology, 334, 388-404, http://dx.doi.org/10.1016/j.jhydrol.2006.10.014, 2007.
Priestley, C. H. B., and Taylor, R. J.: On the Assessment of Surface Heat Flux and Evaporation Using Large-Scale Parameters, Monthly Weather Review, 100, 81-92, 10.1175/1520-0493(1972)100<0081:OTAOSH>2.3.CO;2, 1972.
Rajib, M. A., Merwade, V., and Yu, Z.: Multi-objective calibration of a hydrologic model using spatially distributed remotely sensed/in-situ soil moisture, Journal of Hydrology, 536, 192-207, http://dx.doi.org/10.1016/j.jhydrol.2016.02.037, 2016.
Refsgaard, J. C.: Parameterisation, calibration and validation of distributed hydrological models, Journal of Hydrology, 198, 69-97, http://dx.doi.org/10.1016/S0022-1694(96)03329-X, 1997.
Refsgaard, J. C., Stisen, S., Højberg, A. L., Olsen, M., Henriksen, H. J., Børgesen, C. D., Vejen, F., Kern-Hansen, C., and Blicher-Mathiesen, G.: DANMARKS OG GRØNLANDS GEOLOGISKE UNDERSØGELSE RAPPORT 2011/77, Geological Survey of Danmark and Greenland (GEUS), 2011.
Richey, A. S., Thomas, B. F., Lo, M.-H., Reager, J. T., Famiglietti, J. S., Voss, K., Swenson, S., and Rodell, M.: Quantifying renewable groundwater stress with GRACE, Water Resources Research, 51, 5217-5238, 10.1002/2015WR017349, 2015.
Ridler, M.-E., Madsen, H., Stisen, S., Bircher, S., and Fensholt, R.: Assimilation of SMOS-derived soil moisture in a fully integrated hydrological and soil-vegetation-atmosphere transfer model in Western Denmark, Water Resources Research, 50, 8962-8981, 10.1002/2014WR015392, 2014.
Rientjes, T. H. M., Muthuwatta, L. P., Bos, M. G., Booij, M. J., and Bhatti, H. A.: Multi-variable calibration of a semi-distributed hydrological model using streamflow data and satellite-based evapotranspiration, Journal of Hydrology, 505, 276-290, http://dx.doi.org/10.1016/j.jhydrol.2013.10.006, 2013.
Ringgaard, R., Herbst, M., Friborg, T., Schelde, K., Thomsen, A. G., and Soegaard, H.: Energy Fluxes above Three Disparate Surfaces in a Temperate Mesoscale Coastal Catchment, Vadose Zone Journal, 10, 54-66, 10.2136/vzj2009.0181, 2011.
Rodell, M., Velicogna, I., and Famiglietti, J. S.: Satellite-based estimates of groundwater depletion in India, Nature, 460, 999-1002, 2009.
Rouse, J. W., Haas, R. H., Deering, D. W., and Schell, J. A.: Monitoring the vernal advancement and retrogradation (green wave effect) of natural vegetation, Goddard Space Flight Center, Greenbelt, MD, 87, 1973.
Ruiz-Pérez, G., Koch, J., Manfreda, S., Caylor, K., and Francés, F.: Calibration of a parsimonious distributed ecohydrological daily model in a data scarce basin using exclusively the spatio-temporal variation of NDVI, Hydrol. Earth Syst. Sci. Discuss., 2016, 1-33, 10.5194/hess-2016-573, 2016.
Samaniego, L., Kumar, R., and Attinger, S.: Multiscale parameter regionalization of a grid-based hydrologic model at the mesoscale, Water Resources Research, 46, 10.1029/2008WR007327, 2010.
Savitzky, A., and Golay, M. J. E.: Smoothing and Differentiation of Data by Simplified Least Squares Procedures, Analytical Chemistry, 36, 1627-1639, 10.1021/ac60214a047, 1964.
Schuurmans, J. M., Troch, P. A., Veldhuizen, A. A., Bastiaanssen, W. G. M., and Bierkens, M. F. P.: Assimilation of remotely sensed latent heat flux in a distributed hydrological model, Advances in Water Resources, 26, 151-159, http://dx.doi.org/10.1016/S0309-1708(02)00089-1, 2003.
Stisen, S., Jensen, K. H., Sandholt, I., and Grimes, D. I. F.: A remote sensing driven distributed hydrological model of the Senegal River basin, Journal of Hydrology, 354, 131-148, http://dx.doi.org/10.1016/j.jhydrol.2008.03.006, 2008.
Stisen, S., McCabe, M. F., Refsgaard, J. C., Lerer, S., and Butts, M. B.: Model parameter analysis using remotely sensed pattern information in a multi-constraint framework, Journal of Hydrology, 409, 337-349, http://dx.doi.org/10.1016/j.jhydrol.2011.08.030, 2011.
Stisen, S., Hojberg, A. L., Troldborg, L., Refsgaard, J. C., Christensen, B. S. B., Olsen, M., and Henriksen, H. J.: On the importance of appropriate precipitation gauge catch correction for hydrological modelling at mid to high latitudes, Hydrology and Earth System Sciences, 16, 4157-4176, 10.5194/hess-16-4157-2012, 2012.

Formatted: German (Germany)
Formatted: German (Germany)
Formatted: German (Germany)
Formatted: German (Germany)
Formatted: German (Germany)
Formatted: German (Germany)
Formatted: German (Germany)
Formatted: German (Germany)
Formatted: German (Germany)
Formatted: German (Germany)
Formatted: German (Germany)
Formatted: German (Germany)

Sugita, M., and Brutsaert, W.: Daily evaporation over a region from lower boundary layer profiles measured with radiosondes, Water Resources Research, 27, 747-752, 10.1029/90WR02706, 1991.
Sutanudjaja, E. H., de Jong, S. M., van Geer, F. C., and Bierkens, M. F. P.: Using ERS spaceborne microwave soil moisture observations to predict groundwater head in space and time, Remote Sensing of Environment, 138, 172-188, http://dx.doi.org/10.1016/j.rse.2013.07.022, 2013.
van der Keur, P., Hansen, J. R., Hansen, S., and Refsgaard, J. C.: Uncertainty in Simulation of Nitrate Leaching at Field and Catchment Scale within the Odense River Basin, Vadose Zone Journal, 7, 10-21, 10.2136/vzj2006.0186, 2008.
Vansteenkiste, T., Tavakoli, M., Van Steenbergen, N., De Smedt, F., Batelaan, O., Pereira, F., and Willems, P.: Intercomparison of five lumped and distributed models for catchment runoff and extreme flow simulation, Journal of Hydrology, 511, 335-349, http://dx.doi.org/10.1016/j.jhydrol.2014.01.050, 2014.
Vereecken, H., Pachepsky, Y., Simmer, C., Rihani, J., Kunoth, A., Korres, W., Graf, A., Franssen, H. J. H., Thiele-Eich, I., and Shao, Y.: On the role of patterns in understanding the functioning of soil-vegetation-atmosphere systems, Journal of Hydrology, 542, 63-86, http://dx.doi.org/10.1016/j.jhydrol.2016.08.053, 2016.
Wanders, N., Bierkens, M. F. P., de Jong, S. M., de Roo, A., and Karssenberg, D.: The benefits of using remotely sensed soil moisture in parameter identification of large-scale hydrological models, Water Resources Research, 50, 6874-6891, 10.1002/2013WR014639, 2014.
Wang, D.-C., Zhang, G.-L., Zhao, M.-S., Pan, X.-Z., Zhao, Y.-G., Li, D.-C., and Macmillan, B.: Retrieval and Mapping of Soil Texture Based on Land Surface Diurnal Temperature Range Data from MODIS, PLOS ONE, 10, e0129977, 10.1371/journal.pone.0129977, 2015.
Wang, L., Koike, T., Yang, K., and Yeh, P. J.-F.: Assessment of a distributed biosphere hydrological model against streamflow and MODIS land surface temperature in the upper Tone River Basin, Journal of Hydrology, 377, 21-34, http://dx.doi.org/10.1016/j.jhydrol.2009.08.005, 2009.
Windolf, J., Thodsen, H., Troldborg, L., Larsen, S. E., Bøgestrand, J., Ovesen, N. B., and Kronvang, B.: A distributed modelling system for simulation of monthly runoff and nitrogen sources, loads and sinks for ungauged catchments in Denmark, Journal of Environmental Monitoring, 13, 2645-2658, 10.1039/c1em10139k, 2011.
Yan, J., and Smith, K. R.: SIMULATION OF INTEGRATED SURFACE WATER AND GROUND WATER SYSTEMS - MODEL FORMULATION1, JAWRA Journal of the American Water Resources Association, 30, 879-890, 10.1111/j.1752-1688.1994.tb03336.x, 1994.
Zhang, D., Madsen, H., Ridler, M. E., Kidmose, J., Jensen, K. H., and Refsgaard, J. C.: Multivariate hydrological data assimilation of soil moisture and groundwater head, Hydrol. Earth Syst. Sci., 20, 4341-4357, 10.5194/hess-20-4341-2016, 2016.

Formatted: German (Germany)

Formatted: German (Germany)

Formatted: German (Germany)

Formatted: German (Germany)