Reply to Reviewer Comments

Reviewer #1

Comment 1) This paper deals with an innovative calibration framework that combines temporally aggregated observed spatial patterns with a new spatial performance metric and a flexible spatial parameterisation scheme. An application of the mesoscale Hydrologic Model (mHM) to the Skjern River Basin is used as an example show the effectiveness of the presented calibration framework. This is a very timely topic that fits very well to the scope of HESS.

However, the spatial model parameterizing methodology is not well described. Both the root fraction coefficient and the PET correction factor parameterizations should be presented in more detail including graphical presentations of the underlying relationships. For these reasons, my recommendation is to accept this manuscript with minor revisions.

I have provided specific comments and suggestions for improvement below.

Reply from authors: The authors thank the Dr Heye Bogena (reviewer #1) for his constructive comments on the manuscript. We have replied to each comment below.

Specific comments:

Comment 2) P3L7: Which models?

Reply from authors: "Immerzeel and Droogers (2008) showed that the models can be constrained by using spatially distributed observations with a monthly temporal resolution" is replaced by "Immerzeel and Droogers (2008) showed how a semi distributed model of a basin in Southern India could be constrained by using spatially distributed observations with a monthly temporal resolution"

Comment 3) P3L16-32: This section is redundant with the method section and should be shortened.

Reply from authors: We agree with the comment. This section will be shortened in the revised version of the manuscript.

Reply to Reviewer Comments

Reviewer #3

Review of the Article

Combining satellite data and appropriate objective functions for improved spatial pattern performance of a distributed hydrologic model by Mehmet C. Demirel, Juliane Mai, Gorka Mendiguren, Julian Koch, Luis Samaniego, Simon Stisen

The paper addresses the relevant topic of using spatial data for the calibration of distributed hydrological models. The paper explores the tradeoffs between streamflow-based and spatial-based calibrations, illustrating the benefits of combining separate observation types and objective functions. Given the increasing availability of spatial data, to understand the value of a spatial calibration versus a more traditional calibration is a timing and relevant research topic. Spatial pattern of actual evapotranspiration (AET), obtained with the mesoscale Hydrologic Model (mHM), are calibrated against spatial patterns of AET estimated trough remote sensing (TSEB model). The major novel point of the paper is to present novel bias-insensitive spatial pattern metric, which exploits the key information contained in the observed patterns, useful for spatially distributed models optimization. I found the paper valuable and worth to be published. However, I believe some methodological aspects should be further clarified, as detailed in the specific comments.

I suggest a minor revision for the paper. Please revise also carefully English language.

Reply from authors: We wish to thank Dr. Giacomo Bertoldi (Reviewer #3) for the very constructive comments and suggestions.

General comments:

1.In the paper formulates the hypothesis that "The current calibration framework builds on the assumption that the satellite based estimate of AET patterns approximate an observed pattern that is suitable for model optimization". In my advice, this hypothesis should be better motivated and the implications better discussed.

At the end, six monthly mean "climatological" AET maps have been produced. But spatial patterns are time consistent? If the spatial structure of AET changes a lot from time to time, or there is a great inter-annual variability, then it makes little sense to calibrate the model to such patterns. A preliminary analysis on the temporal variability of the observed AET patterns should be done. Only if spatial AET patterns are quite constant over the different years, the proposed calibration approach could be used. Moreover, the fact that an AET monthly "climatology" has been used, could be one reason of the limited tradeoff found with the discharge. Could you discuss about this? In this way,

you cannot for example identify from AET observation a drought period, which can have an effect on discharge.

I see later that this choice is explicit in the paper (see 4.1 Objective functions). "Since all three terms are bias-insensitive, the spatial efficiency only constrains the model simulations with the pattern information in the satellite data while leaving the water balance (bias) to be constrained by streamflow". However, this implies a hypothesis of spatial invariance of AET patterns. You need to support and discuss this assumption in the paper.

Reply from authors: We have elaborated on the assumption that the RS based AET data contains valuable pattern information that can help constrain the hydrological model regarding simulation of spatial patterns. We did a preliminary analysis of how consistent the spatial patterns were in time. They turned out to be quite consistent with average correlations of 0.82 for any individual daily AET map to the corresponding monthly mean map. This we have added to the manuscript, section 2.2. The overall idea is to utilize the most informative part of the observation data (spatial patterns from RS AET maps and water balance and temporal pattern from discharge). In addition, we are aiming at calibrating mainly the spatial parametrization of soil and vegetation parameters using the bias-insensitive spatial pattern metric, these parameters typically do not vary in time (expect in some cases vegetation, like in our case vegetation dynamics change, but the vegetation related calibration parameters are time-invariant). Therefore, we believe that the temporally averaged spatial patterns are a reasonable way of constraining the spatial model parametrization. In addition, a spatially consistent drought would be captured by the discharge observations; however, admittedly our approach is not designed to capture spatial variability in drought patterns (this would especially be relevant for larger catchments).

2. Another assumption is that "Bias and temporal variability of satellite derived AET estimates could be useful for model optimization, however, in this study, we deliberately limited the information content of the satellite data to address the spatial patterns." You clearly separate in your calibration framework the source of temporal information (discharge) from the source of spatial information (AET). It seems at the end that this second part is not really a calibration framework, since you even introduce some modifications in the model structure to properly incorporate spatial information in modelled AET. It seems to me that the main outcome of the paper is an interesting way to integrate spatially distributed observations in a modeling framework. It is a kind of spatial data assimilation, or better spatial data integration approach, more than a traditional model´s parameters calibration. I suggest in the Introduction and in the Discussion to place your results in the broader perspective of the data integration techniques and not only model calibration literature, and underline this potential of your proposed approach.

Reply from authors: I see the point; however, I still regard this as a classical model parameter calibration approach different from data assimilation schemes. The reason is that we do not update states, fluxes or parameters directly to match the observed pattern, but only use the observed spatial patterns to formulate a classical objective function that is then minimized by changing global parameter values. Some of these global parameters are then part of an upscaling operator or transfer function, but they are essentially still model parameters.

Specific comments:

Introduction

It is important to underline that actual evapotranspiration (AET) estimates from satellites are not observations, but the results of a model (in this case the TSEB). This model shares part of the input information used by the hydrological model (i.e. LAI, fractional vegetation cover, meteorological data as temperature). This does not influence paper´s results, but it is important to clearly state that AET patterns are not observations, but the results of a remote-sensing based model, with a lot of uncertainties.

Reply from authors: True, in the original manuscript we have the following sentence "Despite not being pure observations but estimates from an energy balance model based on satellite observations we will refer to these AET maps as reference observations." We will repeat this later in the paper when the AET patterns are used for defining objective functions.

2.2 Satellite based data

Canopy fraction is a key input information of the TSEB model. How it has been estimated? AET estimations of the TSEB model are relative to the instant of the thermal remote sensing image used. How AET estimations have been extended to the daily and monthly time scale?

Reply from authors: Details of this are given in: Mendiguren, G., Koch, J., and Stisen, S.: Spatial pattern evaluation of a calibrated national hydrological model – a remote-sensing-based diagnostic approach, Hydrol. Earth Syst. Sci., 21, 5987-6005, https://doi.org/10.5194/hess-21-5987-2017, 2017. We will make sure reference is given in the manuscript.

Figure 1 TSEB results are usually highly dependent on LST and canopy fraction maps. It is very interesting how the RS-AET map reflects also well the soil type map, since no soil or soil water content information is given to the TSEB. Your hypothesis to link soil type to root density and

therefore to root water extraction is interesting. However other hypotheses could be done. The simplest is just the areal with lower sand content are wetter, because the soil has a higher water holding capacity, and therefore there is a higher AET. Could you comment on this?

Reply from authors: Yes, our initial thought was that simply soil physical properties could explain the differences. However, this was not enough to explain the differences seen in the RS AET maps. In addition, we had specific knowledge from the literature about the effective rooting depth variations in this region. Also the sensitivity of the root fraction coefficient to AET is very high in the model (and in many other PET based AET models as well), so getting the soil influence on AET right while preserving the original uniform root fraction coefficient for all non-forest areas would not be sufficient or correct.

Distributed root fraction coefficient. The assumed model could be better described. Why land cover does not influence root fraction? I expect that different vegetation types are the main source of the different root fractions. Is the root depth a relevant parameter in the mHM model or only root fraction is used?

Reply from authors: mHM uses root fraction distribution, and land cover is the main driver for differences in root fraction, but it is simplified to forest, non-forest vegetation and impervious (the later not being used in the current study due to very minor area of urbanization). The soil dependency we introduce is only to further distribute the root density across non-forest vegetation (agricultural cropland). The main differences in root density is still between forest and non-forest. This is made clearer in the revised manuscript.

4.1 Objective functions. The application of the new SPAEF index is one of the main elements of novelty in the paper. A schematic figure explaining the concept of "histogram intersection" could help. It is not obvious for me.

Reply from authors: The histogram intersect will be explained better.

Figure 3. Why including AET patterns reduces model performances in the summer months? This is the time when I expect AET counts more for the water budget. Is this a drawback of the choice of not considering biases in AET? What would happen if you include also as calibration target the temporal evolution of the spatial average of AET?

Reply from authors: This is probably a drawback, and the loss in performance in discharge during summer months is seen as a tradeoff resulting from adding another objective function.

Figure 4. This figure shows the potential of this technique of spatial calibration more in terms of remote sensing data integration, than in terms of calibration. In fact, in many cases distributed hydrological models cannot produce detailed spatial patterns, because the coarse spatial patterns of

models input data (except topography, soil and land cover types are usually known at a coarse spatial scale). Could you comment in this? (see also general comments)

Reply from authors: As replied above, we regard this as a classical parameter calibration exercise. However, the flexibility in the spatial model parametrization and the use of distributed input (with same resolutions as the model grid) of soil texture and vegetation enables the model to approach a detailed spatial pattern. It should be noted that the results of the Q-only calibration in figure 4 has exactly the same parametrization scheme and free model parameters as the two other calibrations, so the only reason that it fails in reproducing the spatial pattern is that it is not informed about what the "true" pattern is.

Page 16 Results show how combining Q and AET results in a more robust model parametrization for the validation phase. Could this also contribute to reduce models equifinality?

Reply from authors: We hope so; however, we have not addressed equifinality as such.

Page 16 "It is recognized that traditional downstream discharge measurements contain much more accurate and robust information on the overall water balance compared to the non-continuous remotely sensed estimates". Please provide some references to support this.

Reply from authors: Not sure there are specific references for this quite specific statement. We are here arguing that a continuous time-series of total catchment discharge contains more information on the overall water balance compared to a series instantaneous snapshots of AET for only cloud free conditions. We will rephrase to make this our own statement, unless we find an appropriate reference.

Discussion

Page 17 "Here, we ignore the temporal aspect and focus only on the consistent spatial patterns for each month of the growing season" "The current calibration framework builds on the assumption that the satellite based estimate of AET patterns approximate an observed patterns that is suitable for model optimization". Please motivates better this assumption. See general comments above.

Reply from authors: We have tries to motivate this better in the revised manuscript section 2.2.

The choice of not including temporal information from the AET data is an assumption quite well justified in the discussion. However, this could reduce the information amount coming from AET.

Reply from authors: true, we could miss information actually contained in the AET data, however, as explained throughout the manuscript, we focus on the information content of the RS AET that supplements the discharge time series the best.

"AET estimate is validated against eddy covariance stations (Mendiguren et al., 2017) they only represent specific cloud-free days limiting their value to assess the long term water balance of the catchment" OK, but if AET patters in cloudy days are different? What happens if AET patterns are not stationary in time? From my experience with distributed models, looking to daily model´s AET output maps, I´ve been ever surprised how such patterns are changing from day to day. See also general comments. A preliminary analysis of the time series of the RS-AET map should be done to assess how much time-consistent are the patterns.

Reply from authors: This is definitely a limitation to our approach. But again, we have decided to utilize what we assume is the best information content available from the AET data. This is the spatial pattern on the days where we have LST observations. We could have interpolated these cloud-free "observations" to cloudy and rainy days, however this would not increase our number of "true" observations. In addition, for the current catchment, the impact of soil and vegetation parameters (which we calibrate) on AET on very cloudy and rainy days will be very limited, because in this region, potential Evapotranspiration will occur as soon as the potential evapotranspiration becomes low (as will happen on cloudy and rainy days).

Reply to Reviewer Comments

Reviewer #2

The authors present a methodology aimed to improve predictions of a distributed hydrological model both in time and space. In order to identify model parameters responsible for the spatial predictions, an additional complex objective function is introduced taking into account the match between observed (satellite based) spatial evapotranspiration patterns and those predicted by the model.

The paper is interesting and the approach taken opens a new direction of research towards the use of spatial remote sensed observations to improve predictions of a distributed hydrological model. In particular, the authors applied site-specific parameterizations to increase the flexibility of the description of actual evapotranspiration characteristics (root zone and potential evapotranspiration corrections). The authors formulated the problem in a deterministic framework, which might be a good introduction to a new approach, but the discussion on uncertainty is missing. The paper is well written and requires only some clarification of the presented material and a wider discussion of the assumptions taken. I recommend to publish the paper after minor corrections.

Reply from authors: The authors thank the Dr Renata Romanowicz (reviewer #2) for her positive and constructive comments on the manuscript.

Specific comments

Comment 43) Page 1, line 18 In addition two new site-specific spatial parameter distribution options have been introduced

Reply from authors: Corrected.

Comment 44) Page 2, line 19 This is from the fact grammar should be corrected e.g. ‘this is because..’

Reply from authors: Corrected.

Comment 45) Page 3, Line 17 it is not clear what the authors mean by ‘domain’

Reply from authors: We meant “catchment” with domain. This is corrected in the revised version of the manuscript.

Comment 46) Page 3, line 22 is it for comparing spatial patterns of two continuous variables?

Reply from authors: Corrected.

Comment 47) Page 5, lines 8-11 what is the uncertainty of AET estimates?

Reply from authors: Please see reply to comment 5) by the reviewer #1

Comment 48) Page 5, line 30 Could you please give more detail about the way monthly AET maps are applied in the model and the disaggregation method used?

Reply from authors: Page 5 line 30 is not about the AET maps but the LAI input, but we have elaborated on the monthly averaging of AET maps in the revised manuscript.

Comment 49) Page 6, line 9 The parameterisation introduced is a very interesting way forward and requires a separate paper backed up with field experiments. Could you please give the possible disadvantages of the parameterisation? Even though the parameterisation decreases the number of parameters of a distributed hydrological model, the parameters require a sufficient amount of observations to be properly identified. The question is how to test the parameterisation using very limited and uncertain information obtained from the indirect and fragmented satellite observations. The other question is, how to estimate the uncertainty related to that parameterisation. Some comments would be welcome.

Reply from authors: Our study builds on previous studies on the relations between surface properties of soil and vegetation and model parameters. We are not deriving these relationships ourselves from new field data. We don't see the spatial parametrization as a novel methodology, but more as a flexible spatial parametrization scheme, that allows us to explore the main topic of the paper, which is the use of a new spatial pattern metric and complementary observations of RS AET patterns and stream discharge for model calibration. The idea behind the parametrization is that we use data that is well described spatially, namely soil texture maps (based on a great number of measurements) and vegetation maps (based on satellite remote sensing) to distribute related model parameters spatially while still allowing for some calibration. The disadvantage would be the validity of the proposed relations between soil and vegetation properties and model parameters. But we regard it as a robust approach that avoids over simplification in spatially uniform parametrization and over-parametrization in grid-by-grid calibration.

Comment 50) Page 6, lines 10-24 How sensitive is water storage variability to this parameterisation?

Reply from authors: We don't quite understand the question, perhaps the reviewer could elaborate.

Comment 51) Page 7, lines 2-10 was the model tested on observations and what assumptions must be fulfilled?

Reply from authors: This concept is taken from Danish soil type and root depth studies (see references in manuscript) and is believed to be representative for the study area, where cereal crops are grown on soils ranging from very sandy to loamy. See also reply to reviewer 1.

Comment 52) Page 7, line 24 was this parameterisation tested on observations? What assumptions are imposed?

Reply from authors: Please see reply to comment 20) by reviewer #1.

Comment 53) Page 8, line 7 Since comparison should start from a new line and a new objective function responsible for reproducing spatial patterns should be introduced.

Reply from authors: We agree, this will be corrected.

Comment 54) Page 8, lines 10-11 histograms of what?

Reply from authors: Corrected as below. Also we will elaborate on the histogram matching as requested by reviewer #3.

"In this context, we adopted the structure of the Kling–Gupta efficiency while substituting the standard deviation term by a term based on the coefficient of variation $\left({}^{\sigma_O}/_{\sigma_S}\right)$ *and replacing the bias term with a histogram comparison index to compare the intersection-percentage of two histograms of observed and simulated spatial maps."*

Comment 55) Page 8, line 24 The AET from TSEB have been treated as error free data – a comment is needed on the possible errors involved.

Reply from authors: This is a very good point, and we have given it a lot of thought how to quantify the uncertainty. However, given the way we utilize the AET maps (bias-insensitive pattern performance) the quantitative uncertainty related to comparison with point measurements would not be very useful. What is needed is a quantification of the uncertainty of the RS AET pattern. This is far from trivial, because even if the uncertainty of some of the input to the TSEB model (mainly LST) can be approximated, it is the impact this uncertainty has on the AET pattern that is important. The crude assumption is that errors in the TSEB input are largely uniform for this relatively small catchment (e.g. if the LST input on a given day has an estimated error of +1 Kelvin, this is assumed to apply for the entire area and therefore have a limited effect on the estimated AET pattern. The whole framework is deterministic, and this

is clearly a limitation, however, the study proposes several new ideas, and to put this into a thorough uncertainty framework is beyond the scope, but will be an interesting further development. We have added a discussion on this in the discussion section of the revised manuscript.

Comment 56) Page 9, line 31 That criterion might be very misleading when the response surface is flat. The optimisation algorithm might stop in any part of the optimisation range or, most likely at the edge of the parameter range. The authors are asked for a comment.

Reply from authors: Each iteration is comprised of more than 100 runs with different parameter sets. This means that if the objective function doesn't improve after more than 500 runs then the calibration stops. This is assumed to be a reasonable number of testing in an optimization and generally we don't see parameters that stop at the edge of the parameter range.

Comment 57) Page 10, line 8 A and B and . (- B is confusing)

Reply from authors: Corrected.

Comment 58) Page 13, line 12 The existence of local minima depends on the form of the objective function which defines the parameter response surface. In the case of a model with 26 parameters the objective function will show local minima. Following the equifinality hypothesis, there are many parameter sets which give the same value of the objective function and therefore it is not surprising that many different optimum solutions can be found.

Reply from authors: We agree, we will remove this sentence, it actually does not raise questions about the SCE, but is a result of the multi-objective 26 parameter problem being solved.

Comment 59) Page 14, line 8-9 Spatial calibration constraints the solution rather than reduces its uncertainty – the uncertainty was not evaluated.

Reply from authors: Correct, we will rephrase this from reduces uncertainty to better constrained.

Comment 60) Page 15, lines 23-24 Could it be explained why the improvement occurs?

Reply from authors: We describe a drop in performance, so I guess the question is why the performance drops? And yes we have a good idea why; there was a significant drop in the number of rain gauges in Denmark in 2007-onwards, this has had a direct impact quality of the precipitation input and on the streamflow performance across the country, but we fell it is beyond the scope of this paper to go into that discussion. It is the topic of other ongoing studies.

Comment 61) Page 16, Discussion As I understood, the authors did calibration using different sequences of cases. Might it be advantageous if the optimisations Q-only and Spatial only were applied iteratively?

Reply from authors: Yes, we thought about that and given the limited trade-off and that some model parameters are mainly/only effecting either discharge or SPAEF, we would probably have reached similar results as the combined calibration.

Comment 62) Page 17, line 13 associated with .

Reply from authors: Corrected.

Comment 63) Page 17, line 32 site-specific due to .

Reply from authors: Corrected.

Comment 64) Page 17, line 34 different countries.

Reply from authors: Corrected.

Comment 65) Page 18, line 28 The conclusion on achieving a more robust parameter set because the trade-off disappears is not well founded and too general after only one validation exercise.

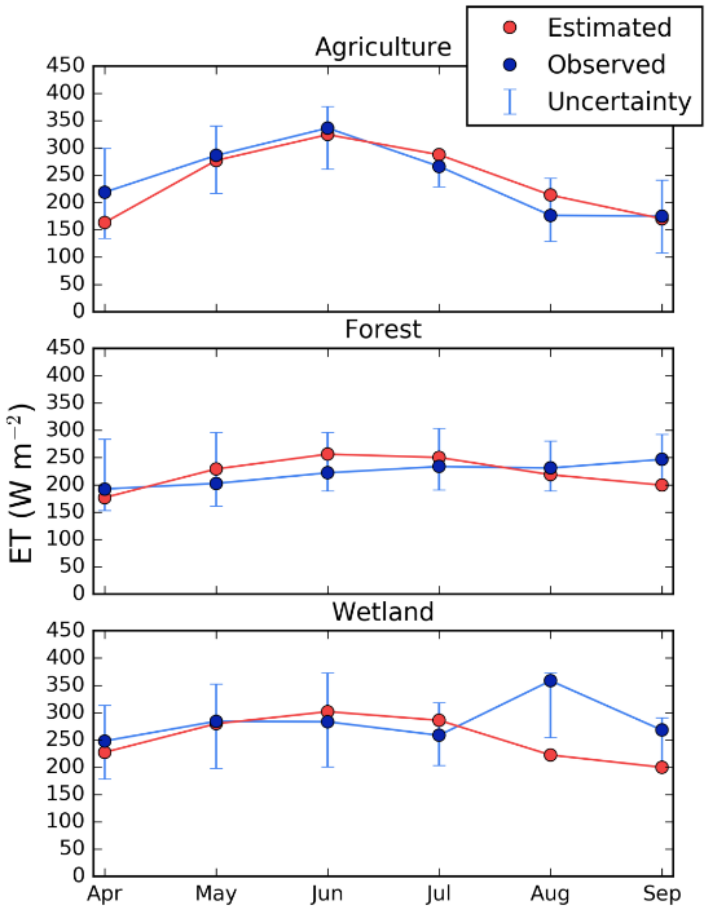
Reply from authors: We are just saying that it "suggests" that the parameter set is more robust. I don't think this is stretching the conclusion, since we actually do the validation for all nine optimized parameter sets and the performance on Discharge is equally good for the Q+spatial calibration compared to the Q-only calibration across all nine optimized parameter sets.

Comment 4) P5L6 Delete “s”

Reply from authors: “s“ is removed.

Comment 5) P5L9-11: How accurate are the AET maps, e.g. in relation to continuous EC-measurements?

Reply from authors: The accuracy of the AET maps has not been quantified in this paper. However, a comparison to observed AET for three EC sites has been included in a previous paper Mendiguren et al 2017 HESS. This comparison focussed on the monthly mean AET estimates for three land-cover types, which were well in agreement with the measurements, although it has to be noted that the measurements themselves were subject to uncertainty due to energy balance closure issues. Below is the figure that appears in the Mendiguren et al. 2017 paper



However, in the current study, the accuracy of the AET maps cannot be completely assessed by a comparison to EC measurements, because only the bias insensitive pattern information is utilized, and the uncertainty of this pattern cannot be fully described by just three EC sites. The

current study therefore relies on an assumption that when performing well for monthly means at three different sites and being mainly driven by remote sensing observations of LST and NDVI/LAI, the satellite based AET estimates contain spatial pattern information that is suitable for constraining the spatial pattern simulations of our distributed model.

Comment 6) P5L30-31: Why should this procedure accelerate model runs?

Reply from authors: The file size of the 1x1km inputs is much larger than 10 or 20 km. Therefore, reading these files to the model at each timestep takes very long time as compared to the discretisation algorithm defined inside the Fortran code. This is a technicality related to either pre-processing the LAI-based input on a daily scale prior to model execution or reading the monthly data and disaggregate to daily data inside mHM. We suggest skipping this line altogether.

Comment 7) P6L8: "stretch the spatial contrast" of what?

Reply from authors: Corrected as "spatial contrast of simulated actual evapotranspiration" in the revised version of the manuscript.

Comment 8) P6L8: ": : :based on soil and vegetation properties: : :"

Reply from authors: Corrected.

Comment 9) P6L11: Please change "domain-specific" into "site-specific" or "local"

Reply from authors: Corrected throughout the manuscript.

Comment 10) P6L11-12: According Feddes et al. (2001) and others, the root fraction coefficient is a vegetation dependent coefficient. Please add more information on why your assumption that this coefficient can be explained by field capacity is justified.

Reply from authors: It is true that generally the literature suggests that root fraction coefficient is mainly dependant on vegetation type. We also separate root fraction coefficient in two main categories: forest and non-forest. However, for the non-forest (in this study agricultural cropland, since this is the only major land cover type) we have introduced an option to distribute root fraction coefficient based on soil type. This is based on several studies carried out by Danish soil scientists who showed that for the very sandy soils of Western Denmark the effective root depth is smaller relative to soils with similar crop types, but higher clay content, although this dependency seizes at a certain clay content. Instead of basing the root fraction

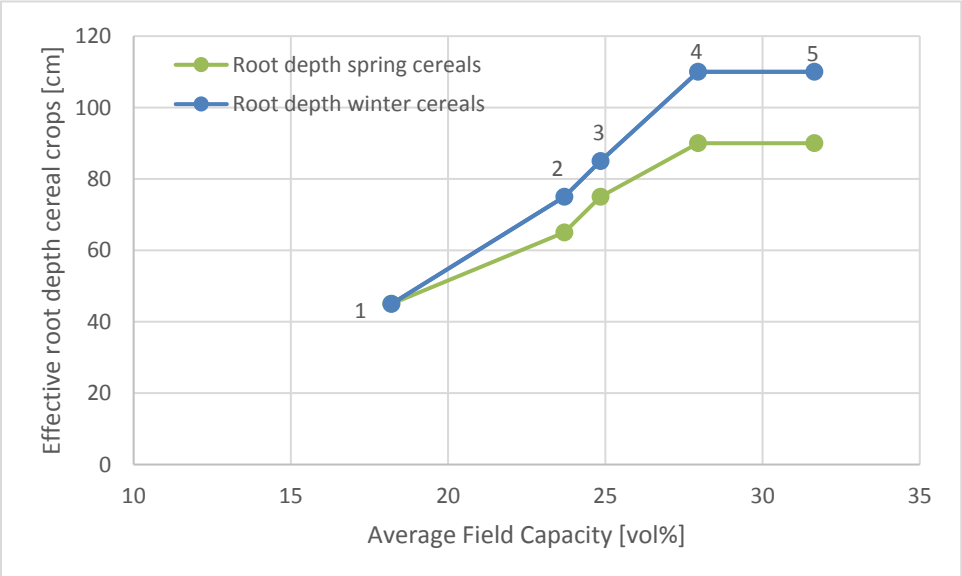
coefficient on soil type, we utilized the FC which in mHM is a function of sand and clay content through the internal pedo transfer functions. We admit that the relation might be specific to the Danish case, where even very poor (very sandy) soils are utilized for crop production and where the relation between root depth and soil type has been established for very sandy to loamy soils. A similar relation is used in the parametrisation of the Danish National Water Resources model and in National crop growth models. In the revised paper we will elaborate on this and make it clear that root depth is also a function of main landcover/vegetation types.

Comment 11) P6L14-15: At this point it unclear why these parameterizations increase the model freedom. Please reorganise the text in a more comprehensively way.

Reply from authors: We will rephrase and organise this paragraph more thoroughly in the revised version of the manuscript.

Comment 12) P6L19-24: It took me some time to understand your method, also because the assumption that the root fraction is linked to FC is counterintuitive. For a better understanding of your method, it would be helpful to graphically show the relationship between FC and rooting depth based on the soil database.

Reply from authors: We will elaborate on this, but the relation builds on previous work by others, so we can refer to their studies and illustrate the relation graphically but not plot their actual data. Here we have produced a graphical representation of the data published in Madsen and Platou 1983, but we will not include this in the paper.



The graph is a representation of the table values in Madsen and Platou 1983 for soil types 1-5 of which only soils 1-4 are present in our study area. The data represents measured average

effective root depths, which is slightly different from the root fraction coefficient in mHM (which describes the root distribution with depth), but which effectively has the same impact on AET simulations.

Comment 13) P6L22: Why do you restrict your method to pervious non-forested areas?

Reply from authors: Basically mHM operates with three land cover types: Forest, pervious and impervious (urban). The urban areas in our catchment is very limited, so we ignore that category, leaving just forest and non-forest. As explained above we acknowledge that the overall vegetation type (forest or cropland) is the dominating controlling factor for root depth, therefore we allow separate root fraction coefficients for forest and non-forest and only apply the FC dependency for non-forest.

Comment 14) P6L22: Please change "domain-specific" into "site-specific" or "local"

Reply from authors: Corrected.

Comment 15) P7L2: Why should soil properties of sandy soils impede root development?

Reply from authors: Impede is the word used in the literature we refer to on the root depth dependency on soil type. Corrected to "influence"

Comment 16) P7L2-3: Which parameters and why is a fine vertical discretization more effective?

Reply from authors: This is related to the way the root fraction coefficient effects AET simulations. When a given soil layer dries out (SM falls below FC) the simulated AET is reduced linearly until it reaches zero at wilting point. Fine vertical discretization of the upper soil layers will allow for a finer representation of the root fraction variation with depth and result in more frequent occasions where SM content reduces AET. This might be a technicality, and the sentence could be omitted, since we used the same vertical discretization through all model runs in this study and we have not examined the impact of vertical discretization.

Comment 17) P7L5-7: By changing the root fraction parameters for maximum and minimum FC, the relationship between FC and rooting depth will be changed. In the extreme case, both parameters have the same value, which means that there is no relationship at all. Did you check whether the optimised model still provides realistic relationships between FC and rooting depth? In addition, it is unclear, how you derived FC values from your digital soil map. Typically, one would sum up the horizontal FC values down to the certain soil depth (e.g. rooting depth). Please provide additional information.

Reply from authors: For the latter question, The FC is estimated in the pre-processing within mHM based on pedo transfer functions. It is a fundamental part of the model concept that only soil texture data is specified and the soil physical properties are derived. This insures a spatially consistent parametrization (given the soil texture data is good) and reduces the calibration parameters to the global pedo transfer functions parameters. We used the same soil texture parametrization (and therefore FC) for the entire soil column. We have looked at the variation in both the derived FC and WP maps and the range of root fraction parameters and they all look reasonable. In addition, we have tied the minimum and maximum root fraction coefficient so they can approach each other but never reverse. If they approach each other through calibration and become identical, that would essentially indicate that the approach did not benefit the spatial pattern performance, but that is not what we see. It clearly improves the spatial pattern performance to separate root fractions based on FC.

Comment 18) P7L16-17: You should not mention equations that have not been already introduced.

Reply from authors: Corrected.

Comment 19) P7L21-25: The section needs to be rewritten in a more comprehensible way.

Reply from authors: We will reorganise this section in the revised version of the manuscript.

Comment 20) P7L26: It is how this equation was derived and why it should be "physically meaningful". In addition, it is unclear how the DSF parameter is used to correct ET_ref.

Reply from authors: We didn't derive this equation. It is a time-space variable implementation of the crop coefficient concept suggested by Allen et.al. The basis is that the climate dependant reference evapotranspiration is given at a coarse scale of 20 km grid for a reference crop. In reality, the vegetation does not reflect a reference crop everywhere. In order to correct the reference ET a scaling factor is applied which is above 1 if the evaporative potential of the vegetation is higher than for the reference crop and below 1 if it is lower. Typical values of crop coefficients are between 0.8 and 1.2. Allen et al. 1998 and others (eg. Hunink et al. (2017)) suggested using LAI or NDVI to estimate the crop coefficient.

Our implementation is simply using the same equation in combination with remote sensing based LAI to create a time-space variable correction factor to convert ETref to ETpot.

*The DFS is simply a multiplication factor. ETpot = DFS*ETref*

mHM originally contains an even simpler correction factor, which is spatially and temporally uniform (although it also has an option to include aspect in mountainous terrain). We have omitted this correction and implicitly included it in the calibration of the DFS, because the average DFS does not necessarily add up to 1. We will be more accurate in the description of the DFS in the revised manuscript.

Comment 21) P7L28: The previous sections also belong to "Methods".

Reply from authors: We don't regard the parametrization of the mHM model as a methodology as such, we are not changing the process descriptions of the model but merely adding spatial flexibility to the parametrization of exiting model parameters. Therefore we feel that it is more appropriate to limit the methodology section to the more novel parts of the manuscript, namely the development of a new spatial performance metric and a multi-objective calibration framework based on a complimentary principle.

Comment 22) P9L1: Only accepted paper should be used a reference.

Reply from authors: We agree with the comment.

Comment 23) P9L4-9: This section can be omitted.

Reply from authors: will be corrected

Comment 24) P9L24-26: Is this statement relevant for this work or Koch et al., 2017a

Reply from authors: This statement is relevant only for Koch et al (2017) and will be omitted.

Comment 25) P9L34: Why are you using the "same" cloud-free days?

Reply from authors: This perhaps should be rephrased; we want to explain that we are making the monthly average AET maps based on simulated daily AET by averaging only the days that were also available for estimating the RS based maps (cloud free days). We have omitted the last part of the sentence, since it was confusing.

Comment 26) P10L5: Either use the Greek symbol for phi or "phi" (here and elsewhere)

Reply from authors: Corrected.

Comment 27) P10L22: Delete "and"

Reply from authors: Corrected.

Comment 28) P10L33-34: I wonder why the parameter "root fraction" for forest area is listed and not for the impervious non-forest areas, since the latter was used for the spatial parameterization.

Reply from authors: Impervious areas are very small and not represented in the land use map for root fraction distribution. Therefore, root fraction for impervious areas is not listed in Table 2.

Comment 29) P10L34: "(SPAEF column in Tab. 2)"

Reply from authors: Corrected.

Comment 30) P11L9-10: Why are you using a second model warm-up period (2005-2008)? This is also not mentioned in the text.

Reply from authors: The second warm-up period is just used to ensure that the validation results are not effected by initial conditions. The text will be updated accordingly in the revised version of the manuscript.

Comment 31) P12L18-20: However, it should be noted that the uncertainty of the spatial AET information from remote sensing is typically larger than the runoff measurements.

Reply from authors: We completely agree that the average AET estimate from RS is more uncertain that the runoff measurement. That is the reason why we use only the runoff to constrain the water balance, while we only use the bias insensitive pattern information of the RS AET to constrain the simulated pattern. It is unclear to us exactly what to correct or rephrase here.

Comment 32) P12L28: Delete "that"

Reply from authors: Corrected.

Comment 33) P13L2-3: This indicates that the spatial AET information either has large uncertainties or that it has very limitated information on the subsurface properties due to extensive irrigation in the catchment. Please discuss.

Reply from authors: We disagree, the poor performance on discharge for the spatial-only calibration is a direct consequence of the way the SPAEF objective function is designed. Since it is bias-insensitive and only contains information on the spatial pattern it cannot be used to constrain the discharge and water balance. The Spatial only calibration is (as stated in the manuscript) not meaningful from a water balance or discharge perspective, it is solely included

as a benchmark for the spatial pattern performance capability of the modelling framework and to illustrate the limited trade-offs between discharge and SPAEF when applied as proposed in the current study.

Comment 34) P15L13: Does Figure 4 present a certain year or monthly averages of several years?

Reply from authors: As mentioned in Page 5 line 6-10, Figure 4 presents averages of cloud free days for a specific month across all years for the model calibration period (2001-2008). We have clarified this in the revised manuscript.

Comment 35) P15L14: The differences seem to vary also from month to month. You should quantify the differences in AET patters, e.g. by presenting the variances or variograms.

Reply from authors: What was meant by P15L14 was simply that the calibrations including the RS AET maps result in simulated patterns that has a much better representation of the variance compared to the Q-only calibration. Table 3 and 4 present month to month differences in SPAEF which also includes the coefficient of variation. It is unclear what a specific presentation of variances across months would add to the analysis, we are open to hear more on this from the reviewer.

Comment 36) P17L21: Awkward sentence. Please reformulate.

Reply from authors: Corrected as below.

"*This is largely attributed to the nature of the metric as the spatial performance metric is bias-insensitive whereas the streamflow metrics have very little sensitivity to spatial redistribution of AET patterns as long as the spatial averages remain unchanged."*

Comment 37) P17L25: "This was: : :"

Reply from authors: Corrected as below.

"*This was done because even though the satellite based AET estimate is validated against eddy covariance stations (Mendiguren et al., 2017) they only represent specific cloud-free days limiting their value to assess the long term water balance of the catchment."*

Comment 38) P17L31-32: Unclear why the sandy soil texture should restrict the model parametrisation method.

Reply from authors: What is meant here is that the root fraction coefficient dependency on soil type might be site-specific due to the uniform land use (agricultural cropland) across soils ranging from very coarse sandy soil to loamier soils. In other words, we do not claim that this parametrisation is generic, but it reflects knowledge of the particular area and fits well with the observed patterns in soil properties and RS AET as illustrated in Figure 1. We will rephrase this sentence in a revised manuscript.

Comment 39) P617L32: Please change "domain-specific" into "site-specific" or "local"

Reply from authors: Corrected.

Comment 40) P18L23: "has proven"

Reply from authors: Corrected.

Figures and Tables:

Comment 41) Figure 1: The RS-AET scale should be reversed (red should indicate high AET values)

Reply from authors: We disagree, we have used green/blue colours for high AET and red/brown for low throughout the manuscript. To us this is the most logical colouring scheme, which intuitively symbolizes wet and dry conditions.

Comment 42) Table 1: Columns 3, 4 and 5 should be removed as they provide only limited information, which is already presented in the text.

Reply from authors: Corrected.

Additional literature

Feddes, R.A., H. Hoff, M. Bruen, T. Dawson, P. de Rosnay, P. Dirmeyer, R.B. Jackson, P. Kabat, A. Kleidon, A. Lilly, and A.J. Pitman, 2001: Modeling Root Water Uptake in Hydrological and Climate Models. Bull. Amer. Meteor. Soc., 82, 2797–2809.

Combining satellite data and appropriate objective functions for improved spatial pattern performance of a distributed hydrologic model

Mehmet C. Demirel[1, 6], Juliane Mai[2, 4], Gorka Mendiguren[1, 5], Julian Koch[1, 3], Luis Samaniego[2], Simon Stisen[1]

[1]Geological Survey of Denmark and Greenland, Øster Voldgade 10, 1350 Copenhagen, Denmark

[2]Department Computational Hydrosystems, UFZ—Helmholtz Centre for Environmental Research, Leipzig

[3]Department of Geosciences and Natural Resource Management, University of Copenhagen, Copenhagen, Denmark

[4]Department of Civil and Environmental Engineering, University of Waterloo, Waterloo, Canada

[5]Department of Environmental Engineering, Technical University of Denmark, 2800 Kgs. Lyngby, Denmark

[6]Department of Civil Engineering, Istanbul Technical University, 34469 Maslak, Istanbul, Turkey

Correspondence to: Simon Stisen (sst@geus.dk)

Abstract. Satellite based earth observations offer great opportunities to improve spatial model predictions by means of spatial pattern oriented model evaluations. In this study, observed spatial patterns of actual evapotranspiration (AET) are utilized for spatial model calibration tailored to target the pattern performance of the model. The proposed calibration framework combines temporally aggregated observed spatial patterns with a new spatial performance metric and a flexible spatial parameterisation scheme. The mesoscale Hydrologic Model (mHM) is used to simulate streamflow and AET and has been selected due to its soil parameter distribution approach based on pedo-transfer functions and the build in multiscale parameter regionalization. In addition two new ~~domain specific~~ spatial parameter distribution options have been incorporated in the model in order to increase the flexibility of root fraction coefficient and potential evapotranspiration correction parameterisations, based on soil type and vegetation density. These parametrisations are utilized as they are most relevant for simulated AET patterns from the hydrologic model. Due to the fundamental challenges encountered when evaluating spatial pattern performance using standard metrics, we developed a simple but highly discriminative spatial metric i.e. comprised of three easily interpretable components measuring co-location, variation and distribution of the spatial data.

The study shows that with flexible spatial model parameterisation used in combination with the appropriate objective functions, the simulated spatial patterns of actual evapotranspiration become substantially more similar to the satellite based estimates. Overall 26 parameters are identified for calibration through a sequential screening approach based on a combination of streamflow and spatial pattern metrics. The robustness of the calibrations is tested using an ensemble of nine calibrations based on different seed numbers using the shuffled complex evolution optimizer. The calibration results reveal ~~a limited trade-offs~~a limited trade-off between streamflow dynamics and spatial patterns illustrating the benefit of combining separate observation types and objective functions. At the same time, the simulated spatial patterns of AET significantly improved when including an objective function based on observed AET patterns and a novel spatial performance metric compared to traditional streamflow only calibration. Since the overall water balance is usually a crucial goal in the hydrologic modelling, spatial pattern oriented optimization should always be accompanied by traditional discharge measurements. In such a multi-objective framework, the current study promotes the use of a novel bias-insensitive spatial pattern metric, which exploits the key information contained in the observed patterns while allowing the water balance to be informed by discharge observations.

1 Introduction

Reliable estimations of spatially-distributed actual evapotranspiration (AET) are useful for various sustainable water resources management practices such as irrigation planning, agricultural drought monitoring and water demand forecasting in large cultivated areas (Wei et al., 2017). Distributed hydrologic models can potentially provide this insight since ET is a major part of the water cycle. In spite of their ability to simulate detailed spatial patterns of a range of hydrological state variable and fluxes, distributed model evaluation remains focused on temporal aspects of the aggregated streamflow variable (Demirel et al., 2013; Schumann et al., 2013). We are interested in including spatial AET patterns in the model calibration using spatial parameterisations and complementary objective functions. Different methods exist that utilize satellite based land surface temperature data to derive spatially detailed estimates of latent heat fluxes from land-surface and canopy at a scale relevant for catchment modelling (Kalma et al., 2008). Since AET cannot be measured directly by satellite, surface energy balance models are developed to estimate AET based on data from a range of spectral and thermal bands (Guzinski et al., 2013; Norman et al., 1995; Su, 2002). While these satellite-based estimates are usually employed as a tool to understand and improve the model parameterisations (Conradt et al., 2013; Hunink et al., 2017; Schuurmans et al., 2011), they can also be used to calibrate models (Crow et al., 2003; Immerzeel and Droogers, 2008; Zhang et al., 2009). Therefore, adding satellite based observations to model calibration is not novel; however, specifically evaluating spatial patterns in the calibration has been rarely done (Stisen et al., 2011b). Interesting examples exist where model calibration could benefit from the spatial pattern information of actual evapotranspiration (Githui et al., 2016; Li et al., 2009; Zhang et al., 2009) and satellite_based_recharge patterns (Hendricks Franssen et al., 2008). This paper utilizes monthly patterns of AET first to understand and organize ET related model spatial parameterisations then to pursue a calibration. This is ~~from the fact that~~because adding only temporal aspect of the spatial observations to the objective function is not sufficient for achieving significant improvements in simulated spatial patterns if model parameterisation is not flexible enough to physically adjust to the observed pattern. Besides, the model structure, parameterisations and calibration schemes have usually been designed for streamflow optimizations (Vazquez et al., 2011; Velázquez et al., 2010). In order to ensure compatibility between the spatial pattern calibration target and model parameterisation, the flexibility of the spatial model parameterisation needs to be reconsidered. Recently, inadequate representation of spatial variability and hydrologic connectivity of a well-known distributed model (VIC) has been reported by Melsen et al. (2016). The mesoscale Hydrologic Model (mHM) has the flexibility to alter the spatial patterns via pedo-transfer function (PTF) parameters and by including a multi-scale parameter regionalization (MPR) scheme (Kumar et al., 2013; Samaniego et al., 2010). Mizukami et al. (2017) incorporated this MPR approach with VIC to estimate parameters for large domain based on geophysical data for 531 basins. The multi-basin calibration results using MPR revealed physically meaningful parameter fields without patchiness (discontinuities). The study by Loosvelt (2013) is one of few other examples that incorporate PTFs for soil texture and moisture components of a hydrologic model.

All calibration strategies rely on the selection of performance metrics indicating the goodness-of-fit of the model to be optimized. Choosing an appropriate set of objective functions is crucial to build a robust calibration strategy since there will be trade-offs between different objective functions or redundant information. In the hydrology literature, there are a range of different temporal metrics for hydrograph match while metrics designed for spatial pattern matching are less common (Koch et al., 2017; Rees, 2008). For distributed models, spatial metrics usually evaluate cell-to-cell correlation and deviations (e.g. Pearson's R and bias). The use of multi-component metrics as described for discharge by Gupta et al. (2009) is however rare for spatial pattern evaluation. An essential feature of our study is introducing a new spatial efficiency (SPAEF) metric that contains three components i.e. correlation, variance and histogram intersection providing reliable bias-insensitive pattern information unlike other traditional metrics focusing on only one aspect like correlation, mean squared error or bias.

Prior to model calibration, sensitivity analysis is usually conducted to attribute response of the model outputs to the changes in model parameters (Shin et al., 2013), which can enhance our understanding of both temporal and spatial model behaviour (Berezowski et al., 2015). In the context of spatial model calibration, the sensitivity analysis should not only identify the

parameters that affect the water balance and hydrograph dynamics but also the parameters that shape the spatial patterns of the simulated states and fluxes. To achieve this, we have to design objective functions that reflect the spatial pattern of the models and utilize these in model parameter sensitivity analysis.

In light of the well-known equifinality problems in model calibration (Beven and Freer, 2001) spatial pattern evaluation can be useful for selecting the most appropriate parameter set from a group of sets leading to both reasonable streamflow performance and physically meaningful AET pattern. Immerzeel and Droogers (2008) showed how a semi distributed model of a basin in Southern India could be constrained by using spatially distributed observations with a monthly temporal resolution~~showed that the models can be constrained by using spatially distributed observations with a monthly temporal resolution~~. Cornelissen et al. (2016) highlighted the need to identify which model parameters influence the simulated spatial pattern and showed that spatial patterns of simulated evapotranspiration were most sensitive to the land-use parameterisation, whereas precipitation was the most sensitive input data with respect to temporal dynamics of the model. Rakovec et al. (2016) used total water storage (TWS) anomaly from the Gravity Recovery and Climate Experiment (GRACE) satellites and evapotranspiration estimates from FLUXNET data (https://fluxnet.ornl.gov/) to improve model parameterisations for discharge simulations. They showed that adding TWS anomalies to the calibration performed reasonably well for continental 83 European basins with different climatology.

The main objectives of this study are to ~~1) determine appropriatespatial model parameterisations and objective functions for the domain and 2)~~ incorporate spatial patterns of satellite based actual evapotranspiration data in the model calibration and validation. ~~For that we investigate the ET related model parameterisations and evaluate both temporal dynamics of streamflow and spatial patterns of AET over the basin.~~ In order to improve AET simulations, we use transfer functions in the spatial model parameterisation that combine a priori maps of soil and vegetation properties with few global calibration parameters in order to enhance the spatial parameterisation flexibility and allow the parameter field to adjust to an observed spatial patterns of AET from the catchment. We also design a new multi-component metric specifically suited for comparing ~~two~~ spatial patterns of two continuous variables.~~This requires bias insensitive componentsevaluating different crucial characteristics of the spatial data~~. Here, we prioritize three main data properties, which are co-location, variation and distribution. ~~In addition to the streamflow observations we use remote sensing AET data to calibrate and evaluate the distributed model for the second research objective.Prior to calibration, important parameters are identified using a sensitivity analysis including both streamflow dynamics and spatial patterns of AET as objective functions. We then use a well-known global search algorithm of SCE-UA (Duan et al., 1992) to calibrate the mHM hydrologic model.~~ The calibration is conducted using three strategies for objective function selection. First, streamflow metrics and spatial pattern metrics are used in isolation during calibration and subsequently they are combined in a more balanced model optimization. In this way we can investigate the trade-offs and robustness of the different approaches by evaluation the performances regarding both streamflow and spatial patterns during calibration and validation.

2 Study Area and Data

2.1 Study area

The Skjern river basin is one of the most popular research basins in Denmark as it is highly instrumented for hydrological monitoring including eddy-flux towers, a dense soil moisture network and other state-of-the-art monitoring of hydrological variables (Jensen and Illangasekare, 2011). The basin area is approximately 2500 km^2 containing mostly sandy soils (Figure 1~~Figure 1Figure 1~~). The river is the largest in Denmark by flow volume and located in the western part of the Jutland peninsula, a region dominated by agriculture and forests together covering ~80% of the domain (Larsen et al., 2016). The basin is mostly flat with a maximum altitude of 130 m and it receives a mean annual precipitation of around 1000 mm (Stisen et al., 2011a). The mean annual streamflow is around 475 mm and monthly mean temperatures varies from 2 up to 17°C (Jensen and

Illangasekare, 2011).

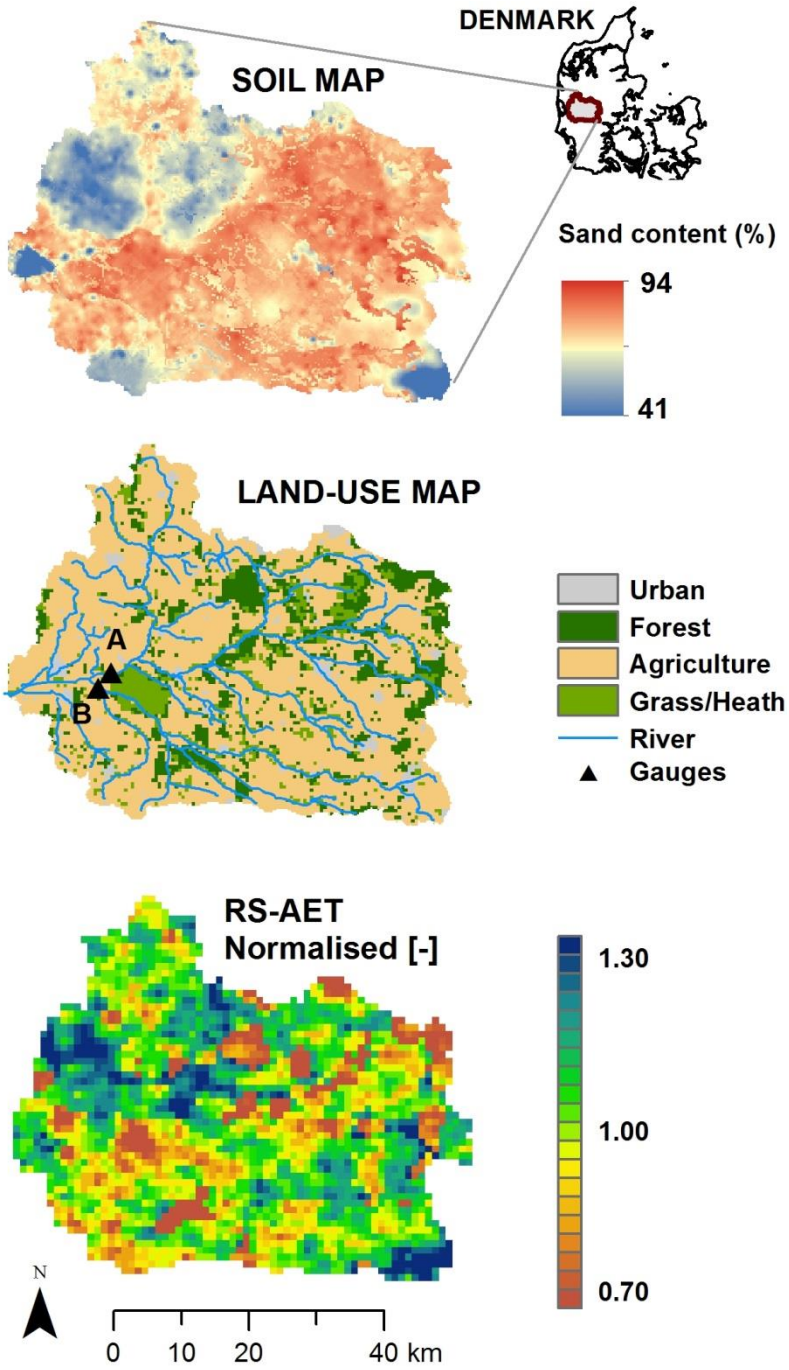


Figure 1Skjern River Basin location, soil type and land-use characteristics. An average pattern of satellite based actual evapotranspiration for June (average of all years from 2001 until 2008) is ~~also~~ presented to illustrate the interaction between soil type and land-use that generate the land surface flux patterns.

2.2 Satellite based data

The Moderate Resolution Imaging Spectroradiometer (MODIS), polar orbiting platforms, Terra and Aqua, observe mid-latitude regions four times per day at a spatial resolution of approximately 1x1km. The Two Source Energy Balance (TSEB) model proposed by Norman et al. (Norman et al., 1995) based on the Priestly-Taylor approximation (Priestley and Taylor, 1972) is used in this study to calculate AET based on MODIS data under cloud-free conditions. The model inputs are land surface temperature (LST), solar zenith angle (SZA), as well as albedo and height of canopy all derived from MODIS observations (Mendiguren et al., 2017). Additional inputs such as climate variables of air temperature and incoming radiation

are obtained from ERA interim reanalysis data (Dee et al., 2011). The main motivation of preparing a new AET dataset based on land surface temperature is that most other available products are based mainly on vegetation index data which may not be sufficient to assess the complicated interplay between climate, soil and vegetation dynamics on the AET patterns especially during the growing season. For more details of our newly produced AET data for Denmark, including equations, parameterisation, calibration and validation, please refer to the recent study by Mendiguren et al. (2017).

In this study, all remote sensing based AET data were averaged for each month ~~s~~ during the growing season across all years for the model calibration period (2001-2008) resulting in~~to~~ six monthly mean maps from April to September representing AET under cloud-free condition. This ensures that in spite of uncertainty in the individual instantaneous midday estimates of AET, the monthly maps represent the general spatial pattern for each month under cloud-free conditions. The individual daily AET patterns are evaluated for temporal consistency by calculating the Pearson correlation between each daily pattern and the monthly mean pattern for the given month. This analysis showed that the overall average correlation between an individual day and the monthly mean was 0.82.The satellite based monthly AET maps are validated against eddy-covariance measurements for three different land cover types, forest, cropland and wetland within the Skjern catchment and display good agreement on the monthly timescale (Mendiguren et al., 2017). Despite not being pure observations but estimates from an energy balance model based on satellite observations we will refer to these AET maps as reference observations. Based on the sensitivity analysis in (Mendiguren et al., 2017), which showed that the TSEB is largely controlled by the satellite input of LST, which is can be considered an observation, it is assumed that the TSEB AET estimates represent spatial patterns of AET that are suitable for pattern evaluation of the hydrological model.

3 Hydrologic Model

The mesoscale Hydrologic Model (mHM) is a distributed model providing various simulated spatial outputs, fluxes and states at different spatiotemporal model resolutions (Samaniego et al., 2010). The model includes pedo-transfer functions for soil parameterisation and originally contains 53 global parameters that can be adjusted during calibration. In this study, some parameters are fixed at a default value and others have been added from the new spatial model parameterisations resulting in a total of 48 global parameters for further analysis. The model simulates major components of the hydrologic cycle i.e. interception, infiltration, snow accumulation and melting, evapotranspiration, groundwater storage, seepage and runoff generation. The readers are referred to the study by Samaniego et al. (2010) for full model description, assumptions, limitations and process formulations.

Table 1~~Table 1~~Table 1 provides a summary of the modelling data used in this study. As shown in the table, meteorological data can be at a different spatial scale than both morphological data and the model scale. This flexibility arises from the fact that mHM incorporates a multi-parameter regionalization technique to swap between different scales while calculating all fluxes and routing streamflow at a preferred model scale. We run the model at 1x1 km spatial scale and daily time step. Some processes like ET are calculated at hourly time step then the final results are aggregated to daily values. All morphological data are prepared at 250x250 m scale. All three meteorological data sets, i.e. P, ET_{ref} and T_{avg}, were originally at 10-20 km resolution. We re-sampled them to 1x1 km using cubic interpolation. This interpolation method is used to avoid patchiness in model simulations due to coarse grids at the native scale of the metrological data. We use 12 monthly LAI maps to represent the climatology for both interception and PET correction for the entire period (2001-2014) and the model warm-up period (1997-2000). ~~There is a disaggregation module in mHM model transforming 12 LAI monthly values to daily values which accelerates the model runs.~~

Table 1 Overview of morphological and meteorological data used as input for mHM. Acronyms: BIOS: BioScience Aarhus University, DMI: Danish Meteorological Institute, GEUS: Geological Survey of Denmark and Greenland, MODIS: Moderate Resolution Imaging Spectroradiometer, DGA: Danish Geodata Agency

Variable	Description	Spatial Resolution	Source
Q (daily)	Streamflow	Point	BIOS
P (daily)	Precipitation	10 km	DMI
ET_{ref} (daily)	Reference evapotranspiration	20 km	GEUS and DMI
T_{avg} (daily)	Average air temperature	20 km	GEUS and DMI
LAI	Fully distributed 12 monthly values based on 8-day time varying Leaf Area Index (LAI) dataset	1 km	MODIS and Mendiguren~~*et al.*~~ *et al.* (2017).
Land cover	Forest, ~~pervious~~ agriculture and urban	250 m	GEUS
DEM related data	Slope, aspect, flow accumulation and direction	250 m	DGA
Geology class	Two main geological formations	250 m	GEUS
Soil class	Fully distributed soil texture data	250 m	Greve et al. (2007)

3.1 Spatial model parameterisation

In order to facilitate a meaningful spatial pattern oriented calibration of a distributed model, we need to compromise between comprehensive (each cell in the basin) and lumped (one cell - one basin) parameterisations as the first approach may require an immense computer resource during calibration and the latter approach usually results in a uniform pattern. For instance, in a detailed calibration study by Corbari et al. (2013),each pixel in the catchment is represented by a parameter whereas, in a coarse parameterisation, a uniform parameter represents the entire catchment (Stisen et al., 2017). In this study, we follow an intermediate level of parameterisation comprised of several flexible spatial parameters and nonlinear equations allowing us to stretch the spatial contrast of simulated actual evapotranspiration based on soil and vegetation ~~in the catchment~~properties. This level of parameterisation is still physically meaningful as the parameters are tied to the land surface characteristics of the basin via transfer functions.

Distributed root fraction coefficient

Root distribution with depth is generally perceived as being a function of vegetation type (Jackson et al., 1996), and our spatial parametrization of root fraction distribution is initially separated based on land covers of forest and agricultural crops. However, ~~F~~following the ~~domain~~site-specific soil and plant~~geo~~ physical literature (Jensen et al., 2001; Madsen and Platou, 1983),we subdivide the root fraction coefficient for agricultural crops ~~we define the root fraction coefficient~~ as a function of field capacity (FC).~~Similarly followingthe concept of crop coefficient (Allen et al., 1998),we define the PET correctionfactor as a function of LAI to represent the vegetation more realistically than a uniform correction factor based on the aspect ratio.These two new spatial model parametrizations areused to increasethe realization capability of the model by increasing the model freedom.~~Here, spatial model parameterisation is implemented to the root fraction calculation in the original mHM

structure which follows the asymptotic equation for vertical root distribution (Eq.(1~~(1(1)~~)) proposed by Jackson ~~et al~~ et al. (1996).

$$Y = 1 - (\beta_c)^d \quad (1)$$

Where Y is the cumulative root fraction from soil surface to depth d (cm), and β_c is the root fraction coefficient. We substituted the root fraction coefficient for ~~pervious areas (non-forest)~~agricultural crops (non-forest) with two new root fraction parameters i.e. one root fraction for maximum FC (clay) and one for minimum FC (sand) which allow for full spatial distribution of root fraction with varying FC. This relation between soil characteristics and effective rooting depth is based on a site~~domain~~-specific database with more than 100 soil and root profiles collected in Denmark (Table 19.4 in Jensen et al., 2001) and the literature focusing on soil texture and effective rooting depths in Denmark (Madsen, 1985, 1986; Madsen and Platou, 1983).The approach is not necessarily globally valid, but designed to the specific region of Western Denmark where very sandy soils (Figure 1) are cultivated for agricultural purposes even though the soil properties ~~impede~~ influence root development.~~The two newly introduced parameters are more effective when fine vertical discretization of soil layers is applied.~~ These parameters are used to form the root fraction coefficient for ~~pervious~~ soil with agriculture ($\beta_{\text{agriculture}}$~~$_{\text{pervious}}$~~) based on field capacity dependent root fraction at Eq.(2~~(2(2~~) and (3~~(3(3~~).

$$FC_{norm} = \frac{FC_i - FC_{min}}{FC_{max} - FC_{min}} \quad (2)$$

whereFC_{norm} is the normalized field capacity ranging from 0 to 1.

$$\beta_{agriculture} = (FC_{norm} * \beta_{max}) + (1 - FC_{norm}) * \beta_{min} \quad (3)$$

Where $\beta_{agriculture}$~~$_{pervious}$~~ is the new root fraction for ~~pervious~~ soil with agriculture comprised of root fraction for clay (β_{max}) and root fraction for sand (β_{min}).

Dynamic ET_{ref} scaling function

As a second spatial ~~parametrization~~parameterisation step, we incorporated remotely sensed vegetation information, to downscale coarse climatological reference evapotranspiration (ET_{ref}) to the model scale. This was done to emphasize the effect of vegetation on the simulated spatial patterns of AET. The original scaling factor in mHM is based on a lumped minimum correction and an aspect driven additional term. Using aspect ratio for ET_{ref} correction makes sense in mountainous areas; ~~however~~however, this is found irrelevant for the Skjern basin which is characterized by a low topographical variation. The dynamic scaling function introduced here allows the modeller to superimpose the imprint of LAI on the simulated AET patterns via a downscaling of the ET_{ref}. The concept of dynamic scaling function (DSF)~~,given in equation(4)~~, is similar to the concept of a crop coefficient used to convert ET_{ref} to a potential evapotranspiration (ET_{pot}) for a given vegetation that differs from the reference crop. ~~as~~O~~o~~ur implementation follows the equation for estimating the crop coefficient for natural vegetation originally proposed by Allen ~~et al~~ et al. (1998). Similarly, Hunink et al. (2017) compared different applications of crop coefficients based on remotely sensed vegetation indices in hydrologic modelling. They found that the effect of crop coefficient parameterisations on the water balance is trivial and constant throughout the year; however, it has a major effect on seasonal evapotranspiration and soil moisture fluxes showing the ~~crucial~~ role of crop coefficient for spatial calibration. The DSF, shown at Eq. (6~~(65~~), is simply a time-space variable implementation of the crop coefficient for natural vegetation ~~(Eq. (4))~~, parameterized through ~~has three dimensionless parameters that can be calibrated and~~ a spatio-temporal LAI (-) component ~~to~~ accounting for the effects of characteristics that ~~distinguish~~ separate the actual vegetation from a reference grass (well-watered 10 cm height and having albedo of 0.23). These characteristics include specific land cover, albedo and aerodynamic resistance (Allen et al., 1998; Liu et al., 2017). This ensures a physically meaningful downscaling from a coarse (here 20 km) ET_{ref} grid to the model resolution (here 1 km).

$$ET_{pot} = DFS \cdot ET_{ref} \quad (4)$$

~~$$DSF = a + b\left(1 - e^{(-c \cdot LAI)}\right)$$~~

$$DSF = a + b\left(1 - e^{(-c \cdot LAI)}\right) \quad (5~~3~~)$$

where a in the model (ET_{ref}-a) is the intercept term representing uniform scaling, b (ET_{ref}-b) represents the vegetation dependent component while c (ET_{ref}-c) describes the degree of nonlinearity in the LAI dependency.

~~These two new spatial model parametrizations are used to increase the realization capability of the model by increasing the model freedom.~~

4 Methods

In this study, we applied a recently developed sequential screening method (Cuntz et al., 2015) to select important parameters for calibration. Since different parameters can be sensitive to different hydrologic processes, we tested three different performance metrics to evaluate process-parameter relationships. Two of these metrics are derived from the hydrograph i.e. Kling-Gupta Efficiency (KGE, Gupta et al. (2009)) and KGE of only below average streamflow (KGE_{low}) whereas the spatial efficiency metric focuses on the spatial pattern of actual evapotranspiration.

4.1 Objective functions

As an objective function for streamflow performance, we chose the Kling–Gupta efficiency, shown at Eq. (~~5~~6), (KGE; Kling and Gupta, 2009) and applied it to both the entire time series and to the low flow part of the hydrograph (below mean discharge).

$$KGE = 1 - \sqrt{(\alpha_Q - 1)^2 + (\beta_Q - 1)^2 + (\gamma_Q - 1)^2} \quad (6~~5~~)$$

$$\alpha_Q = \rho(S, O) \text{ and} \beta_Q = {}^{\sigma_O}/_{\sigma_S} \text{ and } \gamma_Q = \frac{(\mu_S - \mu_O)}{\sigma_O}$$

where α_Q is the Pearson correlation coefficient between observed and simulated discharge time series, β_Q is the relative variability based on the fraction of standard deviation in simulated and in observed valuesand γ_Q is the bias term normalized by the standard deviation in the observed data.

Since comparison of two spatial pattern maps is of obvious importance, a bias-insensitive spatial performance metric is developed and used in this study. In this context, we adopted the structure of the Kling–Gupta efficiency while substituting the standard deviation term by a term based on the coefficient of variation

~~In this context, we adopted the structure of the Kling–Gupta efficiency while substituting the standard deviation term by a term based on the coefficient of variation~~$\left({}^{\sigma_O}/_{\sigma_S}\right)$ and replacing the bias term with a histogram comparison index to compare the intersection-percentage of two histograms of observed and simulated spatial maps. The histogram intersect is performed after normalization of the observed and simulated maps to a mean of 0 and standard deviation of 1 (z-score). This ensures that the histogram comparison is unaffected by any bias or variance differences and solely reflects the agreement in distribution of the variable in space.~~and replacing the bias term with a histogram comparison index tocompare the intersection-percentage of two histograms i.e. observed and simulated histograms~~. The main utility of the histogram comparison is that it distinguishes between different soil and vegetation groups reflected in the spatial pattern results. This unique feature of being sensitive to

clusters in the data compliments the other two components in the equation, in particular the correlation coefficient (α in Eq.(7~~(7~~(6)) since α is highly vulnerable to very distinct clusters of points aligned on a diagonal axis. This can result in high correlation coefficient values in spite of low correlation inside the individual clusters inevitably misleading the model calibration. The separated clusters often occur in environmental models where different land-use classes and soil classes etc. can produce patchy spatial patterns. The new spatial efficiency metric (optimal value equals to 1) is defined as:

$$SPAEF = 1 - \sqrt{(\alpha - 1)^2 + (\beta - 1)^2 + (\gamma - 1)^2} \quad (7~~7~~6)$$

$$\alpha = \rho(A, B) \text{ and} \beta = \left(\frac{\sigma_A}{\mu_A}\right) \Big/ \left(\frac{\sigma_B}{\mu_B}\right) \text{ and } \gamma = \frac{\sum_{j=1}^{n} min(K_j, L_j)}{\sum_{j=1}^{n} K_j}$$

where α is the Pearson correlation coefficient between observed AET map (A) and simulated AET map (B) for a particular month, β is the fraction of coefficient of variations representing spatial variability and γ is the percentage of histogram intersection (Swain and Ballard, 1991). The gamma (γ)is calculatedfor a given histogram *K* of the observed AET map (A) and the histogram *L* of the model simulated AET map (B), each one containing *n* bins i.e. herein 100 bins. The maps are standardized to a mean of 0 and a standard deviation equal to 1 (zscore) to avoid the effect of different units. In this study, we compare AET from TSEB (in W/m^2) based on instantaneous satellite data with daily averaged AET (mm/day) simulated by the model, and regard the satellite based AET maps as the "observation" even though they are more accurately AET "estimates" based on satellite observations. ~~It should be noted that we also examined whether we could improve the discriminative skill of the SPAEF using the slope of the QQ plot instead of histogram intersection~~. ~~However, all~~A ~~a~~ttempts to use ~~and~~ numerous other spatial metrics including Mapcurves, FSS, Goodman and Kruskal's lambda, Theil's Uncertainty, EOF and Cramér's V (Cramér, 1946; Koch et al., 2015; Rees, 2008)did not distinguish the general AET patterns as well as the spatial efficiency (SPAEF) metric. The strength of the spatial efficiency metric is that each component contains different and non-overlapping information. Moreover, the components are straightforward as compared to the aforementioned metrics. While the correlation term (α) expresses only the spatial correlation of AET values, the coefficient of variation term (β) expresses only the range/contrast in the image while the histogram term (γ) only expresses the agreement on histogram shape without considering either variation or correlation. Since all three terms are bias-insensitive, the spatial efficiency only constrains the model simulations with the pattern information in the satellite data while leaving the water balance (bias) to be constrained by streamflow metrics.~~The readers are referred to our subsequent study by Koch et al.(2017a) where weelaborate and rigorously analyseevery component of this metric as well as compare it with other spatial metrics.~~

4.2 Sequential screening of the model parameters

We applied the variance-based sequential screening (SS) method ~~based on Morris elementary effects. This method was firstly~~ introduced by Cuntz et al. (2015)to identify, with a low computational budget, the parameters which are most informative regarding a certain model output M. ~~The method can be understood as a pre processor to different kinds of many query applications such as model calibration, Sobol sensitivity analysis and parameter uncertainty estimation. These methods can then be performed using only the set of informative parameters while the uninformative parameters are discarded and fixed at a default value.~~

For this approach the parameters are sampled in trajectories as initially described by Morris (1991) and improved by Campolongo et al. (2007). Each trajectory consists of (N+1) parameter sets assuming that N is the total number of model parameters. The first parameter set in each trajectory is sampled randomly while all the subsequent sets *i (i>1)* differ to the prior set (*i-1*) in exactly one parameter value. Therefore, the whole trajectory is a path through the parameter space. Trajectories allow us to sample the whole parameter space efficiently and consider parameter interactions to certain extend. In the approach of Cuntz et al. (2015), only a small number (M_1) of such trajectories are sampled to lower the computational burden. The resulting (M_1 x (N+1)) model outputs are derived and the elementary effects (EE) are computed for each parameter. The EE's

are then used to identify the most informative parameters by deriving a threshold splitting the parameters into a set N_u of uninformative and a set N_i of informative ones. In the following, the first parameter set is again sampled randomly but then only the uninformative parameters are perturbed meaning that the new trajectory only ~~consist~~consists of (N_u+1) parameter sets. The derivation of model output and calculation of EE's is repeated. The major step is to determine whether one of the previously uninformative parameters is now above the threshold and if so it is added to the set of informative parameters N_i. These steps are repeated until no further parameter is added to the set N_i. At the end M_2 trajectories are sampled to confirm that the set of uninformative parameters N_u is stable and no further parameter would be found to be informative.~~Also a combination of Latin hypercube global sampling strategy and a local sensitivity method (van Griensven et al., 2006) is tested tofurther reduce the number of effective parameters for our subsequent study by Koch et al. (2017a).~~

4.3 Model calibration and validation

We calibrated the 1 km-daily mHM for the Skjern basin in Denmark using the well-known global search algorithm Shuffled Complex Evolution University of Arizona (SCE-UA) (Duan et al., 1992). The SCE-UA algorithm is configured with two complexes running in parallel with 53 (2n+1) parameter sets in each complex and 27 (n+1) parameter sets per sub-complex. Moreover, the maximum relative objective function change is set to 1% over five iterations as the model convergence criterion. This criterion was usually reached after 3500 runs; in rare cases up to 8000 runs were necessary. We evaluated the differences between monthly AET estimates from the TSEB reference data and simulated AET from the hydrologic model for the calibration period (2001-2008) and validation period (2009-2014)~~using the same cloud-free days in summer~~.

The two streamflow stations are defined separately to follow the improvements in each metric throughout the calibrations. After testing different combinations of streamflow and spatial metrics, we chose two streamflow metrics (KGE and KGE_{low}) and one spatial efficiency metric given by Eq. (~~5~~6) and (~~6~~7), respectively. These objective functions are used individually or combined in three model calibration cases based on (1) only streamflow using equally-weighted KGE and KGE_{low}, (2) only spatial patterns of AET using spatial efficiency, (3) both equally-weighted streamflow and spatial pattern match using all three metrics. It should be noted that the case 2 calibration is designed as a benchmark to explore how good the pattern match can get when not considering streamflow performance, even though the solution might not be interesting from a hydrological perspective, since the bias insensitive spatial pattern metric does not secure a reasonable water balance. To test the overall robustness of the calibration framework we use an ensemble of nine calibrations for case 1 and nine calibrations for case 3 each started from a different seed number. In order to fairly weigh the objective functions, we retrieve the residuals (ε) from the three objective functions based on a random initial model run (Eq. ~~7~~8, ~~8~~9 and ~~9~~10).We calculate the new weights which will result in equal contribution to the total error ~~phi$_{total}$~~(Φ_{total}) i.e. 50% from spatial metric and 50% from the two streamflow metrics. Ideally if it exists the optimizer searches a parameter set resulting in zero Φ_{total}~~phi$_{total}$~~ otherwise the closest point to zero will be considered as optimum solution.

$$\Phi_Q = \sum_{i=1}^{2}\left(\varepsilon_{KGE_i} * \omega_{KGE_i}\right)^2 + \sum_{i=1}^{2}\left(\varepsilon_{KGE_{low,i}} * \omega_{KGE_{low,i}}\right)^2 \quad (8\text{~~7~~}) $$

$$\Phi_{Spatial} = \sum_{m=1}^{6}\left(\varepsilon_{SPAEF_m} * \omega_{SPAEF_m}\right)^2 \quad (9\text{~~8~~}) $$

$$\Phi_{total} = \Phi_Q + \Phi_{Spatial} \quad (10\text{~~9~~}) $$

Where Φ_Q is the total ~~Φphi~~ for streamflow of the two stream flow gauges ~~A and B~~and $\Phi_{Spatial}$ is the total ~~Φphi~~ for spatial performances of six summer months. For Q-only calibration, the weight for SPAEF (ω_{SPAEF}) becomes zero whereas for spatial-only calibration the weights for KGE and KGE_{low} become zero.

5 Results

5.1 Sequential screening of the model parameters

Table 2~~Table 2~~~~Table 2~~ shows the sequential screening results based on KGE, KGE_{low} and SPAEF, respectively. Each objective function reflects on different spatio-temporal dynamics of the catchment. While KGE and KGE_{low} evaluate high and low streamflow dynamics and biases, the bias-insensitive SPAEF focuses on only spatial patterns of AET. From the results it is clear that some of the highly sensitive parameters for streamflow dynamics especially interflow-related parameters, groundwater-related geology parameters and single routing parameter have minor to zero influence on the spatial patterns of AET. The new ET parameters, ET_{ref}–a~~p~~ (non-forest~~pervious~~), –af (forest), –b and –c are identified to be informative based on all objective functions. The root fraction coefficient for forest (rotfrcoffore) appeared to be not very important for streamflow metrics whereas it is crucial for SPAEF. ~~Similarly~~Similarly, the two newly introduced parameters i.e. root fraction coefficient for sand and clay (i.e. rotfrcofs and rotfrcofclay) soil are informative based on all three objective functions. Organic matter for forest (orgmatforest) is especially important for low flows whereas ~~and~~ organic matter for impervious areas (orgmatimper) has zero influence on spatial patterns of AET. The exponent slow interflow (expslwintflw) parameter is found to be most informative for low flows while recharge coefficient (rechargcoef) is most informative for streamflow and ET_{ref}–af is most informative for calibrating spatial patterns of AET.

On average 475 model evaluations are required to split the total number of 48 parameters into informative and uninformative ones. However, the number of iterations is dependent on objective function, therefore; 449 model runs were required for KGE, 431 model runs for KGE_{low} and 544 model runs for SPAEF. This is in close agreement with the computational budget of 10N model evaluations already reported by Cuntz et al. (2015). This also makes the sequential screening method computationally very attractive compared to other global search methods. However, the computational advantage is at the cost of exploring a larger part of the parameter space, hence the sequential screening is mostly valuable for identifying informative/non-informative parameters prior to calibration or further assessment of the parameter behaviour. Overall, these results show that there are 26 parameters above the threshold of 1% of at least one case (Table 2~~Table 2~~~~Table 2~~). In principal the parameters with zero sensitivity (SPAEF column) can be fixed at some value during calibration which may lead to faster convergence with lower degree of freedom. However, we include the same set of 26 parameters in all three calibration cases for consistency.

Table 2 Selected 26 parameters for calibration and their normalized sensitivity indices sorted based on SPAEF column. Zero values are highlighted with a grey shade.

Parameter	Description	Normalized Sensitivity		
		KGE	KGElow	SPAEF
ETref-af	Intercept – forest	0.022	0.117	**0.646**
ETref-c	Exponent coefficient	0.031	0.732	0.490
ETref-b	Base coefficient	0.439	3.013	0.317
rotfrcoffore	Root fraction coefficient for forest areas	0.011	0.013	0.162
ETref-a	Intercept –nonforest ~~(pervious)~~	0.308	3.235	0.157
ptfhigconst	Constant in Pedo-transfer function for soils with sand content higher than 66.5%	0.063	0.223	0.096
rotfrcofclay	Root fraction coefficient ~~Threshold for actual ET reduction~~ for clay in agricultural cropland	0.101	0.274	0.094
ptfhigdb	Coefficient for bulk density in Pedo-transfer function for soils with sand content higher than 66.5%	0.036	0.257	0.070
rotfrcofsand	Root fraction coefficient for sand in agricultural cropland~~Threshold for actual ET reduction for sand~~	0.120	0.439	0.061

canintfact	Canopy interception factor	0.004	0.029	0.018
orgmatforest	Organic matter content for forest	0.136	0.893	0.014
ptfhigclay	Coefficient for clay content in pedo-transfer function	0.008	0.033	0.011
infshapef	Infiltration shape factor	0.103	0.099	0.006
ptfkssand	Coefficient for sand content in pedo-transfer function for hydraulic conductivity	0.415	2.780	0.002
ptfksconst	Constant in pedo-transfer function for hydraulic conductivity of soils with sand content higher than 66.5%	0.236	0.842	0.001
snotrestemp	Snow temperature threshold for rain and snow separation	0.034	0.206	0.000
ptfksclay	Coefficient for clay content in pedo-transfer function for hydraulic conductivity	0.040	0.313	0.000
orgmatimper	Organic matter content for impervious zone	0.009	0.020	0.000
expslwintflw	Exponent slow interflow	0.412	**3.490**	0.000
slwintreceks	Slow interception	0.872	1.296	0.000
intrecesslp	Interflow recession slope	0.602	1.105	0.000
rechargcoef	Recharge coefficient	**0.935**	0.666	0.000
geoparam1	Parameter for geological formation 1	0.328	0.138	0.000
geoparam2	Parameter for geological formation 2	0.558	0.207	0.000
strcelerity	Streamflow celerity for routing	0.364	0.062	0.000
intstorcapf	Interflow storage capacity factor	0.198	0.010	0.000

5.2 Model calibration and validation

The mHM model is calibrated using streamflow records (gauges A and B in Figure 1~~Figure 1~~~~Figure 1~~) from eight years (2001-2008) and validated for a recent period (2009-2014).Preceding four-years ~~of these two periods~~(1997-2000~~) and 2005-2008)~~are used for model warm-up. We prepared remotely sensed monthly averaged AET pattern-maps calculated for these years considering only cloud-free days from summer months. AET patterns of winter months are not considered since it is mostly cloudy and ET is very low and uniform (energy limited) in winter.

The 26 selected parameters from SS are used in the following three calibration strategies: 1) only streamflow oriented (Q-only) calibration using equally-weighted KGE and KGE_{low}, 2) only spatial pattern oriented calibration using SPAEF and 3) streamflow and spatial patterns of AET together using all three objective functions with equal weights of 50% on spatial metric and 50% for the two streamflow metrics (25% each).Table 3~~Table 3~~~~Table 3~~ provides the overall picture of the three different calibration strategies where two of these strategies are based on an ensemble of nine calibrations. Therefore, the basic descriptive statistics are also given as robustness indicators. The results show that the combined calibration (Q and Spatial) produces similar results to both Q-only and Spatial-only calibrations focusing on streamflow and spatial patterns of AET respectively; whereas the single metric calibrations gave very different results for the opposite objective functions e.g. SPAEF vs streamflow metrics. It is interesting that when comparing the calibration ensemble with the median performance there is very limited trade-off between the Q-only and the combined Q and Spatial calibrations which have very similar average KGE values. When looking specifically at the best performing ensemble member with lowest total Φ, there is a more pronounced trade-off between the Q-only and Q and Spatial together calibrations, as the streamflow performance is poorer when SPAEF is included in the group of objective functions. The differences in the streamflow metrics indicate that each objective function carries relevant but slightly conflicting information. Moreover, the results show that the hydrologic model simulates best AET patterns in different months for different ensemble calibrations. In other words, while one ensemble member has the best performance for April, other calibrations may have the best performance for May and June. This is a secondary trade-off which

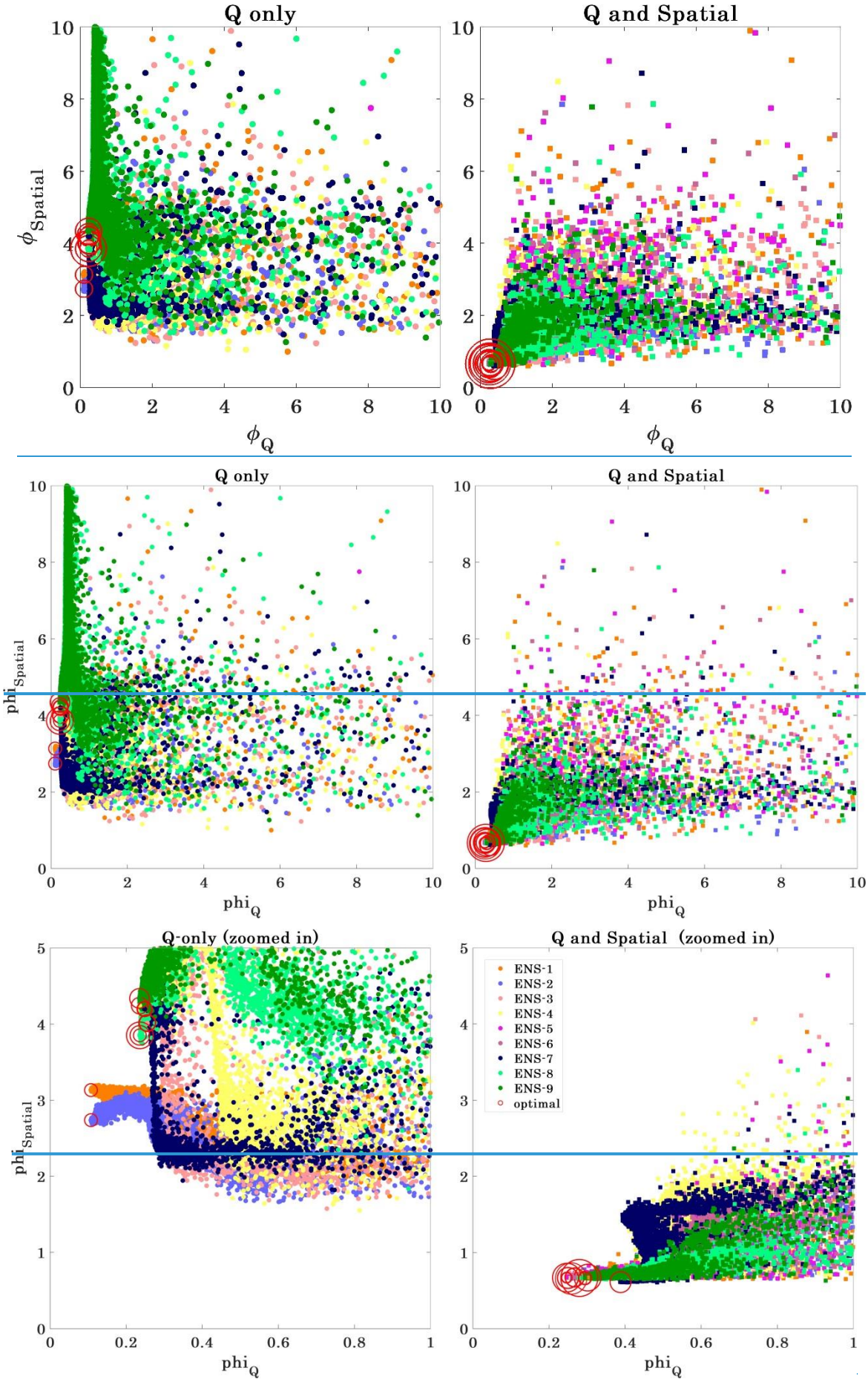
illustrates that the calibration might benefit from temporal variability in the parameters controlling the spatial parametrization scheme. It should be noted that ranking the calibrations within the two ensembles are based on the overall Φphi that is comprised of all objective functions for the corresponding calibration. For that reason, the best member of Q and Spatial calibration holds the lowest (i.e. best) Φphi of 0.91comprised of the highest possible KGE, KGE_{low} and SPAEF at the same time but not necessarily the highest SPAEF alone. This resulted in a slightly lower SPAEF mean of 0.395 for the best member compared to the median member with a SPAEF mean of 0.396 (Table 3Table 3Table 3).

Table 3 Summary of the calibration results for three cases. Median and standard deviation (Std.) refers to the calibration ensemble ranked based on their total Φphi. Lower Φphi indicates lower error.

Metrics / Gauge		Q-Only		Spatial-Only	Q and Spatial	
		Median (Std)	Best	Single Cal.	Median (Std)	Best
KGE [-]	(A)	0.83(0.03)	0.97	-1.47	0.88 (0.01)	0.89
KGE [-]	(B)	0.94(0.02)	0.97	-1.03	0.92 (0.01)	0.93
KGE_{low} [-]	(A)	0.81(0.02)	0.86	-3.12	0.81 (0.02)	0.82
KGE_{low} [-]	(B)	0.81(0.02)	0.85	-2.66	0.79 (0.02)	0.8
BIAS [%]	(A)	-6.24(2.23)	-0.98	38.56	-2.44 (0.62)	-1.25
BIAS [%]	(B)	1.51(0.83)	1.83	46.87	1.42 (0.93)	3.86
Apr - SPAEF		-0.88 (0.33)	-0.17	0.5	0.57 (0.02)	0.51
May - SPAEF		-0.62 (0.26)	-0.12	0.47	0.35 (0.07)	0.35
Jun - SPAEF		-0.59 (0.23)	-0.09	0.36	0.27 (0.11)	0.27
Jul - SPAEF		-0.39 (0.10)	-0.36	0.51	0.38 (0.04)	0.43
Aug - SPAEF		-0.29 (0.10)	-0.36	0.53	0.48 (0.05)	0.49
Sep - SPAEF		0.02(0.19)	0.30	0.40	0.33 (0.02)	0.32
ΦPhi_{total}		0.24 (0.06)	0.11	0.52	0.94 (0.03)	0.91

The results of the Q-only model calibration using only KGE and KGE_{low} reveal very poor simulated patterns of AET, with negative SPAEF for all months. This is not surprising since this atoptimization calibration is not constrained regarding the spatial patterns, but also illustrates that discharge observations alone contain no spatial pattern information of AET. In contrast, the spatial-only calibration using only SPAEF shows a very poor water balance, with negative KGE and large bias. We are aware that spatial-only calibration is not applicable and meaningful for hydrologic studies.

The model performance development through the calibrations (9+9) and optimum points are shown using scatter plots in Figure 2Figure 2Figure 2, which displays all model runs with Φphi values inside the specified plot ranges. The scatter plots illustrate trade-offs between objective functions and consistency among calibration ensemble members. The performance regarding spatial patterns ($\Phi phi_{Spatial}$) displays a high degree of trade-off with all combined calibrations achieving $\Phi phi_{Spatial}$ values around 0.8 whereas the Q-only calibrations achieve $\Phi phi_{Spatial}$ values ranging 2.8 and 4.4.There are two main clusters in the Q-only calibrations one around 0.11 Φphi_{Q} and the other around 0.25 Φphi_{Q} whereas all nine Q and spatial calibrations follow a similar level in y axis ($\Phi phi_{Spatial}$).It is surprising to see that SCE-UA did not always find the same optimum solution with varying seed number, which is the case mainly for the Q-only calibration. This raises the question whether SCE-UA algorithm may have shortcomings e.g. a local minimum sincewe applied the same stopping rulesto all calibrations but selected different seed numbers only.Perhaps more consistent optimum solutions for the Q-only calibrations could have been achieved with tighter stopping rules and the same initial parameter sets.


Q only
Q and Spatial
$\phi_{Spatial}$
ϕ_Q
phi$_{Spatial}$
phi$_Q$
Q-only (zoomed in)
Q and Spatial (zoomed in)
ENS-1
ENS-2
ENS-3
ENS-4
ENS-5
ENS-6
ENS-7
ENS-8
ENS-9
optimal

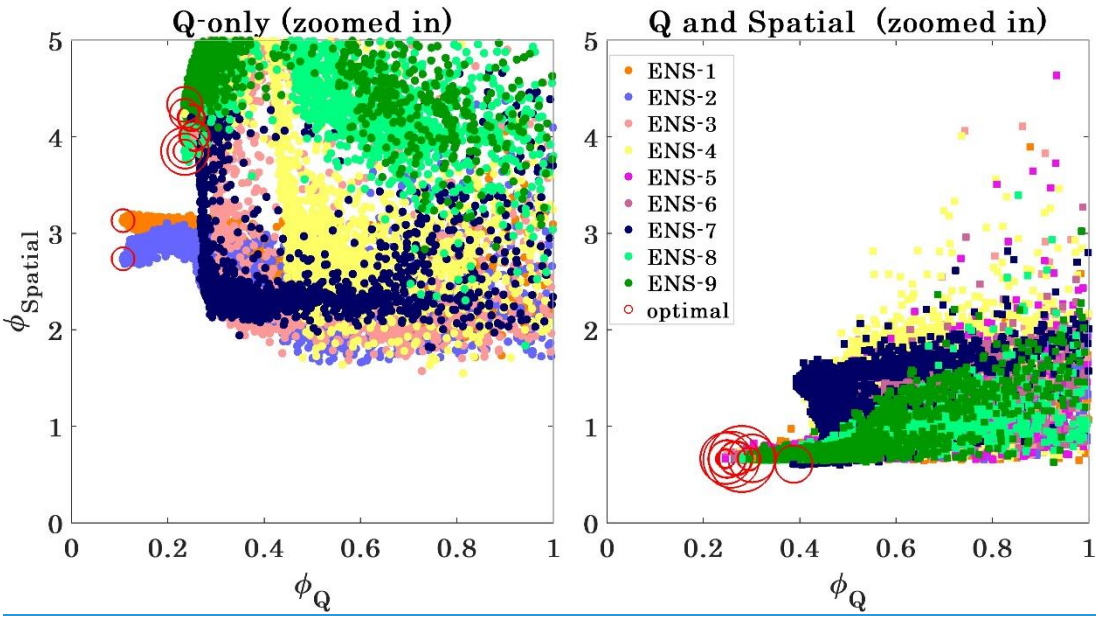


Figure 2. Scatter plots of total ~~spatial~~ $\Phi_{Spatial}$~~phi~~ versus total ~~Q~~ Φ_Q for all nine calibration ensemble members. First and second row subplots are the same figures except for different extent i.e. [10 10] and [1 5] to zoom into the edge of the search space. Different radius of red circles is used to show the optimum points for all nine ensemble members clearly.

Similarly, the grey shades in Figure 3~~Figure 3~~Figure 3 show the ensemble range of simulated hydrographs for the Q-only and Q and Spatial calibrations. From the hydrographs it is clear that the ensemble range for station A is generally larger than station B, indicating larger uncertainty for sub-basin A. Interestingly, Figure 3~~Figure 3~~Figure 3 also illustrates that the Q and Spatial calibration ~~reduces~~ constrains the solution better~~the uncertainties~~, not only in AET simulations, but also in streamflow simulations, as indicated by the slightly narrower range in simulated streamflow for the Q and Spatial calibrations. However, even though the range of hydrographs is slightly narrower the simulations are also further from the observed during summer months.

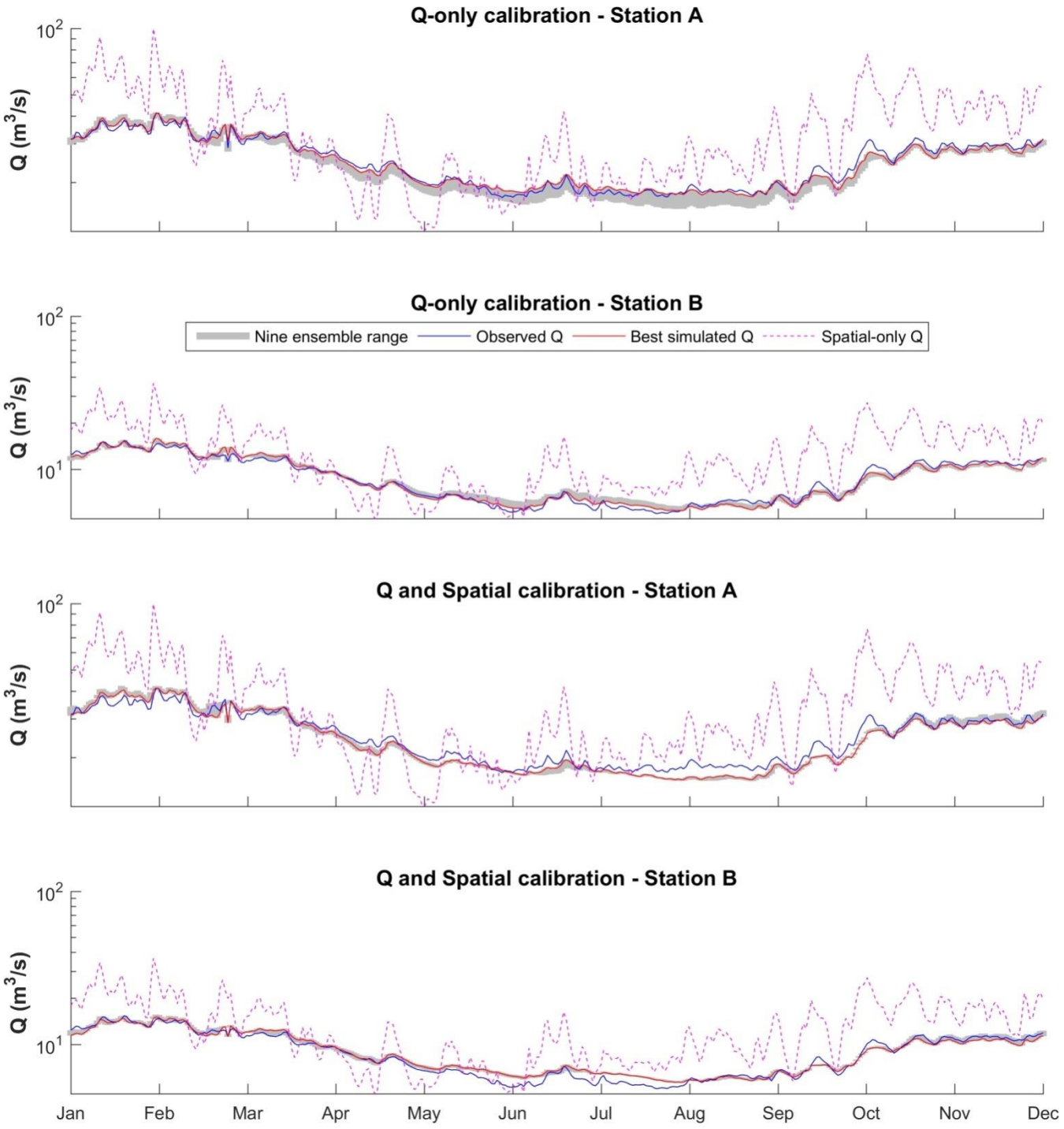


Figure 3. Average hydrograph of all years in the calibration period (2001-2008) to illustrate the ensemble of nine model calibrations with different seed numbers.

The corresponding simulated AET maps for the results presented in Table 3~~Table 3~~~~Table 3~~are shown in Figure 4~~Figure 4~~~~Figure 4~~.This figure illustrate the monthly mean maps across all years of actual evapotranspiration for the cloud-free days available for the remote sensing estimates. Only the best performing members from the two ensembles are presented in this figure. The maps are normalized with their mean value to use one representative colour bar in the legend. As indicated in Table 3~~Table 3~~~~Table 3~~, the resultant maps from Spatial-only (third row in Figure 4~~Figure 4~~~~Figure 4~~) and Q and Spatial calibrations (fourth row in Figure 4~~Figure 4~~~~Figure 4~~) are obviously more similar to the reference monthly maps (first row in Figure 4~~Figure 4~~~~Figure 4~~) than the maps of Q-only calibration (second row in Figure 4~~Figure 4~~~~Figure 4~~). The results clearly show that the model can simulate month-to-month variations in AET patterns reasonably well. The poor AET performance in the Q-only maps is obvious in the second row of Figure 4~~Figure 4~~~~Figure 4~~ where we see only a uniform simulated AET pattern except for the

forest areas revealing very little information about variability in AET and the influence of soil and vegetation. This is from the fact that the KGE and KGE_{low} objective functions contain no information on the patterns of AET resulting in an unconstrained optimization regarding spatial pattern and variability. ~~Therefore~~Therefore, the optimizer randomly moves in the SPAEF solution space and picks the best streamflow performance with no regard to AET patterns. Although not perfect (average SPAEF = 0.46and 0.40), the simulated pattern match in the last two rows of Figure 4~~Figure 4~~~~Figure 4~~ is quite good compared to the remote sensing based estimate since the simulation is able to represent the general pattern influenced by soil, vegetation and land cover while maintaining a similar variance and smoothness.

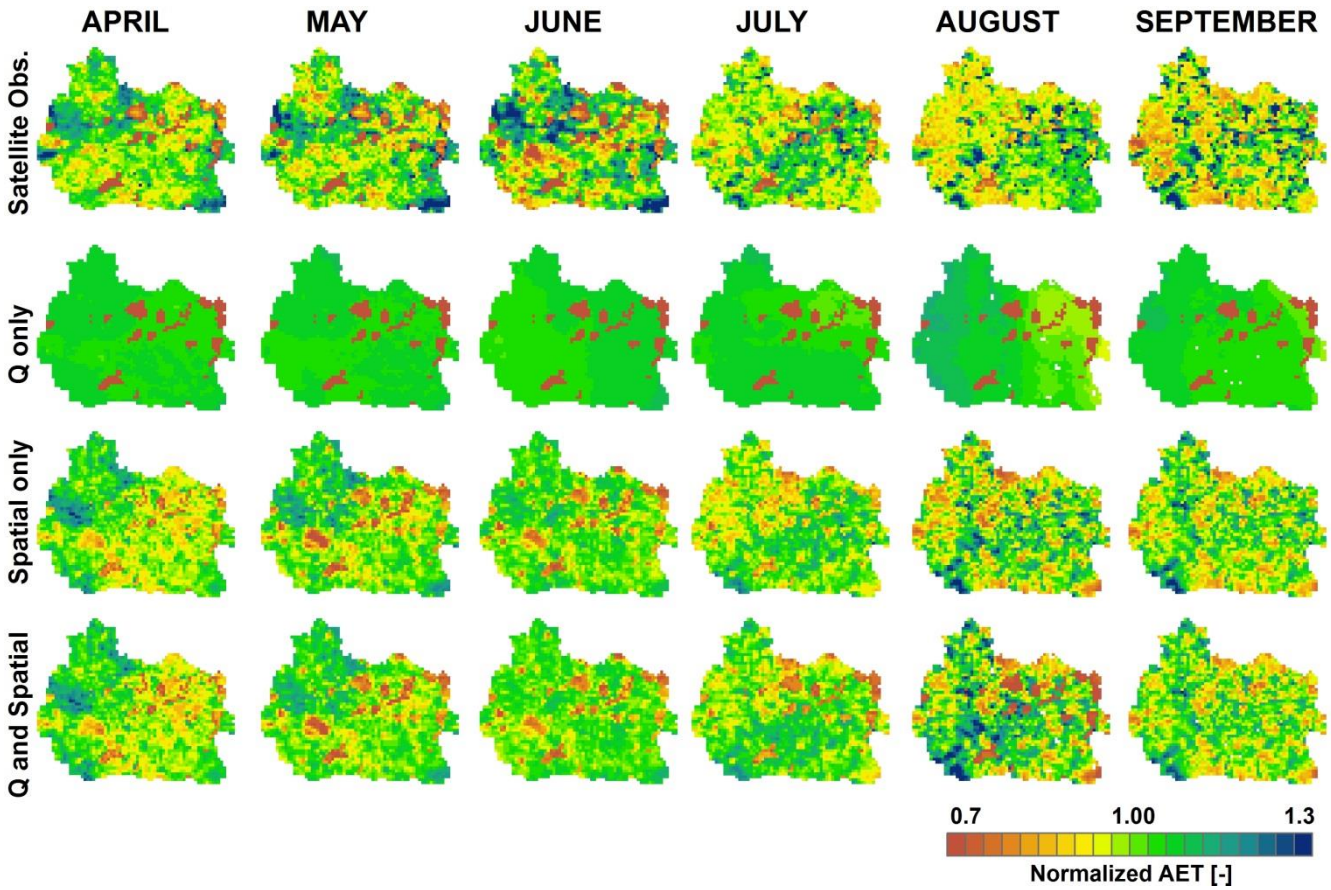


Figure 4. Three different calibration strategies: 1) streamflow-only (second row) 2) spatial-only (third row) and 3) streamflow and spatial together (fourth row) are compared with monthly TSEB estimates (first row). Calibrations are evaluated for monthly averages from April to September using cloud-free days. Note that these maps are normalized with their mean to use one representative colour bar and highlight the pattern information.

Table 4~~Table 4~~~~Table 4~~shows the same results as Table 3~~Table 3~~~~Table 3~~ but for the validation period spanning from 2009 until 2014. Obviously, the results are somewhat poorer than those for the calibration period. A prop in performance for spatial-only and combined metrics is mainly seen for KGE_{low} and the total bias, whereas the SPAEF for Spatial-only and Q and Spatial remain similar to the calibration periods with average SPAEF around 0.4. Interestingly, there is no real trade-off for streamflow metrics between Q-only and Q and Spatial calibrations for the validation period, even for the best performing ensemble member. Although a better streamflow performance could be achieved by Q-only calibration during calibration, this cannot be sustained during validation, indicating some overfitting when using streamflow metric only. In contrast, the SPAEF performance does not drop during validation for the combined Q and Spatial optimization, indicating less overfitting and a more robust model parametrization.

Table 4 Summary of the validation results for three cases. Median and standard deviation (Std.) refers to the validation ensemble ranked based on their total Φ~~phi~~. Lower Φ ~~phi~~ indicates lower error.

Metrics / Gauge		Q-Only Median (Std)	Q-Only Best	Spatial-Only Single Cal.	Q and Spatial Median (Std)	Q and Spatial Best
KGE [-]	**(A)**	0.83 (0.01)	0.86	-1.65	0.86 (0.01)	0.88
KGE [-]	**(B)**	0.89 (0.02)	0.93	-1.40	0.87 (0.01)	0.88
KGE_{low} [-]	**(A)**	0.70 (0.04)	0.79	-3.84	0.72 (0.02)	0.76
KGE_{low} [-]	**(B)**	0.65 (0.02)	0.66	-3.72	0.64 (0.03)	0.65
BIAS [%]	**(A)**	-15.76 (2.26)	-10.38	29.64	-11.89 (1.15)	-8.85
BIAS [%]	**(B)**	5.23 (1.04)	5.39	55.86	5.00 (1.45)	9.13
Apr - SPAEF		-0.76 (0.30)	-0.11	0.47	0.51 (0.02)	0.53
May - SPAEF		-0.65 (0.28)	-0.09	0.56	0.51 (0.03)	0.51
Jun - SPAEF		-0.50 (0.19)	-0.13	0.38	0.27 (0.04)	0.27
Jul - SPAEF		-0.56 (0.17)	-0.30	0.59	0.48 (0.09)	0.50
Aug - SPAEF		-0.15 (0.10)	-0.33	0.18	0.19 (0.07)	0.21
Sep - SPAEF		-0.12 (0.13)	-0.31	0.44	0.35 (0.02)	0.37
Φ ~~Phi~~$_{total}$		0.73 (0.11)	0.50	0.60	1.38 (0.10)	1.34

6 Discussion

In the initial phase of the study numerous flawed calibrations were carried out in an attempt to produce simulated spatial patterns of AET similar to the satellite based reference patterns. However, the inability to produce similar patterns was found to be caused by limitations in spatial model parameterisation and spatial performance metric choice. Regarding the spatial parameterisation, the initial model was based on a spatially uniform parameterisation of root fraction coefficient and PET correction factor, two parameters with major control on the simulated AET. Therefore, more flexible yet physically meaningful parameterisations were implemented where full spatial variability was enabled by combining 2-3 calibration parameters to initial spatial distributions of soil type and LAI. Regarding the use of appropriate spatial performance metrics, the initial attempts using standard metrics of correlation coefficient, Mapcurves (Hargrove et al., 2006), coefficient of variation, Goodman and Kruskal's lambda (Goodman and Kruskal, 1954), agreement coefficient (Ji and Gallo, 2006), Theil's Uncertainty, EOF and Cramér's V (Cramér, 1946; Koch et al., 2015; Rees, 2008)proved to be inadequate in a calibration framework, since undesired visual patterns were achieved e.g. with high correlation, but too low standard deviation or highly separate clusters. Therefore, we developed the SPAEF metric which proved to be very efficient for calibrating the model to a satisfying spatial pattern by combining correlation coefficient, coefficient of variation ratio and histogram overlap in a robust metric that guides the model calibration well. ~~For more details of the testing of the SPAEF in a calibration framework please see the study by Koch et al. (2017a)~~. It is our experience and recommendation that incorporating the spatial dimension in all aspects of the distributed hydrological model development from model structure, parametrization, metric selection, sensitivity analysis and calibration is essential in order to achieve significant improvement in the spatial pattern performance of a model. ~~It is recognized~~ We believe that traditional downstream discharge measurements contain much more accurate and robust information on the overall water balance compared to the non-continuous remotely sensed estimates, and therefore, the model constraint on biases should only originate from these streamflow observations. Conversely, it is well-known that aggregated streamflow measurements contain no information on spatial patterns upstream of the measurement (Stisen et al., 2011b). Therefore, the combination of satellite derived patterns and aggregated streamflow measurements are an ideal way of constraining distributed hydrological models. In ~~fact~~fact, spatial patterns should always be considered when evaluating distributed models. Even if detailed satellite estimates are not available, expert judgments and land cover information should be used to select the most appropriate parameter set (producing the most likely spatial patterns) among equally likely solutions obtained through discharge-only calibration. When a distributed model is applied, it should be desired, that it produces not only satisfying discharge simulations, but at the same time produces realistic spatial patterns of states and fluxes such as AET

and soil moisture. White et al. (2017) also highlighted the importance of getting the spatial patterns right in their study since constraining the model against streamflow alone did not secure a robust land cover change scenario modelling.

The monthly spatial maps are built based on the AET patterns from cloud-free days. Here, we ignore the temporal aspect and focus only on the consistent spatial patterns for each month of the growing season. The advantage of this approach is that only the main information content of the satellite data, their spatial patterns, are utilized while the uncertainty associated with~~to~~ the absolute values of the AET estimates are not influencing the calibrations. In addition, the simulated monthly mean AET maps reflect mainly the model parameterisation and to a lesser degree the day-to-day variation in climate forcing. This is desirable since the aim of the model calibration is to optimize the model parameterisation with a given climate forcing dataset. The current calibration framework builds on the assumption that the satellite based estimate of AET patterns approximate an observed ~~patterns~~pattern that is suitable for model optimization. In general, the calibration approach is deterministic by nature and does not consider error or uncertainties in neither observed discharge nor AET patterns. Future work could add this component to the approach. However, assessment of the uncertainties in the observed spatial patterns are far from straightforward, since the uncertainties of interest with the proposed approach are solely related to the uncertainties related to the spatial patterns and not to biases. Therefore, quantification of pattern uncertainties would require a very dense network of actual evapotranspiration measurements.

The calibration results obtained in the current study where three strategies were tested with varying combinations of objective functions showed that with an appropriate metric design, limited trade-offs can be achieved when combining streamflow and spatial pattern metrics in a joint calibration framework. This is largely attributed to the nature of the metric as the spatial performance metric is bias-insensitive whereas the streamflow metrics have very little sensitivity to spatial redistribution of AET patterns as long as the spatial averages remain unchanged.~~This is largely attributed to the fact that by design, the spatial performance metric is bias-insensitive whereas the streamflow metrics have very little sensitivity to spatial redistribution of AET patterns as long as the spatial averages remain unchanged~~. Bias and temporal variability of satellite derived AET estimates could also be useful for model optimization, however, in this study, we deliberately limited the information content of the satellite data to address the spatial patterns. This was~~is~~ done because even though the satellite based AET estimate is validated against eddy covariance stations (Mendiguren et al., 2017) they only represent specific cloud-free days limiting their value to assess the long term water balance of the catchment. The calibration results using only streamflow metrics revealed that this traditional calibration target cannot guarantee satisfying spatial pattern performance even though the model structure and parametrization framework enables this without much compromise as illustrated by the performance of the combined Q and Spatial calibrations which resulted in very similar performance of both streamflow and spatial patterns as the single objective calibrations individually.

The spatial model parameterisation applied in Skjern catchment can be site~~domain~~- specific due to the uniform land use (agricultural cropland) across soils ranging from very coarse sandy soil to more loamy soils ~~due to itsvery sandy soils~~whereas the calibration framework and SPAEF metric can be applied to any other river basin in the world. ~~Although we found spatial parametrization examples from different counties such as Sweden (Wiklert, 1961) and Germany (Vetter and Scharafat, 1964), we have not tested our field capacity dependent root fraction approach under other conditions yet. This will be a topic of our subsequent study on multiple European catchments.~~Regarding the dynamic scaling function, developed for incorporating remotely sensed LAI in ET_{ref} scaling, it should be noted that the use of LAI to describe the deviation of each grid cell from the assumed reference grass is a simplification. Albedo could also have been included in the dynamic scaling function; however, one could argue that albedo and LAI are somewhat correlated and including one of them is already contributing the information about the other (Chen et al., 2005; Liu et al., 2017; Stisen et al., 2008). Moreover, we limit this study to temporally averaged spatial patterns of AET and deliberately choose to ignore the day-to-day dynamics of AET. In this study, spatially varying but temporally constant field capacity dependent root fraction is utilized, however; it would be more elegant and physically more

sound to represent the seasonality in root-growth dynamics more realistically by implementing a seasonally varying root fraction coefficient (beta) that is similar to the concept of LAI based PET correction using the DSF module.

7 Conclusions

Our study aimed at parameterising a distributed hydrologic model for simulating distributed actual evapotranspiration patterns before an ensemble calibration using satellite based data. This order is crucial for a progressive hydrologic modelling with flexible model structure based on open-source philosophy. All these steps should be suitable for the catchment to give model enough flexibility to adjust to pattern observations. The calibration efforts will have limited effect on spatial patterns if the model parameterisation has not been investigated with pattern performance in mind. Ideally, the models should offer different parametrization schemes or at least have room for development based on open-source philosophy so that we can test different spatial parameterisations for a particular calibration goal. Here, we implemented a field capacity dependent root fraction coefficient determining the root profile over depth for different soil and vegetation types. We introduced a dynamic scaling function which imprints the leaf area index in the potential evapotranspiration. After organizing the spatial parameterisation of the model in a parsimonious manner, we also reduced the number of parameters using sequential screening. Only the informative parameters from the sequential screening are used in the subsequent ensemble calibration exercise. We then assessed the effect of different calibration strategies including monthly spatial patterns of actual evapotranspiration in combination with traditional streamflow observations. In the spatial calibration, the agreement between observed and simulated spatial patterns is added as a part of the objective function used for model optimization. For that a multi-component bias insensitive spatial efficiency metric (SPAEF) is used to evaluate the simulated AET maps. The following conclusions can be drawn from our results:

- Preparing the model parameterisation for spatial calibration is a key element for achieving the calibration objectives. More specifically, the model parameterisation needs to be designed to allow the spatial parameter distribution to be optimized through calibration.
- The newly proposed spatial efficiency metric (SPAEF) ~~is~~ has proven to be robust and easy to interpret due to its three distinct and complementary components of correlation, variance and histogram matching.
- Based on the multi-component calibration results, including spatial pattern information in calibration, significantly improves the spatial model simulations while maintaining similar streamflow performance. For the combined calibration, there is a limited trade-off between streamflow and spatial patterns for the best performing calibration ensemble compared to the Q-only calibration. However, this trade-off disappears in the validation test, indicating that a more robust parameter set is achieved during the combined Q and Spatial calibration.

Overall, the hydrological modelling community can benefit from building familiarity with several aspects of spatial model evaluation, including spatial parameterisation and multi-component spatial performance metrics.

Acknowledgements

We acknowledge the financial support for the SPACE project by the Villum Foundation (http://villumfonden.dk/) through their Young Investigator Programme (grant VKR023443). The TSEB code is retrieved from https://github.com/hectornieto/pyTSEB. All MODIS data was retrieved from the online Data Pool, courtesy of the NASA Land Processes Distributed Active Archive Center (LP DAAC), USGS/Earth Resources Observation and Science (EROS) Center, Sioux Falls, South Dakota, https://lpdaac.usgs.gov/data_access/data_pool.Pre-processing ET with crop coefficient type dynamic scaling function is available in the mHM version 5.7 (www.ufz.de/mhm/).The Python and Matlab scripts for Spatial

Efficiency (SPAEF) and a tutorial are available in the SPACE project website (http://www.space.geus.dk/) and via Researchgate repository (Demirel et al., 2017).

References

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. [online] Available from: http://www.fao.org/docrep/x0490e/x0490e00.htm, 1998.

Berezowski, T., Nossent, J., Chormański, J. and Batelaan, O.: Spatial sensitivity analysis of snow cover data in a distributed rainfall-runoff model, Hydrol. Earth Syst. Sci., 19(4), 1887–1904, doi:10.5194/hess-19-1887-2015, 2015.

Beven, K. and Freer, J.: Equifinality, data assimilation, and uncertainty estimation in mechanistic modelling of complex environmental systems using the GLUE methodology, J. Hydrol., 249(1–4), 11–29, doi:Doi 10.1016/S0022-1694(01)00421-8, 2001.

Campolongo, F., Cariboni, J. and Saltelli, A.: An effective screening design for sensitivity analysis of large models, Environ. Model. Softw., 22(10), 1509–1518, doi:10.1016/j.envsoft.2006.10.004, 2007.

Chen, J. M., Chen, X., Ju, W. and Geng, X.: Distributed hydrological model for mapping evapotranspiration using remote sensing inputs, J. Hydrol., 305(1–4), 15–39, doi:10.1016/j.jhydrol.2004.08.029, 2005.

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, Hydrol. Earth Syst. Sci., 17(7), 2947–2966, doi:10.5194/hess-17-2947-2013, 2013.

Corbari, C., Ravazzani, G., Ceppi, A. and Mancini, M.: Multi-pixel Calibration of a Distributed Energy Water Balance Model Using Satellite Data of Land Surface Temperature and Eddy Covariance Data, Procedia Environ. Sci., 19(November 2015), 285–292, doi:10.1016/j.proenv.2013.06.033, 2013.

Cornelissen, T., Diekkrüger, B. and Bogena, H.: Using High-Resolution Data to Test Parameter Sensitivity of the Distributed Hydrological Model HydroGeoSphere, Water, 8(5), 202, doi:10.3390/w8050202, 2016.

Cramér, H.: Mathematical Methods of Statistics, Princeton University Press., 1946.

Crow, W. T., Wood, E. F., Pan, M., de Wit, M., Stankiewicz, J., Crow, W. T., Coe, M. T. and Birkett, C. M.: Multiobjective calibration of land surface model evapotranspiration predictions using streamflow observations and spaceborne surface radiometric temperature retrievals, J. Geophys. Res., 311(D23), 1917–1921, doi:10.1029/2003WR002543, 2003.

Cuntz, M., Mai, J., Zink, M., Thober, S., Kumar, R., Schäfer, D., Schrön, M., Craven, J., Rakovec, O., Spieler, D., Prykhodko, V., Dalmasso, G., Musuuza, J., Langenberg, B., Attinger, S. and Samaniego, L.: Computationally inexpensive identification of noninformative model parameters by sequential screening, Water Resour. Res., 51(8), 6417–6441, doi:10.1002/2015WR016907, 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, Q. J. R. Meteorol. Soc., 137(656), 553–597, doi:10.1002/qj.828, 2011.

Demirel, M. C., Booij, M. J. and Hoekstra, A. Y.: Effect of different uncertainty sources on the skill of 10 day ensemble low flow forecasts for two hydrological models, Water Resour. Res., 49(7), 4035–4053, doi:10.1002/wrcr.20294, 2013.

Demirel, M. C., Koch, J. and Stisen, S.: SPAEF: SPAtial EFficiency, Researchgate, doi:10.13140/RG.2.2.18400.58884, 2017.

Duan, Q.-Y. Y., Sorooshian, S. and Gupta, V.: Effective and efficient global optimization for conceptual rainfall-runoff models, Water Resour. Res., 28(4), 1015–1031, doi:10.1029/91WR02985, 1992.

Githui, F., Thayalakumaran, T. and Selle, B.: Estimating irrigation inputs for distributed hydrological modelling: a case study from an irrigated catchment in southeast Australia, Hydrol. Process., 30(12), 1824–1835, doi:10.1002/hyp.10757, 2016.

Goodman, L. A. and Kruskal, W. H.: Measures of Association for Cross Classifications*, J. Am. Stat. Assoc., 49(268), 732–

764, doi:10.1080/01621459.1954.10501231, 1954.

Greve, M. H., Greve, M. B., Bøcher, P. K., Balstrøm, T., Breuning-Madsen, H. and Krogh, L.: Generating a Danish raster-based topsoil property map combining choropleth maps and point information, Geogr. Tidsskr. J. Geogr., 107(2), 1–12, doi:10.1080/00167223.2007.10649565, 2007.

Gupta, H. V, Kling, H., Yilmaz, K. K. and Martinez, G. F.: Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling, J. Hydrol., 377(1–2), 80–91, doi:10.1016/j.jhydrol.2009.08.003, 2009.

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, edited by V. i Danmark, Hydrol. Earth Syst. Sci., 17(7), 2809–2825, doi:10.5194/hess-17-2809-2013, 2013.

Hargrove, W. W., Hoffman, F. M. and Hessburg, P. F.: Mapcurves: a quantitative method for comparing categorical maps, J. Geogr. Syst., 8(2), 187–208, doi:10.1007/s10109-006-0025-x, 2006.

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 Resour. Res., 44(1), 224–240, doi:10.1029/2007WR006097, 2008.

Hunink, J. E., Eekhout, J. P. C., de Vente, J., Contreras, S., Droogers, P. and Baille, A.: Hydrological Modelling using Satellite-based Crop Coefficients: a Comparison of Methods at the Basin Scale, Remote Sens., 1–16, doi:10.3390/rs9020174, 2017.

Immerzeel, W. W. and Droogers, P.: Calibration of a distributed hydrological model based on satellite evapotranspiration, J. Hydrol., 349(3–4), 411–424, doi:10.1016/j.jhydrol.2007.11.017, 2008.

Jackson, R. B., Canadell, J., Ehleringer, J. R., Mooney, H. A., Sala, O. E. and Schulze, E. D.: A global analysis of root distributions for terrestrial biomes, Oecologia, 108(3), 389–411, doi:10.1007/BF00333714, 1996.

Jensen, H. E., Jensen, S. E., Jensen, C. R., Mogensen, V. O. and Hansen, S.: Jordfysik og jordbrugsmeteorologi, Jordbrugsforlaget., 2001.

Jensen, K. H. and Illangasekare, T. H.: HOBE: A Hydrological Observatory, Vadose Zo. J., 10(1), 1, doi:10.2136/vzj2011.0006, 2011.

Ji, L. and Gallo, K.: An agreement coefficient for image comparison, Photogramm. Eng. Remote Sensing, 72(7), 823–833 [online] Available from: http://pubs.er.usgs.gov/publication/70028506, 2006.

Kalma, J. D., McVicar, T. R. and McCabe, M. F.: Estimating Land Surface Evaporation: A Review of Methods Using Remotely Sensed Surface Temperature Data, Surv. Geophys., 29(4–5), 421–469, doi:10.1007/s10712-008-9037-z, 2008.

Kling, H. and Gupta, H.: On the development of regionalization relationships for lumped watershed models: The impact of ignoring sub-basin scale variability, J. Hydrol., 373(3–4), 337–351, doi:10.1016/j.jhydrol.2009.04.031, 2009.

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 Resour. Res., 51(2), 1225–1246, doi:10.1002/2014WR016607, 2015.

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, J. Hydrometeorol., 18(4), 1121–1142, doi:10.1175/JHM-D-16-0148.1, 2017.

Kumar, R., Samaniego, L. and Attinger, S.: Implications of distributed hydrologic model parameterization on water fluxes at multiple scales and locations, Water Resour. Res., 49(1), 360–379, doi:10.1029/2012WR012195, 2013.

Larsen, M. A. D., Refsgaard, J. C., Jensen, K. H., Butts, M. B., Stisen, S. and Mollerup, M.: Calibration of a distributed hydrology and land surface model using energy flux measurements, Agric. For. Meteorol., 217(MARCH), 74–88, doi:10.1016/j.agrformet.2015.11.012, 2016.

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, J. Hydrol., 377(1–2), 120–130, doi:10.1016/j.jhydrol.2009.08.012, 2009.

Liu, C., Sun, G., McNulty, S. G., Noormets, A. and Fang, Y.: Environmental controls on seasonal ecosystem evapotranspiration/potential evapotranspiration ratio as determined by the global eddy flux measurements, Hydrol. Earth Syst. Sci., 21(1), 311–322, doi:10.5194/hess-21-311-2017, 2017.

Loosvelt, L., Vernieuwe, H., Pauwels, V. R. N., De Baets, B. and Verhoest, N. E. C.: Local sensitivity analysis for compositional data with application to soil texture in hydrologic modelling, Hydrol. Earth Syst. Sci., 17(2), 461–478, doi:10.5194/hess-17-461-2013, 2013.

Madsen, H. B.: Distribution of spring barley roots in Danish soils, of different texture and under different climatic conditions, Plant Soil, 88(1), 31–43, doi:10.1007/BF02140664, 1985.

Madsen, H. B.: Computerized soil data used in agricultural water planning, Denmark, Soil Use Manag., 2(4), 134–139, doi:10.1111/j.1475-2743.1986.tb00697.x, 1986.

Madsen, H. B. and Platou, S. W.: Land use planning in Denmark: the use of soil physical data in irrigation planning, Hydrol. Res., 14(5), 267 LP-276 [online] Available from: http://hr.iwaponline.com/content/14/5/267.abstract, 1983.

Melsen, L., Teuling, A., Torfs, P., Zappa, M., Mizukami, N., Clark, M. and Uijlenhoet, R.: Representation of spatial and temporal variability in large-domain hydrological models: case study for a mesoscale pre-Alpine basin, Hydrol. Earth Syst. Sci., 20(6), 2207–2226, doi:10.5194/hess-20-2207-2016, 2016.

Mendiguren, G., Koch, J. and Stisen, S.: Spatial pattern evaluation of a calibrated national hydrological model – a remote-sensing-based diagnostic approach, Hydrol. Earth Syst. Sci., 21(12), 5987–6005, doi:10.5194/hess-21-5987-2017, 2017.

Mizukami, N., Clark, M., Newman, A. J., Wood, A. W., Gutmann, E., Nijssen, B., Rakovec, O. and Samaniego, L.: Toward seamless large domain parameter estimation for hydrologic models, Water Resour. Res., doi:10.1002/2017WR020401, 2017.

Morris, M. D.: Factorial Sampling Plans for Preliminary Computational Experiments, Technometrics, 33(2), 161, doi:10.2307/1269043, 1991.

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, Agric. For. Meteorol., 77(3–4), 263–293, doi: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, Mon. Weather Rev., 100(2), 81–92, doi:10.1175/1520-0493(1972)100<0081:OTAOSH>2.3.CO;2, 1972.

Rakovec, O., Kumar, R., Attinger, S. and Samaniego, L.: Improving the realism of hydrologic model functioning through multivariate parameter estimation, Water Resour. Res., 613–615, doi:10.1002/2016WR019430, 2016.

Rees, W. G.: Comparing the spatial content of thematic maps, Int. J. Remote Sens., 29(13), 3833–3844, doi:10.1080/01431160701852088, 2008.

Samaniego, L., Kumar, R. and Attinger, S.: Multiscale parameter regionalization of a grid-based hydrologic model at the mesoscale, Water Resour. Res., 46(5), W05523, doi:10.1029/2008WR007327, 2010.

Schumann, G. J. P., Neal, J. C., Voisin, N., Andreadis, K. M., Pappenberger, F., Phanthuwongpakdee, N., Hall, A. C. and Bates, P. D.: A first large scale flood inundation forecasting model, Water Resour. Res., doi:10.1002/wrcr.20521, 2013.

Schuurmans, J. M., van Geer, F. C. and Bierkens, M. F. P.: Remotely sensed latent heat fluxes for model error diagnosis: a case study, Hydrol. Earth Syst. Sci., 15(3), 759–769, doi:10.5194/hess-15-759-2011, 2011.

Shin, M.-J., Guillaume, J. H. A., Croke, B. F. W. and Jakeman, A. J.: Addressing ten questions about conceptual rainfall–runoff models with global sensitivity analyses in R, J. Hydrol., 503, 135–152, doi:10.1016/j.jhydrol.2013.08.047, 2013.

Stisen, S., Jensen, K. H., Sandholt, I. and Grimes, D. I. F. F.: A remote sensing driven distributed hydrological model of the Senegal River basin, J. Hydrol., 354(1–4), 131–148, doi:10.1016/j.jhydrol.2008.03.006, 2008.

Stisen, S., Sonnenborg, T. O., Højberg, A. L., Troldborg, L. and Refsgaard, J. C.: Evaluation of Climate Input Biases and Water Balance Issues Using a Coupled Surface–Subsurface Model, Vadose Zo. J., 10(1), 37, doi:10.2136/vzj2010.0001,

2011a.

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, J. Hydrol., 409(1–2), 337–349, doi:10.1016/j.jhydrol.2011.08.030, 2011b.

Stisen, S., Koch, J., Sonnenborg, T. O., Refsgaard, J. C., Bircher, S., Ringgaard, R. and Jensen, K. H.: Moving beyond runoff calibration - Multi-constraint optimization of a surface-subsurface-atmosphere model, Hydrol. Process., submitted, 2017.

Su, Z.: The Surface Energy Balance System (SEBS) for estimation of turbulent heat fluxes, Hydrol. Earth Syst. Sci., 6(1), 85–100, doi:10.5194/hess-6-85-2002, 2002.

Swain, M. J. and Ballard, D. H.: Color indexing, Int. J. Comput. Vis., 7(1), 11–32, doi:10.1007/BF00130487, 1991.

Vazquez, J. A., Anctil, F., Ramos, M. H., Perrin, C. and Velázquez, J. A.: Can a multi-model approach improve hydrological ensemble forecasting? A study on 29 French catchments using 16 hydrological model structures, Adv. Geosci., 29, 33–42, doi:10.5194/adgeo-29-33-2011, 2011.

Velázquez, J. A., Anctil, F., Perrin, C. and Vazquez, J. A.: Performance and reliability of multimodel hydrological ensemble simulations based on seventeen lumped models and a thousand catchments, Hydrol. Earth Syst. Sci., 14(11), 2303–2317, doi:10.5194/hess-14-2303-2010, 2010.

Wei, Z., Yoshimura, K., Wang, L., Miralles, D. G., Jasechko, S. and Lee, X.: Revisiting the contribution of transpiration to global terrestrial evapotranspiration, Geophys. Res. Lett., 44(6), 2792–2801, doi:10.1002/2016GL072235, 2017.

White, J., Stengel, V., Rendon, S. and Banta, J.: The importance of parameterization when simulating the hydrologic response of vegetative land-cover change, Hydrol. Earth Syst. Sci., 21(8), 3975–3989, doi:10.5194/hess-21-3975-2017, 2017.

Zhang, Y., Chiew, F. H. S., Zhang, L. and Li, H.: Use of Remotely Sensed Actual Evapotranspiration to Improve Rainfall–Runoff Modeling in Southeast Australia, J. Hydrometeorol., 10(4), 969–980, doi:10.1175/2009JHM1061.1, 2009.