Hydrol. Earth Syst. Sci. Discuss., 12, 5389–5426, 2015
www.hydrol-earth-syst-sci-discuss.net/12/5389/2015/
doi:10.5194/hessd-12-5389-2015

Hydrology and Earth System Sciences
Discussions
Open Access

Accounting for dependencies in regionalized signatures for predictions in ungauged catchments

S. Almeida[1,2], N. Le Vine[2], N. McIntyre[2,3], T. Wagener[1,4], and W. Buytaert[2]

[1]Department of Civil Engineering, University of Bristol, Bristol, UK
[2]Department of Civil and Environmental Engineering, Imperial College London, London, UK
[3]Centre for Water in the Minerals Industry, Sustainable Minerals Institute, The University of Queensland, Brisbane, Australia
[4]Cabot Institute, University of Bristol, Bristol, UK

Received: 19 May 2015 – Accepted: 22 May 2015 – Published: 10 June 2015

Correspondence to: S. Almeida (susana.almeida@bristol.ac.uk)

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

Abstract

A recurrent problem in hydrology is the absence of streamflow data to calibrate rainfall-runoff models. A commonly used approach in such circumstances conditions model parameters on regionalized response signatures. While several different signatures are often available to be included in this process, an outstanding challenge is the selection of signatures that provide useful and complementary information. Different signatures do not necessarily provide independent information, and this has led to signatures being omitted or included on a subjective basis. This paper presents a method that accounts for the inter-signature error correlation structure so that regional information is neither neglected nor double-counted when multiple signatures are included. Using 84 catchments from the MOPEX database, observed signatures are regressed against physical and climatic catchment attributes. The derived relationships are then utilized to assess the joint probability distribution of the signature regionalization errors that is subsequently used in a Bayesian procedure to condition a rainfall-runoff model. The results show that the consideration of the inter-signature error structure may improve predictions when the error correlations are strong. However, other uncertainties such as model structure and observational error may outweigh the importance of these correlations. Further, these other uncertainties cause some signatures to appear repeatedly to be disinformative.

1 Introduction

In many areas of the world the absence of past observational streamflow time series to calibrate rainfall-runoff models limits the ability to apply such models reliably to predict streamflow and inform effective water resources management. Whilst a large and increasing number of regions across the world are insufficiently gauged (Mishra and Coulibaly, 2009), there are also many highly monitored catchments (Gupta et al., 2014). Transferring the knowledge gained in data-rich areas to ungauged catchments -

HESSD

12, 5389–5426, 2015

a process known as regionalization - offers a possible way of overcoming the absence of streamflow observations in data-scarce regions. Several techniques for transferring information are reported in the literature (for an overview of different methods used in continuous streamflow regionalization see He et al. (2011), Peel and Blöschl (2011), and Razavi and Coulibaly (2013), and for a recent comparative assessment of some of the most commonly used methods see Parajka et al. (2013)).

A commonly applied approach is to use response signatures (e.g. the runoff ratio and the base flow index), which can provide insight into the hydrological functional behavior of a catchment (Wagener et al., 2007). Response signatures are calculated from available system output or input-output time series for numerous gauged catchments with known catchment attributes, i.e. physiographic and/or meteorological attributes (e.g. drainage area, latitude and longitude, average annual temperature, average monthly precipitation, etc.). Subsequently, statistical models relating each response signature to a set of catchment attributes can be identified. Given the attributes of an ungauged catchment, the signatures for the ungauged location can then be estimated using the derived statistical models. Numerous regional models of this type can be found in the literature (e.g. Boorman et al., 1995). These regionalized signatures can be used to constrain the prior range of streamflow simulations generated using a pre-selected rainfall-runoff model structure and hence restrict the model parameter space (Yadav et al., 2007; Zhang et al., 2008; Bulygina et al., 2009; Castiglioni et al., 2010).

Different ways of incorporating the regionalized information into a catchment model have been suggested in the literature. This includes set-theoretic approaches (e.g. Yadav et al., 2007; Winsemius et al., 2009) and more formal Bayesian data assimilation frameworks (e.g. Bulygina et al., 2009, 2011; Castiglioni et al., 2010; Singh et al., 2011). Where probability distributions characterizing regionalization quality have been estimated, a Bayesian conditioning procedure is one of the possibilities (Bulygina et al., 2009, 2011). This provides a framework for combining prior knowledge with the regionalized data and/or other sources of information (e.g. small scale physics-based knowledge and hydrological measurements as in Bulygina et al., 2012),which has the potential to formally encompass the nature of the errors arising from the regionalization.

Conditioning a rainfall-runoff model on multiple independent signatures would reflect a spectrum of processes and in principle lead to an accurate prediction of flow time series (Parajka et al., 2013). However, regionalized signatures have correlated errors, for example if the signatures have been estimated using a common dataset of catchment attributes or using the same hydro-climatic data; and in general the correlations are expected to be stronger for pairs of signatures that represent similar functional behaviors of the catchment. This raises the questions of, not only how many and which signatures should be used, but also how to avoid double-counting of the information in signatures with correlated error distributions. Previous applications have tended to use a small number of signatures (e.g. Bulygina et al., 2009, 2011) and/or have tended to select signatures that are considered to be independent (e.g. Yadav et al., 2007). When multiple signatures are used, the correlations between the errors in the different sources of information are commonly disregarded (e.g. Bulygina et al., 2012). To make better use of information in available sets of signatures, a formal way of combining them so that information is neither double-counted nor neglected is required. Using formal methods to include autocorrelated data errors in model calibration is well-researched (e.g. Sorooshian and Dracup, 1980); an application of comparable methods in the regionalization context will allow making more formal and rigorous assessments of the value of correlated information sources.

Formally, in a Bayesian context, it is necessary to distinguish between *correlated signatures* and *correlated signature errors*. It is the correlation between the errors that should be accounted for in the likelihood function to avoid double-counting of information. It is possible to have two highly correlated signatures that are derived from independent information sources and therefore they have uncorrelated errors. In that case it would be valid to include both signatures in the likelihood function without accounting for correlation. This principle is well established when considering Bayesian calibration to a time series of flow observations, where flow values are typically strongly autocor-

HESSD
12, 5389–5426, 2015

related - but it is the observation error autocorrelation that is relevant to the likelihood function derivation (e.g. Sorooshian and Dracup, 1980). The same principle applies to adopting signatures as the observations. In the case study below, the signatures are derived from a common dataset using a common approach, so in practice the signature correlations are comparable to the signature error correlations; nevertheless for the sake of formality, we use the term *signature error correlations* (or *covariance*).

In this paper we introduce and test a method that considers multiple regionalized signatures, explicitly accounting for the signature error correlations. By formally accounting for the error covariance, we hypothesize that accuracy of flow predictions will generally improve and a greater number of signatures can usefully be included without introducing avoidable bias related to the duplication of information. The objective is thus to explore how to get fuller value out of a set of regionalized information than has been achieved in past applications. The method is applied to a set of 84 United States catchments with a broad range of hydro-meteorological characteristics, obtained from the Model Parameter Estimation Experiment (MOPEX) dataset (Duan et al., 2006; Schaake et al., 2006). The impact of signature error covariance is assessed using pairs of signatures to condition a rainfall-runoff model. Along with the real data, synthetic streamflow data are used to isolate the effect of model structural error. Further, the model is conditioned on a variable number of regionalized signatures to evaluate whether an increasing number of signatures is justifiable when formally accounting for the error covariance.

2 Method

2.1 Bayesian method for signature assimilation

Using a simple least-squares regression, observed signatures of catchments' functional responses are related to physical and climatic attributes of the catchments. Assuming that the same catchment attributes are available for an ungauged location, it

HESSD
12, 5389–5426, 2015

is possible to obtain an estimate of the set of signatures for the location. Further, the parametric distribution of regression errors can be directly translated to a response signature(s) likelihood function. The likelihood function can then be used to update the prior available knowledge about model parameters via Bayes' law, which is expressed as

$$p(\mathbf{\Theta}|s^*, \mathbf{I}, M) = \frac{L(s(\mathbf{\Theta})|s^*, \mathbf{I}, M) \times p(\mathbf{\Theta}|\mathbf{I}, M)}{p(s^*|\mathbf{I}, M)} \tag{1}$$

where, for one catchment, s^* represents the regionalized response signature(s); $p(\mathbf{\Theta}|\mathbf{I}, M)$ is the prior distribution of parameters $\mathbf{\Theta}$ for a model structure M and inputs $\mathbf{I}$; $L(s(\mathbf{\Theta})|s^*, \mathbf{I}, M)$ is the likelihood function of the modeled response signature(s) $s(\mathbf{\Theta})$ given s^*, $\mathbf{I}$ and M; $p(s^*|\mathbf{I}, M)$ is the marginal density of s^*; and $p(\mathbf{\Theta}|s^*, \mathbf{I}, M)$ is the posterior distribution of $\mathbf{\Theta}$ given s^*, $\mathbf{I}$ and M. For the purpose of this paper, M is selected in advance and considered to be fixed (as it is the common practice in regionalization studies, Wagener and Montanari, 2011), as is $\mathbf{I}$ for any one catchment, and so both these terms are dropped from (Eq. 1) for convenience, resulting in

$$p(\mathbf{\Theta}|s^*) = \frac{L(s(\mathbf{\Theta})|s^*) \times p(\mathbf{\Theta})}{p(s^*)} \tag{2}$$

Parameter sets are then sampled from the parameter posterior to allow an ensemble of rainfall-runoff simulations and a posterior distribution of flow at each time-step to be estimated and evaluated against observed flow. This can be repeated using different sets of signatures and different assumptions about their error correlations.

2.2 Prior distribution and likelihood function

2.2.1 Prior distribution

To apply Bayes' law (Eq. 2) it is necessary to specify the likelihood function ($L(s(\mathbf{\Theta})|s^*)$ in Eq. 2) and the prior distribution ($p(\mathbf{\Theta})$ in Eq. 2). The prior is defined so that it reflects

HESSD
12, 5389–5426, 2015

our initial lack of knowledge. We follow Almeida et al. (2013) and express our initial state of indifference in the same space that we seek information about – the signature space. This is different from the common assumption of a uniform distribution on the rainfall-runoff model parameter space (e.g. Yadav et al., 2007), and is shown to avoid the problem of bias in predictions (Almeida et al., 2013). As it is usually not possible to sample directly from the uniform signature prior distributions, an importance sampling is utilized to approximate the distributions numerically (Doucet et al., 2000).

2.2.2 Likelihood function approximation

The likelihood functions are defined using joint distributions of respective signature errors obtained from the regionalization model. Errors introduced by the regionalization procedure may come from at least five sources. First, errors are introduced by the fact that the regression model is estimated using a specific sample of catchments rather than the entire population; second, differences may exist between the observed and the true value of the response signature due, for example, to factors such as the discharge record length and time period of record used in the computation (Kennard et al., 2010); third, errors are present due to errors in the catchment properties data; fourth, errors exist due to the incomplete set of catchment properties used as explanatory variables in the regression equations; and, fifth, they exist due to the assumed linear regression structure. It is assumed that the total error model for the regionalized signature(s) s^* can be estimated using the following procedure:

1. Considering all available gauged catchments stepwise regression is applied to each signature independently to determine the predictors to include. The predictors are then fixed for the remaining steps.

2. Considering all available gauged catchments, one catchment is left out and the remaining are used in the fitting of the regression models for each signature.

3. The regression models obtained in Step 2 are used to estimate the signature values for the omitted catchment.

4. The error for each signature is calculated for the omitted catchment by comparing the regionalized and observed signature values.

5. The process is repeated for all catchments.

6. A parametric joint probability distribution is fitted to all the computed errors. Furthermore, the errors are tested for independence that allows (approximately) factorizing a joint distribution into a product of marginal distributions.

The resultant error distribution defines the likelihood function L in Eq. (2). The main assumption here is that the potentially complex nature of errors in the set of signature values can be usefully represented by the fitted error distributions.

2.2.3 Synthetic case and likelihood functions

To avoid masking the potential value of the regionalized signatures with model structure and observational errors, a "perfect model" is first employed. This involves using the pre-selected rainfall-runoff model and the observed forcing data to generate the "observed" catchment signatures. The Nash–Sutcliffe criteria (NSE) (Nash and Sutcliffe, 1970) optimal parameter set is taken to generate a "perfect model" streamflow time series for each catchment. To produce regionalized signature analogues in this case, two types of imposed errors are introduced to these "observed" signatures. The first error type is characterized by a range of standard deviations (1, 5, 10 and 20 % of the signature value range observed over all catchments used in this study) and a range of inter-signature error correlations (Pearson correlation coefficients equal to 0, 0.25, 0.50, 0.75 and 0.90). This allows the sensitivity of the results to the regionalization quality and the regionalization errors' correlations to be evaluated. The second error type is set to be equal to the observation-based likelihood function (Sect. 2.2.2). These

HESSD
12, 5389–5426, 2015

error structures are the likelihoods used in Eq. (2) for the synthetic case when flows are generated by a "perfect model".

2.3 Case study and rainfall-runoff model

2.3.1 Study catchments

A set of 84 medium sized United States catchments (242 to 8657 km^2) from the MOPEX database (Schaake et al., 2006; Duan et al., 2006), for which a variety of regional response signature models have been determined in Almeida et al. (2012), are used to test the method proposed in this paper. Use of more catchments from the MOPEX database would require different regionalization equations due to changing process controls, and would be unnecessary given that the focus of the study is on signature error correlations given a regionalization model. For more details on the motivation for choosing these specific 84 catchments the reader is referred to Almeida et al. (2012) and Almeida (2014).

The 84 catchments are hydrologically varied with a selection of properties summarized in Table 1. Daily time series for the period from 1 October 1949 to 30 September 1959 are employed. As highlighted in Almeida et al. (2012), these 10 years of data, representing only a subset of all the data available, are assumed to be of sufficient length to capture climatic variability, but short enough to avoid effects of long-term climatic trends (Sawicz et al., 2011).

2.3.2 Response signatures

Five response signatures are considered: runoff ratio (RR), base flow index (BFI), streamflow elasticity (SE), slope of flow duration curve (SFDC) and high pulse count (HPC) (Table 1). This specific subset of signatures is selected to cover a wide range of different qualities of regionalized information, and also to ensure that some signature errors are largely uncorrelated, whilst others are strongly correlated (see also Sect. 3.1).

RR reflects the amount of precipitation that becomes streamflow over a certain area and time. It is determined as the ratio of catchment's outlet streamflow and catchment average precipitation over the 10 years used in this study. BFI gives the proportion of streamflow that is considered to be base flow. A simple one-parameter single-pass digital filter method is used to derive BFI (Arnold and Allen, 1999). SE provides a measure of the sensitivity of streamflow to changes in precipitation (Sankarasubramanian et al., 2001). It is calculated as a median of the inter-annual variation in total annual streamflow to the inter-annual variation in total annual precipitation ratios normalized by the long-term runoff ratio (Sawicz et al., 2011; Sankarasubramanian et al., 2001). SFDC gives an indication of the streamflow variability and is calculated as the slope of the flow duration curve between the 33 and 66 % flow exceedance values in a semi-log scale (Sawicz et al., 2011). HPC reflects aspects of the high flow regime and catchment flashiness, and is calculated as the average number of events per year that exceed three times the median daily flow (Clausen and Biggs, 2000; Yadav et al., 2007).

2.3.3 Rainfall-runoff model choice

The probability distributed moisture (PDM) model (Moore, 2007) together with two parallel linear routing stores and a simple snow model (Hock, 2003) is selected with two major motivations (for a detailed description of the model see Kollat et al., 2012; Almeida, 2014). First, this type of model has been shown to have a suitable complexity for modelling daily rainfall-runoff over a large sample of the MOPEX catchments (Wagener and McIntyre, 2012). Second, the model has been successfully applied in other regionalization studies across a wide range of climate and physiographic conditions, for example Calver et al. (1999), Lamb and Kay (2004), McIntyre et al. (2005), Young (2006), and De Vleeschouwer and Pauwels (2013). Even though other model structures may be better suited for some specific catchments, it is prohibitively difficult to vary model structure between catchments and no single model structure will ever

HESSD
12, 5389–5426, 2015

be best for all catchments (Lidén and Harlin, 2000; Clark et al., 2008; van Werkhoven et al., 2008). Consequently, the selected model structure is believed to be a sufficient choice for the purposes of this paper. Most importantly, the general framework is independent of the rainfall-runoff model choice.

2.4 Posterior distribution and performance assessment

Employing Bayes' law (Eq. 2), the rainfall-runoff model is conditioned on different combinations of signatures: (1) assuming independence between the signature regionalization errors (setting the correlation values to zero in the joint probability function); and (2) accounting for the inter-signature error correlations (using the estimated covariance in the joint probability function).

Two metrics are used to assess the effectiveness of the parameter conditioning procedure: (1) the Bayes factor (Jeffreys, 1961) to assess convergence of the parameter posteriors to known parameter values; (2) the probabilistic Nash–Sutcliffe efficiency (Bulygina et al., 2009) to assess convergence of the flow ensembles to the observed flows.

The Bayes factor BF is defined as the ratio between two marginal distributions of the data $\boldsymbol{y}$ (e.g. observed streamflow time series) for two competing hypotheses (H_1 and H_2) (Kass and Raftery, 1995) (more detail is given in Appendix A):

$$\mathrm{BF} = \frac{p(\boldsymbol{y}|H_1)}{p(\boldsymbol{y}|H_2)} \tag{3}$$

Thus, to test the impact of representing the error correlations, the hypothesis H_1 corresponds to the inter-signature errors being treated as correlated, while the hypothesis H_2 corresponds to the inter-signature errors assumed to be independent. If the resulting Bayes factor is greater than 1, there is more support for hypothesis H_1, and the inter-signature error correlation is worth considering.

When using synthetic streamflow data ("perfect model" approach), with the streamflow time series generated by a pre-selected parameter set, $p(\boldsymbol{y}|H)$ in Eq. (3) can be

HESSD
12, 5389–5426, 2015

seen as either the posterior probability of the known observed streamflow time series under hypothesis H or the probability of the known parameter set that generated that particular flow time series under hypothesis H. As in a "perfect model" approach there is no observational error, $p(\boldsymbol{y}|H)$ is the probability estimated for the known value of the parameter set that generated the observed streamflow under each of the hypotheses H_1 and H_2. Since there is no known parameter value corresponding to the real data, the application of the Bayes factor is less useful in this situation. In this case, defining $\boldsymbol{y}$ as an NSE-optimal parameter set allows an indication of the relative degree of convergence around the chosen point.

The probabilistic Nash–Sutcliffe efficiency NSEprob (Bulygina et al., 2009) is a probabilistic analogue of the traditional Nash-Sutcliffe efficiency coefficient (Nash and Sutcliffe, 1970), and allows both prediction accuracy and precision to be summarized by a single statistic. The incremental improvement in the NSEprob can be used to measure the value of adding signatures into the conditioning or otherwise changing the likelihood function. An NSEprob of 1 indicates a perfect fit, i.e. the results are both accurate and precise.

For model validation, we use a jack-knife approach (or leave-one-out strategy), commonly employed in regionalization studies (e.g. Merz and Blöschl, 2004; Shu and Ouarda, 2012). One catchment at a time is removed as a test "ungauged" catchment and the remaining gauged catchments are used to support the regionalization process, including Steps 2–6 listed in Sect. 2.2.2. The procedure is repeated for each of the available catchments.

3 Results and discussion

3.1 Regionalized signature errors and likelihood functions

The regionalization error probability distributions (that define the likelihoods) are generated following Steps 2 to 6 in Sect. 2.2.2 and are shown in Fig. 1. The marginal error distributions, shown on the Fig. 1 diagonal, are approximated using histograms, and parameters of normal distributions are fitted using the method of moments. The univariate Kolmogorov-Smirnov test shows that the marginal distribution normality cannot be rejected at the 95 % confidence level for each of the five signatures. The off-diagonal shows the regionalization errors for different signature pairs (lower off-diagonal), the corresponding correlation coefficient values and their statistical significance (upper off-diagonal). The joint error distributions are approximated using multivariate normal distributions that are fitted using estimates of the marginal normal distribution parameters and the inter-signature error correlations. These marginal and joint distributions define the likelihood functions in Eq. (2). Note that Fig. 1 represents the regionalization errors based on all 84 catchments. Meanwhile, the jack-knife procedure (see Sect. 2.4) utilized in the performance assessment employs only 83 catchments at a time.

3.2 The impact of inter-signature error correlations (Pairs of signatures)

This section considers the role of inter-signature error correlation on model parameter estimation when pairs of signatures are used. First, different imposed error variances and correlations together with synthetic streamflow data are employed to test the impact of inter-signature error correlation without the impact of model structural error. Then, the results obtained using the observation-based error structure, for both synthetic and observed data streamflow, are analyzed.

3.2.1 Synthetic streamflow data (Imposed likelihoods)

Synthetic streamflow data are generated as described in Sect. 2.2.3, and the imposed likelihood functions are defined as described in Sect. 2.2.3. The imposed likelihoods are considered to have standard deviations equal to 1, 5, 10, 20 % of the signature value range observed over all catchments. A comparison of the imposed error structures under the different levels of variance and the observed error structure is given in

HESSD
12, 5389–5426, 2015

Table 2. Furthermore, different inter-signature error correlations are also tested, namely 0 (linear independence), 0.25, 0.50, 0.75 and 0.90.

Ten possible pairs of the five response signatures are used in parameter conditioning, and the median Bayes factor, calculated over the 84 MOPEX catchments, is calculated for each pair. The Bayes factor (Eq. 3) compares the two following hypothesis: H_1) the inter-signature error correlation is to be taken into account, and H_2) the errors between the different sources of information can be assumed independent. The Bayes factor is found to be relatively insensitive to the selection of response signature pairs (Kruskal–Wallis test). Table 3 summarizes the 95 % pooled confidence intervals for the median Bayes factor across all catchments and across all 10 signature pairs, for each choice of the likelihood (i.e. 20 likelihoods). This provides reference values indicative of the error interdependency importance in model regionalization depending on the signature pair correlations and marginal distribution variances. As it would be expected, the median Bayes factor is equal to 1 when signatures errors are not correlated (i.e. $\rho = 0$). However, as correlations between signatures errors increase the median Bayes factor increases noticeably. This suggests that considering error correlations allocates higher likelihoods to parameter sets that capture a considered signature pair. Furthermore, the results shown in Table 3 also imply that the median Bayes factor is relatively insensitive to the precision with which the signatures are regionalized.

3.2.2 Synthetic and observed streamflow data (Observation-based likelihoods)

Figure 2 shows the distribution of the Bayes factor values obtained across the 84 catchments for each of the 10 possible different pairs of signatures, when the observation-based error structure is used for each catchment. Figure 2a shows the results for the observed streamflow data with regionalized signatures calculated from the derived regressions; Fig. 2b shows the results for the synthetic streamflow data with regionalized signatures calculated by adding noise to the exact signature values. The Tukey boxplots in red correspond to pairs of signatures whose errors are statistically significantly correlated (see Fig. 1). The upper whisker represents the upper quartile plus one and

HESSD
12, 5389–5426, 2015

a half times the interquartile range, and the lower whisker represents the lower quartile minus one and a half times the interquartile range. The matrix below Fig. 2b shows the pairs of signatures used.

The signature pair [SFDC, HPC] shows the strongest correlation between errors ($\rho = 0.65$, Fig. 1). A likelihood function with a standard deviation equal to 10 % of the observed signature ranges and $\rho = 0.75$ in Table 3 is comparable to the observation-based likelihood of the pair [SFDC, HPC] (Table 2), with Table 3 indicating [1.45, 1.53] as a 95 % confidence interval for the median Bayes factor. However, a median Bayes factor of 2.17 is obtained for the observed streamflow data (Fig. 2a). Similar differences are found for the other pairs of signatures, although the comparison with the reference table (Table 3) becomes challenging, as the individual signatures have not been regionalized necessarily with similar quality. On the other hand, Fig. 2b shows that the Bayes factors for the synthetic study (when there is no model structural error) are consistent with the values provided in the look-up Table 3. The difference between the median Bayes factor for the two cases is likely to be caused by the model structure error, or may be related to the location of the NSE-optimal in the parameter space.

Nevertheless, it is clear from Fig. 2 that those pairs of signatures whose errors are significantly correlated (i.e. [SFDC, HPC], [BFI, HPC], [BFI, SFDC] and [BFI, SE]) have wider interquartile ranges. Furthermore, the pair of signatures with the strongest correlation between errors [SFDC, HPC] presents the greatest interquartile range. Therefore the inclusion of significant correlations in the likelihood function matters, but whether or not it is beneficial to conditioning the parameters seems to depend on the interplay between model structure error, parameter space and likelihood function. Only strong correlations (as in the [SFDC, HPC] case) can be expected to result in a median Bayes factor clearly above 1.

HESSD
12, 5389–5426, 2015

3.3 The impact of inter-signature error correlations (Multiple signatures)

Multiple signatures are used for parameter constraining and flow prediction. The information value of multiple signatures and its dependence on inter-signature error correlations is explored in this section.

3.3.1 Synthetic streamflow data (Observation-based likelihood)

Figure 3 shows Bayes factors derived for the synthetic streamflow data (generated using the NSE-optimal parameter set) when the observation-based likelihood is used. The Bayes factor considers $p(.|H_2)$ to be the prior parameter distribution, and $p(.|H_1)$ to be one of the parameter posteriors that includes or ignores the inter-signature error correlations. Figure 3 summarizes the variability in the Bayes factor for the different combinations of signatures for all 84 catchments. Boxplots are color coded by the total number of signatures combined, when the inter-signatures error correlation is considered in the likelihood function definition. The grey dashed boxplots correspond to the results obtained assuming that the inter-signature errors are independent when defining the likelihood function. Although the colored boxplots visually seem to have higher values than the grey dashed boxplots, these differences are not statistically significant at a 95 % confidence level (Kolmogorov–Smirnov two-sided tests).

To better evaluate whether the incorporation of additional sources of information improves parameter identification, one-sided Kolmogorov-Smirnov tests are applied between any combination of certain signatures (e.g. [SE, SFDC]) and any other combination that contains the same signatures and a new one (e.g. [SE, SFDC, HPC]). It is found that adding more signatures improves parameter identification in 82.5 % of the cases (66 out of 80 cases) at a 95 % confidence level).

Figure 4 summarizes the variability in the analog Nash-Sutcliffe efficiency measure NSEprob for different combinations of signatures for all 84 catchments. The colored boxplots correspond to the results obtained when the inter-signature error correlations are considered in the likelihood definition, and the grey dashed boxplots correspond to

HESSD
12, 5389–5426, 2015

the results when the inter-signature errors are assumed to be independent. There is no visual or statistical (two-sided Kolmogorov Smirnov tests) difference between the colored boxplots and the grey dashed boxplots in Fig. 4. Moreover, visually, adding more response signatures seems to improve streamflow predictions in terms of accuracy and precision when no model structure error exists. However, only in 59 % of the cases (47 out of 80 cases) more signatures contribute to improved streamflow predictions at a 95 % confidence level (one-sided Kolmogorov–Smirnov test). The other 33 cases always involve the inclusion of the most poorly regionalized signatures (with the highest variance from the five regionalized signatures) – SE, SFDC or HPC – as additional sources of information (see Table 2).

3.3.2 Observed streamflow data (Observation-based likelihood)

Figure 5 shows the results when the same methodology as in the Sect. 3.3.1 is applied using the observed streamflow data. As in the synthetic streamflow case, the differences between the Bayes factor distributions when inter-signature error correlations are considered and when inter-signature errors are assumed to be independent are not statistically significant at a 95 % confidence level (Kolmogorov–Smirnov two-sided tests).

Further, by comparing Fig. 5 with Fig. 3, it becomes clear that the signatures contribute less information, and there is a smaller increase in performance as more signatures are added. It is found that adding more signatures tends to improve parameter identification only in half of the cases when compared to the synthetic streamflow case at a 95 % confidence level (42.5 % versus 82.5 % in the synthetic streamflow case). Furthermore, and contrastingly to the case where no structural error exists, in five situations adding more signatures contributes to a decrease in performance. These five cases always involve adding either SFDC or HPC as an additional source of information. This performance deterioration can be attributed to model structure and observational error. Overall, a statistically significant drop in performance with regard to the Bayes factor is observed most of the time when model structural error is present.

Figure 6 presents the results in terms of NSEprob using the observed streamflow data. As in the synthetic study in Sect. 3.3.1, there is no statistically significant difference at a 95 % confidence difference between the NSEprob distributions when the inter-signature error correlations are considered and when the errors are treated independently (Kolmogorov–Smirnov two-sided tests).

Figure 6 shows that better results in terms of NSEprob are not necessarily achieved when all five signatures are used simultaneously. It is found that adding more signatures tends to improve parameter identification only in 36 % of the cases at a 95 % confidence level (compared to 59 % when there is no model structure error). Furthermore, and contrasting the case where no model structure error exists, in two situations, adding more signatures may contribute to a decrease in performance (when we start with [RR, BFI] and add HPC, and when we start with [RR, BFI] and add SFDC). This might be due to regionalization biases in SFDC and HPC and/or due to the inability of the PDM model to maintain a satisfactory overall performance when conditioned on high peak flow and medium flow information. This negative impact is not observed when synthetic streamflow data are used (Fig. 4), indicating that the decrease in performance may be due to model structural deficiencies. Moreover, a statistically significant drop in performance with regard to NSEprob is observed most of the time when there is model structural error.

In summary, unless there is no model structural error, an all-round performance improvement is not guaranteed by adding more signatures. Furthermore, model structure uncertainty seems to have a much bigger effect on the performance than the explicit inclusion of the inter-signature error correlations.

3.4 Limitations and applicability

The main feature of the method suggested in this paper lies in the possibility of allowing a large number of signatures to be added to the conditioning process, without worrying about double-counting of information or degree of uncertainty in signature estimates, and avoiding subjective decisions about removal of possibly nonindepen-

HESSD
12, 5389–5426, 2015

dent information. Although the proposed framework can be applied to any number of signatures, the limited sample size (i.e. number of gauged catchments available) can have an impact on the definition of the likelihood distribution. For this specific study 83 samples were available to define that distribution. When a single response signature is used to condition the hydrological model this sample size is likely to be sufficient to confidently judge whether the normal distribution assumption is sufficient. However, when moving to multidimensional problems, in which various signatures may be used simultaneously to condition the hydrological model, it is increasingly difficult to judge the adequacy of any multivariate parametric distribution and to judge which catchments are outliers. This implies that as more signatures are used simultaneously in the conditioning of the hydrological model, the more gauged catchments should be used to define the likelihood function. As stressed by Gupta et al. (2014), large samples are of great importance to support statistical regionalization of uncertainty estimates, and this is particularly the case if dependencies between information sources are to be specified.

While the work presented in this paper addresses a number of issues associated with model regionalization, it is important to highlight some additional areas for future research. An important source of uncertainty comes from model structure error (Gupta et al., 1998; Kuczera et al., 2006). The conditioning framework suggested here is independent of the selected model, and, in principle, Figs. 5 and 6 could be created by using the model structure that is considered suitable for each catchment rather than using a model structure that we consider good for generalizing. Further research is needed to diagnose the relative importance of different model structures in various climate regimes and for different catchment characteristics (Clark et al., 2008; Hrachowitz et al., 2013). This is crucial to both identifying the most appropriate model structure for an ungauged location and quantifying the uncertainty in the model structure that should be integrated into the likelihood, thus allowing virtually any model choice. Similarly, other sources of uncertainty, namely observational error (e.g. rainfall error), should ideally be evaluated and integrated into the likelihood function. By accounting for all the

HESSD
12, 5389–5426, 2015

important sources of uncertainty, further insight should be achieved into the information value of sets of signatures and the value of including their dependencies in the likelihood function.

Some of the results presented may be sensitive to the response signatures used. The relationship between value of signatures and catchment type remains ambiguous and an interesting aspect for posterior evaluation would be how the value of signatures depends on catchment type. Other aspects that are worth further research include whether a similar framework could be applied to different types of information source, e.g. can some discharge measurements be added into the model conditioning process? While Bulygina et al. (2012) suggests a framework capable of combining multiple sources of knowledge, namely physically based information, regionalized signatures and spot observations to identify parameters for models of ungauged catchments, the errors between them were assumed to be independent in their case study. A combination of the framework suggested by Bulygina et al. (2012) and the method proposed in this paper may be the way forward to maximizing the value of the available information within a framework of uncertainty reduction.

4 Conclusions

Uncertainty in streamflow estimation in ungauged catchments originates not only from the traditional sources of error generally identified in rainfall-runoff modelling (i.e. model structural, parameter and data errors), but also by errors introduced by the transposition of information from data-rich areas and use of this information to condition model simulations. To identify which and how many types of signatures can usefully be included in model conditioning, it is critical to understand the effects of all these uncertainties. Moreover, when multiple signatures are used simultaneously to condition model simulations, inter-signature error dependencies may also introduce uncertainty and affect decisions about the value of information. While error and uncertainty analyses are quite common in regionalization studies, the question of how much information can be taken

HESSD
12, 5389–5426, 2015

from a set of uncertain signatures and determining how many and which signatures should be used given their error dependencies has not been extensively studied.

The method suggested in this paper allows the specification of a signature error structure. A common reason for not including large numbers of signatures in regionalization studies is the potential for under-estimation of uncertainty due to duplication of information. This study helps to justify the inclusion of larger sets of signatures in the regionalization procedure if their error correlations are formally accounted for and thus enables a more complete use of all available information. The results show that adding response signatures to constrain the hydrological model, while accounting for inter-signature error correlations, can contribute to a stronger identification of the optimum parameter set when the error correlations between different sources of information are strong. Furthermore, the results show that assuming independency of errors does not result in significant deterioration in model performance, unless the error correlation is very strong. The results also show that the effect of error correlations is likely to be overwhelmed by model structure and observation errors. The method suggested here can therefore become more relevant if observational and structural errors are reduced. In addition, it is illustrated that using more signatures, with and without considering their error correlations, may lead to deterioration in performance. In our case, there were particular problems when adding the slope of the flow duration curve and/or the high pulse count. As this is likely to be specific to the rainfall-runoff model used, the selected performance criteria and the set of catchments, it is recommended that the disinformative information sources are identified as part of any regionalization study, in a similar manner as has been done here.

Appendix A: The Bayes factor

When evaluating the impact of inter-signature error correlations on model parameter identification, results are assessed in terms of Bayes factor (Jeffreys, 1961). This form of assessment is preferred to the most commonly used QQ plots (Laio and Tamea, 2007), due to the particular nature of the problem under analysis. When signature(s) (either regionalized for the case of an ungauged catchment, or derived from actual observations for the case of gauged catchments) is employed to reduce uncertainty beyond what is possible by defining the priors on model parameters, QQ plots may not be the most effective form of assessment. Although response signatures are measures of theoretically relevant system process behaviors (Gupta et al., 2008; Wagener et al., 2007), they reflect fragmented knowledge as different signatures capture different catchment processes. Consequently, the quantiles of observed flows are not conditioned to follow a uniform distribution, as QQ plots assess. Rather, quantiles of response signatures should follow this condition (for all catchments considered – Almeida et al., 2013). Therefore, an alternative performance measure that more adequately reflects the aim of this particular application (i.e. the reproduction of certain aspects of the hydrograph) is used. The Bayes factor BF is particularly relevant in the current context as it allows comparison of predictions based on two competing theories (Jeffreys, 1961). It is defined as the ratio between the marginal distributions of the data $\boldsymbol{y}$ for the two hypotheses (H_1 and H_2) being compared (Kass and Raftery, 1995):

$$\mathrm{BF} = \frac{p(\boldsymbol{y}|H_1)}{p(\boldsymbol{y}|H_2)} \tag{A1}$$

When the two hypotheses are equally likely a priori, the Bayes factor is the posterior odds in favor of H_1 (Kass and Raftery, 1995). In other words, a value of BF greater than 1 means that H_1 is more strongly supported by the data than H_2. For example, a Bayes factor equal to 2 implies that H_1 is favored over H_2 with 2 : 1 odds given the evidence provided by the data.

For a given hypothesis H, parameterized by model parameter set $\boldsymbol{\Theta}$, the marginal density $p(\boldsymbol{y}|H)$ represents the likelihood of the data and it is given by

$$p(\boldsymbol{y}|H) = \int p(\boldsymbol{y}|\boldsymbol{\Theta}, H)\, p(\boldsymbol{\Theta}|H)\, \mathrm{d}\boldsymbol{\Theta} \tag{A2}$$

HESSD
12, 5389–5426, 2015

where $p(\boldsymbol{y}|\boldsymbol{\Theta},H)$ is the conditional density function given parameters $\boldsymbol{\Theta}$ under hypothesis H and $p(\boldsymbol{\Theta}|H)$ is the distribution of parameters under H. Hypothesis H may represent different model and parameter distributions. In this paper, the same model structure is considered. However, different parameter distributions are used in Eq. (A2) to enable prediction comparison based on two theories about parameter distributions.

The above integral can be numerically approximated as,

$$\int p(\boldsymbol{y}|\boldsymbol{\Theta},H)p(\boldsymbol{\Theta}|H)\mathrm{d}\boldsymbol{\Theta} \approx \frac{1}{N}\sum_{i=1}^{N} p(\boldsymbol{y}|\boldsymbol{\Theta}^{(i)},H)p(\boldsymbol{\Theta}^{(i)}|H) \tag{A3}$$

where $\boldsymbol{\Theta}^{(i)}$ is the ith of N draws from $p(.|\boldsymbol{\Theta})$, and N is the size of the Monte Carlo sample (in this paper N is equal to 10 000).

In a "perfect model" study, data $\boldsymbol{y}$ are generated by a model with parameter set $\boldsymbol{\Theta}^*$, so that there is no model structural or observational error. This means that $p(\boldsymbol{y}|\boldsymbol{\Theta}^{(i)},H)$ is always equal to zero, except when $\boldsymbol{\Theta}^{(i)} = \boldsymbol{\Theta}^*$. Mathematically this is expressed as $p(\boldsymbol{y}|\boldsymbol{\Theta}^{(i)},H) = \delta_{\boldsymbol{\Theta}^{(i)}=\boldsymbol{\Theta}^*}$, where δ is the Dirac delta function. Therefore Eq. (A3) is equal to $1/N$ times $p(\boldsymbol{\Theta}^{(i)} = \boldsymbol{\Theta}^*|H)$ and the Bayes factor is given by

$$\mathrm{BF} = \frac{\frac{1}{N}\sum_{i=1}^{N}\delta_{\boldsymbol{\Theta}^{(i)}=\boldsymbol{\Theta}^*}p(\boldsymbol{\Theta}^{(i)}|H_1)}{\frac{1}{N}\sum_{i=1}^{N}\delta_{\boldsymbol{\Theta}^{(i)}=\boldsymbol{\Theta}^*}p(\boldsymbol{\Theta}^{(i)}|H_2)} = \frac{p(\boldsymbol{\Theta}^{(i)} = \boldsymbol{\Theta}^*|H_1)}{p(\boldsymbol{\Theta}^{(i)} = \boldsymbol{\Theta}^*|H_2)} \tag{A4}$$

While other choices can be made, two cases are considered in this paper. First, the two distributions in Eq. (A4) are posterior distributions, but with different assumptions about the likelihood functions. Given that we are particularly interested in evaluating the impact of considering the inter-signature error correlations versus ignoring them, H_1 will correspond to the joint likelihood defined such that inter-signature error correlations are considered, while H_2 corresponds to the likelihood when inter-signature error correlations are ignored. For the Bayes factor defined in this way, a value greater than 1 supports the idea that considering inter-signature error correlations contributes to an

HESSD
12, 5389–5426, 2015

Accounting for dependencies in regionalized signatures

S. Almeida et al.

improved specification of the optimum parameter set. In this paper we are also interested in the value of adding/not adding more signatures in model conditioning, and so the Bayes factor will be also calculated for $p(.|H_2)$ set to be the prior parameter distribution, and $p(.|H_1)$ set to one of the derived parameter posteriors. For the Bayes factor defined in this way, a value greater than 1 supports the idea that additional sources of information contribute to a stronger identification of the optimum parameter set.

Acknowledgements. The authors would like to acknowledge the support of Fundação para a Ciência e a Tecnologia (FCT), Portugal, sponsor of the PhD program of S. Almeida at Imperial College London, under the grant SFRH/BD/65522/2009. This work was also partially supported by the Natural Environment Research Council [Consortium on Risk in the Environment: Diagnostics, Integration, Benchmarking, Learning and Elicitation (CREDIBLE); grant number NE/J017450/1]. The authors would also like to acknowledge Keith Sawicz for advice and support relating to the data used in this study.

References

Almeida, S. M. C. L.: The Value of Regionalised Information for Hydrological Modelling, PhD thesis, Imperial College London, London, UK, 2014. 5397, 5398

Almeida, S., Bulygina, N., McIntyre, N., Wagener, T., and Buytaert, W.: Predicting flows in ungauged catchments using correlated information sources, in: British Hydrological Society's Eleventh National Hydrology Symposium, Hydrology for a Changing World, Dundee, UK, doi:10.7558/bhs.2012.ns02, 2012. 5397

Almeida, S., Bulygina, N., McIntyre, N., Wagener, T., and Buytaert, W.: Improving parameter priors for data-scarce estimation problems, Water Resour. Res., 49, 6090–6095, doi:10.1002/wrcr.20437, 2013. 5395, 5410

Arnold, J. G. and Allen, P. M.: Automated methods for estimating baseflow and ground water recharge from streamflow records, J. Am. Water Resour. As., 35, 411–424, doi:10.1111/j.1752-1688.1999.tb03599.x, 1999. 5398

Boorman, D. B., Hollis, J. M., and Lilly, A.: Hydrology of soil types: a hydrologically-based classification of the soils of the United Kingdom, Tech. rep., Institue of Hydrology, Wallingford, UK, 1995. 5391

HESSD

12, 5389–5426, 2015

Bulygina, N., McIntyre, N., and Wheater, H.: Conditioning rainfall-runoff model parameters for ungauged catchments and land management impacts analysis, Hydrol. Earth Syst. Sci., 13, 893–904, doi:10.5194/hess-13-893-2009, 2009. 5391, 5392, 5399, 5400

Bulygina, N., McIntyre, N., and Wheater, H.: Bayesian conditioning of a rainfall-runoff model for predicting flows in ungauged catchments and under land use changes, Water Resour. Res., 47, W02503, doi:10.1029/2010wr009240, 2011. 5391, 5392

Bulygina, N., Ballard, C., McIntyre, N., O'Donnell, G., and Wheater, H.: Integrating different types of information into hydrological model parameter estimation: application to ungauged catchments and land use scenario analysis, Water Resour. Res., 48, W06519, doi:10.1029/2011wr011207, 2012. 5391, 5392, 5408

Calver, A., Lamb, R., and Morris, S. E.: River flood frequency estimation using continuous runoff modelling, P. I. Civil Eng.-Water, 136, 225–234, doi:10.1680/iwtme.1999.31986, 1999. 5398

Castiglioni, S., Lombardi, L., Toth, E., Castellarin, A., and Montanari, A.: Calibration of rainfall-runoff models in ungauged basins: a regional maximum likelihood approach, Adv. Water Resour., 33, 1235–1242, doi:10.1016/j.advwatres.2010.04.009, 2010. 5391

Clark, M. P., Slater, A. G., Rupp, D. E., Woods, R. A., Vrugt, J. A., Gupta, H. V., Wagener, T., and Hay, L. E.: Framework for Understanding Structural Errors (FUSE): A modular framework to diagnose differences between hydrological models, Water Resour. Res., 44, W00B02, doi:10.1029/2007wr006735, 2008. 5399, 5407

Clausen, B. and Biggs, B. J. F.: Flow variables for ecological studies in temperate streams: groupings based on covariance, J. Hydrol., 237, 184–197, doi:10.1016/S0022-1694(00)00306-1, 2000. 5398

De Vleeschouwer, N. and Pauwels, V. R. N.: Assessment of the indirect calibration of a rainfall-runoff model for ungauged catchments in Flanders, Hydrol. Earth Syst. Sci., 17, 2001–2016, doi:10.5194/hess-17-2001-2013, 2013. 5398

Doucet, A., Godsill, S., and Andrieu, C.: On sequential Monte Carlo sampling methods for Bayesian filtering, Stat. Comput., 10, 197–208, doi:10.1023/a:1008935410038, 2000. 5395

Duan, Q., Schaake, J., Andréassian, V., Franks, S., Goteti, G., Gupta, H., Gusev, Y., Habets, F., Hall, A., Hay, L., Hogue, T., Huang, M., Leavesley, G., Liang, X., Nasonova, O., Noilhan, J., Oudin, L., Sorooshian, S., Wagener, T., and Wood, E.: Model Parameter Estimation Experiment (MOPEX): An overview of science strategy and major results from the second and third workshops, J. Hydrol., 320, 3–17, doi:10.1016/j.jhydrol.2005.07.031, 2006. 5393, 5397

HESSD
12, 5389–5426, 2015

Accounting for dependencies in regionalized signatures

S. Almeida et al.

Gupta, H. V., Sorooshian, S., and Yapo, P. O.: Toward improved calibration of hydrologic models: Multiple and noncommensurable measures of information, Water Resour. Res., 34, 751–763, doi:10.1029/97WR03495, 1998. 5407

Gupta, H. V., Wagener, T., and Liu, Y.: Reconciling theory with observations: elements of a diagnostic approach to model evaluation, Hydrol. Process., 22, 3802–3813, doi:10.1002/hyp.6989, 2008. 5410

Gupta, H. V., Perrin, C., Blöschl, G., Montanari, A., Kumar, R., Clark, M., and Andréassian, V.: Large-sample hydrology: a need to balance depth with breadth, Hydrol. Earth Syst. Sci., 18, 463–477, doi:10.5194/hess-18-463-2014, 2014. 5390, 5407

He, Y., Bárdossy, A., and Zehe, E.: A review of regionalisation for continuous streamflow simulation, Hydrol. Earth Syst. Sci., 15, 3539–3553, doi:10.5194/hess-15-3539-2011, 2011. 5391

Hock, R.: Temperature index melt modelling in mountain areas, J. Hydrol., 282, 104–115, doi:10.1016/S0022-1694(03)00257-9, 2003. 5398

Hrachowitz, M., Savenije, H., Blöschl, G., McDonnell, J., Sivapalan, M., Pomeroy, J., Arheimer, B., Blume, T., Clark, M., Ehret, U., Fenicia, F., Freer, J., Gelfan, A., Gupta, H., Hughes, D., Hut, R., Montanari, A., Pande, S., Tetzlaff, D., Troch, P., Uhlenbrook, S., Wagener, T., Winsemius, H., Woods, R., Zehe, E., and Cudennec, C.: A decade of Predictions in Ungauged Basins (PUB) – a review, Hydrol. Sci. J., 58, 1198–1255, doi:10.1080/02626667.2013.803183, 2013. 5407

Jeffreys, H.: Theory of Probability, Oxford University Press, Oxford, 1961. 5399, 5409, 5410

Kass, R. E. and Raftery, A. E.: Bayes Factors, J. Am. Stat. Assoc., 90, 773–795, doi:10.2307/2291091, 1995. 5399, 5410

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, River Res. Appl., 26, 137–156, doi:10.1002/rra.1249, 2010. 5395

Kollat, J. B., Reed, P. M., and Wagener, T.: When are multiobjective calibration trade-offs in hydrologic models meaningful?, Water Resour. Res., 48, W03520, doi:10.1029/2011wr011534, 2012. 5398

Kuczera, G., Kavetski, D., Franks, S., and Thyer, M.: Towards a Bayesian total error analysis of conceptual rainfall-runoff models: Characterising model error using storm-dependent parameters, J. Hydrol., 331, 161–177, doi:10.1016/j.jhydrol.2006.05.010, 2006. 5407

Laio, F. and Tamea, S.: Verification tools for probabilistic forecasts of continuous hydrological variables, Hydrol. Earth Syst. Sci., 11, 1267–1277, doi:10.5194/hess-11-1267-2007, 2007. 5409

Lamb, R. and Kay, A. L.: Confidence intervals for a spatially generalized, continuous simulation flood frequency model for Great Britain, Water Resour. Res., 40, W07501, doi:10.1029/2003WR002428, 2004. 5398

Lidén, R. and Harlin, J.: Analysis of conceptual rainfall-runoff modelling performance in different climates, J. Hydrol., 238, 231–247, doi:10.1016/S0022-1694(00)00330-9, 2000. 5399

McIntyre, N., Lee, H., Wheater, H., Young, A., and Wagener, T.: Ensemble predictions of runoff in ungauged catchments, Water Resour. Res., 41, doi:10.1029/2005WR004289, 2005. 5398

Merz, R. and Blöschl, G.: Regionalisation of catchment model parameters, J. Hydrol., 287, 95–123, doi:10.1016/j.jhydrol.2003.09.028, 2004. 5400

Mishra, A. K. and Coulibaly, P.: Developments in hydrometric network design: A review, Rev. Geophys., 47, RG2001, doi:10.1029/2007RG000243, 2009. 5390

Moore, R. J.: The PDM rainfall-runoff model, Hydrol. Earth Syst. Sci., 11, 483–499, doi:10.5194/hess-11-483-2007, 2007. 5398

Nash, J. E. and Sutcliffe, J. V.: River flow forecasting through conceptual models. Part I. A discussion of principles, J. Hydrol., 10, 282–290, doi:10.1016/j.bbr.2011.03.031, 1970. 5396, 5400

Parajka, J., Andréassian, V., Archfield, S. A., Bárdossy, A., BlÖschl, G., Chiew, F., Duan, Q., Gelfan, A., Hlavcova, K., Merz, R., McIntyre, N., Oudin, L., Perrin, C., Rogger, M., Salinas, J. L., Savenije, H. G., Skøien, J. O., Wagener, T., Zehe, E., and Zhang, Y.: Prediction of runoff hydrographs in ungauged basins, in: Runoff Prediction in Ungauged Basins: Synthesis across Processes, Places and Scales, edited by: Blöschl, G., Sivapalan, M., Wagener, T., Viglione, A., and Savenije, H., 53–69, Cambridge University Press, Cambridge, 2013. 5391, 5392

Peel, M. C. and Blöschl, G.: Hydrological modelling in a changing world, Prog. Phys. Geog., 35, 249–261, doi:10.1177/0309133311402550, 2011. 5391

Razavi, T. and Coulibaly, P.: Streamflow Prediction in Ungauged Basins: Review of Regionalization Methods, J. Hydrol. Eng., 18, 958–975, doi:10.1061/(ASCE)HE.1943-5584.0000690, 2013. 5391

HESSD

12, 5389–5426, 2015

Sankarasubramanian, A., Vogel, R. M., and Limbrunner, J. F.: Climate elasticity of streamflow in the United States, Water Resour. Res., 37, 1771–1781, doi:10.1029/2000WR900330, 2001. 5398

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, 2895–2911, doi:10.5194/hess-15-2895-2011, 2011. 5397, 5398

Schaake, J., Cong, S., and Duan, Q.: U.S. MOPEX data set, Tech. Rep. UCRL-JRNL-221228, Lawrence Livermore National Laboratory, 2006. 5393, 5397

Shu, C. and Ouarda, T. B. M. J.: Improved methods for daily streamflow estimates at ungauged sites, Water Resour. Res., 48, W02523, doi:10.1029/2011WR011501, 2012. 5400

Singh, R., Wagener, T., van Werkhoven, K., Mann, M. E., and Crane, R.: A trading-space-for-time approach to probabilistic continuous streamflow predictions in a changing climate – accounting for changing watershed behavior, Hydrol. Earth Syst. Sci., 15, 3591–3603, doi:10.5194/hess-15-3591-2011, 2011. 5391

Sorooshian, S. and Dracup, J. A.: Stochastic parameter estimation procedures for hydrologie rainfall-runoff models: Correlated and heteroscedastic error cases, Water Resour. Res., 16, 430–442, doi:10.1029/WR016i002p00430, 1980. 5392, 5393

van Werkhoven, K., Wagener, T., Reed, P., and Tang, Y.: Characterization of watershed model behavior across a hydroclimatic gradient, Water Resour. Res., 44, W01429, doi:10.1029/2007WR006271, 2008. 5399

Wagener, T. and McIntyre, N.: Hydrological catchment classification using a data-based mechanistic strategy, in: System Identification, Environmental Modelling, and Control System Design, edited by: Wang, L. and Garnier, H., 483–500, Springer, London, 2012. 5398

Wagener, T. and Montanari, A.: Convergence of approaches toward reducing uncertainty in predictions in ungauged basins, Water Resour. Res., 47, W06301, doi:10.1029/2010WR009469, 2011. 5394

Wagener, T., Sivapalan, M., Troch, P., and Woods, R.: Catchment Classification and Hydrologic Similarity, Geogr. Compass, 1, 901–931, doi:10.1111/j.1749-8198.2007.00039.x, 2007. 5391, 5410

Winsemius, H. C., Schaefli, B., Montanari, A., and Savenije, H. H. G.: 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, 2009. 5391

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, 1756–1774, doi:10.5194/hess-15-3539-2011, 2007. 5391, 5392, 5395, 5398
Young, A. R.: Stream flow simulation within UK ungauged catchments using a daily rainfall-runoff model, J. Hydrol., 320, 155–172, doi:http://dx.doi.org/10.1016/j.jhydrol.2005.07.017, 2006. 5398
Zhang, Z., Wagener, T., Reed, P., and Bhushan, R.: Reducing uncertainty in predictions in ungauged basins by combining hydrologic indices regionalization and multiobjective optimization, Water Resour. Res., 44, doi:10.1029/2008WR006833, 2008. 5391

HESSD
12, 5389–5426, 2015

Table 1. Summary of general catchment properties and response signatures of the 84 MOPEX catchments.

Catchment property	Units	Range
Average annual streamflow	($mm\,yr^{-1}$)	208–896
Average annual precipitation	($mm\,yr^{-1}$)	758–1495
Average annual maximum temperature	(°C)	12–23
Average annual minimum temperature	(°C)	0–10
Average annual potential evaporation	($mm\,yr^{-1}$)	679–1112
Aridity index*	(-)	0.5–1.2
Average elevation	(m)	176–1056
Runoff ratio	(-)	0.16–0.76
Base flow index	(-)	0.36–0.90
Streamflow elasticity	(-)	0.02–4.34
Slope of flow duration curve	(-)	0.01–0.08
High pulse count	(yr^{-1})	2.10–120.80

* Long-term ratio of potential evaporation over precipitation.

HESSD
12, 5389–5426, 2015

Table 2. Tested variance values for the data-based and imposed error structures.

	Observed error structure	1 % observed signature ranges	5 % observed signature ranges	10 % observed signature ranges	20 % observed signature ranges
RR residuals	0.054^2	0.005^2	0.027^2	0.055^2	0.109^2
BFI residuals	0.044^2	0.006^2	0.030^2	0.060^2	0.121^2
SE residuals	0.635^2	0.023^2	0.116^2	0.232^2	0.464^2
SFDC residuals	0.006^2	0.0005^2	0.002^2	0.005^2	0.010^2
HPC residuals	10.687^2	0.977^2	4.883^2	9.767^2	19.533^2

HESSD
12, 5389–5426, 2015
Accounting for dependencies in regionalized signatures
S. Almeida et al.

Table 3. Reference table showing the 95% confidence interval for the median Bayes factor. The correlation coefficient ρ and the standard deviation of the marginal distributions σ are shown.

		σ			
		1 %	5 %	10 %	20 %
	0	1	1	1	1
	0.25	1.01–1.03	1.03–1.04	1.02–1.04	1.04–1.05
ρ	0.50	1.09–1.15	1.16–1.19	1.14–1.17	1.14–1.18
	0.75	1.41–1.51	1.50–1.57	1.45–1.53	1.40–1.49
	0.90	1.94–2.11	2.11–2.32	2.12–2.26	2.20–2.34

HESSD
12, 5389–5426, 2015

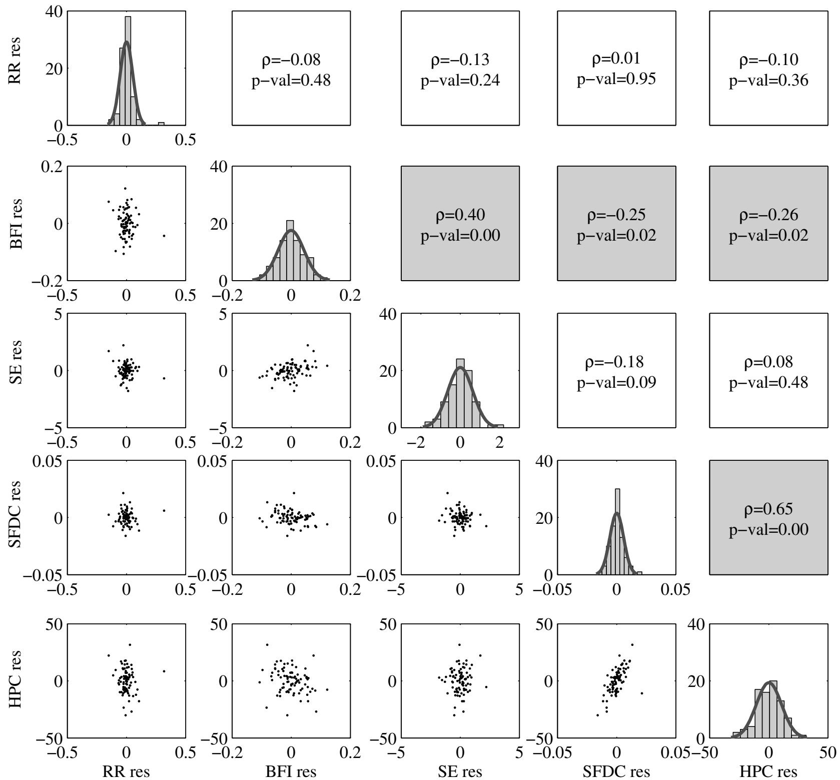

Figure 1. Distribution of individual signature residuals (res) are approximated as histograms and normal distributions. The scatterplots and correlation coefficients (ρ) show correlation between the signature residuals.

HESSD
12, 5389–5426, 2015

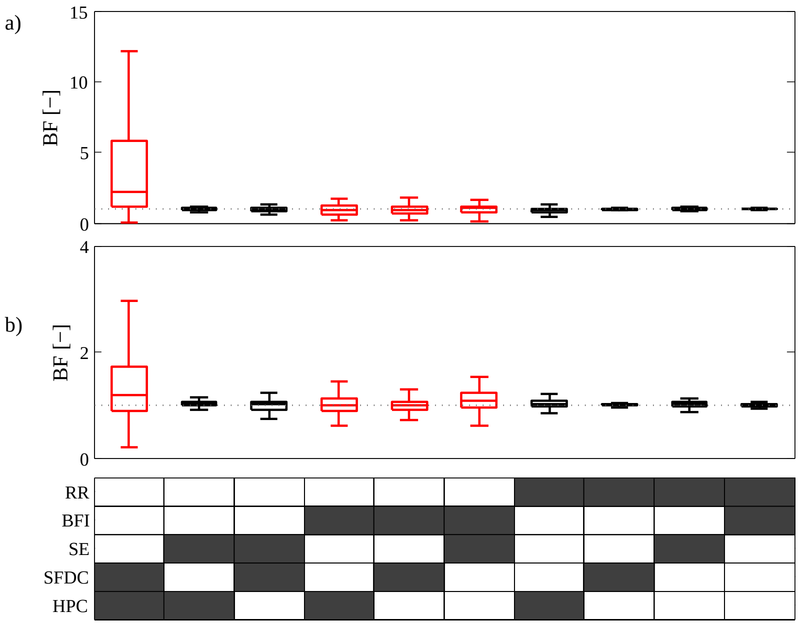

Figure 2. The Bayes factor for the 10 pairs of signatures over the 84 catchments when the observation-based error structure is used with **(a)** observed streamflow data, **(b)** synthetic streamflow data. The upper whisker represents the upper quartile plus one and a half times the interquartile range, and the lower whisker represents the lower quartile minus one and a half times the interquartile range. The dashed line represents BF = 1.

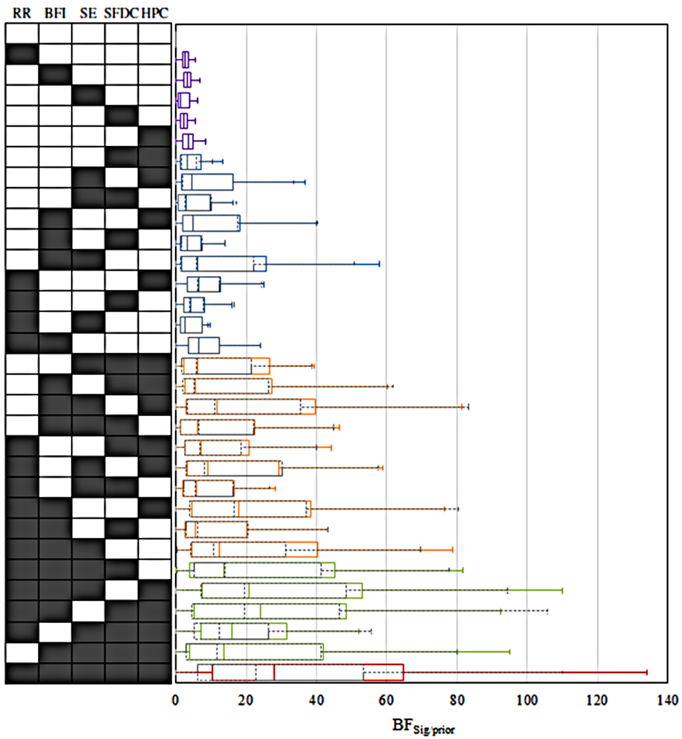

Figure 3. Boxplots representing the distribution of the Bayes factor for each combination of signatures for synthetic streamflow data. The colored boxplots correspond to the results obtained when inter-signature error correlations are considered in the likelihood function, whereas the grey dashed boxplots correspond to the results obtained assuming that the inter-signature errors are independent.

HESSD
12, 5389–5426, 2015

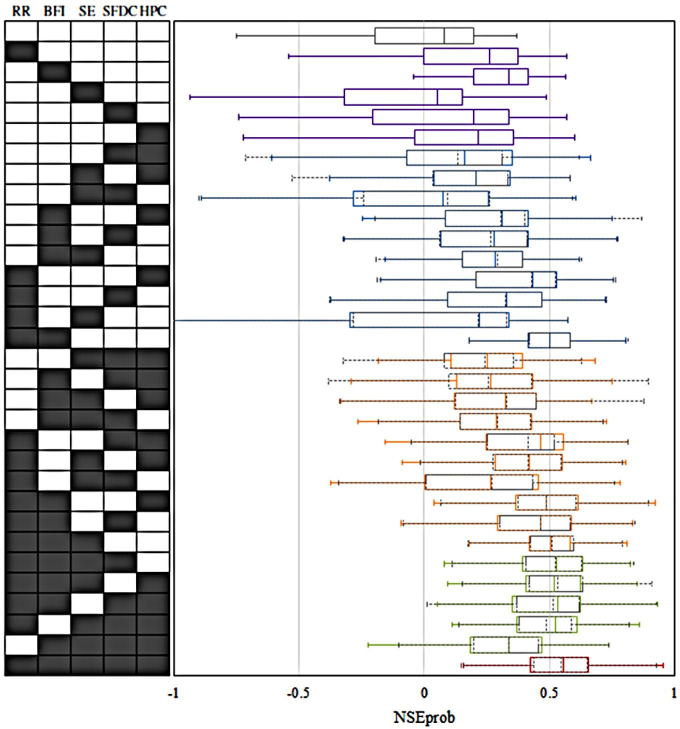

Figure 4. Boxplots representing the distribution of NSEprob values for each combination of signatures for synthetic streamflow data. The colored boxplots correspond to the results obtained when inter-signature error correlations are considered in the likelihood function, whereas the grey dashed boxplots correspond to the results obtained assuming that the inter-signature errors are independent.

HESSD
12, 5389–5426, 2015

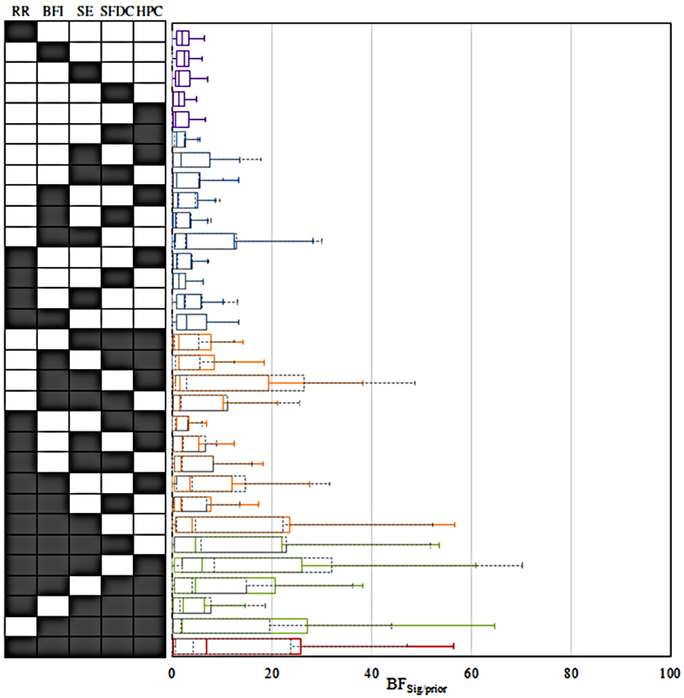

Figure 5. Boxplots representing the distribution of the Bayes factor for each combination of signatures for observed streamflow data. The colored boxplots correspond to the results obtained when inter-signature error correlations are considered in the likelihood function, whereas the grey dashed boxplots correspond to the results obtained assuming that the inter-signature errors are independent.

HESSD
12, 5389–5426, 2015

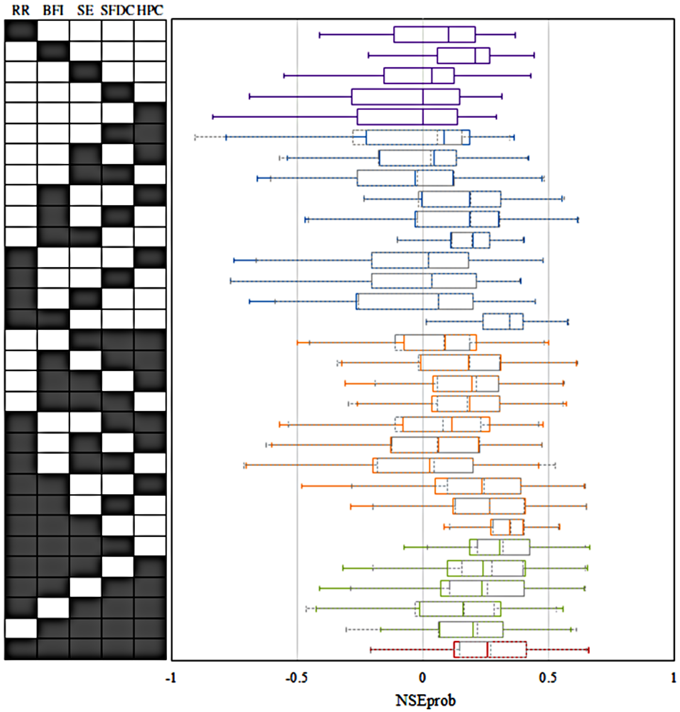

Figure 6. Boxplots representing the distribution of NSEprob values for each combination of signatures for observed streamflow data. The colored boxplots correspond to the results obtained when inter-signature error correlations are considered in the likelihood function, whereas the grey dashed boxplots correspond to the results obtained assuming that the inter-signature errors are independent.