Response to Referee #1

1 *Samaniego et al. propose MPR to be a practical and robust method that provides consistent (seamless) parameter and flux fields across scales owing to the inconsistent and unrealistic parameter fields for land surface geophysical properties in many existing land surface and large-scale hydrological models. Although this study is properly motivated, I am having a hard time to understand what are the new advances from this manuscript comparing to Samaniego et al., WRR 2011 and Mizukami et al., 2017, particularly given that Mizukami et al. is submitted to WRR and perhaps under review.*
Mizukami, N., Clark, M., Newman, A., Wood, A., Gutmann, E., Nijssen, B., Samaniego, L., and Rakovec, O.: Towards seamless large domain parameter estimation for hydro- logic models, Water Resources Research, submitted., 2017

Thank you for the comment. We are sorry for not making clear enough the differences between Mizukami et al. (under review) (hereafter [MCN+2017]) and this manuscript. In the revised manuscript we show clear differentiations between [MCN+2017] and our study, see P8 L7ff, P12 L16ff.
[MCN+2017] is aiming at the development of "a model agnostic MPR system called MPR-flex which is applied to the Variable Infiltration Capacity (VIC) model to produce hydrologic simulations over the contiguous USA (CONUS)". In [MCN+2017] no attempt has been made to verify the flux-matching condition of ET obtained with VIC using the MPR-flex parameterization across scales.

In this manuscript (hereafter [SKT+2017]) we attempt to describe the progress towards seamless parameterizations in land surface or hydrological models. We present a short description of what has been made (the literature on the topic is extensive) and provide a simple example to visualize how many of the existing models are estimating a fundamental parameter such as soil porosity differently. We postulate, based on our own experience, a way forward that uses MPR, provide a "Protocol for evaluation of model parameterization" (which is not publish before), implement it to PCR-GLOBWB (also new and unpublished) and carry out a series of experiments (based on the spirit of the E. Wood's recommendation) to demonstrate how to spot faulty parameterizations (also not publish before). We also compare the effects of the parameterization on three models (mHM, WaterGAP, and PCR-GLOBWB) as part of these experiments (all using the same forcings and underlaying data). It should be clearly noted that none of these basic components are part of [MCN+2017].

2 *Another reason for my trouble of identifying new advances may be that lots of previous concepts and methods (REA, REW, HRU etc.) are touched but in a rather scattered manner, i.e., without a coherent synthesis, thus making it difficult to follow the authors' logic chain to lead to the new contributions from this study.*

These topics were excluded of the manuscript due to space restrictions and to improve the flow of the manuscript. We only refer to the HRU concept because it is commonly used for parameterization of HMs.

3 *... By briefly glancing through Samaniego et al., WRR 2011 I was guessing that perhaps in this study the major contribution is to introduce MPR as a robust parameter estimation approach for land surface and/or large-scale hydrological models, which in my mind are not really the same as those watershed-scale or highly-distributed hydrological models. For example, the application of MPR to PCR-GLOBWB has been largely illustrated in this manuscript. However, I am then confused again realizing there is another manuscript (Mizukami et al.) where MPR has also been applied to PCR-GLOBWB.*

The hydrological process implemented in a land surface model (LSM) can be similar to those of a hydrological model (HM). We agree that LSMs and HMs are not the same because they aim at different purposes, and that the former ones tend to be much more complex than the latter ones. The parameterization of soil parameters (e.g., soil porosity), however, can be based on the same principles of soil physics, and is often found in a large number of LSM/HM as shown in this manuscript.

The reviewer's confusion may have been originated by weak formulations in P9 L2 or in P18 L16 (old). We clarified these sentences in the revised manuscript. MPR has been applied to PCR-GLOBWB **only** in this manuscript up to now. The MPR-flex development presented in [MCN+2017] is applied **only** to VIC up to now. See P8 L7ff, P12 L16ff.

4 *I therefore strongly encourage the authors clearly articulate the major advancements in this study. That said, I have a few specific comments as below.*

Thank you for the recommendations. We explicitly point out the innovations of this study in the revised manuscript. The text of the introduction was drastically reduced to focus only on the state of the art that may lead to seamless parameterizations.

5 L2, Page 2. "must made" - "must be made"
Done

6 L6, P10. It is not a good practice to jump from Fig. 2 to Fig. 7 (whilst Fig. 3-6 not introduced yet)
Thank you for spotting this error. It was amended in the revised manuscript.

7 L6-8, P13. I don't think the argument so far can support this conclusion. Given the numerous processes controling the propagation from soil porosity to evapotranspiration and the fact these processes are very often presented & parameterized in different models with varying levels of complexity (i.e., model structure uncertainty), I could not really make sense out of this conclusion from my own experience (in both watershed modeling and land surface modeling) either.

We did not intent to claim that MMS is better or worse than MPR. We were only comparing the values obtained by MMS w.r.t. those estimated by MPR and estimate the differences. For sure we do not now at this scale which values are more close to reality, the only fact we know is that the MPR estimates used in two HMs are good enough to close the water balance in relatively well in over 300 basins over Pan-EU as shown in Rakovec et al. 2016. `http://doi.org/10.1175/jhm-d-15-0054.1`.
Since we do not really need the MMS data set in this study, and the comparison may lead to controversies, we decided to exclude these paragraphs from the revised manuscript.

8 L9-11, P13. As a modeler I could not agree with this conclusion either. A good parameter estimation method should never alter the true value of a parameter with very clear physical meaning, such as soil porosity. A parameter, no matter at what resolution(s). Rather, the so-called predictive uncertainties mentioned here should be used a signature to diagnose whether the model itself is sufficiently robust, not the other way around. Otherwise, we are playing with the parameters to get the right anwer for the wrong reasons.

Depends at the scale at which the PTF is applied. An effective paramater at 5 km resolution or coarser is an "effective" parameter representing the heterogeneity of the underlying land surface and hence cannot be observed or measured directly but can only be estimated. Because of this fact, the effective values of porosity cannot compared directly with field samples. Binley, A., Elgy, J., & Beven, K. (1989). doi:10.1029/2008WR007695 and many other publications from Beven and Blöschl, Wood etc. make this fact very clear. If a model is applied at point scale (at most meters) then a parameter estimation method would lead to parameter vales that can be obtained in laboratory. Since the MMS data set comparison was removed, this sentence does not appear in the revised manuscript.

9 L26-27, P15. Why is this well-accepted fact (among modelers at least) being used as a hypothesis?

Thank for mentioning it. Our intention in this experiment is a sensitivity analysis rather than a hypothesis testing. The text in the revised manuscript was revised accordingly.

10 L10-11, P16. Don't follow the logic. According to L6-7, the majority-based approach in Noah-MP is giving 2.3% HIGHER mean porosity than MPR. Why now the porosity field estimated by Noah-MP tends to have lower water holding capacity values?

Thank for pointing out this inconsistency. L6-7 refers to the mean over whole Pan-EU. L10-11 refers to a analysis in Germany whose results are reported in Fig. 6. The text in the revised manuscript was amended.

11 L19-21, P16. Does not read well. How could "dynamic(s)" be enhanced or constrained?

"Enhancing" and "constrained" are inappropriate terms. We should have written increasing or reducing the variance of soil moisture over time. This text was improved in the revised manuscript.

12 L3-4, P17. Not so apparent to me. It appears to me PCR-GLOBWB does not perform bad either. But this may be due to the difficulty to link the flux-matching test with the spatial patterns here.

If the parameters for both models are estimated based on streamflow only, then the model performance as reported in Table 2 tend to be comparable. ET estimates, however, differ greatly as shown in Fig.7 . In this case, both models in this experiment use collocated grids so that a cell at a coarser scale (30 arc min) have exactly the same number of underlying cells at finer resolutions (5 arc min), everywhere and for all models. Consequently, flux matching of ET made on two different resolutions is not a problem, and what is reported here is not an artifact of matching "spatial patterns". The text was improved to explain how the test was performed.

Response to Referee #2

1 *The authors make a nice case for the value of their multiscale parameter regionalization (MPR) method, analysing several aspects (and advantages) of the method. This is in principle a laudable thing to do. The manuscript itself, however, is quite frustrating to read. One the one hand it remains completely unclear what the novelty is. Large parts of the manuscript essentially repeat what has already been published earlier (as also acknowledged in the references provided).*

We are saddened by the fact that the reviewer consider that this manuscript lacks novelty. We have rewritten the introduction and modify large parts of the manuscript. The novelty of the manuscript is made more clear in the introduction and is based on the following key elements (see also the response to Ref.1):

1.1 Attempt to synthesize the progress towards seamless parameterizations in land surface(LSM) or hydrological models(HM). We provide examples to visualize how existing LSMs/HMs are estimating a fundamental parameter such as soil porosity (not found in literature) to make the case.

1.2 We propose, based on our own experience, a way forward that uses MPR and systematize its application by providing a "Protocol for evaluation of model parameterization" (This has not been publish before)

1.3 We implement this protocol to PCR-GLOBWB (also new piece of work and unpublished)

1.4 Carry out a series of experiments (inspired by E. Wood's recommendation) to demonstrate how to spot faulty parameterizations (also not published before).

1.5 Compare the effects of the parameterization on three models (mHM, WaterGAP, and PCR-GLOBWB) as part of these experiments (all using the same forcings and underlaying data). Also unpublished and novel material.

It should be clearly noted that none of these key elements belong to Mizukami et al. (under review) (hereafter [MCN+2017]). The main differences between [MCN+2017] and this manuscript can be found in P8 L7ff, P12 L16ff.

2 *On the other hand, the argument remains in places quite imprecise with a lot of quite sweeping (and not necessarily well substantiated) generalizations.*

We have removed generalizations that are not fully substantiated with our experiments. We would appreciate to know, which parts of our manuscript —according to the reviewer— need to be further substantiated, in the case that our justifications are not yet satisfactory.

3 *In addition, other approaches to parameter selection are quite outrightly dismissed while essentially no critical discussion on potential drawbacks or limitations of MPR are provided.*

There are a number of parameterizations approaches that have been tested in the literature. We provide a long list of references of comparisons between MPR and the most common techniques found in literature. These evaluations (HRUs, Standard regionalization, etc.) have been carried out in independent studies which are cited in our manuscript. Here a short list:

3.1 MPR vs. k-NN regionalization:
Samaniego, L., Bardossy, A., & Kumar, R. (2010). Streamflow prediction in ungauged catchments using copula-based dissimilarity measures. Water Resources Research, 46(2), `http://doi.org/10.1029/2008WR007695`

3.2 MPR vs. standard regionalization (no scaling)
Samaniego, L., Kumar, R., & Attinger, S. (2010). Multiscale parameter regionalization of a grid-based hydrologic model at the mesoscale. Water Resources Research, 46(5), http://doi.org/10.1029/2008WR007327

3.3 Lumped HRU, Distributed HRU, vs. MPR:
Kumar, R., Samaniego, L., & Attinger, S. (2010). The effects of spatial discretization and model parameterization on the prediction of extreme runoff characteristics. Journal of Hydrology, 392(1-2), 54-69. http://doi.org/10.1016/j.jhydrol.2010.07.047

3.4 MPR with satellite data (ungauged basin)
Samaniego, L., Kumar, R., & Jackisch, C. (2011). Predictions in a data-sparse region using a regionalized grid-based hydrologic model driven by remotely sensed data. Hydrology Research, 42(5), 338-355. http://doi.org/10.2166/nh.2011.156

3.5 MPR vs. HRU
Kumar, R., Samaniego, L., & Attinger, S. (2013). Implications of distributed hydrologic model parameterization on water fluxes at multiple scales and locations. Water Resources Research, 49(1), 360-379. http://doi.org/10.1029/2012WR012195

3.6 MPR across scales US basins
Kumar, R., Livneh, B., & Samaniego, L. (2013). Toward computationally efficient large-scale hydrologic predictions with a multiscale regionalization scheme. Water Resources Research, 49(9), 5700-5714. http://doi.org/10.1002/wrcr.20431

3.7 MPR in Pan-EU (transferability test, evaluation of states and fluxes more than 300 basins)
Rakovec, O., Kumar, R., Mai, J., Cuntz, M., Thober, S., Zink, M., et al. (2016). Multiscale and Multivariate Evaluation of Water Fluxes and States over European River Basins. Journal of Hydrometeorology, 17(1), 287-307. http://doi.org/10.1175/jhm-d-15-0054.1

Limitations and drawbacks of MPR w.r.t. to other methods have been mentioned in all our publications (see above). Based on your recommendation, we provide in the revised manuscript a summary of the limitations of MPR (see new section 3.5).

4 *In an exaggerated way, the authors present their MPR method, which I think has formidable potential, like in a product promotion folder.*

We never intended that this manuscript is considered as a "promotion" folder because we conducted a series of new experiments, showed a novel application to PCR-GLOBWB and give recommondations regarding the MPR application for the scientific community. We make very clear in the introduction the aims and scope of the manuscript.

5 *I think the manuscript would strongly benefit from (1) considerably reducing the redundancies with previous work (sections 1-3 can be *substantially* shortened) and (2) taking on a more critical perspective towards MPR. I think that many in the community will agree that it is a great tool. Instead of highlighting this over and over again, it would be more instructive to learn were its limitations are to allow further improvement.*

We reduced redundancies and improved the introduction greatly. Introduction was shortened to focus on the main issue of the manuscript. We recapitulated the MPR technique to have a self-consistent manuscript, if we move this section to an appendix, or refer to MPR to other manuscripts, perhaps is not the optimal solution for the reader. We summarized as much as possible. The flux matching postulation has not been published in present form before. As indicated above, limitations of MPR will be clearly written in section 3.5 of the revised manuscript .

6 *In general, I think it may be more interesting for a wider audience if the MPR technique was scrutinized and compared to other parameter selection and regionalization approaches *independent* of the model it is used for. In this manuscript it is applied exclusively with mhm if I understand correctly. In my understanding, it is a stand-alone method that should be*

applicable to any model. Would it not be fairer to be more consistent in the comparisons here, i.e. compare mhm with/without mpr and/or other models with/without mpr?

We agree that the MPR method should be implemented to other models as well as model comparisons with and without MPR should be conducted. The first point is addressed within this manuscript by implementing MPR into PCR-GLOBWB using the herein proposed and developed protocol. Further implementations are underway such as MCN+2017 to VIC. We think the reviewer will admit that implementing a new parameterization technique like MPR goes along with a substantial adoption of model code which needs a lot of experience and knowledge of the model it is applied to. Therefore, several groups in the world are working on that a literature about such comparisons will increase. As you may admit it is already a substantial contribution to implement MPR to one model as shown herein (PCR-GLOBWB). Comparisons with/and without MPR in mHM have been done, please see Kumar et al 2010, Samaniego et al 2010ab, Kumar et al 2013, etc. and there will be soon a manuscript from Rakovec et al. over 500 US basins showing the effects of MPR/NO-MPR with mHM and VIC. Regarding the general comparison of parameterization techniques we consider a set of 11 different models (CABLE, CLM, CHTESSEL, JULES, LISFLOOD, mHM, Noah-MP, PCR-GLOBWB, WaterGAP2, WaterGAP3, and HBV) as a significant number of models.

7 *The bottom-line is that I have the feeling that two quite independent things are not clearly separated here: the regionalization technique (MPR) and the models (mhm, etc). Here the text needs to become much more precise. Right now it seems to the reader that MPR is compared to e.g. the HBV model. This is not a valid comparison as these are completely different things. In contrast, it would be excellent to make the fact that MPR is a standalone tool clearer, as this may result in more modellers actually picking up the idea for their very own models (which they may not do at the moment due to its perceived exclusive association with mhm).*

Within this study we never intended to compare MPR to other models such HBV. We adapted the text such that this possible confusions vanished. One of our intentions here is, however, to compare parameters obtained with different models and thus with different parameterizations. We have decided to take the water holding capacity and/or porosity of the top soil as an example since this parameter is used in all LSMs/HMs and is somehow physically interpretable. We are further remarking that we have a problem with our LSMs/HMs that we need to solve if we would like to have scale invariant parameterizations and consistent model simulations. MPR is a possible avenue, a hypothesis that we are scrutinizing over and over again in thousands of river basins across the globe. You are right that MPR is freely accessible. It comes with the mHM software package which is available at `www.ufz.de/mhm` as an open source code. Currently, a model-agnostic version of MPR (called MPR-FLEX) is developed and presented by [MCN+2017] and has been applied to VIC up to now. Consequently, MPR is NOT exclusive from mHM now. We have also improved the text so that this “impression” that “MPR is compared to e.g. the HBV model” has vanished.

Specific comments

1 *p.2,l.5: why only over time and not also over space?*

This paragraph is not appearing in the revised manuscript.

2 *p.2,l.10-12: please avoid subjective terms as "elaborate" or “sophisticated”*

Done.

3 *p.2,l.28-29: is this actually true? Why would process dynamics that emerge at larger scales and that integrate several processes necessarily reduce "realism"? It is surely possible, but I do not think that it is a physical necessity. In any case, what is the meaning of "realism" in a situation where most of the system is de facto unobservable? How do we know if something is "realistic"?*

This sentence was removed from the introduction. But, in this respect, we have a pragmatic approach, if a model is able to reproduce surrogate observations in evaluation mode, then we consider that the unobservable states may be plausible. For this reason, we carried out the study reported in Rakovec, et al. 2016 JHM (see above).

4 *p.2,l.33-34: this is a sweeping generalization. What is actually meant by that? Why should an observed quantity, such as for example the stream flow recession constant have no physical meaning? Of course it has, albeit on the scale of the observation.*

We are referring here to transfer function parameters, for example those constants of the Clapp-Horberger PTF, which are basically found empirically and then used to link soil texture values (observable) with soil properties that may or not be observable (e.g., porosity). We are not referring to streamflow recession constant. The word streamflow was removed to avoid confusions.

5 *p.5,l.15ff and elsewhere: many things are mixed together here and the logic is not convincing. For a meaningful argument they need to be carefully disentangled. Is this about models? About parameter selection/calibration procedures? Parameter region- alization? It reads as if MPR does not rely on calibration, which is not correct. and why should lumped and/or semi-distributed models not be run with MPR-derived pa- rameters? Would this for a, say 100km2 catchment, not be the same as if running a distributed model with a 10x10km2 grid in mhm?*

Based on this comment, we consider that our text is not clear enough. We reformulated the text to highlight the most common regionalization techniques used in recent literature. Please refer to Samaniego et al. WRR 2010 to see a diagram that represent the steps done to estimate parameter for a given model. A simple conceptual model whose parameters are calibrated fail, in general, to perform well at cross-validation. This is what we are referring to. MPR improves transferability across scales and locations as shown in previous studies. In fact, this is what we demonstrated in Kumar et al. 2010 JoH. MPR could be used to estimate lumped parameters if a single cell covers the whole basin. In Kumar et al., mHM-MPR always performed better than a lumped mHM with no MPR.

6 *p.5,l.19 and elsewhere in the manuscript: much is made of "discontinuities". However, the authors do not provide a clear definition of what they mean. Nature is, in places, discontinuous (e.g. forest vs. grassland, north vs. south aspect, sharp transitions in geology, breaks topography, etc). thus it is not clear why models should not represent these discontinuities. I suppose that the authors want to say that between individually calibrated catchments discontinuities can occur, where there are in reality no discontinuities. But this needs to be made clearer.*

Thank you for the comment. In the revised manuscript we clarify our definition to avoid confusions. An example of artificially induced discontinuities by parameter calibration is shown is Fig. 4. We agree that there are natural discontinuities, we expect however, that it is unlikely that everywhere the model parameter and fluxes/state fields follow exactly the boundaries of the drainage area at a given location (see Fig.1 below). We call this negative effect calibration imprint, and we attempt to remove it with MPR. This artificial boundaries is what we call discontinuities. Nevertheless, we provide references to literature in P4 L19ff to illustrate our definition. Please see also the obtained parameter fields in Fig.1 (rebuttal) (below) as obtained by Merz and Bloeschl 2004 and by MPR from the study Rakovec et al. 2016 JHM.

7 *p.5,l.21-23: sure, but is this not also the case for distributed models and dependent on the calibration/parameter selection method?*

This is the case for any model even if one uses MPR on a single basin. This is the reason for showing the Fig.4a. Parameter estimation implies to have a representative sample. For this reason we attempt always to perform parameter estimation on several basins simultaneously, see Fig.4b. Single basin calibration is disadvantageous for any parameterization method because artifacts of the data can be “over-learned” which, in-turn, induce large bias somewhere else.

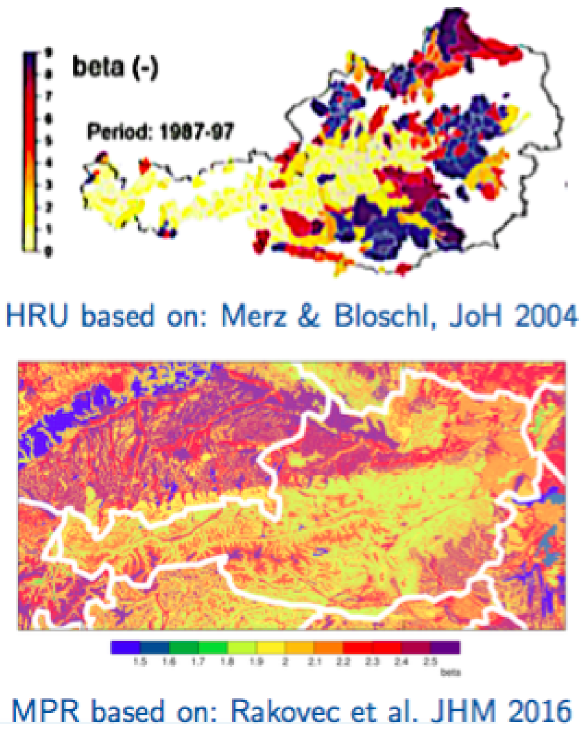

Figure 1: Non-seamless vs. seamless parameter fields

8 *p.6,l.29 and elsewhere: "CONUS": not necessarily every reader will be exposed to large scale studies employing these terms. Thus please avoid the use of fashionable abbreviations without first defining them.*

Sorry for not defining it before as it should be. Done.

9 *p.8,l.7: a question cannot be postulated. Please rephrase.*

Thank you for this remark. We mean "put forward". It will be rephrased in the revised manuscript.

10 *p.8,l.10-11: what is meant by "poor". How do you define it?*

A poor parameterization does not lead to flux-matching, exhibits low model performances (say KGE) in cross-validation experiments across scales and locations, and exhibits artificial "discontinuities", i.e. non-seamless fields. This definition is be clearly mention in the revised manuscript (e.g., Introduction).

11 *p.9,l.3: over-parameterization is only addressed in MPR if simultaneously calibrated to a high number of catchments and/or objective functions. Thus, it depends on how MPR is implemented and applied. Please rephrase.*

This sentence is now amended as suggested in the revised manuscript.

12 *p.10,l.17-18: how do you know that the parameters are "realistic"? See also comment above. Does this not also strongly depend on the assumptions in the upscaling relationships? It is always a question of how MPR (or other parameter selection techniques) are implemented and not a defining proprietary feature of MPR.*

This is a good question. The word "realistic" was removed to avoid confusion. The text was amended to improve clarity. The application of MPR involve many assumptions, PTFs, upscaling relationships, parameter estimation methods, etc. Visual impression of parameter fields may be useful but it is subjective. For these reasons, we need a formalized approach such as that described in Sec. 3.3: **Protocol for evaluation of model parameterization**, which was put forward in this manuscript, and depicted in Fig.2. The experiments presented in Sec. 4 were introduced to addresses this question.

13 *p.13, section 4: in many parts of the section it is unclear what is meant: the individual models or rather the parameter selection/regionalization techniques in the different model applications? These are different pairs of shoes and need to be carefully separated.*

We renamed the experiments to help explain the intention of this section. We also added a short paragraph at the beginning of section 4 to elucidate the aim of these experiments. In summary, we performed these experiments to help identify poor parameterization techniques using several models.

Response to Referee #3

1 *The main points of the paper are: (i) state-of-the-art LSMs and HMs do not have consistent and realistic parameter fields for land surface geophysical properties, and as a result do not satisfy a flux- matching condition (ii) the MPR technique can be used as a generic parameter estimation technique to greatly reduce these limitations (iii) a specific case of this improvement is demonstrated using the PCR-GLOBWB model. In my view the innovation is in the recognition of the problem across multiple models, the wider breadth of application of MPR, and the protocol needed to achieve this. To some extent the purpose of the manuscript is to demonstrate the very significant consequences of different parameter estimation approaches in large-scale LSMs/HMs, and to show the advantages of using MPR. In my view this is a relevant objective for scientific publishing, in relation to relatively new techniques such as MPR, because such examples provide specific examples to which the hydrological modelling community can more easily relate (as opposed to reading about the MPR technique in the abstract, or in relation to its application to a specific model). The main uncertainty for me is the extent to which this material is also contained in the submitted manuscript by Mizukami et al, as that manuscript is cited in relation to many of the main points made here. I leave this point for the Editor to consider.*

Thank you for the valuable comments and recommendations.

We described in detail the extend of Mizukami et al. (under review) (hereafter [MCN+2017]) and this manuscript in the Response to Referee #1 and Referee #2. The main differences between [MCN+2017] and this manuscript can be found in P8 L7ff, P12 L16ff.

[MCN+2017] is aiming at the development of "a model agnostic MPR system called MPR-flex, which is applied to the Variable Infiltration Capacity (VIC) model to produce hydrologic simulations over the contiguous USA (CONUS)". In [MCN+2017] no attempt has been made to verify the flux-matching condition of ET obtained with VIC using the MPR-flex parameterization across scales.

In this manuscript (hereafter [SKT+2017]) we:

1.1 Attempt to describe the progress towards seamless parameterizations in land surface(LSM) or hydrological models(HM). We present a short description of what has been made (the literature on the topic is quite extensive) and provide a simple example to visualize how existing LSMs/HMs are estimating a fundamental parameter such as soil porosity (not found in literature),

1.2 Propose, based on our own experience, a way forward that uses MPR and systematizes its application by providing a "Protocol for evaluation of model parameterization" (This has not been publish before),

1.3 Implement this protocol to PCR-GLOBWB (also new piece of work and unpublished),

1.4 Carry out a series of experiments (based on the spirit of the E. Wood's recommendation) to demonstrate how to spot faulty parameterizations (also not publish before), and

1.5 Compare the effects of the parameterization on three models (mHM, WaterGAP, and PCR-GLOBWB) as part of these experiments (all using the same forcings and underlaying data)

It should be clearly noted that none of these key elements belong to Mizukami et al. (under review) (hereafter [MCN+2017]) .

The scope of [MCN+2017] and [SKT+2017] are now clearly described in the first paragraph of Section 3.2 and in Section 3.5 (limitations of MPR). We also amended the conclusion that was misleading.

Specific comments

1 *Title: "Toward seamless hydrologic predictions across scales" This might be interpreted by readers as referring to seamless predictions across temporal scales, i.e. the linking of nowcasting with NWP. Perhaps "Toward seamless hydrologic predictions across spatial scales"?*

Thank you for the good suggestion. Done.

2 *P2 L2 "trade-offs that must be made to reach a final objective" missing word*

Done.

3 *P2 L9 "numerical weather prediction, land surface schemes, and hydrologic models" It would help to provide a reference or some text to enable readers to distinguish among these three terms. Many would know two of these terms, but far fewer could reliably distinguish all three.*

Key references are provided in the revised manuscript.

4 *P2 L29 "In this case, one states that a physical process is parameterized." It would be helpful to introduce the concept of sub-grid phenomena here, to distinguish between phenomena which are resolved by a given grid resolution, and those that are parameterised. Otherwise, the concept of parameterisation and references to "the missing (complex) processes" remains rather vague. The missing processes should all be sub-grid – anything else that is missing is simply a missing process.*

Thank for the recommendation. The concept of sub-grid phenomena that are not modeled will be introduced in the the revised manuscript.

5 *5. P3 L1 "Parameterizations in land surface models have increased in their complexity during the past decades, but the procedures to estimate constants for the parameterizations have not changed much." Has anything changed as grid sizes got smaller? Did any processes become resolved that were formerly parameterized?*

By comparing versions of land surface models, for example, multi-processes (parameterizations) have been introduced, e.g., in Noah-MP. Phenological processes and radiative transfer schemes have become extremely detailed in the new versions of Noah-MP and other LSMs. Runoff generation mechanisms, on the other hand, have not changed much in most LSMs/HMs. We make this clear in the Introduction (P2, L20ff) of the revised manuscript.

6 *P3 L7 "The reasons for the lack of progress in creating scale-invariant parameterizations are manifold." At this stage you have not established that scale-invariant parameterizations are either desirable or feasible (also relevant to P4 L24). From this point on in the paper it seems that the parameterization problem can be solved by scale-invariant parameterizations, but that there are no other credible paths being explored. I would like to see some mention in the Introduction of non-MPR approaches to parameter estimation which are also taking a serious approach to the problem. Alternative methods are unlikely to satisfy the flux- matching criteria, but they might be partly competitive, e.g. (i) other spatial scaling attributes (e.g. sidestepping the scaling problem by assuming scale-independent distribution functions), (ii) strong links to mapped geophysical attributes (e.g. regularisation), (iii) strong links to observed functional responses of hydrological systems (e.g. Yadav et al (Advances in Water Resources 30 (2007) 1756–1774)).*

Good point. Potential alternative ways are mentioned in the Introduction P3 L8ff. We consider, however, that exploring these alternative paths ways is out the scope of this manuscript to test them.

7 *P4 L19 "The numerical constants can be specified with a great level of precision, but the physical constants and parameters cannot be because they must be treated as random variables (Nearing et al., 2016)" I don't know the Nearing et al paper in detail, but I am surprised to hear that something termed a "physical constant" really requires treatment as a random variable. Surely if it is well enough defined to earn the moniker physical constant, then it can be determined experimentally to relatively high accuracy for practical purposes? Are the authors suggesting we should treat g as a random variable in hydrology because it is determined by measurement, which is subject to error? On the other hand, I accept that parameters may usefully be described as random variables.*

Depends on the accuracy and precision with which we know a physical "constant". Its description can be done by a density function having a know mean and quite small standard deviation. For example, we know the value of the standard acceleration due to gravity with high accuracy (no bias) and precision(very small stdev). In this case and for practical purposes of parameter estimation, we could treat it as a constant. This is not necessarily the case for other physical constants such as the thermal conductivity of a given soil type. In this case with need a transfer-function of infer it based on soil texture fields and other predictors. This section was, however, removed from the revised manuscript because the introduction was drastically cut to better focus on the topic of the manuscript.

8 *P4 I would like to see the term "seamless" defined in the introduction (the abstract provides this, but not the introduction), and particularly an argument made for why seamlessness is (in principle and/or in practice) a desirable attribute.*

Good point. We provided a definition in the introduction P2, L13 of the revised manuscript for consistency and to avoid miss interpretations.

9 *P9 The paragraph starting on L3 seems misplaced. The rest of the section is a description of MPR, whereas this paragraph is an assessment against criteria.*

This section was restructured in the revised manuscript to better explain the MPR approach.

10 *P9 L3 "Currently, MPR is the only method that consistently and simultaneously addresses the scale, nonlinearity and over-parameterization issues" If scale, nonlinearity and over- parameterization issues are the key criteria for assessment, then I would expect them to all be mentioned in the introduction; however, only scale really features in the introduction.*

Good point. These issues are briefly mentioned in the introduction. Many of these issues were introduced in other publications related to MPR (e.g., Samaniego et al. 2010b).

11 *P9 L26 This whole paragraph (slightly rewritten) might sit well in the introduction if there was some material there on regularization procedures.*

In this paragraph we analyze the effects of MPR on over-parameterzation. The Introduction was entirely rewritten, hence, we decided to keep this paragraph in this section. Thank you anyway for the suggestion.

12 *P9 L33 "Consequently, greater care should be taken in their selection." It is unclear what "greater" refers to. Are regularization functions being imposed without care? In which cases?*

If a regularization function is poorly chosen, or lack important predictors, the resulting parameter value might be badly estimated and its posterior distribution could be poorly estimated. For example, the Cosby et al. 1984 PTF is a very simple one (used in SCA-SMA) that relates porosity to sand content only. The application of this regularization function will under/over predict porosity in soils having low sand and high clay/loam fractions. See P2 L25, we mention this example to make clear our point in the revised manuscript.

13 *P11 L19 "Kling-Gupta efficiency (KGE) of the compromise solution ¿ 0.6" Some justification is needed for any threshold on KGE, as it is much easier to do well in some environments than others.*

This part of the protocol remains still subjective. It depends of on many factors such as the input forcings and quality of the land-surface properties. It is difficult to give a justification, it is only provided as a reference, and depends on the data quality, model, etc.

14 *P12 L3 "minimize the occurrence of discontinuities and ease the transferability of model parameters across scales and locations" These criteria for success should both have been outlined much earlier in the paper, either in the Introduction or at the end of the review.*

Good point. We introduced them in the introduction P2, L14.

15 *P12 L17 "which constitute the basis for the EDgE project" Needs a reference to the project, or delete if not relevant.*

We add the reference to `http://edge.climate.copernicus.eu` in the first reference to this project P11 L25.

16 *P18 L28 "MPR ... is feasible to implement in existing LSM/HMs whose goal should be seamless parameter fields across scales." The authors need to add an additional clause to this sentence (based on material from earlier in the paper) so it is clear WHY seamless parameter fields across scales are essential.*

Good point. We added a small clause to make the point. P19 L2

Toward seamless hydrologic predictions across spatial scales

Luis Samaniego[1], Rohini Kumar[1], Stephan Thober[1], Oldrich Rakovec[1], Matthias Zink[1], Niko Wanders[2,7], Stephanie Eisner[3,4], Hannes Müller Schmied[5,6], Edwin H. Sutanudjaja[7], Kirsten Warrach-Sagi[8], and Sabine Attinger[1]

[1]Department of Computational Hydrosystems, UFZ-Helmholtz Centre for Environmental Research, Leipzig, Germany
[2]Department of Civil and Environmental Engineering, Princeton University, USA
[3]Center for Environmental Systems Research, University of Kassel, Kassel, Germany
[4]Now at Norwegian Institute of Bioeconomy Research, Ås, Norway
[5]Institute of Physical Geography, Goethe-University Frankfurt, Frankfurt, Germany
[6]Senckenberg Biodiversity and Climate Research Centre (BiK-F), Frankfurt, Germany
[7]Universiteit Utrecht, Department of Physical Geography, Utrecht, The Netherlands
[8]Institute of Physics and Meteorology, University of Hohenheim, Stuttgart, Germany

Correspondence to: L. Samaniego (luis.samaniego@ufz.de)

Abstract. Land surface and hydrologic models (LSM/HM) are used at diverse spatial resolutions ranging from catchment-scale (1-10 km~~in catchment-scale applications to~~) to global-scale (over 50 km~~in global-scale applications. Application of~~) applications. Applying the same model structure at different spatial scales requires that the model estimates similar fluxes independent of the ~~model resolutionand~~ chosen resolution, i.e., fulfills a flux-matching condition across scales. An analysis of state-of-the-art LSMs and HMs reveals that most do not have consistent ~~and realistic parameter fieldsfor land surface geophysical properties~~hydrologic parameter fields. Multiple experiments with the mHM, Noah-MP, PCR-GLOBWB and WaterGAP models ~~are conducted to~~ demonstrate the pitfalls of ~~poor~~ deficient parameterization practices currently used in most operational models, which are insufficient to satisfy the flux-matching condition. These examples demonstrate that J. Dooge's 1982 statement on the unsolved problem of parameterization in these models remains true. ~~We provide a short~~ Based on a review of existing parameter regionalization techniques~~and discuss a method for obtaining seamless hydrological predictions of water fluxes and states across multiple spatial resolutions. The~~ , we postulate that the multiscale parameter regionalization (MPR) technique ~~is~~ offers a practical and robust method that provides consistent (seamless) parameter and flux fields across scales. ~~A~~ Herein, we develop a general model protocol ~~is presented~~ to describe how MPR can be applied to a ~~specific model, with an example of this~~ particular model, and present an example application using the PCR-GLOBWB model. ~~Applying MPR to PCR-GLOBWB substantially improves the flux-matching condition. Estimation of evapotranspiration without MPR at 5arcmin and 30arcmin spatial resolutions for the Rhine river basin results in a difference of approximately 29. Applying MPR reduce this difference to 9. For total soil water, the differences without and with MPR are 25and 7, respectively.~~ Finally, we discuss potential advantages and limitations of MPR in obtaining the seamless prediction of hydrological fluxes and states across spatial scales.

1 Introduction

... "If it disagrees with experiment, it's wrong". Richard P. Feynman

~~Developing a theory and translating it into a numerical (computational) model is a complex task because of the crucial trade-offs that must made to reach a final objective: to be able to reproduce the observations of a set of variables of interest. According to , modeling is an interactive research process that starts with the observation of a natural system (e.g.,~~

Land surface and hydrologic models (LSM/HM) are currently used at diverse spatial resolutions ranging from 1-10 km in catchment-scale impact analysis and forecasting (Christensen and Lettenmaier, 2007; Addor et al., 2014) to over 50 km in global-scale climate change simulations to estimate land surface boundary conditions of key state variables (Haddeland et al., 2011; Bierkens, 2015; Wanders and Wada, 2015). The fundamental conditions behind the applicability of the same LSM/HM model structure at different spatial scales requires that the model parameterizations are scale-invariant and that the model estimates similar fluxes across a range of spatial resolutions, i.e., it must fulfill the flux-matching condition across scales. A parameterization is a simplified and idealized representation of sub-grid physical phenomenon that is either "too small, too brief, too complex, or too poorly understood" to be explicitly represented by a model at a given resolution (Edwards, 2010). Parameterizations require variables called predictors, effective parameters and constants also called transfer-, global- or super-parameters (Pokhrel and Gupta, 2010). Superparameters are often parameters in empirical relationships that have been found with measurements in the field or in the laboratory. For example, regression parameters in pedo-transfer functions (Cosby et al., 1984). They are often tuned to represent observed variables and often have no physical meaning. These parameters constitute simplified surrogates to compensate for the ~~water cycle over a river basin) and aims to create a "mental" abstraction of the main elements necessary to faithfully describe the evolution of the system over time. Abstraction implies a reduction of system complexity, often formalized by a set of equations, which we call a model . Consequently, a model constitutes an elaborate hypothesis of the dynamics of the system that should be falsifiable . Model predictions should be confronted with new data to assess the ability of the model to reproduce them.~~ missing sub-grid processes that are not accounted for within a modeling system (Brynjarsdottir and O'Hagan, 2014).

~~The various types of environmental models (e.g., numerical weather prediction, land surface schemes, and hydrologic models) that exist today are the result of this elaborate formalization process, which led to a set of equations based on fundamental physical principles whose numerical solution is possible using sophisticated numeric algorithms and increasing computational resources. The scientific method used for modeling the water cycle has retained its fundamental structure since 1922, when L. Richardson wrote his seminal book in which the foundations for numerical weather forecasting were developed . A similar framework was employed by four decades later in formulation of the blueprint for a "distributed" hydrologic model. Interestingly, "distributed" was introduced to distinguish this new type of model from the lumped, black-box hydrologic model that neglected the spatial variability of the input forcing data, state variables and fluxes, or the semi-distributed hydrologic models that partly accounted for this variability by subdividing the basin domain into sub-units (e. g., the Stanford Watershed Model) linked by river reaches.~~

~~The speed at which numerical algorithms can be executed has changed dramatically since the mid 1940s~~An analysis of state-of-the-art LSMs and HMs reveals that most LSMs/HMs do not have consistent patterns of effective parameter fields for land surface geophysical properties across spatial scales, which indicates that their parameterizations are not scale-invariant. Parameter fields often exhibit artificial spatial "discontinuities" such as calibration imprints circumscribing river basins boundaries, and consequently they are not seamless. There are several reasons explaining this parameterization deficiency. With the advent of electronic computers~~in 1945~~, the performance of general circulation models (~~GCM~~GCMs), numerical weather prediction (NWP) models (Pielke Sr, 2013), land surface models ~~(LSMs)~~(Liang et al., 1994; Sellers et al., 1997; Niu et al., 2011), and hydrologic models ~~(HMs)~~ (Batjes, 1996; Lindstrom et al., 1997; van Beek et al., 2011; Samaniego et al., 2010b) has been increased mainly by improving model conceptualization (i.e., the number of process descriptions) and/or spatial resolution ~~if~~ since the storage capacity and computational power allowed for it ~~. For example, the GCMs employed for the Assessment Reports of the Intergovernmental Panel on Climate Change (IPCC) have doubled their spatial resolution every five years since 1990 . Notably, increasing the model resolution by a factor of two implies approximately ten times as much computing power. The hyper-resolution initiative in land surface/hydrologic modeling also opted for this pathway .~~

~~Despite the above mentioned improvements in model development, "there are scales and physical processes that cannot be represented (or resolved) by a numerical model, regardless of the resolution" . Consequently, we simplify the process representations in our environmental models at the expense of physical realism. In this case, one states that a physical process is parameterized. A parameterization is a simplified and idealized representation of the physical phenomenon at a given scale. These simplifications require variables called predictors and constants, also called transfer-, global- or super-parameters which are often tuned to represent observed variables (e.g., streamflow); and have no physical meaning. These constants often constitute simplified surrogates to represent the missing (complex) processes that are not accounted for within a modeling system .~~

~~Parameterizations in land surface models have~~ (Le Treut et al., 2007; Wood et al., 2011; Bierkens et al., 2014). As a result, parameterizations in LSMs have also increased in their complexity during the past decades ~~, but the~~ (Sellers et al., 1997; Fisher et al., 2014). The procedures to estimate ~~constants~~ effective parameters required for the parameterizations~~have not changed much. It is possible to assert that model parameterization is an old, ubiquitous, and recurring problem in land surface and hydrologic modeling for which no final solution has been found. This lack of coherent development has induced to conclude that the "parameterization of hydrologic processes to the grid scale of general circulation models is a problem that has not been approached, let alone solved. " A short survey of existing LSM/HMs presented in Section 2 allowed us to conclude that this statement is still valid~~, however, remained unchanged. For example, LSMs evolved from simple aerodynamic bulk transfer schemes with uniform description of surface parameters during the 1970s, to detailed LSMs having consistent description of the exchange of energy and matter between the atmosphere, the vegetation, and the land surface (Sellers et al., 1997). State-of-the-art LSMs, such as the Community Land Model version 4 (Bonan et al., 2011) and Noah-MP (Niu et al., 2011), however, still use quite simple pedotransfer-functions based on work of Clapp and Hornberger (1978) and Cosby et al. (1984) to estimate fundamental soil properties such as porosity (Oleson et al., 2013).

~~The reasons for the lack of progress in creating scale-invariant parameterizations aremanifold. The most important is likely~~ Among the reasons that have prevented the improvement of parameterization techniques are: 1) the lack of procedures and theories for linking physical properties (e.g., soil porosity) that can be measured at the field scale with “effective” parameter values that represent the aggregate behavior of the land characteristics at the scale of a grid cell required in LSMs or HMs. ~~recognized that the theory and the “constants” required in the dynamic equations (hereafter called effective parameters) “must be appropriate to the size” of the grid element but suggested that these constants should be found experimentally (p.108), if possible. Decades later, suggested that parameters should be selected so that simulations could be extended to ungauged areas, and stated that, even with very detailed representative measurements, it will be “necessary to extrapolate results ...of physical parameters to other points of the basin.” Linking effective parameters with point observations across a range of scales implies a proper knowledge of scaling laws governing the phenomena at hand, the certainty of its invariance, and detailed knowledge of the spatial distribution of geologic formations and soil properties. Due to the non-linearity of the involved processes, extrapolation across scales, which are orders of magnitude apart, is very problematic. In this regard, concluded in his seminal paper that the state-of-the-art regarding “linking phenomena at field scales (10-100 ha) and catchment scales (10-1000 km²) is an unresolved problem.” There have been many attempts to bridge this gap, but the results have not been very successful .~~

~~Another reason for this lack of progress is related to the approach to understanding the scaling problem mentioned by and~~ 2) Poor understanding of the scaling of parameters (Dooge, 1982) and its influence on the ~~hydrologic~~ hydrological response of the ~~entire system . As stated by : “the reason for this lack of progress is due in part to designing these experiments without regard to resolving this scaling question.” The representative elementary area (REA) concept was introduced , but others concluded that this concept has limited utility in hydrology and that there are many shortcomings related to its applicability . There is no general agreed upon theory that resolve this issue. The inclusion of sub grid-scale and surface~~ system (Wood, 1997; Wood et al., 1988). 3) Limited inclusion of sub-grid heterogeneity in hydrological parameterizations and multi-scale modeling of hydrologically relevant variables ~~have also been attempted . To the best of our knowledge, linking the REA concept and the concept of multi-scale modeling in LSM/HMs has not been attempted.~~

~~In vadose zone hydrology, scaling attempts were pioneered by and followed by seminal works on fractal approaches , stochastic perturbation methods , stream tube approaches , and connectivity-based methods . These theories have allowed for finding relationships to scale hydraulic conductivity, pressure head, total porosity and other soil properties from the pore scale to the field scale and have shown that effective parameters may be scale-dependent . They have only been used to upscale evaporation and transpiration fluxes at the field scale and have not been used at larger scales until very recently. applied the scaling proposed by to generate a global database of soil hydraulic properties. This data set, however, has not been used by any LSM/HMs up to now. The inverse modeling approach is frequently used to estimate soil-related parameters at regional scales . Stochastic and geostatistical theories have also been applied in saturated porous media for upscaling measured point-scale geophysical properties to the aquifer scale, including volume averaging theories and pre-asymptotic and asymptotic expansion theories. More recently, coarse graining methods were introduced to reduce the complexity of complex groundwater models and use effective aquifer parameters. Many of these seminal methodological and mathematical developments in scaling issues,~~

as suggested by Famiglietti and Wood (1995, 1994); Liang et al. (1996). 4) Lack of significant progress on the applicability of seminal upscaling theories (Miller and Miller, 1956; Dagan, 1989; Gelhar, 1993; Neuman, 2010; Kitanidis and Vomvoris, 2010) developed for sub-surface ~~hydrological problems , have not been applied or incorporated into regional-scale~~ hydrologic problems into LSM/HMs.

~~A third reason is related to the~~ And 5) lack of transparency in most of the existing LSM/HM source ~~code of~~ codes with respect to the meaning, origin and uncertainty associated with the hard-coded numerical values (i.e., parameters) either in the code or in the look-up tables ~~. It has been shown that these hidden parameters constrain the agility of the numerical model because they hinder the possibility of exploring their sensitivity on model outputs and the possibility of inferring them using observations . Model source code is often mixed, with no clear distinctions between physical or numerical constants (e.g., the acceleration of gravity g or $\frac{\pi}{2}$) and empirical effective parameters such as the soil porosity of a given soil type . noted that model output fluxes in the NOAH-MP model are sensitive to two-thirds of its applicable standard parameters, but most are hidden in the source code. From a statistical point of view, parameters and numerical constants are categorically very different. The numerical constants can be specified with a great level of precision, but the physical constants and parameters cannot be because they must be treated as random variables .~~(Mendoza et al., 2015; Cuntz et al., 2016).

~~In this study, we provide a short overview of the challenges and limitations of existing HM~~Consequently, it is possible to assert that model parameterization is an old, ubiquitous, and recurring problem in land surface and hydrologic modeling. Considering this lack of coherent development during the past decades, we can still concur with Dooge (1982, p.269) and say that the "parameterization of hydrologic processes to the grid scale of general circulation models is a problem that has not been approached, let alone solved."

There are potential methods available in the literature that may lead toward coherent parameterizations and prediction of water and energy fluxes in LSMs/~~LSM parameterizations in providing seamless predictions of water fluxes and states across multiple spatial resolutions. Through several control experiments, we demonstrate that a large portion of the predictive uncertainty in existing LSM/~~HMs~~originates from the poor estimation of effective parameters , which leads to a lack of scale invariance and thus to their poor transferability across scales and locations.~~. For example: 1) sidestepping the scaling problem of key model parameters by assuming scale-independent distribution functions with regionalized distribution parameters (Intsiful and Kunstmann, 2008), 2) find strong links between model parameters to mapped geophysical attributes via regularization procedures (Pokhrel and Gupta, 2010), and 3) find strong links between of observed functional responses of hydrological systems and geophysical characteristics (Yadav et al., 2007). These methods, however, alone may not satisfy the flux-matching criteria.

~~We~~ In contrast with these existing methods, we argue that the multiscale parameter parameterization (MPR) technique (Samaniego et al., 2010b) offers a ~~method to address the challenges of linking the~~ framework to link the field scale (observations) with the catchment scale (Dooge, 1982)~~and to account~~. MPR also accounts for the effect of the spatial variability and non-linearity of geophysical characteristics in the parameterization of hydrologic processes that operate at a range of spatial resolutions (Dooge, 1982; Wood et al., 1988). Depending on the conditions imposed to the parameter estimation technique, MPR can lead to parameterizations that satisfy the flux-matching criteria and hence contributes to obtain seamless parameter and water flux fields. Because MPR relies on empirical transfer functions and upscaling operators to link geophysical properties with model parameters, it provides a very effective procedure to transfer "global parameters" to scales and locations other than those used in ~~model~~ calibration (Samaniego et al., 2010a, b; Kumar et al., 2013b). This dependency on several transferable coefficients also contributes to minimize a serious drawback of spatially explicit models called "over-parametrization" (Beven, 1995). ~~Finally, we provide a modeling protocol that can be used as a guide to address the key question stated by :~~

In this study, we analyze to which extend existing LSM/HM parameterizations are limited to obtain seamless predictions of water fluxes and states across multiple spatial resolutions. Through several modeling experiments addressing Wood (1990)'s query (i.e., "What modeling experiments need to be performed to resolve the scale question ~~and what is the trade-off among model complexity, the physical basis for land parameterizationsand observational data for estimating modelparameters?"~~ ..."), we demonstrate that a large portion of the predictive uncertainty in existing LSM/HMs originates from the deficient estimation of effective parameters, which leads to a lack of scale invariance and thus to their poor transferability across scales and locations. These experiments also aim to help the modeler to reveal poor performing parameterizations, i.e., those that exhibit non-seamless fields. Finally, based on our past experiences and aiming to address the challenges stated above, we develop a protocol that systematizes the application of the MPR technique for any LSM/HM and demonstrate its effectiveness by implementing it into the PCR-GLOBWB model.

2 ~~Parameterization of hydrologic and land surface models~~Current parameterization techniques

2.1 ~~Representing spatial heterogeneity in HMs/LSMs: a brief review~~The state-of-the-art

The ~~core of the blueprint is embracing the spatial distribution of geophysical land-surface attributes in an HM~~most common parameterization techniques found in literature are: 1) look-up-tables (LUT), 2) manual or automatic calibration, 3) hydrologic response units (HRU), 4) representative elementary watersheds (REW), 5) a priori regularization functions, 6) simultaneous regionalization/~~LSM. A large part of the literature in hydrology is devoted to streamflow prediction, which, according to , is a hydrological variable that exhibits low dimensionality and represents the integral signal of a basin. As a result,~~ regularization functions, 7) dissimilarity-based metrics to transfer model parameters.

The simplest technique to assign a parameter value to a modeling unit (e.g., grid cell, HRU, sub-catchment) is based on a LUT. In this case, a categorical index associated with a modeling unit links it with information taken from an external reference file (i.e., the LUT) which maps this index with parameter values that are usually taken from the literature. This technique is commonly used in most of the ~~"precipitation-runoff" models were conceived as lumped models or as semi-distributed models and thus lead to poor spatial representation of geophysical parameters, state variables and fluxes~~(operational) LSMs such as CABLE, CHTESSEL, CLM, JULES, Noah-MP (Kowalczyk et al., 2006; Viterbo and Beljaars, 1995; ECMWF, 2016; Oleson et al., 2013; Best et al., 2011; Niu, 2011). A disadvantage of this method is the difficulty to perform sensitivity analysis (Cuntz et al., 2016). Moreover, the number of classes defined in LUT is often limited to a few (e.g., 13 soil classes in Noah-MP) resulting in non-seamless parameter fields that are not continuous.

~~Lumped~~ Manual or automatic calibration is a commonly used technique to parameterize spatially lumped hydrologic models (e.g., Crawford and Linsley, 1966; Burnash et al., 1973; Lindstrom et al., 1997; Edijatno et al., 1999; Fenicia et al., 2011; Martina et al., 2011; Andréassian et al., 2014; Singh et al., 2014) and semi-distributed ~~models often exhibit reasonably good efficiency in predicting streamflow observations at determined locations, but their performance is largely dependent on the parameter calibration. Due to their excessive reliance on parameter calibration , there is a trade-off: the performance at interior points of the basin or at other locations becomes worse than that for calibrated outlets . A second~~ hydrologic models (e.g., Leavesley et al., 1983; Kavetski et al., 2003; Lindström et al., 2010; Hundecha and Bárdossy, 2004; Merz and Blöschl, 2004; Hundecha et al., 2016). The aim is to minimize the disagreement between model simulations and observations. In the majority of the cases, the target variable is streamflow. The main drawback of this ~~type of model~~ parameterization technique is that the parameter fields, which are obtained by ~~combining~~ colocating lumped model parameters from sub-basins, are ~~unrealistic~~ doubtful because they exhibit sharp discontinuities along ~~the~~ individually calibrated sub-basin boundaries despite having spatial continuity in basin physical attributes like soil, vegetation and geological properties that govern spatial dynamics of hydrological processes (Li et al., 2012; Blöschl et al., 2013; Merz and Blöschl, 2004). In addition, these “~~patch~~ patchwork quilt” parameter fields exhibit significant sensitivity to the calibration conditions (Merz and Blöschl, 2004). Thus, ~~these models~~ models that are parameterized with this technique may exhibit (1) poor predictability of state variables and fluxes at locations and periods not considered in calibration and (2) sharp discontinuities along sub-basin boundaries in state, flux and parameter fields (e.g., Merz and Blöschl, 2004; Lindström et al., 2010). Parameter fields derived ~~using~~ from basin-wise “calibrated” lumped models lack spatial seamlessness, and thus are “inadequate representations of real-world systems” (Savenije and Hrachowitz, 2017). Moreover, excessive reliance on parameter calibration leads to deficient performance at interior points of the basin or at other locations at which the model was not calibrated (Pokhrel and Gupta, 2010; Lerat et al., 2012; Brynjarsdottir and O’Hagan, 2014).

There have been many attempts to improve the ~~realism and performance~~ parameterization of lumped and semi-distributed models by further discretizing the sub-basins into a given number of regions that exhibit nearly similar hydrologic ~~responses~~behavior, i.e., the so-called hydrologic response units (HRU) concept initially proposed by Leavesley et al. (1983) and further developed by others (e.g., Flügel, 1995; Beldring et al., 2003; Blöschl et al., 2008; Viviroli et al., 2009; Zehe et al., 2014). Unfortunately, ~~the~~ results obtained in these ~~modeling~~ parameterization attempts have not been very successful in realistically representing the spatial variability of model parameters, states and fluxes because of the lack of regionalized parameters and the unabridged reliance on parameter calibration to improve model performance ~~. Many attempts have been tested to enforce the continuity and monotony of the~~ (Kumar et al., 2010). Commonly, the effective parameters estimated for the HRUs are found by automatic calibration. Efforts have been made to enforce continuity on parameter fields (Gotzinger and Bárdossy, 2007; Singh et al., 2012), but with somewhat limited success ~~in~~ during the transferability of parameters across scales and locations. In addition, models parameterized using ~~the HRU concept~~ HRUs do not lead to mass conservation of water fluxes (i.e., flux-matching) when applied to scales other than those used ~~in~~ for calibration (Kumar et al., 2010, 2013b). Recent attempts have been made to improve the HRU concept to increase the seamless representation of parameters, states and fluxes (Chaney et al., 2016a). However, this concept has not been tested for scalability and seamlessness of the estimated fields at coarse ~~resolution.~~ resolutions. Lately, a thermodynamic reinterpretation of the HRU concept was proposed by Zehe et al. (2014), but to date, the implementation of this approach has not found its way into meso- to macro-scale LSMs/HMs.

Representative elementary watersheds (Reggiani et al., 1998) are an interesting theoretical concept, which scales mass and momentum balance equations that, to the best of our knowledge, have not been used to estimate effective parameters at meso- and regional scales. ~~Recently, a thermodynamic reinterpretation of the HRU concept was proposed by , but to date, the implementation of this approach has not found its way into meso- to macro-scale LSMs/HMs.~~

A priori regularization functions (e.g., pedo-transfer functions) were introduced by Koren et al. (2013) to ~~correct~~ ensure the “inappropriate randomness in the spatial patterns of model parameters”, i.e., the lack of seamlessness. Unfortunately, in this case, the parameters (or coefficients) of regularization functions were not subject to parameter estimation or to the verification of their ability to predict fluxes and states across various scales. The use of empirical point-scale-based relationships to link geophysical characteristics with LSM/HM parameters and the ~~generalized~~ assumption that their coefficients are universally applicable with certainty (e.g., the coefficients in the Clapp and Hornberger (1978) pedo-transfer functions) ~~is a major reason~~ are the major reasons for the proliferation of hidden parameters in LSM/HM code (Mendoza et al., 2015; Cuntz et al., 2016). It is of pivotal importance to understand that these point-scale relationships should not be applied beyond the scale at which they were derived.

Many types of regionalization (or regularization) approaches have been tested for semi-distributed and distributed models. According to Samaniego et al. (2010b), these approaches can be broadly classified into post-regionalization and simultaneous regionalization approaches, depending on if the regionalization function parameters (or global parameters) are estimated after (Abdulla and Lettenmaier, 1997; Seibert, 1999; Wagener and Wheater, 2006; Livneh and Lettenmaier, 2013) or during the model calibration (Fernandez et al., 2000; Hundecha and Bárdossy, 2004; Gotzinger and Bárdossy, 2007; Pokhrel and Gupta, 2010). None of these procedures consider the sub-grid variability of the model parameters or geophysical characteristics. Livneh and Lettenmaier (2013) noted that most of these regionalization procedures exhibit limited transferability because of the use of discrete soil texture classes as predictors, and very likely discontinuous parameter fields.

Recently, a dissimilarity-based regionalization technique was used by Beck et al. (2016) to generate an ensemble of global parameters of the HBV model at a 0.5° resolution for global-scale hydrological modeling. A shortcoming of this approach is the use of ad hoc nearest-neighbor interpolation of parameter fields to fill gaps where no donor basins are available in (geographically) surrounding regions~~(e. g., over the majority of Eurasia, Africa, South America, and South Europe).~~. Following a similar concept of that of Beck et al. (2016), the ~~HRU~~ parameterization method proposed by Bock et al. (2016) for the ~~CONUS~~ Contiguous United States (CONUS) will likely lead to discontinuous parameter fields ~~because the calibration regions are fixed to one hundred. The authors did not report the parameter fields obtained using this method. Consequently, it is very likely that LSM/HMs using this type of regionalized parameters would not exhibit realistic spatial patterns. Unfortunately, this hypothesis cannot be properly tested because of the lack of studies revealing the spatial patterns of regionalized parameters obtained using these regionalization techniques.~~ for reasons similar to those mentioned above.

Many attempts have been made in the land surface modeling community to address Dooge’s challenges, especially with respect to the transferability of model parameters across locations and scales and to obtain seamless parameter fields. One of the earliest prominent experiments was conducted in the Project for Intercomparison of Land-surface Parameterizations (PILPS) (Wood et al., 1998). In this project, calibrated LSM parameters were transferred from small catchments to their nearest computational grid cells. The results indicated that LSMs exhibited poor transferability across space, leading to significant differences in the partitioning of water and energy fluxes. For instance, Troy et al. (2008) used calibrated VIC model parameters from small basins to generate parameter fields for continental-scale land surface modeling by "linearly interpolating to fill in those grid cell not calibrated" on a sparse grid. As noted by Samaniego et al. (2010b), this type of regionalization is ~~not adequate~~ inadequate because of the nonlinearity of soil and geological formations. The spatial patterns of model parameters that would be obtained by ad-hoc extrapolations based on calibrated parameters ~~at~~ from small basins or grid cells would most likely lead to unrealistic parameter fields with ~~significant discontinuities , as in those shown by the VIC parameters obtained in a CONUS climate change projection assessment~~ spatial discontinuities circumscribing river basins, as shown in recent studies by Wood and Mizukami (2014) and Mizukami et al. (2017) for the VIC model parameters.

Recent community-driven efforts, such as the Protocol for the Analysis of Land Surface Models (PALS) and the Land Surface Model Benchmarking Evaluation Project (PLUMBER) (Haughton et al., 2016), indicate that the hurdles noted in PILPS have not been overcome. Thus, it is ~~possible to postulate that the poor~~ required to gain understanding on whether the inferior predictability of many LSMs evaluated with empirical benchmarks in the PLUMBER project (e.g., CABLE, CHTESSEL, JULES, Noah) may be the result of ~~poor~~ deficient parameterizations, among other factors.

2.2 ~~The state-of-the-art~~ Parameterization of ~~LSM~~soil porosity and available water capacity in selected LSMs/~~HM parameterization~~HMs

~~Recently, the~~ Above mentioned challenges that we face in estimating key physical parameters in LSM/HMs has been intensively discussed ~~by many authors~~ in many studies (Gupta et al., 2014; Bierkens et al., 2014; Bierkens, 2015; Clark et al., 2016, 2017; Mizukami et al., 2017; Peters-Lidard et al., 2017). To further visualize the problems ~~discussed above~~and to understand the deficiencies of current parameterization techniques, we selected a representative sample of ~~LSM~~LSMs/HMs used for research and/or operational purposes, namely: CABLE, CLM, JULES, LISFLOOD, Noah-MP, mHM, PCR-GLOBWB, WaterGAP2 (30 arcmin), WaterGAP3 (5 arcmin), CHTESSEL, and HBV. These models vary in process complexity and spatial resolution.

~~Soil porosity~~ We selected soil porosity as an example to visualize existing shortcomings because it is one of the most common parameters in many ~~LSM~~LSMs/HMs. This parameter controls the dynamic of several state variables and fluxes such as soil moisture, latent heat, and soil temperature, and its sensitivity has been demonstrated in various studies (Goehler et al., 2013; Cuntz et al., 2015; Mendoza et al., 2015; Cuntz et al., 2016). A representation of the porosity of the top 2 m soil column in these models over the Pan-EU is shown in Figure 1. The Pan-EU domain was selected for depiction, but we note that the problem is general and persistent across other domains (Mizukami et al., 2017). For cases in which ~~an~~ a HM does not use this parameter, the "available water capacity" (WaterGAP) or the "field capacity" (HBV) were selected as a surrogate due to their similarity with porosity. Both surrogate fields are normalized (in space) to ease their comparison with the porosity fields. Soil porosity is expressed in m^3m^{-3} ~~and can be compared~~ to ease the comparison among different models.

The following lessons can be learned from Figure 1: ~~(~~1) ~~There is a serious deficiency~~ there is a large variability in the parameterization of this key physical parameter because none of the analyzed models have comparable spatial patterns or comparable estimates at a given location. ~~Thus, its uncertainty is very large.~~ It should be noted that the definition of ~~this~~ the selected parameter is rather simple: it represents the ratio of the volume of voids to the total volume in the soil column. ~~We~~ One can now wonder how large the uncertainty of other ~~complex parameters whose relationship with soil properties is very nonlinear~~ parameters would be (e.g., hydraulic conductivity) ~~. (~~whose relationship with soil properties is very nonlinear. 2) The degree of seamlessness strongly depends on the level of aggregation and the upscaling of underlying soil texture fields. ~~(~~For example, porosity for WaterGAP is substantially different in spatial pattern and magnitude for 30arcmin and 5arcmin simulations. On the contrary, the spatial pattern and magnitude for porosity used in mHM remains almost unchanged for application at 30 arcmin and 5 arcmin resolution. 3) A parameter field becomes highly discontinuous and patchy when, for a given model, the parameter is calibrated in a limited domain (or basins) and then extrapolated to other regions (e.g., as shown in ~~panel (k) of Figure 1,~~ the panel corresponding to the HBV~~model). (~~). 4) These experimental results confirm the postulation of Dooge (1982) that the parameterization of the existing state-of-the-art LSM/HMs at large and continental scales is *still* an unsolved problem.

~~These facts~~ The analysis of current parameterization techniques allow us to ~~postulate~~ put forward the following questions: (1) Why ~~there are~~ are there such large differences between models in estimating a ~~physically interpretable parameter~~ parameter that has a physical meaning? (2) What are the consequences of poor parameterizations on the spatio-temporal dynamics of state variables and fluxes? (3) What are the consequences of model calibration on parameter fields? (~~3~~4) Are current model parameterizations scale invariant? (~~4~~5) Do the fluxes estimated with these models at various scales satisfy the fundamental mass conservation criterion (hereafter denoted the flux-matching test)? ~~(5) What are the consequences of poor parameterizations on the spatio-temporal dynamics of state variables and fluxes?~~

3 ~~The MPR~~ Seamless parameterization framework

3.1 The flux-matching postulation

The key postulation aiming at obtaining scalable (global) parameters that are transferable across locations and scales was proposed by Samaniego et al. (2010b) and further tested in Kumar et al. (2013b, a) and Rakovec et al. (2016b). We hypothesize that

Flux matching across scales leads to quasi scale-invariant global parameters $\hat{\gamma}$, thus:

$$\sum_i \sum_t \left| W_i(\hat{\gamma},t)a_i - \sum_{k\in i} w_k(\hat{\gamma},t)a_k \right| \to 0, \quad \forall i \in \Omega. \tag{1}$$

Here, k denotes the sub-grid elements constituting a given modeling cell i with area a_k. i denotes a modeling grid cell i with area a_i. W_i and w_k denote fluxes at two modeling scales ℓ_1 and ℓ_1', respectively, with $\cancel{\frac{\ell_1}{\ell_1'} = \frac{a_i}{a_k}} \left(\frac{\ell_1}{\ell_1'}\right)^2 = \frac{a_i}{a_k}$. Ω denotes the modeling domain, e.g., a river basin, and t a point in time. It should be noted that the topology of the cells at either level is not specified. Normally, rectangular grid cells are used for convenience, but this is not a necessary condition. This ~~stronger~~ strong flux-matching condition can be used as a penalty function or as an additional test to discriminate parameter sets obtained with conventional parameter estimation approaches.

3.2 ~~Regionalization and upscaling~~The MPR approach

Multiscale parameter regionalization (MPR), proposed by Samaniego et al. (2010b), aims to estimate model parameters that are seamless across scales, satisfy the flux-matching conditions (see Section 3.1), and enable the transferability of global or transfer-function parameters across scales and locations (Samaniego et al., 2010a, b; Kumar et al., 2013a; Wöhling et al., 2013; Livneh et al., 2015; Rakovec et al., 2016a). The development of MPR is ongoing work. Regionalization functions used in MPR for the mHM model (www.ufz.de/mhm) by Samaniego et al. (2010a) were further improved by Kumar et al. (2013b)~~and more~~. More recently, a model-agnostic implementation of MPR has been proposed by Mizukami et al. (2017) ~~.~~

~~Currently, MPR is the only method that consistently and simultaneously addresses the scale, nonlinearity and over-parameterization issues . The MPR approach also addresses the principle of scale-dependent subgrid parameterization (i.e., "net fluxes must satisfy the conservation of mass") proposed by but does not adhere to Beven's other principles, such as that sub-grid parameterizations may be data-~~ and ~~scale-dependent (principle 3 and 4), because exhaustive tests reported in the above mentioned references carried out over hundreds of river basins do not appear to support them. We find MPR to be a robust technique that has the ability to provide "effective parameters" and is capable of addressing the scaling problem; in this sense, it diverges from the Beven's view that these "effective parameters" are an "inadequate approach to the scale problem". Furthermore, MPR differs on the regionalization and aggregation scheme (i.e., patch model areal weighting) proposed by~~ tested in the VIC model in over 500+ basins in the CONUS. The study of Mizukami et al. (2017), in contrast to the present study, does not include flux-matching tests nor the evaluation of model skill across different spatial scales.

The scaling problem in MPR is ~~approached by addressing the existence of~~ addressed by using process specific representative elementary areas (REA) that determine the minimum computational grid size ℓ_1 at which the continuum assumptions can be used without explicit knowledge of the actual patterns of the topography, soil, or rainfall fields (Wood et al., 1988). The REA of a specific process, such as streamflow, can be determined by conducting a careful sensitivity analysis as shown by Samaniego et al. (2010b). To estimate an "effective" model parameter (e.g., total soil porosity) at the selected modeling scale, it is first necessary to estimate its variability at a much finer scale $\ell_0 \ll \ell_1$ ~~so~~ such that the effects of its spatial heterogeneity can be adequately represented. In other words, the parameter at the fine scale ℓ_0 represents the minimum support at which the proposed equations are still valid. Barrios and Francés (2011) indicated that a suitable estimate of ℓ_0 for a given parameter could be near its correlation length. The sub-grid variability of a parameter β_0 depends, in turn, on the spatial heterogeneity of geo- and bio-physical characteristics ($\mathbf{u}_0$), such as terrain elevation, slope and aspect, soil texture, geological formation, and land cover, which are now available at hyper-resolution for the entire globe. The mathematical relationships that link model parameters with these characteristics at the finer resolution are called pedo-transfer, regionalization or regularization functions f (Clapp and Hornberger, 1978; Cosby et al., 1984; Wösten et al., 2001). The constants required in these functions are usually denoted as global parameters $\hat{\gamma}$, thus $\boldsymbol{\beta}_0 = f(\mathbf{u}_0, \hat{\gamma})$. Note that the fields $\boldsymbol{\beta}_0$ and $\mathbf{u}_0$ are dependent on space and time, but the vector $\hat{\gamma}$ is not.

Regularization functions are commonly used in mathematics and statistics to solve ill-posed problems (which is the case when the parameters of a distributed LSM/HM are determined by calibration) and/or to prevent over-fitting. The direct consequence of the regularization is the substantial decrease in degrees of freedom of the optimization problem because the cardinality of the gridded parameter fields $\cancel{\#\boldsymbol{\beta}_0}$ $\#\{\boldsymbol{\beta}_0\}$ is orders of magnitude larger than that of the vector of the global parameters $\cancel{\#\hat{\gamma}}\#\{\hat{\gamma}\}$. Hence, MPR is a parsimonious ~~(not over-parameterized)~~ parameterization technique that offers spatially continuous ~~model~~ parameter fields and removes spatial discontinuities in water fluxes and states, as observed by Gotzinger and Bárdossy (2007) and discussed by Mizukami et al. (2017). From the Bayesian point of view, the regularization functions impose a prior distribution on the model parameters. Consequently, greater care should be taken in their selection.

The second step of the MPR approach consists of upscaling the sub-grid distribution of ~~every~~ a regionalized parameter to the modeling scale. In other words, $\boldsymbol{\beta}_1 = \langle \boldsymbol{\beta}_0 \rangle$. Here, the symbol $\langle \cdot \rangle$ represents an averaging or scaling operator that is parameter-specific, and thus $\boldsymbol{\beta}_1$ denotes the upscaled effective parameter field. It is important to note that this scaling operator is not necessarily the arithmetic mean.

A schematic representation of the MPR procedure can be seen in Figure 2. In short, the motto of MPR is "estimate first, then average" whereas other existing regionalization methods follow the opposite approach of "average first, then estimate." Because the processes in LSM/HMs are highly nonlinear, this sequence of operations does not commute. The consequences can be dramatic (~~see Figure 7~~to be shown in the results section). The latter, which is the standard approach, does not preserve fluxes/states across scales, whereas MPR does to a ~~great~~ considerable extent. The key question here is in finding the right scaling rule for the model parameters such that the fluxes/states are preserved across a range of spatial scales.

Model parameters at the ℓ_1 scale (i.e., 1 km to 100 km) are called "effective" parameters because they cannot be measured by physical means at this resolution and can only be inferred by heuristic relationships $f(\cdot)$. Thus, it is essential that the inequality $\ell_0 \ll \ell_1$ is fulfilled so that the law of large numbers leads to stable estimates of the effective parameter $\boldsymbol{\beta}_1$ having low uncertainty. Since every LSM/HM (e.g., those mentioned in Section 2) contains "effective" model parameters, depending on heuristic relationships (that are hidden in the source code in many cases (Mendoza et al., 2015; Cuntz et al., 2016)), it is logical that existing LSM/HMs are subject to parameter uncertainty. These models can be treated as stochastic models, even though their governing equations are deterministic in nature and based on physical principles such as the conservation of mass and energy (Clark et al., 2015; Nearing et al., 2016). Effective parameters should not be the pure result of a blind calibration algorithm. MPR varies from other regionalization approaches in that the introduced relationships ~~lead to realistic parameter fields~~ may lead to seamless parameter fields and model simulations fulfilling the flux-matching condition.

Currently, MPR is the only method that consistently and simultaneously addresses the scale, nonlinearity and over-parameterization issues if global parameters are estimated simultaneously at multiple locations (i.e., basins). The MPR approach also addresses the principle of scale-dependent subgrid parameterization (i.e., "net fluxes must satisfy the conservation of mass" proposed by Beven (1995)) but does not adhere to Beven's other principles, such as that sub-grid parameterizations may be data- and scale-dependent (principle 3 and 4 in Beven (1995)), because exhaustive tests reported in the above mentioned references carried out over hundreds of river basins do not appear to support them. We find MPR to be a robust technique that has the ability to provide "effective parameters" and is capable of addressing the scaling problem; in this sense, it diverges from the Beven's view (Beven, 1995, p.507) that these "effective parameters" are an "inadequate approach to the scale problem". Furthermore, MPR differs on the regionalization and aggregation scheme (i.e., patch model areal weighting) proposed by Beven (1995, p.520).

The selection of regionalization functions and scaling operators is fundamental to ensure the transferability of global parameters across scales and to guarantee the seamlessness of parameter fields across scales, e.g., from ℓ_1 to $2\ell_1$ ... and so on. Samaniego et al. (2010b) proposed that the key to determining them is the flux-matching condition mentioned above. A seamless parameter field $\boldsymbol{\beta}_1$ can be interpreted as the corollary of the flux-matching condition. Moreover, MPR employs geophysical properties at ℓ_0 that allow for a representative sample at the hyper-resolution promoted by Wood et al. (2011) and Bierkens et al. (2014).

3.3 Protocol for ~~evaluation of model parameterization~~implementing the MPR approach

The development of LSM/HMs and their parameterizations should be guided by a strict hypothesis driven framework (Nearing et al., 2016) that aims at finding parsimonious and robust parameter sets that fulfill the flux-matching condition and a number of efficiency metrics that are not used during the parameter estimation phase. A multivariate, multiscale evaluation assessing the reliability of model simulations should follow the scheme presented in Rakovec et al. (2016a). Based on our ~~experience, we suggest the following procedure~~ previous experiences, we synthesize a formalized scheme (i.e., protocol) for systematically implement the MPR technique in other LSMs/HMs with the aim to obtain a robust and seamless parameterization. A graphical depiction of the estimation procedure at multiple scales is shown in Figure 2.

1. Retrofit the source code of an LSM/HM so that all model parameters are exposed to analysis algorithms. Parameters are the values of a model that can be considered random variables, i.e., those that are subject to various outcomes and can be fully defined by a probability density function. Parameters should not be confused with numerical or physical constants.

2. Determine a set of the most sensitive model parameters through a sensitivity analysis (SA). For computationally expensive LSMs such as CLM or Noah-MP, computationally frugal methods such as the elementary Effects method (Morris, 1991), its enhanced version such as that proposed by Cuntz et al. (2015), or the distributed evaluation of local sensitivity analysis (DELSA, Rakovec et al., 2014; Mendoza et al., 2015) are of particular interest because use of the popular standard Sobol' method (Sobol', 2001) can be computationally expensive although still possible (Cuntz et al., 2016).

3. Regionalize sensitive model parameters ~~the~~ that exhibit marked spatial variabilities. The selection of the regionalization function $f(\cdot)$ can be guided by existing literature or by step-wise methods (e.g., Samaniego and Bárdossy, 2005). This regularization step should be conducted at the highest available spatial resolution for all predictor fields. This resolution is denoted as level ℓ_0. The output of the regularization is the parameter field $\boldsymbol{\beta}_0$.

4. Estimate effective parameter fields β_1 using upscaling operators based on the underlying sub-grid variability β_0. The scale ℓ_1 is determined by synthetic experiments aimed at finding the optimal REA for processes related to the parameter in question (Samaniego et al., 2010b; Kumar et al., 2013b).

5. Estimate the global-parameters $\hat{\gamma}$ using standard optimization algorithms (simulated annealing, shuffled complex evolution (SCE), dynamically dimensioned search (DDS)) by minimizing a compromise metric that includes observations at multiple scales and locations (Duckstein and Opricovic, 1980; Rakovec et al., 2016a). The compromise metric could also include hydrologic signatures to extract as much information from a time series as possible (Nijzink et al., 2016).

6. Perform multi-basin, multi-scale, multi-variate cross-validation tests to evaluate the robustness of the regionalization functions, scaling operators, and global parameters (Rakovec et al., 2016a).

7. If the cross-validation tests provide satisfactory results (e.g., Kling-Gupta efficiency (KGE) of the compromise solution > 0.6), then evaluate the flux-matching condition given by eq. 1. If the total error is too large to be tolerated, repeat steps 3 to 8.

8. Evaluate the parameter seamlessness and the preservation of the statistical moments of fluxes and states across scales (seamless prediction step in Figure 2).

It should be noted that any of the steps above can be tested within a sequential hypothesis testing framework (Clark et al., 2016). A substantial difference from a standard model optimization exercise is that the transfer function $f(\cdot)$ (step 3) and the upscaling operator (step 4) can also be modified in the modeling protocol.

Failure to satisfy the imposed condition, such as the flux-matching test, after exhaustively testing the options in steps 3 to 6 may indicate deficits in process understanding and/or poor data. Consequently, the evaluation step should also provide guidance on detecting and separating the errors stemming from process conceptualization (modeling) and input data. ~~A graphical depiction of the estimation procedure at multiple scales is shown in Figure 2.~~

3.4 Seamless parameter fields across multiple scales using MPR

In Section 3.2, it was postulated that the MPR technique aims at estimating seamless parameter fields across scales which minimize the occurrence of artificial discontinuities and ease the transferability of model parameters across scales and locations. The latter has been tested and reported in many studies in Europe, USA, and other basins worldwide (Samaniego et al., 2011; Kumar et al., 2013b, a; Rakovec et al., 2016b, a). In this study, we provide evidence in favor of the former postulation.

To achieve this goal, the mHM model is parameterized using the multiscale parameter regionalization technique (MPR) (Samaniego et al., 2010b) with hyper-resolution fields of geophysical characteristics at $\ell_0 = 500$ m resolution as input. Among them, the land cover data was obtained from the Corine data sets (land.copernicus.eu/pan-european/corine-land-cover), and the soil texture information was derived from SoilGrids (soilgrids.org). These very detailed and homogenized soil texture fields provide the fractions of clay and sand, mineral bulk density, and fraction of organic matter for six soil horizons up to 2 m deep. A hyper-resolution digital elevation model (DEM) over Europe (approximately 30 m) from the GMES RDA project

(EU-DEM; www.eea.europa.eu/data-and-maps/data/eu-dem) was used to derive terrain characteristics such as slope, aspect, flow direction. The underlying hydro-geological characteristics are based on the International Hydrogeological Map of Europe (IHME; www.bgr.bund.de/ihme1500), available at a 1:1 500 000 scale. Details on the pedo-transfer function used for these simulations can be found in Livneh et al. (2015). mHM global parameters were obtained by closing the water balance over selected river basins in Europe (Rakovec et al., 2016a).

Based on these settings, which constitute the basis for the EDgE project (edge.climate.copernicus.eu), we estimated porosity fields at three modeling resolutions of ℓ_1= 5, 10, and 25 km, based on the same ℓ_0 support information. Following the MPR procedure depicted in Figure 2, the parameter fields for the mHM model at these three resolutions can be estimated. Results are shown in Figure 3.

The results illustrate that the MPR approach can preserve the spatial pattern of the porosity fields (see panels (a), (b) and (c) in Figure 3) and the first and second moments of its probability density function shown in panels (e)-(g). Two-sample Kolmogorov-Smirnov tests indicate that there is insufficient evidence to reject the null hypothesis that any of the three possible pairs of empirical distributions were drawn from the same unknown distribution. ~~Due to the similarity of the empirical distribution functions (EDF) of the porosity fields~~ This highlights that the MPR approach leads to consistent parameter fields across scales. In this case, the mean ~~value estimated at resolutions of 5, 10, and 25km is approximately~~ porosity is estimated to be 0.42 ~~m^3m^{-3} in all three cases.~~ m3 m-3 independent of the scale.

~~The recently derived soil porosity field at 15arcmin (≈25km)over the Pan-EU , hereafter denoted as Miller-Miller Scaling (MMS) , is included~~

3.5 Limitations of the MPR approach

The MPR approach, as any method, has some limitations. One of the crucial aspects of MPR is the selection of transfer functions and upscaling operators. Existing theories could be the first guess, but in the case that nothing is available, the protocol proposed in Section 3.3 could be used to guide the search of robust transfer functions. Testing the model parameterization for flux matching conditions across a range of basin and spatial scales may help to identify adequate upscaling operators. This procedure, although tedious, is the only solution for the moment.

In the case that some state variables change over time (e.g., land cover/use), or during parameter estimation, the MPR algorithm have to be linked to the model because every time a global parameter ($\hat{\gamma}$) is re-estimated, all related model parameters model (β_1) have to be updated as illustrated in Figure ~~3d to visualize the effects of the pedo-transfer functions and the scaling operators. It should be noted, however, that both approaches~~ 2. The computational cost of performing MPR is therefore larger than other parameterization method discussed before.

Another limitation of the applicability of the MPR technique until recently was its availability only as an intrinsic module of the mHM model (www.ufz.de/mhm). This implies that tailored algorithms (i.e., ~~MPR and MMS) use the same input SoilGrids data set at $\ell_0 = 1$km. The main differences between both approaches stem from the type of regionalization functions and upscaling operators used, and the procedure to estimate the PTF parameters. It should be noted that the flux-matching test was only verified in MPR. To generate the MMS dataset, used the ROSETTA-based pedo-transfer-function (PTF) proposed by~~

~~and the upscaling method described by . MPR, on the other hand, employs the PTF proposed by and the harmonic upscaling operator proposed by (see for more details) . Both methods lead to seamless fields because they follow the principle “first estimate then average” which considers the~~ source code) to perform the regionalization and upscaling of parameters for a target LSM/HM have to be developed from scratch, as it is demonstrated here as a case study for the PCR-GLOBWB model. This activity is of course time consuming and not pleasing due to its complexity. For this reason, Mizukami et al. (2017) have started a community effort to develop a model-agnostic MPR implementation (MPR-flex), which has been so-far evaluated for the VIC model.

The availability of high resolution bio-physical characteristics at the spatial scale ℓ_0 constitutes another limitation of the applicability of MPR. Since the sub-grid ~~variability of the soil porosity. The differences between the MPR and MMS derived porosity fields (panels (c) and (d) of Figure 3), however, are striking. In general, MMS tends to have lower porosity values than MPR. The spatial mean of MMS is 5lower than MPR but its standard deviation is in the same range as that of MPR at 10km resolution.In particular,~~ variability is fundamental to estimate robust effective parameter values at coarser scales, the ~~porosity obtained by MMS over the northeastern part of Germany, eastern Europe, and Scandinavia are underestimated compared with MPR .~~ minimum scale at which a model can be applied (ℓ_1) is strongly determined by the data availability. For example, if the soil data is available for the Pan-EU domain at ℓ_0=250 m, the ℓ_1 should not be lower than 1000 m, so that each modeling cell (ℓ_1) have a representative number of underlying sub-grid cells (ℓ_0).

~~The main advantage of MPR over MMS is that the former has the ability, if used in a LSM/HM, to preserve the mass balance of water fluxes (e~~MPR has been mainly developed for a hydrologic model representing the water cycle. However, land surface models also include the energy and carbon cycles and thus have greater complexity. In particular, they have more detailed representation of vegetation. It is a topic for future research to develop a MPR approach (i.e., transfer functions and upscaling operators) for plant functional type-specific parameters such as carboxylation rate and the slope of the Ball-Berry equation for stomatal conductance (Ball et al., 1987), which are required for a successful implementation of MPR in LSMs. ~~g., evapotranspiration) and states across scales, such that the resulting fields have similar spatial patterns and moments across scales as shown in Figure 3(a-e~~

Finally, the computational effort for MPR is also considerably larger in comparison with other methods, because of its requirement to estimate model parameters (β_0) at the highest resolution at which the bio-physical characteristics are available. The computational time, however, could be substantially reduced by using a restart file (i.e., ~~e-g). The method of , has yet to be tested for these fundamental scaling properties when applied to LSMs/HMs. Nevertheless, this experiment allows to conclude that the PTFs and the upscaling operators should not be applied independently from a LSM/HM because the resulting effective parameterswould have large biases that would likely lead to large predictive uncertainties.~~ a data set containing a copy of all parameters, state variables, and fluxes of a model at a given point in time). If this capability is available, the MPR estimation can be greatly reduced for operational simulations because the effective parameter fields and past modeled states do not need to be estimated often.

4 ~~Experiments that reveal inadequate parameterizations~~

4 Experiments to reveal non-seamless parameterizations

4.1 ~~Visualization of artifacts in state and flux fields~~

In this section we perform four modeling experiments, inspired on Wood (1990)'s recommendation, to investigate: 1) the effects of the over-calibration of global parameters on the spatial patterns of modeled state variables. 2) The effects of a parameterization technique on the spatial pattern of effective parameters. 3) The effects of a parameterization technique on the dynamics of a state variable. And, 4) the effects of not satisfying the flux-matching condition on simulated flux across different spatial scales. In these experiments four models are employed: mHM, Noah-MP, PCR-GLOBWB, and WaterGAP.

4.1 Effects of on-site model calibration

As noted in the introduction, on-site (basin-specific) parameter estimation based on HRU or similar techniques (~~based on~~ such as clustering grid cells or sub-basins into regions that exhibit quasi-similar hydrological behavior) leads to non-seamless parameter fields such as those reported in Merz and Blöschl (2004). Here, we go one step further to show the consequences of this common practice on state variables such as soil moisture. Our postulation is that an on-site calibration of global-parameters $\hat{\gamma}$ leads to biased state variables even with ~~sophisticated~~ regularization techniques such as MPR. To falsify this postulation, we performed two model simulations denoted "on-site" and "multi-site" calibration schemes. In both cases, we used the mHM setup described in Rakovec et al. (2016b) over the Pan-EU domain at a 0.25° resolution.

In the first ~~experiment, we performed~~ simulation, we perform on-site calibrations at 400 river basins in the Pan-European domain. Subsequently, the respective optimized parameter sets are used in each corresponding basin to generate the target variable, in this case, the daily soil moisture of the top 1 m soil column. Lastly, daily soil moisture fields ~~were~~ are assembled using the independent basin simulations for the entire Pan-EU domain. The results of this experiment are shown in panel (a) of Figure 4 for a day in August 2005. In the second ~~experiment~~simulation, the global-parameters $\hat{\gamma}$ are estimated simultaneously for a set of 13 basins covering various hydro-climatic regimes in the Pan-EU domain. The corresponding soil moisture field for the same point in time is depicted in panel (b) of Figure 4.

The first ~~experiment~~ simulation shows clear evidence of strong spatial ~~discontinuity~~ imprint in the soil moisture fields that is easily identifiable because the shapes of the constituent river basins (Figure 4a) are apparent. Another interesting feature is a strong wet bias in a basin located in center of the Iberian Peninsula compared to its neighboring regions. Wet soils during this period are very unlikely because the entire region was enduring a prolonged and extreme drought. Moderate dry-bias is apparent in basins in southwest Germany, and a strong dry-bias was detected in basins in west Croatia, south Lithuania, south Hungary and north Bosnia and Herzegovina. Conversely, the soil moisture field obtained with the multi-basin parameter estimation does not exhibit these nuisances and thus can be regarded as a spatially seamless field. In this case, parameter estimation with a large sample of geophysical characteristics and many streamflow times series to estimate efficiency measures leads to a well-posed parameter estimation problem.

Based on these results, it can be concluded that parameter sets obtained using the on-site parameter estimation technique does not lead to seamless parameter fields or state variables. Moreover, automatic optimization algorithms, such as SCE or DDS, tend to over-learn from time series with large observational errors, which in turn leads to poor identifiability of parameters (Brynjarsdottir and O'Hagan, 2014) and biased simulations, as demonstrated above. Consequently, parameter estimation should be performed with a representative sample of basins that adequately cover the variability of hydrological regimes and geophysical properties (e.g., soil types) (Kumar et al., 2015). It is worth noting that if the parameters of a model are estimated in a small basin with very few soil types, a single geological formation, or very flat terrain, then it is very likely that ~~the estimated parameters are biased to a specific basin setting andare not useful~~ some parameters cannot be constrained during calibration. The obtained parameter set is biased to the specific basin in which it has been estimated and, hence, it is not skillful for seamless and continental scale simulations.

4.2 ~~Detecting effects~~ Effects of ~~ad hoc regionalization techniques~~a parameterization technique on spatial patterns of effective parameters

The effects of the commonly used parameterization techniques ~~used~~ to generate the porosity fields of LSMs (such as CHT-ESSEL and Noah-MP depicted in Figure 1) are important to investigate. These fields ~~were~~ are obtained by combining the majority (or dominant) upscaling operator and a look-up table containing categorical values of model parameters tabulated for a limited set of dominant soil types (e.g., Niu (2011, p.20.), ECMWF (2016, p.137)). The ~~majority upscaling is represented as $\beta_{1,i} = \beta_{0,k}$, where $\beta_{1,i}$ denotes the upscaled parameter at grid cell i and $\beta_{0,k}$ is the parameter value corresponding to the most common soil type class k in level-0 within i. The soil types are given in a deterministic look-up table with m classes, and $k \in m$. The~~ majority-based operator is mostly used for estimating grid-specific vegetation classes in LSMs (Li et al., 2013).

The porosity field, based on a majority upscaling for the Noah-MP model used in EURO-CORDEX (www.euro-cordex.net) at an approximately ~~12-km~~ 12 km resolution, is depicted in Figure 1~~g~~. Compared with the other model derived porosity fields, the Noah-MP field appears to be most homogeneously distributed in space~~, and it~~. It is very likely that the spatial heterogeneity is under-represented in this case as the default soil LUT contains only thirteen soil classes. It should be noted that a model such as CABLE that uses a porosity field with an approximately ~~100-km~~ 100 km resolution has a larger variability than that of Noah-MP at 12 ~~km. However, one can argue that the underlying soil texture map , and not the upscaling method, might cause the lack of variability~~ km.

~~To test this hypothesis, the~~ The following experiment is carried out to evaluate whether the variability of the soil map or the upscaling operator has a larger effect on the derived porosity field. The highest resolution soil map available for Europe ~~was~~ is used and applied in the same manner to derive porosity fields as described above. The texture field is provided by the SoilGrids dataset (soilgrids.org) at ~~1000-m~~ 1000 m resolution (level-0). The upscaled porosity field ~~was~~ is generated at 5 km for the EDgE project~~(edge. climate.copernicus.eu).~~. The soil characteristics for Noah-MP ~~were~~ are estimated using the same look-up table as in the EURO-CORDEX-Noah-MP case. The comparison of both parameter fields (i.e., EDgE-Noah-MP and EURO-CORDEX-Noah-MP) and the main statistical moments describing the spatial variability of the porosity fields are shown in Figure 5. The results clearly indicate the inappropriateness of the majority-based upscaling operator for this parameter in both cases. It ~~led~~ leads to reduction of the variance of the porosity field and ~~to~~ thus can be considered the least sensitive operator~~at the resolution of the level-0 data~~. This means that the informational content of the hyper-resolution soil maps, commonly available globally, is almost lost.

Notably, although the overall mean of the porosity estimated using MPR over the Pan-EU domain for mHM (Figure 3a) is only 6.6% lower than that calculated using the majority-based approach for Noah-MP (Figure 5a), the spatial patterns obtained by both models are very different. The evidence of this remarkable dissimilarity can also be visualized by comparing the empirical density functions shown in Figures 3d and 5c, both corresponding to a field at $\ell_1 = 5$ km and with the same input data. A detailed evaluation conducted by Samaniego et al. (2012) in Germany showed that large porosity values estimated with the majority-based approach could over-estimate those obtained with MPR by up to 40%, whereas in other locations, under-estimation up to 15% from those estimated by MPR can be found.

Other upscaling operators, such as the weighted arithmetic mean, are commonly used in LSMs in combination with the mosaic approach. For example, in CLM (Oleson et al., 2013, see p. 160) the texture class of the sub-units of the cell, called tiles, are provided in a look-up table. The upscaled porosity field obtained using this approach is shown in Figure 1 ~~b~~ at a 1° (100 km) resolution. Methods based on the majority and weighted arithmetic mean operators exhibit some similarity and lack spatial variability. In both cases, the spatial mean is approximately 0.43 $m^3 m^{-3}$.

Hydrologic models that do not use soil porosity tend to use a similar conceptualization and values denoted as the total available water capacity (TAWC, WaterGAP versions 2 and 3) and field capacity (FC, HBV). For these ~~type~~ types of conceptual models, normalized values of these parameters are used as surrogates for soil porosity. The consistency of the spatial patterns of TAWC and FC are compared here instead of their actual values. A distinctive difference in the patterns can be observed. For example, WaterGAP3 exhibits lower values than WaterGAP2, whereas the pattern of the normalized FC in HBV is the opposite in many locations (e.g., Spain, Germany, Scandinavia).

Details of the parameterization schemes used to estimate TAWC and FC are beyond the scope of this study. Interested readers may refer to Müller Schmied et al. (2014) or Beck et al. (2016), respectively. However, the TAWC in WaterGAP is obtained by linking the soil type provided by the FAO soil map with available water capacity values estimated by Batjes (1996). Thus, no scaling rule or form of regularization is used in this case. The field capacity parameters used in HBV were determined using an ad hoc nearest-neighbor interpolation technique that relies on calibrated parameters from nearby similar donor basins that might exhibit very different geophysical characteristics. The parameter fields obtained for two versions of WaterGAP (30 arcmin, 5 arcmin) and HBV are depicted in Figure 1~~, panels i, j, and k, respectively~~. It can be concluded that the ~~procedure~~ parameterization technique employed is not scale invariant ~~by comparing the~~ as revealed by distinct parameter sets from ~~both~~ WaterGAP model versions, which ~~were~~ are operated at different resolutions. The regionalization proposed by Beck et al. (2016) leads to a ~~patch-quilt~~ patchwork-quilt field that does not resemble to any other field presented. ~~The~~ Evident from the Figure Figure 1, the HBV field lacks seamlessness ~~and becomes conclusive evidence that calibrated model parameters cannot be transferred to other locations or interpolated in space by heuristic algorithms~~that may result in non-seamless fields of water fluxes and states.

4.3 ~~Changes in~~ Effects of a parameterization technique on the dynamics of ~~water fluxes and states~~a state variable

There is a complex interplay between soil moisture (SM) and latent heat (LH) in LSM/HMs. Improving our understanding of soil-land-atmosphere feedback is fundamental ~~to~~ for making reliable predictions of water and energy fluxes~~on land systems~~. In this context, ~~it can be hypothesized that the parameterization~~ we carry out a sensitivity experiment to investigate the effects of soil related ~~parameters~~ parameterizations (e.g., soil porosity) ~~has significant effects on related state variables and fluxes. To falsify this hypothesis, two~~ on latent heat and soil moisture. Two contrasting modeling paradigms (Noah-MP and mHM) ~~were~~ are employed.

The WRF/Noah-MP system is forced with ERA-interim at the boundaries of the rotated CORDEX-Grid (www.meteo.unican.es/wiki/cordexwrf) at a spatial resolution of 0.11° covering Europe from 1989 to 2009. To ease the comparison, the process-based hydrological model mHM (www.ufz.de/mhm) ~~was~~ is driven with daily precipitation and temperature fields generated by the WRF/Noah-MP system during the same period. The spatial resolution of mHM ~~was~~ is fixed at $5{\times}5$ km^2. The main geophysical characteristics in WRF/Noah-MP of land cover and soil texture are represented with a $1{\times}1$ km^2 MODIS and a single-horizon, coarse-resolution FAO soil map with 16 soil texture classes, respectively. The porosity field of Noah-MP is estimated by applying a majority-based operator to values for different soil classes, as shown in Figure ~~1g~~5b.

The settings of the mHM model used in this experiment are described in Section 3.4. In contrast with those of Noah-MP, the global parameters of mHM estimated using the MPR technique are obtained by closing the water balance over selected river basins in Europe (Rakovec et al., 2016a). The porosity fields obtained for mHM over the Pan-EU ~~is~~ are depicted in Figure ~~1f.~~

~~Notably, although the overall mean of the porosity estimated using MPR is only 2.3lower than that calculated using the majority-based approach in Noah-MP (Figure 5), the spatial patterns obtained by both models are very different. The evidence of this remarkable dissimilarity can also be visualized by comparing the empirical density functions shown in Figures 3d and 5c, both corresponding to a field at $\ell_1 = 5$ km and with the same input data~~3.

~~In general, it can be concluded that the porosity field estimated by Noah-MP tends to have lower water holding capacity values than that of the mHM. Deviations of up to -6.8were detected in Germany, where a detailed evaluation was conducted by . Within the same domain, the porosity values in Noah-MP could be up to 15less than those estimated by mHM. However, at very few locations the opposite can happen, reaching values of 40over-estimation.~~

The phase diagrams of the monthly fraction of soil water saturation fSM $= \frac{\theta}{\theta_s}$ (i.e., plots of monthly fSM(t) vs. fSM$(t+1)$) are subsequently ~~estimated to investigate how the~~ investigated to understand the effect of differences in porosity estimates of the top 2 m soil column ~~affect~~ on the soil moisture dynamics (Figure 6). Two locations in Germany are selected in which Noah-MP ~~significantly~~ systematically over- or underestimated the latent heat fluxes with respect to mHM (the latitude and longitude coordinates of the center of the selected Noah-MP grids are A: $(54^\circ N, 10^\circ E)$ and B: $(51^\circ N, 7^\circ E)$, respectively). At location A, the majority-based approach underestimates the MPR soil porosity by -10%, whereas in location B, it overestimates it by 40%. This experiment unambiguously shows that, at locations where Noah-MP over-estimates latent heat with respect to mHM, the ~~dynamic~~ temporal variance (i.e., dynamic) of the monthly SM time series ~~is enhanced~~simulated by Noah-MP is almost doubled compared to that of mHM, leading to much lower soil moisture values (~~extreme droughts) (~~Figure 6a). Conversely, underestimation of latent heat greatly ~~constrains the~~ reduces the variance of the soil moisture dynamics (Figure 6b).

4.4 ~~Flux~~ Effects of not satisfying the flux matching ~~test~~condition

In Section ~~??, it was~~ 2, we postulated that ad hoc parameterization schemes do not necessarily fulfill the flux-matching test performed with a flux simulated by a given model at two modeling resolutions (~~e.g.,~~ $\ell_1 = 5$ ~~,~~ and 30 arcmin). A detailed description of how to perform this test is provided in Samaniego et al. (2010b). The following experiment ~~was~~ is conducted with three models: mHM, PCR-GLOBWB, and WaterGAP ~~at the resolutions above to try to falsify this strong~~ in an attempt to falsify the above postulation. All models use the same forcings and geophysical information. The ~~experiment was~~ simulations are conducted in the Rhine River upstream of the Lobith gauging station. All three models are driven by daily forcing with a spatial resolution of 5 km, which was kindly provided by the EFAS team at JRC (www.eea.europa.eu). Additional details of the modeling settings of this experiment are provided in Sutanudjaja et al. (2015) and www.hyperhydro.org/. The KGE and bias values of these three models obtained for both scales at the Lobith station during 2003 are reported in Table 2. The ~~streamflow~~ daily streamflow time series during this year ~~was~~ is selected for evaluation because it ~~exhibited~~ exhibits strong temporal dynamics, with wet conditions in the beginning of the year followed by a drought during the summer ~~. These efficiency metrics~~ and fall seasons. The performances obtained for the three models are satisfactorily, but the results shown in Table 2 indicate that mHM is the only model that can have higher KGE ~~efficiencies~~ values regardless of the ~~scale~~spatial modelling resolution.

The flux-matching test presented in Section 3.1 ~~was~~ is performed with simulated evapotranspiration (ET) because it is the largest flux in the water cycle besides precipitation, and is prone to the largest predictive uncertainties ~~.~~(Mueller et al., 2013). To ease the comparison, collocated grids are employed for every model such that every coarser scale grid cell has exactly the same number of underlying cells at finer resolution (5 arc min). The results of this test are shown in Figure 7. ~~These results~~ They reveal that mHM ~~is the only model able to fulfill the~~ exhibits the best flux-matching ~~test, as apparently in this figure~~between these two scales. This experiment also shows that the MPR technique implemented in mHM leads to ET fields that ~~satisfy mass conservation across scales and exhibit seamlessness that becomes apparent at higher resolutions (see~~ are of similar magnitude at both scales indicating a close conservation of mass (Figure 7a).

The PCR-GLOBWB and WaterGAP models reveal large inconsistencies in ~~annual ET~~preserving the spatial pattern of annual ET across two modeling scales, although the streamflow performance at the outlet is good (greater than 0.83 in both cases). These results also confirm the postulation that "streamflow-related metrics are a necessary but not sufficient condition to warrant the proper partitioning of incoming precipitation P into various spatially distributed water storage components (e.g., SM) and fluxes (e.g., ET)" (Rakovec et al., 2016b). Because all models are forced with ~~exactly~~ the same forcings, share the same geophysical information, and have almost similar hydrological process descriptions, it can be safely concluded that the parameterization method used in the models ~~should cause~~ caused the ET mismatch. To falsify this postulation, the MPR parameterization protocol proposed in Section 3.3 ~~was~~ is next applied to PCR-GLOBWB.

5 Implementation of the parameterization protocol in PCR-GLOBWB

To evaluate the consistency of land surface fluxes before and after MPR implementation, we ~~performed a sensitivity analysis to study~~ analyze the impact of MPR on evaporative fluxes and soil moisture content in PCR-GLOBWB (van Beek et al., 2011; Wada and Bierkens, 2014; Sutanudjaja et al., 2016) over the Rhine River basin during 2003. The model ~~was~~ is used to simulate the hydrological states at two different spatial resolutions (ℓ_1 =5 ~~,~~and 30 arcmin), and the sensitivity to MPR implementation ~~was~~ is evaluated using a field difference method (in line with eq. 1):

$$\Delta = \sqrt{\frac{1}{T}\sum_{t=1}^{T}\left(100\frac{W_c(t) - w_f(t)}{w_f(t)}\right)^2} \tag{2}$$

where W_c and w_f are the coarse (c) and fine (f) resolution simulations of variable W, respectively, and T is the total time series length.

The original PCR-GLOBWB parameterization does not include consistency in upscaling as enforced by MPR, leading to a larger difference in soil properties. Figure 8 depicts the porosity fields of this model before and after the implementation of MPR. Panels (a) and (b) of this figure show clearly the problems mentioned in section 2, for example lack of coherence in spatial patterns and the existence of spatial discontinuities of parameter fields at two scales. The porosity fields obtained with the MPR technique shown in panels (c) and (d), on the contrary, exhibit a typical seamless spatial structure in which the main features of the field can be distinguished across scales. It is worth noting that differences seen between Figure 8a and Figure 8c are not only due to the improved upscaling procedure, but also due to a modified pedo-transfer function. The parameters of the pedo-transfer function have also been included in the calibration within the MPR approach.

These differences in soil hydraulic properties influence the derived hydrological properties, leading to changes in saturated conductivity and storage capacity in the unsaturated zone. The considerable differences in ET fluxes are shown in panels (~~c~~a) and (~~d~~b) of Figure ~~7~~ 9 and are the result of these changes.

When MPR is employed, we ~~see that the~~ observe that the difference in actual average Rhine basin evapotranspiration between the two scales Δ drops from 29% to 9.4% (Figure9). For the total column soil moisture, we find a stronger decrease in Δ from 25% to 6.9%, clearly indicating the benefits of MPR implementation. We also observe ~~an~~ a slight increase in the discharge ~~KGE performance compared with observations~~ performance (KGE) at Lobith. The original KGEs ~~were~~ are 0.86 (ℓ_1 =5 arcmin) and 0.93 (ℓ_1 =30 arcmin), whereas the KGEs with MPR implementation are 0.91 and 0.93, respectively. Another advantage is that PCR-GLOBWB is calibrated at a coarser resolution, whereas this model ~~was~~ is calibrated for each spatial resolution individually in the original set-up and with lower consistency in the discharge simulation.

From these evaluations, we conclude that MPR implementation leads to significant improvement in the flux matching and discharge simulations across scales, allowing for more consistency across scales for hydrological model simulations. Notably, additional parameters in PCR-GLOBWB still need to be regionalized within the MPR framework, which could potentially lead to better performance and transferability.

6 Conclusions

Hyper-resolution modeling initiatives (Wood et al., 2011; Bierkens et al., 2014) challenge the hydrological community to intensify efforts to make water (quantity and quality) and energy flux predictions “everywhere" and for these predictions to be “locally relevant.” The predictions should have small uncertainties to be useful for the end-users. These grand challenges also imply that the next generation of land surface and hydrologic models must incorporate probabilistic descriptions of the sub-grid variability of geophysical land surface properties — such as POLARIS (Chaney et al., 2016b) and SoilGrids (Hengl et al., 2017) — to cope with the large uncertainties that characterize the related process below the ~~REA~~ Representative Elementary Area (REA) scale. Consequently, great efforts should be made in hyper-resolution monitoring at the global scale, in improving the computational efficiency of LSM/HMs, and in the development of scale-invariant parameterizations for these models. In this study, we have shown that the state-of-the-art ~~is totally inadequate to~~ parameterizations need to be improved to address this grand challenge, especially ~~scale-invariant parameterization~~with respect to better fulfill the flux-matching condition.

We ~~presented and tested~~ revisited a technique called ~~multiscale parameter regionalization~~ Multiscale Parameter Regionalization (MPR) (Samaniego et al., 2010b), ~~currently~~ originally available only in mHM but recently implemented in PCR-GLOBWB ~~and VIC~~ as a part of this study. Moreover, we proposed a *Parameterization Protocol* as a guideline to apply MPR and to retrofit existing LSM/HMs to ~~make them sufficient for the task of addressing these grand challenges.~~ ease the implementation of MPR in the latter. We also discuss the advantages and limitations of MPR which should be considered while applying this concept to other LSM/HMs.

This study has shown that two models that use ad-hoc parameterizations can have reasonable efficiency with respect to simulated streamflow but poor performance with respect to distributed fluxes such as evapotranspiration. The implementation of this protocol in PCR-GLOBWB in this study increased the model efficiency by almost 6% and improved the consistency of simulated ET fields across scales. ~~We have~~ For example, the estimation of evapotranspiration without MPR at 5 arcmin and 30 arcmin spatial resolutions for the Rhine river basin resulted in a difference of approximately 29%. Applying MPR reduced this difference to 9%. For total soil water, the differences without and with MPR are 25% and 7%, respectively. We have also shown that the PCR-GLOBWB global parameters can be transferred across scales with a consistent ET patterns and model efficiency.

In general, it can be concluded that the estimation of global parameters is feasible with MPR and that these scalars are transferable across scales and locations. The successful application of MPR implies that the averaging procedure of geophysical properties matters and that having the right physics with incorrect “effective” parameters leads to ~~incorrect fluxes , states and feedbacks in the soil-vegetation-atmosphere continuum~~inconsistent fluxes and states. Consequently, MPR is a step forward to quasi scale-invariant parameterizations and is feasible to implement in existing LSM/HMs whose goal should be seamless parameter fields across scales that do not exhibits artificial spatial “discontinuities” such as calibration imprints, and that lead to consistent predictions across scales. We consider that this feature is the key for the next generation of LSM and NWP models such as the “model for prediction across scales” (MPAS) (www.mmm.ucar.edu) and the “nested-domain ~~ICOS~~ICON” (www.earthsystemcog.org/projects/dcmip-2012/icon-mpi-dwd). Furthermore, a proper implementation of MPR in process based (conceptual) models may contribute to recent efforts towards identifying their "effective" parameters through observational datasets at the scale of interest (Savenije and Hrachowitz, 2017).

Finally, we would like to reiterate that a flux obtained from a land surface/hydrologic model should always be evaluated with local observations when available and across scales. If "it disagrees with experiment, it's wrong."

Acknowledgements. We kindly acknowledge our data providers: Noah-MP: Kirsten Warrach-Sagi (University of Hohenheim), PCR-GLOBWB: Niko Wanders (Princeton University), WaterGAP: Hannes Mueller-Schmied (University of Frankfurt), JULES: Anne Verhoef (The University of Reading), LISFLOOD: Peter Salamon (JRC), CABLE: Matthias Cuntz (formerly UFZ, now INRA), CLM: David Lawrence (UCAR) and E. Sutanudjaja and M. Bierkens, et al. for providing results from the HyperHydro WG1 Workshop 9-12 June 2015, Utrecht. This study was carried out within the Helmholtz-Association climate initiative REKLIM (www.reklim.de). This work has been partly funded by Helmholtz Alliance EDA - Remote Sensing and Earth System Dynamics, through the Initiative and Networking Fund of the Helmholtz Association, Germany. This study was partially performed under a contract for the Copernicus Climate Change Service (edge.climate.copernicus.eu). ECMWF implements this service and the Copernicus Atmosphere Monitoring Service on behalf of the European Commission. We thank Martin Schrön for kindly contributing to the artwork of Figure 2. Rens van Beek is acknowledged for providing support with the MPR implementation in PCR-GLOBWB.

References

Abdulla, F. and Lettenmaier, D.: Development of regional parameter estimation equations for a macroscale hydrologic model, J. Hydrol., 197, 230–257, 1997.

Addor, N., Rössler, O., Köplin, N., Huss, M., Weingartner, R., and Seibert, J.: Robust changes and sources of uncertainty in the projected hydrological regimes of Swiss catchments, Water Resources Research, 50, 7541–7562, 2014.

Andréassian, V., Bourgin, F., Oudin, L., Mathevet, T., Perrin, C., Lerat, J., Coron, L., and Berthet, L.: Seeking genericity in the selection of parameter sets: Impact on hydrological model efficiency, Water Resources Research, 50, 8356–8366, 2014.

Ball, J. T., Woodrow, I. E., and Berry, J. A.: A Model Predicting Stomatal Conductance and its Contribution to the Control of Photosynthesis under Different Environmental Conditions, in: Progress in Photosynthesis Research, pp. 221–224, Springer Netherlands, Dordrecht, 1987.

Barrios, M. and Francés, F.: Spatial scale effect on the upper soil effective parameters of a distributed hydrological model, Hydrological Processes, 26, 1022–1033, 2011.

Batjes, N. H.: Development of a world data set of soil water retention properties using pedotransfer rules, Geoderma, 71, 31–52, 1996.

Beck, H. E., van Dijk, A. I. J. M., de Roo, A., Miralles, D. G., McVicar, T. R., Schellekens, J., and Bruijnzeel, L. A.: Global-scale regionalization of hydrologic model parameters, Water Resources Research, 52, 3599–3622, 2016.

Beldring, S., Engeland, K., Roald, L. A., Saelthun, N. R., and Vokso, A.: Estimation of parameters in a distributed precipitation-runoff model for Norway, Hydrol. Earth Syst. Sci., 7, 304–316, 2003.

Best, M. J., Pryor, M., Clark, D. B., Rooney, G. G., Essery, R. L. H., Ménard, C. B., Edwards, J. M., Hendry, M. A., Porson, A., Gedney, N., Mercado, L. M., Sitch, S., Blyth, E., Boucher, O., Cox, P. M., Grimmond, C. S. B., and Harding, R. J.: The Joint UK Land Environment Simulator (JULES), model description – Part 1: Energy and water fluxes, Geoscientific Model Development, 4, 677–699, 2011.

Beven, K.: Linking parameters across scales: Subgrid parameterizations and scale dependent hydrological models, Hydrological Processes, 9, 507–525, 1995.

Bierkens, M. F. P.: Global hydrology 2015: State, trends, and directions, Water Resources Research, 51, 4923–4947, 2015.

Bierkens, M. F. P., Bell, V. A., Burek, P., Chaney, N., Condon, L. E., David, C. H., de Roo, A., Döll, P., Drost, N., Famiglietti, J. S., Flörke, M., Gochis, D. J., Houser, P., Hut, R., Keune, J., Kollet, S., Maxwell, R. M., Reager, J. T., Samaniego, L., Sudicky, E., Sutanudjaja, E. H., van de Giesen, N., Winsemius, H., and Wood, E. F.: Hyper-resolution global hydrological modelling: what is next? "Everywhere and locally relevant", Hydrological Processes, pp. n/a–n/a, 2014.

Blöschl, G., Reszler, C., and Komma, J.: A spatially distributed flash flood forecasting model, Environ. Model. Softw., 23, 464–478, 2008.

Blöschl, G., Sivapalan, M., Wagener, T., Viglione, A., and H, S., eds.: Runoff Prediction in Ungauged Basins: Synthesis Across Processes, Places and Scales, Cambridge University Press, ISBN: 978-1107028180, 2013.

Bock, A. R., Hay, L. E., McCabe, G. J., Markstrom, S. L., and Atkinson, R. D.: Parameter regionalization of a monthly water balance model for the conterminous United States, Hydrology and Earth System Sciences, 20, 2861–2876, 2016.

Bonan, G. B., Lawrence, P. J., Oleson, K. W., Levis, S., Jung, M., Reichstein, M., Lawrence, D. M., and Swenson, S. C.: Improving canopy processes in the Community Land Model version 4 (CLM4) using global flux fields empirically inferred from FLUXNET data, Journal of Geophysical Research-Atmospheres, 116, GB1008, 2011.

Brynjarsdottir, J. and O'Hagan, A.: Learning about physical parameters: the importance of model discrepancy, Inverse Problems, 30, 2014.

Burnash, R. J. C., Ferral, R. L., and McGuire, R. A.: A generalized streamflow simulation system: Conceptual modeling for digital computers, U.S. Dept. of Commerce, National Weather Service, 1973.

Chaney, N. W., Metcalfe, P., and Wood, E. F.: HydroBlocks: a field-scale resolving land surface model for application over continental extents, Hydrological Processes, 30, 3543–3559, 2016a.

Chaney, N. W., Wood, E. F., McBratney, A. B., Hempel, J. W., Nauman, T. W., Brungard, C. W., and Odgers, N. P.: POLARIS: A 30-meter probabilistic soil series map of the contiguous United States, Geoderma, 274, 54–67, 2016b.

Christensen, N. S. and Lettenmaier, D. P.: A multimodel ensemble approach to assessment of climate change impacts on the hydrology and water resources of the Colorado River Basin, Hydrology and Earth System Sciences, 11, 1417–1434, 2007.

Clapp, R. B. and Hornberger, G. M.: Empirical equations for some soil hydraulic properties, Water Resources Research, 14, 601–604, 1978.

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

Clark, M. P., Schaefli, B., Schymanski, S. J., Samaniego, L., Luce, C. H., Jackson, B. M., Freer, J. E., Arnold, J. R., Dan Moore, R., Istanbulluoglu, E., and Ceola, S.: Improving the theoretical underpinnings of process-based hydrologic models, Water Resources Research, pp. n/a–n/a, 2016.

Clark, M. P., Bierkens, M. F. P., Samaniego, L., Woods, R. A., Uijenhoet, R., Bennet, K. E., Pauwels, V. R. N., Cai, X., Wood, A. W., and Peters-Lidard, C. D.: The evolution of process-based hydrologic models: Historical challenges and the collective quest for physical realism, Hydrology and Earth System Sciences Discussions, pp. 1–14, 2017.

Cosby, B. J., Hornberger, G. M., Clapp, R. B., and Ginn, T. R.: A Statistical Exploration of the Relationships of Soil Moisture Characteristics to the Physical Properties of Soils, Water Resources Research, 20, 682–690, 1984.

Crawford, N. H. and Linsley, R. K.: Digital simulation in hydrology: Stanford Watershed Model IV, Tech. Rep. 39, Stanford Univ. Dept. of Civil Engineering, 1966.

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 Resources Research, 51, 6417–6441, 2015.

Cuntz, M., Mai, J., Samaniego, L., Clark, M., Wulfmeyer, V., Branch, O., Attinger, S., and Thober, S.: The impact of standard and hard-coded parameters on the hydrologic fluxes in the Noah-MP land surface model, Journal of Geophysical Research-Atmospheres, 121, 10,676–10,700, 2016.

Dagan, G.: Flow and transport in porous media, Springer Verlag, New York, 1989.

De Roo, A. and Wesseling, C. G.: Physically based river basin modelling within a GIS: the LISFLOOD model, Hydrological Processes, 14, 1981–1992, 2000.

Dooge, J.: Parameterization of hydrologic processes, in: Proceedings of the Greenbelt Study Conference, edited by Eagleson, P., p. 243 288, Cambridge University Press, new York, N.Y., 1982.

Duckstein, L. and Opricovic, S.: Multiobjective optimization in river basin development, Water Resources Research, 16, 14–20, 1980.

ECMWF: IFS DOCUMENTATION – Cy41r2 Operational implementation 8 March 2016, Tech. rep., European Centre for Medium-Range Weather Forecasts, http://www.ecmwf.int/search/elibrary/part?solrsort=sort_label%20asc&title=part&secondary_title=41r1&f[0]=ts_biblio_year%3A2016, Access date: 2017/02/02, 2016.

Edijatno, de Oliveira Nascimento, N., Yang, X., Makhlouf, Z., and Michel, C.: GR3J: a daily watershed model with three free parameters, Hydrological Sciences Journal-Journal Des Sciences Hydrologiques, 44, 263–277, 1999.

Edwards, P. N.: A Vast Machine: Computer Models, Climate Data, and the Politics of Global Warming, The MIP Press, 2010.

Famiglietti, J. and Wood, E.: Multiscale modeling of spatially variable water and energy balance processes, Water Resources Research, 30, 3061–3078, 1994.

Famiglietti, J. S. and Wood, E. F.: Effects of Spatial Variability and Scale on Areally Averaged Evapotranspiration, Water Resources Research, 31, 699–712, 1995.

Fenicia, F., Kavetski, D., and Savenije, H. H. G.: Elements of a flexible approach for conceptual hydrological modeling: 1. Motivation and theoretical development, Water Resources Research, 47, n/a–n/a, 2011.

Fernandez, W., Vogel, R., and Sankarasubramanian, A.: Regional calibration of a watershed model, Hydrolog. Sci. J., 45, 689–707, 2000.

Fisher, J. B., Huntzinger, D. N., Schwalm, C. R., and Sitch, S.: Modeling the Terrestrial Biosphere, dx.doi.org, 39, 91–123, 2014.

Flügel, W. A.: Delineating hydrological response units by geographical information system analyses for regional hydrological modelling using PRMS/MMS in the drainage basin of the river Bröl, Germany, Hydrol. Process., 9, 423–436, 1995.

Gelhar, L. W.: Stochastic Subsurface Hydrology, Prentice Hall, 1993.

Goehler, M., Mai, J., and Cuntz, M.: Use of eigendecomposition in a parameter sensitivity analysis of the Community Land Model, Journal of Geophysical Research-Biogeosciences, 118, 904–921, 2013.

Gotzinger, J. and Bárdossy, A.: Comparison of four regionalisation methods for a distributed hydrological model, J. Hydrol., 333, 374–384, 2007.

Gupta, H. V., Perrin, C., Blöschl, G., Montanari, A., Kumar, R., Clark, M., and Andréassian, V.: Large-sample hydrology: a need to balance depth with breadth, Hydrology and Earth System Sciences, 18, 463–477, 2014.

Haddeland, I., Clark, D. B., Franssen, W., Ludwig, F., Voss, F., Arnell, N. W., Bertrand, N., Best, M., Folwell, S., Gerten, D., Gomes, S., Gosling, S. N., Hagemann, S., Hanasaki, N., Harding, R., Heinke, J., Kabat, P., Koirala, S., Oki, T., Polcher, J., Stacke, T., Viterbo, P., Weedon, G. P., and Yeh, P.: Multimodel Estimate of the Global Terrestrial Water Balance: Setup and First Results, Journal of Hydrometeorology, 12, 869–884, 2011.

Haughton, N., Abramowitz, G., Pitman, A. J., Or, D., Best, M. J., Johnson, H. R., Balsamo, G., Boone, A., Cuntz, M., Decharme, B., Dirmeyer, P. A., Dong, J., Ek, M., Guo, Z., Haverd, V., van den Hurk, B. J. J., Nearing, G. S., Pak, B., Santanello Jr., J. A., Stevens, L. E., and Vuichard, N.: The Plumbing of Land Surface Models: Is Poor Performance a Result of Methodology or Data Quality?, Journal of Hydrometeorology, 17, 1705–1723, 2016.

Hengl, T., Mendes de Jesus, J., Heuvelink, G. B. M., Ruiperez Gonzalez, M., Kilibarda, M., Blagotić, A., Shangguan, W., Wright, M. N., Geng, X., Bauer-Marschallinger, B., Guevara, M. A., Vargas, R., MacMillan, R. A., Batjes, N. H., Leenaars, J. G. B., Ribeiro, E., Wheeler, I., Mantel, S., and Kempen, B.: SoilGrids250m: Global gridded soil information based on machine learning, PLOS ONE, 12, 1–40, doi:10.1371/journal.pone.0169748, 2017.

Hundecha, Y. and Bárdossy, A.: Modeling of the effect of land use changes on the runoff generation of a river basin through parameter regionalization of a watershed model, Journal of Hydrology, 292, 281–295, 2004.

Hundecha, Y., Arheimer, B., Donnelly, C., and Pechlivanidis, I.: A regional parameter estimation scheme for a pan-European multi-basin model, Journal of Hydrology: Regional Studies, 6, 90–111, 2016.

Intsiful, J. and Kunstmann, H.: Upscaling of Land-Surface Parameters Through Inverse Stochastic SVAT-Modelling, Boundary-Layer Meteorology, 129, 137–158, 2008.

Kavetski, D., Kuczera, G., and Franks, S. W.: Semidistributed hydrological modeling: A “saturation path” perspective on TOPMODEL and VIC, Water Resources Research, 39, n/a–n/a, 2003.

Kitanidis, P. K. and Vomvoris, E. G.: A geostatistical approach to the inverse problem in groundwater modeling (steady state) and one-dimensional simulations, Water Resources Research, 19, 677–690, 2010.

Koren, V., Smith, M., and Duan, Q.: Use of a Priori Parameter Estimates in the Derivation of Spatially Consistent Parameter Sets of Rainfall-Runoff Models, pp. 239–254, American Geophysical Union, doi:10.1002/9781118665671.ch18, http://dx.doi.org/10.1002/9781118665671.ch18, 2013.

Kowalczyk, E. A., Wang, Y. P., and Law, R. M.: The CSIRO Atmosphere Biosphere Land Exchange (CABLE) model for use in climate models and as an offline model, CSIRO. Marine and Atmospheric Research, 13, 2006.

Kumar, R., Samaniego, L., and Attinger, S.: The effects of spatial discretization and model parameterization on the prediction of extreme runoff characteristics, Journal of Hydrology, 392, 54–69, 2010.

Kumar, R., Livneh, B., and Samaniego, L.: Toward computationally efficient large-scale hydrologic predictions with a multiscale regionalization scheme, Water Resources Research, 49, 5700–5714, 2013a.

Kumar, R., Samaniego, L., and Attinger, S.: Implications of distributed hydrologic model parameterization on water fluxes at multiple scales and locations, Water Resources Research, 49, 360–379, 2013b.

Kumar, R., Mai, J., Rakovec, O., Cuntz, M., Thober, S., Zink, M., Attinger, S., Schaefer, D., Schrön, M., and Samaniego, L. E.: Regionalized Hydrologic Parameters Estimates for a Seamless Prediction of Continental scale Water Fluxes and States, AGU Fall Meeting Abstracts, 2015.

Le Treut, H., Somerville, R., Cubasch, U., Ding, Y., Mauritzen, C., Mokssit, A., Peterson, T., and Prather, M.: Historical Overview of Climate Change, in: Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change, edited by Solomon, S., Qin, D., Manning, M., Chen, Z., Marquis, M., Averyt, K., Tignor, M., and H.L., M., chap. 1, pp. 1–36, Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA., 2007.

Leavesley, G. H., Lichty, R. W., Troutman, B. M., and Saindon, L. G.: Precipitation-Runoff Modeling System: User's Manual, U.S. Geological Survey Water-Resources Investigations, Denver, Colorado, 83-4238 edn., 1983.

Lerat, J., Andréassian, V., Perrin, C., Vaze, J., Perraud, J.-M., Ribstein, P., and Loumagne, C.: Do internal flow measurements improve the calibration of rainfall-runoff models?, Water Resources Research, 48, 2012.

Li, D., Bou-Zeid, E., Barlage, M., Chen, F., and Smith, J. A.: Development and evaluation of a mosaic approach in the WRF-Noah framework, Journal of Geophysical Research-Atmospheres, 118, 11,918–11,935, 2013.

Li, H., Sivapalan, M., and Tian, F.: Comparative diagnostic analysis of runoff generation processes in Oklahoma DMIP2 basins: The Blue River and the Illinois River, Journal of Hydrology, 418-419, 90–109, 2012.

Liang, X., Lettenmaier, D., Wood, E., and Burges, S.: A Simple Hydrologically Based Model of Land-Surface Water and Energy Fluxes for General-Circulation Models, Journal of Geophysical Research-Atmospheres, 99, 14 415–14 428, 1994.

Liang, X., Lettenmaier, D. P., and Wood, E. F.: One-dimensional statistical dynamic representation of subgrid spatial variability of precipitation in the two-layer variable infiltration capacity model, Journal of Geophysical Research-Atmospheres, 101, 21 403–21 422, 1996.

Lindstrom, G., Johansson, B., Persson, M., Gardelin, M., and Bergström, S.: Development and test of the distributed HBV-96 hydrological model, Journal of Hydrology, 201, 272–288, 1997.

Lindström, G., Pers, C., Rosberg, J., Strömqvist, J., and Arheimer, B.: Development and testing of the HYPE (Hydrological Predictions for the Environment) water quality model for different spatial scales, Hydrology research, 41, 295–26, 2010.

Livneh, B. and Lettenmaier, D. P.: Regional parameter estimation for the unified land model, Water Resources Research, 49, 100–114, 2013.

Livneh, B., Kumar, R., and Samaniego, L.: Influence of soil textural properties on hydrologic fluxes in the Mississippi river basin, Hydrological Processes, 29, 4638–4655, 2015.

Martina, M. L. V., Todini, E., and Liu, Z.: Preserving the dominant physical processes in a lumped hydrological model, Journal of Hydrology, 399, 121–131, 2011.

Mendoza, P. A., Clark, M. P., Barlage, M., Rajagopalan, B., Samaniego, L., Abramowitz, G., and Gupta, H.: Are we unnecessarily constraining the agility of complex process-based models?, Water Resources Research, 2015.

Merz, R. and Blöschl, G.: Regionalisation of catchment model parameters, Journal of Hydrology, 287, 95–123, 2004.

Miller, E. E. and Miller, R. D.: Physical Theory for Capillary Flow Phenomena, Journal of Applied Physics, 27, 324–332, 1956.

Mizukami, N., Clark, M., Newman, A., Wood, A., Gutmann, E., Nijssen, B., Samaniego, L., and Rakovec, O.: Towards seamless large domain parameter estimation for hydrologic models, Water Resources Research, submitted., 2017.

Morris, M. D.: Factorial Sampling Plans for Preliminary Computational Experiments, Technometrics, 33, 161–174, 1991.

Mueller, B., Hirschi, M., Jimenez, C., Ciais, P., Dirmeyer, P. A., Dolman, A. J., Fisher, J. B., Jung, M., Ludwig, F., Maignan, F., Miralles, D. G., McCabe, M. F., Reichstein, M., Sheffield, J., Wang, K., Wood, E. F., Zhang, Y., and Seneviratne, S. I.: Benchmark products for land evapotranspiration: LandFlux-EVAL multi-data set synthesis, Hydrology and Earth System Sciences, 17, 3707–3720, 2013.

Müller Schmied, H., Eisner, S., Franz, D., Wattenbach, M., Portmann, F. T., Flörke, M., and Doll, P.: Sensitivity of simulated global-scale freshwater fluxes and storages to input data, hydrological model structure, human water use and calibration, Hydrology and Earth System Sciences, 18, 3511–3538, 2014.

Nearing, G. S., Tian, Y., Gupta, H. V., Clark, M. P., Harrison, K. W., and Weijs, S. V.: A philosophical basis for hydrological uncertainty, Hydrological Sciences Journal-Journal Des Sciences Hydrologiques, 61, 1666–1678, 2016.

Neuman, S. P.: Universal scaling of hydraulic conductivities and dispersivities in geologic media, Water Resources Research, 26, 1749–1758, 2010.

Nijzink, R. C., Samaniego, L., Mai, J., Kumar, R., Thober, S., Zink, M., Schäfer, D., Savenije, H. H. G., and Hrachowitz, M.: The importance of topography-controlled sub-grid process heterogeneity and semi-quantitative prior constraints in distributed hydrological models, Hydrology and Earth System Sciences, 20, 1151–1176, 2016.

Niu, G.-Y.: THE COMMUNITY NOAH LAND-SURFACE MODEL (LSM) WITH MULTI-PHYSICS OPTIONS, Tech. rep., National Centers for Environmental Prediction (NCEP), Oregon State University, Air Force, and Hydrology Lab - NWS, https://www.jsg.utexas.edu/noah-mp/users-guide/, Access date: 2017/02/02, 2011.

Niu, G.-Y., Yang, Z.-L., Mitchell, K. E., Chen, F., Ek, M. B., Barlage, M., Kumar, A., Manning, K., Niyogi, D., Rosero, E., Tewari, M., and Xia, Y.: The community Noah land surface model with multiparameterization options (Noah-MP): 1. Model description and evaluation with local-scale measurements, Journal of Geophysical Research, 116, 1381, 2011.

Oleson, K., Lawrence, D., Bonan, G., Drewniak, B., Huang, M., Koven, C., Levis, S., Li, F., Riley, W., Subin, Z., Swenson, S., Thornton, P., Bozbiyik, A., Fisher, R., Kluzek, E., Lamarque, J.-F., Lawrence, P., Leung, L., Lipscomb, W., Muszala, S., Ricciuto, D., Sacks, W., Sun, Y., Tang, J., and Yang, Z.-L.: Technical Description of version 4.5 of the Community Land Model (CLM), Tech. rep., Ncar Technical Note NCAR/TN-503+STR, National Center for Atmospheric Research, Boulder, CO, http://www.cesm.ucar.edu/models/cesm1.2/clm/, Access date: 2017/02/02, 2013.

Peters-Lidard, C. D., Clark, M., Samaniego, L., Verhoest, N. E. C., van Emmerik, T., Uijlenhoet, R., Achieng, K., Franz, T. E., and Woods, R.: Scaling, Similarity, and the Fourth Paradigm for Hydrology, Hydrology and Earth System Sciences Discussions, pp. 1–21, 2017.

Pielke Sr, R.: Mesoscale meteorological modeling, Academic Press, Elsevier, International Geophysics, 3 Rev ed, 2013.

Pokhrel, P. and Gupta, H. V.: On the use of spatial regularization strategies to improve calibration of distributed watershed models, Water Resources Research, 46, 2010.

Rakovec, O., Hill, M. C., Clark, M. P., Weerts, A. H., Teuling, A. J., and Uijlenhoet, R.: Distributed Evaluation of Local Sensitivity Analysis (DELSA), with application to hydrologic models, Water Resour. Res., 50, 1–18, doi:10.1002/2013WR014063, 2014.

Rakovec, O., Kumar, R., Attinger, S., and Samaniego, L.: Improving the realism of hydrologic model functioning through multivariate parameter estimation, Water Resources Research, 52, 7779–7792, 2016a.

Rakovec, O., Kumar, R., Mai, J., Cuntz, M., Thober, S., Zink, M., Attinger, S., Schäfer, D., Schrön, M., and Samaniego, L.: Multiscale and Multivariate Evaluation of Water Fluxes and States over European River Basins, Journal of Hydrometeorology, 17, 287–307, 2016b.

Reggiani, P., Sivapalan, M., and Majid Hassanizadeh, S.: A unifying framework for watershed thermodynamics: balance equations for mass, momentum, energy and entropy, and the second law of thermodynamics, Advances in Water Resources, 22, 367–398, 1998.

Samaniego, L. and Bárdossy, A.: Robust parametric models of runoff characteristics at the mesoscale, Journal of Hydrology, 303, 136–151, 2005.

Samaniego, L., Bárdossy, A., and Kumar, R.: Streamflow prediction in ungauged catchments using copula-based dissimilarity measures, Water Resources Research, 46, n/a–n/a, 2010a.

Samaniego, L., Kumar, R., and Attinger, S.: Multiscale parameter regionalization of a grid-based hydrologic model at the mesoscale, Water Resources Research, 46, n/a–n/a, 2010b.

Samaniego, L., Kumar, R., and Jackisch, C.: Predictions in a data-sparse region using a regionalized grid-based hydrologic model driven by remotely sensed data, Hydrology research, 42, 338–355, 2011.

Samaniego, L. E., Warrach-Sagi, K., Zink, M., and Wulfmeyer, V.: Verification of High Resolution Soil Moisture and Latent Heat in Germany, AGU Fall Meeting Abstracts, http://adsabs.harvard.edu/abs/2012AGUFM.H23G..02S, provided by the SAO/NASA Astrophysics Data System, 2012.

Savenije, H. H. G. and Hrachowitz, M.: HESS Opinions "Catchments as meta-organisms — a new blueprint for hydrological modelling", Hydrology and Earth System Sciences, 21, 1107–1116, 2017.

Seibert, J.: Regionalisation of parameters for a conceptual rainfall-runoff model, Agric. Fores. Meteorol., 98-99, 279 – 293, 1999.

Sellers, P. J., Dickinson, R. E., Randall, D. A., Betts, A. K., Hall, F. G., Berry, J. A., Collatz, G. J., Denning, A. S., Mooney, H. A., Nobre, C. A., Sato, N., Field, C. B., and Henderson-Sellers, A.: Modeling the exchanges of energy, water, and carbon between continents and the atmosphere, Science, 275, 502–509, 1997.

Singh, R., Archfield, S. A., and Wagener, T.: Identifying dominant controls on hydrologic parameter transfer from gauged to ungauged catchments – A comparative hydrology approach, Journal of Hydrology, 517, 985–996, 2014.

Singh, S. K., Bárdossy, A., Götzinger, J., and Sudheer, K. P.: Effect of spatial resolution on regionalization of hydrological model parameters, Hydrological Processes, 26, 3499–3509, 2012.

Sobol', I. M.: Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates, Math. Comput. Simulation, 55, 271–280, doi:10.1016/S0378-4754(00)00270-6, 2001.

Sutanudjaja, E., Bosmans, J., Chaney, N., Clark, M. P., Condon, L. E., David, C. H., De Roo, A. P. J., Doll, P. M., Drost, N., Eisner, S., Famiglietti, J. S., Floerke, M., Gilbert, J. M., Gochis, D. J., Hut, R., Keune, J., Kollet, S. J., Maxwell, R. M., Pan, M., Rakovec, O., Reager, II, J. T., Samaniego, L. E., Mueller Schmied, H., Trautmann, T., Van Beek, L. P., Van De Giesen, N., Wood, E. F., Bierkens, M. F., and Kumar, R.: The HyperHydro (H^2) experiment for comparing different large-scale models at various resolutions, AGU Fall Meeting Abstracts, http://adsabs.harvard.edu/abs/2015AGUFM.H23E1622S, 2015.

Sutanudjaja, E., van Beek, R., Wada, Y., Bosmans, J., Drost, N., de Graaf, I., de Jong, K., Lopez Lopez, P., Pessenteiner, S., Schmitz, O., Straatsma, M., Wanders, N., Wisser, D., and Bierkens, M.: PCR-GLOBWB_model: PCR-GLOBWB version v2.1.0_alpha, doi:10.5281/zenodo.60764, https://doi.org/10.5281/zenodo.60764, Note that this is still a 'pre-release' version. Please test it and, if you find any, please report any bugs and/or issues., 2016.

Troy, T. J., Wood, E. F., and Sheffield, J.: An efficient calibration method for continental-scale land surface modeling, Water Resources Research, 44, 2008.

van Beek, L. P. H., Wada, Y., and Bierkens, M. F. P.: Global monthly water stress: 1. Water balance and water availability, Water Resources Research, 47, n/a–n/a, 2011.

Viterbo, P. and Beljaars, C. M.: An improved land surface parameterization scheme in the ECMWF model and its validation, Journal of Climate, 8, 2716–2748, 1995.

Viviroli, D., Zappa, M., Gurtz, J., and Weingartner, R.: An introduction to the hydrological modelling system PREVAH and its pre- and post-processing-tools, Environ. Model. Softw., 24, 1209–1222, 2009.

Wada, Y. and Bierkens, M. F. P.: Sustainability of global water use: past reconstruction and future projections, Environmental Research Letters, 9, 104 003–15, 2014.

Wagener, T. and Wheater, H. S.: Parameter estimation and regionalization for continuous rainfall-runoff models including uncertainty, Journal of Hydrology, 320, 132–154, 2006.

Wanders, N. and Wada, Y.: Human and climate impacts on the 21st century hydrological drought, Journal of Hydrology, 526, 208–220, 2015.

Wöhling, T., Samaniego, L., and Kumar, R.: Evaluating multiple performance criteria to calibrate the distributed hydrological model of the upper Neckar catchment, Environmental Earth Sciences, 69, 453–468, 2013.

Wood, A. and Mizukami, N.: Project Summary Report: CMIP5 1/8 Degree Daily Weather and VIC Hydrology Datasets for CONUS, Tech. rep., B. o. R. U.S. Department of the Interior, Technical Services Center, Denver, Colorado, http://www.corpsclimate.us/docs/cmip5.hydrology.2014.final.report.pdf, access date: 2017/01/24, 2014.

Wood, E., ed.: Land Surface, atmosphere interactions for climate modelling: observations. models, and analysis, Kluwer, 1990.

Wood, E.: Effects of soil moisture aggregation on surface evaporative fluxes, in: Journal of Hydrology, pp. 397–412, 1997.

Wood, E. F., Sivapalan, M., Beven, K., and Band, L.: Effects of Spatial Variability and Scale with Implications to Hydrologic Modeling, Journal of Hydrology, 102, 29–47, 1988.

Wood, E. F., Lettenmaier, D. P., Liang, X., Lohmann, D., Boone, A., Chang, S., Chen, F., Dai, Y., Dickinson, R. E., Duan, Q., Ek, M., Gusev, Y. M., Habets, F., Irannejad, P., Koster, R., Mitchel, K. E., Nasonova, O. N., Noilhan, J., Schaake, J., Schlosser, A., Shao, Y., Shmakin, A. B., Verseghy, D., Warrach, K., Wetzel, P., Xue, Y., Yang, Z.-L., and Zeng, Q. C.: The project for intercomparison of land-surface parameterization schemes (PILPS) phase 2(c) Red-Arkansas River basin experiment: 1. Experiment description and summary intercomparisons, Global and Planetary Change, 19, 115–135, 1998.

Wood, E. F., Roundy, J. K., Troy, T. J., van Beek, L. P. H., Bierkens, M. F. P., Blyth, E., de Roo, A., Doell, P., Ek, M., Famiglietti, J., Gochis, D., van de Giesen, N., Houser, P., Jaffé, P. R., Kollet, S., Lehner, B., Lettenmaier, D. P., Peters-Lidard, C., Sivapalan, M., Sheffield, J., Wade, A., and Whitehead, P.: Hyperresolution global land surface modeling: Meeting a grand challenge for monitoring Earth's terrestrial water, Water Resources Research, 47, 2011.

Wösten, J. H. M., Pachepsky, Y. A., and Rawls, W. J.: Pedotransfer functions: bridging the gap between available basic soil data and missing soil hydraulic characteristics, Journal of Hydrology, 251, 123–150, 2001.

Yadav, M., Wagener, T., and Gupta, H.: Regionalization of constraints on expected watershed response behavior for improved predictions in ungauged basins, Advances in Water Resources, 30, 1756–1774, 2007.

Zehe, E., Ehret, U., Pfister, L., Blume, T., Schroeder, B., Westhoff, M., Jackisch, C., Schymanski, S. J., Weiler, M., Schulz, K., Allroggen, N., Tronicke, J., van Schaik, L., Dietrich, P., Scherer, U., Eccard, J., Wulfmeyer, V., and Kleidon, A.: HESS Opinions: From response units to functional units: a thermodynamic reinterpretation of the HRU concept to link spatial organization and functioning of intermediate scale catchments, Hydrology and Earth System Sciences, 18, 4635–4655, 2014.

Table 1. Data sources and parameterization method used by models used in this study

Model	Parameterization Method	References	Source code & Projects
CABLE	Pedo-transfer functions, look-up table, dominant soil type	Kowalczyk et al. (2006)	www.cawcr.gov.au/publications/technicalreports/CTR_057.pdf
CLM	Pedo-transfer functions, look-up table, mosaic approach	Oleson et al. (2013)	www.cesm.ucar.edu/models/cesm1.2/clm/
CHTESSEL	Look-up table, dominant soil type	Viterbo and Beljaars (1995); ECMWF (2016)	www.ecmwf.int/search/elibrary
HBV	k-NN interpolation, calibrated parameter	Beck et al. (2016)	www.gloh2o.org/hbv-simreg/
JULES	Look-up table, dominant soil type	Best et al. (2011)	jules.jchmr.org
LISFLOOD	Pedo-transfer functions, mosaic approach, arithmetic mean	De Roo and Wesseling (2000)	ec.europa.eu/jrc/en/publication/eur-scientific-and-technical-research-reports/lisflood-distributed-water-balance-and-/flood-simulation-model-revised-user-manual-2013
mHM	MPR	Samaniego et al. (2010b)	edge.climate.copernicus.eu www.ufz.de/mhm
Noah-MP	Look-up table, dominant soil type	Niu (2011)	www.jsg.utexas.edu/noah-mp www.meteo.unican.es/wiki/cordexwrf
PCR-GLOBWB	(Original) pedo-transfer functions with averaged predictors (New) MPR	van Beek et al. (2011); Wada and Bierkens (2014) Samaniego et al. (2010b)	pcraster.geo.uu.nl/projects/applications/pcrglobwb/
WaterGAP (2,3)	Look-up tables	Müller Schmied et al. (2014); Batjes (1996)	www.uni-kassel.de/einrichtungen/en/cesr/research/projects/active/watergap.html www.uni-frankfurt.de/45218063/WaterGAP

Table 2. Efficiency of mHM, PCR-GLOBWB and WaterGAP obtained for the Rhine basin at Lobith station during 2003 for spatial resolutions of 5 and 30 arcmin.

Model	5 arcmin		30 arcmin	
	KGE	Bias [m^3s^{-1}]	KGE	Bias [m^3s^{-1}]
mHM	0.96	61.19	0.96	21.74
PCR-GLOBWB	0.93	-20.61	0.86	248.09
WaterGAP (3,2)	0.83	143.02	0.90	-41.99