Hydrol. Earth Syst. Sci. Discuss., 10, 13191–13229, 2013
www.hydrol-earth-syst-sci-discuss.net/10/13191/2013/
doi:10.5194/hessd-10-13191-2013

Hydrology and Earth System Sciences Discussions
Open Access

Do land parameters matter in large-scale terrestrial water dynamics? – Toward new paradigms in modelling strategies

L. Gudmundsson and S. I. Seneviratne

Institute for Atmospheric and Climate Science, ETH Zurich, Universitaetstrasse 16, 8092 Zurich, Switzerland

Received: 14 October 2013 – Accepted: 25 October 2013 – Published: 4 November 2013

Correspondence to: L. Gudmundsson (lukas.gudmundsson@env.ethz.ch) and S. I. Seneviratne (sonia.seneviratne@env.ethz.ch)

Published by Copernicus Publications on behalf of the European Geosciences Union.

Abstract

Large-scale variations of terrestrial water storages and fluxes are key aspects in the Earth system, as they control ecosystem processes, feed back on weather and climate, and form the basis for water resources management. However, relevant observations are limited and process models used for estimation are highly uncertain. These models rely on approximations of terrestrial processes as well as on location-specific parameters (e.g. soil types, topography) to translate atmospheric forcing (e.g. precipitation, net radiation) into terrestrial water variables (e.g. soil moisture, river flow). To date it is unclear which processes and parameters should be included to model terrestrial water systems on regional to global scales. Using a data driven approach we show, that skillful estimates of monthly water dynamics in Europe can be derived from information on atmospheric drivers alone and that the inclusion of land parameters does not improve the estimate. The results highlight that substantial parts of terrestrial water dynamics are controlled by atmospheric forcing, which dominates over land parameters. This is not reflected in current model developments, which are striving at incorporating an increasing number of small scale processes and related parameters. Our results thus point at major potential for theory and model development, with important implications for water resources modelling, seasonal forecasting and climate change projections.

1 Introduction

Models of terrestrial water systems, commonly termed “Land Surface Models” (LSMs) or “Global Hydrological Models” (GHMs), are currently used to estimate terrestrial water dynamics at continental and global scales (Koster and Milly, 1997; Oki and Kanae, 2006; Dirmeyer et al., 2006; Haddeland et al., 2011; Mueller et al., 2011b; Gudmundsson et al., 2012a, b; Seneviratne et al., 2012) – either in offline mode (focusing on terrestrial water) or embedded into Global Climate Models. These models relate atmospheric forcing variables (e.g. precipitation, net radiation) to terrestrial water variables such as soil moisture, evapotranspiration or runoff. Their common structure can be summarised using the vertical water balance equation, which describes temporal changes in $S_{x,t}$, the water stored in a land unit at location x and time step t, as

$$\frac{\mathrm{d}S_{x,t}}{\mathrm{d}t} = P_{x,t} - E_{x,t} - Q_{x,t} \,, \tag{1}$$

where $P_{x,t}$ is precipitation, $E_{x,t}$ is evapotranspiration and $Q_{x,t}$ is runoff, the water that drains out of the system and eventually feeds into rivers. In LSMs/GHMs, evapotranspiration and runoff are modelled as functions of the terrestrial water storage such that $E_{x,t} = G(S_{x,t}, \mathrm{Ep}_{x,t}, \Pi_x)$ and $Q_{x,t} = H(S_{x,t}, P_{x,t}, \Pi_x)$, where $\mathrm{Ep}_{x,t}$ is the atmospheric water demand, also referred to as potential evapotranspiration, and Π_x is a set of locally varying land parameters related e.g. to soil types, topography, or vegetation cover. Different LSMs/GHMs are then defined by differences in the functions $G(\cdot)$ and $H(\cdot)$, which are representations of the relevant physical processes. These processes include e.g. the percolation of water through soils or root water uptake by plants. In practice, Eq. (1) is solved for discreet land units, so called grid cells. For offline simulations, these grid cells have typically dimensions comparable to the size of headwater catchments, i.e. hundreds to thousands square kilometres (Gudmundsson et al., 2012a). Consequently many of the processes that determine terrestrial water movements cannot be represented explicitly but have to be parametrised for the scale of implementation. Uncertainties in both the parametrisation (Sivapalan et al., 2003; Koster and Milly, 1997; Clark et al., 2011; Gupta et al., 2012) as well as in mapped values of land parameters (Sivapalan et al., 2003; Bastidas et al., 2006; Teuling et al., 2009; Fischer et al., 2011) are regularly highlighted as key limitations of model performance. While possible ways to improve LSMs/GHMs are vividly debated by the community (van den Hurk et al., 2011; Wood et al., 2011, 2012; Beven and Cloke, 2012), it remains unclear whether current limitations in model performance are mostly related to uncertain estimates of land parameters, or if a refinement of model formulation is necessary.

In order to discriminate between the impacts of model structure and land parameters on LSM/GHM performance, we confront here the above outlined representation of

HESSD
10, 13191–13229, 2013

terrestrial water systems with the radical hypothesis that hydrological variability at any location in space does solely depend on present and past atmospheric conditions – and not on locally varying land parameters. This so-called "Constant Land Parameter Hypothesis" (CLPH) assumes that one single set of model parameters is valid at every location in space (Fig. 1). This Hypothesis is based on observations that suggest that runoff dynamics in Europe can be separated in small- and large-scale phenomena (Sect. 2). The CLPH is formally introduced in Sect. 3 and tested in Sect. 4, where we first introduce a CLPH based null-model which we subsequently try to reject using a suite of alternative models. Finally, the results are discussed, emphasising possible implications for large-scale hydrological modelling.

2 Separation of scales

2.1 Theory

It is a common notion that hydrological phenomena have characteristic space and time scales (Klemeš, 1983; Blöschl and Sivapalan, 1995). Although the term scale is only vaguely defined in the context of hydrology, we adopt here the qualitative definition that the spatial (temporal) scale of a process is the characteristic length (duration) of the related phenomenon (Storch and Zwiers, 1999). For non periodic phenomena the characteristic length (duration) can be quantified using the de-correlation distance (time), i.e. the distance (time interval) at which two observations do not share common variations (Skøien et al., 2003; Storch and Zwiers, 1999). Empirically the de-correlation distance (time) is often estimated as the lag at which the spatial (temporal) autocorrelation function reaches $1/e$.

For the analysis of terrestrial water systems, it is useful to recall that early considerations on the related physics have shown that a statistical model of soil moisture corresponds to a first-order Markov process with the autocorrelation function

$$r(\tau) = \exp\left(-\frac{\tau}{T_c}\right) \tag{2}$$

where τ is a time lag and T_c is the de-correlation time (Delworth and Manabe, 1988).

Later, empirical investigations (Vinnikov et al., 1996; Robock et al., 1998; Entin et al., 2000) have led to the development of the hypothesis that soil moisture dynamics has two separated space and time scales: A short scale (order of 10 of meters and 1 day), where heterogeneous land properties dominate soil moisture dynamics and a large scale (order of 100 km and 1 month) where large features of the atmospheric forcing are dominating. More formally, the separation of time scales can be expressed as a mixture of two autocorrelation functions such that

$$r(\tau) = \zeta \exp\left(-\frac{\tau}{T_L}\right) + (1-\zeta)\exp\left(-\frac{\tau}{T_A}\right) \tag{3}$$

where T_L is the time scale related to heterogeneous land properties, T_A the time scale related to the atmospheric forcing and $\zeta \in [0,1]$ is the fraction of variance related to T_L. Note also that $T_L < T_A$. Similarly the separation of space scales can be expressed as

$$r(\lambda) = \eta \exp\left(-\frac{\lambda}{L_L}\right) + (1-\eta)\exp\left(-\frac{\lambda}{L_A}\right) \tag{4}$$

where λ is the lag distance, L_L is the length scale related to heterogeneous land properties, L_A the length scale related to the atmospheric forcing and $\eta \in [0,1]$ is the fraction of variance related to L_L.

2.2 Testing the separation of scales for runoff in Europe

While the separation of scales is well documented for soil moisture (Vinnikov et al., 1996; Robock et al., 1998; Entin et al., 2000; Crow et al., 2012; Mittelbach and Seneviratne, 2012), its application for other terrestrial water storages and fluxes is less clear.

The analysis of river flow time series in North America (Lins, 1997) and Europe (Gudmundsson et al., 2011a, b) has revealed continental scale regions of temporally homogeneous runoff variability. These correspond to features in the atmospheric forcing variables (Barlow et al., 2001; Tootle and Piechota, 2006; Gudmundsson et al., 2011b) and not to continental scale drainage basins (Lins, 1997). For small scales, some studies (e.g. Wood et al., 1988; Blöschl et al., 1995; Woods et al., 1995) highlight that locally heterogeneous land parameters start to dominate runoff production at spatial scales smaller than one kilometre. In contrast, a recent investigation (Skøien et al., 2003) suggested that the spatial scale of runoff lies between 18 and 59 kilometres. Note, however, that that Skøien et al. (2003) did not assess the existence of two different characteristic scales, which limits the interpretation of the results in the present context.

To check whether the hypothesised separation of scales is a valid assumption for runoff, we test whether Eqs. (3) and (4) are applicable for daily streamflow observations in Europe. The investigation is based on 426 streamflow series from small undisturbed catchments, which are a subset (Stahl et al., 2010) of the European Water Archive (EWA). The EWA is collected by the European Flow Regimes from International Experimental and Network Data (Euro-FRIEND) project (http://ne-friend.bafg.de/servlet/is/7413/, accessed: 26 September 2013) and held by the Global Runoff Data Centre (GRDC, http://grdc.bafg.de, accessed: 26 September 2013). Following a previous study (Skøien et al., 2003), daily runoff rates were log transformed and seasonal effects were removed. The deseasonalisation strictly follows recommendations on an optimal removal of the seasonal cycle in the mean and the variance using harmonic regression (Hipel and McLeod, 1994; McLeod and Gweon, 2013). Temporal correlation was first estimated for each gauging station separately. The maximum time lag was limited to 120 days to reduce effects of climate induced interannual variability, which is reportedly strong in the data under investigation (Gudmundsson et al., 2011b). The estimated temporal autocorrelation functions from the individual stations were finally averaged as in previous studies (Entin et al., 2000; Skøien et al., 2003; Vinnikov et al., 1996) to obtain an estimate of the mean runoff autocorrelation function in Europe. Spatial correlation was estimated using Morans I (Moran, 1950; Legendre and Legendre, 1998) for each time step separately with a spatial bin width of 10 km. This bin width is a compromise between having enough station pairs per bin and the ability to resolve small scale processes (the first bin contains 31 pairs, the median number of pairs: 490). The analysis of spatial correlation was limited to a maximum lag distance of 400 kilometres to reduce the effect of large scale climate gradients, which impact European runoff dynamics (Gudmundsson et al., 2011a, b). Finally the spatial correlation functions were then averaged over all time steps, resulting in an estimate of mean spatial correlation for the time period under investigation.

Figure 2 shows the estimated temporal and spatial correlation functions for runoff in Europe and Table 1 reports the parameters of Eqs. (3) and (4) fitted to the data. Note, however, that the lower limit of L_L was set to 10 km for the estimation procedure, to account for the limited resolution of the observed spatial autocorrelation function. The small p values of all parameters show that the hypothesised separation of scales is supported by observations.

3 The constant land parameter hypothesis

Several processes underlying terrestrial water dynamics are best understood on small scales. The associated equations (e.g. Richards equation) depend on locally varying land parameters (e.g. hydraulic conductivity) and capture phenomena occurring on a length scale of meters (e.g.: infiltration of water into soils). This reductionistic understanding, however, comes with a series of issues if larger scale phenomena such as the catchment or grid-cell scale water balance are targeted: Many of the necessary land parameters cannot be observed with the required spatial resolution, and the computational demand makes the implementation at spatial scales resolving these processes infeasible. Consequently, developing macroscopic theories that capture spatial and temporal averages of the target phenomena is necessary. Such developments will in

HESSD
10, 13191–13229, 2013

the following be referred to as process up-scaling. Following previous studies (Blöschl and Sivapalan, 1995; Wood, 1998) process up-scaling can be expressed as follows: A small scale representation of a water flux $v_{s,t}$ (e.g. evapotranspiration, drainage or runoff) at the spatial location x and the time t is given by

$$v_{x,t} = f(s_{x,t}, i_{x,t}, \pi_x) \tag{5}$$

where $s_{x,t}$, $i_{x,t}$ and π_x stand for storage (e.g. soil moisture), input (e.g. precipitation) and land parameters (e.g. hydraulic conductivity). An up-scaled process representation can then be deduced by integrating over a large spatial domain (X) and a long time interval (T) such that:

$$V = \int\limits_{t \epsilon T} \int\limits_{x \epsilon X} f(s_{x,t}, i_{x,t}, \pi_x) \mathrm{d}t \, \mathrm{d}x = F(S, I, \Pi) \tag{6}$$

where $F(\cdot)$ is the up-scaled process description, V is the total water flux, S the total storage, I the total input and Π a large scale summary of land parameters relevant for $F(\cdot)$. For some processes, Eq. (6) can be solved analytically, but the nonlinearity of the system have limited these approaches to special cases with strong assumptions (see e.g. Beven, 2006b, and references therein). An alternative (and often considered) approach is to specify $F(\cdot)$ in an ad-hoc fashion based on simple abstractions of the system. A classical example for this are (linear) reservoir (or bucket) models. (See e.g. Clark et al. (2008, 2011) for a comprehensive overview on popular formulations). Such models often appeal through their simplicity but lack a rigorous physical basis. Moreover the resulting land parameters do often not have an unambiguous physical interpretation, making it difficult to derive them from observable land properties.

Assuming that $F(\cdot)$ is known and that S and I can be satisfactorily approximated by other means, the major challenge for application is the identification of Π. Unfortunately many of the relevant land parameters are neither directly observable nor spatially homogenous, hampering precise estimation. Therefore hydrology has traditionally used inverse modelling approaches, identifying Π by calibration. While this can provide practical solutions for specific systems, it comes with several scientific caveats. It has for example long been noted that the calibration procedure can render an unambiguous physical interpretation impossible (Beven, 2006a). Moreover, the inverse problem cannot be solved for systems that have insufficient observations for calibration (Sivapalan et al., 2003), limiting the application to well monitored systems.

The above outlined separation of scales of terrestrial water storages and fluxes (Sect. 2, Fig. 2) raises question of the role of land parameters in process up-scaling. In particular, one can ask which effect this separation of scales has on the upscaling integral (Eq. 6) if the integration time and/or the area of integration exceed the small scale at which the influence of land parameters is dominating (i.e.: $T \gg T_\mathrm{L}$ and/or $\sqrt{X} \gg L_\mathrm{L}$). Here we hypothesise that the impact of changes in land parameters on $F(\cdot)$ is small compared to the influence of the atmospheric input if the integration domain exceeds the space or time scale at which land properties dominate terrestrial water dynamics. More formally this implies

$$O\left(\left|\frac{\partial F(\cdot)}{\partial \Pi}\right|\right) \ll O\left(\left|\frac{\partial F(\cdot)}{\partial I}\right|\right) \tag{7}$$

and suggests that process up-scaling may result in parametrisations that are independent of locally varying land parameters such that

$$V = F^{*}(S, I), \tag{8}$$

where $F^{*}(\cdot)$ is an up-scaled process description that is independent of locally varying land parameters. We refer to Eq. (8) as the Constant Land Parameter Hypothesis (CLPH, see also Fig. 1).

4 Testing the CLPH

4.1 Setup

In the following we test the CLPH by first constructing a null-model, based on the assumption that terrestrial water dynamics can be predicted on the basis of atmospheric forcing only. The explanatory power of the resulting model is then compared to the power of more sophisticated models that explicitly account for spatially varying land parameters. We reject the CLPH only if any of the alternative models performs significantly better. The rationale underlying this approach is motivated by Occam's razor, which emphasises that theories based on parsimonious assumptions are more powerful than their complex counterparts. In other words, an increased level of detail in system description needs to be justified by a higher explanatory power of the resulting model.

4.1.1 The CLPH model

To quantitatively assess the CLPH, a few pragmatic assumptions have to be made: (1) As systematic model evaluation relies on the abundance of observations, we focus on runoff from small catchments in Europe, a quantity that has been monitored for decades with relatively high spatial coverage; (2) Spatial and temporal resolution are chosen to be well above the space and time scales at which land properties are expected to have a dominating influence on terrestrial water dynamics (Fig. 2). Spatial resolution is set to $0.5^\circ \times 0.5^\circ$, which is a typical resolution of LSMs/GHMs and corresponds to the resolution of global scale estimators of atmospheric variables. Temporal resolution is chosen to be monthly, which is well above the typical LSM/GHM resolution but also a common resolution for many large scale applications. Having these assumptions in mind, it is possible to formulate the hypothesis that runoff ($Q_{x,t}$) at a land unit (x) and at any time step (t) solely depends on present and past atmospheric conditions, and

Discussion Paper | Discussion Paper | Discussion Paper | Discussion Paper |

HESSD
10, 13191–13229, 2013

not on land properties. More specifically this hypothesis is expressed as:

$$Q_{x,t} = F^{*}(\tau_n(I^1_{x,t}), \tau_n(I^2_{x,t}), \ldots, \tau_n(I^p_{x,t})), \tag{9}$$

where $I^1_{x,t}, \ldots, I^p_{x,t}$ are atmospheric forcing variables such as precipitation, temperature or humidity. The time lag operator τ_n is defined as $\tau_n(I_{x,t}) = [I_{x,t}, I_{x,t-1}, \ldots, I_{x,t-n}]$ and gives access to atmospheric conditions over the past n time steps (months), allowing to approximate the time integration in Eq. (6). The practical challenge is then to identify F^*, which is achieved using a machine learning tool called Random Forest (Breiman, 2001), an approach that is inspired by recent advances in estimating land-atmosphere fluxes using similar techniques (Jung et al., 2009, 2010, 2011; Zeng et al., 2012). (Note that these previous studies rely on location specific vegetation indices and are thus not compatible with the CLPH). The application of machine learning tools such as Random Forests to identify F^* has the advantage that no prior assumptions on the relevance of specific processes has to be made. The resulting model is referred to as the CLPH-Random Forest Model (CLPH-RFM).

The CLPH-RFM is set up assuming Eq. (9) and taking the atmospheric conditions of the past year into account ($n = 11$ months). We found that the results were more stable if the observed runoff-rates were log-transformed before model training. Details on Random Forests (Breiman, 2001) and specific parameter choices can be found in the Appendix A.

The CLPH-RFM is trained using runoff estimates based on streamflow observations from the EWA. As in previous studies (Gudmundsson et al., 2012a, b) streamflow observations from the 426 catchments where first converted into runoff rates per unit area and the coordinates of the corresponding gauging stations were assigned to the 0.5° grid cells defined by the atmospheric forcing data. If more than one gauging station occurred in one catchment the area weighted average runoff rate was used. This procedure results in 298 grid-cells with observed runoff. Estimates of atmospheric near-surface variables were taken from the WATCH (Water and Global Change) project (http://www.eu-watch.org/, accessed: 26 September 2013) Forcing Data (WFD) (Wee-

HESSD
10, 13191–13229, 2013

don et al., 2011). The analysis is based on the full WFD covering the following set of variables: Rainfall, snowfall, air temperature, incoming long and short wave radiations, humidity, surface pressure and wind speed.

4.1.2 Alternative models

The CLPH is tested by comparing the skill of the CLPH-RFM to an alternative model, which considers land parameters such that

$$Q_{x,t} = F(\tau_n(I^1_{x,t}), \tau_n(I^2_{x,t}), \ldots, \tau_n(I^p_{x,t}), \Pi_x), \tag{10}$$

where Π_x incorporates slope and information on soil texture, both being widely used in LSM/GHM parameterisations. Median grid-cell slope was derived from the HYDRO1k dataset available from the US Geological Survey. Information on soil texture for each grid-cell (median fraction of clay, silt, sand, gravel) were taken from the Harmonized World Soil Database (version 1.2) (FAO et al., 2012). The setup of this alternative model is identical to the CLPH-RFM, except for the inclusion locally varying land parameters.

The CLPH-RFM is further confronted with runoff simulations from nine state-of-the-art LSMs/GHMs (Gudmundsson et al., 2012a, b), developed by the WATCH project. Details on the simulation setup, key features of the participating models and further model validation can be found in the literature (Gudmundsson et al., 2012a, b). All participating models were forced using the WFD which guarantees a fair comparison with the CLPH-RFM introduced in this study. The LSM runoff simulations were augmented by the multi-model mean (MMM).

4.1.3 Model skill and hypothesis testing

Model skill of all participating models at locations with observations is evaluated using the skill score

$$S = \frac{A - A_{\text{ref}}}{A_{\text{perf}} - A_{\text{ref}}}, \tag{11}$$

HESSD
10, 13191–13229, 2013

where A is a measure of accuracy between the observed and the modelled values, A_{ref} is the measure of accuracy computed for a reference model, and A_{perf} is the perfect accuracy (Wilks, 2011). Here the mean annual cycle of monthly runoff observations is used as reference. The accuracy (A) is quantified using the mean square error computed on the logarithm of the observed and modelled values. The logarithmic transformation is necessary, as runoff has a very skewed distribution which would bias model skill toward extremely wet conditions.

The skill of the Random Forest Models (RFM) (Eqs. 9 and 10) in estimating runoff variability at locations that were not used for model training is assessed using a ten-fold cross validation. For this, the grid cells with observations were randomly stratified into ten subsamples. The CLPH-RFM was then trained using only nine of the subsamples and its skill was quantified using the grid cells of the remaining subsample. This procedure was repeated ten times, until model skill at each grid cell was estimated. Note that this validation strategy makes the analysis compatible with the Prediction of Ungauged Basins (PUB) initiative (Sivapalan et al., 2003; Hrachowitz et al., 2013; Parajka et al., 2013) of the International Association of Hydrological Sciences.

The CLPH is tested by comparing the skill of the CLPH-RFM to the skill of all alternative models. To do so, the median of the pairwise difference in skill, $\delta_{skill} = \text{median}(S_x^{CLPH} - S_x^{ALT})$, is assessed, where S_x^{CLPH} is the skill of the CLPH-RFM and S_x^{ALT} is the skill of an alternative model at locations with observations (x). Positive values indicate that the performance of the CLPH-RFM is superior in more than halve of the instances. Significance is assessed using 95 % confidence intervals of δ_{skill} (bias corrected and accelerated bootstrap estimates, (Mudelsee, 2010), 10^4 replications). The CLPH is rejected if $\delta_{skill} < 0$ (including confidence intervals) for any of the alternative models. The analysis of median difference in skill is augmented by $\Pr_{skill} = \Pr(S_x^{CLPH} > S_x^{ALT})$, the proportion of locations with observations where the CLPH-RFM outperforms the alternative model. Values of $\Pr_{skill} > 0.5$ indicate that that the CLPH-RFM has superior performance in more than 50 % of the cases. Signifi-

HESSD
10, 13191–13229, 2013

cance is assessed using 95 % bootstrap confidence intervals. The CLPH is rejected if $Pr_{skill} < 0.5$.

4.2 Results

4.2.1 Validating the CLPH-RFM

The CLPH-RFM (Eq. 9) proves, despite its limited assumptions, to be a reasonable estimator of monthly runoff at locations that were not used for model training (Fig. 3a–c, Appendix B). In 69 % of the grid cells it is found to be closer to the observations than a primitive model based on repetitions of the mean annual cycle and has an average skill well above zero (median skill: 0.35). This shows that the CLPH-RFM captures important aspects of runoff dynamics in Europe, even though some features remain unexplained. The CLPH-RFM proves further to be a reliable basis for estimating river discharge from pan-European river basins (Fig. 3d–g). (Observed continental river discharge and corresponding drainage areas are a subset of a previously assembled collection, Mueller et al., 2011a.)

In order to get an impression of the physical integrity of the CLPH-based runoff estimates with respect to other variables the long-term difference between the forcing precipitation and CLPH-RFM runoff was compared to a comprehensive estimate of land evapotranspiration from the LandFlux-EVAL synthesis product (Mueller et al., 2013). Figure 4 shows the mean evapotranspiration derived from the CLPH-RFM and the LandFlux-EVAL synthesis product. Overall the two products agree well ($R^2 = 0.66$), and the CLPH-RFM based estimate lies in the majority of the cases within the uncertainty bounds of the LandFlux-EVAL product. Note that the CLPH-RFM estimate does have small negative values in some parts of Scandinavia, which is related to a previously documented bias in the precipitation forcing (Gudmundsson et al., 2012b; Kauffeldt et al., 2013).

The reasonable performance of the CLPH-RFM with respect to (1) grid cell runoff, (2) discharge from continental drainage basins and (3) large-scale Evapotranspiration demonstrates the fidelity of the CLPH-RFM out of its expected comfort zone. Consequently these results suggest that the CLPH-RFM is a suitable null-hypothesis to assess the added value of more complex models. Note also that the CLPH-RFM can be used as a pragmatic estimator of continental scale terrestrial water dynamics (Fig. 5). Such reconstructions can be produced in near real time, as they solely rely on estimates of atmospheric variables, which are readily available from weather services.

4.2.2 Testing the CLPH-RFM

The CLPH is finally tested by comparing the skill of the CLPH-RFM (Eq. 9) to the skill of the alternative models. The skill of the Random Forest model taking land properties into account (Eq. 10) cannot be distinguished from the CLPH-RFM (Fig. 6, Appendix B and C), implying that the CLPH cannot be rejected on the basis of this experiment. This result shows that the impact of the tested land parameters on terrestrial water dynamics is small compared to the influence of atmospheric forcing at the space and time scales considered.

Overall, the CLPH-RFM displays a significantly higher skill in capturing observed monthly runoff rates than any of the LSM/GHM-based estimates (Fig. 6, Appendix B and C). Also this result does not allow us to reject the CLPH and suggests that LSMs/GHMs do not fully exploit the information available in the atmospheric forcing.

5 Discussion and conclusions

We note that issues common to all statistical applications can limit the interpretation of the presented results. Uncertainty in the used data, correlations between atmospheric forcing and land parameters, as well as an incomplete list of possible explanatory variables can influence the analysis. However, these limitations do also imply that the effect of the considered land parameters on large scale features of terrestrial water dynamics may have a similar order of magnitude as the mentioned disturbing factors.

HESSD
10, 13191–13229, 2013

The fact that the CLPH cannot be rejected on the basis of the presented experiments suggests that the influence of typically considered land parameters on runoff dynamics is small compared to the impact of atmospheric forcing on the considered temporal (monthly) and spatial (regional ≈ 50 km) scales. The physics underlying this result can be understood by recognising the scale dependence of terrestrial water dynamics (Sects. 2 and 3, Fig. 2). Only phenomena with small space and time scales are expected to be strongly controlled by processes that depend on land properties, whereas large scale phenomena are expected to be dominated by atmospheric drivers. Consequences of this scale dependence are twofold: (1) only models with very high spatial and temporal resolution can explicitly account for the influence of land parameters on terrestrial water dynamics; (2) but, conversely, the representation of large-scale terrestrial water dynamics is apparently little affected by these high resolution processes.

Current-generation LSMs/GHMs target terrestrial water dynamics at a temporal resolution (sub-daily) that requires consideration of small scale processes. However, the mismatch between their temporal and spatial resolution raises the question of whether this can be successful. To date, hydrology has not yet established unambiguous relations of model parameters to observable land properties (Sivapalan et al., 2003; Beven and Cloke, 2012; Hrachowitz et al., 2013; Parajka et al., 2013). This, together with the difficulties associated with a detailed mapping of the sub surface, implies that an increase of model resolution is unlikely to resolve this issue. In addition, these models are often used at scales not requiring this detailed process representation (e.g. within Earth System Models used for climate change scenarios, see Fig. 2).

The good performance of the CLPH-RFM suggests that the influence of land parameters may be negligible for phenomena occurring at monthly and regional scales. This is not reflected in current LSM/GHM development, which is dominated by the attempt to incorporate an increasing number of small-scale processes (van den Hurk et al., 2011). Consequently the presented results open new avenues for theory and model development. The option to neglect a large number of small scale processes can e.g. facilitate the analytic assessment and can possibly help to establish well understood

HESSD
10, 13191–13229, 2013

L. Gudmundsson and S. I. Seneviratne

limits of predictability. Scale analysis of the relevant equations and observations can help to clarify meaningful model resolutions. Likewise the predictive uncertainty may be substantially reduced as the number of observationally ill-constrained parameters decreases.

In conclusion, our findings highlight that land parameters are less influential on large scale features of terrestrial water systems than commonly assumed and imply that a rethinking of optimal strategies in land surface and hydrological model development is necessary. The comparatively high skill of the CLPH-RFM suggests that substantial progress in this field can be achieved in the coming years. Eventually, a better understanding of terrestrial water systems has important implications in a number of research applications ranging from water resources modelling to seasonal forecasting and climate change projections.

Appendix A

Random Forests

Random Forests, RF, (Breiman, 2001) are based on large ensembles of a modified version of Classification and Regression Trees, each grown on a bootstrap sample of the data. Despite its considerable complexity, the RF algorithm (Breiman, 2001; Liaw and Wiener, 2002; Hastie et al., 2009) can be summarised in a simplified manner as:

1. Draw B bootstrap samples from the data.
2. For each bootstrap sample, grow a Random Forest tree by recursively repeating the following steps:
   (a) Select m of the available predictor variables at random.
   (b) Among the m selected variables: find the one with the split point that best partitions the data.

HESSD
10, 13191–13229, 2013

(c) Split the data into two nodes and repeat the two previous steps on each node until the terminal node has reached the minimum node size n.

3. The RF prediction for new data is the average of the predictions of the B individual trees.

The free parameters of RFs need to be specified by the user. We opted for $B = 1000$, $n = 10$, and $m = p/3$, where p is the number of predictor variables, following recommendations in the literature (Hastie et al., 2009). In general, we found the results to be little sensitive to the parameter choice as long as the number of grown trees (B) was large enough.

An important feature of RF is that they allow for estimating the error rate using the training data only:

1. At each of the B bootstrap iterations: Predict the data that are not in the bootstrap sample (also called "out-of-bag" or OOB data).

2. Average the OOB predictions of the B trees. Note that on average each data point will be 36 % of the instances OOB.

3. Calculate the OOB error (in the context of this study: skill score, Eq. (11) by comparing the OOB predictions to the observations.

OOB errors are approximately equal to K-fold cross validation errors (Breiman, 2001, see also Appendix B) and are computationally more efficient to estimate, as they do not depend on K training iterations. However, in the context of this study, the OOB estimate was not suitable as it does not allow to systematically evaluate model performance at spatial locations that were not used for model training.

Appendix B

Additional model validation: grid-cell scale

In order to increase confidence in the results, a series of additional validation experiments were made, aiming at assessing the stability of the Random Forest Models (RFMs) with respect to the influence of (a) the chosen performance metric, (b) the (cross-) validation strategy, and (c) the impact of forcing variables. Section B1 briefly describes the additional validation experiments and Sect. B2 summarises the most important results.

B1 Setup of additional validation experiments

B1.1 Performance metrics

It can happen that models perform well with respect to a particular performance metric, but less well with respect to others. To draw robust conclusions it is thus important to assess the model performance using several metrics. Therefore we augment the analysis presented in this study using additional performance metrics. Table B1 provides an overview of the additional performance metrics that are considered here, most of which are based on Eq. 11.

B1.2 Validation strategy

In the main body of the article, model skill is estimated using 10-fold cross-validation, leaving subsequently 10 % of the grid cells out (later referred to as "cv-space"). This approach is chosen as it allows for a direct testing of the CLPH. However, the ability of the model to estimate runoff variability over time periods that were not used for model training is also of interest. Therefore a second cross validation experiment has been conducted, where the data were split into ten continuous time blocks (later referred to as "cv-time").

HESSD
10, 13191–13229, 2013

For future applications of Random Forests in large-scale hydrological modelling, it is interesting to know whether the OOB estimates of model performance (see Appendix A) are comparable to cross-validation estimates. The answer to this is not straightforward, as the strong spatial and temporal correlation of the considered variables could lead to overly optimistic OOB errors.

B1.3 Influence of forcing variables

In the main body of article (Figs. 3–6) the full WATCH Forcing Data (Weedon et al., 2011) (WFD) are used (later referred to as "FULL forcing"), including: rainfall, snowfall, air temperature, incoming long- and shortwave radiation, humidity, surface pressure and wind speed. We opted to focus on the FULL forcing, as LSMs/GHMs typically take most of these variables into account. However, catchment-scale hydrological modelling is usually based on precipitation (the sum of rainfall and snowfall) and temperature only, which raises the question whether the RFM can be skillful if only these variables are considered (later referred to as "PT forcing").

B2 Results of additional validation experiments

Figure B1 shows the results of the additional validation experiments and puts them into context by comparing them to the performance of the LSMs/GHMs under consideration.

B2.1 Performance metrics

The RFM-based estimates of runoff variability in Europe systematically outperform the LSM/GHM based estimates for the majority of the available performance metrics (Fig. B1). The fact that BIAS and $\mathrm{BIAS}_{\log}$ computed for the RFM are hardly distinguishable from the LSMs/GHMs shows that biases are only a secondary issue in model performance. For R^2 the Multi Model Mean (MMM) of all LSMs/GHMs has a higher median performance than one of the RFM based estimators. Note, however, that the spread in R^2 is much larger for MMM, indicating issues with the stability of the MMM predictions.

The difference between the RFM and the LSMs/GHMs is largest for climatology – R^2 and climatology – $R^2_{\log}$. This is consistent with previously documented issues (Gudmundsson et al., 2012b) and shows that major improvements of LSM/GHM performance can be made by correctly accounting for factors controlling the seasonality of terrestrial water dynamics. The picture is less clear for the anomaly correlations (anomaly – R^2 and $R^2_{\log}$). In this case some LSMs/GHMs have a performance that is comparable with that of the RFM.

B2.2 Validation strategy

Overall there are only marginal differences in the performance metrics estimated on the basis of spatial (cv-space) or temporal (cv-time) cross validation.

Only for the RFM, taking locally varying land parameters into account, major differences occur. The fact that the temporal cross-validation leads to systematically lower performances points to issues with over-fitting if (non-informative) static predictors (i.e. land parameters) are included. Overall the results indicate that the RFM is not only suitable for estimating runoff variability at locations without observations, but also for estimating runoff variability at non observed time intervals (e.g. for filling missing values in time series or for extrapolating the runoff records)

The OOB performance metrics are on average always larger than the ones based on cross-validation estimates. This indicates that the OOB estimates of model performance are positively biased. However, the biases are small, suggesting that OOB estimates of model performance can be used in future applications of Random Forests for modelling hydrological variablity on continental scales.

B2.3 Influence of forcing variables

In a direct comparison, estimates based on the FULL set of forcing variables do always outperform the estimates only based on precipitation and temperature (PT forcing). However, the differences are often small, suggesting that the additional variables in the FULL forcing do only add little extra information. It is further noteworthy, that the RFM forced with precipitation and temperature only does also compare favourably to the LSM/GHM predictions. This result shows that skillfull estimates of continental scale runoff dynamics in Europe can be produced based on precipitation and temperature data only.

Appendix C

Additional model validation: continental scale river discharge

Table C1 compares the ability of the CLPH-RFM to estimate continental scale river discharge with the performance of the considered alternative models. The performance of the CLPH-RFM is not distinguishable from that of the alternative RFM taking land parameters into account.

Differences between the CLPH-RFM and the LSMs/GHMs are more pronounced. The CLPH-RFM overestimates river discharge in most of the cases, whereas the LSMs/GHMs underestimate this quantity. This difference does suggest that biases in the forcing data may have an influence on model performance. Note, however, that the large differences among the LSMs/GHMs do also point towards issues with their parametrisations. Correlation between observed and modelled monthly river-discharge is not sensitive to biases in the forcing data and reflects the models' ability to capture important dynamical features. The fact that the correlations of the LSMs/GHMs are systematically lower if compared to the CLPH-RFM shows that differences in performance are not only related to biases in the forcing data, but also related to model formulation.

Acknowledgements. This research contributes to the European Union (FP7) funded project DROUGHT-R&SPI (contract no. 282769). The effort to assemble the European Water Archive (EWA) by of the UNESCOIHP VII FRIEND programme, the generation of the WFD and the LSM/GHM ensemble by members of the European Union WATCH project (FP6) are gratefully acknowledged.

References

Barlow, M., Nigam, S., and Berbery, E. H.: ENSO, pacific decadal variability, and US summertime precipitation, drought, and stream flow, J. Climate, 14, 2105–2128, doi:10.1175/1520-0442(2001)014<2105:EPDVAU>2.0.CO;2, 2001. 13196

Bastidas, L. A., Hogue, T. S., Sorooshian, S., Gupta, H. V., and Shuttleworth, W. J.: Parameter sensitivity analysis for different complexity land surface models using multicriteria methods, J. Geophys. Res., 111, D20101, doi:10.1029/2005JD006377, 2006. 13193

Beven, K.: A manifesto for the equifinality thesis, J. Hydrol., 320, 18–36, doi:10.1016/j.jhydrol.2005.07.007, 2006a. 13199

Beven, K.: Searching for the Holy Grail of scientific hydrology: $Q_t = H(S, R, \Delta t)A$ as closure, Hydrol. Earth Syst. Sci., 10, 609–618, doi:10.5194/hess-10-609-2006, 2006b. 13198

Beven, K. J. and Cloke, H. L.: Comment on "Hyperresolution global land surface modeling: Meeting a grand challenge for monitoring Earth's terrestrial water"; by Eric F. Wood et al., Water Resour. Res., 48, W01801, doi:10.1029/2011WR010982, 2012. 13193, 13206

Blöschl, G. and Sivapalan, M.: Scale issues in hydrological modelling: a review, Hydrol. Process., 9, 251–290, doi:10.1002/hyp.3360090305, 1995. 13194, 13198

Blöschl, G., Grayson, R. B., and Sivapalan, M.: On the representative elementary area (REA) concept and its utility for distributed rainfall-runoff modelling, Hydrol. Process., 9, 313–330, doi:10.1002/hyp.3360090307, 1995. 13196

Breiman, L.: Random forests, Mach. Learn., 45, 5–32, doi:10.1023/A:1010933404324, 2001. 13201, 13207, 13208

Clark, M. P., Slater, A. G., Rupp, D. E., Woods, R. A., Vrugt, J. A., Gupta, H. V., Wagener, T., and Hay, L. E.: Framework for Understanding Structural Errors (FUSE): a modular framework to diagnose differences between hydrological models, Water Resour. Res., 44, W00B02, doi:10.1029/2007WR006735, 2008. 13198

Clark, M. P., McMillan, H. K., Collins, D. B. G., Kavetski, D., and Woods, R. A.: Hydrological field data from a modeller's perspective: Part 2: process-based evaluation of model hypotheses, Hydrol. Process., 25, 523–543, doi:10.1002/hyp.7902, 2011. 13193, 13198

Crow, W. T., Berg, A. A., Cosh, M. H., Loew, A., Mohanty, B. P., Panciera, R., de Rosnay, P., Ryu, D., and Walker, J. P.: Upscaling sparse ground-based soil moisture observations for the validation of coarse-resolution satellite soil moisture products, Rev. Geophys., 50, RG2002, doi:10.1029/2011RG000372, 2012. 13195

Delworth, T. L. and Manabe, S.: The influence of potential evaporation on the variabilities of simulated soil wetness and climate., J. Climate, 1, 523–547, doi:10.1175/1520-0442(1988)001<0523:TIOPEO>2.0.CO;2, 1988. 13195

Dirmeyer, P. A., Gao, X., Zhao, M., Guo, Z., Oki, T., and Hanasaki, N.: GSWP-2: multimodel analysis and implications for our perception of the land surface, B. Am. Meteorol. Soc., 87, 1381–1397, doi:10.1175/BAMS-87-10-1381, 2006. 13192

Entin, J. K., Robock, A., Vinnikov, K. Y., Hollinger, S. E., Liu, S., and Namkhai, A.: Temporal and spatial scales of observed soil moisture variations in the extratropics, J. Geophys. Res., 105, 865–877, doi:10.1029/2000JD900051, 2000. 13195, 13196

FAO, IIASA, ISRIC, ISSCAS, and JRC: Harmonized World Soil Database (version 1.2), Tech. rep., FAO, Rome, Italy and IIASA, Laxenburg, Austria, 2012. 13202

Fischer, E., Lawrence, D., and Sanderson, B.: Quantifying uncertainties in projections of extremes – a perturbed land surface parameter experiment, Clim. Dynam., 37, 1381–1398, doi:10.1007/s00382-010-0915-y, 2011. 13193

Gudmundsson, L., Tallaksen, L. M., and Stahl, K.: Spatial cross-correlation patterns of European low, mean and high flows, Hydrol. Process., 25, 1034–1045, doi:10.1002/hyp.7807, 2011a. 13196, 13197

Gudmundsson, L., Tallaksen, L. M., Stahl, K., and Fleig, A. K.: Low-frequency variability of European runoff, Hydrol. Earth Syst. Sci., 15, 2853–2869, doi:10.5194/hess-15-2853-2011, 2011b. 13196, 13197

Gudmundsson, L., Tallaksen, L. M., Stahl, K., Clark, D. B., Dumont, E., Hagemann, S., Bertrand, N., Gerten, D., Heinke, J., Hanasaki, N., Voss, F., and Koirala, S.: Comparing large-scale hydrological model simulations to observed runoff percentiles in Europe, J. Hydrometeorol., 13, 604–620, doi:10.1175/JHM-D-11-083.1, 2012a. 13192, 13193, 13201, 13202

Gudmundsson, L., Wagener, T., Tallaksen, L. M., and Engeland, K.: Evaluation of nine large-scale hydrological models with respect to the seasonal runoff climatology in Europe, Water Resour. Res., 48, W11504, doi:10.1029/2011WR010911, 2012b. 13192, 13201, 13202, 13204, 13211

Gupta, H. V., Clark, M. P., Vrugt, J. A., Abramowitz, G., and Ye, M.: Towards a comprehensive assessment of model structural adequacy, Water Resour. Res., 48, W08301, doi:10.1029/2011WR011044, 2012. 13193

Haddeland, I., Clark, D. B., Franssen, W., Ludwig, F., Voß, F., Arnell, N. W., Bertrand, N., Best, M., Folwell, S., Gerten, D., Gomes, S., Gosling, S. N., Hagemann, S., Hanasaki, N., Harding, R., Heinke, J., Kabat, P., Koirala, S., Oki, T., Polcher, J., Stacke, T., Viterbo, P., Weedon, G. P., and Yeh, P.: Multimodel estimate of the global terrestrial water balance: setup and first results, J. Hydrometeorol., 12, 869–884, doi:10.1175/2011JHM1324.1, 2011. 13192

Hastie, T., Tibshirani, R., and Friedman, J. H.: The Elements of Statistical Learning – Data Mining, Inference, and Prediction, Second Edition, Springer Series in Statistics, Springer, 2 edn., 2009. 13207, 13208

Hipel, K. and McLeod, A.: Time series modelling of water resources and environmental systems, vol. 45 of Developments in Water Science, Elsevier, 1994. 13196

Hrachowitz, M., Savenije, H. H. G., Blöschl, G., McDonnell, J. J., Sivapalan, M., Pomeroy, J. W., Arheimer, B., Blume, T., Clark, M. P., Ehret, U., Fenicia, F., Freer, J. E., Gelfan, A., Gupta, H. V., Hughes, D. A., Hut, R. W., Montanari, A., Pande, S., Tetzlaff, D., Troch, P. A., Uhlenbrook, S., Wagener, T., Winsemius, H. C., Woods, R. A., Zehe, E., and Cudennec, C.: A decade of Predictions in Ungauged Basins (PUB) – a review, Hydrolog. Sci. J., 58, 1198–1255, doi:10.1080/02626667.2013.803183, 2013. 13203, 13206

Jung, M., Reichstein, M., and Bondeau, A.: Towards global empirical upscaling of FLUXNET eddy covariance observations: validation of a model tree ensemble approach using a biosphere model, Biogeosciences, 6, 2001–2013, doi:10.5194/bg-6-2001-2009, 2009. 13201

Jung, M., Reichstein, M., Ciais, P., Seneviratne, S. I., Sheffield, J., Goulden, M. L., Bonan, G., Cescatti, A., Chen, J., de Jeu, R., Dolman, A. J., Eugster, W., Gerten, D., Gianelle, D., Gobron, N., Heinke, J., Kimball, J., Law, B. E., Montagnani, L., Mu, Q., Mueller, B., Oleson, K., Papale, D., Richardson, A. D., Roupsard, O., Running, S., Tomelleri, E., Viovy, N., Weber, U., Williams, C., Wood, E., Zaehle, S., and Zhang, K.: Recent decline in the global land evapotranspiration trend due to limited moisture supply, Nature, 467, 951–954, doi:10.1038/nature09396, 2010. 13201

Jung, M., Reichstein, M., Margolis, H. A., Cescatti, A., Richardson, A. D., Arain, M. A., Arneth, A., Bernhofer, C., Bonal, D., Chen, J., Gianelle, D., Gobron, N., Kiely, G., Kutsch, W., Lasslop, G., Law, B. E., Lindroth, A., Merbold, L., Montagnani, L., Moors, E. J., Papale, D., Sottocornola, M., Vaccari, F., and Williams, C.: Global patterns of land-atmosphere fluxes of carbon dioxide, latent heat, and sensible heat derived from eddy covariance, satellite, and meteorological observations, J. Geophys. Res., 116, G00J07, doi:10.1029/2010JG001566, 2011. 13201

Kauffeldt, A., Halldin, S., Rodhe, A., Xu, C.-Y., and Westerberg, I. K.: Disinformative data in large-scale hydrological modelling, Hydrol. Earth Syst. Sci., 17, 2845–2857, doi:10.5194/hess-17-2845-2013, 2013. 13204

Klemeš, V.: Conceptualization and scale in hydrology, J. Hydrol., 65, 1–23, doi:10.1016/0022-1694(83)90208-1, 1983. 13194

Koster, R. D. and Milly, P. C. D.: The Interplay between transpiration and runoff formulations in land surface schemes used with atmospheric models, J. Climate, 10, 1578–1591, doi:10.1175/1520-0442(1997)010<1578:TIBTAR>2.0.CO;2, 1997. 13192, 13193

Legendre, P. and Legendre, L.: Numerical ecology, Elsevier, New York, 1998. 13197

Liaw, A. and Wiener, M.: Classification and regression by randomForest, R News, 2, 18–22, 2002. 13207

Lins, H. F.: Regional streamflow regimes and hydroclimatology of the United States, Water Resour. Res., 33, 1655–1667, doi:10.1029/97WR00615, 1997. 13196

McLeod, A. I. and Gweon, H.: Optimal deseasonalization for monthly and daily geophysical time series, Journal of Environmentl Statistics, 4, available at: http://www.jenvstat.org/v04/i11/ (last access: October 2013), 2013. 13196

Mittelbach, H. and Seneviratne, S. I.: A new perspective on the spatio-temporal variability of soil moisture: temporal dynamics versus time-invariant contributions, Hydrol. Earth Syst. Sci., 16, 2169–2179, doi:10.5194/hess-16-2169-2012, 2012. 13195

Moran, P. A. P.: Notes on continuous stochastic phenomena, Biometrika, 37, 17–23, 1950. 13197

Mudelsee, M.: Climate Time Series Analysis: Classical Statistical and Bootstrap Methods, vol. 42 of Atmospheric and Oceanographic Sciences Library, Springer Science+Business Media B. V., Dordrecht, doi:10.1007/978-90-481-9482-7, 2010. 13203

Mueller, B., Hirschi, M., and Seneviratne, S. I.: New diagnostic estimates of variations in terrestrial water storage based on ERA-Interim data, Hydrol. Process., 25, 996–1008, doi:10.1002/hyp.7652, 2011a. 13204

HESSD

10, 13191–13229, 2013

Do land parameters matter?

L. Gudmundsson and S. I. Seneviratne

Mueller, B., Seneviratne, S. I., Jimenez, C., Corti, T., Hirschi, M., Balsamo, G., Ciais, P., Dirmeyer, P., Fisher, J. B., Guo, Z., Jung, M., Maignan, F., McCabe, M. F., Reichle, R., Reichstein, M., Rodell, M., Sheffield, J., Teuling, A. J., Wang, K., Wood, E. F., and Zhang, Y.: Evaluation of global observations-based evapotranspiration datasets and IPCC AR4 simulations, Geophys. Res. Lett., 38, L06402, doi:10.1029/2010GL046230, 2011b. 13192

Mueller, B., Hirschi, M., Jimenez, C., Ciais, P., Dirmeyer, P. A., Dolman, A. J., Fisher, J. B., Jung, M., Ludwig, F., Maignan, F., Miralles, D. G., McCabe, M. F., Reichstein, M., Sheffield, J., Wang, K., Wood, E. F., Zhang, Y., and Seneviratne, S. I.: Benchmark products for land evapotranspiration: LandFlux-EVAL multi-data set synthesis, Hydrol. Earth Syst. Sci., 17, 3707–3720, doi:10.5194/hess-17-3707-2013, 2013. 13204, 13226

Nash, J. E. and Sutcliffe, J. V.: River flow forecasting through conceptual models part I – a discussion of principles, J. Hydrol., 10, 282–290, doi:10.1016/0022-1694(70)90255-6, 1970. 13221

Oki, T. and Kanae, S.: Global hydrological cycles and world water resources, Science, 313, 1068–1072, doi:10.1126/science.1128845, 2006. 13192

Parajka, J., Viglione, A., Rogger, M., Salinas, J. L., Sivapalan, M., and Blöschl, G.: Comparative assessment of predictions in ungauged basins – Part 1: Runoff-hydrograph studies, Hydrol. Earth Syst. Sci., 17, 1783–1795, doi:10.5194/hess-17-1783-2013, 2013. 13203, 13206

Robock, A., Schlosser, C. A., Vinnikov, K. Y., Speranskaya, N. A., Entin, J. K., and Qiu, S.: Evaluation of the AMIP soil moisture simulations, Global Planet. Change, 19, 181–208, doi:10.1016/S0921-8181(98)00047-2, 1998. 13195

Seneviratne, S. I., Nicholls, N., Easterling, D., Goodess, C., Kanae, S., Kossin, J., Luo, Y., Marengo, J., McInnes, K., Rahimi, M., Reichstein, M., Sorteberg, A., Vera, C., and Zhang, X.: Changes in climate extremes and their impacts on the natural physical environment, in: Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaptation, Cambridge University Press, Cambridge, UK, and New York, NY, USA, a Special Report of Working Groups I and II of the Intergovernmental Panel on Climate Change (IPCC), 109–230, 2012. 13192

Sivapalan, M., Takeuchi, K., Franks, S. W., Gupta, V. K., Karambiri, H., Lakshmi, V., Liang, X., Mc Donnell, J. J., Mendiondo, E. M., O'Connell, P. E., Oki, T., Pomeroy, J. W., Schertzer, D., Uhlenbrook, S., and Zehe, E.: IAHS decade on Predictions in Ungauged Basins (PUB), 2003–2012: shaping an exciting future for the hydrological sciences, Hydrolog. Sci. J., 48, 857–880, doi:10.1623/hysj.48.6.857.51421, 2003. 13193, 13199, 13203, 13206

Skøien, J. O., Blöschl, G., and Western, A. W.: Characteristic space scales and timescales in hydrology, Water Resour. Res., 39, 1304, doi:10.1029/2002WR001736, 2003. 13194, 13196

Stahl, K., Hisdal, H., Hannaford, J., Tallaksen, L. M., van Lanen, H. A. J., Sauquet, E., Demuth, S., Fendekova, M., and Jódar, J.: Streamflow trends in Europe: evidence from a dataset of near-natural catchments, Hydrol. Earth Syst. Sci., 14, 2367–2382, doi:10.5194/hess-14-2367-2010, 2010. 13196

Storch, H. V. and Zwiers, F. W.: Statistical Analysis in Climate Research, Cambridge University Press, Cambridge, 1999. 13194

Teuling, A. J., Uijlenhoet, R., van den Hurk, B., and Seneviratne, S. I.: Parameter sensitivity in LSMs: an analysis using stochastic soil moisture models and ELDAS soil parameters, J. Hydrometeorol., 10, 751–765, doi:10.1175/2008JHM1033.1, 2009. 13193

Tootle, G. A. and Piechota, T. C.: Relationships between Pacific and Atlantic ocean sea surface temperatures and US streamflow variability, Water Resour. Res., 42, W07411, doi:10.1029/2005WR004184, 2006. 13196

van den Hurk, B., Best, M., Dirmeyer, P., Pitman, A., Polcher, J., and Santanello, J.: Acceleration of land surface model development over a decade of GLASS, B. Am. Meteorol. Soc., 92, 1593–1600, doi:10.1175/BAMS-D-11-00007.1, 2011. 13193, 13206

Vinnikov, K. Y., Robock, A., Speranskaya, N. A., and Schlosser, C. A.: Scales of temporal and spatial variability of midlatitude soil moisture, J. Geophys. Res., 101, 7163–7174, doi:10.1029/95JD02753, 1996. 13195, 13196

Weedon, G. P., Gomes, S., Viterbo, P., Shuttleworth, W. J., Blyth, E., Österle, H., Adam, J. C., Bellouin, N., Boucher, O., and Best, M.: Creation of the WATCH forcing data and its use to assess global and regional reference crop evaporation over land during the twentieth century, J. Hydrometeorol., 12, 823–848, doi:10.1175/2011JHM1369.1, 2011. 13201, 13210

Wilks, D. S.: Statistical Methods in the Atmospheric Sciences, vol. 100 of International Geophysics Series, Academic Press, 3 edn., 2011. 13203

Wood, E. F.: Scale analyses for land-surface hydrology, in: Scale Dependence and Scale Invariance in Hydrology, Cambridge University Press, doi:10.1017/CBO9780511551864.002, 1–29, 1998. 13198

Wood, E. F., Sivapalan, M., Beven, K., and Band, L.: Effects of spatial variability and scale with implications to hydrologic modeling, J. Hydrol., 102, 29–47, doi:10.1016/0022-1694(88)90090-X, 1988. 13196

Wood, E. F., Roundy, J. K., Troy, T. J., van Beek, L. P. H., Bierkens, M. F. P., Blyth, E., de Roo, A., Döll, 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 Resour. Res., 47, W05301, doi:10.1029/2010WR010090, 2011. 13193

Wood, E. F., Roundy, J. K., Troy, T. J., van Beek, R., Bierkens, M., Blyth, E., de Roo, A., Döll, P., Ek, M., Famiglietti, J., Gochis, D., van de Giesen, N., Houser, P., Jaffe, P., Kollet, S., Lehner, B., Lettenmaier, D. P., Peters-Lidard, C. D., Sivapalan, M., Sheffield, J., Wade, A. J., and Whitehead, P.: Reply to comment by Keith, J. Beven and Hannah, L. Cloke on "Hyperresolution global land surface modeling: Meeting a grand challenge for monitoring Earth's terrestrial water", Water Resour. Res., 48, W01802, doi:10.1029/2011WR011202, 2012. 13193

Woods, R., Sivapalan, M., and Duncan, M.: Investigating the representative elementary area concept: an approach based on field data, Hydrol. Process., 9, 291–312, doi:10.1002/hyp.3360090306, 1995. 13196

Zeng, Z., Piao, S., Lin, X., Yin, G., Peng, S., Ciais, P., and Myneni, R. B.: Global evapotranspiration over the past three decades: estimation based on the water balance equation combined with empirical models, Environ. Res. Lett., 7, 014026, doi:10.1088/1748-9326/7/1/014026, 2012. 13201

HESSD
10, 13191–13229, 2013

Do land parameters matter?

L. Gudmundsson and S. I. Seneviratne

Table 1. Temporal and spatial scales of daily runoff in Europe: estimate, standard error and p value (t test) of the scaling models (Eqs. 3 and 4) fitted to observed temporal and spatial correlation functions using nonlinear least squares regression. Note, that the lower limit of L_L was set to the resolution of the empirical spatial correlation function (10 km).

	Temporal			Spatial		
	ζ [–]	T_L [days]	T_A [days]	η [–]	L_L [km]	L_A [km]
Estimate	0.50	7.4	68.3	0.51	≤ 10	180.5
Standard Error	3.8×10^{-3}	0.1	0.6	0.04	2.9	19.6
p value	< 0.001	< 0.001	< 0.001	< 0.001	0.002	< 0.001

HESSD
10, 13191–13229, 2013

Table B1. Additional performance metrics.

Notation	Description
MSE-skill:	Skill score (Eq. 11) using the Mean Square Error (MSE) as measure of accuracy and the mean annual cycle as reference model.
MSE_{log}-skill:	Same as MSE-skill, but computed on the logarithm of the data. This score is also used in the main body of the article.
MAE-skill:	Skill score (Eq. 11) using the Mean Absolute Error (MAE) as measure of accuracy and the mean annual cycle as reference model.
MAE_{log}-skill:	Same as MAE-skill, but computed on the logarithm of the data.
MEf:	Skill score (Eq. 11) using the MSE as measure of accuracy and the longterm mean as a reference model. MEf is also known as the model efficiency (Nash and Sutcliffe, 1970) which is a popular choice in hydrological modelling.
MEf_{log}:	Same as MEf, but computed on the logarithm of the data.
BIAS:	Difference between modelled and observed mean runoff. Negative (positive) values indicate that the models underestimate (overestimate) runoff.
$BIAS_{log}$:	Same as BIAS, but computed on the logarithm of the data.
R^2:	The square of the product-moment correlation coefficient.
R^2_{log}:	Same as R^2, but computed on the logarithm of the data.
climatology – R^2:	The square of the product-moment correlation coefficient between the observed and the modelled mean annual cycle.
climatology – R^2_{log}:	Same as climatology – R^2, but computed on the logarithm of the data.
anomaly – R^2:	The square of the product-moment correlation coefficient between the observed and the modelled monthly anomalies.
anomaly – R^2_{log}:	Same as anomaly – R^2, but computed on the logarithm of the data.

HESSD
10, 13191–13229, 2013

Table C1. Ability of the models under consideration to estimate monthly river discharge from several continental scale basins in Europe: displayed are the longterm Bias, the correlation of observed and modelled monthly river discharge (R^2) and the correlation of observed and modelled monthly anomalies (R^2_{ano}).

	Ebro	Elbe	Garonne	Loire	Po	Rhine	Rhone	Seine	Weser
					Bias				
CLPH	0.32	0.25	0.42	0.26	−0.34	0.11	0.21	0.16	0.14
VLPH	0.32	0.27	0.45	0.25	−0.32	0.11	0.21	0.18	0.13
GWAVA	−0.15	−0.22	−0.57	−0.38	−1.23	−0.77	−0.77	−0.31	−0.44
H08	−0.04	−0.17	−0.10	−0.08	−1.07	−0.35	−0.36	−0.10	−0.19
Htessel	−0.22	−0.40	−0.70	−0.58	−1.37	−1.03	−1.03	−0.62	−0.68
Jules	−0.35	−0.45	−1.01	−0.69	−1.80	−1.24	−1.46	−0.65	−0.73
LPJmL	0.11	0.01	−0.23	−0.03	−1.16	−0.48	−0.59	−0.01	−0.11
MATSIRO	−0.29	−0.32	−0.82	−0.56	−1.31	−1.06	−1.16	−0.53	−0.57
MPI-HM	0.12	0.05	0.03	−0.03	−0.50	−0.13	−0.08	−0.16	−0.10
Orchidee	0.19	0.01	−0.00	−0.03	−0.79	−0.38	−0.29	0.09	−0.23
WaterGAP	0.06	−0.20	−0.41	−0.30	−0.27	−0.43	−0.37	−0.37	−0.38
					R^2				
CLPH	0.69	0.79	0.82	0.89	0.40	0.85	0.83	0.84	0.90
VLPH	0.68	0.78	0.82	0.89	0.35	0.84	0.79	0.83	0.90
GWAVA	0.30	0.17	0.38	0.34	0.25	0.40	0.37	0.31	0.35
H08	0.38	0.26	0.51	0.51	0.44	0.44	0.55	0.43	0.46
Htessel	0.31	0.13	0.42	0.28	0.40	0.29	0.49	0.23	0.22
Jules	0.00	0.01	0.02	0.05	0.21	0.04	0.08	0.03	0.00
LPJmL	0.24	0.09	0.40	0.28	0.36	0.35	0.43	0.26	0.27
MATSIRO	0.64	0.58	0.73	0.67	0.54	0.67	0.72	0.69	0.74
MPI-HM	0.57	0.47	0.78	0.73	0.38	0.62	0.72	0.69	0.67
Orchidee	0.09	0.00	0.20	0.06	0.27	0.11	0.19	0.09	0.06
WaterGAP	0.48	0.37	0.61	0.54	0.40	0.45	0.59	0.40	0.45
					R^2_{ano}				
CLPH	0.66	0.72	0.80	0.82	0.57	0.84	0.86	0.77	0.87
VLPH	0.65	0.71	0.80	0.81	0.57	0.83	0.85	0.76	0.87
GWAVA	0.28	0.19	0.42	0.34	0.37	0.46	0.49	0.29	0.36
H08	0.35	0.26	0.54	0.51	0.42	0.50	0.64	0.38	0.43
Htessel	0.28	0.17	0.43	0.35	0.36	0.37	0.55	0.26	0.27
Jules	0.06	0.05	0.07	0.05	0.26	0.22	0.30	0.04	0.09
LPJmL	0.22	0.14	0.39	0.29	0.31	0.39	0.49	0.27	0.32
MATSIRO	0.59	0.58	0.68	0.62	0.52	0.66	0.70	0.67	0.74
MPI-HM	0.48	0.33	0.68	0.62	0.42	0.58	0.69	0.55	0.54
Orchidee	0.14	0.08	0.33	0.21	0.28	0.27	0.40	0.19	0.21
WaterGAP	0.38	0.35	0.52	0.47	0.38	0.48	0.63	0.32	0.40

HESSD
10, 13191–13229, 2013

Do land parameters matter?

L. Gudmundsson and S. I. Seneviratne

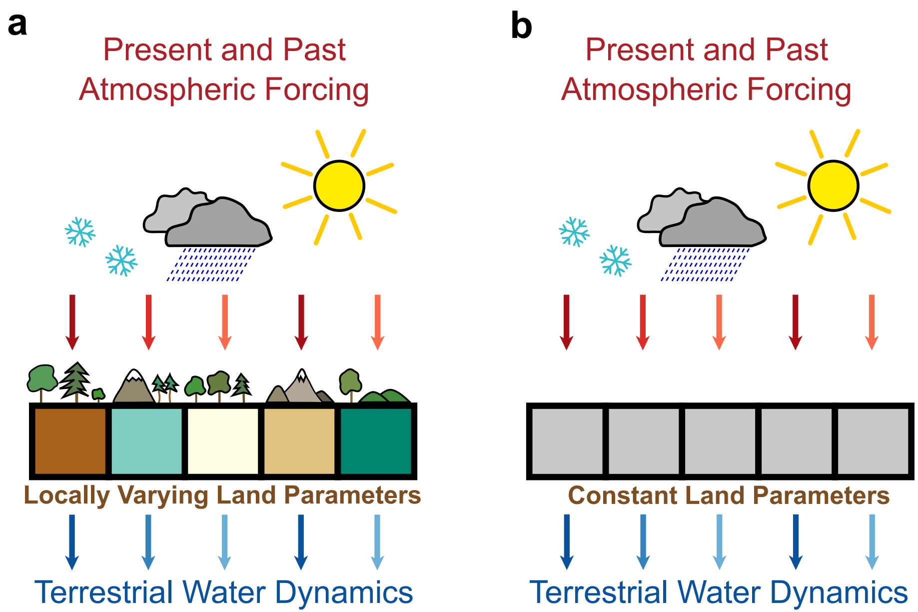

Fig. 1. The structure of LSMs/GHMs compared to that of models based on the Constant Land Parameter Hypothesis (CLPH): (**a**) LSMs/GHMs; (**b**) CLPH-based models. For both model types atmospheric forcing variables are processed by a terrestrial system, resulting in terrestrial water dynamics. LSMs/GHMs attempt to exploit information on locally varying land parameters to accurately simulate hydrological dynamics. Models built on the CLPH assume that terrestrial water dynamics depend only on present and past atmospheric forcing and not on locally varying land parameters.

HESSD
10, 13191–13229, 2013

Do land parameters matter?

L. Gudmundsson and S. I. Seneviratne

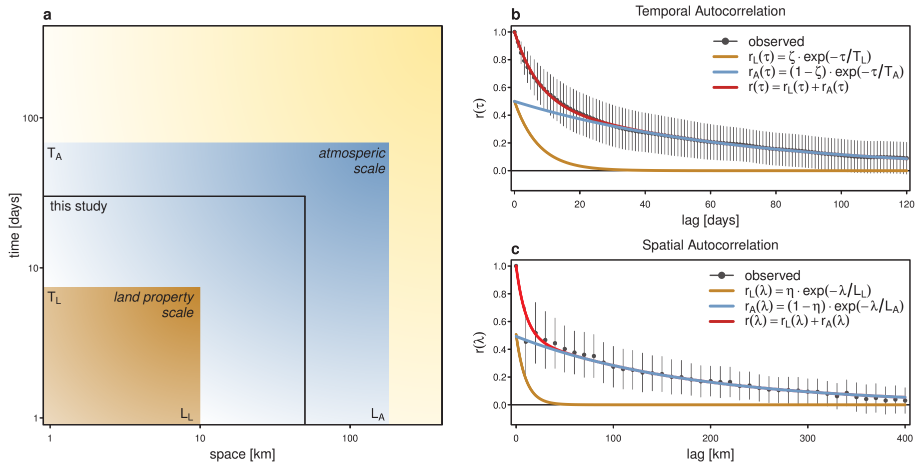

Fig. 2. Time and space scales of runoff in Europe: (**a**) empirical results show that runoff in Europe has two space and time scales. A small scale (T_L: time scale; L_L: space scale), at which runoff dynamics is strongly influenced by locally varying land properties, and a large scale (T_A: time scale; L_A: space scale) at which runoff dynamics is dominated by atmospheric forcing. Both the spatial and temporal resolution of this study are located well above the scales at which land properties are expected to have a strong influence on runoff dynamics. (**b, c**) Small and large scales are estimated from observed autocorrelations of daily runoff anomalies in Europe. See text for details.

HESSD
10, 13191–13229, 2013

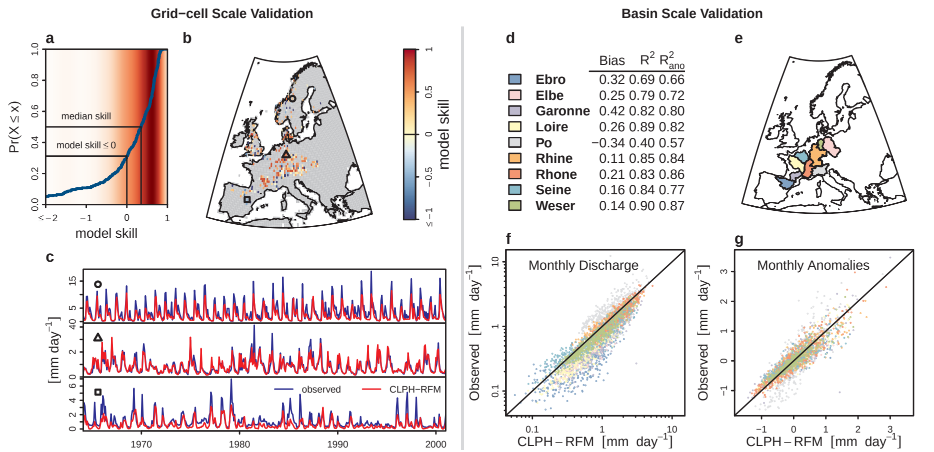

Fig. 3. Validation of the CLPH-based Random Forest Model (CLPH-RFM): the left panel shows validation against grid-cell scale runoff observations: (**a**) distribution of the CLPH-RFM skill in estimating monthly runoff rates at locations that were not used for model training. Model skill is quantified as the relative improvement over the climatology (best value: one, values smaller than zero indicate that replications of the observed mean annual cycle are a better estimator, see text). Blue line: cumulative distribution. Shading: indication of density, estimated using a kernel density estimator with Gaussian kernel. (**b**) Spatial distribution of model skill. The uniform distribution indicates the absence of regional biases. (**c**) Example series, comparing runoff observations and CLPH-RFM predictions (locations indicated on panel (**b**); square: arid region; triangle: temperate region; circle: cold region). The right panel displays the models' ability to estimate discharge from continental scale river basins (**d, e**: Bias: mean bias [mm day^{-1}], R^2: correlation between observed and modeled monthly discharge (**f**), R^2_{ano}: correlation between observed and modeled monthly discharge anomalies (**g**), values smaller than the 0.1 % percentile and larger than the 99.9 % percentile are excluded for visualization purposes only). Monthly discharge is estimated as the mean runoff of all grid-cells within the drainage area of a large river basin.

HESSD
10, 13191–13229, 2013

L. Gudmundsson and S. I. Seneviratne

Title Page
Abstract
Introduction
Conclusions
References
Tables
Figures
◀
▶
◀
▶
Back
Close
Full Screen / Esc
Printer-friendly Version
Interactive Discussion

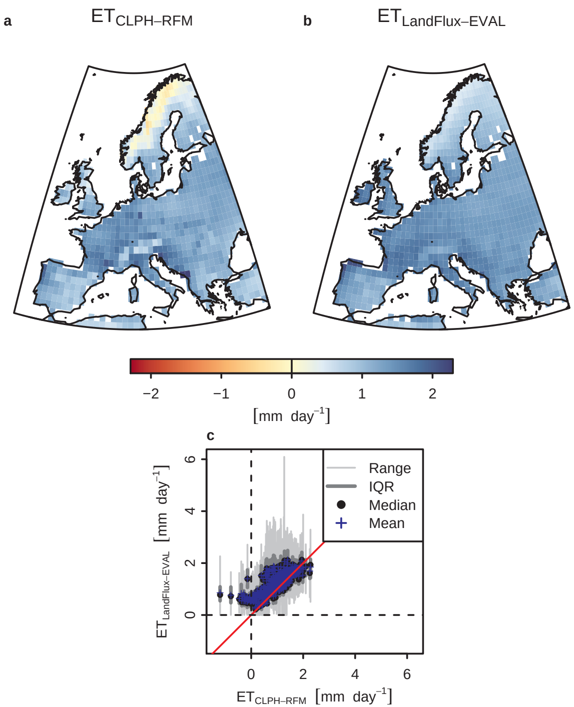

Fig. 4. Comparison of mean evapotranspiration (1989–1995) derived from the CLPH-RFM and the LandFlux-EVAL synthesis product: (**a**) Mean evapotranspiration computed as the mean difference between precipitation and runoff derived from the CLPH-RFM. (**b**) Mean evapotranspiration from the LandFlux-Eval synthesis product (Mueller et al., 2013). (**c**) Comparison of the CLPH-RFM and the LandFux-EVAL estimates of mean evapotranspiration. The vertical bars denote the interquartile range (IQR) and the range of all 40 data sets entering the LandFux-EVAL product. The points and crosses indicate the median and mean evapotranspiration of the LandFlux-EVAL product.

HESSD
10, 13191–13229, 2013

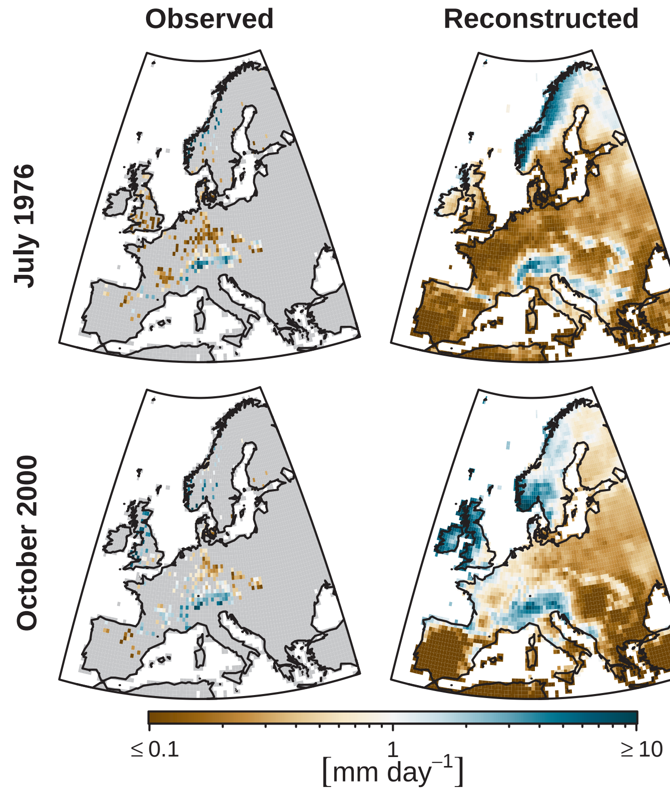

Fig. 5. Observed and reconstructed (CLPH-RFM) runoff in Europe. In July 1976 large areas of Europe where stricken by a severe drought, resulting in extremely low runoff rates. In October 2000 damaging floods occurred in several European regions, including Great Britain. Both events are captured by the observed and the modelled runoff rates.

HESSD
10, 13191–13229, 2013

Do land parameters matter?

L. Gudmundsson and S. I. Seneviratne

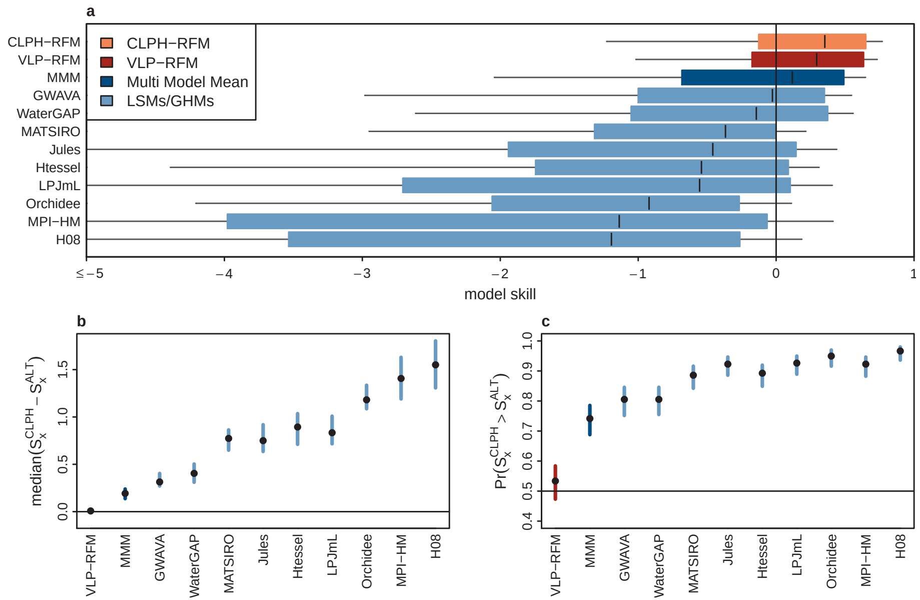

Fig. 6. Testing the constant land parameter hypothesis (CLPH): **(a)** Skill distribution of the CLPH Random Forest Model (CLPH-RFM, Eq. 9), compared to the skill of an alternative Random Forest Model, considering spatially varying land parameters (VLP-RFM, Eq. 10) and to an ensemble of LSMs/GHMs. Skill is computed for grid cells with observations. MMM: multi-model mean of all LSM/GHM simulations. The models are ranked according to the median skill of each model. Box-plots: the whiskers cover the region between the 10th and the 90th percentile; the box covers the inter-quartile range; the bar is the median. **(b)** Median of the difference in skill of the CLPH-RFM (S_x^{CLPH}) and the skill of the alternative models (S_x^{ALT}) at each location (x) with observations. Values larger than zero indicate that the CLPH-RFM outperforms the alternative models. **(c)** Proportion of $S_x^{\mathrm{CLPH}} > S_x^{\mathrm{ALT}}$. Values > 0.5 indicate that the CLPH-RFM outperforms the alternative models. The vertical bars are 95 % bootstrap confidence intervals.

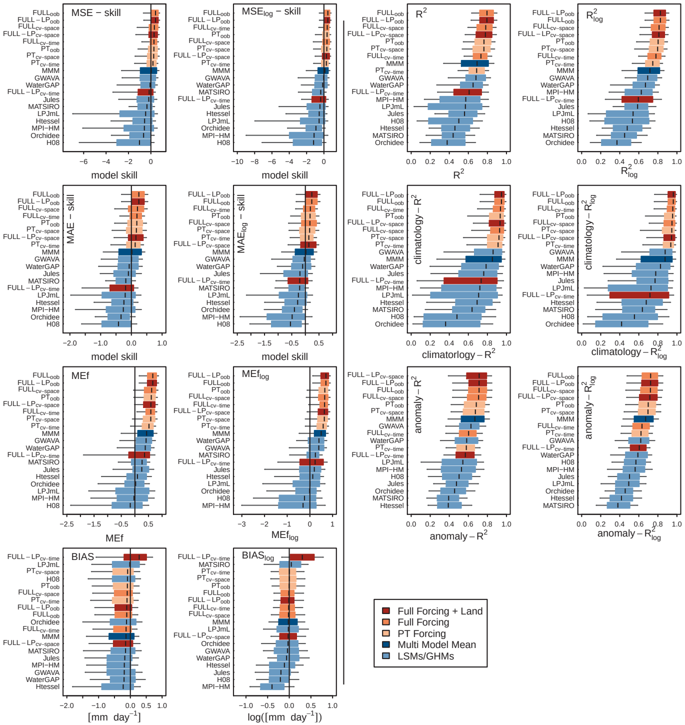

Fig. B1. Additional model validation: the individual panels correspond to the performance metrics introduced in Table B1. FULL: RFM forced with all variables available form the WFD. PT: RFM forced with precipitation and temperature only (see Sect. B1.3). The subscripts of the model names indicate different validation strategies (seel Sect. B1.2). cv-space: cross validation in space; cv-time: cross validation in time; oob: out-of-bag estimates of the performance metrics. Box-plots: the whiskers cover the region between the 10th and the 90th percentile; the box covers the inter-quartile range; the bar is the median.