

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 #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.

C1

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 felt 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 will 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 will highlight and clarify 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.

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.). Such an approach will ensure a physically consistent and realistic spatial distribution and has already been implemented by Samaniego et al.(2010). This should be combined with an objective function which does not evaluate the individual error of each grid, but evaluates the spatial pattern across the model domain.

Samaniego, L., Kumar, R., Attinger, S., 2010. Multiscale parameter regionalization of a grid-based hydrologic model at the mesoscale. *Water Resour. Res.* 46. doi:10.1029/2008wr007327

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 in spatial patterns between the satellites based estimates and models and between the two models. We will rephrase that part to make it clearer to:

“...The model evaluation will be 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. By this approach it is aimed at identifying which parts of the model parametrization are creating these differences in the spatial pattern...”

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 patterns information and not on the absolute values the study we conduct here. 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 and Remote sensing derived inputs

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

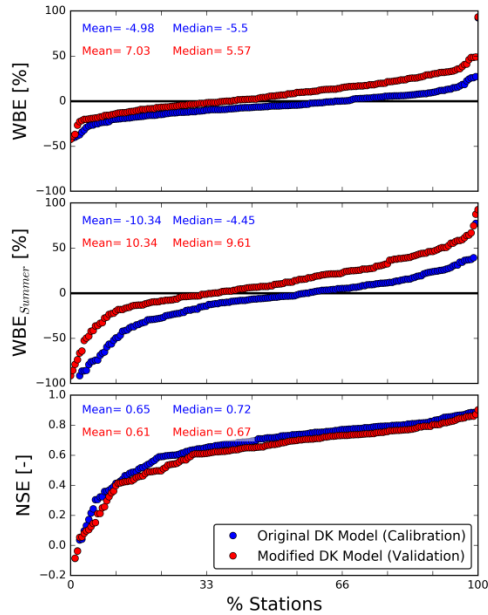
2.3.- Spatial patterns analysis: Empirical orthogonal functions

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?

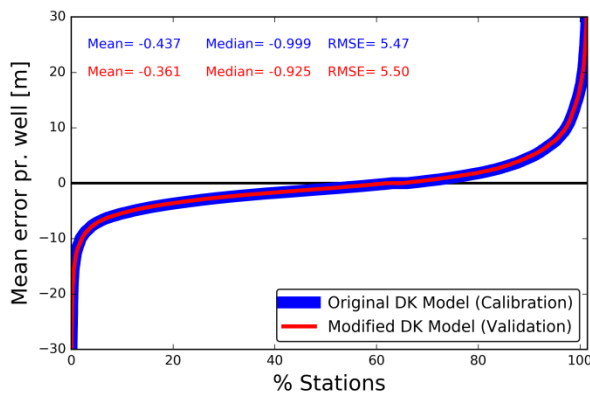
Reply: Reviewer 1 has pointed to this and we agree that this information should be included in the manuscript. To reduce the size of the response, please read the response give to reviewer 1 in the last comment before the technical corrections.

We include the figures for guidance:

Water balance error:



Ground Water heads:



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.

$T_{\text{Rad}}(\theta) = [f(\theta)T_C^n + (1 - f(\theta))T_S^n]^{1/n}$, where θ is the observation angle, T_{Rad} is the radiometric temperature, T_s is the soil temperature, T_c 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 m^2/m^2 , alfa is m^2/m^2 and beta 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:

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

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.

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 F_g 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:

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

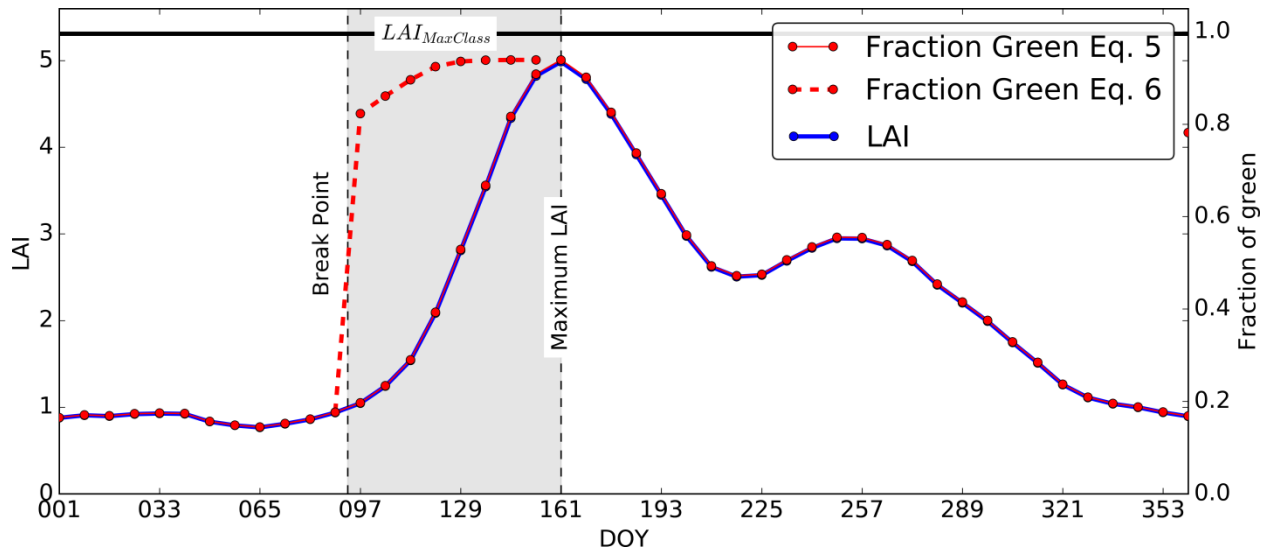
Where Fg_i indicates the Fraction of green for a certain pixel i , LAI_i indicates the LAI value for a pixel i and $LAI_{MaxClass}$ is the maximum LAI value for an specific land cover type. This equation was applied to needle leaf forest land cover type.

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 F_g will be allowed to increase rapidly just after crop emergence by substituting EQ 5 by EQ 6.

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

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 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 is maximum), 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. 9/10.

Equations 11 and 12: Please put some references.

Reply: We cannot provide any reference for equation 11 as it was created by us in 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 equations more clearly.

The reference for equation 12 has been included in the 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 the change in variable/parameter. For temperatures these are changes based on their °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 PT_{meadow} and PT_{agri} , because there is not really any physical reason that it should deviate from the original value of 1.28. In contrast literature suggests that PT_{forest} 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. We aim at normalize the maps from figures 5,6 and 6 as was conducted in figure 8, however that will mean that the seasonal component of the ET might disappear. The results and discussion were adjusted to 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.

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 will include it in the revise manuscript.

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 domains. 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.