Toward seamless hydrologic predictions across scales

Luis Samaniego[1], Rohini Kumar[1], Stephan Thober[1], Oldrich Rakovec[1], Matthias Zink[1], Niko Wanders[2], 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 1-10 km in catchment-scale applications to over 50 km in global-scale applications. Application of the same model structure at different spatial scales requires that the model estimates similar fluxes independent of the model resolution and 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

5   parameter fields for land surface geophysical properties. Multiple experiments with the mHM, Noah-MP, PCR-GLOBWB and WaterGAP models are conducted to demonstrate the pitfalls of poor 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 review of existing parameter regionalization techniques and discuss a method for obtaining seamless hydrological predictions of water fluxes and

10   states across multiple spatial resolutions. The multiscale parameter regionalization (MPR) technique is a practical and robust method that provides consistent (seamless) parameter and flux fields across scales. A general model protocol is presented to describe how MPR can be applied to a specific model, with an example of this application using the PCR-GLOBWB model. Applying MPR to PCR-GLOBWB substantially improves the flux-matching condition. Estimation of evapotranspiration without MPR at 5 arcmin and 30 arcmin spatial resolutions for the Rhine river basin results in a difference of approximately 29%.

15   Applying MPR reduce this difference to 9%. For total soil water, the differences without and with MPR are 25% and 7%, respectively.

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 Popper (1935), modeling is an interactive research process that starts with the observation of a natural system (e.g., 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 (Popper, 1935; Clark et al., 2016). Model predictions should be confronted with new data to assess the ability of the model to reproduce them.

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 (Richardson, 1922). A similar framework was employed by Freeze and Harlan (1969) 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. With the advent of electronic computers in 1945, the performance of general circulation models (GCM), numerical weather prediction (NWP) models, land surface models (LSMs), and hydrologic models (HMs) has been increased by improving model conceptualization (i.e., the number of process descriptions) and/or spatial resolution if 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 (Le Treut et al., 2007). 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 (Wood et al., 2011; Bierkens et al., 2014).

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" (Stensrud, 2007). 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 (Pokhrel and Gupta, 2010) 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 (Brynjarsdottir and O'Hagan, 2014).

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. 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 Dooge (1982, p.269) 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.

The reasons for the lack of progress in creating scale-invariant parameterizations are manifold. The most important is likely the lack of procedures and theories for linking physical properties (e.g., soil porosity) that can be measured at the field scale with "effective" values that represent the aggregate behavior of the land characteristics at the scale of a grid cell required in LSMs or HMs. Richardson (1922, p.9) 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, Crawford and Linsley (1966) suggested that parameters should be selected so that simulations could be extended to ungauged areas, and Freeze and Harlan (1969) 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, Dooge (1982, p.30) 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 km^2) is an unresolved problem." There have been many attempts to bridge this gap, but the results have not been very successful (Wood et al., 1998).

Another reason for this lack of progress is related to the approach to understanding the scaling problem mentioned by Dooge (1982) and its influence on the hydrologic response of the entire system. As stated by Wood (1997): "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 (Wood et al., 1988), but others concluded that this concept has limited utility in hydrology (Fan and Bras, 1995) and that there are many shortcomings related to its applicability (Blöschl et al., 1995; Seyfried and Wilcox, 1995; Woods et al., 1995). There is no general agreed upon theory that resolve this issue. The inclusion of sub grid-scale and surface heterogeneity in hydrological parameterizations and multi-scale modeling of hydrologically relevant variables have also been attempted (Famiglietti and Wood, 1995, 1994; Liang et al., 1996). 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 Miller and Miller (1956) and followed by seminal works on fractal approaches (Rieu and Sposito, 1991; Tyler and Wheatcraft, 1992; Pachepsky et al., 2000), stochastic perturbation methods (Yeh et al., 1985; Russo et al., 1997), stream tube approaches (Zhu and Mohanty, 2002; Or and Rubin, 2010), and connectivity-based methods (Vogel and Roth, 1998; Neuweiler and Cirpka, 2005). 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 (Beven, 1995). They have only been used to

upscale evaporation and transpiration fluxes (e.g., Kozak et al., 2005) at the field scale and have not been used at larger scales (Vereecken et al., 2007) until very recently. Montzka et al. (2017) applied the scaling proposed by Miller and Miller (1956) 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 (Kitanidis and

5   Vomvoris, 2010). Stochastic and geostatistical theories (Dagan, 1989; Gelhar, 1993; Neuman, 2010) 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 (Attinger, 2003) 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, developed for sub-surface hydrological

10   problems, have not been applied or incorporated into regional-scale LSM/HMs.

A third reason is related to the lack of transparency in most of the existing LSM/HM source code of with respect to the meaning, origin and uncertainty associated with the hard-coded numerical values (i.e., parameters) 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 (Mendoza

15   et al., 2015). 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 (Oleson et al., 2013, see p. 160). Cuntz et al. (2016) 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,

20   but the physical constants and parameters cannot be because they must be treated as random variables (Nearing et al., 2016).

In this study, we provide a short overview of the challenges and limitations of existing HM/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.

25   We argue that the multiscale parameter parameterization (MPR) technique (Samaniego et al., 2010b) offers a method to address the challenges of linking the field scale (observations) with the catchment scale (Dooge, 1982) and to account for the effect of the spatial variability of geophysical characteristics in the parameterization of hydrologic processes that operate at a range of spatial resolutions (Dooge, 1982; Wood et al., 1988). Because MPR relies on empirical transfer functions to link geophysical properties with model parameters, it provides a very effective procedure to transfer "global parameters"

30   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 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 Wood (1990):

"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 parameterizations and observational data for estimating model parameters?"

2 Parameterization of hydrologic and land surface models

2.1 Representing spatial heterogeneity in HMs/LSMs: a brief review

The core of the Freeze and Harlan (1969) blueprint is embracing the spatial distribution of geophysical land-surface attributes in an HM/LSM. A large part of the literature in hydrology (Blöschl et al., 2013) is devoted to streamflow prediction, which, according to Jakeman and Hornberger (1993), is a hydrological variable that exhibits low dimensionality and represents the integral signal of a basin. As a result, most of the "precipitation-runoff" models were conceived as lumped 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) or as semi-distributed 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) and thus lead to poor spatial representation of geophysical parameters, state variables and fluxes.

Lumped 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 (Brynjarsdottir and O'Hagan, 2014), there is a trade-off: the performance at interior points of the basin or at other locations becomes worse than that for calibrated outlets (Pokhrel and Gupta, 2010; Lerat et al., 2012). A second drawback of this type of model is that the parameter fields, which are obtained by combining lumped model parameters from sub-basins, are unrealistic because they exhibit sharp discontinuities along the sub-basin boundaries (Li et al., 2012; Blöschl et al., 2013; Merz and Blöschl, 2004). In addition, these "patch quilt" parameter fields exhibit significant sensitivity to the calibration conditions (Merz and Blöschl, 2004). Thus, these models 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 basin-wise "calibrated" lumped models lack spatial seamlessness, and thus are "inadequate representations of real-world systems" (Savenije and Hrachowitz, 2017).

There have been many attempts to improve the realism and performance of lumped and semi-distributed models by further discretizing the sub-basins into a given number of regions that exhibit nearly similar hydrologic responses, 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 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 parameter fields (Gotzinger and Bárdossy, 2007; Singh et al., 2012), but with somewhat limited success in the transferability of param-

eters across scales and locations. In addition, models parameterized using the HRU concept do not lead to mass conservation of water fluxes (i.e., flux-matching) when applied to scales other than those used in 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

5 at coarse resolution. 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 Zehe et al. (2014), 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 the "inap-

10 propriate 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 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 for the

15 proliferation of hidden parameters in LSM/HM code (Mendoza et al., 2015; Cuntz et al., 2016).

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 model

20 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.

Recently, a dissimilarity-based regionalization technique was used by Beck et al. (2016) to generate an ensemble of global

25 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 will likely lead to discontinuous parameter fields because the calibration regions are fixed to one hundred. The authors

30 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.

Many attempts have been made in the land surface modeling community to address Dooge's challenges, especially with

35 respect to the transferability of model parameters across locations and scales. 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

5   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 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 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 (Wood

10 and Mizukami, 2014; Mizukami et al., 2017).

Recent community-driven efforts, such as the Protocol for the Analysis of Land Surface Models (PALS) 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 predictability of many LSMs evaluated with empirical benchmarks in the PLUMBER project (e.g., CABLE, CHTESSEL, JULES, Noah) may be the result of poor parameterizations, among other

15 factors.

2.2   The state-of-the-art of LSM/HM parameterization

Recently, the challenges that we face in estimating key physical parameters in LSM/HMs has been intensively discussed by many authors (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 visualize the problems discussed above, we selected a representative sample of LSM/HMs used

20 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 is one of the most common parameters in many LSM/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

25 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 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

30 comparison with the porosity fields. Soil porosity is expressed in $m^3 m^{-3}$ and can be compared among different models.

The following lessons can be learned from Figure 1: (1) There is a serious deficiency 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 parameter is rather simple: it represents the ratio of the volume of voids to the total volume in the soil column. We can now wonder how large the uncertainty of

other complex parameters whose relationship with soil properties is very nonlinear would be (e.g., hydraulic conductivity). (2) The degree of seamlessness strongly depends on the level of aggregation and the upscaling of underlying soil texture fields. (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, corresponding to the

5 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 an unsolved problem.

These facts allow us to postulate the following questions: (1) Why there are such large differences between models in estimating a physically interpretable parameter? (2) What are the consequences of model calibration on parameter fields? (3) Are current model parameterizations scale invariant? (4) Do the fluxes estimated with these models at various scales satisfy

10 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 framework

3.1 The flux-matching postulation

The key postulation aiming at obtaining scalable (global) parameters that are transferable across locations and scales was

15 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

20 with area a_i. W_i and w_k denote fluxes at two modeling scales ℓ_1 and ℓ_1', respectively, with $\frac{\ell_1}{\ell_1'} = \frac{a_i}{a_k}$. 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 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.

25 ### 3.2 Regionalization and upscaling

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, 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 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 (Beven, 1995). The MPR approach also addresses the principle of scale-dependent subgrid parameterization (i.e., "net
5 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), 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 ap-
10 proach 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 scaling problem in MPR is approached by addressing the existence of 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,
15 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 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.
20 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
25 $\beta_0 = f(\mathbf{u}_0, \hat{\gamma})$. Note that the fields β_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 $\#\beta_0$ is orders of magnitude larger than that of the vector of the global parame-
30 ters $\#\hat{\gamma}$. Hence, MPR is a parsimonious (not over-parameterized) parameterization technique that offers spatially continuous model parameter fields and removes spatial discontinuities, 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 regionalized parameter to the
35 modeling scale. In other words, $\beta_1 = \langle \beta_0 \rangle$. 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."

5 Because the processes in LSM/HMs are highly nonlinear, this sequence of operations does not commute. The consequences can be dramatic (see Figure 7). The latter, which is the standard approach, <u>does not</u> preserve fluxes/states across scales, whereas MPR does to a great extent. The key question here is in finding the right scaling rule for the model parameters such that the fluxes/states are preserved.

Model parameters at the ℓ_1 scale (i.e., 1 km to 100 km) are called "effective" parameters because they cannot be measured

10 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 2.2) 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

15 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.

The selection of regionalization functions and scaling operators is fundamental to ensure the transferability of global param-

20 eters across scales and to guarantee the seamlessness of parameter fields across scales, e.g., from ℓ_1 to $2\ell_1$ ... 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

25 The development of LSM/HMs 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 to obtain a robust and seamless parameterization.

30   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 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 $\boldsymbol{\beta}_1$ using upscaling operators based on the underlying sub-grid variability $\boldsymbol{\beta}_0$. 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 $\hat{\boldsymbol{\gamma}}$ 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.

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 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;

5   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

10   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

15   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, we estimated porosity fields at three modeling resolutions of $\ell_1 =$ 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.

20   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, the mean value estimated at resolutions of 5, 10, and 25 km is approximately

25   0.42 $m^3 m^{-3}$ in all three cases.

The recently derived soil porosity field at 15 arcmin ($\approx$25 km) over the Pan-EU (Montzka et al., 2017), hereafter denoted as Miller-Miller Scaling (MMS), is included in Figure 3d to visualize the effects of the pedo-transfer functions and the scaling operators. It should be noted, however, that both approaches (i.e., MPR and MMS) use the same input SoilGrids data set (Hengl et al., 2017) at $\ell_0 = 1$ km. The main differences between both approaches stem from the type of regionalization functions and

30   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, Montzka et al. (2017) used the ROSETTA-based pedo-transfer-function (PTF) proposed by Schaap et al. (2001) and the upscaling method described by Miller and Miller (1956). MPR, on the other hand, employs the PTF proposed by Zacharias and Wessolek (2007) and the harmonic upscaling operator proposed by Zhu and Mohanty (2002) (see Livneh et al. (2015) for more details). Both methods lead to seamless fields because they follow the

principle "first estimate then average" which considers 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 5% lower than MPR but its standard deviation is in the same range as that of MPR at 10 km resolution. In particular, the porosity obtained by MMS over the northeastern part of Germany, eastern Europe, and Scandinavia are underestimated compared with MPR.

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.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-c,e-g). The method of Montzka et al. (2017), 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 parameters would have large biases that would likely lead to large predictive uncertainties.

4 Experiments that reveal inadequate parameterizations

4.1 Visualization of artifacts in state and flux fields

As noted in the introduction, on-site (basin-specific) parameter estimation based on HRU or similar techniques (based on 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 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 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, 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 shows clear evidence of strong spatial discontinuity 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 and are not useful for seamless and continental scale simulations.

4.2 Detecting effects of ad hoc regionalization techniques

The effects of the commonly used parameterization techniques used to generate the porosity fields of LSMs (such as CHTES-SEL and Noah-MP depicted in Figure 1) are important to investigate. These fields were 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 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 resolution, is depicted in Figure 1g. Compared with the other model derived porosity fields, the Noah-MP field appears to be most homogeneously distributed in space, and it is very likely that the spatial heterogeneity is under-represented in this case. It should be noted that a model such as CABLE that uses a porosity field with an approximately 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

To test this hypothesis, the highest resolution soil map available for Europe was 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 resolution (level-0). The upscaled porosity field was generated at 5 km for the EDgE project (edge.climate.copernicus.eu). The soil characteristics for Noah-MP were 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 to reduction of the variance of the porosity field and to 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.

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 1b 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 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. 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 employed is not scale invariant by comparing the parameter sets from both WaterGAP model versions, which were operated at different resolutions. The regionalization proposed by Beck et al. (2016) leads to a patch-quilt field that does not resemble to any other field presented. 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.

4.3 Changes in the dynamics of water fluxes and states

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 making reliable predictions of water and energy fluxes on land systems. In this context, it can be hypothesized that the parameterization of soil related parameters (e.g., soil porosity) has significant effects on related state variables and fluxes. To falsify this hypothesis, two contrasting modeling paradigms (Noah-MP and mHM) were 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 comparison, the process-based hydrological model mHM (www.ufz.de/mhm) was driven with daily precipitation and temperature fields generated by the WRF/Noah-MP system during the same period. The spatial resolution of mHM was fixed at 5×5 km². The main geophysical characteristics in WRF/Noah-MP of land cover and soil texture are represented with a 1×1 km² 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.

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

5 basins in Europe (Rakovec et al., 2016a). The porosity fields obtained for mHM over the Pan-EU is depicted in Figure 1f.

Notably, although the overall mean of the porosity estimated using MPR is only 2.3% lower 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.

10 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 -6.8% were detected in Germany, where a detailed evaluation was conducted by Samaniego et al. (2012). Within the same domain, the porosity values in Noah-MP could be up to 15% less than those estimated by mHM. However, at very few locations the opposite can happen, reaching values of 40% over-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($t+1$))

15 are subsequently estimated to investigate how the differences in porosity estimates of the top 2 m soil column affect the soil moisture dynamics (Figure 6). Two locations in Germany are selected in which Noah-MP significantly 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 $(54°N, 10°E)$ and $(51°N, 7°E)$, respectively). This experiment unambiguously shows that, at locations where Noah-MP over-estimates latent heat with respect to mHM, the dynamic of the monthly SM time series is enhanced, leading to much

20 lower soil moisture values (extreme droughts) (Figure 6a). Conversely, underestimation of latent heat greatly constrains the soil moisture dynamics (Figure 6b).

4.4 Flux matching test

In Section 2.1, it was postulated that ad hoc parameterization schemes do not fulfill the flux-matching test performed with a flux simulated by a given model at two modeling resolutions (e.g., $\ell_1 = 5, 30$ arcmin). A detailed description of how to perform

25 this test is provided in Samaniego et al. (2010b). The following experiment was conducted with three models: mHM, PCR-GLOBWB, and WaterGAP at the resolutions above to try to falsify this strong postulation. All models use the same forcings and geophysical information. The experiment was 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.

30 (2015) and www.hyperhydro.org/. The KGE and bias of these three models obtained for both scales at the Lobith station during 2003 are reported in Table 2. The streamflow during this year was selected for evaluation because it exhibited strong temporal dynamics, with wet conditions in the beginning of the year followed by a drought during the summer. These efficiency metrics indicate that mHM is the only model that can have higher KGE efficiencies regardless of the scale.

The flux-matching test presented in Section 3.1 was 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. The results of this test are shown in Figure 7. These results reveal that mHM is the only model able to fulfill the flux-matching test, as apparently in this figure. This experiment also shows that the MPR technique implemented in mHM leads to ET fields that satisfy mass

5   conservation across scales and exhibit seamlessness that becomes apparent at higher resolutions (see Figure 7a).

The PCR-GLOBWB and WaterGAP models reveal large inconsistencies in annual ET, 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

10   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 the ET mismatch. To falsify this postulation, the MPR parameterization protocol proposed in Section 3.3 was 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

15   to study 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 used to simulate the hydrological states at two different spatial resolutions ($\ell_1 =$5, 30 arcmin), and the sensitivity to MPR implementation was evaluated using a field difference method (in line with eq. 1):

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

20   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. 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

25   fluxes are shown in panels (c) and (d) of Figure 7 and are the result of these changes.

When MPR is employed, we see that the actual average Rhine basin evapotranspiration Δ drops from 29% to 9.4% (Figure8). 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 increase in the discharge KGE performance compared with observations at Lobith. The original KGEs were 0.86 ($\ell_1 =$5 arcmin) and 0.93 ($\ell_1 =$30 arcmin), whereas the KGEs with MPR implementation are 0.91

30   and 0.93, respectively. Another advantage is that PCR-GLOBWB is calibrated at a coarser resolution, whereas this model was 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.

5 ## 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 10 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 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 address this grand challenge, especially scale-invariant parameterization.

15 We presented and tested a technique called multiscale parameter regionalization (Samaniego et al., 2010b), currently available only in mHM but recently implemented in PCR-GLOBWB and VIC (Mizukami et al., 2017). 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.

This study has shown that two models that use ad-hoc parameterizations can have reasonable efficiency with respect to 20 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 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 trans-25 ferable 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. 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. We consider that this feature is the key for the next generation of LSM and NWP models such as the "model for prediction 30 across scales" (MPAS) (www.mmm.ucar.edu) and the "nested-domain ICOS" (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."

References

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

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.

Attinger, S.: Generalized Coarse Graining Procedures for Flow in Porous Media, Computational Geosciences, 7, 253–273, 2003.

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., Grayson, R. B., and Sivapalan, M.: On the representative elementary area (REA) concept and its utility for distributed rainfall-runoff modelling, Hydrological Processes, 9, 313–330, 1995.

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.

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.

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.

5 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

10 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.

15 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,

20 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,

25 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.

30 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.

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,

35 31, 699–712, 1995.

Fan, Y. and Bras, R. L.: On the concept of a representative elementary area in catchment runoff, Hydrological Processes, 9, 821–832, 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.

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.

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

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.

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.

Jakeman, A. and Hornberger, G.: How much complexity is warranted in a rainfall-runoff model?, Water Resources Research, 29, 2637–2650, 1993.

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.

Kozak, J. A., Ahuja, L. R., Ma, L., and Green, T. R.: Scaling and estimation of evaporation and transpiration of water across soil textures, Vadose Zone Journal, 4, 418, 2005.

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.

5 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.

10 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. Geolog-
15 ical 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.

20 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. 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
25 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, Hydrolog-
30 ical 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.

35 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.

Montzka, C., Herbst, M., Weihermüller, L., Verhoef, A., and Vereecken, H.: A global data set of soil hydraulic properties and sub-grid variability of soil water retention and hydraulic conductivity curves, Earth System Science Data Discussions, pp. 1–25, 2017.

Montzka, C., Herbst, M., Weihermüller, L., Verhoef, A., and Vereecken, H.: A global data set of soil hydraulic properties and sub-grid variability of soil water retention and hydraulic conductivity curves, link to model result files in NetCDF format,
5   doi:10.1594/PANGAEA.870605, https://doi.pangaea.de/10.1594/PANGAEA.870605, supplement to: Montzka, C et al. (in prep.): A global data set of soil hydraulic properties and sub-grid variability of soil water retention and hydraulic conductivity curves. Earth System Science Data Discussions, 2017.

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

Müller Schmied, H., Eisner, S., Franz, D., Wattenbach, M., Portmann, F. T., Flörke, M., and Doll, P.: Sensitivity of simulated global-scale
10   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,
15   2010.

Neuweiler, I. and Cirpka, O. A.: Homogenization of Richards equation in permeability fields with different connectivities, Water Resources Research, 41, n/a–n/a, 2005.

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,
20   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.

Oleson, K., Lawrence, D., Bonan, G., Drewniak, B., Huang, M., Koven, C., Levis, S., Li, F., Riley, W., Subin, Z., Swenson, S., Thornton, P.,
25   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.

Or, D. and Rubin, Y.: Stochastic modeling of unsaturated flow in heterogeneous media with water uptake by plant roots: Tests of the parallel
30   columns model under two-dimensional flow conditions, Water Resources Research, 29, 4109–4119, 2010.

Pachepsky, Y. A., Giménez, D., Crawford, J. W., and Rawls, W. J.: Conventional and fractal geometry in soil science, in: Fractals in Soil Science, edited by Y. Pachepsky, J. C. and Rawls, W., vol. 27 of *Developments in Soil Science*, pp. 7 – 18, Elsevier, 2000.

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.

35   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.

Popper, K.: The Logic of Scientific Discovery, Routledge Classics, Taylor & Francis, 1935.

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.

5 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.

Richardson, L. F.: Weather Prediction by Numerical Process, Cambridge University Press, 1922.

10 Rieu, M. and Sposito, G.: Fractal fragmentation, soil porosity, and soil water properties: I. Theory, Soil Science Society of America Journal, 1991.

Russo, D., Russo, I., and Laufer, A.: On the spatial variability of parameters of the unsaturated hydraulic conductivity, Water Resources Research, 33, 947–956, 1997.

Samaniego, L. and Bárdossy, A.: Robust parametric models of runoff characteristics at the mesoscale, Journal of Hydrology, 303, 136–151, 15 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.

20 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.

25 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.

Schaap, M. G., Leij, F. J., and van Genuchten, M. T.: ROSETTA: a computer program for estimating soil hydraulic parameters with hierarchical pedotransfer functions, Journal of Hydrology, 251, 163–176, 2001.

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

30 Seyfried, M. S. and Wilcox, B. P.: Scale and the Nature of Spatial Variability: Field Examples Having Implications for Hydrologic Modeling, Water Resources Research, 31, 173–184, 1995.

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, 35 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.

Stensrud, D. J.: Parameterization Schemes: Keys to Understanding Numerical Weather Prediction Models, Cambridge University Press, 2007.

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.,
5 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,
10 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.

Tyler, S. W. and Wheatcraft, S. W.: Fractal scaling of soil particle-size distributions: analysis and limitations, Soil Science Society of America
15 ..., 1992.

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.

Vereecken, H., Kasteel, R., Vanderborght, J., and Harter, T.: Upscaling Hydraulic Properties and Soil Water Flow Processes in Heterogeneous Soils, Vadose Zone Journal, 6, 1, 2007.

20 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.

Vogel, H. J. and Roth, K.: A new approach for determining effective soil hydraulic functions, European Journal of Soil Science, 49, 547–556,
25 1998.

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.

30 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.

35 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
5   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.

10   Woods, R., Sivapalan, M., and Duncan, M.: Investigating the representative elementary area concept: An approach based on field data, Hydrological Processes, 9, 291–312, 1995.

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.

Yeh, T. C. J., Gelhar, L. W., and Gutjahr, A. L.: Stochastic Analysis of Unsaturated Flow in Heterogeneous Soils: 2. Statistically Anisotropic
15   Media With Variable α, Water Resources Research, 21, 457–464, 1985.

Zacharias, S. and Wessolek, G.: Excluding Organic Matter Content from Pedotransfer Predictors of Soil Water Retention, Soil Sci. Soc. Am. J., 71, 43–50, 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
20   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.

Zhu, J. and Mohanty, B. P.: Spatial Averaging of van Genuchten Hydraulic Parameters for Steady-State Flow in Heterogeneous Soils: A Numerical Study, Vadose Zone Journal, 1, 261–272, 2002.
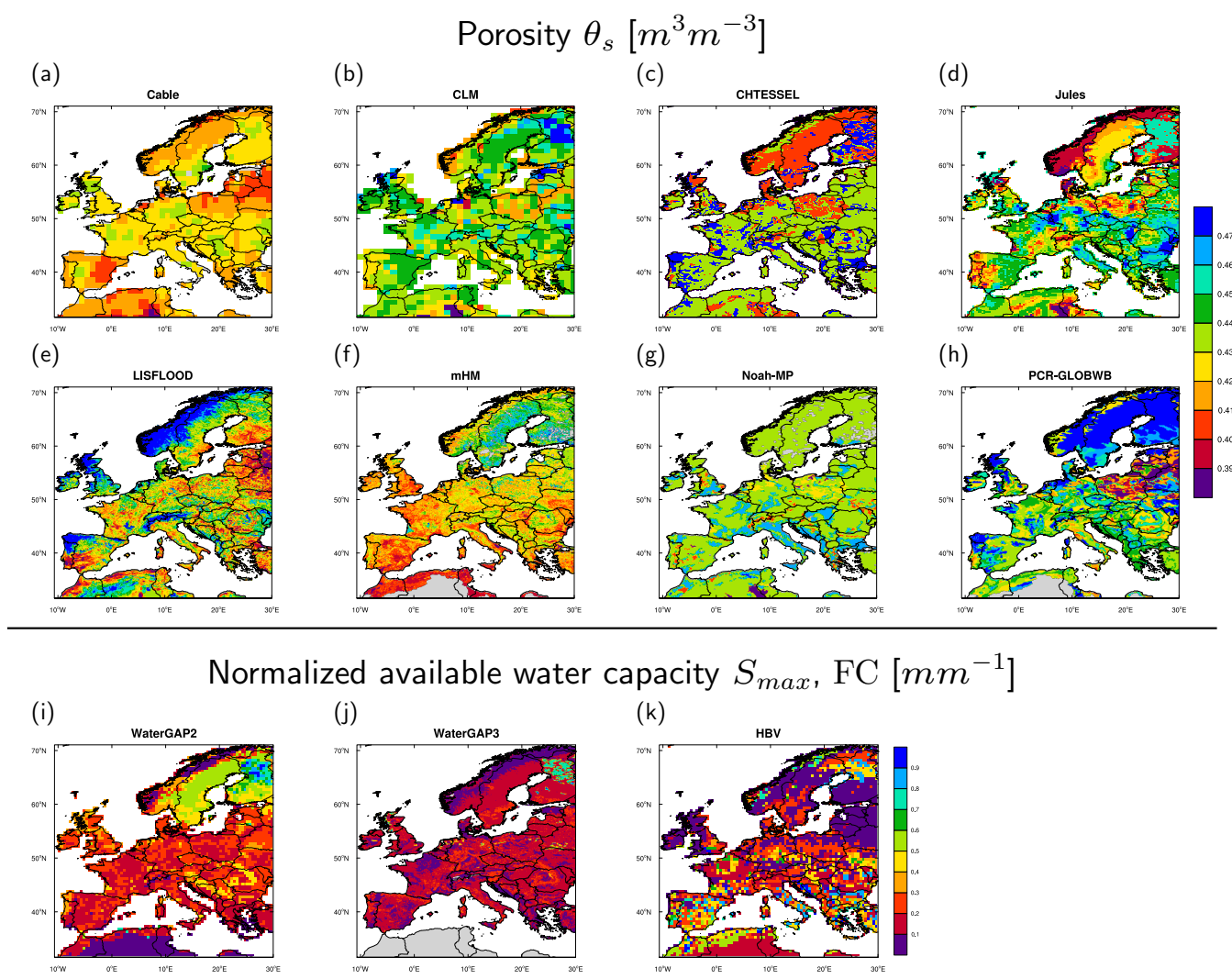
Figure 1. Porosity fields (top 2m) of typical LSM/HM over Pan-EU at various resolutions: (from left-right, top-down) (a) CABLE ($1°$), (b) CLM ($1°$), (c) CHTESSEL ($0.11°$), (d) JULES (35 km), (e) LISFLOOD (EFAS, 5 km), (f) mHM (EDgE-C3S, 5 km), (g) Noah-MP (CORDEX-EU, $0.11°$), and (h) PCR-GLOBWB (EDgE-C3S, 5 km). Normalized available water capacity of (i) WaterGAP2 (HyperHydro, 30 arcmin), [3,536] mm, (j) WaterGAP3 (HyperHydro, 5 arcmin), [1,960] mm, and (k) HBV [50,698] mm. The normalization values, denoted as [min,max], are provided only for panels (i)-(k). In parenthesis is provided the name of the project (if available) and the LSM/HM resolution.
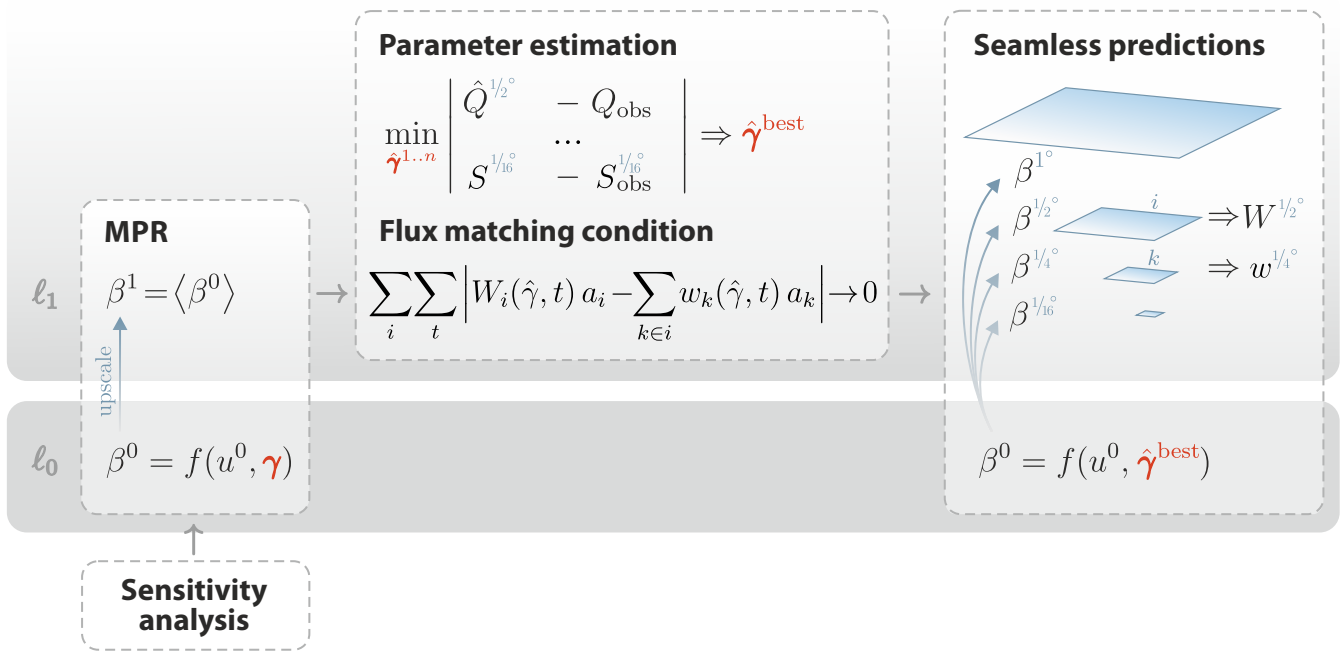
Figure 2. Schematic representation of the proposed seamless prediction framework based on Rakovec et al. (2016b). It includes a preliminary sensitivity analysis, MPR estimation, global-parameter estimation, a flux matching test, and multiscale seamless prediction. W_i and w_k are the fluxes at the i and k cells of the 1/2° and 1/4° resolutions, respectively (as an example). Q_{obs} and S_{obs} are the observed time series of streamflow and soil moisture, respectively. The operator $|\cdot|$ is a compromise dissimilarity metric composed of many independent observations at various scales.
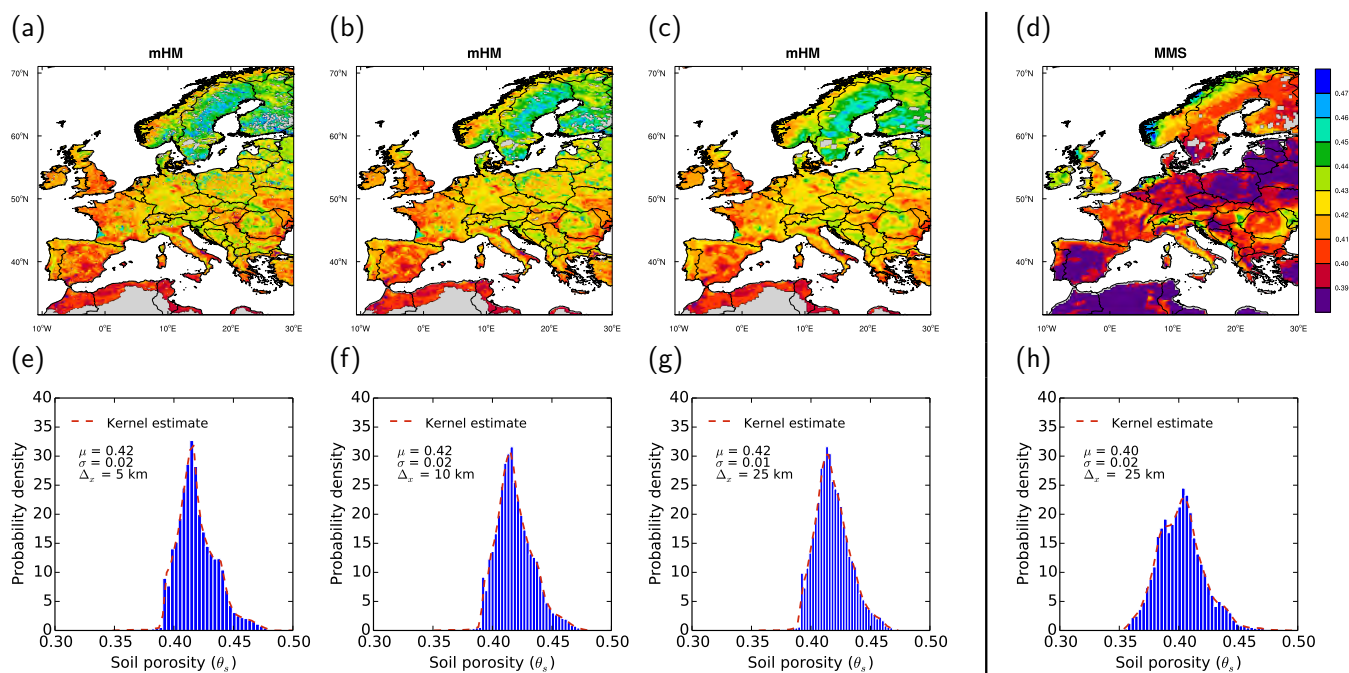
Figure 3. Seamless soil porosity (top 2 m) fields obtained using MPR at three spatial resolutions ℓ_1 (a) 5 km, (b) 10 km, and (c) 25 km, respectively. The soil porosity (denoted as MMS) estimated by Montzka et al. (2017) at 15 arcmin (top 2m) ($\approx$25 km) is depicted in panel (d). MMS shows the effects of estimating an effective parameter without checking the flux-matching condition in a LSM/HM. Lower panels (e)-(h) show the empirical distribution function of porosity at the respective resolution and method. The MMS soil porosity (top 2 m) was obtained as a weighed average of the porosity values provided for soil horizons at 0, 5, 15, 30, 60, 100, and 200 cm.
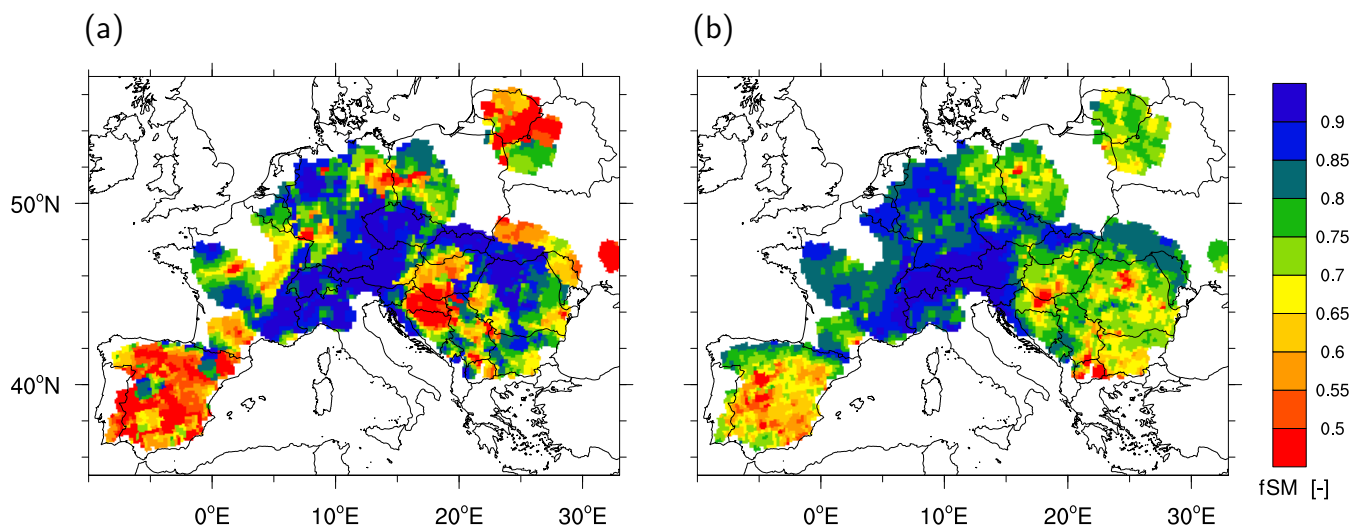
Figure 4. mHM simulations of soil moisture as the fraction from saturation $\frac{\theta}{\theta_s}$ for a day in August/2005 conducted with (a) on-site estimated parameters and (b) multi-basin estimated parameters. Panel (b) is a seamless soil moisture field.
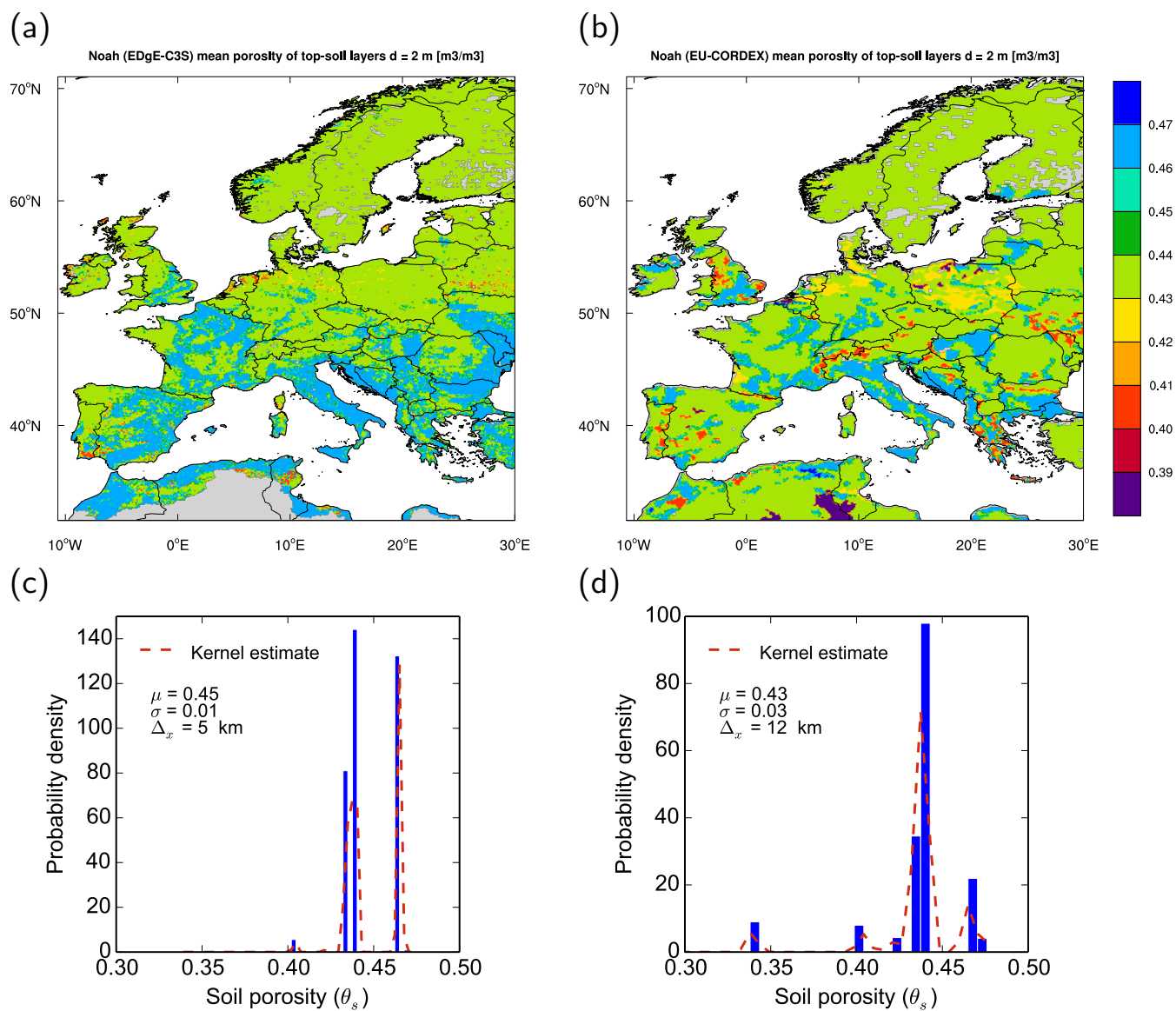
Figure 5. Porosity fields obtained using the majority upscale operator for spatial resolutions of (a) 5 km and (b) 12 km with the Noah-MP model used in the EDgE and EURO-CORDEX projects, respectively. Lower panels (c)-(d) show the empirical distribution function of porosity at the respective resolution and method.
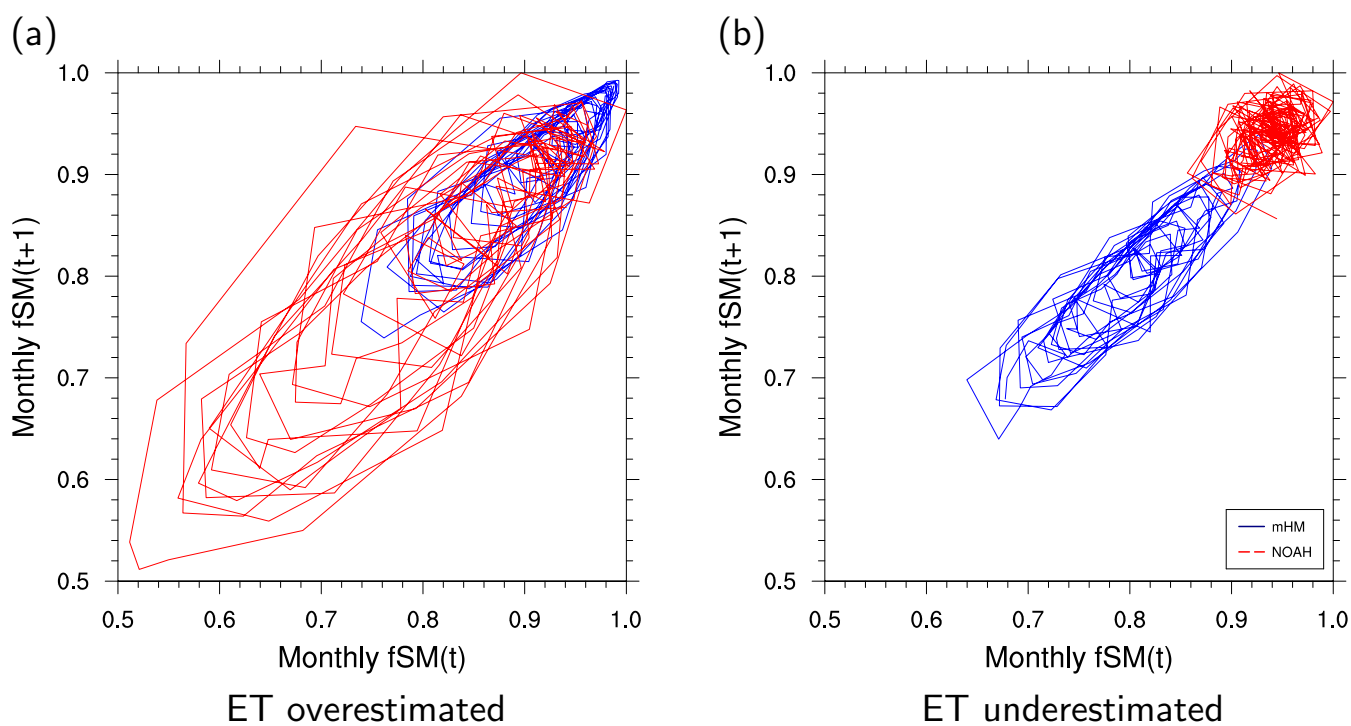
Figure 6. Phase diagrams of monthly soil moisture fraction for two locations in Germany, (a) $54°N,10°E$ and (b) $51°N,7°E$, in which the latent heat estimated by Noah-MP is over- or under-estimated with respect to corresponding estimates of mHM. The models have identical forcings.
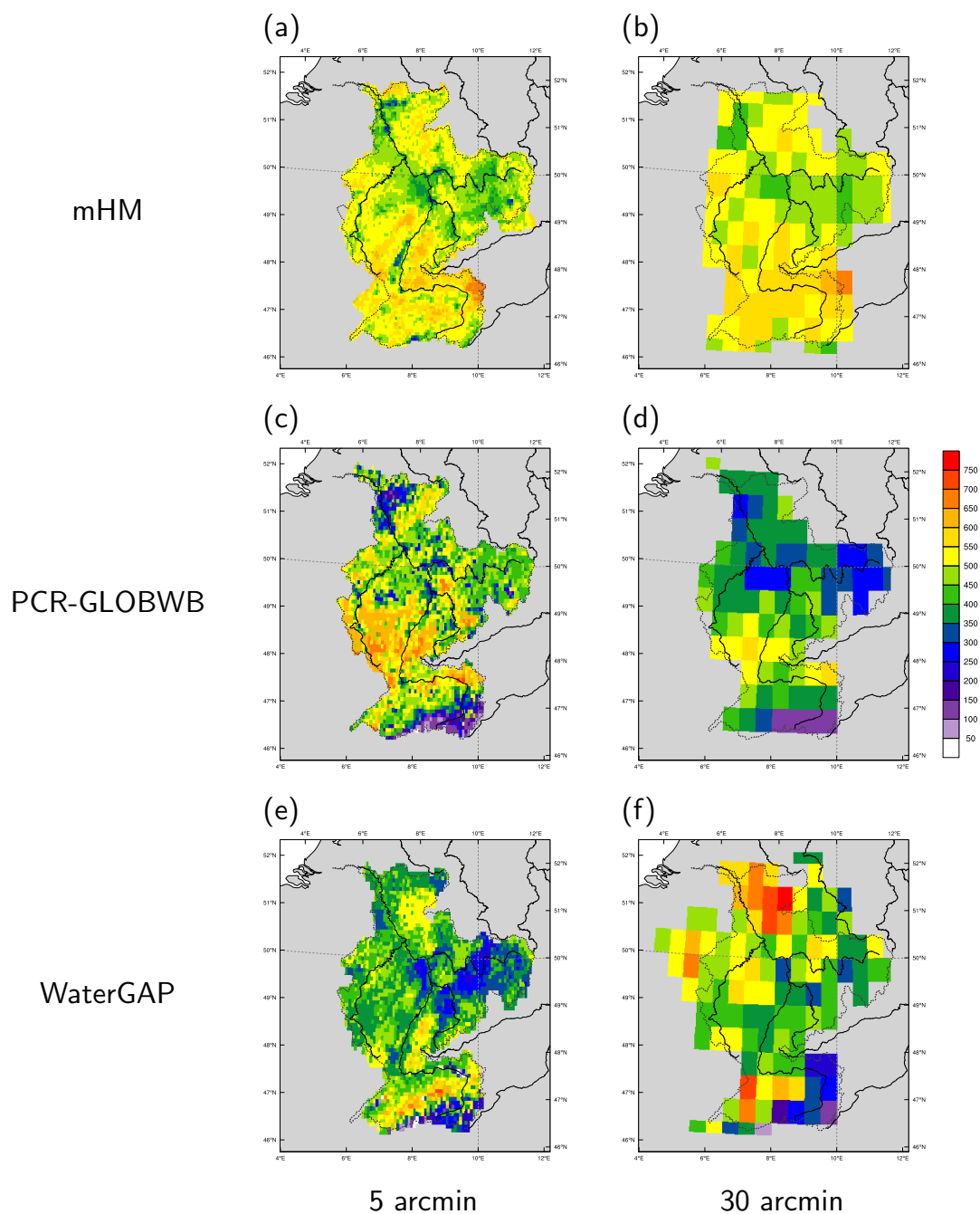
Figure 7. Multiscale simulation of ET for the Rhine River with mHM, WaterGAP (versions 3 and 2) and PCR-GLOBWB at spatial resolutions ℓ_1 of 5 and 30 arcmin, respectively.
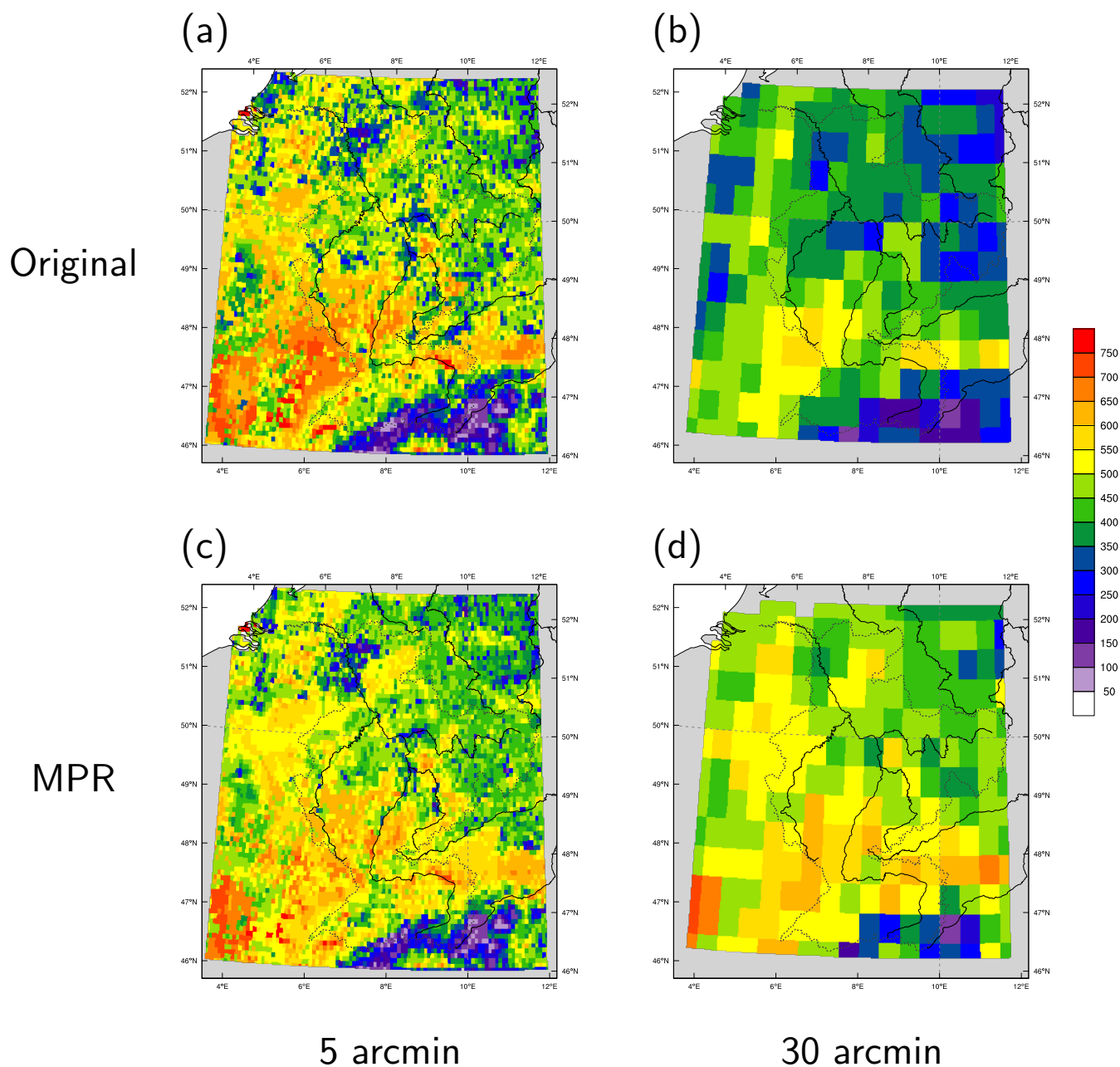
Figure 8. ET fields of PCR-GLOBWB before (panels a and b) and after implementing MPR (c and d) for two spatial resolutions of 5 and 30 arcmin. Dotted lines denote the Rhine basin and the continuous line is the main EU river basin network.

Hydrology and
Earth System
Sciences
Discussions

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