A ranking of hydrological signatures based on their predictability in space

N. Addor[1,2,*], G. Nearing[3], C. Prieto[4], A. J. Newman[1], N. Le Vine[5], M. P. Clark[1]

[1] Hydrometeorological Applications Program, Research Applications Laboratory, National Center for Atmospheric Research, Boulder, USA
[2] Climatic Research Unit, School of Environmental Sciences, University of East Anglia, Norwich, UK
[3] University of Alabama, Tuscaloosa, USA
[4] Environmental Hydraulics Institute "IHCantabria", University of Cantabria, Santander, Spain
[5] Department of Civil and Environmental Engineering, Imperial College, London, UK
*Corresponding author: N.Addor@uea.ac.uk

Key points

We used machine learning (random forests) and a hydrological model to simulate 15 hydrological signatures over 671 catchments in the US.
The predictability of the signatures is highly correlated with the smoothness of their spatial pattern, which we quantified using Moran's *I*.
Poorly-predicted signatures vary abruptly in space, are sensitive to streamflow errors and their links to catchment attributes are elusive.

Abstract

Hydrological signatures are now used for a wide range of purposes, including catchment classification, process exploration and hydrological model calibration. The recent boost in the popularity and number of signatures has however not been accompanied by the development of clear guidance on signature selection. Here we propose that exploring the predictability of signatures in space provides important insights into their drivers, their sensitivity to data uncertainties, and is hence useful for signature selection. We use three complementary approaches to compare and rank 15 commonly-used signatures, which we evaluate in 671 US catchments from the CAMELS data set (Catchment Attributes and MEteorology for Large-sample Studies). Firstly, we employ machine learning (random forests) to explore how attributes characterizing the climatic conditions, topography, land cover, soil and geology influence (or not) the signatures. Secondly, we use simulations of a conceptual hydrological model (Sacramento) to benchmark the random forest predictions. Thirdly, we take advantage of the large sample of CAMELS catchments to characterize the spatial auto-correlation (using Moran's *I*) of the signature field. These three approaches lead to remarkably similar rankings of the signatures. We show i) that signatures with the noisiest spatial pattern tend to be poorly captured by hydrological simulations, ii) that their relationship to catchments attributes are

elusive (in particular they are not correlated to climatic indices) and iii) that they are particularly sensitive to discharge uncertainties. We suggest that a better understanding of their drivers and better characterization of their uncertainties would increase their value in hydrological studies.

1 Introduction

Hydrological signatures (indices characterizing hydrologic behavior) are now commonly used to understand space-time variability in hydrological processes (Troch et al., 2009; Sawicz et al., 2011) and to diagnose weaknesses in hydrological models (Gupta et al., 2008; Euser et al., 2013; Vrugt and Sadegh, 2013). Signatures can be computed using a wide range of data sources, but in practice they are most often computed using discharge time series (e.g., Yilmaz et al., 2008) and referred to as streamflow indices or streamflow characteristics. Hydrological signatures are particularly useful to characterize and compare the dynamics of large samples of catchments, for which observations are typically limited only to streamflow (streamflow is measured, but evapotranspiration, snow water equivalent, tracer concentrations or water table level are usually not measured). In a sense, hydrological signatures are an indirect way to explore hydrological processes, when those processes cannot be isolated because of the lack of measured data. This enables in particular catchment classification (Sawicz et al., 2011) and provides insights into hydrological behavior in places where little to no data are available apart from streamflow (Kuentz et al., 2017). Hydrological signatures are also increasingly used for model calibration (Euser et al., 2013; Hrachowitz et al., 2014) and model selection (Clark et al., 2011; McMillan et al., 2011; Schaefli, 2016).

A profusion of hydrological signatures already exists, and more are being developed. The diversity of hydrologic signatures enables the characterization of a wide variety of hydrological features, but at the same time, makes selecting appropriate signatures challenging (McMillan et al., 2017). There are some general selection criteria; for instance, it is desirable that i) signatures can be related to hydrological processes to enable a better understanding of particular aspects of catchment behavior, ii) they are sensitive to processes occurring over different periods (from the sub-daily to the decadal time scale, e.g., Shamir et al., 2005), and iii) they are not redundant (i.e., they do not share information content). Yet, signature selection is essentially dealt with on a case-by-case basis, different studies invariably use different signatures, and the same signatures may be computed in different ways (e.g., the baseflow index). While it is normal that each study selects signatures to meet its specific needs, there is a need to develop general guidance on the selection of hydrologic signatures.

Here we propose that signature selection can be informed by considering a key, yet usually overlooked, aspect: the spatial predictability of signatures. Signatures in ungauged basins have been widely predicted based on climatic and physiographic attributes (i.e., regionalized, see Hrachowitz et al., 2013) typically following one of these three approaches: i) by employing a statistical model using catchment attributes as predictors (Nathan and McMahon, 1992; Lacey and Grayson, 1998; Yadav et al., 2007; Beck et al., 2013, 2015; Kuentz et al., 2017), ii) by transferring signatures from gauged catchments considered similar based on their attributes (Burns, 1990; Holmes et al., 2002; Westerberg et al., 2016) or iii) by running a hydrological model with parameters regionalized based on catchment attributes (Sefton and Howarth, 1998; Andréassian et al., 2006; Wagener and Wheater, 2006; Young, 2006; He et al., 2011; Ragettli et al., 2017). Over the last decades, the scope of regionalization studies expanded, and the number of signatures, catchment attributes and

basins increased. Three of these studies (Beck et al., 2015; Westerberg et al., 2016; Kuentz et al., 2017) assess the regionalization of several signatures over a large number of gauging stations (3000 to 4000 depending on the signature, 43 and 1366, respectively) in different regions (global, UK and Europe) and using different techniques (neural networks ensembles, weighted-pooling-group approach and multiple regression models). Despite these different setups, these three studies demonstrate that there are significant differences in the predictability of hydrological signatures. The number of signatures they have in common is modest, but some signatures were considered by the three studies and interestingly, there is some consistency in their ranking: mean streamflow and high flow indices tend to the better predicted than low flow indices and the baseflow index. This study aims to synthesize the factors leading to good and poor predictability for a wide range of signatures, to explore how the predictability can be improved, and to discuss the implications of poor predictability for signature use in hydrological studies. The experimental setup was designed to address three interrelated research questions:

1. How well can signatures be predicted using landscape characteristics? We used a machine learning algorithm (random forests) to relate catchments attributes to hydrological signatures. To explore how the interplay of landscape attribute shapes hydrological behavior, we quantified the relative influence of the landscape attributes in the random forests (i.e., how strongly landscape attributes influence the predictions of the hydrological signatures). We leveraged the hydrometeorological times series and catchment attributes of recently-released and particularly exhaustive data set covering 671 basins in the USA (the CAMELS data set, Newman et al., 2015; Addor et al., 2017b).

2. How well can signatures be simulated by a calibrated conceptual hydrological model? We ask whether explicitly accounting for hydrological processes (instead of adopting a purely statistical, data-driven approach) improves the signature predictions. There is a growing recognition of the utility of benchmarks in hydrology (Seibert, 2001; Best et al., 2015; Newman et al., 2017; Seibert et al., 2018), we use these hydrological simulations to evaluate the predictions of our random forests, keeping in mind that random forests are significantly quicker to setup and run than calibrated hydrological models.

3. Why are some signatures better regionalized than others? It is known that the accuracy of the regionalization varies from signature to signature, but it is unclear why this is the case. Here we rank the signatures based on their predictability in space and explore whether their predictability is related to data uncertainty and lack of process understanding.

The remainder of this paper is organized as follows: The data and methods are presented in Section 2; the ranking of signatures is presented in Section 3; the reasons behind this ranking and implications for signature use are discussed in Section 4; conclusions and future research needs are presented in Section 5.

2   Data and methods

2.1   The CAMELS data set

All the data used in this study come from the CAMELS data set (Catchment Attributes and MEteorology for Large-sample Studies). The CAMELS data set covers 671 catchments in the contiguous US (CONUS) and consists of two types of data: daily time series of observed atmospheric forcing, observed discharge, and simulated discharge (Newman et al., 2014, 2015) and catchment attributes selected to provide a quantitative description of landscape features likely to influence hydrological processes (Addor et al., 2017a, 2017b). CAMELS is a unique combination of a large number of diverse catchments (671 US catchments with long streamflow time series, minimally impacted by human activities, and covering a wide range of hydro-climatic conditions), 15 hydrological signatures characterizing a wide range of hydrograph features, and 43 climatic and physiographic attributes based on recent and well-documented data sets, and computed in order to enable direct basin comparisons (the same data sets were used for all the catchments and the attributes were computed over the same period). The hydrometeorological time series and catchment attributes are described in Sections 2.2 and 2.3, respectively.

2.2   CAMELS hydrometeorological time series

The hydrometeorological time series include both daily meteorological forcing and observed discharge time series, as well as daily hydrological simulations. Precipitation and temperature at the catchment scale were retrieved from the Daymet data set (Thornton et al., 2012). Potential evapotranspiration was estimated based on Priestley and Taylor (1972). The hydrologic simulations were produced using the Sacramento Soil Moisture Accounting model (Burnash et al., 1973) combined with the SNOW-17 snow accumulation and ablation model (Anderson, 1973), with streamflow being routed using a unit-hydrograph model. Hereafter this modeling setup is referred to as SAC. SAC was calibrated using the shuffled complex evolution (SCE, Duan et al., 1992) global optimization routine, minimizing the root mean squared error (RMSE) of the discharge simulations. Simulations started on October 1st 1980 for the 598 basins (out of 671) for which discharge measurements started on or before that date. For the other basins, simulations started on the first October 1st after the start of the discharge records. SAC was calibrated over the first 15 years of the simulation for each catchment, meaning that different periods were used for different catchments. For each catchment, SCE was started from 10 different random seeds, which led to 10 optimized parameter sets. Further details on the hydrometeorological time series are provided in Newman et al. (2015).

The selection of SAC for this study was motivated by the availability of SAC simulations for the CAMELS catchments. We do not claim that SAC is more adapted to these catchments than other hydrological models, and we recognize that using RMSE as an objective function favors the simulation of specific signatures (such as the mean and peak discharge). We think, however, that such a modelling setup constitutes a useful benchmark for the random forest predictions, since hydrological models of similar complexity and calibrated in a similar way are widely used.

2.3 CAMELS catchment attributes

The landscape of each catchment was described using a wide range of attributes, which can be divided into five classes:

1. Topographic characteristics: features such as catchment area and mean elevation, extracted from the United States Geologial Survey (USGS) data base.
2. Climatic indices: indices such as aridity and the frequency of high precipitation events, computed using the Daymet (Thornton et al., 2012) daily time series extracted by Newman et al. (2015).
3. Land cover characteristics: attributes such as the maximum leaf area index and the rooting depth, estimated using MODIS imagery.
4. Soil characteristics: variables such as the soil depth and the sand fraction, extracted from the State Soil Geographic Database (STATSGO, Miller and White, 1998) and from Pelletier et al. (2016).
5. Geological characteristics: characteristics such as the dominant geology class and the subsurface permeability, retrieved from Global Lithological Map (GLiM, Hartmann and Moosdorf, 2012) and GLobal HYdrogeology MaPS (GLHYMPS, Gleeson et al., 2014).

The complete list of catchment attributes, as well as details on the methods and data used to compute them, is provided in Table 1. Note that not all of the CAMELS attributes were used. We excluded the following attributes to avoid redundant information and clarify the result of the statistical analysis: the leaf area index difference and green vegetation fraction difference (both are highly correlated with the leaf area index maximum), soil porosity and conductivity (both are highly correlated with the sand fraction because of their estimation relying on sand fraction) and the second dominant geological class of the GLiM data set (as it is unavailable for 138 catchments, which are entirely covered by a single class). Further details on the data and methods used to compute the catchment attributes and hydrological signatures are provided in Addor et al. (2017b).

2.4 Signature regionalization using random forests and hydrological modeling

To characterize the hydrological behavior of the catchments, we used 15 hydrological signatures computed for the CAMELS data set. Those signatures were selected because they characterize different parts of the hydrograph and they are sensitive to processes occurring over different time scales. They are also commonly employed in the literature, so we used this study as an opportunity to compare them. The signatures we considered are described in Table 2. We computed them using the observed discharge and the discharge simulated using the SCE parameter set leading to the lowest RMSE - we also performed the analysis using the mean of the 10 simulations (one per parameter set), the results did not change significantly. We also predicted these signatures based on catchment attributes using random forests (Section 2.5). We evaluated the signatures simulated by calibrated SAC and predicted by random forests by computing the fraction of variance (R^2) of the observed signatures that they explain. The number of stations used for R^2 computation varies slightly from signature to signature, because in some specific situations, for instance when rivers are dry for significant periods, the signatures cannot be computed. The number of catchments for each signature is however always greater than 600. R^2 is unitless, which enables the direct comparison of different signatures. All the signatures were computed using daily discharge

data scaled by the catchment area.

2.5 Random forests to predict hydrological signatures using catchment attributes

We used random forests to predict hydrological signatures using catchment attributes. Random forests are a machine-learning algorithm relying on a large number of regression trees to produce an ensemble of predictions. They have been successfully used in a various fields of geosciences, for instance to predict hydrological signatures (Snelder et al., 2009; Booker and Woods, 2014) and soil characteristics (Chaney et al., 2016; Hengl et al., 2017). We provide a brief introduction to random forests in Appendix 1. For more detailed information, we refer the reader to Breiman (2001). We developed random forests in R (R Core Team, 2017) using the package randomForest (Liaw and Wiener, 2002). For an introduction to random forests using R, we recommend James et al. (2013).

We selected random forests for the data mining of the CAMELS data set for the following reasons:

1. Random forests allow for multiple predictors and non-linear relationships: It is common to use a single characteristic (typically aridity or the baseflow index) to summarize hydrological behavior and differentiate between catchments. Yet, catchment behavior is never determined by a single attribute, but instead reflects the interplay of numerous attributes. Beck et al. (2015) explored streamflow characteristics for thousands of catchments and concluded that "the individual relationships between catchment attributes and Q characteristics were generally weak, suggesting the need for models incorporating multiple predictors to estimate Q characteristics". Random forests are well-adapted for this task because they allow for multiple predictors, and since they are constructed using a series of thresholds, they are more adapted to capture the non-linear relationships between attributes and hydrological signatures than classical multiple linear regressions.

2. Random forests are not limited by our understanding of catchment behavior: Random forests are a flexible statistical model, which is not constrained by any physical principles or assumptions on hydrological processes. We see it as an advantage, as data exploration using random forest can potentially reveal relationships, which are not commonly acknowledged, although they can be explained a posteriori from a physical perspective.

3. Reduced risk of data overfitting: Random forests are an ensemble of regression trees, which gives them more robustness than individual regression trees. Randomness is introduced when they are constructed so that their predictions are not overly influenced by specific catchments or predictors (Appendix 1).

4. Transparency and interpretability: When producing multi-variable predictions, it is important to be able to assess which predictors have the greatest influence on the response variables. The interpretation of the influence of each predictor in the random forest using IncMSE is straight-forward (IncMSE is the relative increase in the MSE of the prediction when the values of the predictor of interest are shuffled, see Appendix 1).

5. Good performance in prediction mode and reliable uncertainty estimates: Random forests and similar machine-learning techniques (such as neural network, e.g. Beck et al., 2015) can deliver accurate predictions for little computation effort (growing each forest takes a few seconds). Further, each random forest relies on an ensemble of trees, that can be used

to estimate the uncertainty of the prediction (those uncertainty estimates can be very reliable, see Figure A1d).

We argue that these advantages justify the use of random forests in our study. It is however fair to acknowledge that random forests also have drawbacks. Critically, they are highly parameterized, as each regression tree uses on the order of 10 thresholds. In this study, we used 500 trees to predict each of the 15 hydrological signatures, which leads to about 70,000 parameters (thresholds on predictors). This number of parameter is impractical to analyze on an individual basis, but the relative influence of the predictors on each signature can be quantified using the IncMSE (see Appendix 1).

The random forest predictions were evaluated using a ten-fold cross-validation: a random forest was trained using 90% of the basins and its predictions were evaluated using the remaining basins, this procedure was then repeated nine additional times in order to cover all the basins. The results showed hereafter are for the validation phase, not for the training phase. Random forests require all the predictors to be available for each catchment. This restricted our analysis to the 643 catchments for which all the attributes could be quantified. For the slope of the flow duration curve, catchments for which a third or more of the streamflow values were equal to 0 (causing the signature to be undefined) had to be excluded, reducing the number of catchments to 617.

To assess the value of each class of attributes for the prediction of hydrological signatures, we produced the random forests for 5 predictor groups (PGs). We started using only the climatic indices (PG1), then added the topographic indices (PG2), the soil characteristics (PG3), the landcover characteristics (PG4) and finally the geology characteristics (PG5). We started with climatic indices because the data required for their computation are comparatively easy to access: many atmospheric forcing data sets are available, they are well documented and evaluated, and often, catchment-averaged time series can be obtained from previous studies. Further, climatic indices have been shown to be good predictors of streamflow indices (Beck et al., 2015; Kuentz et al., 2017). In contrast, soil, land cover and geological characteristics are usually less readily available at the catchment scale, and their predictive power tends to be weaker. Using these five PGs for the signatures predictions is a way to assess how far we can get using climatic indices only, and to estimate the added-value of the other attribute classes. We consider the predictions based on climatic data only as a baseline, i.e., what we expect the signature to be solely based on climate. When the observed signature significantly departs from this baseline, one explanation is that the signature is significantly influenced by other attributes which were not considered (e.g., soil attributes). Progressively adding attributes and monitoring if they significantly improve the predictions enables us to assess the value of different data sets for signature regionalization, and can provide insights into which hydrological processes drive the signatures.

3   Results

The presentation of the results is organized as follows. We first present spatial maps for a subset of commonly used signatures (mean discharge, slope of the flow duration curve, and the baseflow index), and then we present statistics for the full set of 15 signatures. Finally, we show the influence of individual catchment attributes on random forest predictions of different signatures.

3.1 Simulation, prediction and spatial smoothness of hydrological signatures - introduction

Figure 1 illustrates predictions of three example hydrologic signatures (mean annual discharge, slope of the flow duration curve, and the baseflow index) from both random forests and the SAC model. Mean discharge can be predicted very well by a random forest based on catchment descriptors ($R^2 = 0.92$) and can be also simulated remarkably well by the conceptual hydrological model SAC calibrated by minimizing the RSME ($R^2 = 0.98$). In contrast, the performance of both the random forest and SAC is poor when it comes to the slope of the flow duration curve ($R^2 = 0.29$ and $R^2 = 0.15$, respectively). The baseflow index is predicted (by the random forest) and simulated (by SAC) better than the slope of the flow duration curve, but worse than the mean annual discharge ($R^2 = 0.64$ and $R^2 = 0.84$, respectively). Note that for these three signatures, the performance of the random forest and of SAC are related: both methods perform well for the mean annual discharge, reasonably well for the baseflow index, and poorly for the slope of the flow duration curve.

Interestingly, the performance of both the random forest and SAC is related to the spatial smoothness of the hydrological signatures. Note how the mean discharge field varies smoothly across space, whereas the slope of the flow duration curve exhibits large changes over short distances (first row of Figure 1). To quantify the spatial smoothness, we used Moran's I to measure the spatial auto-correlation (Appendix 2). I enables us to quantify features that are clear visually, and to compare signatures based on the spatial smoothness of their field. The spatial smoothness is the highest for the mean discharge ($I = 0.51$), intermediate for the baseflow index ($I = 0.16$) and the lowest for the slope of the flow duration curve ($I = 0.09$). This ranking is the same as the ranking based on the performance of the random forest and SAC. In other words, Figure 1 suggests that signatures with lower spatial smoothness may be harder to relate to catchment characteristics and to simulate using our modelling setup.

3.2 Simulation, prediction and spatial smoothness of hydrological signatures - evaluation for 15 signatures

Figure 2 shows that there is a strong three-way relationship between how well signatures can be predicted based on catchment attributes, how well they can be simulated by SAC, and the smoothness of their spatial variability over the CONUS. The signatures in Figure 2 are ordered from left to right based on how well they can be predicted using a random forest. Like for Figure 1, we compared the observed and predicted signatures from the random forest by computing the coefficient of determination R^2, shown in light blue in Figure 2. R^2 varies from 0.92 (mean annual discharge) to 0.29 (slope of the flow duration curve). The performance of the random forest is compared to that of SAC, shown in dark blue in Figure 2. It is clear that hydrological signatures that can be accurately predicted from catchment attributes by the random forest can also be well simulated by SAC. Indeed, the performance of the random forest and that of SAC, each described by 15 R^2 values, are highly correlated ($\rho = 0.90$). Note that several signatures we considered were also predicted by Beck et al. (2015) using characteristics from thousands of catchments from across the world and neural networks. They also find that some signatures are better predicted than others and interestingly, it appears that if they had ranked signatures based on the R^2 they report in their Figure 5, the ranking would have been very similar to what we propose (with the mean annual flow and half-flow date being best predicted, followed by the high-flow quantile, and finally the low flow quantile and the baseflow index). Note that the hydrological simulations

stem from a single model structure calibrated using a single objective function, and the results for each signature may be different for another modeling setup.

The random forest predictions of the mean annual discharge, mean winter discharge, mean half-flow date, Q95 and runoff ratio are all very good, with $R^2 > 0.8$. They are almost as good as the simulations produced by the SAC, although SAC was calibrated for each basin individually while the same random forest is used for the entire country. The baseflow index and Q5 are satisfactory predicted ($R^2 > 0.6$) but are better captured by SAC. For the other signatures, the predictions of the random forests are worse. SAC performs particularly poorly for the mean duration of low-flow events, the slope of the flow duration curve and the no flow frequency, reflecting that using a general metric such as RMSE can deliver a good overall performance according this specific metric, but does not provide enough constrains to capture specific parts of the hydrograph defining catchment behavior (De Boer-Euser et al., 2017). For these metrics, the random forests perform better than SAC, but can only explain a limited fraction of the observed variability.

Furthermore, the spatial smoothness measured by Moran's *I* (shown in green in Figure 2, see Fig. S1 for the Moran's *I* computed for each signature) is almost systematically greater for signatures that can be accurately predicted by the random forest and well simulated by SAC. In fact, the correlation between the performance of the random forest and spatial smoothness is strong ($\rho = 0.90$). This suggests that random forests fail to capture sudden (small-scale) changes in hydrological signatures over short distances. The spatial smoothness also appears to be a good predictor of how well hydrological signatures are captured by SAC ($\rho = 0.78$). The relationship between spatial smoothness and predictability is discussed in Section 4.2.

3.3 Strong and weak predictors of hydrological signatures

Recall from Figures 1 and 2 that hydrological signatures well predicted by random forests tend to have a smooth pattern. This can be explained by the strength of the climate signal: climatic indices have a smooth pattern over the CONUS, and when they are highly correlated to signatures, those signatures inherit their smooth pattern. This is clear in Figure 3: the spatial patterns of climate indices shown in the first row (originally selected by Berghuijs et al., 2014) are similar to the signatures in the second row. The maps of mean annual discharge and the runoff ratio show very similar patterns to that of the aridity map, while the half-flow date principally reflects the precipitation seasonality and the fraction of precipitation falling as snow. In contrast, the maps in the bottom row of poorly predicted signatures show a noisier spatial pattern and lack a clear relationship to the climatic indices shown in the first row.

To better understand why some signatures were better predicted than others, we explored which predictors were preferentially used by the random forest. To this end, we consider the IncMSE, the increase in the MSE of the prediction when the value for each predictor were shuffled. IncMSE is indicated by the size of the dots in Figure 4. The color of the dots indicates the Spearman rank correlation coefficient between each attribute and signature. Most of the influential predictors in the random forest are climatic variables. If we restrict attention to the 14 pairs of catchment attributes-hydrological signatures with IncMSE > 20%, 11 of them involve a climatic variable (aridity alone accounts for 6 pairs). In this respect, the climatic indices exert a stronger influence on hydrological signatures than the topographic, soil, land cover and geological attributes combined.

We found that climatic indices have by far the greatest influence on selected hydrological signatures, while the attributes characterizing the land cover, soil, geology and topography have a much weaker influence. The lack of dark colors in the corresponding columns of Figure 4 indicate that those attributes, when considered individually, are not strongly correlated to hydrological signatures. Even when those attributes are combined with other attributes using a random forest, their influence, beyond the influence that is already captured indirectly by climate indices, is generally insignificant, as shown by the lack of the large circles in the same columns. The relative strength of climatic variables when compared to other catchment attributes has the following implication. When a hydrological signature is strongly linked to one or several climate indices, it is well predicted, and conversely, weak links lead to poor predictions. Hence, climatic attributes strongly condition how well hydrological signatures can be predicted by the random forest. Some signatures like the slope of the flow duration curve are not well constrained by climate variables, and the random forest is not able to extract relevant information from the predictors we are using.

Figure 5 summarizes the value of the five classes of attributes for the regionalization of signatures using random forests. For most signatures, the predictions using climatic indices only (PG1) are only marginally poorer than the predictions using all the attributes (PG5), meaning that the value of topographic, soil, land cover and geological attributes is low. For well-predicted signatures (left part of Figure 5), using climatic indices only already delivers good predictions, and for poorly predicted signatures (right part of Figure 5), adding non-climatic attributes (PG2 to PG5) once climatic indices have been considered is not enough to significantly improve the predictions. The baseflow index is the signature benefiting most from the addition of non-climatic attributes, yet none of these attribute classes contribute as much to the predictions as the climatic attributes.

4   Discussion

4.1   How to improve the predictability of signatures not directly related to climatic indices?

The key influence of climatic conditions on hydrological behavior is not new. Aridity is commonly regarded as the main driver of water partitioning at the land surface (Budyko, 1974; Hrachowitz et al., 2013). The influence of climate on hydrological regimes (Berghuijs et al., 2014) and the water balance (Padrón et al., 2017) is well acknowledged, yet it is debated whether this influence is direct, via the water balance, or indirect, via the long-term influence of climate on the landscape (Harman and Troch, 2014). Importantly, climatic variables do not only drive current, but probably also trends induced by climate change (Rice et al., 2016). Figure 4 reflects the control aridity exerts on the water balance, yet it also reveals that several hydrological variables, which reflect key aspects of hydrological dynamics, are poorly predicted by aridity alone, or even by a combination of several climatic indices. For instance, random forests were unable to clearly relate climate indices to the precipitation-streamflow elasticity, the slope of the flow duration curve or the no-flow frequency. In other words, the number of hydrological signatures that can be well predicted based on climatic indices alone is limited.

Interestingly, signatures for which climatic indices are weak predictors do not have other strong predictors. In particular, the land cover, soil, geology are, for most variables, poor predictors, their added value is low (Figure 5). These results are consistent with those of Beck et al. (2015), who predicted a range of hydrological signatures using catchment attributes and reported that climate indices exerted the strongest influence, while predictors related to soils and geology were less important. Merz and Bloschl (2009) similarly showed that event runoff coefficients in 459 Austrian catchments were barely influenced by land cover, soil types, and geology, and were better explained by climate-related indices. In contrast, when exploring and classifying 116 near-natural catchments in the UK, Chiverton et al. (2015) found that geology, the depth to gleyed layer in soils and the percentage of arable land were good discriminants. Likewise, Singh et al. (2014) found geology and land use do matter when choosing donor catchments, but their influence depend on the region.

We are concluding that non-climatic drivers are irrelevant, but rather that it challenging to capture their influence on hydrological signatures at the catchment scale. The influence of land cover, soil and geology attributes can be missed by the random forests for several reasons. Firstly, there is currently a lack of data uncertainty estimates in large-sample hydrological data sets, which comes in part from the lack of uncertainty estimates in standard products (exceptions being SoilGrids (Hengl et al., 2017) and POLARIS (Chaney et al., 2016), currently not used for CAMELS). Data quality has been brought up to explain why soil and geological data are not good predictors of hydrological signatures (Beck et al., 2015). It is indeed likely that issues related to data collection (see discussion in Addor et al., 2017b) limit the predictive power of soil data. Accounting for the uncertainty of soil characteristics may for instance make the influence of soils on water dynamics clearer and improve the predictions. This highlights the urgent need for uncertainty estimates to be provided with time series (e.g., streamflow observations, see Section 4.1) and landscape attributes (e.g., land cover and geological variables), and to be included in large-sample hydrological data sets. Secondly, the scale on which vegetation, soil, geological processes occur is several orders of magnitude smaller than what our finest data sets or models can capture. Key properties are difficult to upscale in a way that preserves their influence on water dynamics, which stresses the importance of upscaling methods capture landscape properties across scales (Samaniego et al., 2010; Rakovec et al., 2016). Note that in CAMELS all attributes are basin averages, and the heterogeneity within each catchment is not considered. Thirdly, we have not included predictors that have been shown to influence catchment behavior. For instance, we are not considering attributes characterizing the network or the shape of the catchment, because their computation and validation for the CAMELS catchment is still ongoing. Finally, predictions could be improved by training the random forests over smaller regions (Nearing et al., 2016; Kuentz et al., 2017), but this would be done at the expense of the generality of the statistical model.

4.2 Poorly predicted signatures are particularly variable in space. Is there a link?

Using a large sample of catchment enabled us to consider the smoothness of the spatial field of signatures and to quantify it using Moran's *I*. We showed that signatures with a noisy pattern tend to be poorly predicted by random forests and poorly simulated by SAC. We are not aware of other studies quantifying the spatial smoothens of signatures and showing that it is related to their predictability. As discussed above, signatures with a smooth pattern tend to be highly correlated with climatic indices (Figure 3). We propose that signature with a noisy pattern tend to i) be particularly affected by errors in observed signatures and ii) result from competing processes.

In this study, we do not explicitly characterize errors in discharge time series resulting from rating curve uncertainties, nor how those uncertainties propagate into hydrological signatures. These aspects were however investigated by Westerberg et al. (2016) for 43 UK catchments. They report that some signatures, such as the mean discharge, are far less sensitive to rating curve uncertainty than others, such as the slope of the flow duration curve (as illustrated by their Figure 6). Similarly, low flow signatures are more sensitive to data errors than high flow signatures. Here we show that the signatures they identified as sensitive to rating curve uncertainty tend to vary abruptly over short distances (low Moran's *I*, Figure 2). This suggests that part of the spatial variability is noise, i.e. stems from variations caused by the data collection and processing (formulation of the signature) and do not reflect differences in the hydrological behavior of the catchments. These errors in streamflow measurements impact the evaluation of the SAC simulations and the random forest predictions, which contributes to explain why both methods encounter similar difficulties in capturing the observed signatures and why the ranking of signatures using each method is so similar (Figure 2).

It is noteworthy that signatures varying abruptly in space have been shown to be difficult to regionalize. Westerberg et al. (2016) relied on a weighted-pooling-group approach, in which each signature was estimated using the weighted mean of its value in similar catchments. Their regionalization performs better for high flows than for low flows, and better for the mean discharge than for the slope of the flow duration curve (their Figure 8). This is not only consistent with the sensitivity of the signatures to rating curve uncertainties they determined, but also the spatial smoothness of their field. Signatures with a smooth field would be well regionalized when selecting the closest catchments as donors. It is likely that the sudden variations over space for some signatures, which we argue make regionalization difficult, come in part from discharge uncertainties, as discussed above.

Noisy spatial patterns are not only caused by errors in streamflow measurements. An indication of that, is that signatures computed using SAC simulations can also vary significantly over short distances (see e.g., Figure 1k), although they are free of streamflow measurement errors. This suggests that instead of being driven by a limited number of first order processes (e.g., water balance prescribed by aridity, or snow accumulation/melt), the signature is the result of a myriad of processes interacting in complex ways and difficult to disentangle (think about the range of processes influencing the slope of the flow duration curve). These interactions can lead to sudden changes in space, and it is not surprising that they are not satisfactorily captured by a random forest not accounting for any hydrological process, or by a hydrological model calibrated by optimizing an objective function that does not account for the internal consistency of the system. As a consequence of this diversity of processes, it is difficult to establish clear links between landscape attributes and hydrological signatures.

4.3   Many signatures are poorly predictable. How well do we understand them?

Part of our motivation to assess the spatial predictability of signatures comes from the idea that "the ability to accurately predict behavior is a severe test of the adequacy of knowledge in any subject" put forward by Crawford and Linsley (1966). Our study showed that many signatures commonly used in hydrological applications are poorly predictable based on catchment attributes. This makes it difficult to explain why these signatures vary in space and makes us wonder how well we understand them. To give one example, the precipitation-discharge elasticity is commonly used to anticipate the future impacts of climate change on

discharge, yet even recent research recognizes that "it is difficult to identify physical reasons, for the spatial variations in elasticity values" (Andréassian et al., 2016). We believe that a better understanding of the drivers of elasticity would improve the reliability of the projections and would be useful to assess whether climate and hydrological models correctly capture the impacts of changes in precipitation on streamflow (Vano et al., 2015).

Another way to approach the lack of predictability of signatures, and the implications for signature interpretation, is to consider the noise in maps of signatures. For instance, the slope of the flow duration curve is poorly predictable (Figure 2), and is also known to be particularly sensitive to streamflow uncertainties (Westerberg et al., 2016). A crucial question is to which extent the abrupt changes between two neighboring catchments (see Figure 1i) truly reflect hydrological differences between these catchments, and to which extent they are artifacts resulting from data errors (Kennard et al., 2011). Although the slope of the flow duration curve is commonly used in catchment classification (Sawicz et al., 2011) and model evaluation studies (Euser et al., 2013), its discriminatory power is debated (McMillan et al., 2017), and it is our impression that the community would benefit from a better understanding of what it reflects.

We are not advocating against the use of the poorly predictable signatures. In contrast, we think they deserve more attention. Low flow metrics (frequency of low flow, persistence of low flow, slope of FDC) are not well captured by our setup, but it is possible that some studies may choose to focus on these signatures specifically (e.g., if their particular focus was on ecology, or if their aim was to improve rainfall-runoff structures by focusing on simulation deficiencies). Our study illustrates that these signatures are less constrained by landscape characteristics than other signatures, and that they a more challenging to explain, so there is a need for a better understanding of the processes driving them. Yet these signatures have also been shown to be more sensitive to streamflow uncertainty, so for progress to be made, these uncertainties have to be characterized and accounted for.

5 Conclusions and outlook

We systematically explored how landscape attributes influence (or not) hydrological signatures. We described the landscape of 671 catchments in the contiguous USA using five classes of attributes (topography, climatology, land cover, soil and geology) and summarized catchment behaviour using 15 hydrological signatures. Random forests allowed us to combine these landscape characteristics in non-linear ways and to quantitatively explore their relative influence on hydrological signatures. We found that climatic attributes are by far the most influential predictors for signatures that can be well-predicted based on catchment attributes (such as the mean annual discharge or the half-flow date), with land cover, soil and geology attributes playing secondary roles. In contrast, several other signatures, such as the slope of the flow duration curve or the streamflow-precipitation elasticity are poorly predicted based on catchments attributes, and in particular, could not be satisfactorily predicted by climatic indices alone.

Using a large sample of catchments enabled us to explore the spatial patterns of hydrological signatures over the CONUS, and to characterize their spatial smoothness (auto-correlation) using Moran's *I*. We found that spatial smoothness is a simple yet powerful way to gain insights into a variety of aspects of large-sample studies. Signatures with smooth spatial variations are typically those with a high spatial predictability. In contrast, when signatures exhibit abrupt changes over short distances, those changes usually cannot be related to

catchment attributes using random forests and they are also poorly captured by hydrological simulations from a conceptual model. These sudden variations make signature regionalization difficult if neighbouring catchments are used as donors.

In summary, we found strong relationships between i) our ability to capture hydrological signatures using simulations from a conceptual hydrological model (SAC), ii) our ability to predict them using catchment characteristics as predictors in a machine-learning algorithm (random forests), iii) the spatial smoothness of the maps of these signatures and iv) the strength of the climate influence on those signatures. The strong consistency between these four aspects enabled us to rank hydrological signatures. Signatures poorly related to catchment attributes, are also poorly captured by SAC, their spatial pattern is noisy, and based on results from other studies, they are particularly susceptible to discharge uncertainties and difficult to regionalize. We propose these poorly predicted signatures deserve more attention, as signatures related to the water balance are already well explained by climatic variables. This relies on an improved understanding of their drivers, a better exploitation of the value of non-climatic attributes (such as soil, land cover and geology) and a more systematic characterization of the uncertainties in both signatures and catchment attributes.

Appendix 1: An introduction to regression trees and random forests

We chose to use a machine-learning tool (random forests, Breiman, 2001) to explore how the interplay between landscape attributes shapes hydrological behavior. Machine-learning algorithms are gaining in popularity as the quantity and diversity of data to process increase. Machine-learning algorithms have been shown to be powerful prediction techniques, including in hydrologic studies (e.g., Gudmundsson and Seneviratne, 2013; Beck et al., 2015). Here we present a brief introduction to random forests, which may be useful for the interpretation of our results.

A random forest relies on an ensemble of regression trees to relate predictors (here catchment attributes) to a response variable (here a hydrological signature). In a regression tree, the prediction is made based on a series of threshold-based conditions on the predictors. The prediction scheme is initiated at the top of the tree (in the example shown in Figure A1a, the question at the top split is whether the mean elevation is greater than 1151m). The prediction is then refined using other thresholds on other (and sometimes the same) predictors at lower levels of the tree. The influence of each predictor on the response variable can be estimated based on its position in the regression tree: predictors appearing higher in the tree have a higher separating/predictive power (Figure A1a indicates that mean elevation is a strong predictor of the base flow index, likely because it conditions the formation a snow pack, which will increase the baseflow index when it melts). Note that regression trees are typically not symmetrical (different variables are used in different parts of the tree).

Regression trees are grown following a "recursive binary splitting" approach. The procedure starts at the top of the tree and at each split, one variable and one threshold are selected in order to minimize the mean squared error (MSE) of the prediction. The prediction is the mean value of the predictor for all the elements (catchments) falling in each class. As a consequence, the predictions of a decision tree are discrete values (one per terminal node, such as 0.4801 for the left-most terminal node of the tree shown in Figure A1a, which leads to the horizontally aligned back points in Figure A1b). Trees are grown and then pruned by minimizing the cross-validated MSE in order to reduce the risk of overfitting. While

regression trees are intuitive to interpret and can deal with non-linear relationships between variables, they typically lack robustness. We found that regression trees produced by randomly excluding half of the catchments to be quite different in the predictive variables they selected and in the position of these variables in the tree.

To overcome this limitation, we used random forests instead of single regression trees. Random forests are an ensemble of regression trees (here we used 500 trees per forest). The robustness of the forest comes from the way each tree is grown. At each split, a subsample of predictors is randomly excluded and the prediction must be done using solely the remaining. This implies that strong predictors, which otherwise might have been used for this specific split, will be excluded. This introduces differences between the trees, making the prediction more robust than if all the trees were similar. The number of trees N and the number of predictors P excluded at each split are variables defined by the user. We found that variations around the default value for P (a third of the total number of predictors) has little influence on our predictions, and that $N = 500$ is adequate because it leads to better predictions than small forests, but more trees did not improve the predictions.

Since it is not practical to inspect each tree to determine which variables are used for the prediction, the relative influence of the predictors of a random forest is measured in an automated way. Once the forest has been grown, each predictor is considered individually and its values are shuffled (their statistical distribution remains the same but their order is now random). The relative drop in prediction accuracy (expressed in %) indicates how influential this predictor is (large increases in MSE indicate influential predictors). Figure A1c shows that for the prediction of the baseflow by a random forest, the fraction of precipitation falling as snow is the most influential predictor.

An advantage of growing a random forest is that the ensemble of trees can be used to characterize the uncertainty in the prediction. We used QQ plots to assess the reliability of the ensembles and found that for all the hydrological signatures except the fraction of no flow, the ensembles are remarkably reliable (Figure A1d). Although this is not a feature we use in this study, we consider important to stress this finding, as it can be relevant in other contexts, for instance for parameter estimation based on regionalized hydrological signatures. Finally, note that because the deterministic prediction of each random forest is the mean prediction of its regression trees, the predictions are continuous values. This reduces the granularity of the predictions when compared to regression trees, which only predict a limited number of discrete values (Figure A1b).

Appendix 2: Moran's *I* as a measure of spatial smoothness

When a variable is plotted on a map for numerous catchments, spatial patterns can appear and help with the formulation of starting hydrological hypotheses. A fundamental advantage of large-sample hydrology over small-sample hydrology is that, when maps are produced using hundreds of catchments, those insights are likely to be clearer than if the maps were based on a handful of catchments, because those tend to be patchier.

In this study, we explore and quantify regional variability in hydrological signatures using a measure of spatial smoothness. Addor et al. (2017b) observed that maps of climate indices generally exhibit smoother patterns than maps of hydrological signatures, whose patterns tend to be noisier (with potentially strong differences between adjacent catchments). Similar differences in spatial variability can also be observed among hydrological signatures: some signatures vary gradually across the landscape, while others exhibit abrupt changes over short distances. This is already apparent in earlier studies. Figure 2 of Sawicz et al. (2011) indicates for instance that the runoff ratio over the Eastern United States varies more smoothly in space than the slope of the flow duration curve.

To quantify the smoothness of spatial patterns in maps of hydrological signatures, we measure the spatial autocorrelation using Moran's I (Moran, 1950; Legendre and Legendre, 1998):

$$I = \frac{\frac{1}{W} \sum_{i=1}^{N} \sum_{j=1}^{N} w_{i,j} (x_i - \overline{x})(x_j - \overline{x})}{\frac{1}{N} \sum_{i=1}^{N} (x_i - \overline{x})^2}$$

where x is the variable of interest with N elements (here $N = 671$ catchments), $\overline{x}$ is its mean, w is the weight associated with each pair of catchments (here $w = 1/d$, where d is the distance along a great circle between the two catchments, the diagonal elements of the matrix w being set to 0) and W is the sum of all the weights. Spatial correlation can be related to temporal autocorrelation: if all the pairs of data points close in space (in time) have a similar value, then the field is spatially (temporally) auto-correlated. Differences (or similarities) between points far apart have a comparatively small influence on I because of the distance-based weighting system selected. I values close to 0 indicate no spatial correlation. The higher the value I, the greater the spatial auto-correlation and the smother the spatial patterns (compare Figures 1a, e and i for an example). Note that in contrast to correlation coefficients, $|I|$ can exceed 1 (de Jong et al., 1984).

Note that the random forests in this study do not directly account for spatial structure, since we are not using the basins spatial coordinates as predictors. In other words, the spatial proximity of the basins cannot be used by the random forests, and nor can the spatial smoothness of any field.

Acknowledgements

References

Addor, N., Newman, A. J., Mizukami, N. and Clark, M. P.: Catchment attributes for large-sample studies, , doi:10.5065/D6G73C3Q, 2017a.

Addor, N., Newman, A. J., Mizukami, N. and Clark, M. P.: The CAMELS data set: catchment attributes and meteorology for large-sample studies, Hydrol. Earth Syst. Sci., 21(10), 5293–5313, doi:10.5194/hess-21-5293-2017, 2017b.

Anderson, E. A.: National Weather Service River Forecast System - Snow accumulation and ablation model, Tech. Memo. NWS HYDRO-17., 1973.

Andréassian, V., Coron, L., Lerat, J. and Le Moine, N.: Climate elasticity of streamflow revisited - An elasticity index based on long-term hydrometeorological records, Hydrol. Earth Syst. Sci., 20(11), 4503–4524, doi:10.5194/hess-20-4503-2016, 2016.

Andréassian, V., Hall, A., Schaake, J. and Chahinian, N., Eds.: Large sample basin experiments for hydrological model parameterization, International Association of Hydrological Sciences., 2006.

Beck, H. E., Van Dijk, A. I. J. M., Miralles, D. G., De Jeu, R. A. M., Bruijnzeel, L. A., McVicar, T. R. and Schellekens, J.: Global patterns in base flow index and recession based on streamflow observations from 3394 catchments, Water Resour. Res., 49, 7843–7863, doi:10.1002/2013WR013918, 2013.

Beck, H. E., de Roo, A. and van Dijk, A. I. J. M.: Global maps of streamflow characteristics based on observations from several thousand catchments, J. Hydrometeorol., 16, 1478–1501, doi:10.1175/JHM-D-14-0155.1, 2015.

Berghuijs, W. R., Sivapalan, M., Woods, R. A. and Savenije, H. H. G.: Patterns of similarity of seasonal water balances: A window into streamflow variability over a range of time scales, Water Resour. Res., 50, 5638–5661, doi:10.1002/2014WR015692, 2014.

Best, M. J., Abramowitz, G., Johnson, H. R., Pitman, A. J., Balsamo, G., Boone, A., Cuntz, M., Decharme, B., Dirmeyer, P. a., Dong, J., Ek, M., Guo, Z., Haverd, V., van den Hurk, B. J. ., Nearing, G. S., Pak, B., Peters-Lidard, C., Santanello, J. a., Stevens, L. and Vuichard, N.: The plumbing of land surface models: benchmarking model performance, J. Hydrometeorol., 16, 1425–1442, doi:10.1175/JHM-D-14-0158.1, 2015.

De Boer-Euser, T., Bouaziz, L., De Niel, J., Brauer, C., Dewals, B., Drogue, G., Fenicia, F., Grelier, B., Nossent, J., Pereira, F., Savenije, H., Thirel, G. and Willems, P.: Looking beyond general metrics for model comparison - Lessons from an international model intercomparison study, Hydrol. Earth Syst. Sci., 21, 423–440, doi:10.5194/hess-21-423-2017, 2017.

Booker, D. J. and Woods, R. A.: Comparing and combining physically-based and empirically-based approaches for estimating the hydrology of ungauged catchments, J. Hydrol., 508, 227–239, doi:10.1016/j.jhydrol.2013.11.007, 2014.

Breiman, L.: Random forests, Mach. Learn., 45(1), 5–32, doi:10.1023/A:1010933404324, 2001.

Budyko, M. I.: Climate and life, New York Academic Press., 1974.

Burnash, R. J. C., Ferral, R. L. and McGuire, R. A.: A generalized streamflow simulation system: Conceptual modeling for digital computers., 1973.

Burns, D. H.: Evaluation of regional flood frequency analysis with a region of influence approach, Water Resour. Res., 26(10), 2257–2266, doi:10.1029/90WR01192, 1990.

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, doi:10.1016/j.geoderma.2016.03.025, 2016.

Chiverton, A., Hannaford, J., Holman, I., Corstanje, R., Prudhomme, C., Bloomfield, J. and Hess, T. M.: Which catchment characteristics control the temporal dependence structure of daily river flows?, Hydrol. Process., 29(6), 1353–1369, doi:10.1002/hyp.10252, 2015.

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(4), 523–543, doi:10.1002/hyp.7902, 2011.

Clausen, B. and Biggs, B. J. F.: Flow variables for ecological studies in temperate streams: Groupings based on covariance, J. Hydrol., 237(3–4), 184–197, doi:10.1016/S0022-1694(00)00306-1, 2000.

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 Resour. Res., 20(6), 682–690, doi:10.1029/WR020i006p00682, 1984.

Court, A.: Measures of streamflow timing, J. Geophys. Res., 67(11), 4335–4339, doi:10.1029/JZ067i011p04335, 1962.

Crawford, N. H. and Linsley, R. K.: Digital simulation in hydrology: Stanford Watershed Model IV., 1966.

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

Euser, T., Winsemius, H. C., Hrachowitz, M., Fenicia, F., Uhlenbrook, S. and Savenije, H. H. G.: A framework to assess the realism of model structures using hydrological signatures, Hydrol. Earth Syst. Sci., 17(5), 1893–1912, doi:10.5194/hess-17-1893-2013, 2013.

Falcone, J. A.: GAGES-II: Geospatial Attributes of Gages for Evaluating Streamflow. [online] Available from: https://water.usgs.gov/GIS/metadata/usgswrd/XML/gagesII_Sept2011.xml, 2011.

Gleeson, T., Moosdorf, N., Hartmann, J. and van Beek, L. P. H.: A glimpse beneath earth's surface: GLobal HYdrogeology MaPS (GLHYMPS) of permeability and porosity, Geophys. Res. Lett., 41, 3891–3898, doi:10.1002/2014GL059856, 2014.

Gudmundsson, L. and Seneviratne, S. I.: Do land parameters matter in large-scale terrestrial water dynamics? – Toward new paradigms in modelling strategies, Hydrol. Earth Syst. Sci. Discuss., 10(11), 13191–13229, doi:10.5194/hessd-10-13191-2013, 2013.

Gupta, H. V., Wagener, T. and Liu1, Y.: Reconciling theory with observations: elements of a diagnostic approach to model evaluation, Hydrol. Process., 22, 3802–3813, doi:10.1002/hyp.6989, 2008.

Harman, C. and Troch, P. A.: What makes Darwinian hydrology "darwinian"? Asking a different kind of question about landscapes, Hydrol. Earth Syst. Sci., 18(2), 417–433, doi:10.5194/hess-18-417-2014, 2014.

Hartmann, J. and Moosdorf, N.: The new global lithological map database GLiM: A representation of rock properties at the Earth surface, Geochemistry, Geophys. Geosystems, 13(12), 1–37, doi:10.1029/2012GC004370, 2012.

He, Y., Bárdossy, A. and Zehe, E.: A review of regionalisation for continuous streamflow simulation, Hydrol. Earth Syst. Sci., 15(11), 3539–3553, doi:10.5194/hess-15-3539-2011, 2011.

Hengl, T., Mendes De Jesus, J., Heuvelink, G. B. M., Gonzalez, M. R., Kilibarda, M., Blagotí, 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(2), e0169748, doi:10.1371/journal.pone.0169748, 2017.

Holmes, M. G. R., Young, A. R., Gustard, A. and Grew, R.: A region of influence approach to predicting flow duration curves within ungauged catchments, Hydrol. Earth Syst. Sci., 6(4), 721–731, doi:10.5194/hess-6-721-2002, 2002.

Hrachowitz, M., Fovet, O., Ruiz, L., Euser, T., Gharari, S., Nijzink, R., Freer, J., Savenije, H. H. G. and Gascuel-Odoux, C.: Process consistency in models: The importance of system

signatures, expert knowledge, and process complexity, , 7206–7230, doi:10.1002/2013WR014956.Received, 2014.

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, Hydrol. Sci. J., 58(6), 1198–1255, doi:10.1080/02626667.2013.803183, 2013.

James, G., Witten, D., Hastie, T. and Tibshirani, R.: An Introduction to Statistical Learning with Applications in R, edited by Springer., 2013.

de Jong, P., Sprenger, C. and van Veen, F.: On Extreme Values of Moran's I and Geary's c, Geogr. Anal., 16(1), 17–24, doi:10.1111/j.1538-4632.1984.tb00797.x, 1984.

Kennard, M. J., Mackay, S. J., Pusey, B. J., Olden, J. D. and Marsh, N.: Quantifying uncertainty in estimation of hydrologic metrics for ecohydrological studies, Limnetica, 26, 137–156, doi:10.1002/rra, 2011.

Kuentz, A., Arheimer, B., Hundecha, Y. and Wagener, T.: Understanding hydrologic variability across Europe through catchment classification, Hydrol. Earth Syst. Sci., 21, 2863–2879, doi:10.5194/hess-21-2863-2017, 2017.

Lacey, G. C. and Grayson, R. B.: Relating baseflow to catchment properties in south-eastern Australia, J. Hydrol., 204(1–4), 231–250, doi:10.1016/S0022-1694(97)00124-8, 1998.

Ladson, A., Brown, R., Neal, B. and Nathan, R.: A standard approach to baseflow separation using the Lyne and Hollick filter, Aust. J. Water Resour., 17(1), 25–34, doi:http://dx.doi.org/10.7158/W12-028.2013.17.1., 2013.

Laio, F. and Tamea, S.: Verification tools for probabilistic forecasts of continuous hydrological variables, Hydrol. Earth Syst. Sci., 11(4), 1267–1277, doi:10.5194/hess-11-1267-2007, 2007.

Lawrence, D. M. and Slater, A. G.: Incorporating organic soil into a global climate model, Clim. Dyn., 30(2–3), 145–160, doi:10.1007/s00382-007-0278-1, 2008.

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

Liaw, A. and Wiener, M.: Classification and Regression by randomForest, R News, 2(3), 18–22, 2002.

McMillan, H. K., Clark, M. P., Bowden, W. B., Duncan, M. and Woods, R. A.: Hydrological field data from a modeller's perspective: Part 1. Diagnostic tests for model structure, Hydrol. Process., 25(4), 511–522, doi:10.1002/hyp.7841, 2011.

McMillan, H., Westerberg, I. and Branger, F.: Five guidelines for selecting hydrological signatures, Hydrol. Process., 1–5, doi:10.1002/hyp.11300, 2017.

Miller, D. A. and White, R. A.: A Conterminous United States Multilayer Soil Characteristics Dataset for Regional Climate and Hydrology Modeling, Earth Interact., 2(2), doi:10.1175/1087-3562(1998)002<0002:CUSMS>2.0.CO;2, 1998.

Moran, P. A. P.: Notes on Continuous Stochastic Phenomena, Biometrika, 37(1/2), 17, doi:10.2307/2332142, 1950.

Nathan, R. J. and McMahon, T. A.: Estimating low flow characteristics in ungauged catchments, Water Resour. Manag., 6(2), 85–100, doi:10.1007/BF00872205, 1992.

Nearing, G. S., Mocko, D. M., Peters-Lidard, C. D., Kumar, S. V. and Xia, Y.: Benchmarking NLDAS-2 Soil Moisture and Evapotranspiration to Separate Uncertainty Contributions, J. Hydrometeorol., 17(3), 745–759, doi:10.1175/JHM-D-15-0063.1, 2016.

Newman, A. J., Clark, M. P., Sampson, K., Wood, A., Hay, L. E., Bock, A., Viger, R., Blodgett, D., Brekke, L., Arnold, J. R., Hopson, T. and Duan, Q.: Development of a large-sample watershed-scale hydrometeorological dataset for the contiguous USA: dataset characteristics and assessment of regional variability in hydrologic model performance,

Hydrol. Earth Syst. Sci., 19, 209–223, doi:10.5194/hess-19-209-2015, 2015.

Newman, A. J., Mizukami, N., Clark, M. P., Wood, A. W., Nijssen, B. and Nearing, G.: Benchmarking of a physically based hydrologic model, J. Hydrometeorol., JHM-D-16-0284.1, doi:10.1175/JHM-D-16-0284.1, 2017.

Newman, A. J., Sampson, K., Clark, M. P., Bock, A., Viger, R. J. and Blodgett, D.: A large sample watershed-scale hydrometeorological dataset for the contiguous USA, , doi:10.5065/D6MW2F4D, 2014.

Olden, J. D. and Poff, N. L.: Redundancy and the choice of hydrologic indices for characterizing streamflow regimes, River Res. Appl., 19(2), 101–121, doi:10.1002/rra.700, 2003.

Padrón, R. S., Gudmundsson, L., Greve, P. and Seneviratne, S. I.: Large-Scale Controls of the Surface Water Balance Over Land: Insights From a Systematic Review and Meta-Analysis, Water Resour. Res., 1–20, doi:10.1002/2017WR021215, 2017.

Pelletier, J. D., Patrick D. Broxton, Hazenberg, P., Zeng, X., Troch, P. A., Niu, G.-Y., Williams, Z., Brunke, M. A. and Gochis, D.: A gridded global data set of soil, intact regolith, and sedimentary deposit thicknesses for regional and global land surface modeling, J. Adv. Model. Earth Syst., 8, doi:10.1002/2015MS000526, 2016.

Priestley, C. H. B. and Taylor, R. J.: On the Assessment of Surface Heat Flux and Evaporation Using Large-Scale Parameters, Mon. Weather Rev., 100(February), 81–92, doi:10.1175/1520-0493(1972)100<0081:OTAOSH>2.3.CO;2, 1972.

R Core Team: R: A Language and Environment for Statistical Computing, [online] Available from: http://www.r-project.org/, 2017.

Ragettli, S., Zhou, J., Wang, H., Liu, C. and Guo, L.: Modeling flash floods in ungauged mountain catchments of China: A decision tree learning approach for parameter regionalization, J. Hydrol., 555, 330–346, doi:10.1016/j.jhydrol.2017.10.031, 2017.

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, J. Hydrometeorol., 17(1), 287–307, doi:10.1175/JHM-D-15-0054.1, 2016.

Renard, B., Kavetski, D., Kuczera, G., Thyer, M. and Franks, S. W.: Understanding predictive uncertainty in hydrologic modeling: The challenge of identifying input and structural errors, Water Resour. Res., 46(5), 1–22, doi:10.1029/2009WR008328, 2010.

Rice, J. S., Emanuel, R. E. and Vose, J. M.: The influence of watershed characteristics on spatial patterns of trends in annual scale streamflow variability in the continental U.S., J. Hydrol., 540, 850–860, doi:10.1016/j.jhydrol.2016.07.006, 2016.

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

Sawicz, K., Wagener, T., Sivapalan, M., Troch, P. a. and Carrillo, G.: Catchment classification: Empirical analysis of hydrologic similarity based on catchment function in the eastern USA, Hydrol. Earth Syst. Sci., 15(9), 2895–2911, doi:10.5194/hess-15-2895-2011, 2011.

Schaefli, B.: Snow hydrology signatures for model identification within a limits-of-acceptability approach, Hydrol. Process., doi:10.1002/hyp.10972, 2016.

Sefton, C. E. M. and Howarth, S. M.: Relationships between dynamic response characteristics and physical descriptors of catchments in England and Wales, J. Hydrol., 211, 1–16, doi:10.1016/S0022-1694(98)00163-2, 1998.

Seibert, J.: On the need for benchmarks in hydrological modelling, Hydrol. Process., 15(6), 1063–1064, doi:10.1002/hyp.446, 2001.

Seibert, J., Vis, M., Lewis, E. and van Meerveld, I.: Upper and lower benchmarks in

hydrological modeling, Hydrol. Process., 1–8, doi:10.1002/hyp.11476, 2018.

Shamir, E., Imam, B., Gupta, H. V. and Sorooshian, S.: Application of temporal streamflow descriptors in hydrologic model parameter estimation, Water Resour. Res., 41(6), 1–16, doi:10.1029/2004WR003409, 2005.

Singh, R., Archfield, S. A. and Wagener, T.: Identifying dominant controls on hydrologic parameter transfer from gauged to ungauged catchments - A comparative hydrology approach, J. Hydrol., 517, 985–996, doi:10.1016/j.jhydrol.2014.06.030, 2014.

Snelder, T. H., Lamouroux, N., Leathwick, J. R., Pella, H., Sauquet, E. and Shankar, U.: Predictive mapping of the natural flow regimes of France, J. Hydrol., 373(1–2), 57–67, doi:10.1016/j.jhydrol.2009.04.011, 2009.

Thornton, P. E., Thornton, M. M., Mayer, B. W., Wilhelmi, N., Wei, Y. and Cook, R. B.: Daymet: Daily surface weather on a 1km grid for North America, 1980–2012., 2012.

Troch, P. A., Martinez, G. F., Pauwels, V. R. N., Durcik, M., Sivapalan, M., Harman, C., Brooks, P. D., Gupta, H. and Huxman, T.: Climate and vegetation water use efficiency at catchment scales, Hydrol. Process., 23, 2409–2414, doi:10.1002/hyp.7358, 2009.

Vano, J. A., Nijssen, B. and Lettenmaier, D. P.: Seasonal hydrologic responses to climate change in the Pacific Northwest, Water Resour. Res., 51(4), 1959–1976, doi:10.1002/2014WR015909, 2015.

Viger, R. J. and Bock, A.: GIS Features of the Geospatial Fabric for National Hydrologic Modeling., 2014.

Vrugt, J. A. and Sadegh, M.: Toward diagnostic model calibration and evaluation: Approximate Bayesian computation, Water Resour. Res., 49(7), 4335–4345, doi:10.1002/wrcr.20354, 2013.

Wagener, T. and Wheater, H. S.: Parameter estimation and regionalization for continuous rainfall-runoff models including uncertainty, J. Hydrol., 320(1–2), 132–154, doi:10.1016/j.jhydrol.2005.07.015, 2006.

Westerberg, I. K. and McMillan, H. K.: Uncertainty in hydrological signatures, Hydrol. Earth Syst. Sci., 19(9), 3951–3968, doi:10.5194/hess-19-3951-2015, 2015.

Westerberg, I. K., Wagener, T., Coxon, G., McMillan, H. K., Castellarin, A., Montanari, A. and Freer, J.: Uncertainty in hydrological signatures for gauged and ungauged catchments, Water Resour. Res., 52, 1847–1865, doi:10.1002/2015WR017635, 2016.

Woods, R. A.: Analytical model of seasonal climate impacts on snow hydrology: Continuous snowpacks, Adv. Water Resour., 32, 1465–1481, doi:10.1016/j.advwatres.2009.06.011, 2009.

Yadav, M., Wagener, T. and Gupta, H.: Regionalization of constraints on expected watershed response behavior for improved predictions in ungauged basins, Adv. Water Resour., 30(8), 1756–1774, doi:10.1016/j.advwatres.2007.01.005, 2007.

Yilmaz, K. K., Gupta, H. V. and Wagener, T.: A process-based diagnostic approach to model evaluation: Application to the NWS distributed hydrologic model, Water Resour. Res., 44(May), 1–18, doi:10.1029/2007WR006716, 2008.

Young, A. R.: Stream flow simulation within UK ungauged catchments using a daily rainfall-runoff model, J. Hydrol., 320(1–2), 155–172, doi:10.1016/j.jhydrol.2005.07.017, 2006.

Zeng, X.: Global Vegetation Root Distribution for Land Modeling, J. Hydrometeorol., 2(5), 525–530, doi:10.1175/1525-7541(2001)002<0525:GVRDFL>2.0.CO;2, 2001.

(Court, 1962; Cosby et al., 1984; Clausen and Biggs, 2000; Zeng, 2001; Olden and Poff, 2003; Lawrence and Slater, 2008; Woods, 2009; Falcone, 2011; Hartmann and Moosdorf, 2012; Ladson et al., 2013; Gleeson et al., 2014; Viger and Bock, 2014; Westerberg and McMillan, 2015)

Table 1: Catchment attributes

camels_topo - Topography and location			
Description	*Unit*	*Data source*	*References*
gauge latitude	° north	N15 - USGS data	
gauge longitude	° east	N15 - USGS data	
catchment mean elevation	meter above sea level	N15 - USGS data	
catchment mean slope	m/km	N15 - USGS data	
catchment area (GAGESII estimate)	km²	N15 - USGS data	Falcone (2011)
catchment area (Geospatial Fabric estimate)	km²	N15 - Geospatial Fabric	Viger (2014), Viger and Bock (2014)

camels_clim - Climate indices - *: Computed over the period 1989/10/01 to 2009/09/30			
Description	*Unit*	*Data source*	*References*
mean daily precipitation	mm/day	N15 - Daymet*	
mean daily PET [estimated by N15 using Priestley-Taylor formulation calibrated for each catchment]	mm/day	N15 - Daymet*	
aridity (PET/P, ratio of mean PET [estimated by N15 using Priestley-Taylor formulation calibrated for each catchment] to mean precipitation)	-	N15 - Daymet*	
seasonality and timing of precipitation (estimated using sine curves to represent the annual temperature and preciptiation cycles, positive [negative] values indicate that precipitation peaks in summer [winter], values close to 0 indicate uniform precipitation throughout the year)	-	N15 - Daymet*	Eq. 14 in Woods et al. (2009)
fraction of precipitation falling as snow (i.e., on days colder than 0°C)	-	N15 - Daymet*	
frequency of high precipitation days (>= 5 times mean daily precipitation)	days/year	N15 - Daymet*	
average duration of high precipitation events (number of consecutive days >= 5 times mean daily precipitation)	days	N15 - Daymet*	
season during which most high precipitation days (>= 5 times mean daily precip.) occur	season	N15 - Daymet*	
frequency of dry days (<1 mm/day)	days/year	N15 - Daymet*	
average duration of dry periods (number of consecutive days <1 mm/day)	days	N15 - Daymet*	
season during which most dry days (<1 mm/day) occur	season	N15 - Daymet*	

camels_vege - Land cover characteristics - *: Period 2002 to 2014			
Description	*Unit*	*Data source*	*References*
forest fraction	-	N15 - USGS data	
maximum monthly mean of the leaf area index (based on 12 monthly means)	-	MODIS*	
difference between the maximum and mimumum monthly mean of the leaf area index (based on 12 monthly means)	-	MODIS*	
maximum monthly mean of the green vegetation fraction (based on 12 monthly means)	-	MODIS*	
difference between the maximum and mimumum monthly mean of the green vegetation fraction (based on 12 monthly means)	-	MODIS*	
dominant land cover type (Noah-modified 20-category IGBP-MODIS land cover)	-	MODIS*	
fraction of the catchment area associated with the dominant land cover	-	MODIS*	
root depth (percentiles XX = 50 and 99% extracted from a root depth distribution based on IGBP land cover)	m	MODIS*	Eq. 2 and Table 2 in Zeng (2001)

Table 1 continued: Catchment attributes

camels_soil - Soil characteristics - *: Only covers the top 1.5 m			
Description	**Unit**	**Data source**	**References**
depth to bedrock (maximum 50m)	m	Pelletier et al.	
soil depth (maximum 1.5m, layers marked as water and bedrock were excluded)	m	Miller and White (1998) - STATSGO*	
volumetric porosity (saturated volumetric water content estimated using a multiple linear regression based on sand and clay fraction for the layers marked as USDA soil texture class and a default value [0.9] for layers marked as organic material, layers marked as water, bedrock and "other" were excluded)	-	Miller and White (1998) - STATSGO*	Table 4 in Cosby et al. (1984), Lawrence and Slater (2008)
saturated hydraulic conductivity (estimated using a multiple linear regression based on sand and clay fraction for the layers marked as USDA soil texture class and a default value [36cm/hr] for layers marked as organic material, layers marked as water, bedrock and "other" were excluded)	cm/hr	Miller and White (1998) - STATSGO*	Table 4 in Cosby et al. (1984), Lawrence and Slater (2008)
maximum water content (combination of porosity and soil_depth_statgso, layers marked as water, bedrock and "other" were excluded)	m	Miller and White (1998) - STATSGO*	
sand fraction (of the soil material smaller than 2 mm, layers marked as oragnic material, water, bedrock and "other" were excluded)	%	Miller and White (1998) - STATSGO*	
silt fraction (of the soil material smaller than 2 mm, layers marked as oragnic material, water, bedrock and "other" were excluded)	%	Miller and White (1998) - STATSGO*	
clay fraction (of the soil material smaller than 2 mm, layers marked as oragnic material, water, bedrock and "other" were excluded)	%	Miller and White (1998) - STATSGO*	
fraction of the top 1.5m marked as water (class 14)	%	Miller and White (1998) - STATSGO*	
fraction of soil_depth_statsgo marked as organic material (class 13)	%	Miller and White (1998) - STATSGO*	
fraction of soil_depth_statsgo marked as other (class 16)	%	Miller and White (1998) - STATSGO*	

camels_geol - Geological characteristics			
Description	**Unit**	**Data source**	**References**
most common geologic class in the catchment	-	GLiM	Hartmann and Moosdorf (2012)
fraction of the catchment area associated with its most common geologic class	-	GLiM	Hartmann and Moosdorf (2012)
2nd most common geologic class in the catchment	-	GLiM	Hartmann and Moosdorf (2012)
fraction of the catchment area associated with its 2nd most common geologic class	-	GLiM	Hartmann and Moosdorf (2012)
fraction of the catchment area characterized as "Carbonate sedimentary rocks"	-	GLiM	Hartmann and Moosdorf (2012)
subsurface porosity	-	GLHYMPS	Gleeson et al. (2014)
subsurface permeability (log10)	m^2	GLHYMPS	Gleeson et al. (2014)

Table 2: Hydrological signatures

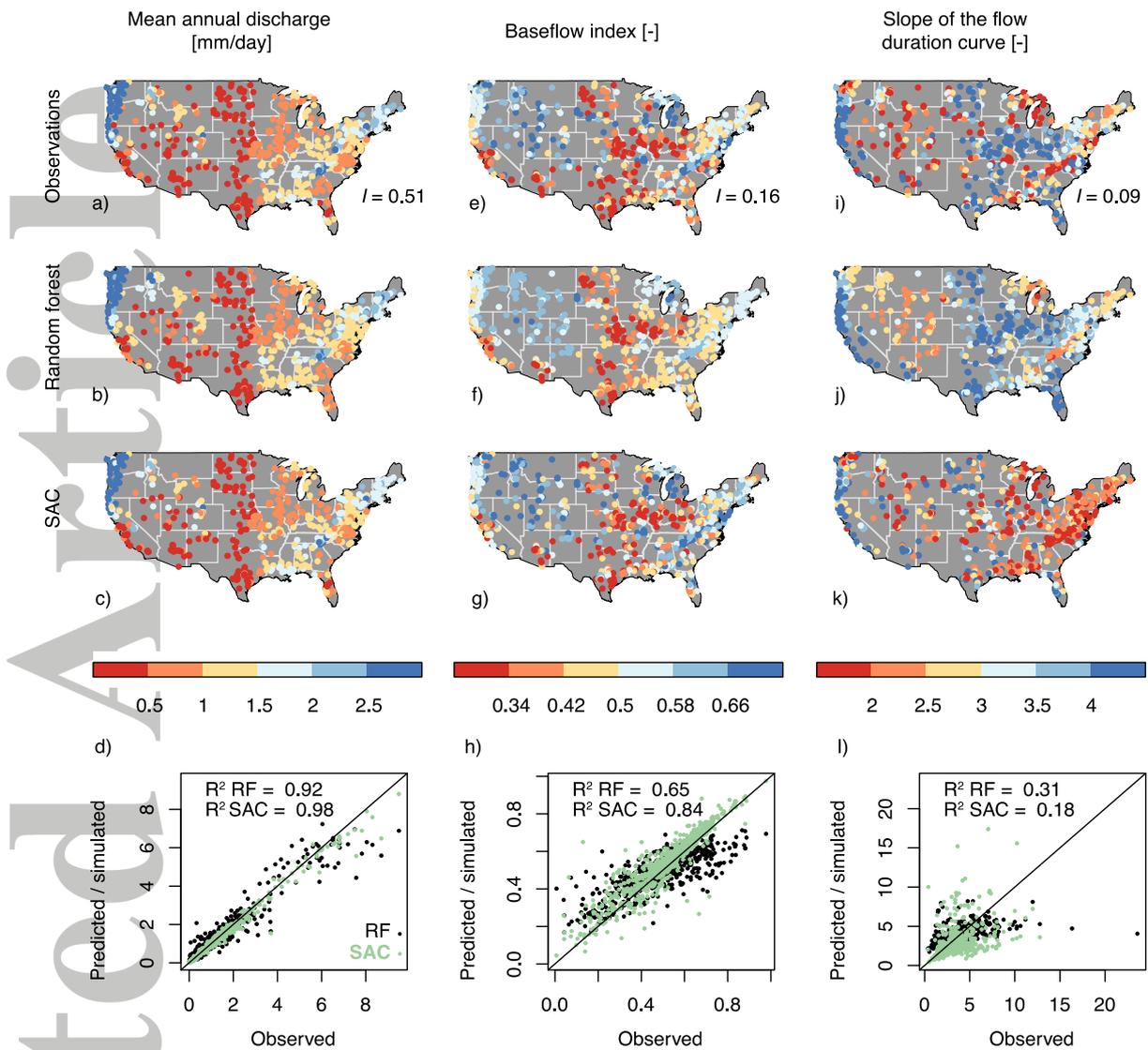| Description	Unit	Data source	References
camels_hydro - Hydrological signatures - *: Period 1989/10/01 to 2009/09/30			
mean annual discharge	mm/day	N15 - USGS data*	
mean winter (DJF) discharge	mm/day	N15 - USGS data*	
mean summer (JJA) discharge	mm/day	N15 - USGS data*	
runoff ratio (ratio of mean daily discharge to mean daily precipitation)	-	N15 - USGS data*	Eq. 2 in Sawicz et al. (2011)
streamflow-precipitation elasticity (sensitivity of streamflow to changes in precipitation at the annual time scale)	-	N15 - USGS data*	Eq. 7 in Sankarasubramanian et al. (2001), the last element being P/Q not Q/P
slope of the flow duration curve (between the log-transformed 33rd and 66th streamflow percentiles)	-	N15 - USGS data*	Eq. 3 in Sawicz et al. (2011)
baseflow index (ratio of mean daily baseflow to mean daily discharge, hydrograph separation performed using Ladson et al. [2013] digital filter)	-	N15 - USGS data*	Ladson et al. (2013)
mean half flow date (date on which the cumulative discharge since October 1st reaches half of the annual discharge)	day of year	N15 - USGS data*	Court (1962)
5% flow quantile (low flow)	mm/day	N15 - USGS data*	
95% flow quantile (high flow)	mm/day	N15 - USGS data*	
frequency of high-flow days (> 9 times the median daily flow)	days/year	N15 - USGS data*	Clausen and Biggs (2000), Table 2 in Westerberg and McMillan (2015)
mean duration of high-flow events (number of consecutive days > 9 times the median daily flow)	days	N15 - USGS data*	Clausen and Biggs (2000), Table 2 in Westerberg and McMillan (2015)
frequency of low-flow days (< 0.2 times the mean daily flow)	days/year	N15 - USGS data*	Olden and Poff (2003), Table 2 in Westerberg and McMillan (2015)
mean duration of low-flow events (number of consecutive days < 0.2 times the mean daily flow)	days	N15 - USGS data*	Olden and Poff (2003), Table 2 in Westerberg and McMillan (2015)
frequency of days with Q = 0 mm/day	%	N15 - USGS data*	

Figure 1: Comparison of the observed, predicted and simulated (first, second and third row, respectively) mean annual discharge, baseflow index and slope of the flow duration curve (first, second and third column, respectively). The spatial auto-correlation quantified using Moran's I is indicated for the maps of top row. The last row combines and compares the data from the three maps of the same column and indicates the coefficient of determination R^2 for the random forest predictions and SAC simulations computed over all the catchments.
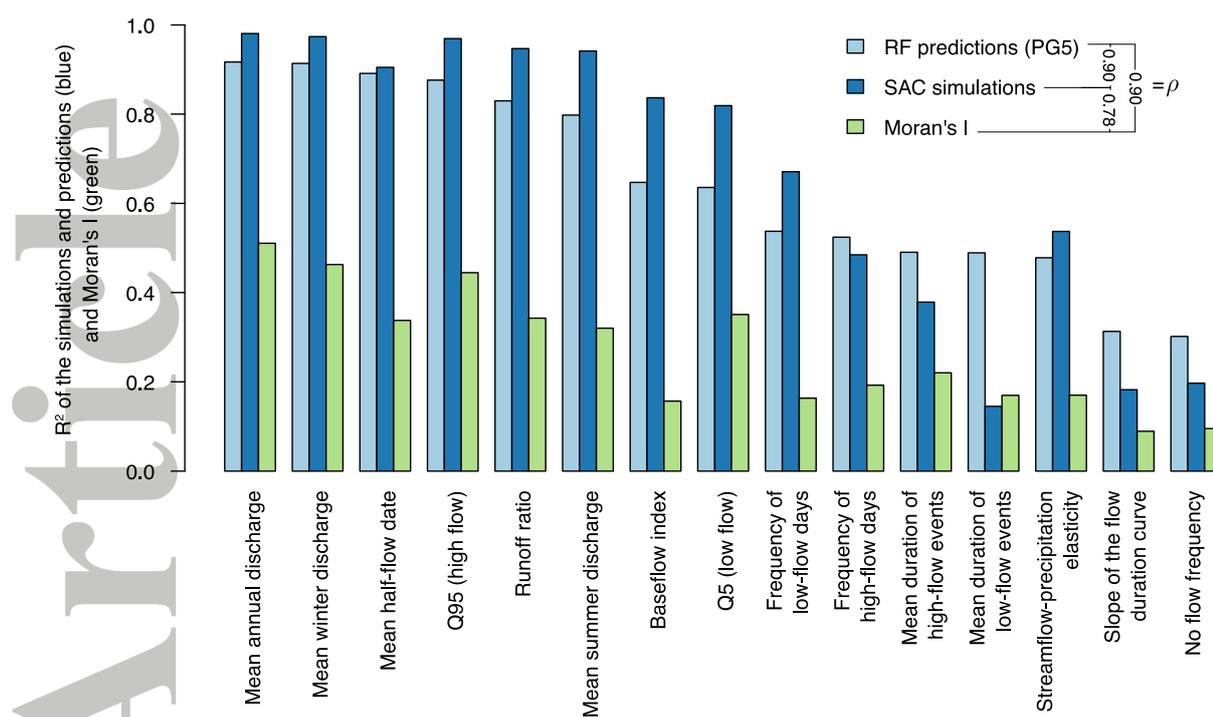
Figure 2: Illustration of the strong three-way relationship between how well signatures can be predicted based on catchment attributes using a random forest (R^2 between the observed and predicted signatures, light blue), how well they can be simulated by SAC (R^2 between the observed and simulated signatures, dark blue), and the smoothness of their spatial variability over the CONUS (Moran's I, green). The correlations between these variables are indicated in the upper-right corner. The signatures are ordered from left to right based on how well they can be predicted using a random forest based on all the predictors (PG5).
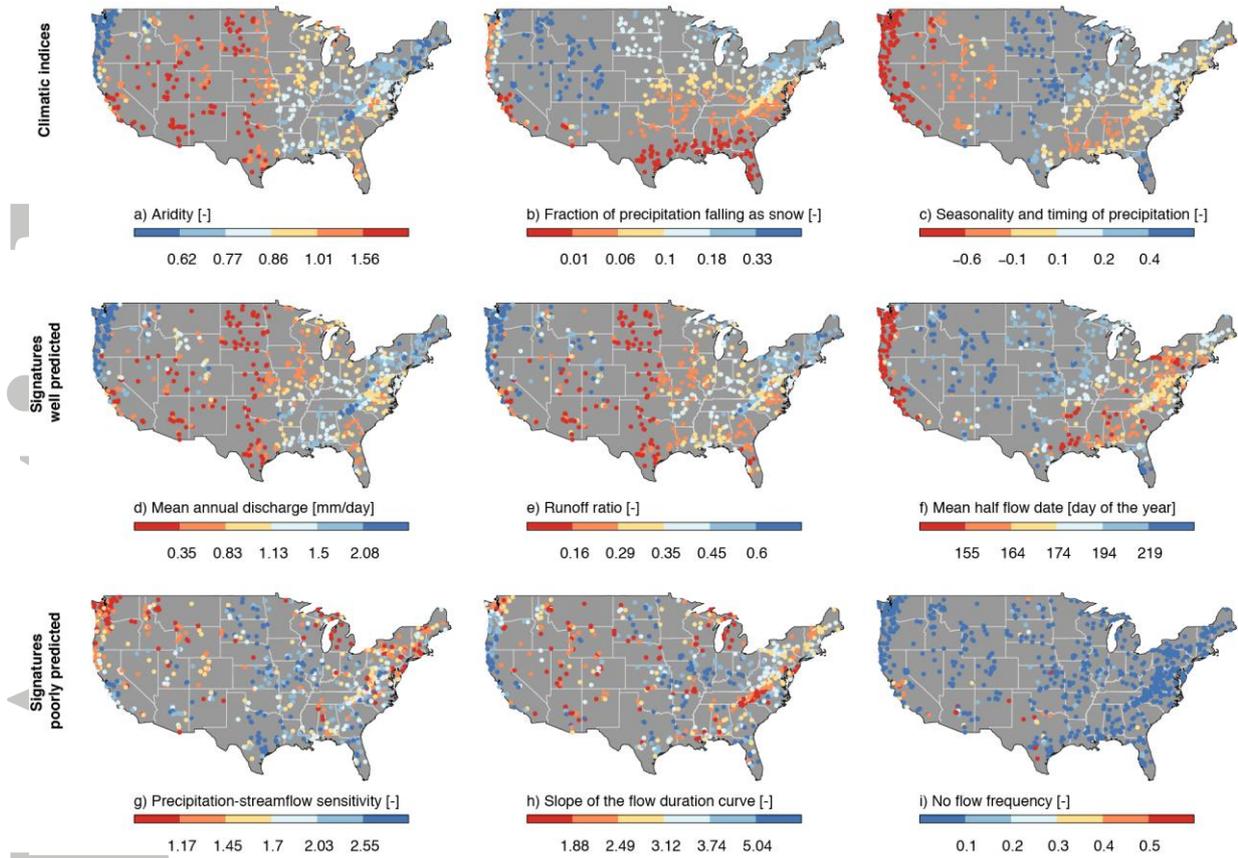
Figure 3: Comparison of the spatial patterns in climatic indices (top row), well-predicted hydrological signatures (middle row) and poorly-predicated hydrological signatures (bottom row). We used the same color scheme for all the maps to underscore similarities between them. Note that units and break values vary. The break values were chosen so that each color class encompasses about one sixth of the total number of catchments (except for the no flow frequency).
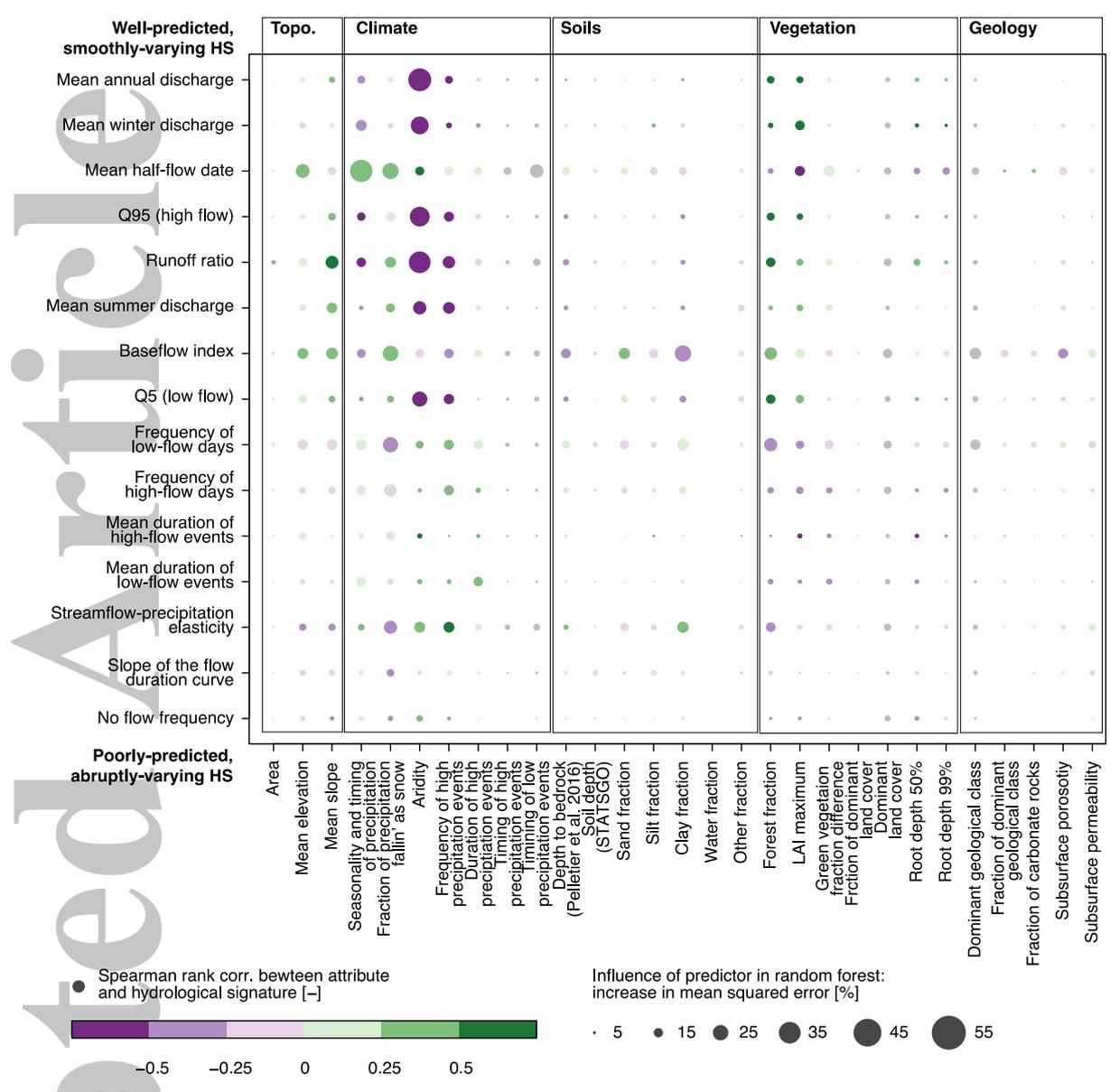
Figure 4: Comparison of the influence of each catchment attributes (x-axis) on each signature (y-axis) in the random forest. Their influence is measured using IncMSE and is proportional to the size of the dots. The signatures are ordered vertically based on how well they are predicted by random forests (same order as for Figure 2).
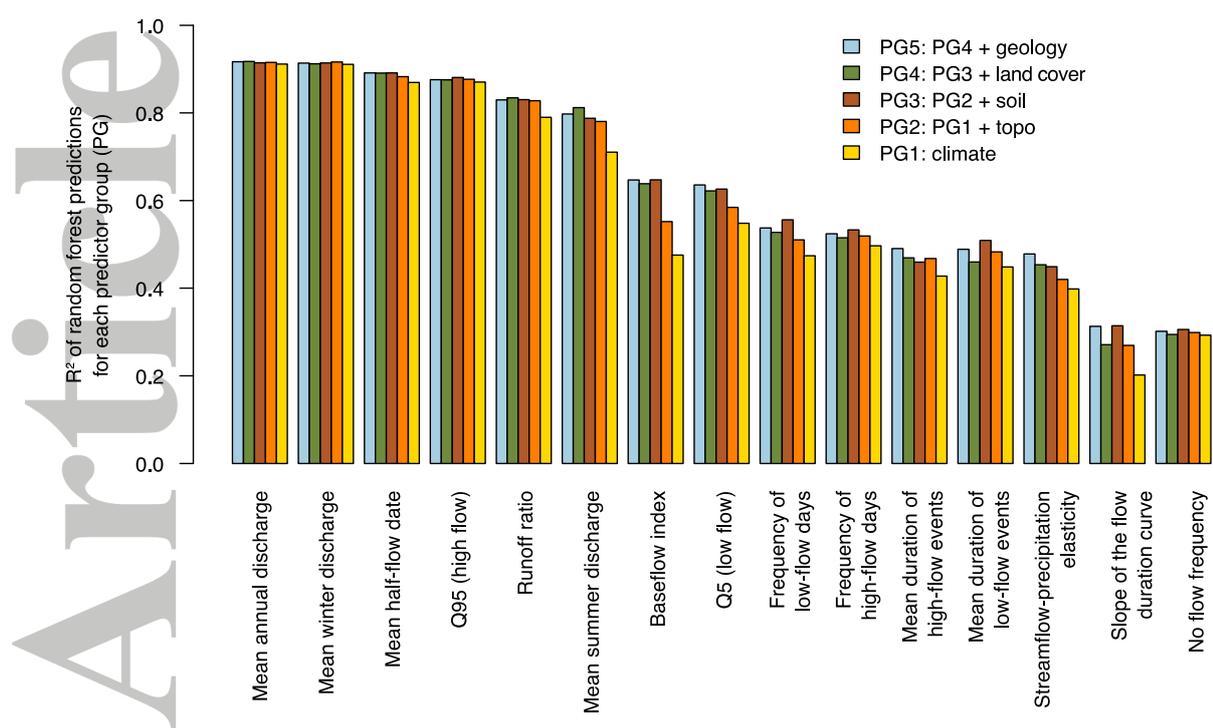
Figure 5: Exploration of the predictive value added by each class of attributes. The signatures are ordered based on their predictability using all the predictors (PG5, same as in Figure 2). The light blue bars show the same information as the light blue bars in Figure 2.
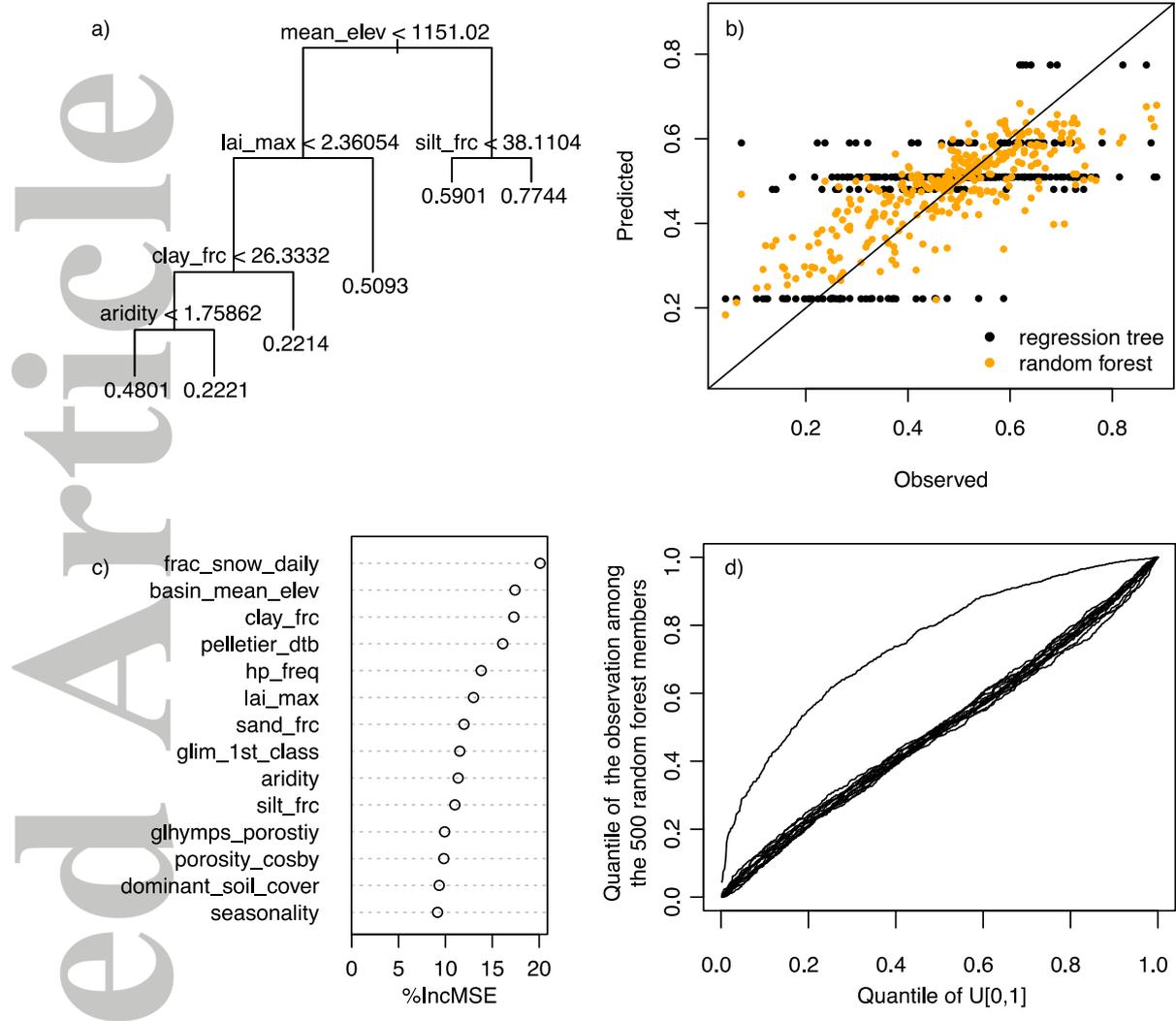
Figure A1: a) Example of a pruned regression tree trained to predict the baseflow index. b) Comparison of baseflow index observations to predictions from the regression tree shown in a) and from a random forest, whose most influential predictors are shown in c). c) Assessment of the relative influence of the random forest variables for the prediction of the baseflow index, the predictors are ordered from the most to least influential (top to bottom). d) QQplot for the 15 hydrological variables, lines close to the diagonal indicate reliable ensembles, the only line significantly departing from the diagonal is the fraction of no flow, see Laio and Tamea (2007) or Renard et al. (2010) for more details on how to interpret this plot.

camels_hydro - Hydrological signatures - *: Period 1989/10/01 to 2009/09/30				
Attribute	*Description*	*Unit*	*Data source*	*References*
q_mean	mean daily discharge	mm/day	N15 - USGS data*	
runoff_ratio	runoff ratio (ratio of mean daily discharge to mean daily precipitation)	-	N15 - USGS data*	Eq. 2 in Sawicz et al. (2011)
stream_elas	streamflow precipitation elasticity (sensitivity of streamflow to changes in precipitation at the annual time scale)	-	N15 - USGS data*	Eq. 7 in Sankarasubramanian et al. (2001), the last element being P/Q not Q/P
slope_fdc	slope of the flow duration curve (between the log-transformed 33rd and 66th streamflow percentiles)		N15 - USGS data*	Eq. 3 in Sawicz et al. (2011)
baseflow_index	baseflow index (ratio of mean daily baseflow to mean daily discharge, hydrograph separation performed using Ladson et al. [2013] digital filter)	-	N15 - USGS data*	Ladson et al. (2013)
hfd_mean	mean half flow date (date on which the cumulative discharge since October 1st reaches half of the annual discharge)	day of year	N15 - USGS data*	Court (1962)
Q5	5% flow quantile (flow flow)	mm/day	N15 - USGS data*	
Q95	95% flow quantile (high flow)	mm/day	N15 - USGS data*	
high_q_freq	frequency of high-flow days (> 9 times the median daily flow)	days/year	N15 - USGS data*	Clausen and Biggs (2000), Table 2 in Westerberg and McMillan (2015)
high_q_dur	average duration of high-flow events (number of consecutive days > 9 times the median daily flow)	days	N15 - USGS data*	Clausen and Biggs (2000), Table 2 in Westerberg and McMillan (2015)
low_q_freq	frequency of low-flow days (< 0.2 times the mean daily flow)	days/year	N15 - USGS data*	Olden and Poff (2003), Table 2 in Westerberg and McMillan (2015)
low_q_dur	average duration of low-flow events (number of consecutive days < 0.2 times the mean daily flow)	days	N15 - USGS data*	Olden and Poff (2003), Table 2 in Westerberg and McMillan (2015)
zero_q_freq	frequency of days with Q = 0 mm/day	%	N15 - USGS data*	