University of Bristol - Explore Bristol Research
General rights

RESEARCH ARTICLE

Uncertainty in hydrological signatures for gauged and ungauged catchments

Ida K. Westerberg[1,2], Thorsten Wagener[1,3], Gemma Coxon[4], Hilary K. McMillan[5,6], Attilio Castellarin[7], Alberto Montanari[7], and Jim Freer[3,4]

[1]Department of Civil Engineering, University of Bristol, Bristol, UK, [2]IVL Swedish Environmental Research Institute, Stockholm, Sweden, [3]Cabot Institute, University of Bristol, Bristol, UK, [4]School of Geographical Sciences, University of Bristol, Bristol, UK, [5]National Institute of Water and Atmospheric Research, Christchurch, New Zealand, [6]Department of Geography, San Diego State University, San Diego, California, USA, [7]Department of Civil, Chemical, Environmental, and Materials Engineering (DICAM), University of Bologna, Bologna, Italy

Abstract Reliable information about hydrological behavior is needed for water-resource management and scientific investigations. Hydrological signatures quantify catchment behavior as index values, and can be predicted for ungauged catchments using a regionalization procedure. The prediction reliability is affected by data uncertainties for the gauged catchments used in prediction and by uncertainties in the regionalization procedure. We quantified signature uncertainty stemming from discharge data uncertainty for 43 UK catchments and propagated these uncertainties in signature regionalization, while accounting for regionalization uncertainty with a weighted-pooling-group approach. Discharge uncertainty was estimated using Monte Carlo sampling of multiple feasible rating curves. For each sampled rating curve, a discharge time series was calculated and used in deriving the gauged signature uncertainty distribution. We found that the gauged uncertainty varied with signature type, local measurement conditions and catchment behavior, with the highest uncertainties (median relative uncertainty ±30–40% across all catchments) for signatures measuring high- and low-flow magnitude and dynamics. Our regionalization method allowed assessing the role and relative magnitudes of the gauged and regionalized uncertainty sources in shaping the signature uncertainty distributions predicted for catchments treated as ungauged. We found that (1) if the gauged uncertainties were neglected there was a clear risk of overconditioning the regionalization inference, e.g., by attributing catchment differences resulting from gauged uncertainty to differences in catchment behavior, and (2) uncertainty in the regionalization results was lower for signatures measuring flow distribution (e.g., mean flow) than flow dynamics (e.g., autocorrelation), and for average flows (and then high flows) compared to low flows.

1. Introduction

Reliable information about the hydrological behavior of both gauged and ungauged catchments is needed for a wide range of scientific and water-resources management purposes. Such information is often summarized as an index value – or a hydrological signature – calculated from data time series in gauged catchments. Examples include the base-flow index and flow descriptors such as flow percentiles or statistics of high and low flow behavior. Signatures have a long history of use in eco-hydrology [*Olden and Poff*, 2003] and hydrology for, e.g., change detection [*Archer and Newson*, 2002; *Juston et al.*, 2014; *Sawicz et al.*, 2014], model evaluation [*Hrachowitz et al.*, 2014; *Montanari and Toth*, 2007; *Refsgaard and Knudsen*, 1996; *Sugawara*, 1979], model-structure diagnostics [*Coxon et al.*, 2013; *Gupta et al.*, 2008; *Jothityangkoon et al.*, 2001; *McMillan et al.*, 2011], and catchment classification [*Sawicz et al.*, 2011]. In particular, they have been widely used for transferring information about hydrological behavior from gauged to ungauged catchments [*Bloeschl et al.*, 2013]. In this paper we consider regionalization procedures that transfer flow signature information directly from gauged to ungauged catchments (i.e., without using a hydrological model), and the uncertainties that affect such procedures.

Uncertainty in signature values for gauged catchments stems from the observed data from which they are calculated and, for more complex signatures such as recession parameters, from the choice of calculation method [*Westerberg and McMillan*, 2015]. Such uncertainties reduce the information gained from the signature values for hydrological analyses and thus also the reliability of those analyses, e.g., when used to study

differences in catchment behavior [*Wagener and Montanari*, 2011]. It is therefore important to understand the magnitude and characteristics of signature uncertainty under different conditions. The main sources of data uncertainty are the measurements' accuracy, precision and representativeness for the studied variable [*McMillan et al.*, 2012], but also data postprocessing [*Hamilton and Moore*, 2012]. Studies have shown that rating curve uncertainty propagates to uncertainty in flood-frequency estimates and in signatures used for model calibration or change detection for individual catchments [*Blazkova and Beven*, 2009; *Juston et al.*, 2014; *Kuczera*, 1996; *Westerberg et al.*, 2011]. *Westerberg and McMillan* [2015] found that rainfall-runoff signature uncertainty as a result of observational uncertainty for two catchments in the UK and New Zealand were on the order of $\pm 10-40\%$ and varied between the signatures. However, there have been no large-scale studies investigating these uncertainties across multiple catchments and multiple signature types.

When regionalizing signature values to an ungauged catchment, the uncertainty in the regionalized signatures have several sources; 1) uncertainty in the signatures calculated for the gauged catchments, 2) uncertainty stemming from the regionalization procedure, and where the latter may include 3) uncertainty in catchment characteristics data (e.g., geomorphological descriptors like elevation and soils) used to describe catchment similarity for the transfer of information. There is a long tradition of regionalization of flow signatures to ungauged basins [*Bloeschl et al.*, 2013], with common approaches including those based on regression against catchment descriptors [e.g., *Almeida et al.*, 2012; *Bardossy*, 2007; *Castiglioni et al.*, 2010; *Nathan and McMahon*, 1992], donor catchments or pooling groups [*Burn*, 1990; *Holmes et al.*, 2002; *Kjeldsen et al.*, 2014], and, more recently, geostatistics [*Pugliese et al.*, 2014; *Viglione et al.*, 2013]. *Yadav et al.* [2007] investigated uncertainty in the regionalization procedure when using regression to regionalize signature values for 30 UK catchments. They found that regionalization performance varied widely between signatures and that the most useful independent catchment characteristics were climate, topography and geology characteristics. *Hannaford et al.* [2013] evaluated the utility of the hydrometric network in England and Wales for regionalization based on catchment descriptors and gauging station data quality. They found that for low (high) flows 22% (45%) of the catchments with the highest regionalization potential have low utility because of inadequate hydrometric data quality. *Westerberg et al.* [2014] investigated signature uncertainty resulting from observed data and the regionalization procedure for 36 Central American catchments. They regionalized flow duration curves (FDCs) using a typical estimate of discharge uncertainty for the region, and found that the majority of the predicted uncertainty bounds encompassed the observed values. However, discharge uncertainty is known to vary with flow range depending on site-specific measurement conditions [*Le Coz et al.*, 2014; *McMillan and Westerberg*, 2015; *Morlot et al.*, 2014]. Further investigation using data sets that allow site-specific uncertainty estimates is therefore needed to gain a better understanding of uncertainty for a wider range of flow signatures compared to earlier studies for both gauged catchments and in regionalization for ungauged catchments.

The aim of this study was to investigate uncertainty in flow signatures for gauged and ungauged catchments. In particular the objectives were to; 1) regionalize signatures while accounting for discharge uncertainty in the gauged donor catchments as well as uncertainty in the regionalization procedure, and 2) investigate the role and relative magnitude of the different uncertainty sources in defining the predicted signature uncertainties. Uncertainty in the catchment characteristics data used to describe catchment similarity was not included in this study.

2. Data

The study was performed using a comprehensive data set consisting of 15 min water level time series (1 October 2003 to 30 September 2008) in combination with rating curve and gauging data for 43 catchments in England and Wales ranging in size from 8 to 1480 km^2 (Figure 1). The catchment characteristics that we used for the regionalization procedure were: mean annual precipitation for the study period, BFIHOST (a baseflow index from the UK Flood Estimation Handbook derived from soil characteristics in the hydrology of soil types (HOST) classification [*Institute of Hydrology*, 1999]), and the 90th percentile of the catchment elevation distribution (see also section 3.3). The BFIHOST and elevation indices were obtained from the UK Hydrometric Register [*Marsh and Hannaford*, 2008]. We selected catchments that fulfilled a number of criteria to ensure reliable discharge data uncertainty estimates and that the regionalization performance was not affected by anthropogenic factors or nested catchment locations. The criteria were: (1) the station was active, (2) it was classified as having a natural flow regime in the UK Hydrometric Register,
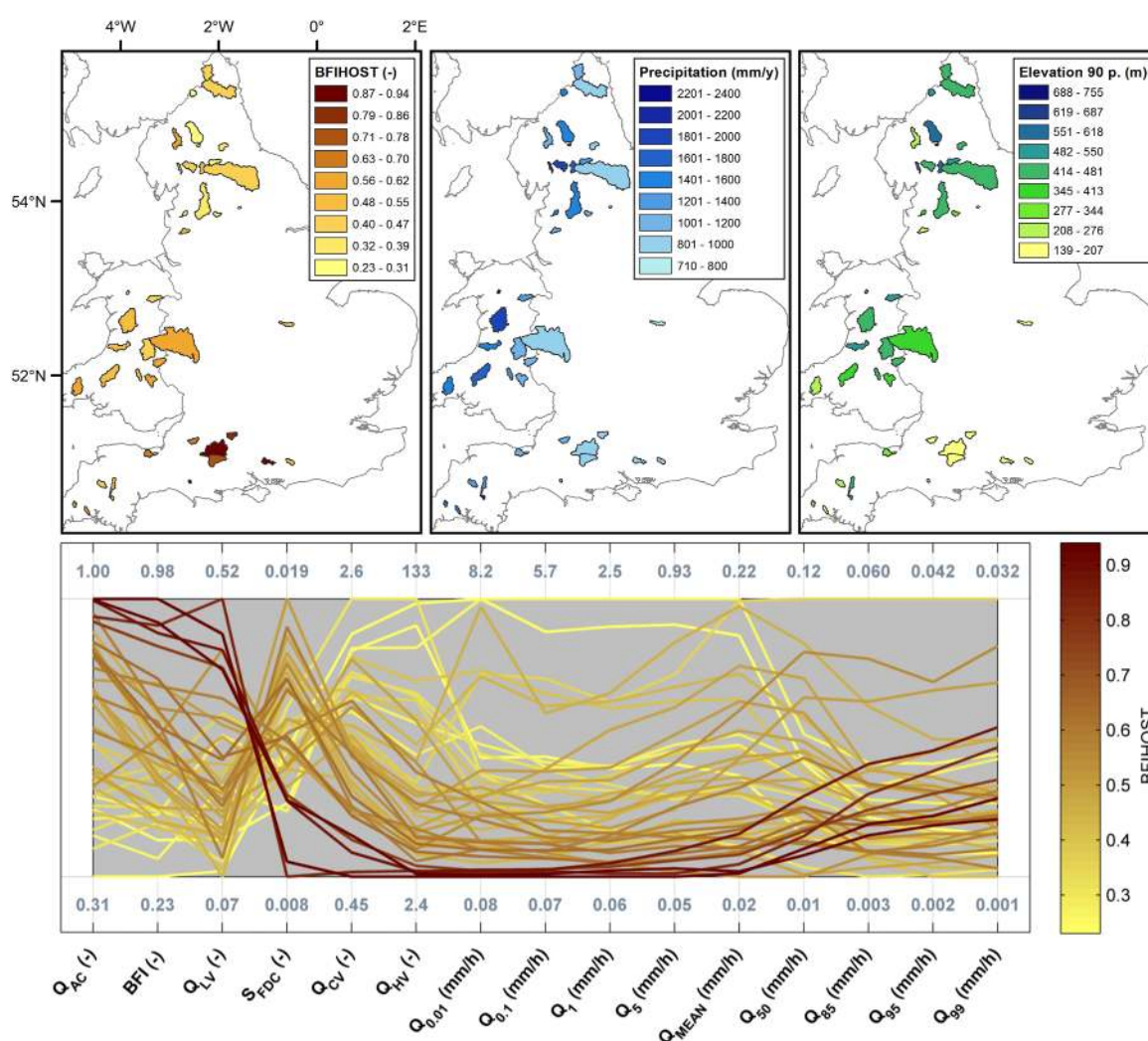
Figure 1. The 43 catchments in England and Wales used in the study and their range of characteristics (top, where Elevation 90 p. is the 90th percentile of the catchment elevation distribution) and signature values (bottom, where the parallel coordinate plots shows the range of signature values with the minimum (maximum) values below (above) the plot). The signatures were calculated with the optimal rating curves from the uncertainty analysis (signature definitions in Table 1).

(3) the station was classified as having a Service License Agreement in the register (part of a strategic monitoring network subject to more rigorous quality control), (4) data suitable for reliable discharge uncertainty analyses were available (e.g., sufficient information about out-of-bank rating, no stilling-well problems, etc.), (5) it was a gauged weir and/or velocity-area station, and 6) it was not upstream/downstream of another catchment in the data set. Only water level data not classified as suspect by the data provider were used; other uncertainties in the water level data series were not considered. There were five stations with 5–12% missing water level data, 32 stations had less than 2% missing data and the rest were in-between 2% and 5%. The chosen catchments spanned a wide range of hydrological behavior (Figure 1), representing most of the range of catchments classified as having a natural flow regime in England and Wales. In central and Eastern England, there are few catchments with natural flow regimes, and only one of these fulfilled all the selection criteria for this study.

3. Methods

3.1. Choice and Calculation of Signatures

We used nine signatures describing the flow distribution and six signatures that describe flow dynamics (Table 1). These signatures describe the magnitude and dynamics of high and low flows together with

Table 1. Runoff Signatures Included in the Study[a]

Signature	Signature Name	Description	Unit
Flow Distribution			
Q_{MEAN}	Mean flow	Mean flow for the analysis period	mm/h
$Q_{0.01}, Q_{0.1}, Q_1, Q_5,$ $Q_{50}, Q_{85}, Q_{95}, Q_{99}$	Flow percentiles	Low and high flow exceedance percentiles from the FDC	mm/h
Flow Dynamics			
Q_{BFI}	Base Flow Index	Contribution of baseflow to total streamflow. The index was calculated from daily flows according to the Flood Estimation Handbook methodology [*Gustard et al.*, 1992].	
S_{FDC}	Slope of normalized FDC	Slope of the FDC between the 33% and 66% flow exceedance values of streamflow normalized by their means [*Yadav et al.*, 2007].	
Q_{CV}	Overall flow variability	Coefficient of variation in streamflow (standard deviation divided by mean flow). Used by *Clausen and Biggs* [2000] and *Jowett and Duncan* [1990].	
Q_{LV}	Low flow variability	Mean of annual minimum flow divided by the median flow [*Jowett and Duncan*, 1990].	
Q_{HV}	High flow variability	Mean of annual maximum flow divided by the median flow [*Jowett and Duncan*, 1990].	
Q_{AC}	Flow autocorrelation	Autocorrelation for 1 day (24 h). Used by *Euser et al.* [2013] and *Winsemius et al.* [2009].	

[a]All signatures were calculated on hourly data unless otherwise specified, annual flow was calculated for hydrological years.

average flow conditions and overall flow variability. They represent signature information of interest for a wide range of applications and illustrate the effect of data uncertainty across a range of flow behavior. Most of these have been used in previous studies aiming to regionalize specific signature information separately [*Castellarin et al.*, 2004; *Holmes et al.*, 2002], when using multiple signatures to constrain rainfall-runoff models [*Euser et al.*, 2013; *Hrachowitz et al.*, 2014; *Yadav et al.*, 2007; *Yilmaz et al.*, 2008], and in eco-hydrological studies [*Clausen and Biggs*, 2000; *Jowett and Duncan*, 1990].

3.2. Discharge and Signature Uncertainty for Gauged Catchments

The uncertainty in the signatures for the gauged catchments was estimated as follows. Discharge uncertainty was estimated for each gauging station. The uncertainty in the rating curve parameters was estimated from the stage-discharge gauging data, obtaining 40,000 equally likely rating curves for each station (see below). Each rating curve was used to calculate a discharge time series from the water level data. The resulting set of 40,000 discharge time series were aggregated to hourly time scale, converted to specific discharge (i.e., per unit area, expressed in mm/h) and each used to calculate a signature value, thus obtaining an uncertainty distribution for each signature. A detailed description of this method is given by *Westerberg and McMillan* [2015].

The uncertainty in the rating curve parameters was estimated in a Markov Chain Monte Carlo (MCMC) analysis with the Voting Point likelihood method [*McMillan and Westerberg*, 2015]. This method accounts for random and epistemic uncertainty sources. Random (aleatory) gauging measurement uncertainty was estimated as logistic distribution functions for UK conditions for a set of stations where uncertainty due to temporal rating curve variability was assumed negligible [*Coxon et al.*, 2015]. Epistemic uncertainty related to the rating curve approximation of the true stage-discharge relationship is important to consider at many gauging stations. This approximation may be uncertain outside the gauged range where the curve is extrapolated, or where processes such as erosion, seasonal weed growth, hysteresis, and variable backwater induce nonstationarity in the stage-discharge relationship. Such epistemic uncertainties imply that all gauging points are not compatible with the same "true" rating curve, and the voting point likelihood was therefore defined in terms of the fraction of time that a candidate rating curve could have been representative of the channel conditions (see definition and equations in *McMillan and Westerberg* [2015]). The analysis was constrained to the functional form of the official rating curves used in the study period, which was a power-law function often containing multiple segments. The priors for the rating curve parameters were set to standardized ranges defined relative to the official parameter values. These were adjusted for some stations where visual inspection showed that these ranges did not fully capture the gauging data uncertainty. Only
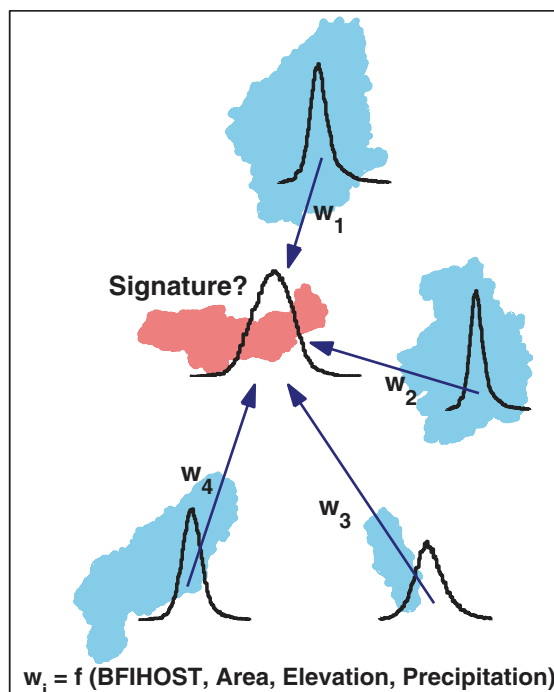
Figure 2. Schematic illustration of the signature regionalization procedure. The signature distribution for the target ungauged catchment is estimated by sampling from the signature distributions for the most hydrologically similar gauged catchments proportional to their hydrological similarity weight (w_i). The hydrologic similarity was calculated as a function of the catchment descriptors BFIHOST, the 90th percentile of the elevation distribution, catchment area and mean annual precipitation.

the gaugings that were representative for the rating curve were used, where the gauging data were pooled based on deviations between historical rating curves [*Coxon et al.*, 2015].

3.3. Regionalization of Signatures to Ungauged Catchments With Uncertainty

Traditional regionalization methods, e.g., regression of signature values using catchment descriptors as independent variables, allow estimation of predictive uncertainty, but involve strong assumptions on the signature error distribution (e.g., normality). Those assumptions may not be compatible with uncertainties estimated in site-specific analyses of gauged catchments. Instead, our regionalization method is based on hydrologic similarity, allowing for different empirical distributions for the gauged signature uncertainties, drawing on previous studies by *Holmes et al.* [2002] and *Westerberg et al.* [2014]. Hydrologic similarity was expressed as the Euclidean distance d_{it} in the standardized catchment descriptor space between each gauged catchment, i, and the target catchment, t:

$$d_{it} = \sqrt{\sum_{m=1}^{M} (X_{mi} - X_{mt})^2} \qquad (1)$$

where X_{mi} is the standardized catchment descriptor m (normalized by the standard deviation for all stations) for catchment i, and M the number of descriptors. The catchment descriptors were chosen in a correlation analysis aimed at finding descriptors that were highly correlated with the signature values, but weakly correlated with each other (following *Yadav et al.* [2007] and *Westerberg et al.* [2014]). The chosen descriptors were mean annual precipitation in the study period, the 90 percentile catchment elevation, the BFIHOST base-flow index and catchment area (section 2 and Figure 1). These describe climate, topography and geology, similar to descriptors previously found to explain most of the observed daily streamflow behavior for UK catchments [*Yadav et al.*, 2007]. We found it useful to also include catchment area [*Kjeldsen et al.*, 2014; *McIntyre et al.*, 2005], which can, for example, explain differences in flow peak attenuation that are more pronounced in hourly data.

A dynamic region of influence (i.e., a pooling group) was defined as the N catchments that were most similar to the target catchment [*Burn*, 1990]. The signature PDFs for each target catchment were then estimated by sampling from each pooling catchment's signature PDF, with the number of samples proportional to the catchment's similarity weight, w_{it} (Figure 2):

$$w_{it} = \frac{\frac{1}{d_{it}}}{\sum_{i=1}^{N} \frac{1}{d_{it}}} \qquad (2)$$

This method of deriving the predicted signature distribution makes the assumption that the weights, based on the similarity quantified in the catchment descriptor space through (1), can be interpreted as the probability that each donor catchment is the "nearest neighbor" of the target catchment in terms of the streamflow signatures. That is, that the catchment descriptor similarity is proportional to the probability that the gauged catchment signatures are representative of the ungauged catchment signatures. Our method is equivalent to making multiple draws from the gauged catchment signature uncertainty distributions, and

each time selecting the "probable nearest neighbor" according to those probabilities. Regionalization uncertainty was thus represented by the weighted pooling group variability (equivalent to a nearest-neighbor method with uncertainty). This reflects the expectation that hydrologic similarity is always approximate [*Olden et al.*, 2012; *Oudin et al.*, 2010; *Reichl et al.*, 2009; *Wagener et al.*, 2007] and that there is no ideal donor catchment [*Beven*, 2000]. The method provides a direct visualization of the gauged versus the regionalization uncertainty components. We also present "gauging-uncertainty-only" regionalized results, where only the gauged uncertainty component was considered, to illustrate the effect of excluding uncertainty stemming from the regionalization procedure. These were obtained by randomly sampling values from each observed signature distribution in the *N* catchment pooling group and calculating the predicted value as a linear weighted combination [e.g., *Holmes et al.*, 2002].

4. Results

4.1. Uncertainty in Rating Curves and Flow Percentiles for Gauged Catchments
The rating curve estimation method succeeded in capturing the uncertainty in the gauging data over the diverse range of gauging data and rating curve characteristics for the 43 stations. Results from five stations with different flow magnitudes, rating curve sections and gauging data error characteristics are shown in Figure 3 as examples that illustrate the range of rating curve uncertainties across the data set. The first station (Figure 3a) is affected by seasonal weed growth at low flows and extrapolation uncertainty due to lack of high-flow gaugings. The second station (Figure 3b) has low uncertainty; it is well gauged for almost the whole flow range with little gauging data scatter. The third station (Figure 3c) was the "worst case" in the data set. It has a considerable gauging scatter for the whole flow range as a result of tidal influence and heavy weed growth, and the gauging authority has therefore downgraded it to a level-only station with a high-flow rating. The fourth station (Figure 3d) is a velocity-area station with large scatter at low flows and a gauged out-of-bank section with high flows reaching large magnitudes. In contrast, the fifth station (Figure 3e) has very low flow magnitudes, and a rating curve that appears to underestimate discharge for the whole flow range.

Figure 4 shows how rating curve uncertainties propagate to uncertainty in hourly flow percentiles, with the results for the five stations in Figure 3 highlighted in blue. The relative uncertainties were calculated with respect to the optimal rating curve from the MCMC estimation. The signature uncertainties result from the combination of the rating curve uncertainty distribution and the variability of the flow time series during the period. An extrapolated and uncertain high-flow part of a rating curve will therefore have more/less impact on the signature uncertainties depending on how often the highest gauged discharge was exceeded. This is illustrated in Figure 5 for a station which had one of the largest rating curve extrapolations in the data set (about 2 m). However, the large extrapolation mainly affects one 5 h peak-flow event that is more than twice the size of the other annual maximum flows. This demonstrates that the time series variability needs to be considered when determining the effect of the rating curve uncertainty on the flow signatures. The largest relative uncertainties occurred at high and low flows (Figure 4) where uncertainty in the rating-curve is normally highest (section 3.2), similar to results from a large Norwegian rating curve study [*Petersen-Overleir et al.*, 2009].

4.2. Uncertainty in Signature Values for Gauged Catchments
The relative uncertainty ranges were calculated as the half-width of the 5–95 percentile range for each signature and catchment. Then the 5, 50 and 95 percentiles of these ranges were calculated to illustrate typical values for catchments with low, medium and high uncertainty respectively (Table 2). The uncertainties were in general lowest (median values $\sim \pm 10$–15%) for Q_{BFI} measuring average groundwater contribution and Q_{MEAN} and Q_5 measuring average flow behavior, and highest (median values $\sim \pm 30$–40%) for signatures measuring low and high flow magnitude (e.g., Q_{99} and $Q_{0.01}$) and dynamics (Q_{LV} and Q_{HV}). This shows that careful consideration of data uncertainty is needed when using the latter type of signatures, e.g., in flood and drought studies. Signatures measuring flow variability across the time series (Q_{AC} and Q_{CV}) had uncertainty magnitudes similar to or somewhat higher than the average flow signatures, where Q_{AC} had generally higher uncertainty in the above-median range, but the lowest uncertainty in the below-median range. The four catchments with the highest BFIHOST values (>0.8) had the lowest maximum uncertainties across all the signatures. This might be because these catchments have a dampened flow variability where low and
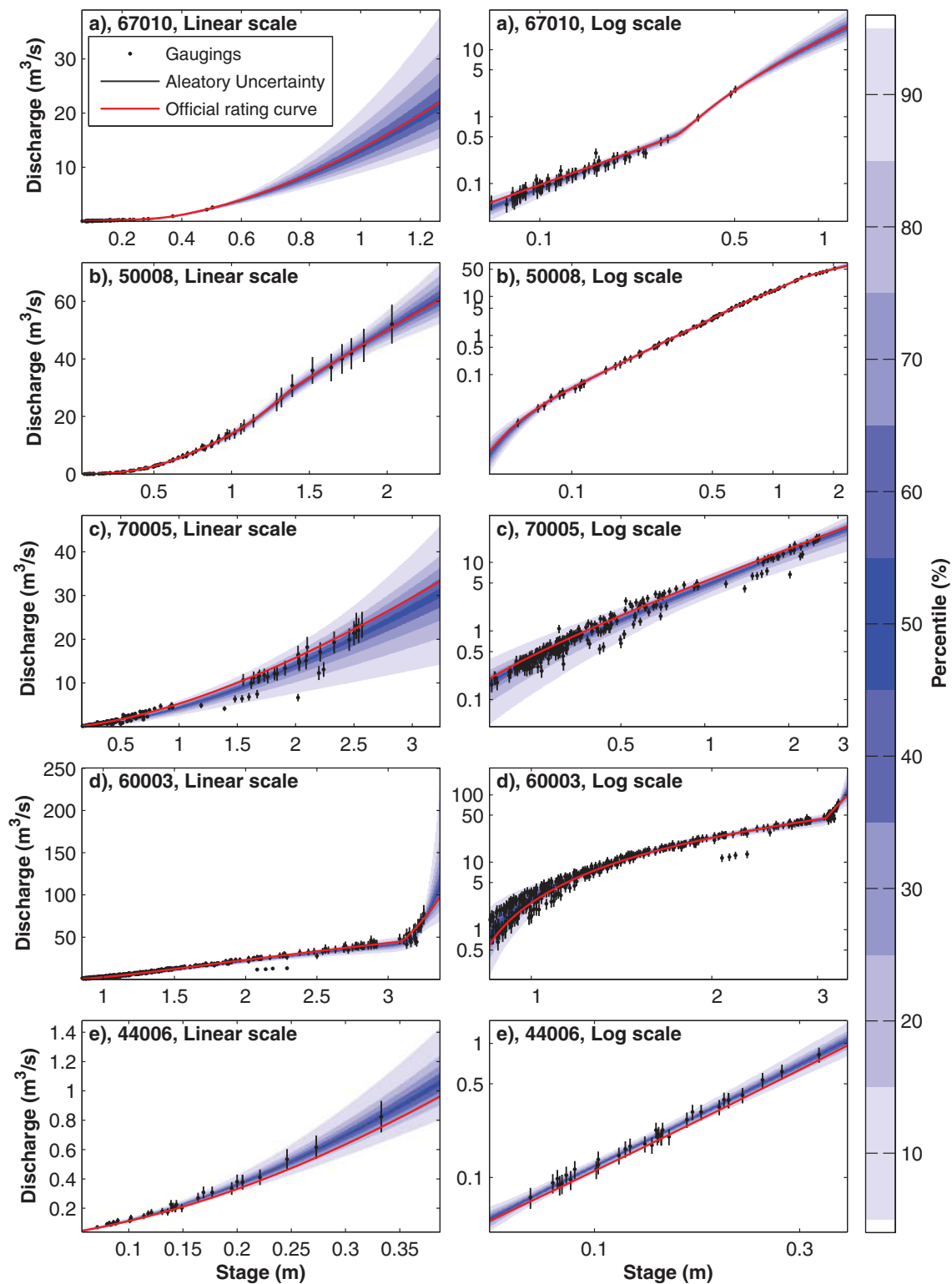
Figure 3. Gauging data with estimated aleatory discharge measurement uncertainties and rating curve sample quantiles for five catchments with diverse error characteristics shown in linear and log space. The quantiles are calculated for each stage interval from the distribution of flow values from all rating curve samples. Narrower colored intervals indicate higher density of rating curve samples and therefore higher probability density for the flow. The symbols (a) to (e) refer to the stations highlighted in blue in Figure 4, and the 5-digit numbers are the gauging station reference numbers.
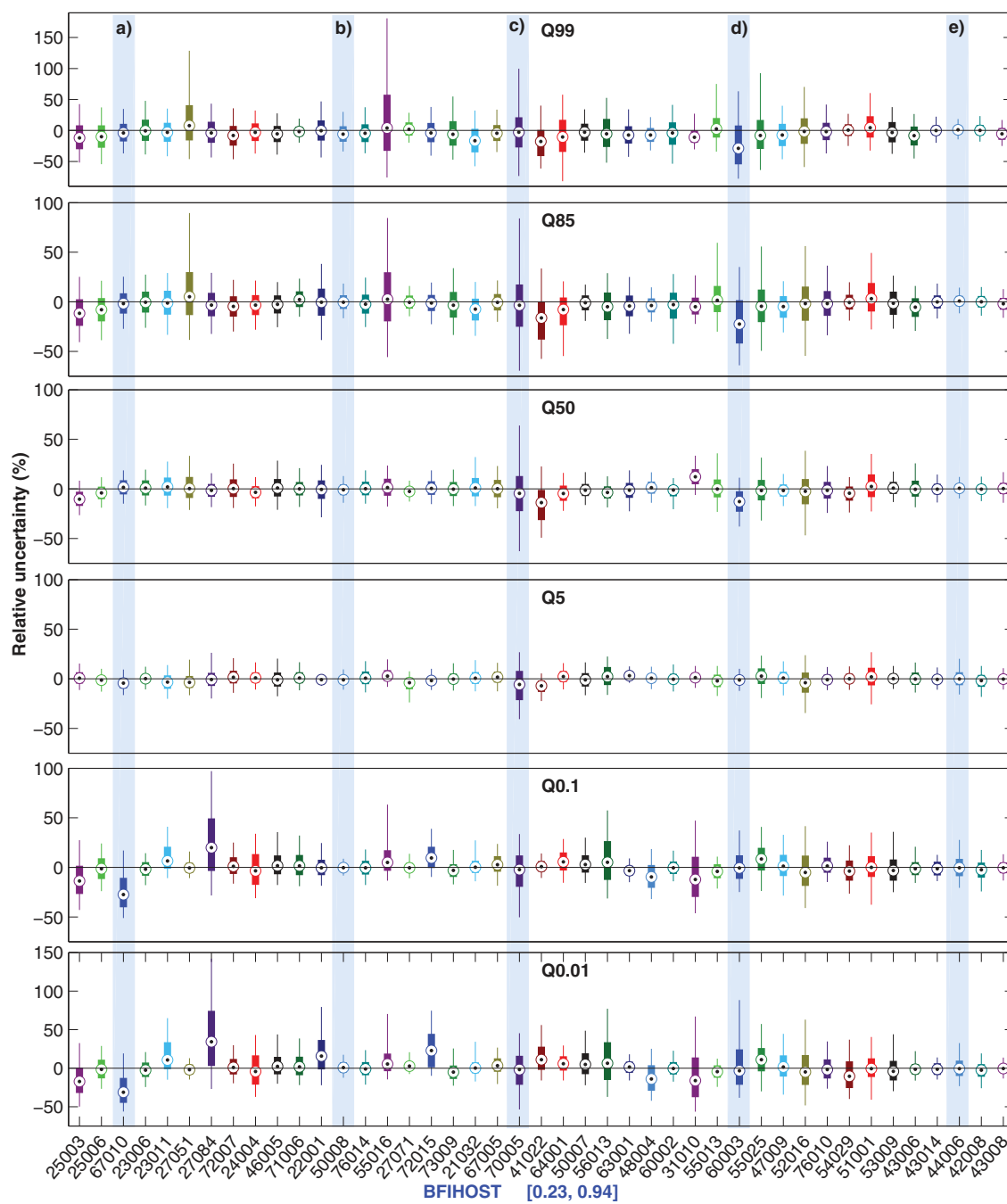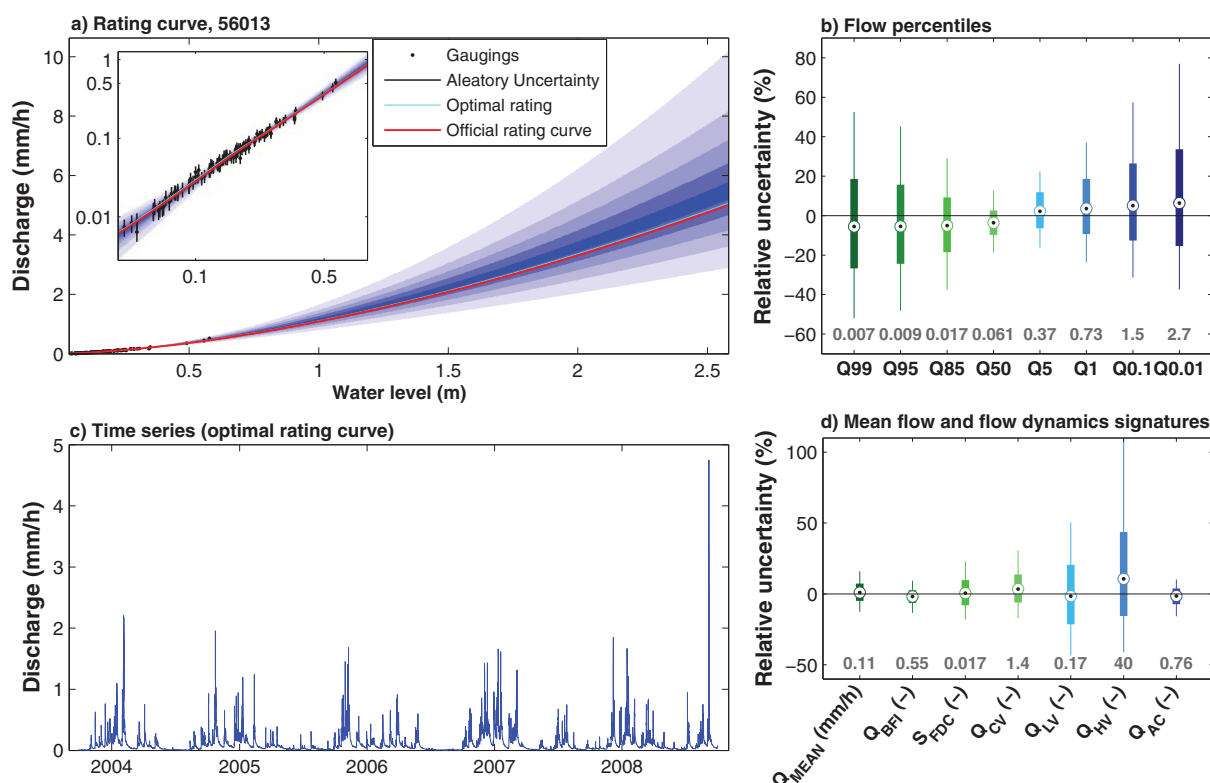
Figure 4. Uncertainty in hourly flow percentiles for all gauging stations sorted by catchment BFIHOST value from left to right. Note that the scale is different for the top and the bottom graph for visualization purposes. The stations (a) to (e) that are highlighted in blue correspond to the stations with rating curve uncertainties in Figure 3. The boxplot whiskers extend to the 5 and 95 percentiles, and the box covers the interquartile range.

high flows are less extreme, thus facilitating gauging of the whole flow range. In addition, the slow water-level dynamics would control the values of the Q_{AC} and Q_{BFI} signatures, rather than rating curve uncertainty.

The uncertainty magnitudes were correlated within the low/high flow signature groups, but poorly correlated between these two groups, which is expected given the different factors affecting the uncertainty of high and low flows (Table 2). The uncertainty magnitude for S_{FDC} was poorly correlated (correlation

Figure 5. (a) Rating curve in linear and log scale (inset plot), see explanation to Figure 3, and here the optimal curve is also shown (next to the official curve), (b) relative uncertainty in the flow percentiles, (c) hourly time series for the study period calculated from the optimal rating curve, and (d) relative uncertainty in mean flow and flow dynamics signatures. The signature values for the optimal rating curve are given in gray text at the bottom of the (b) and (d) figures, in (b) the unit is mm/h.

coefficient <0.5) against all the other signatures. We found that the uncertainty in this signature was high for most stations that had a breakpoint in the rating curve in the 33–66 percentile flow range.

The relative signature uncertainties were in some cases considerable even for signatures measuring average behavior, such as Q_{MEAN} (±30–40% for three catchments in Figure 6b). This demonstrates the systematic nature of discharge uncertainties caused by rating curve uncertainty; they do not cancel out when averaging over longer time periods. These large uncertainties in Q_{MEAN} occurred for the stations that had a large gauging scatter over the whole flow range (e.g., 70005 in Figure 3c that was affected by tides and heavy weed growth), and when there was large uncertainty in the flow range that contributed the largest flow volumes (Table 2). This was also illustrated by a high correlation between the magnitude of the Q_{MEAN} uncertainty range and that for Q_{50} and Q_5.

For cross-catchment comparisons, signature uncertainty is important when the absolute uncertainties overlap and therefore impede the interpretation of differences between catchments. This situation was in particular found for S_{FDC} (Figure 6c, note the contrast to Q_{MEAN} in Figure 6a), but would also be of concern for several other signatures (e.g., Q_{LV} and Q_{AC}, supporting information Figures S1–S5). We found no clear links between the type of gauging station and the uncertainty magnitudes; these are likely more controlled by local gauging conditions such as weed growth and backwater causing variability in the stage-discharge relation.

4.3. Uncertainty in Signature Predictions for Ungauged Catchments
To evaluate the signature predictions for ungauged catchments we compared the uncertainty distributions for the gauged signatures with those for the regionalized predictions. The comparisons were made 1) across all signatures and catchments using summary measures, 2) by analyzing differences between signatures and catchments, and 3) by analyzing the contributions of the different uncertainty sources to the predicted PDFs.

Table 2. Uncertainty Magnitudes for Different Signature Types and the Factors That Affect the Uncertainty Magnitude

| Signatures	Factors Affecting Uncertainty	Uncertainty Magnitudes[a]: Low, Medium, High
Q_{MEAN}, Q_5	Largest uncertainty where there was a large scatter for the whole flow range (e.g., Figure 3c), or when there was large uncertainty in the range of flows that contribute most of the total flow volume.	Q_{MEAN}: 9%, 12%, 29% Q_5: 10%, 14%, 27%
Q_{BFI}	Generally low uncertainty. The station with the largest uncertainty had a large scatter for the whole flow range.	Q_{BFI}: 2%, 9%, 17%
S_{FDC}	Uncertainty in the 33–66 percentile flow range, including change in rating curve (i.e., breakpoint in a multisection curve), in combination with uncertainty in the mean flow used for normalization.	S_{FDC}: 3%, 17%, 31%
Q_{LV}	High uncertainty in the low flow range	Q_{LV}: 13%, 31%, 59%
Q_{99}, Q_{95}, Q_{85}	Scatter in low flow gaugings (e.g., because of weed growth or riverbed change)	Q_{99}: 20%, 39%, 87% Q_{95}: 17%, 34%, 78% Q_{85}: 14%, 27%, 66%
Q_{50}	Largest uncertainty where there was gauging scatter in the median flow range	Q_{50}: 12%, 18%, 38%
Q_1, $Q_{0.1}$, $Q_{0.01}$	High flow uncertainty. Extrapolation above highest gauging and/or scatter in high flow gaugings (e.g., because of back-water effects or high gauging uncertainty).	Q_1: 10%, 17%, 32% $Q_{0.1}$: 12%, 24%, 45% $Q_{0.01}$: 14%, 32%, 62%
Q_{AC}, Q_{CV}, Q_{HV}	High flow uncertainty (see Q_1, $Q_{0.1}$, and $Q_{0.01}$ above), and flashiness of runoff response	Q_{AC}: 0.2%, 12%, 30% Q_{CV}: 9%, 16%, 37% Q_{HV}: 18%, 38%, 94%

[a]The low, medium and high values for the half widths of the 5–95 percentile uncertainty ranges were calculated as the 5, 50 and 95 percentiles from the distribution of half widths at all stations.

4.3.1. Performance of the Regionalization Method and Size of the Pooling Group

We first evaluated the performance of the regionalization method in a leave-one-out cross-validation by comparing the overlap between the 5–95 percentile ranges of the gauged and regionalized signature distributions. This comparison accounts for the uncertainty in the observed as well as the regionalized values and was made in terms of reliability (the overlapping range as a percentage of the gauged range) and precision (the overlap as a percentage of the regionalized range, see definitions in Figure 7c). Ideally predictions should have both high reliability and precision, but high reliability is more important than high precision. These measures have previously been used by, e.g., *Westerberg et al.* [2014] and are similar to those used by *Yadav et al.* [2007]. We also compared the distributions using two standard metrics, the Kolmogorov-Smirnov distance (K-S, the maximum absolute distance between the CDFs) and the earth mover's distance (EM, the sum of the absolute distances between the CDFs).

Reliability (across all signatures and catchments, Figure 7a) increased with the number of stations, while precision decreased (the regionalized distributions became wider as the pooling group size increased). The average reliability was high (>80%) even for a pooling group of 4–5 stations, and there was a marked increase in the reliability for poor performance stations (10th percentile) when the group size increased, with a smaller drop in overall precision. The K-S and the EM distance metrics showed similar results with a large initial decrease in the distances when the number of pooling catchments increased (Figure 7d). This was followed by a slight increase in the distances as more and more samples were taken from catchments with low similarity weights with often somewhat wider and flatter distributions as a result. The number of stations in the pooling group was chosen to be 10 as a trade-off between increase in reliability and decrease in precision across all the signatures. Increasing the number of stations means that the pooling group is less homogenous, thus allowing more reliable predictions for catchments with low hydrologic similarity. However, for catchments near the extremes of the signature distributions, this implies less precise predictions [*Burn*, 1990; *Holmes et al.*, 2002].

With 10 pooling catchments the average reliability (precision) varied between 83 and 93% (24–48%) for the different signatures, while the 10th percentile of the overall reliability (measuring poor performance) was 54%. The pooling group size is similar to that used by *Westerberg et al.* [2014] who used 8 donor catchments in regionalization of FDCs with uncertainty, while *Holmes et al.* [2002] also used 10 donors in deterministic regionalization of Q_{95} in the UK. Previous studies using catchment similarity as a basis for conceptual model
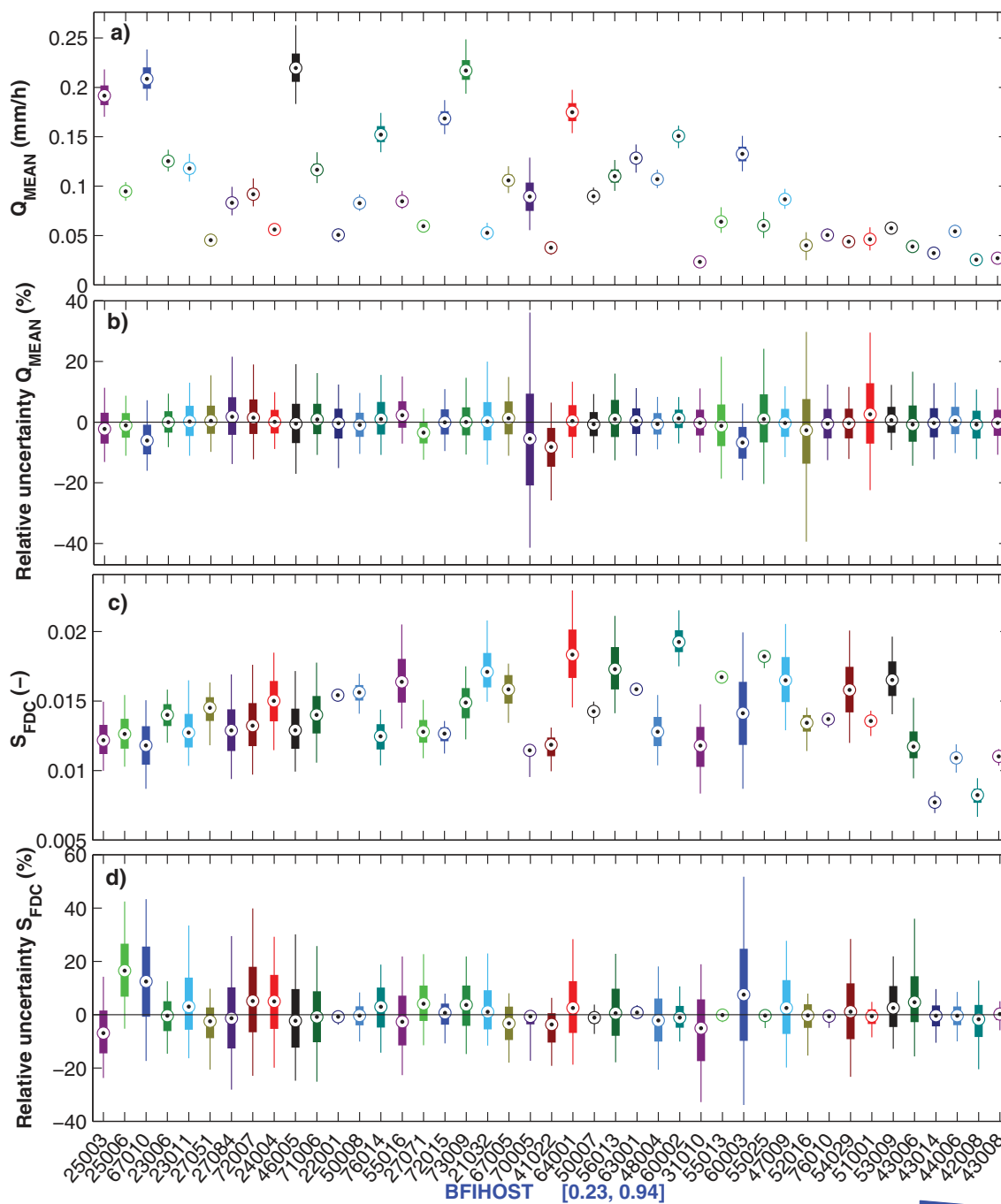
Figure 6. (a, b) Absolute and relative uncertainty in mean flow, and (c, d) slope of the flow duration curve for all catchments, the catchments are sorted from left to right according to their BFIHOST values.

regionalization, have similarly found that multiple donors are needed [*Reichl et al.*, 2009], e.g., *McIntyre et al.* [2005] also used 10 donors for weighted ensemble predictions of the PDM model in the UK.

For comparison we investigated the reliability and precision of a gauging-uncertainty-only regionalization where regionalization uncertainty was not included (section 3.3). This resulted in predicted regionalized distributions that were generally narrower than the observed distributions and much less reliable (average reliability 18% for 10 pooling stations). The gauging-uncertainty-only predicted distributions became narrower as more pooling stations were included in the weighted average predictions, leading to a decrease in
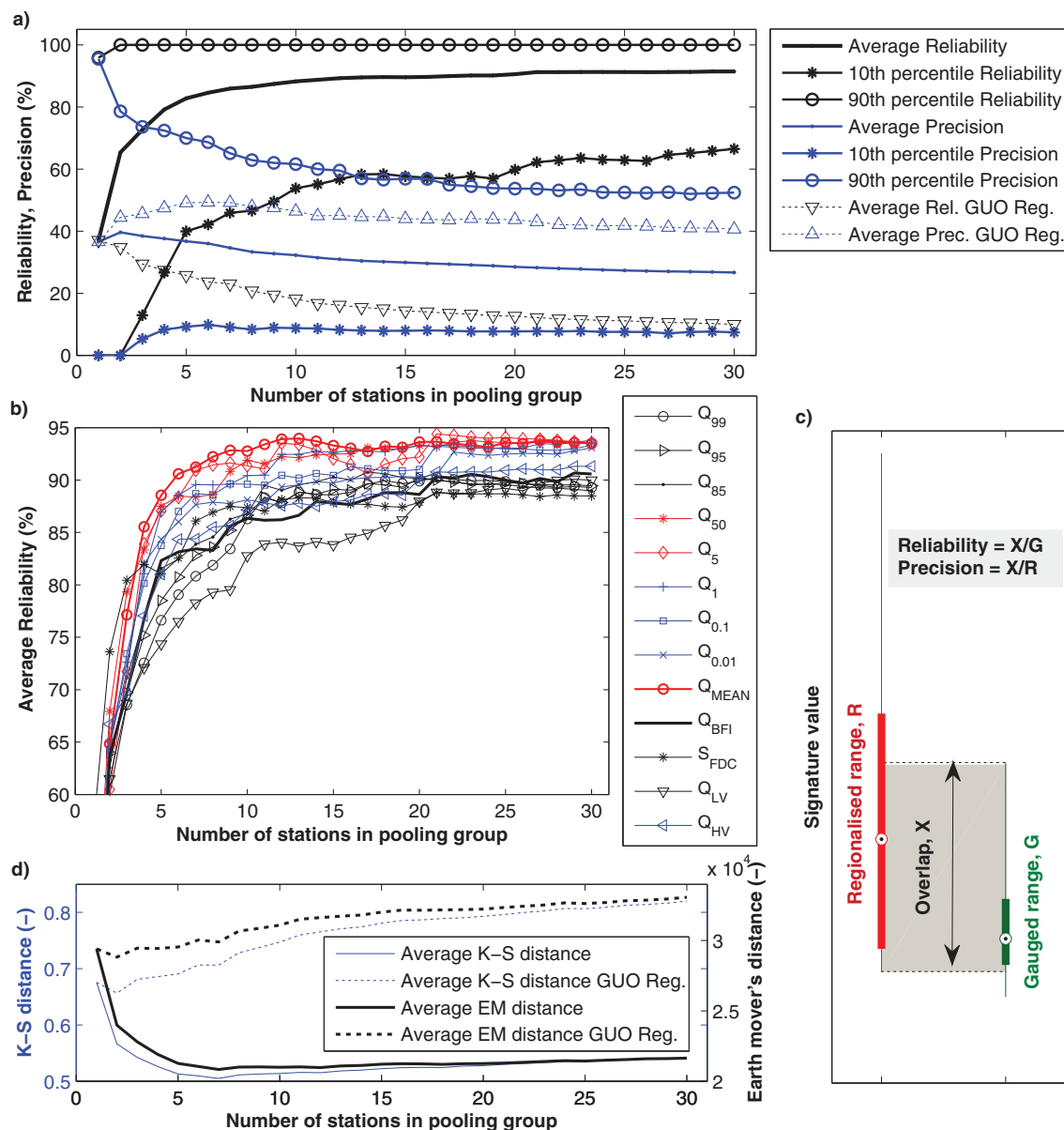
Figure 7. Performance of the regionalization method measured by reliability and precision plotted against the number of stations in the pooling group for (a) statistics of the values for all signatures and catchments, and (b) average reliability for all catchments for low flow (black), medium flow (red), and high flow (blue) signatures separately. (c) The calculation of the reliability and precision measures, where the boxplots extend to the 5 and 95 percentiles of the signature distributions. (d) The K-S distance and the earth mover's (EM) distance for all signatures and catchments, where GUO Reg. denotes the "gauging-uncertainty-only" regionalization.

average reliability with increase in pooling group size (Figure 7a). This was also seen in the K-S and EM distances that increased with the number of pooling stations (Figure 7d).

4.3.2. Differences in Regionalization Performance Between Signatures

The average flow signatures had the highest reliability followed by the high flow signatures, whereas the low flow signatures had the worst performance (Figure 7b). In general, the results were reliable except for the most extreme signature values that were not captured well (Figures 7 and 8). This is expected as the regionalization predictions were constrained to the observed variability among the pooling catchments. In the best performance range, there were 7 (2) catchments where all 15 signatures had a reliability of 90% (100%), and 31 (27) catchments where 10 or more signatures had a reliability of 90% (100%). In the poor performance range, there was one catchment for which 7 signatures had a reliability of 0, and an additional 9 catchments where 1–3 signatures had a reliability of 0. The station that had the poorest results had the
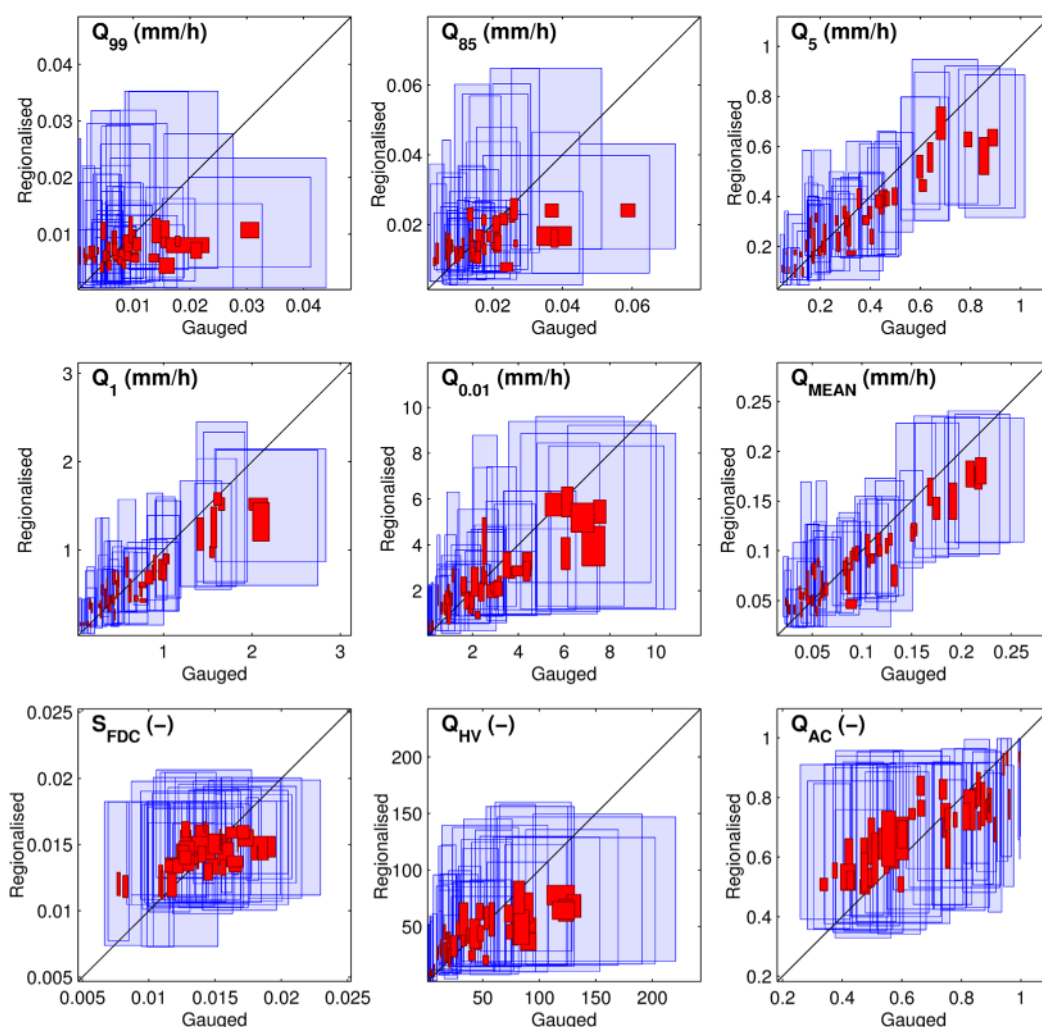
Figure 8. Gauged and regionalized signature uncertainty ranges (5–95 percentile range in blue, 40–60 percentile range in red) for each catchment. Flow distribution signatures are shown in the top and middle row and flow dynamics signatures are shown in the bottom row.

most extreme Q_{BFI} and high flow signature values, and a groundwater catchment that differed from the topographical catchment. Uncertainties in the contributing catchment area add uncertainty to both the signature values and their regionalization, but were not explicitly considered in this study.

Both the magnitude of the gauged uncertainties and the explanatory strength of the regionalization method need to be considered when evaluating the results (Figure 8). The average signatures (Q_{MEAN}, and Q_5) that both had low gauged uncertainty and strong correlations with the catchment characteristics, had the highest number of stations (>84%) with high reliability (>95%). The opposite was seen for S_{FDC} that had the lowest number of stations (49%) with high reliability. This was caused by high gauged uncertainty in relation to the signature range across the data set (Figures 6c and 8), in combination with poor correlation with the catchment characteristics. This poor correlation may partly be a result of the high gauged uncertainties, exemplifying how consideration of gauged uncertainty is important when interpreting the regionalization results.

When using the regionalized signature information, high reliability is important, but high precision is also desirable. Regionalized values with large uncertainty compared to the gauged range give little information about the regional signature variability (e.g., Q_{AC}, Figure 8). The signatures measuring flow dynamics (Table 1, Q_{AC}, Q_{HV}, etc.), had less precise regionalized results than those measuring flow distribution (Q_{MEAN} and the flow percentiles, Figure 8). Better results might be obtained for some flow dynamics signatures by

tailoring the regionalization to each signature separately, e.g., by giving a higher weight to the BFIHOST characteristic in predicting Q_{BFI}.

4.3.3. Contribution of the Different Uncertainty Sources

The regionalized distributions provide additional information about the success of the regionalization and the role of the different uncertainty sources (Figure 9). The color of the distributions illustrates the contributions from catchments with different hydrologic similarity to the target catchment. For example, where the whole distribution is light-blue (e.g., 25003 and 72015 in Figure 9), all the pooling group catchments have a low hydrologic similarity with the ungauged catchment, indicating that the regionalization is likely to be imprecise.

The shape of the distributions varied from unimodal (where gauged uncertainty dominates) to multimodal distributions (where regionalization uncertainty dominates). In general, the low flow signatures were the most unimodal (gauged uncertainty was high), and the average flow signatures were the most multimodal (gauged uncertainty was low). The widths of the distributions are also important. For example, if there is both a wide range of signature values and multiple separated peaks (e.g., Q_5 for 27051) this reflects a large variability within the pooling group and that the regionalization uncertainty dominates. This contrasts with other cases where the regionalized distribution is more compact and the gauged uncertainty dominates over the regionalization uncertainty (e.g., S_{FDC} for 27084). Where stations have different levels of gauged uncertainty this gave regionalized distributions with multiple peaks of different width (e.g., S_{FDC} for 60003 and Q_5 for 41022). These results clearly illustrate the risk of attributing differences between catchments that are a result of gauged uncertainty to differences in catchment behavior. In other cases, disregarding gauged uncertainty may lead to underestimation of signature variability between catchments. Neglecting the gauged uncertainties thus leads to overconditioning of the regionalization inference, i.e., the domain of possible predicted values is too constrained because the full range of possible data values is not taken into account.

A comparison with a gauging-uncertainty-only regionalized simulation (i.e., not including regionalization uncertainty, section 3.3) was also made (grey lines in Figure 9). These distributions were often narrower than the gauged uncertainties and often completely outside the gauged distributions, showing that the regionalization uncertainty needs to be considered to obtain reliable results. In addition, the optimal gauged and regionalized signature values (the optimal values from the MCMC rating curve estimation) are shown as black and grey dots respectively on the x-axes in the figures. These illustrate how overconditioned analyses that do not consider gauged uncertainty can be (e.g., $Q_{0.01}$ for 25003, Q_{CV} for 41022 and S_{FDC} for 27084).

5. Discussion

5.1. Rating Curve Uncertainty

Our rating curve uncertainty estimation method captured the uncertainty for diverse gauging data sets with different epistemic errors (e.g., weed growth or high-flow extrapolation), and different multisection power-law rating curves. Site-specific hydraulic information might reduce the uncertainties, but would require detailed information and investigation [e.g., *Le Coz et al.*, 2014], that is typically not available across large catchment data sets. We found it important to check the estimates for each station against available metadata, including information about nonideal conditions such as weed growth, backwater and out-of-bank flow ranges. The last was especially important to avoid unreliable extrapolation where there was insufficient information about the out-of-bank rating, and we excluded several such catchments. It is important to note that rating curve uncertainty may vary with time. This means that our estimates are not necessarily representative for other time periods as a result of station modifications, rating shifts, and/or different flow ranges (in particular out-of-bank flows). The rating curve estimations involved a considerable effort, suggesting that similar estimations for hundreds of stations are not easily achievable and that depth may need to be balanced with breath for large-sample hydrology also in terms of data uncertainty estimation [*Gupta et al.*, 2014].

5.2. Gauged Signature Uncertainty

The medium-level gauged signature uncertainty magnitudes (Table 2) we found were similar to the ±10–40% range found in two catchments by *Westerberg and McMillan* [2015]. However, we found a large
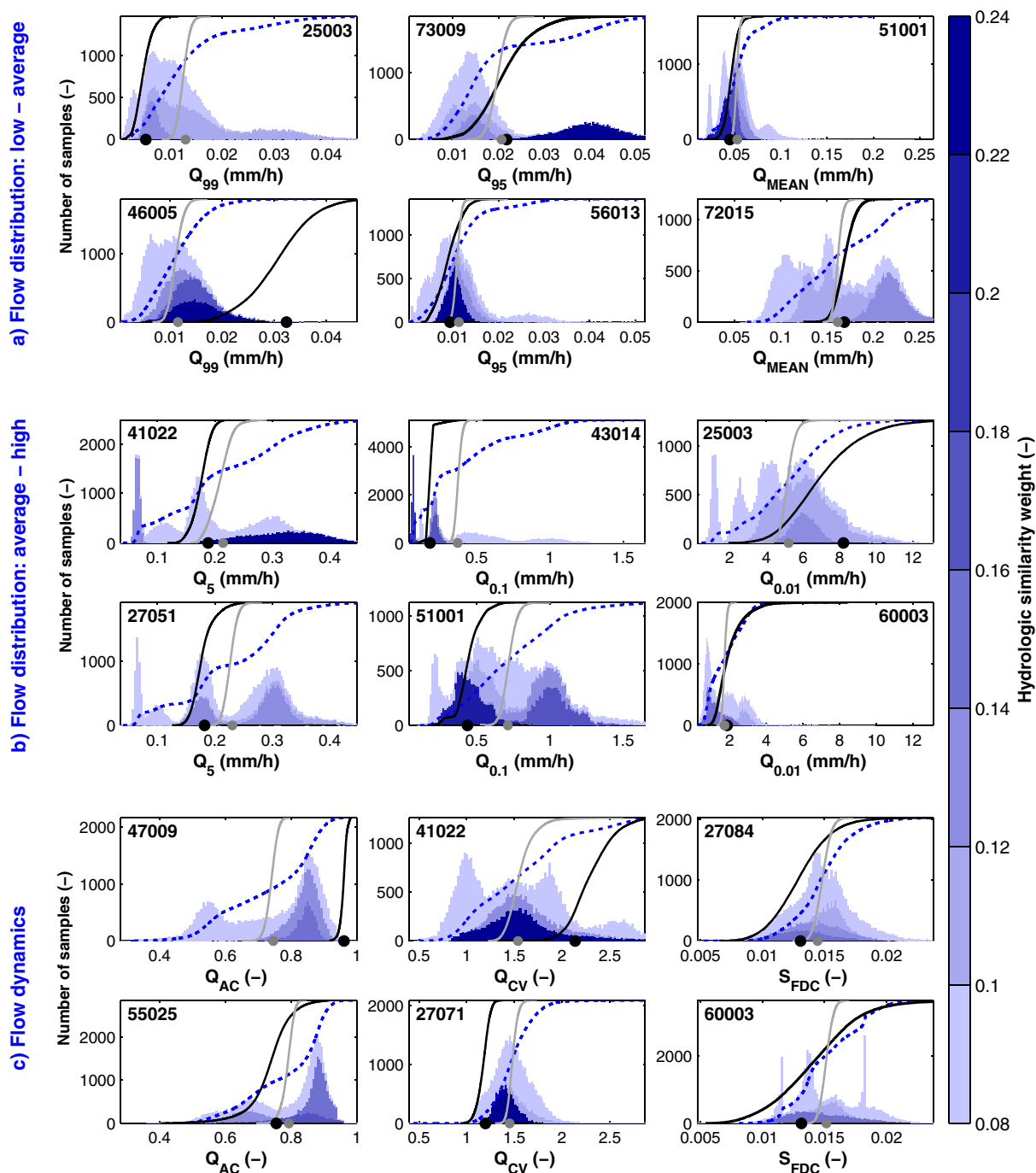
Figure 9. (a, b) Regionalized distributions for flow distribution and (c) flow dynamics signatures plotted as blue histograms, where the sampled values from each pooling group catchment are colored according to the hydrologic similarity weight with the target catchment. The CDFs of the gauged (black), regionalized (blue, dotted) and gauging-uncertainty-only regionalized (grey) uncertainty distributions are shown as lines. The deterministically regionalized (grey) and gauged (black) optimal values are shown as dots. Signatures of the same type are plotted with the same scale on the x-axis.

range in uncertainty magnitudes across the 43 catchments for all signatures, including those measuring average conditions (e.g., ±10–30% range in typical low and high uncertainty values for Q_{MEAN}). This large variability, and the absence of clear links between station types and uncertainty magnitudes, illustrate the importance of site-specific factors in controlling uncertainty. We found factors linked to high uncertainty in particular signature types, such as uncertainty in breakpoint location for multisection rating curves affecting S_{FDC}. Our analyses could be extended in the future to include catchments with greater human impacts, and

to include signatures requiring both rainfall and runoff data such as runoff ratio, which may have different uncertainty characteristics. Uncertainties in the water level time series may be important, but were not considered here other than removing suspect, flagged data and excluding stations with documented problems (e.g., blocked stilling-well intake pipes). In addition to data uncertainty, the data time step determines the information gained from signature values, as temporal averaging leads to loss of information about short-term response patterns [*Hrachowitz et al.*, 2013]. We used an hourly time scale as catchments in England and Wales are small, and we could see a clear loss of flow-peak information when averaging data to a daily time scale.

5.3. Signature Regionalization

The signatures quantifying low-flow magnitude and dynamics had the poorest regionalization. These signatures had high gauged uncertainty, the lowest correlations with the independent catchment descriptors, and they may be more susceptible to water level time series uncertainty (e.g., from moderate human impacts like sluices) than average and high flow signatures. As discussed by *Olden et al.* [2012], limitations in data and process understanding (e.g., surface water–groundwater connectivity) make it difficult to accurately characterize spatial variation in low flow magnitude and duration using catchment descriptors. However, including additional geologic data might improve prediction [*Holmes et al.*, 2002]. Similar to the Austrian study by *Viglione et al.* [2013], the average flow signatures had the best regionalization results. In contrast, they found poorer results for high than low flows, but they did not account for gauged uncertainty that is often large at high flows.

Our regionalization method enabled visualization of how uncertainties in the gauged data and the regionalization contribute to the predicted uncertainty; thus providing valuable additional information about the reliability of each prediction. For example, if the highest weighted catchments have distinct signature peaks, this shows that regionalization uncertainty is large and that the hydrologic similarity definition might be improved. The influence of the gauged uncertainties in the individual pooling catchments are also made explicit. Catchments with high gauged uncertainty contribute less information to the regionalization, but do not compromise the reliability of the predictions when their uncertainties are accounted for.

Our results clearly show that regionalization uncertainty is important: the gauging-uncertainty-only regionalized distributions were much less reliable and often completely outside the gauged distributions. We represented the regionalization uncertainty by the weighted pooling group variability, which is a simple and straight-forward method that enabled us to incorporate the site-specific gauged uncertainty distributions. However, the predicted signature distributions are conditional on both the hydrologic similarity weights and the assumption that the signature value in the target catchment can be selected according to the uncertain nearest neighbor method that we describe in section 3.3. This assumption could be explored in the future by developing methods to incorporate site-specific gauged uncertainty estimates in other regionalization methods, such as the Bayesian regression technique based on conditional probabilities, suggested by *Muller et al.* [1996]. Such methods could also enable prediction of "extreme" signature values outside the observed variability in the pooling data set (which here covered most of the variability among catchments with a natural flow regime in England and Wales, see section 2). Our method could be further developed by considering uncertainty in the definition of hydrologic similarity, both the type of measure/weighting used, and in the catchment characteristics data used to calculate it [*Burn*, 1990; *Reichl et al.*, 2009]. The latter was found to be important in snow-dominated catchments [*Arsenault and Brissette*, 2014] and may be difficult to estimate, e.g., uncertainty in catchment area where groundwater and surface water catchments differ, but sensitivity analyses could be used for a first investigation of their impact. For the characteristics we used, we expect elevation to have low uncertainty, uncertainty in catchment area to be important in flat and karstic areas, uncertainty in BFIHOST to be dependent on the underlying model, and uncertainty in mean annual precipitation to depend strongly on the number of rain gauges [*Westerberg and McMillan*, 2015]. The latter study found that uncertainty in mean annual precipitation was around $\pm 10\%$ in two catchments (50 and 135 km^2) using 1 rain gauge, which would not have a large effect on the relative weightings given the range of values in our data set (710–2400 mm/y, Figure 1). The choice and/or weighting of the catchment characteristics could also be tailored to the different signatures, e.g., using the most correlated characteristics for each signature, or enabling dynamic regions where the number of catchments depend on within-region similarity as suggested by *Holmes et al.* [2002].

5.4. Implications of Signature Uncertainty

Uncertainties in discharge data and derived signature values affect analyses such as catchment classification, eco-hydrological analyses, change detection and model calibration [e.g., *Juston et al.*, 2014; *Kennard et al.*, 2010; *McMillan et al.*, 2012]. Signatures have in particular been used to understand differences in catchment function, and to reduce predictive uncertainty in ungauged catchments [*Hrachowitz et al.*, 2013; *Wagener and Montanari*, 2011]. Understanding whether differences in signature values are a result of data uncertainty or a difference in catchment behavior is fundamental to comparative analyses [*Kennard et al.*, 2010], and when transferring information to ungauged catchments [*Hrachowitz et al.*, 2013; *Olden et al.*, 2012]. To understand the impact of data uncertainty the signature type needs to be considered, e.g., signatures describing the magnitude and dynamics of extreme flows are more susceptible to data uncertainty than those describing average behavior. *Viglione et al.* [2013] find that regionalization performance decreases with catchment area and discuss the generally poorer results found in arid regions [*Bloeschl et al.*, 2013] as a result of greater nonlinearity in runoff processes and larger space-time variability. In addition to these factors, discharge uncertainties likely play an important role: in arid catchments there are few high flow events with rapid flow variability, which impedes reliable gauging of the high-flow rating curve. Similarly, greater discharge uncertainty can be expected in small catchments because greater flow variability and shorter rainfall–runoff lag times impede reliable gauging of the full flow range. Understanding the sources of data uncertainty, at what conditions they are active, and how they affect different types of signatures and analyses is therefore important for reliable estimation of predictive uncertainty.

The gauged signature uncertainty distributions varied in size and shape between the stations and over the flow range in a site-specific way. This means that regression-based regionalization that assumes normally distributed errors [e.g., *Yadav et al.*, 2007], or fuzzy methods that use general discharge uncertainty estimates [e.g., *Westerberg et al.*, 2014], do not fully represent the nature of these errors. Using the regionalized signatures from this study for constraining model predictive uncertainty in ungauged basins (as in these previous studies) could therefore provide valuable insights. Valuable further information would also be gained by investigating the effects of rating curve uncertainty on signature analyses in other regions as they are determined by measurement practices in combination with natural conditions such as topography, catchment size, geology and climate (e.g., snow and ice conditions would introduce different uncertainties [*Hamilton and Moore*, 2012]).

6. Conclusions

This study has demonstrated how rating curve uncertainty propagates to uncertainty in hydrological signatures and their regionalization across a large set of catchments with diverse flow series characteristics and across multiple signature types. The gauged uncertainty varied with signature type and for each station local measurement conditions (e.g., weed growth, backwater, and station design) in combination with flow variability determined the uncertainty magnitudes. The catchments with the most dampened flow variability had the lowest signature uncertainties in our data set. The highest uncertainty magnitudes were found for signatures measuring high/low flow magnitude and dynamics (relative uncertainty ±30–40% as the median across all catchments). Signatures measuring average flow behavior had lower uncertainty (median relative uncertainty ±10–15%), but there was a large range in uncertainty magnitudes across the 43 catchments for all signatures. Our regionalization method allowed us to assess the role and relative magnitudes of the gauged and regionalized uncertainty sources in shaping the signature uncertainty distributions predicted for catchments treated as ungauged. We found that 1) if the gauged uncertainties were neglected there was a clear risk of overconditioning the regionalization inference, e.g., by attributing differences between catchments resulting from gauged uncertainty to differences in catchment behavior, and 2) the uncertainty in the regionalization results was lower for signatures measuring flow distribution than flow dynamics, as well as for average flows (and then high flows) compared to low flows.

Our results provide a strong demonstration of the need to investigate data uncertainties in analyses where signatures are used. Consideration of data uncertainty may often make these analyses more complex. But, as emphasized by *Juston et al.* [2013], moving beyond deterministic frameworks by recognizing these inherent data uncertainties increases our possibilities to draw robust conclusions about present and future hydrologic behavior – in gauged and ungauged catchments.

References

Almeida, S. L., N. Bulygina, N. McIntyre, T. Wagener, and W. Buytaert (2012), Predicting flows in ungauged catchments using correlated information sources, paper presented at Eleventh National Hydrology Symposium: Hydrology for a Changing World, Br. Hydrol. Soc., Dundee, U. K.

Archer, D., and M. Newson (2002), The use of indices of flow variability in assessing the hydrological and instream habitat impacts of upland afforestation and drainage, *J. Hydrol.*, *268*(1-4), 244–258, doi:10.1016/S0022-1694(02)00171-3.

Arsenault, R., and F. P. Brissette (2014), Continuous streamflow prediction in ungauged basins: The effects of equifinality and parameter set selection on uncertainty in regionalization approaches, *Water Resour. Res.*, *50*, 6135–6153, doi:10.1002/2013WR014898.

Bardossy, A. (2007), Calibration of hydrological model parameters for ungauged catchments, *Hydrol. Earth Syst. Sci.*, *11*(2), 703–710.

Beven, K. J. (2000), Uniqueness of place and process representations in hydrological modelling, *Hydrol. Earth Syst. Sci.*, *4*(2), 203–213.

Blazkova, S., and K. J. Beven (2009), A limits of acceptability approach to model evaluation and uncertainty estimation in flood frequency estimation by continuous simulation: Skalka catchment, Czech Republic, *Water Resour. Res.*, *45*, W00B16, doi:10.1029/2007WR006726.

Bloeschl, G., M. Sivapalan, T. Wagener, A. Viglione, and H. H. G. Savenije (Eds.) (2013), *Runoff Prediction in Ungauged Basins: Synthesis Across Processes, Places and Scales*, Cambridge Univ. Press, Cambridge, U. K.

Burn, D. H. (1990), Evaluation of regional flood frequency-analysis with a region of influence approach, *Water Resour. Res.*, *26*(10), 2257–2265.

Castellarin, A., G. Galeati, L. Brandimarte, A. Montanari, and A. Brath (2004), Regional flow-duration curves: Reliability for ungauged basins, *Adv. Water Resour.*, *27*(10), 953–965, doi:10.1016/j.advwatres.2004.08.005.

Castiglioni, S., L. Lombardi, E. Toth, A. Castellarin, and A. Montanari (2010), Calibration of rainfall-runoff models in ungauged basins: A regional maximum likelihood approach, *Adv. Water Resour.*, *33*(10), 1235–1242, doi: 10.1016/j.advwatres.2010.04.009.

Clausen, B., and B. J. F. Biggs (2000), 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.

Coxon, G., J. E. Freer, T. Wagener, N. Odoni, and M. P. Clark (2013), Diagnostic evaluation of multiple hypotheses of hydrological behaviour in a limits-of-acceptability framework for 24 UK catchments, *Hydrol. Processes*, *28*, 6135–6150, doi:10.1002/hyp.10096.

Coxon, G., J. Freer, I. K. Westerberg, T. Wagener, R. Woods, and P. J. Smith (2015), A novel framework for discharge uncertainty quantification applied to 500 UK gauging stations, *Water Resour. Res.*, *51*, 5531–5546, doi:10.1002/2014WR016532.

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

Gupta, H. V., T. Wagener, and Y. Q. Liu (2008), Reconciling theory with observations: Elements of a diagnostic approach to model evaluation, *Hydrol. Processes*, *22*(18), 3802–3813, doi: 10.1002/hyp.6989.

Gupta, H. V., C. Perrin, G. Bloschl, A. Montanari, R. Kumar, M. Clark, and V. Andreassian (2014), Large-sample hydrology: A need to balance depth with breadth, *Hydrol. Earth Syst. Sci.*, *18*(2), 463–477, doi:10.5194/hess-18-463-2014.

Gustard, A., A. Bullock, and J. M. Dixon (1992), Low flow estimation in the United Kingdom, *Rep. 108*, 88 pp, Inst. of Hydrol., Wallingford, U. K.

Hamilton, A. S., and R. D. Moore (2012), Quantifying uncertainty in streamflow records, *Can. Water Resour. J.*, *37*(1), 3–21, doi:10.4296/cwrj3701865.

Hannaford, J., M. G. R. Holmes, C. L. R. Laize, T. J. Marsh, and A. R. Young (2013), Evaluating hydrometric networks for prediction in ungauged basins: A new methodology and its application to England and Wales, *Hydrol. Res.*, *44*(3), 401–418.

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

Hrachowitz, M., et al. (2013), A decade of Predictions in Ungauged Basins (PUB): A review, *Hydrol. Sci. J.*, *58*(6), 1198–1255, doi:10.1080/02626667.2013.803183.

Hrachowitz, M., O. Fovet, L. Ruiz, T. Euser, S. Gharari, R. Nijzink, J. Freer, H. H. G. Savenije, and C. Gascuel-Odoux (2014), Process consistency in models: The importance of system signatures, expert knowledge, and process complexity, *Water Resour. Res.*, *50*, 7445–7469, doi: 10.1002/2014WR015484.

Institute of Hydrology (1999), *Flood Estimation Handbook*, 5 vols., Wallingford, U. K.

Jothityangkoon, C., M. Sivapalan, and D. L. Farmer (2001), Process controls of water balance variability in a large semi-arid catchment: Downward approach to hydrological model development, *J. Hydrol.*, *254*(1-4), 174–198, doi:10.1016/S0022-1694(01)00496-6.

Jowett, I. G., and M. J. Duncan (1990), Flow variability in New-Zealand rivers and its relationship to in-stream habitat and biota, *N. Z. J. Mar. Freshwater Res.*, *24*(3), 305–317.

Juston, J., A. Kauffeldt, B. Q. Montano, J. Seibert, K. J. Beven, and I. K. Westerberg (2013), Smiling in the rain: Seven reasons to be positive about uncertainty in hydrological modelling, *Hydrol. Processes*, *27*(7), 1117–1122, doi:10.1002/hyp.9625.

Juston, J., P.-E. Jansson, and D. Gustafsson (2014), Rating curve uncertainty and change detection in discharge time series: Case study with 44-year historic data from the Nyangores River, Kenya, *Hydrol. Processes*, *28*, 2509–2523, doi:10.1002/hyp.9786.

Kennard, M. J., S. J. Mackay, B. J. Pusey, J. D. Olden, and N. Marsh (2010), Quantifying uncertainty in estimation of hydrologic metrics for ecohydrological studies, *River Res. Appl.*, *26*(2), 137–156, doi:10.1002/rra.1249.

Kjeldsen, T. R., D. A. Jones, and D. G. Morris (2014), Using multiple donor sites for enhanced flood estimation in ungauged catchments, *Water Resour. Res.*, *50*, 6646–6657, doi:10.1002/2013WR015203.

Kuczera, G. (1996), Correlated rating curve error in flood frequency inference, *Water Resour. Res.*, *32*(7), 2119–2127.

Le Coz, J., B. Renard, L. Bonnifait, F. Branger, and R. Le Boursicaud (2014), Combining hydraulic knowledge and uncertain gaugings in the estimation of hydrometric rating curves: A Bayesian approach, *J. Hydrol.*, *509*, 573–587, doi:10.1016/j.jhydrol.2013.11.016.

Marsh, T. J., and J. Hannaford (Eds.) (2008), *UK Hydrometric Register*, 210 pp., Cent. for Ecol. & Hydrol., Wallingford.

McIntyre, N., H. Lee, H. Wheater, A. Young, and T. Wagener (2005), Ensemble predictions of runoff in ungauged catchments, *Water Resour. Res.*, *41*, W12434, doi:10.1029/2005WR004289.

McMillan, H. K., and I. K. Westerberg (2015), Rating curve estimation under epistemic uncertainty, *Hydrol. Processes*, *29*(7), 1873–1882, doi:10.1002/hyp.10419.

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

McMillan, H. K., T. Krueger, and J. Freer (2012), Benchmarking observational uncertainties for hydrology: Rainfall, river discharge and water quality, *Hydrol. Processes*, *26*(26), 4078–4111, doi:10.1002/hyp.9384.

Montanari, A., and E. Toth (2007), Calibration of hydrological models in the spectral domain: An opportunity for scarcely gauged basins?, *Water Resour. Res.*, *43*, W05434, doi:10.1029/2006WR005184.

Morlot, T., C. Perret, A. C. Favre, and J. Jalbert (2014), Dynamic rating curve assessment for hydrometric stations and computation of the associated uncertainties: Quality and station management indicators, *J. Hydrol.*, *517*, 173–186, doi:10.1016/j.jhydrol.2014.05.007.

Muller, P., A. Erkanli, and M. West (1996), Bayesian curve fitting using multivariate normal mixtures, *Biometrika*, *83*(1), 67–79, doi:10.1093/biomet/83.1.67

Nathan, R. J., and T. A. McMahon (1992), Estimating low flow characteristics in ungauged catchments, *Water Resour. Manage.*, *6*, 85–100.

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

Olden, J. D., M. J. Kennard, and B. J. Pusey (2012), A framework for hydrologic classification with a review of methodologies and applications in ecohydrology, *Ecohydrology*, *5*(4), 503–518, doi:10.1002/eco.251.

Oudin, L., A. Kay, V. Andreassian, and C. Perrin (2010), Are seemingly physically similar catchments truly hydrologically similar?, *Water Resour. Res.*, *46*, W11558, doi:10.1029/2009WR008887.

Petersen-Overleir, A., A. Soot, and T. Reitan (2009), Bayesian rating curve inference as a streamflow data quality assessment tool, *Water Resour. Manage.*, *23*(9), 1835–1842, doi:10.1007/s11269-008-9354-5.

Pugliese, A., A. Castellarin, and A. Brath (2014), Geostatistical prediction of flow-duration curves in an index-flow framework, *Hydrol. Earth Syst. Sci.*, *18*(9), 3801–3816, doi:10.5194/hess-18-3801-2014.

Refsgaard, J. C., and J. Knudsen (1996), Operational validation and intercomparison of different types of hydrological models, *Water Resour. Res.*, *32*(7), 2189–2202.

Reichl, J. P. C., A. W. Western, N. R. McIntyre, and F. H. S. Chiew (2009), Optimization of a similarity measure for estimating ungauged streamflow, *Water Resour. Res.*, *45*, W10423, doi:10.1029/2008WR007248.

Sawicz, K., T. Wagener, M. Sivapalan, P. A. Troch, and G. Carrillo (2011), 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.

Sawicz, K., C. Kelleher, T. Wagener, P. A. Troch, M. Sivapalan, and G. Carrillo (2014), Characterizing hydrologic change through catchment classification, *Hydrol Earth Syst. Sci.*, *18*, 273–285.

Sugawara, M. (1979), Automatic calibration of the tank model, *Hydrol. Sci. Bull.*, *24*(3), 375–388.

Viglione, A., J. Parajka, M. Rogger, J. L. Salinas, G. Laaha, M. Sivapalan, and G. Bloschl (2013), Comparative assessment of predictions in ungauged basins––Part 3: Runoff signatures in Austria, *Hydrol. Earth Syst. Sci.*, *17*(6), 2263–2279, doi:10.5194/hess-17-2263-2013.

Wagener, T., and A. Montanari (2011), Convergence of approaches towards reducing uncertainty in Predictions in Ungauged Basins (PUB), *Water Resour. Res.*, *47*, W06301, doi:10.1029/2010WR009469.

Wagener, T., M. Sivapalan, P. A. Troch, and R. Woods (2007), Catchment classification and hydrologic similarity, *Geogr. Compass*, *1*(4), 901–931.

Westerberg, I. K., and H. K. McMillan (2015), Uncertainty in hydrological signatures, *Hydrol. Earth Syst. Sci.*, *19*, 3951–3968.

Westerberg, I. K., J. L. Guerrero, P. M. Younger, K. J. Beven, J. Seibert, S. Halldin, J. E. Freer, and C. Y. Xu (2011), Calibration of hydrological models using flow-duration curves, *Hydrol. Earth Syst. Sci.*, *15*(7), 2205–2227, doi:10.5194/hess-15-2205-2011.

Westerberg, I. K., L. Gong, K. Beven, J. Seibert, A. Semedo, C. Y. Xu, and S. Halldin (2014), Regional water balance modelling using flow-duration curves with observational uncertainties, *Hydrol. Earth Syst. Sci.*, *18*, 2993–3013.

Winsemius, H. C., B. Schaefli, A. Montanari, and H. H. G. Savenije (2009), On the calibration of hydrological models in ungauged basins: A framework for integrating hard and soft hydrological information, *Water Resour. Res.*, *45*, W12422, doi:10.1029/2009wr007706.

Yadav, M., T. Wagener, and H. Gupta (2007), 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.

Yilmaz, K. K., H. V. Gupta, and T. Wagener (2008), A process-based diagnostic approach to model evaluation: Application to the NWS distributed hydrologic model, *Water Resour. Res.*, *44*, W09417, doi:10.1029/2007WR006716.