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 and Stisen, 2017) 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.

Regarding the description of the differences between model domains, we agree that this has not been written very clearly and this will be 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 will be exemplified better.

Regarding Mu et al., YEAR 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. Oppesed 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 will add an elaborated description of the TSEB-method and discuss how it differs from other regional AET methods and argue why we chose the TSEB.

Koch, J., and Stisen, S.: Citizen science: A new perspective to advance spatial pattern evaluation in hydrology, PLOS ONE, 12, e0178165, 10.1371/journal.pone.0178165, 2017.

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 will be 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:

"2.2 TSEB setup and remotely sensed derived inputs

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:

$$LE = R_n - H - G$$

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.

The sensible heat flux (H) is calculated as:

$$H = H_C + H_S = \rho C_P \left[\frac{T_C - T_A}{R_A} + \frac{T_S - T_A}{R_A + R_S}\right]$$

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, ρC_P is the volumetric heat capacity of air, R_S is the resistance to heat flow in the boundary layer above the soil surface and R_A is the aerodynamic resistance expressed as:

$$R_A = \frac{\left[ln\left(\frac{z_U - d}{z_M}\right) - \Psi_M\right]\left[ln\left(\frac{z_T - d}{z_M}\right) - \Psi_H\right]}{0.16U}$$

where z_U and z_T are the height of the wind speed and air temperature (U and Ta)is the aerodynamic resistance, R_S is the resistance to heat flow in the boundary layer immidiatelly above the soil layer. T_A represents the air temperature, T_C is the canopy temperature, T_S is the soil temperature. Ψ_M and Ψ_H are the diabatic 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), 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.

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

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:

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

And where f_c is calculated as:

$$f_c = 1 - exp\left(\frac{-0.5LAI}{cos\,\theta}\right)$$

Where θ is the viewing angle.

The model iterates until the energy balance equations are satisfied for soil and canopy. The readers 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.

Several inputs to TSEB are directly obtained from the LST product 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 TERRA and AQUA 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 used to select only those pixels with the best observation possible. LAI is derived using an empirical relationship with the Normalized Difference Vegetation Index (NDVI) (Rouse et al., 1973)."

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.

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.

• 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 included in the text what BRDF means: Bidirectional Reflectance Distribution Function

• 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.

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

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

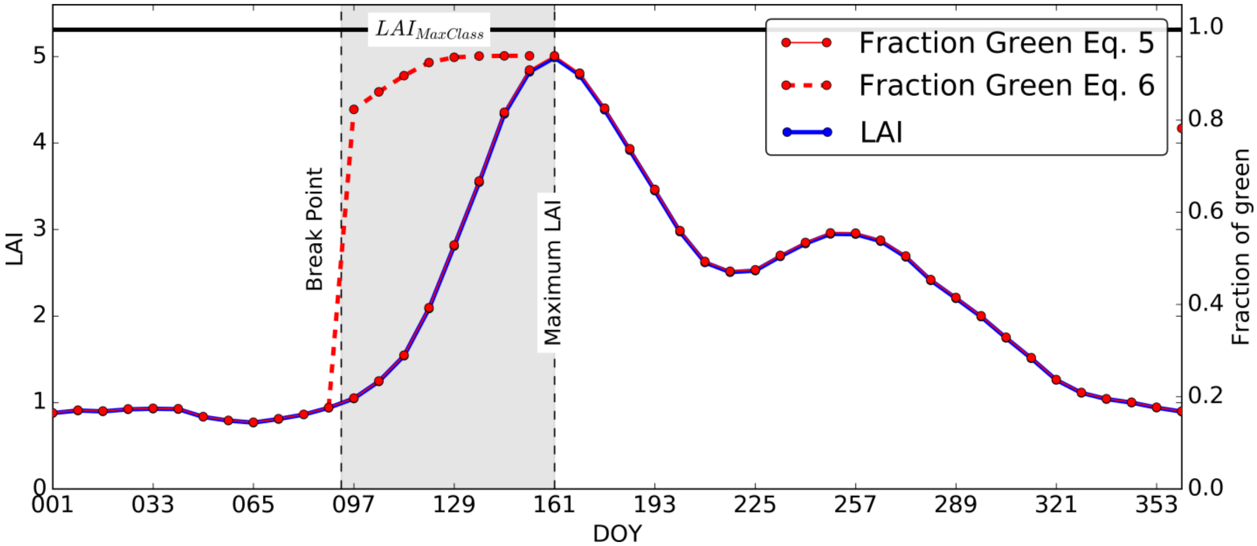
Albedo was used in the study to calculate the net radiation.

• 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.

• 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. Data presented corresponds to an Agricultural pixel from the dataset. Grey area shows the region where Eq. 6 is used instead of Eq.5

• 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 increased by 20% compared to the pre-growing season low. Once the date where that happened was found we assumed that all the fraction of the vegetation should present most of the fraction green and therefore we considered that a rapid change reaching values close to 0.9 were meaningful. Then 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 approach is not found in the literature, we will explain better why we apply it. 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:

$$Fg_i = \frac{LAI_i}{LAI_{MaxClass}} \text{ (EQ 5)}$$

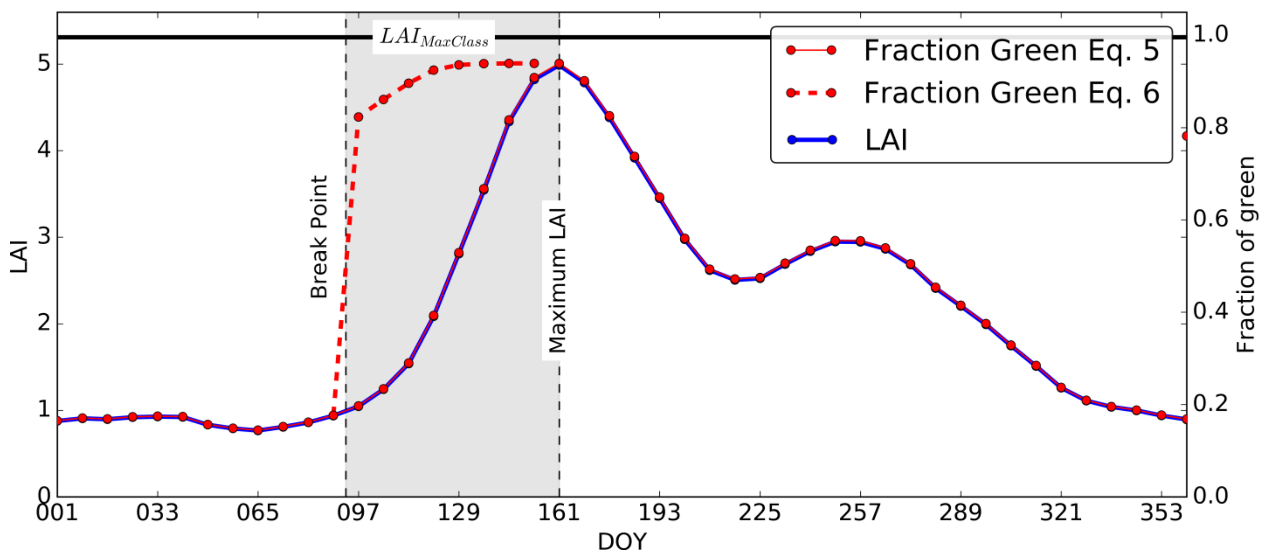
Where Fg_i indicates the Fraction of green for a certain pixel i, LAI_i indicates the LAI value for a pixel i and $LAI_{MaxClass}$ is the maximum LAI value for an specific land cover type. This equation was applied to needle leaf forest land cover type.

For the other land cover types (deciduous, grasslands, crops…) equation xxx was modified adding another term. These land cover types show a stronger seasonality. 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 F_g to non-needle leaf forest land covers, Figure xx. For these types of vegetation Fg will be allowed to increase rapidly just after crop emergence by substituting EQ 5 by EQ 6.

$$Fg_i = \frac{LAI_{i,max}}{LAI_{MaxClass}} \cdot (1 - e^{(-2 \cdot LAI_i)}) \qquad \textit{(EQ 6)}$$

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 defined by an increase in 20% increase in LAI compared to the winter low and until the to the time at which LAI reaches its maximum (see next Figure)."



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 is maximum), senescence starts and more non- photosynthetically active regions start to appear in the plant, what is translated in lower Fg values.

• 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.

• 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:

"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 (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: In the first stages of the study we used the approach based on LAI but we found some problems when utilizing that equation to translate LAI to root depth. In this study the LAI was developed 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. 9/10.

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

Reply: I think you mean eq 10**.** Sorry for not including it. Now it is included in the manuscript.

Now the text is as follows:

"Where $NDVI_i$is the value of the NDVI for a pixel i, and $NDVI_{max}$indicates the maximum NDVI value for the pixel in the time series...."

We have double checked this error in the manuscript and correct it where it has been necessary.

• 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. They are generated from soil property maps based on the relations between soil properties and effective rooting depth described on page 25 of Refsgaard et al. (2011) (http://vandmodel.dk/xpdf/77-2011_vandbalance.pdf). The RDmax values are scaled 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.

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.

• 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 and 6. (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 a 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 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.

• 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. 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 cloudfree 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 cloudfree conditions. Most probably the AET maps for all weather conditions would look different.

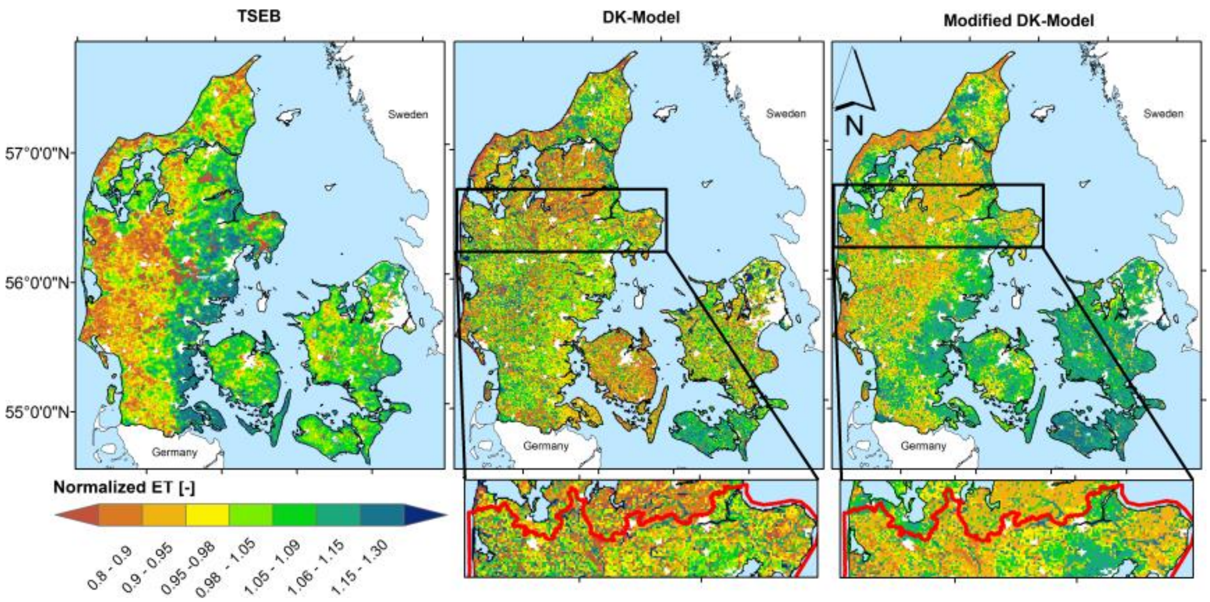
• P10L17: are causing differences in area 2 in the model domain

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 ...

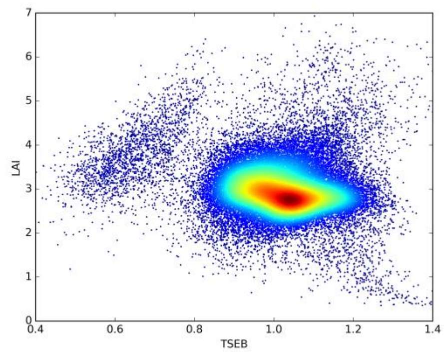
Reply: Thanks for the suggestion. Text has been modified.

• 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.

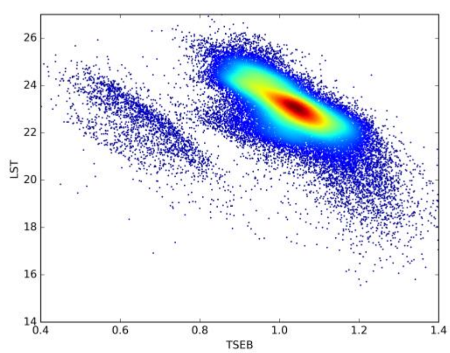
Reply: We agree that the visual interpretation should be backed by some quantification; we will add correlation coefficients between variables and TSEB AET to the maps in figure 9.

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:

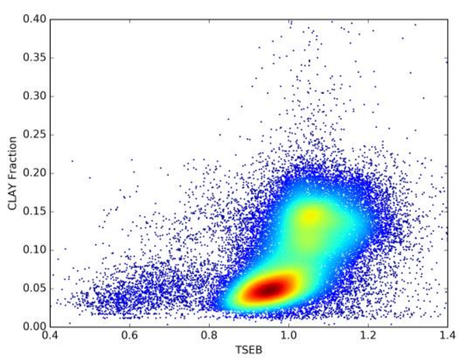
Density scatter plot TSEB- LAI (r =-0.15)



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)

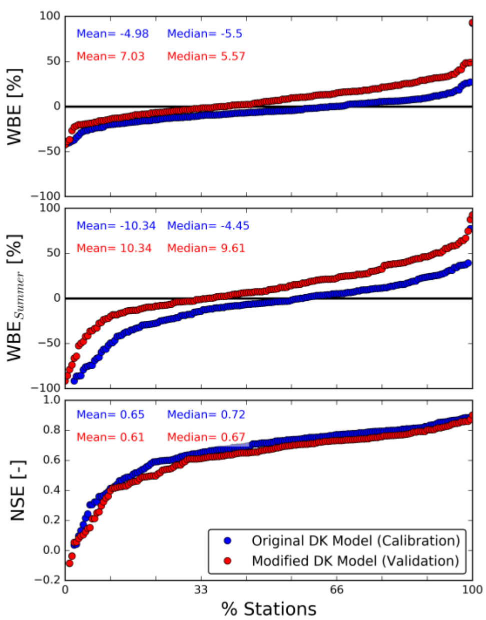


Figure 12. Model performance statistics showing the results of the original and modified versions of the DK model. Stations have been ranked by performance and presented in the x axis as a percentage. Figure shows the Nash-Sutcliffe Efficiency (NSE), water balance error (WBE), and water balance error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

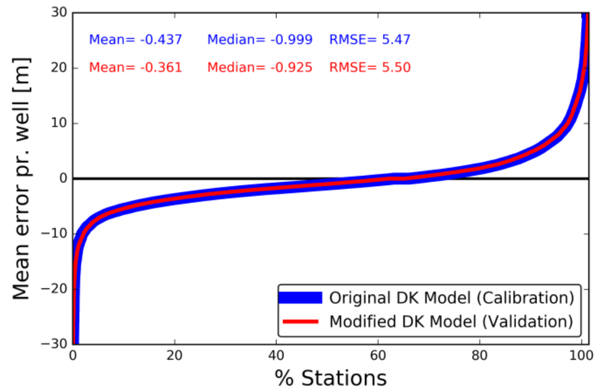


Figure 13. The figure shows the error and statistics in the ground water heads estimated from the original DK model and the modified version of the DK model. Stations have been ranked by performance and presented in the x axis as a percentage.

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 performanceWe 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 DK model performance

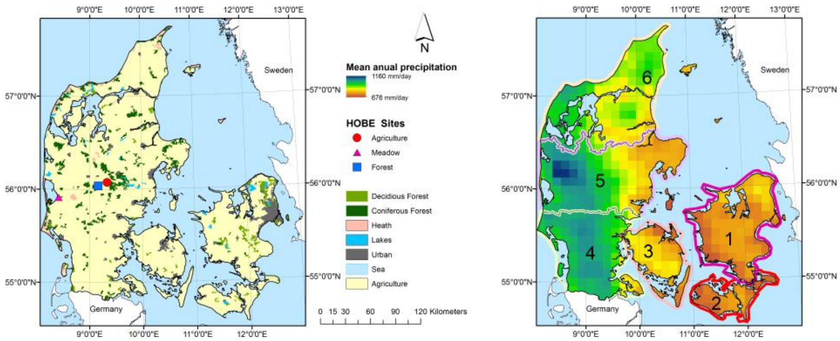
Results showing the water balance error (WBE) for all year and summer (from xxx to xxx) as well as NSE (Nash-Sutcliffe Efficiency) are presented in figure 12. The performance statistics of the DK model has change compared to its original calibrated setup. The first noticeable thing that can be concluded 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 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, which sounds like a large error, however, the median, is below 1 m.

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 in model bias 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: No, LAIiMax is for the current grid cell, while LAImax (should be named LAImaxclass) is 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