

Water Resources Research®



RESEARCH ARTICLE

10.1029/2022WR031929

Key Points:

- Approximate Bayesian Computation (ABC) employed for signature-domain calibration in ungauged catchments
- A new methodology, based on ABC, proposed to calibrate a precipitation-streamflow model to stochastic regionalized signatures
- Streamflow uncertainty in ungauged catchments is estimated reliably, precision worsens compared to gauged scenario

Correspondence to:

F. Fenicia,
fabrizio.fenicia@eawag.ch

Citation:

Dal Molin, M., Kavetski, D., Albert, C., & Fenicia, F. (2023). Exploring signature-based model calibration for streamflow prediction in ungauged basins. *Water Resources Research*, 59, e2022WR031929. <https://doi.org/10.1029/2022WR031929>

Received 3 JAN 2022

Accepted 21 MAY 2023

Author Contributions:

Conceptualization: Marco Dal Molin, Dmitri Kavetski, Carlo Albert, Fabrizio Fenicia

Formal analysis: Marco Dal Molin

Funding acquisition: Fabrizio Fenicia

Methodology: Marco Dal Molin, Dmitri Kavetski, Carlo Albert, Fabrizio Fenicia

Software: Marco Dal Molin

Supervision: Dmitri Kavetski, Fabrizio Fenicia

Writing – original draft: Marco Dal Molin

Writing – review & editing: Dmitri Kavetski, Carlo Albert, Fabrizio Fenicia

Exploring Signature-Based Model Calibration for Streamflow Prediction in Ungauged Basins

Marco Dal Molin^{1,2}, Dmitri Kavetski^{1,3} , Carlo Albert¹ , and Fabrizio Fenicia¹ 

¹Eawag, Swiss Federal Institute of Aquatic Science and Technology, Dübendorf, Switzerland, ²Centre of Hydrogeology and Geothermics (CHYN), University of Neuchâtel, Neuchâtel, Switzerland, ³School of Civil, Environmental and Mining Engineering, University of Adelaide, Adelaide, SA, Australia

Abstract Calibration of precipitation-streamflow models to streamflow signatures is a promising approach for streamflow prediction in ungauged basins (PUB). The estimation of parameter and prediction uncertainty in this case is not trivial because: (a) calibration takes place in the signature domain, while predictions are required in the time domain, and (b) streamflow signatures are estimated (e.g., from donor catchments) rather than “observed” (computed from observed streamflow in the target catchment), and therefore particularly uncertain. This study investigates model calibration using estimated signatures in an Approximate Bayesian Computation framework. First, we construct a stochastic signature transfer model, based on seasonal flow duration curves. Second, we calibrate a precipitation-streamflow model to the estimated signatures, accounting for their uncertainty. The proposed method is tested in six catchments of the Thur basin, Switzerland. Three data availability scenarios are considered: (a) concomitant scenario, where signatures are “observed,” (b) non-concomitant scenario, where signatures are transferred from a different time period, and (c) regionalization scenario, where signatures are transferred from (neighboring) donor catchments. In this study, the switch from observed to regionalized signatures increases predictive streamflow uncertainty by 38% and worsens the (deterministic) fit to observations by 17% (in terms of Nash-Sutcliffe efficiency). Despite this deterioration, posterior predictive uncertainty remains lower than prior predictive uncertainty generated using uniform priors over representative parameter ranges (“uncalibrated” model), which demonstrates the effectiveness of the proposed signature-based calibration. More importantly, uncertainty is reliably estimated at the ungauged catchments, which represents a key advance in stochastic streamflow PUB.

1. Introduction

Prediction in ungauged basins (PUB) is an important research theme in catchment hydrology. Many human activities (e.g., flood protection, water resources management, ecosystem preservation, etc.) require reliable and precise hydrological predictions, yet such predictions are difficult to obtain when observed data is not available for model calibration. PUB has motivated a large body of work (e.g., review studies of Parajka et al., 2013; Spence et al., 2013), including during the scientific decade 2003–2012 of the International Association of Hydrological Sciences (Hrachowitz et al., 2013; Sivapalan et al., 2003). Another decade later, PUB continues to attract research attention, due to its practical importance and its unresolved challenges (e.g., Kratzert et al., 2019; Prieto et al., 2019).

One of the main objectives of PUB is predicting streamflow and its uncertainty in locations where historical observations are not available (e.g., He et al., 2011; Parajka et al., 2013). The common approach for generating streamflow predictions is through a precipitation-streamflow model calibrated at the catchment of interest. When streamflow observations are available (in a gauged catchment), model parameters can be calibrated to this data. This approach has typically employed residual-based goodness-of-fit measures (e.g., McInerney et al., 2017; Schoups & Vrugt, 2010; Sorooshian & Gupta, 1983, and many others), though the use of hydrological signatures as calibration objectives has also been explored (e.g., Shafii & Tolson, 2015). Here, the term “signatures” refers to summary statistics that reflect meaningful traits of catchment streamflow time series, such as relative base-flow volume, shape of hydrograph recessions, timing of peaks, etc. (e.g., Addor et al., 2017; Sivapalan, 2006). Streamflow signatures based on the flow duration curve (FDC) have received particular attention, as they convey key properties of the marginal distribution of streamflow and are used widely in hydrological engineering applications (e.g., Castellarin et al., 2013).

© 2023. The Authors.

This is an open access article under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

In “ ungauged ” catchments, observed streamflow time series are not available, which considerably complicates the estimation of model parameters. Existing approaches can be classified into three main categories:

1. *Calibration to estimated signatures.* Provided that signatures at the ungauged location can be estimated, which generally requires a modeling approach (as discussed below), model calibration can proceed by optimizing an objective function that quantifies the mismatch between regionalized and simulated signatures (e.g., Kim et al., 2017; Lombardi et al., 2012; Shafii & Tolson, 2015; Yadav et al., 2007; Yu & Yang, 2000). Bayesian inference approaches have also been proposed, using the estimated signatures either to construct the prior distribution of model parameters (e.g., Bulygina et al., 2009, 2011) or to create an approximation of the likelihood function (e.g., Almeida et al., 2016; Bulygina et al., 2012; Castiglioni et al., 2010; Prieto et al., 2019).
2. *Regionalization of model parameters.* This approach relates precipitation-streamflow model parameters to observable catchment characteristics. This association can be based on the direct measurement of model parameters in the field (e.g., Abdulla & Lettenmaier, 1997), which in principle is a viable approach when using physically based models (e.g., Dornes et al., 2008; Fang et al., 2010). Alternatively, it can be inferred by calibration to other gauged catchments, which is the typical approach when using conceptual models (e.g., Fenicia et al., 2016; Merz & Blöschl, 2004; Samaniego et al., 2010; Singh et al., 2014) or machine learning models (e.g., Jiang et al., 2020; Kratzert et al., 2019).
3. *Constraining model parameters using complementary sources of data.* Earlier work in this direction included the use of remotely sensed observations of soil moisture (e.g., Li et al., 2018), radar altimetry (e.g., Sun et al., 2012), GRACE gravity data (e.g., Nijzink et al., 2018), satellite-based evaporation (e.g., Winsemius et al., 2008), spot measurements of streamflow (e.g., Pool et al., 2017), crowd sourced water level data (e.g., Mazzoleni et al., 2017), and process based constraints (e.g., Gharari et al., 2014).

Note that all these approaches assume that meteorological forcings, namely precipitation and potential evaporation, are available at the “ ungauged ” catchment. This scenario is common, because precipitation and temperature data—from which potential evaporation is typically estimated—are available with wide spatial coverage, either from spatial extrapolation of ground data (e.g., Cornes et al., 2018) or from remote sensing platforms (e.g., Xiang et al., 2021). Further note that the three approaches mentioned above are not mutually exclusive, as it can be beneficial to combine multiple sources of information in order to constrain model predictions (e.g., Nijzink et al., 2018; Winsemius et al., 2009).

This study focuses on the first approach, namely the estimation of streamflow signatures in the ungauged catchment using a signature model followed by the calibration of a precipitation-streamflow model to these estimated signatures. Within this approach, it is convenient to distinguish three “ data availability ” scenarios:

1. *Non-overlapping time series,* where observed inputs (e.g., precipitation) and observed outputs (e.g., streamflow) are available in the target catchment but in different time periods. This scenario, which we refer to as “ non-concomitant calibration ” (NCC), represents arguably the simplest ungauged scenario. In this scenario, the signatures are “ transferred ” (extrapolated) in time, in the simplest case by assuming signatures are constant in time. NCC is relevant to catchments where old streamflow time series are available without corresponding precipitation records (e.g., Winsemius et al., 2009), as well as to catchments where streamflow characteristics have been heavily influenced by recent hydraulic infrastructures (e.g., Hingray et al., 2010). NCC was explored by Montanari and Toth (2007), who proposed the use of a likelihood based on the spectral properties of the time series.
2. *Streamflow data available from donor catchments.* In this scenario, streamflow signatures in the target catchment are estimated by the modeler using data from donor catchments, that is, the signatures are transferred (extrapolated) in space, and possibly also in time, depending on the setup. These approaches are typically based on regression between signatures and catchment attributes such as climatic and landscape properties (e.g., Addor et al., 2018; Berger & Entekhabi, 2001; Castiglioni et al., 2009; Prieto et al., 2019; Yadav et al., 2007).
3. *Streamflow data not available.* This scenario is even more restrictive than scenario 2, and requires the regression relating signatures and catchment attributes to be constructed offline as a standalone model, so that, in the given application, the modeler need not find donor catchments. A notable example of such models was developed by Botter et al. (2009) and used later by Doulatyari et al. (2015). This model expresses the seasonal flow regime as a function of four physically based parameters that embed the geomorphic and climate features

of the contributing catchment. Similar signature models were developed by Booker and Woods (2014) and Betterle et al. (2017).

An important challenge associated with data availability and model approximations is the treatment of streamflow predictive uncertainty. Predictive uncertainty quantification when the model is calibrated to *estimated* signatures is particularly challenging. First, calibration takes place in the signature domain, while predictions are necessary in the time domain. Second, streamflow signatures are extrapolated from other catchments rather than calculated directly from local observed streamflow data, and hence may be particularly uncertain.

Uncertainty estimation in the context of PUB has relied mainly on heuristic “limits of acceptability” approaches. For example, Westerberg et al. (2011) searched for precipitation-streamflow model parameters that produce FDCs similar to the observed FDCs at selected quantiles. Winsemius et al. (2009) used observed signatures to impose hard and soft constraints to select suitable parameter sets. Yadav et al. (2007) used regionalized signatures (RS) and accepted only model parameters that produce signatures that fall into an a priori estimated range of variability around the RS. A limitation of these approaches from our perspective in this study is that they yield predictive uncertainty estimates that are not interpretable in a statistical sense.

Although the Bayesian approach could in principle address these limitations, its applications to-date have not focused on the quantification of the different sources of streamflow uncertainty in PUB. For example, previous studies have explored methodologies for conditioning model parameters on RS (e.g., Bulygina et al., 2009, 2011; Castiglioni et al., 2010; Prieto et al., 2019) but have considered mainly the uncertainty related to the regionalization process and neglected other sources (e.g., uncertainty related to the precipitation-streamflow model); usually, these studies weigh model parameter sets based on their ability to produce simulated signatures that are close to the regionalized ones and, then, report the parametric predictive uncertainty of the simulated hydrographs. Other studies (e.g., Almeida et al., 2016) report general model performance (e.g., Nash-Sutcliffe efficiency (NSE)) and its ability to represent the signatures without attempting to estimate predictive uncertainty.

The estimation of uncertainty in the time domain while performing calibration in the signature domain can be pursued using Approximate Bayesian Computation (ABC), as demonstrated in previous publications (e.g., Fenicia et al., 2018; Kavetski et al., 2018; Nott et al., 2014; Vrugt & Sadegh, 2013). Efficient implementations of the ABC approach have been proposed, including the SABC algorithm (Albert et al., 2015), DREAM_(ABC) (Sadegh & Vrugt, 2014), and others. However, these previous studies focused on scenarios where signatures are computed directly from available streamflow observations, which is not the case for ungauged catchments.

In this study, we focus on streamflow prediction and uncertainty estimation at ungauged locations. In order to calibrate a precipitation-streamflow model to uncertain signatures, we propose a Bayesian inference framework implemented using ABC.

Our aims are as follows:

1. Introduce a novel Bayesian approach for the calibration of precipitation-streamflow models to signatures that are estimated, with substantial uncertainty, using a signature transfer model. This advance builds on previous work where the signatures were computed directly from observed streamflow in the catchment of interest.
2. Assess the ability of the proposed signature-domain calibration to provide reliable and precise predictions, in the following data availability scenarios:
 - a) Calibration to concomitant signatures (CS), that is, using observed FDCs. This scenario serves as a baseline;
 - b) Calibration to non-concomitant signatures (NCS), that is, extrapolating the FDCs from another time period;
 - c) Calibration to RS, that is, extrapolating the FDCs from neighboring donor catchments.
3. Assess the performance of the proposed signature-domain calibration in broader contexts, including in comparison to:
 - a) Classical time-domain calibration, in order to appraise potential loss of quality in model predictions;
 - b) Prediction using prior parameter ranges, in order to appraise the extent to which calibration to RS is able to constrain model predictions.

The proposed approach provides separate treatment of two key sources of uncertainty that arise in the modeling process, namely: (a) signature transfer uncertainty, resulting from uncertainty in the data, structure and parameters of the signature transfer model and (b) hydrological model uncertainty, resulting from the data, structure and

parameters of the precipitation-streamflow model. As part of the development, the SABC algorithm proposed in earlier work is generalized to accommodate the case of stochastic (rather than fixed) signatures. A key focus of our evaluation is on the (statistical) reliability of the predictions, that is, on the ability to estimate the magnitude of uncertainty and characterize its distributional properties.

The case study is based on data from six catchments of the Thur basin in Switzerland, which are used to simulate multiple ungauged scenarios. A lumped conceptual precipitation-streamflow model is applied separately in each of the catchments. Our choice of signatures is based on seasonal FDCs. The signature model is based on Doulatyari et al. (2017), but here we enhance it with its own error model (i.e., the signature model is enhanced from deterministic to stochastic), in order to represent uncertainty in the signature regionalization.

The paper is organized as follows. Section 2 presents the theory behind the proposed methodology and Section 3 details the algorithm implementation. Section 4 describes the case study setup, Section 5 reports the case study results, and Section 6 discusses these results and their implications. Section 7 draws the conclusions.

2. Theory

2.1. General Overview

Figure 1 shows a schematic of the signature-based inference framework proposed in this study. The following scenarios are central to this study:

- CS, where streamflow observations are available at the location and time period of interest. The signatures are calculated directly from this observed data. This “gauged catchment” scenario provides the baseline for the other two scenarios.
- NCS, where streamflow observations are available at the location of interest but not in the time period of interest. This scenario corresponds to data availability scenario 1 in Section 1. The signatures in the time period of interest are transferred in time using a stochastic signature model, based on data from the available time period and a random error model to describe associated uncertainty.
- RS, where streamflow observations at the location of interest are not available. The signatures at the location of interest are transferred in space using a stochastic signature model and data from donor catchments, once again with the inclusion of a random error model to describe uncertainty. This scenario corresponds to data availability scenario 2 in Section 1.

In Figure 1, panel (a) describes scenario CS, where signatures are observed, and panel (b) describes the scenarios NCS and RS, where in both cases, the signatures are estimated. We use the term *target signatures* to indicate, regardless of the scenario considered, the signatures that the precipitation-streamflow model has to “match” during the calibration. The method used to estimate the target signatures varies depending on the scenario.

In scenario CS (panel a), the target signatures $\tilde{\mathbf{y}}$ are calculated directly from observed streamflow data $\tilde{\mathbf{q}}$ hence $\tilde{\mathbf{y}} = \mathbf{g}(\tilde{\mathbf{q}})$, where \mathbf{g} is a (vector-valued) deterministic function that transforms a streamflow time series into a vector of target signatures (e.g., FDC quantiles, baseflow index, or other signatures at the catchment of interest) (Section 2.3.1). The parameters $\boldsymbol{\theta}^{(H)}$ of the precipitation-streamflow model H that generates streamflow predictions $\mathbf{Q}^{(H)}$ are inferred in the signature domain, by seeking parameter values that minimize the distance between the modeled signatures $\mathbf{Y}_{\text{prior}}^{(H)} = \mathbf{g}(\mathbf{Q}^{(H)})$ and the observed signatures $\tilde{\mathbf{y}} = \mathbf{g}(\tilde{\mathbf{q}})$. The resulting set of samples from the posterior distribution of the parameters, $\boldsymbol{\theta}_{\text{posterior}}^{(H)}$ can then be used to generate posterior streamflow distributions $\mathbf{Q}_{\text{posterior}}^{(H)}$.

In scenarios NCS and RS (panel b), a signature transfer model $\mathbf{Y}^{(T)}$ ($\boldsymbol{\theta}^{(T)}$) is employed to estimate the target signatures (refer to Sections 2.3.2 and 2.3.3). The signature model generates the same signatures as the function \mathbf{g} but with two caveats: (a) it takes a different set of inputs when streamflow time series are not available and (b) a random error term is included to represent estimation error. The inference process is carried out in two sequential steps. In the first step (shown in the red dashed box), the signature model is calibrated (using data from a different time period in Scenario NCS and data from donor catchments in Scenario RS) and then used to estimate the target signatures $\mathbf{Y}_{\text{prior}}^{(T)}$ at the catchment of interest. In the second step the precipitation-streamflow model is calibrated to the estimated signatures $\mathbf{Y}_{\text{posterior}}^{(T)}$, generated using the pre-calibrated signature model. This

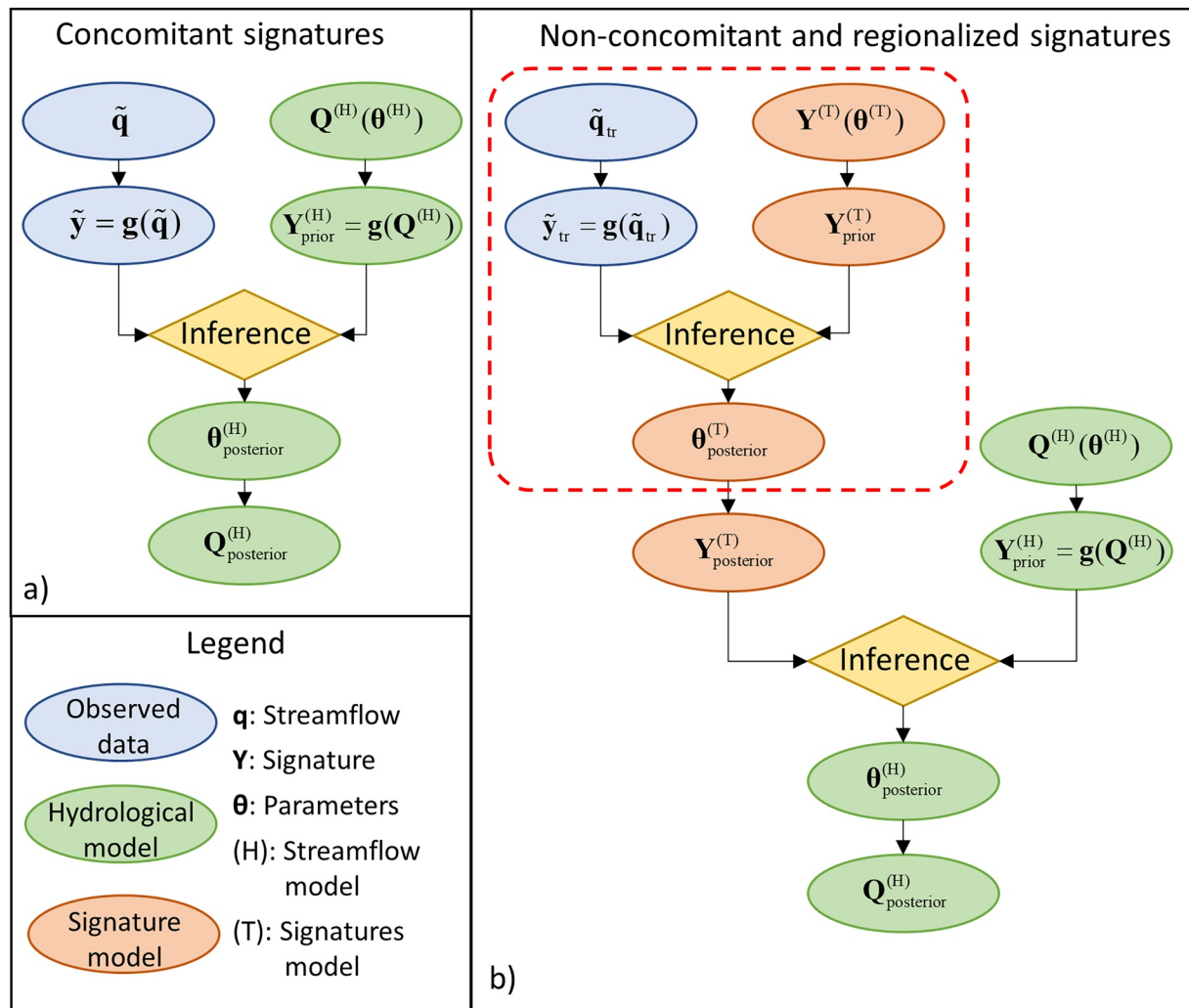


Figure 1. Schematic of the inference setup used in this work to estimate the streamflow model parameters $\theta^{(H)}$. Panel (a) illustrates the inference in a scenario where signatures can be computed directly from observed streamflow. Panel (b) illustrates the inference in scenarios where signatures are estimated using a regionalization model, which itself requires calibration (red dashed box).

step produces samples $\theta^{(H)}_{posterior}$ from the posterior distribution of the parameters of the precipitation-streamflow model, and in turn, samples $Q^{(H)}_{posterior}$ from the posterior (predictive) distribution of streamflow time series. A key distinction of scenarios NCS and RS from scenario CS is that the target signatures form a distribution rather than a fixed value—a difference with important theoretical and algorithmic implications, as explained in Section 3.

In addition to the three main scenarios, the case study considers two auxiliary scenarios, namely concomitant hydrograph (CH) and prior simulation (PS). Scenario CH employs classical time-domain calibration where the parameters of the precipitation-streamflow model are inferred using observed streamflow time series. Scenario PS simulates streamflow using model parameters sampled from their prior distribution (i.e., without any “inference”); note that in this scenario residual errors are ignored, leaving (prior) parameter uncertainty of the precipitation-streamflow model as the sole source of predictive uncertainty. Scenarios CH and PS represent, respectively, the most and least constrained parameter estimation setups, and provide further context for the analysis, as described in Section 4.6.

The procedures depicted in Figure 1 are detailed next.

2.2. Precipitation-Streamflow Model

The predictive distribution $\mathbf{Q}^{(H)} = \{\mathbf{Q}_t^{(H)}; t = 1, \dots, N_T\}$, over N_T time steps is generated by a stochastic precipitation-streamflow model. Typical stochastic models in hydrological modeling are obtained by combining a deterministic model $\mathbf{m}^{(H)}$ with a random residual error term $\epsilon^{(H)}$. In this prototypical example, the stochastic model equation can be written as

$$z[\mathbf{Q}^{(H)}(\boldsymbol{\theta}^{(H)}, \tilde{\mathbf{x}}^{(H)}); \boldsymbol{\theta}_z^{(H)}] = z[\mathbf{m}^{(H)}(\boldsymbol{\theta}_m^{(H)}, \tilde{\mathbf{x}}^{(H)}); \boldsymbol{\theta}_z^{(H)}] + \epsilon^{(H)}(\boldsymbol{\theta}_e^{(H)}) \quad (1)$$

where z is a transformation (e.g., Box-Cox) that accounts for the heteroscedasticity and skew in the residuals (e.g., McInerney et al., 2017); its parameters are denoted as $\boldsymbol{\theta}_z^{(H)}$. The precipitation-streamflow model parameters are denoted as $\boldsymbol{\theta}^{(H)}$, and in turn are composed by parameters of the deterministic model $\boldsymbol{\theta}_m^{(H)}$, parameters of the residual error model $\boldsymbol{\theta}_e^{(H)}$, and transformation parameters $\boldsymbol{\theta}_z^{(H)}$. The term $\tilde{\mathbf{x}}^{(H)}$ denotes all fixed inputs of the model, such as observed precipitation, potential evaporation, etc.

The derivation below is presented for general stochastic models and transformations. Specific modeling choices taken in the case studies are detailed in Section 4.2.

2.3. Streamflow Signatures

Consider a set of signatures \mathbf{y} , defined from the underlying streamflow time series \mathbf{q} through the deterministic function \mathbf{g} , such that $\mathbf{y} = \mathbf{g}(\mathbf{q})$.

The choice of signatures represents a modeling choice analogous to the choice of calibration variables. Our choice of signatures for the case study is given by selected quantiles of seasonal FDCs, as detailed in Section 4.3. For consistency, the choice of signatures is kept constant in all scenarios (though their value will naturally change depending on the estimation method).

2.3.1. Scenario “Concomitant Signatures” (CS)

In scenario CS, the signatures are computed directly from observed streamflow time series in the calibration period and the target catchment,

$$\tilde{\mathbf{y}} = \mathbf{g}(\tilde{\mathbf{q}}) \quad (2)$$

where \mathbf{g} is used to transform observed streamflow $\tilde{\mathbf{q}}$ into “observed” signatures $\tilde{\mathbf{y}}$.

2.3.2. Scenario “Non-Concomitant Signatures” (NCS)

In scenario NCS, the signatures are derived from observed streamflow time series $\tilde{\mathbf{q}}_{tr}$ in the target catchment but in a different time period (with respect to rainfall forcing). The signature transfer model, which combines a deterministic term and an additive random error term to describe uncertainty, is given by:

$$z[\mathbf{Y}^{(T)}(\boldsymbol{\theta}^{(T)}, \tilde{\mathbf{q}}_{tr}); \boldsymbol{\theta}_z^{(T)}] = z[\mathbf{g}(\tilde{\mathbf{q}}_{tr}); \boldsymbol{\theta}_z^{(T)}] + \epsilon^{(T)}(\boldsymbol{\theta}_e^{(T)}) \quad (3)$$

where $\mathbf{Y}^{(T)}$ are the estimated signatures, $\boldsymbol{\theta}_e^{(T)}$ are the parameters of the signature transfer error model $\epsilon^{(T)}$, and $\boldsymbol{\theta}^{(T)}$ comprises $\boldsymbol{\theta}_z^{(T)}$ and $\boldsymbol{\theta}_e^{(T)}$.

The choice of transformation z and error model $\epsilon^{(T)}$ in Equation 3 are application-specific; our choices for the case study are detailed in Section 4.4.2.

2.3.3. Scenario “Regionalized Signatures” (RS)

In scenario RS, the streamflow signatures are obtained using a stochastic signature model. For simplicity, we again assume that the stochastic model comprises a deterministic term $\mathbf{m}^{(T)}$ and an additive random error term $\epsilon^{(T)}$ to describe its uncertainty.

The stochastic signature model is given by

$$z[\mathbf{Y}^{(T)}(\boldsymbol{\theta}^{(T)}, \tilde{\mathbf{x}}^{(T)}); \boldsymbol{\theta}_z^{(T)}] = z[\mathbf{m}^{(T)}(\boldsymbol{\theta}_m^{(T)}, \tilde{\mathbf{x}}^{(T)}); \boldsymbol{\theta}_z^{(T)}] + \epsilon^{(T)}(\boldsymbol{\theta}_e^{(T)}) \quad (4)$$

Similar to Equation 1, the term $\tilde{\mathbf{x}}^{(T)}$ denotes all fixed inputs of the signature model, for example, observed precipitation statistics and landscape characteristics of the target catchment. The model parameters $\boldsymbol{\theta}^{(T)}$ comprise $\boldsymbol{\theta}_m^{(T)}$, $\boldsymbol{\theta}_z^{(T)}$, and $\boldsymbol{\theta}_e^{(T)}$.

The specific choices of model $\mathbf{m}^{(T)}$, transformation z , and residual error model $\boldsymbol{\epsilon}^{(T)}$ made in our case studies are detailed in Section 4.4.3.

Note that the deterministic term $\mathbf{m}^{(T)}$ estimates the signatures without the use of streamflow (which is unavailable in the donor catchment). Hence the function \mathbf{g} does not appear in Equation 4.

2.4. Bayesian Inference Approach

The posterior distribution of precipitation-streamflow model parameters, $p(\boldsymbol{\theta}^{(H)}|\tilde{\mathbf{y}}, \tilde{\mathbf{x}})$, is given by Bayes equation as follows,

$$p(\boldsymbol{\theta}^{(H)}|\tilde{\mathbf{y}}, \tilde{\mathbf{x}}) = \frac{p(\tilde{\mathbf{y}}|\boldsymbol{\theta}^{(H)}, \tilde{\mathbf{x}})p(\boldsymbol{\theta}^{(H)})}{p(\tilde{\mathbf{y}}|\tilde{\mathbf{x}})} \quad (5)$$

where $p(\tilde{\mathbf{y}}|\boldsymbol{\theta}^{(H)}, \tilde{\mathbf{x}})$ is the likelihood function, $p(\boldsymbol{\theta}^{(H)})$ is the prior distribution of the model parameters, and $p(\tilde{\mathbf{y}}|\tilde{\mathbf{x}})$ is a normalization constant (given by the marginal distribution of the observed signatures).

In scenario CS, where the target signatures are deterministic, Equation 5 can be used directly to estimate the posterior of $\boldsymbol{\theta}^{(H)}$; the algorithmic implementation is described in Section 3.

In contrast, in scenarios NCS and RS, where the target signatures are themselves explicitly formulated as the output of a stochastic model, Equation 5 cannot be applied directly.

To accommodate scenarios NCS and RS, we formulate the posterior distribution in “expanded” form using the total probability integral,

$$p(\boldsymbol{\theta}^{(H)}|\tilde{\mathbf{y}}, \tilde{\mathbf{x}}) = \int p(\boldsymbol{\theta}^{(H)}|\mathbf{z}, \tilde{\mathbf{x}})p(\mathbf{z}|\tilde{\mathbf{y}}, \tilde{\mathbf{x}})d\mathbf{z} \quad (6)$$

where $p(\boldsymbol{\theta}^{(H)}|\tilde{\mathbf{y}}, \tilde{\mathbf{x}})$ and $p(\boldsymbol{\theta}^{(H)}|\mathbf{z}, \tilde{\mathbf{x}})$ are the posterior distributions of the precipitation-streamflow model parameters $\boldsymbol{\theta}^{(H)}$ with respect to the observed signatures $\tilde{\mathbf{y}}$ and the simulated signatures \mathbf{z} , respectively. The term $p(\mathbf{z}|\tilde{\mathbf{y}}, \tilde{\mathbf{x}})$ represents the distribution of modeled (transferred) signatures in the target catchment given signatures $\tilde{\mathbf{y}}$ observed in donor catchments.

Expanding the first term in Equation 6 using Bayes equation yields

$$p(\boldsymbol{\theta}^{(H)}|\tilde{\mathbf{y}}, \tilde{\mathbf{x}}) = \int \frac{p(\mathbf{z}|\boldsymbol{\theta}^{(H)}, \tilde{\mathbf{x}})p(\boldsymbol{\theta}^{(H)})}{p(\mathbf{z}|\tilde{\mathbf{x}})}p(\mathbf{z}|\tilde{\mathbf{y}}, \tilde{\mathbf{x}})d\mathbf{z} \quad (7)$$

The three scenarios differ in the choice of the term $p(\mathbf{z}|\tilde{\mathbf{x}}, \tilde{\mathbf{y}})$ in Equation 7.

In scenario CS, the uncertainty in observed signatures is treated as part of the “lumped” error term $\boldsymbol{\epsilon}^{(H)}$ already included in the precipitation-streamflow model in Equation 1. In this case, the probability distribution $p(\mathbf{z}|\tilde{\mathbf{y}}, \tilde{\mathbf{x}})$ collapses to a Dirac function and the integral in Equation 7 simplifies to the usual Bayesian formulation in Equation 5,

$$p(\boldsymbol{\theta}^{(H)}|\tilde{\mathbf{y}}, \tilde{\mathbf{x}}) = \int \frac{p(\mathbf{z}|\boldsymbol{\theta}^{(H)}, \tilde{\mathbf{x}})p(\boldsymbol{\theta}^{(H)})}{p(\mathbf{z}|\tilde{\mathbf{x}})}\delta(\mathbf{z} - \tilde{\mathbf{y}})d\mathbf{z} = \frac{p(\tilde{\mathbf{y}}|\boldsymbol{\theta}^{(H)}, \tilde{\mathbf{x}})p(\boldsymbol{\theta}^{(H)})}{p(\tilde{\mathbf{y}}|\tilde{\mathbf{x}})} \quad (8)$$

In scenarios NCS and RS, this simplification cannot be made because the uncertainty of the estimated signatures is not part of the “lumped” error term $\boldsymbol{\epsilon}^{(H)}$. Therefore, the term $p(\mathbf{z}|\tilde{\mathbf{y}}, \tilde{\mathbf{x}})$ represents the probability density function of the stochastic signature model $\mathbf{Y}^{(T)}$ (denoted in full by $p_{Y^{(T)}}(\mathbf{z}|\hat{\boldsymbol{\theta}}^{(T)}, \tilde{\mathbf{y}}, \tilde{\mathbf{x}})$ in Equation 9 below), with parameters $\hat{\boldsymbol{\theta}}^{(T)}$ pre-calibrated offline to complementary data (refer to Sections 2.3.2 and 2.3.3). Equation 7 can then be written as

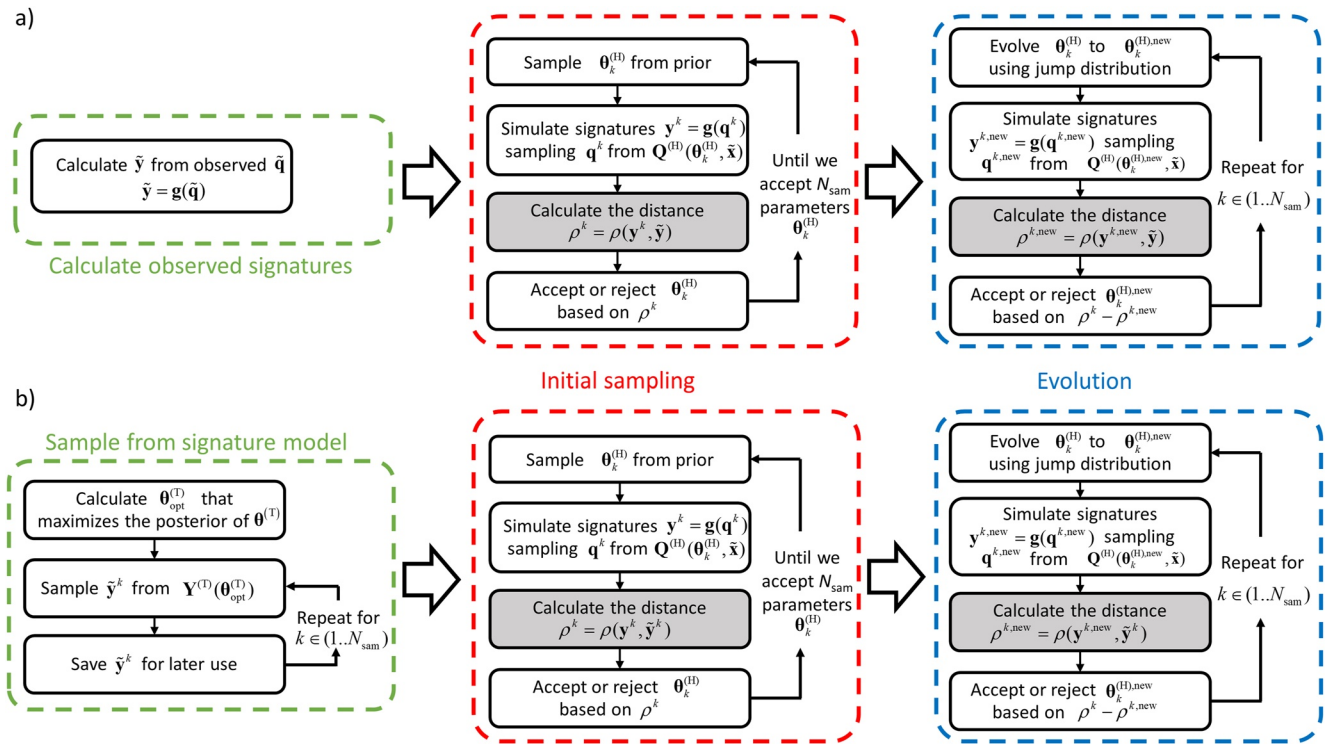


Figure 2. Schematic of the algorithm used to infer the parameters of the precipitation-streamflow model. Panel (a) shows the algorithm used in scenario CS; panel (b) shows the algorithm used in scenarios non-concomitant signatures (NCS) and regionalized signatures (RS). Gray background is used to indicate the step of calculating the distance metric, which is the only difference between the initial sampling and evolution in scenarios concomitant signatures versus NCS and RS.

$$p(\theta^{(H)}|\tilde{y}, \tilde{x}) = \int \frac{p(z|\theta^{(H)}, \tilde{x})p(\theta^{(H)})}{p(z|\tilde{x})} p_{Y^{(T)}}(z|\hat{\theta}^{(T)}, \tilde{y}, \tilde{x}) dz \quad (9)$$

The posterior parameter distribution in Equations 8 and 9 is sampled using an ABC procedure as described next in Section 3.

3. Sampling Algorithm to Implement Parameter Inference: Modified SABC Algorithm

This section describes the sampling algorithm used to estimate the posterior distribution of parameters $\theta^{(H)}$ of the precipitation-streamflow model. Compared to classical time-domain calibration, signature-domain calibration brings the complication that the likelihood function is not available in closed form. Moreover, a signature itself can be treated either as a fixed given value (scenario CS) or as a random variable to represent estimation uncertainty (scenarios NCS and RS).

The ABC sampling algorithm used in this work is based on the SABC algorithm proposed in Albert et al. (2015), but employs a modification to take into account the use of stochastic (rather than fixed) target signatures.

Figure 2 illustrates the algorithm. Panel (a) shows the workflow applied in the case of deterministic signatures (scenario CS); for this scenario, the procedure is the same as in the earlier study by Kavetski et al. (2018) and Fenicia et al. (2018). Panel (b) shows the algorithm modifications developed to handle stochastic signatures (scenarios NCS and RS).

In both cases, the algorithm is composed by three separate stages:

1. Calculation of the target signatures (green box). This step is carried out using observed streamflow data in scenario CS and using the signature models in scenarios NCS and RS;

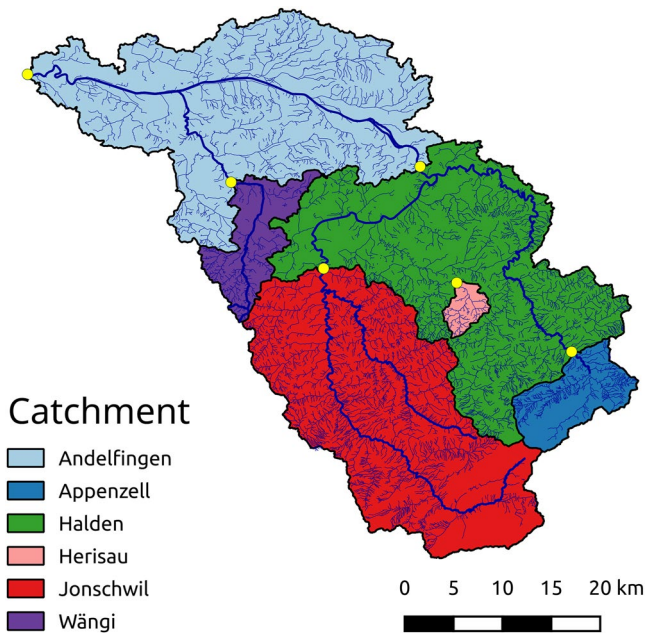


Figure 3. Map of the Thur basin indicating its catchments. Streamflow gauging stations are indicated with yellow dots.

2. Determination of an initial population of N_{sam} model parameters $\theta_k^{(H)}$, which is obtained by sampling parameter sets from the prior distribution and retaining parameter sets where the distance ρ between simulated and target signatures is below a prescribed tolerance (red box). Note that this procedure inevitably yields different initial populations for different scenarios (due to differences in signature values, and hence in distance metric values and in the satisfaction of the prescribed tolerance);
3. Evolution of the population of particles $\theta_k^{(H)}$ using a Metropolis step with the acceptance/rejection tolerance evolving according to the annealing schedule described in Albert et al. (2015) (blue box).

Intuitively, the SABC algorithm works as follows. We seek to generate samples (parameter sets and corresponding streamflow time series and signatures) that meet a signature distance tolerance that is as tight as possible (ideally zero though this is impossible in practice). To obtain such samples, Stage 2 generates an initial population of samples that meet a loose tolerance and then Stage 3 progressively “evolves” these samples toward values that meet tighter tolerances. For the overall algorithm to be computationally viable, the tolerance in Stage 2 must be sufficiently loose that the initial population can be generated in a reasonable time. As the SABC algorithm progresses toward convergence, the tolerance is tightened and the influence of the distance metric on the parameter inference decreases (see Albert et al. (2015) and Kavetski et al. (2018) for details).

A detailed description of the standard SABC algorithm shown in panel (a) is provided in Albert et al. (2015); see also Kavetski et al. (2018). The presentation here focuses on the modifications designed in this work to adapt the SABC algorithm to the case of stochastic target signatures (panel b). Further technical details and specific algorithmic choices are provided in Section 4.

The key difference is in the computation of the target signatures. In scenario CS (panel a, green box), a single fixed set of signatures is calculated from observed streamflow data. In scenarios NCS and RS (panel b, green box), a population of N_{sam} sets of target signatures is sampled from the pre-calibrated signature model. This step is indicated with a gray background in Figure 2.

This population is kept fixed throughout the SABC computation. Each particle $\theta_k^{(H)}$ is associated to one of these target signature values \tilde{y}_k , and is evolved by the SABC algorithm to minimize the distance of its corresponding simulated signature set to this target signature set. This modification allows for the incorporation of the uncertainty of the stochastic signatures into the SABC sampling algorithm.

Once sampling is complete, the predictive streamflow distribution can be generated for any period of interest. See Appendix A for the computational procedure.

Once sampling is complete, the predictive streamflow distribution can be generated for any period of interest. See Appendix A for the computational procedure.

4. Case Study Materials and Methods

4.1. Catchments

The proposed inference approach is tested in the Thur basin, which is an alpine and pre-alpine catchment located in north-eastern Switzerland, south of Lake Constance. Within this basin we selected six catchments, as shown in Figure 3.

Although some of the catchments are nested, they present substantial variability in streamflow signatures including average streamflow (1.64–4.14 mm/d), baseflow index (0.42–0.57) and seasonality (the mean half streamflow date varying between 168 and 221 days), as shown in previous work (Dal Molin, Schirmer, et al., 2020).

The catchment has high quality observation data, having been studied intensively in the last 40 years. Detailed physical characteristics of the Thur basin and a summary of previous investigations can be found in Dal Molin, Schirmer, et al. (2020). The work of Doulatyari et al. (2017) on the estimation of FDCs is particularly relevant to this case study, as detailed in Section 4.4.3.

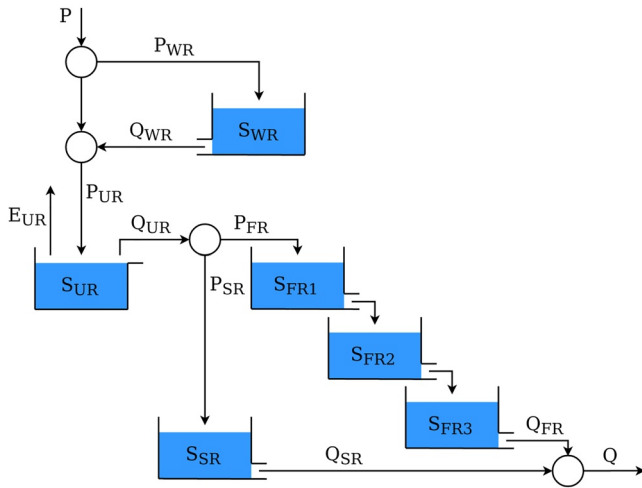


Figure 4. Schematic representation of the lumped streamflow model used in the case study. “ P ” represents the precipitation entering in the reservoirs, “ E ” the evaporation, and “ Q ” the outflow from the reservoirs. The subscripts indicate the reservoirs: WR, snow reservoir; UR, unsaturated reservoir; FR, fast reservoir; SR, slow reservoir. The governing equations are reported in Appendix B.

The streamflow time series are obtained from the Swiss Federal Office for the Environment. The precipitation, temperature, and potential evaporation (PET) time series are derived from data provided by the Swiss Federal Office of Meteorology and Climatology (MeteoSwiss); the calculation of catchment averages from meteorological ground station data is detailed in Section 2 of Dal Molin, Schirmer, et al. (2020).

All the time series used in this study are daily and span the period 1981–2005, with only one gap in the streamflow data in Herisau, from 31 December 1982 to 9 May 1983. The 24 years of data are divided into three periods of 8 years, which are used to create “virtual” ungauged scenarios (see Section 4.5 for details).

4.2. Precipitation-Streamflow Model

The deterministic precipitation-streamflow model $\mathbf{m}^{(H)}$ used in Equation 1 is based on the popular HyMod model (Boyle, 2003), with the additional inclusion of a snow reservoir. This model provides reasonable fits to the observed data in the Thur basin (Dal Molin, Schirmer, et al., 2020). Importantly, as a lumped model, it has low computational and data requirements.

The model has four elements, shown schematically in Figure 4. The snow reservoir (WR) intercepts the incoming precipitation and releases it according to the input temperature, in order to simulate snow accumulation and melting. The unsaturated reservoir (UR) partitions the combined precipitation and snowmelt into a flux that builds storage, which eventually evaporates, and a flux that is propagated to the downstream elements and eventually produces streamflow. The latter flux is partitioned between a cascade of three fast reservoirs (FR) and a slow reservoir (SR). The FR are intended to generate the peaks of the hydrograph and their offset, while the SR is intended to produce the baseflow.

The model is implemented using the SUPERFLEX modeling framework (Dal Molin, Kavetski, & Fenicia, 2020; Fenicia et al., 2011). A fixed step implicit Euler time stepping scheme is used for numerical stability. The equations and the calibrated parameters are detailed in Appendix B.

The streamflow residual error model is built to take into account heteroscedasticity, skew and autocorrelation. The Box-Cox transformation (Box & Cox, 1964),

$$z_{BC}[x; \lambda] = \frac{x^\lambda - 1}{\lambda} \quad (10)$$

is used in Equation 1. The power parameter λ is fixed to 0.2 (McInerney et al., 2017).

The autocorrelation is represented using a first-order autoregressive model (AR1),

$$\varepsilon_t^{(H)} = \phi^{(H)} \varepsilon_{t-1}^{(H)} + W_t^{(H)} \quad (11)$$

where $\varepsilon_t^{(H)}$ is the residual error at the time step t , $\phi^{(H)}$ is the autoregressive parameter and $W_t^{(H)}$ is the innovation (random noise term).

The innovations are assumed to follow a lower-truncated Gaussian distribution,

$$W_t^{(H)} \sim \mathcal{N}_{LT}(0, \sigma_W^{(H)}, L_{W,t}) \quad (12)$$

with zero mean, variance $\sigma_W^{(H)}$, and lower bound $L_{W,t}$ set such that $Q_t^{(H)} \geq 0$.

The streamflow residual error model is, therefore, defined by two parameters $\theta_e^{(H)} = [\phi^{(H)}, \sigma_W^{(H)}]$; the value of $\phi^{(H)}$ is fixed to 0.8 while $\sigma_W^{(H)}$ is inferred. These choices follow the recommendations of Evin et al. (2014) and McInerney et al. (2017).

4.3. Definition of Streamflow Signatures

The streamflow signatures are derived from the seasonal FDCs. To avoid ambiguity, we define a seasonal FDC, $\psi_s(q)$, as the cumulative distribution function (CDF) of a streamflow time series in season s ,

$$\psi_s(q) = \Pr(Q_s \leq q) \quad (13)$$

where Q_s is a random variable representing seasonal streamflow.

Simulated FDCs are calculated directly from simulated streamflow time series. Target FDCs are either calculated from an observed streamflow time series (scenario CS) or estimated using a model (scenarios NCS and RS). For example, in scenario CS, the streamflow time series \mathbf{q} is divided in four seasonal partitions \mathbf{q}_s , where $s = 1, \dots, N_{\text{season}}$ is the season and $N_{\text{season}} = 4$. Season 1 (autumn) is defined as September, October and November; Season 2 (winter) is defined as December, January and February; and so forth.

The signatures \mathbf{y} are defined by the slopes between selected (consecutive) quantiles of the seasonal FDCs,

$$y_{[j,s]} = \begin{cases} \psi_s^{-1}(\zeta_j)/\zeta_j & j = 1 \\ \frac{\psi_s^{-1}(\zeta_j) - \psi_s^{-1}(\zeta_{j-1})}{\zeta_j - \zeta_{j-1}} & j > 1 \end{cases} \quad (14)$$

where $\psi^{-1}(\zeta)$ is the inverse FDC, that is, the streamflow at quantile ζ . The “matrix subscript” notation $[j, s]$ is used in Equation 14 to refer to the j th signature in season s .

The quantiles ζ_j in Equation 14 are selected according to the following expression, which provides higher resolution of quantiles for higher streamflow (e.g., Westerberg et al., 2011),

$$\zeta_j = 1 - \left(\frac{j+2}{N_{\text{FDC}} + 3} \right)^2, \quad \text{for } j = 1, \dots, N_{\text{FDC}} \quad (15)$$

Following Westerberg et al. (2011), a total of $N_{\text{FDC}} = 19$ slopes are calculated according to Equation 14: $N_{\text{FDC}} - 1$ slopes between the quantiles in Equation 15, and additionally the slope between (0,0) and $(\zeta_1, \psi_s^{-1}(\zeta_1))$. The choice of N_{FDC} provides a balance between the benefits of a high resolution of the FDC versus the increased computational costs. Note that the extremal quantiles 0 and 1, which correspond to the maximum and minimum streamflow, are excluded due to their particular volatility and uncertainty.

Since in other equations the signatures are assumed to be concatenated into a single vector $\mathbf{y} = \{y_i; i = 1, \dots, N_y\}$, the mapping $y_i = y_{[j,s]}$ holds for $s = \text{floor}(i/N_{\text{FDC}}) + 1$ and $j = i - (s - 1)N_{\text{FDC}}$, where floor is the integer rounding down function.

The use of FDC slopes instead of FDC quantiles, as shown in Equation 14, yields a better-behaved signature error model in scenarios NCS and RS. In particular, from an algorithmic perspective, monotonicity of the FDC is specified more easily by requiring the slopes to be positive rather than by imposing relational constraints between the quantiles.

Finally, note that the use of FDC slopes instead of FDC quantiles does not result in a loss of information, as there is a one-to-one correspondence between slopes and quantiles. The definition of the first slope, $j = 1$ in Equation 14 distinguishes “parallel” FDCs (with matching slopes everywhere except the first segment).

The distance metric is set to the maximum of the seasonally averaged distances between observed and simulated FDC slopes. A normalization is used to scale the slopes by their estimated variability. See Appendix C for details.

4.4. Signature Models

4.4.1. Scenario CS (Concomitant Signatures)

The signatures are computed as

$$\tilde{\mathbf{y}} = \mathbf{g}(\tilde{\mathbf{q}}) \quad (16)$$

where the function \mathbf{g} is as defined earlier in Section 2.3.

4.4.2. Scenario NCS (Non-Concomitant Signatures)

The model for NCS is formulated based on Equation 3 as follows,

$$z_{BC} [Y^{(T)}(\theta_e^{(T)}, \tilde{q}_{tr}), \lambda_{BC}^{(T)}] = z_{BC} [g(\tilde{q}_{tr}), \lambda_{BC}^{(T)}] + \mathcal{N}(0, \sigma_e^{(T)}) \quad (17)$$

where \tilde{q}_{tr} is streamflow from the same catchment and season but different time period. The Box-Cox parameter is set as $\lambda_{BC}^{(T)} = 0.2$. The error standard deviation $\sigma_e^{(T)}$ is inferred in the same catchment in a period where streamflow is available (see Section 4.5).

4.4.3. Scenario RS (Regionalized Signatures)

The deterministic FDC model in this study is based on the model proposed by Botter et al. (2009) and subsequently applied in the Thur basin (Doulatyari et al., 2017). This parametric model defines the probability density function (pdf) of streamflow q , up to a constant, as follows,

$$D(q; \alpha, L, a, k) = \frac{d\psi}{dq} \propto q^{-a} \exp\left(-\frac{q^{2-a}}{\alpha k(2-a)} + \frac{Lq^{1-a}}{k(1-a)}\right) \quad (18)$$

where parameters α and L represent, respectively, the mean precipitation and the frequency of effective precipitation events (i.e., precipitation events that generate streamflow); k and a represent the coefficient and the exponent of the hydrograph recession.

This model does not make use of streamflow data, and therefore can be used for cases where streamflow data is not available (data availability scenario 3 in Section 1). As described by Doulatyari et al. (2017), parameters α and L are estimated from daily rainfall and snowfall time series (the model uses precipitation without distinguishing its form). Parameters k and a are estimated from geomorphic properties of the catchment.

The model in Equation 18 is applied on a seasonal basis. We used the seasonal parameter values reported in Table 2, column “Modeled,” of Doulatyari et al. (2017) for the Thur basin, where parameter λ corresponds to parameter L as defined in this presentation.

The deterministic model $m^{(T)}$ used in Equation 4 is obtained by numerical integration of the PDF in Equation 18 with respect to q in order to calculate the corresponding CDF and hence the (seasonal) FDC, which in turn is used in Equation 14 to calculate the signatures.

The specific choices regarding transformation z and error model $\epsilon^{(T)}$ in Equation 4 are motivated by residual analysis of model $m^{(T)}$ in the gauged catchments. Based on these analyses, it was observed that residuals are heteroscedastic, and characterized by a consistent difference (bias) between observed and simulated FDCs (e.g., Figure 4 in Doulatyari et al., 2017). Here, we chose to represent heteroscedasticity by making the standard deviation of the residuals proportional to the magnitude of simulated signatures,

$$\epsilon^{(T)}(b_e^{(T)}, m^{(T)}) = \mathcal{N}(0, b_e^{(T)} m^{(T)}) \quad (19)$$

where $b_e^{(T)}$ is a proportionality constant that can be interpreted loosely as a relative error. As a consequence of this parameterization, the transformation z is not needed.

The signature model is “bias-corrected” by applying a multiplier $a_e^{(T)}$ to the modeled signatures. Equation 4 is therefore simplified to

$$Y^{(T)}(\theta_e^{(T)}, \tilde{x}^{(T)}) = a_e^{(T)} m^{(T)}(\theta_m^{(T)}, \tilde{x}^{(T)}) + \epsilon^{(T)}(\theta_e^{(T)}, m^{(T)}) \quad (20)$$

The inferred model parameters are $\theta_e^{(T)} = [a_e^{(T)}, b_e^{(T)}]$.

In order to estimate the error term, we use data from neighboring catchments, meaning that the stochastic signature transfer model belongs to the data availability scenario 2 in Section 1.

4.5. Construction of Ungauged Scenarios for the Calibration of the Signature Models

In order to construct “virtual” ungauged scenarios, the 24 years-long period of streamflow data is divided into three periods: P1, from 1 September 1981 to 31 August 1989; P2, from 1 September 1989 to 31 August 1997; P3, from 1 September 1997 to 31 August 2005.

The following procedure is then applied for the three scenarios:

- Scenario CS: when $\theta^{(H)}$ are needed for a given catchment and a given period, the model is calibrated to observed signatures in that catchment and that period.
- Scenario NCS: when $\theta^{(H)}$ are needed in P1 for a given catchment, the signature transfer model in Equation 3 uses inputs \hat{q}_t from P2, and is conditioned on signatures \hat{y} observed in P3. We then rotate the three periods in order to run the inference for all the time series of a given catchment. We then apply this procedure to all catchments.
- Scenario RS: when $\theta^{(H)}$ are needed in a specific catchment for a given period, the signature model in Equation 4 was conditioned on signatures \hat{y} observed in the other five catchments. We then rotated the six catchments in order to run the inference in all catchments for a given period. We then applied this procedure to all periods.

This procedure generates a set of calibrated parameters $\theta_{opt}^{(T)}$ that is specific for each catchment and each time period. These parameters are kept fixed at their calibrated values when inferring $\theta^{(H)}$ with the procedure described in Section 3. The model diagnostics metrics described in Section 4.7 are calculated on the entire 24 years-long period, by concatenating the predictions in periods P1, P2, and P3 into a single time series, and comparing this concatenated time series to the corresponding (single) observed time series.

The number of SABC samples (see Section 3) is set to $N_{sam} = 5,000$ following our previous work (Fenicia et al., 2018), where it was found appropriate to characterize the posterior distribution of a HyMod-like model. Note that, with these settings, convergence of the SABC algorithm requires approximately 4 million hydrological model runs.

The CS scenario serves as a baseline of ideal performance given the selection of model and signatures. The NCS and RS scenarios use signature transfer in time and in space respectively, therefore they can be considered as validation scenarios. It would be in principle possible to create an even more challenging validation scenario where signatures are transferred both in time and in space, but we did not test this scenario in this work.

4.6. Experiments

Three experiments are carried out:

- Experiment 1 compares model performance in the three signature-domain calibration scenarios (CS, NCS, and RS) in order to appraise potential loss of quality in model predictions when moving from gauged to ungauged conditions.
- Experiment 2 compares signature-domain calibration in scenario CS with time-domain calibration (scenario CH), in order to appraise potential loss of quality in model predictions when moving from time domain calibration to signature domain calibration.
- Experiment 3 compares model performance in scenario RS with scenario PS, where the prior predictive distribution is used, that is, hydrographs generated using parameters sampled from the prior distribution. This experiment helps assess the extent to which calibration to RS is informative beyond what is already known a priori.

Experiment 1 is the main experiment of this study. Experiments 2 and 3 are auxiliary experiments that represent, respectively, the most and least constrained parameter estimation setups, and provide additional context to Experiment 1. Note that, apart from the modifications described in Section 3, the SABC algorithmic settings were the same in all scenarios.

The results are reported using two levels of detail. First, we report the result for a representative catchment and time period, namely Andelfingen and time period P3 (see Section 4.5), where we provide details including hydrographs and annual FDCs. Second, we consider all catchments and time periods, and report performance using the metrics defined in Section 4.7 applied to the concatenated time series as described in Section 4.5.

For Experiment 1, we include a comparison of the posterior distribution of the model parameters in the three scenarios. Specifically, we illustrate cases of parameters being well identified or weakly identified in all scenarios and cases where parameter identifiability varies by scenario.

4.7. Performance Metrics for a Posteriori Evaluation of Streamflow Predictions

The quality of streamflow predictions in all experiments and scenarios is evaluated using a set of performance metrics. We report model performance in terms of both streamflow time series (hydrograph) and their FDC (expressed as a cumulative distribution). We distinguish metrics that characterize the fit to the data from metrics that characterize the uncertainty.

The fit of the simulations to the observations is quantified using:

- NSE of the median of the predictive distribution. The NSE is chosen because it is a standard metric of model performance in hydrological applications;
- Volumetric bias, which measures the long-term water balance error of simulations.

The uncertainty of the streamflow predictions is assessed through the following metrics:

- Reliability, which measures the statistical consistency between the observations and the predictive distribution (i.e., the degree to which observations are consistent with being samples from the predictive distribution);
- Precision, which measures the (average) spread of the predictive distribution.

The NSE and bias metrics are common in hydrology, and the uncertainty quantification metrics are also well established (e.g., Ehlers et al., 2019; McInerney et al., 2017; Oliveira et al., 2018). The metrics are detailed further in Appendix D.

To avoid confusion, we emphasize that the performance metrics in this section are used to evaluate model predictions a posteriori (after the calibration), and are not used in the model calibration itself. Model calibration is instead based on the ABC approach, which matches the selected signatures (FDC slopes) according to the ABC distance metric, as described in Section 3 (general principles) and Appendix C (specific case study choices).

5. Results

5.1. Experiment 1: Signature Calibration in Multiple Scenarios

5.1.1. Representative Catchment: Hydrograph and FDC Representation

Figure 5 shows the results of scenarios CS, NCS, and RS for the representative Andelfingen catchment in period P3. Panel (a) shows the simulated hydrographs and panel (b) shows the simulated FDCs. Panels (c and d) show the values of the performance metrics.

The hydrographs in panel (a) show a consistent difference in the width of the streamflow uncertainty bounds between scenario CS and the other two scenarios. Scenario CS has the narrowest predictive uncertainty, scenario RS comes second, followed closely by scenario NCS, where uncertainty bounds enclose the other two. This difference in (relative) magnitude of uncertainty is also confirmed by the precision metric of the hydrograph (panel d), which is the lowest in the scenario CS (0.42) and assumes comparable values in the other two scenarios (0.57 for scenario NCS and 0.56 for scenario RS). The predictions are generally reliable, with best values of the reliability metric in scenario CS (0.049), followed by scenario RS (0.12) and then scenario NCS (0.13).

Panel (a) also shows that the median simulated hydrograph (dashed lines) in Scenario CS provides much better capture of the observed hydrograph “dynamics” (i.e., reaction to rainfall events), especially for high streamflow. In scenarios NCS and RS the median hydrograph is less dynamic, with deeper recession and limited ability to reach the observed peaks. This qualitative assessment is confirmed by the value of the NSE (panel c), which is highest in scenario CS (0.64), followed by scenarios RS and NCS, both with a value of 0.59.

In terms of FDC performance, all scenarios tend to suffer from underestimated high streamflow. Apart from that, the comparison of FDCs shows results consistent with the comparison of hydrographs. In particular, scenario CS produces the most precise predictive distributions and has the median closest to the observed values. Scenarios NCS and RS produce wider predictive distributions with comparable precision to each other. Note that, due to

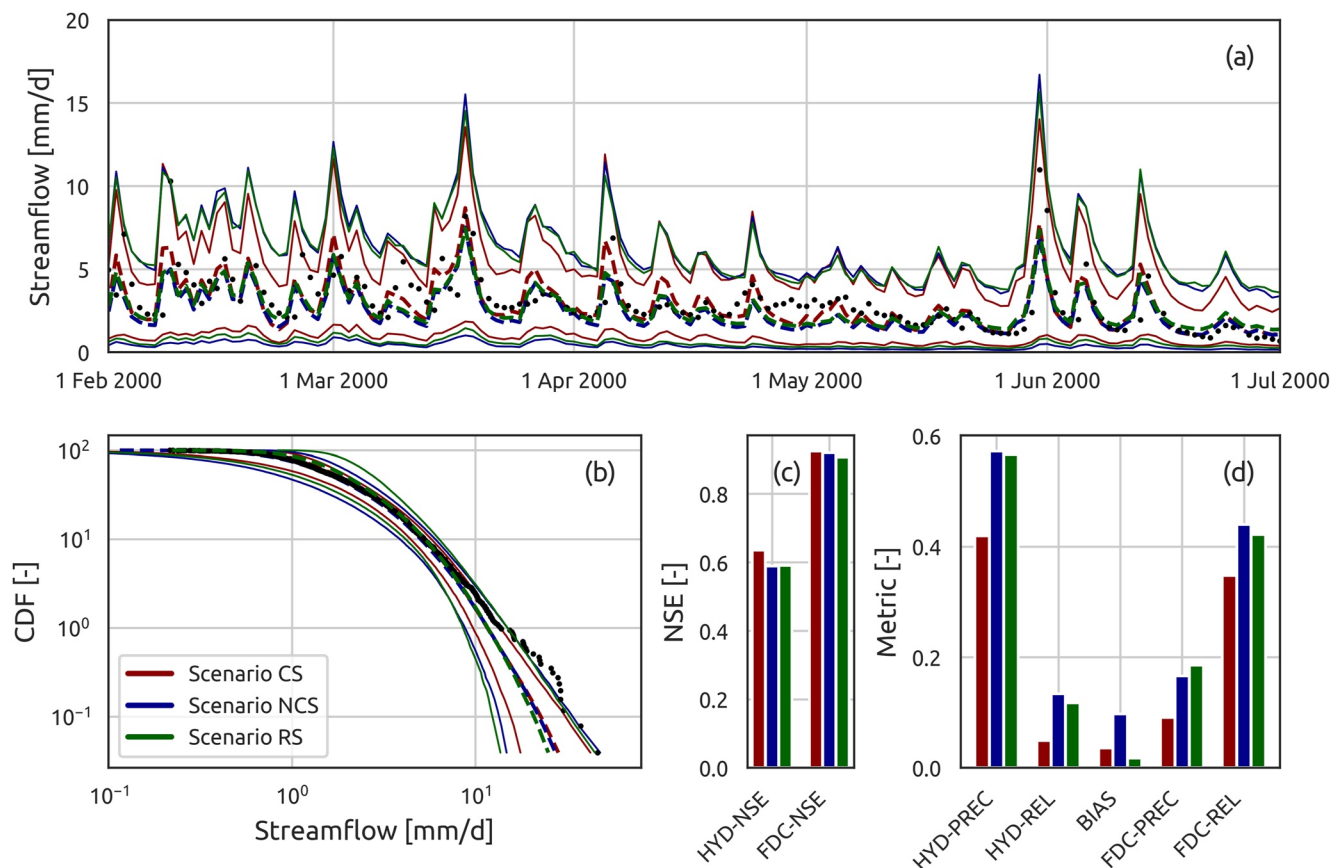


Figure 5. Results of Experiment 1 for a representative catchment (Andelfingen) and time period (P3). Panel (a) shows the hydrograph, panel (b) the annual flow duration curve, panels (c and d) the metrics used for evaluating the simulations. In panels (a and b), the dashed line represents the median simulation; the solid lines represent 95% uncertainty bounds; the black dots represent the observed data.

the logarithmic scale of the vertical axis, the highest 1% of the streamflow values occupies almost half of the plot area—hence a large visual mismatch may result in a limited difference in the metrics.

These qualitative findings are confirmed by the quantitative metrics shown in panels (c and d):

- The reliability of the FDC predictions is generally worse than for the hydrographs, with best values obtained for scenario CS (0.35) followed by RS (0.42) and NCS (0.44).
- The FDC precision metric has similar values for scenarios NCS and RS (0.17 and 0.18) while it improves for scenario CS (0.09).
- The FDC NSE slightly declines when going from scenario CS (0.92) to NCS (0.92) and RS (0.91).
- The volumetric bias is substantially higher in scenario NCS and confirms the general tendency of the simulations, seen in the FDCs, to underestimate streamflow.

5.1.2. Performance Over All Catchments

Figure 6 shows the performance metrics calculated for all the catchments in all periods. The figure is subdivided into seven panels, with the left column showing the metrics calculated on the hydrograph and the right column showing the FDC metrics. Each point in the plots represents the value of the metric for the simulations of a specific catchment and period. Scenarios are evaluated based on the mean position and spread of the corresponding group of points (metric values). A position toward the right-hand side of the plot indicates better mean performance (note that for precision, reliability, and volumetric bias, the horizontal axes are flipped so that “better” results are on the right, for consistency with the other metrics). The spread of metric values reflects the consistency of simulations for the given scenario.

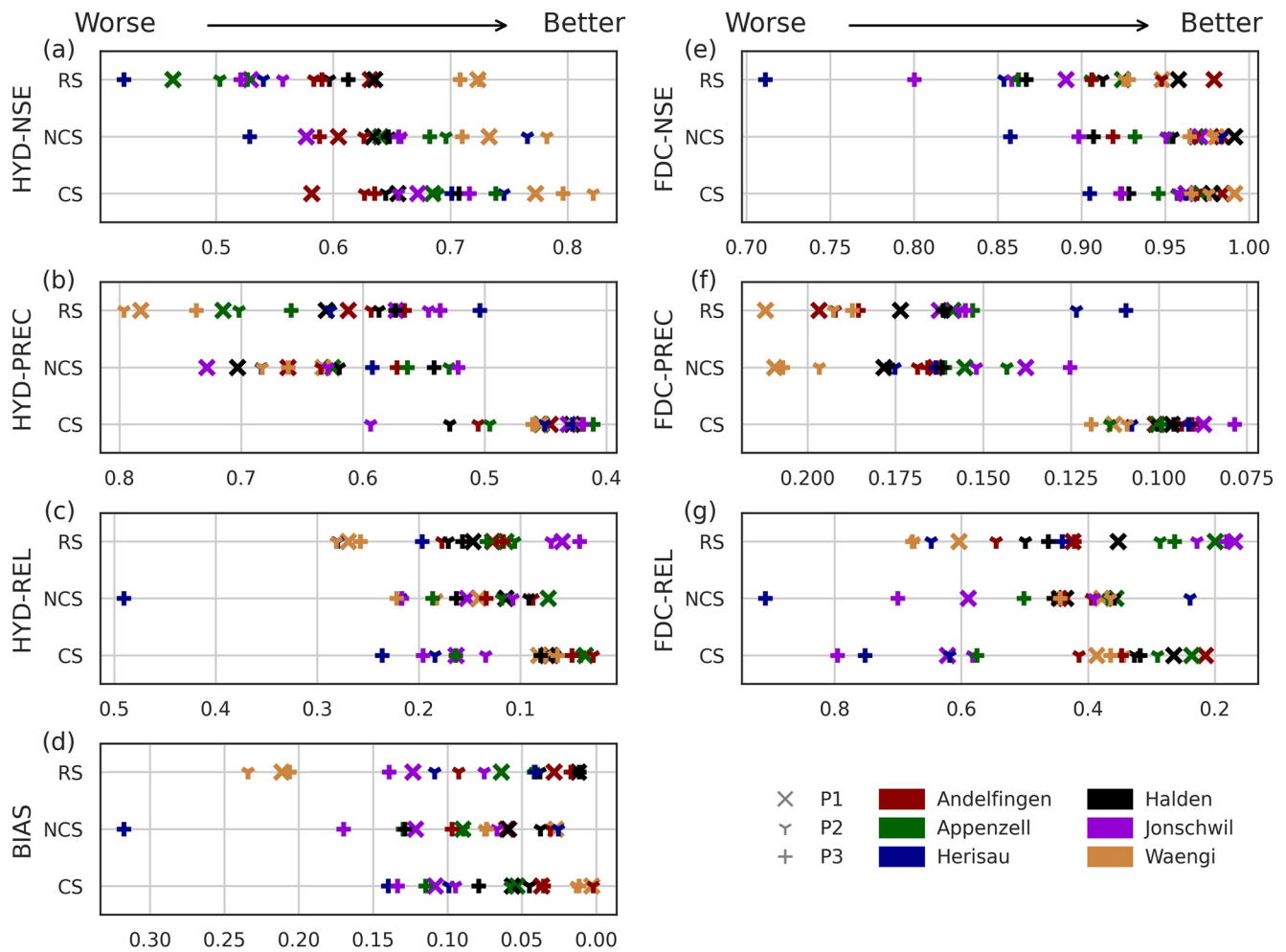


Figure 6. Performance metrics achieved by the simulations in Experiment 1. The left column reports the metrics calculated on the hydrograph; the right column reports the metrics calculated on the annual flow duration curve. Catchments are distinguished by color and periods are distinguished by symbol. Note that some horizontal axes are reversed so that “better” metric values appear consistently on the right in all plots.

Panel (a) shows the NSE of the median hydrograph. All scenarios produce a similar spread, but the mean position of the points shows a clear ranking: scenario CS in the first place (mean value of 0.70), NCS second (0.66), and RS third (0.58). In all scenarios, the points that refer to the same catchment (same color) tend to be clustered together.

Panel (b) shows the precision metric of the hydrograph. In this case, there is a clear trend among the scenarios, both in terms of mean position and spread, with scenario CS achieving the “best” performance (mean precision of 0.46 with standard deviation of 0.047), NCS coming second (mean of 0.62, standard deviation 0.058) and RS third (mean 0.63, standard deviation 0.084).

Panel (c) and (d) show the reliability metric of the hydrograph and the volumetric bias. For both metrics, all three scenarios perform similar in terms of mean value and spread. It is noted that the simulations in the catchment Herisau in period P3 in scenario NCS have considerably worse performance metrics than the other catchments in the same scenario.

Panel (e) shows NSE of the median FDC. There is a clear trend in the scenarios: CS achieves the highest mean (0.96) with low spread (standard deviation 0.023), NCS achieves a similar mean (0.95) but increases the spread (standard deviation 0.035), and RS deteriorates somewhat in the mean (0.89) and in the standard deviation (0.062).

Panel (f) shows the precision metric of the FDC. Scenarios NCS and RS achieve a similar performance in terms of mean (0.17 for both) and spread (standard deviation 0.022 and 0.025), with scenario RS that has two outliers

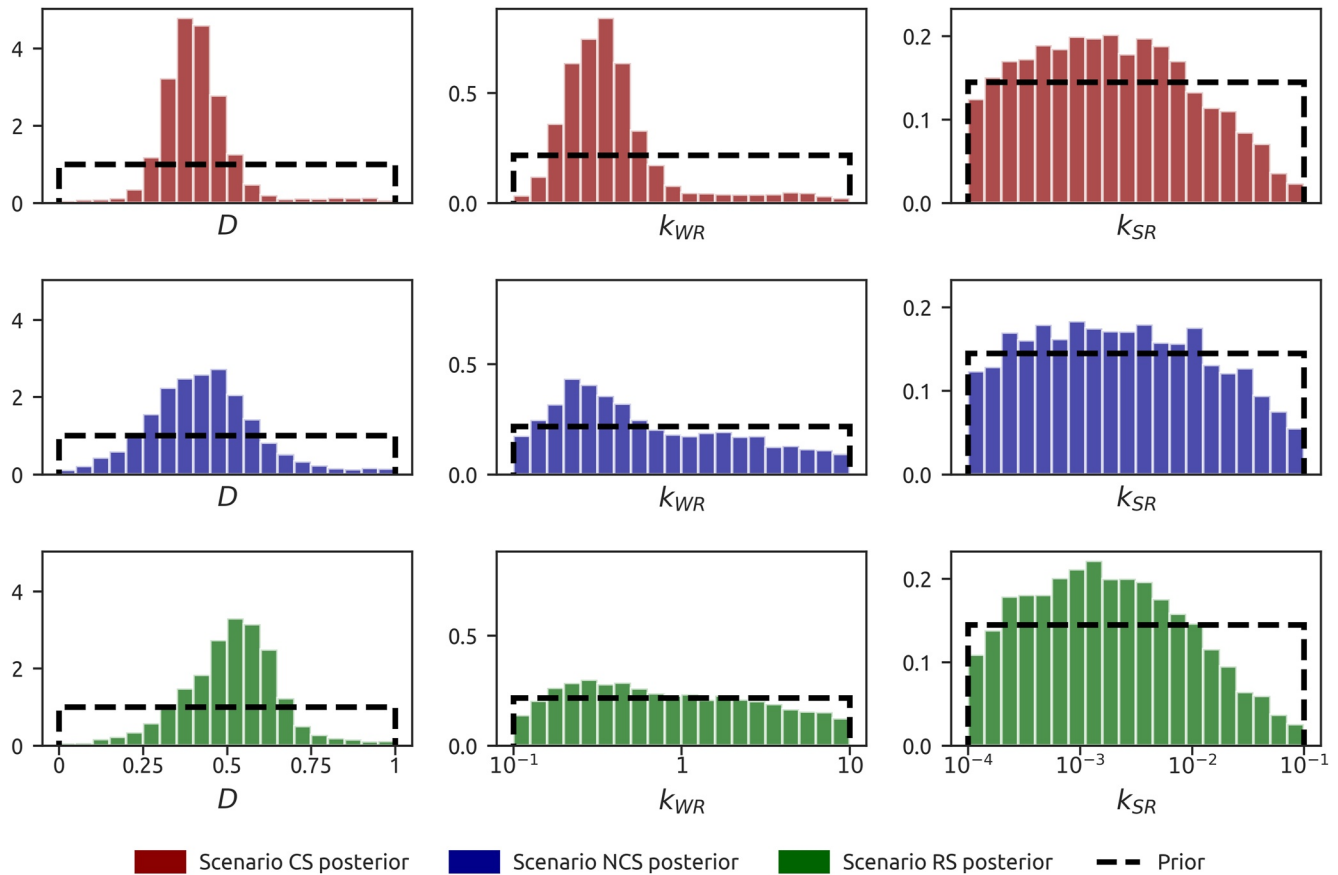


Figure 7. Posterior distributions of three selected parameters, demonstrating qualitatively different behaviors in the scenarios. The histogram bars represent the posterior distributions. The black dashed lines represent the prior distributions.

that perform substantially better than the rest. Both scenarios NCS and RS are strongly outperformed by scenario CS (mean 0.10, standard deviation 0.01), with almost no simulation performing better in scenarios NCS and RS than in scenario CS.

Panel (g), finally, shows the reliability metric of the FDC. All the scenarios perform similarly, with the presence of one bad outlier in the scenario NCS (Herisau, period P3).

5.1.3. Posterior Distribution of the Model Parameters

Figure 7 shows the posterior distributions of selected parameters in the three scenarios. The parameters have been selected as representative of three common behaviors:

1. Parameters that are identifiable in all scenarios, with posterior distributions that are clearly distinct from the priors. An example is given by parameter D , which controls the split of outflow from the UR between the fast and the slow reservoirs.
2. Parameters that are well identified in scenario CS but lose identifiability in scenarios NCS and RS. An example is given by parameter k_{WR} , which controls the outflow of the snow reservoir.
3. Parameters that are weakly identified by all the scenarios, with the posterior distributions that differ little to moderately from the priors. An example is given by parameter k_{SR} , which controls the outflow of the SR.

In terms of the other model parameters, β_{UR} and k_{FR} follow the behavior of the first group, and S_{UR}^{\max} follows the behavior of the second group.

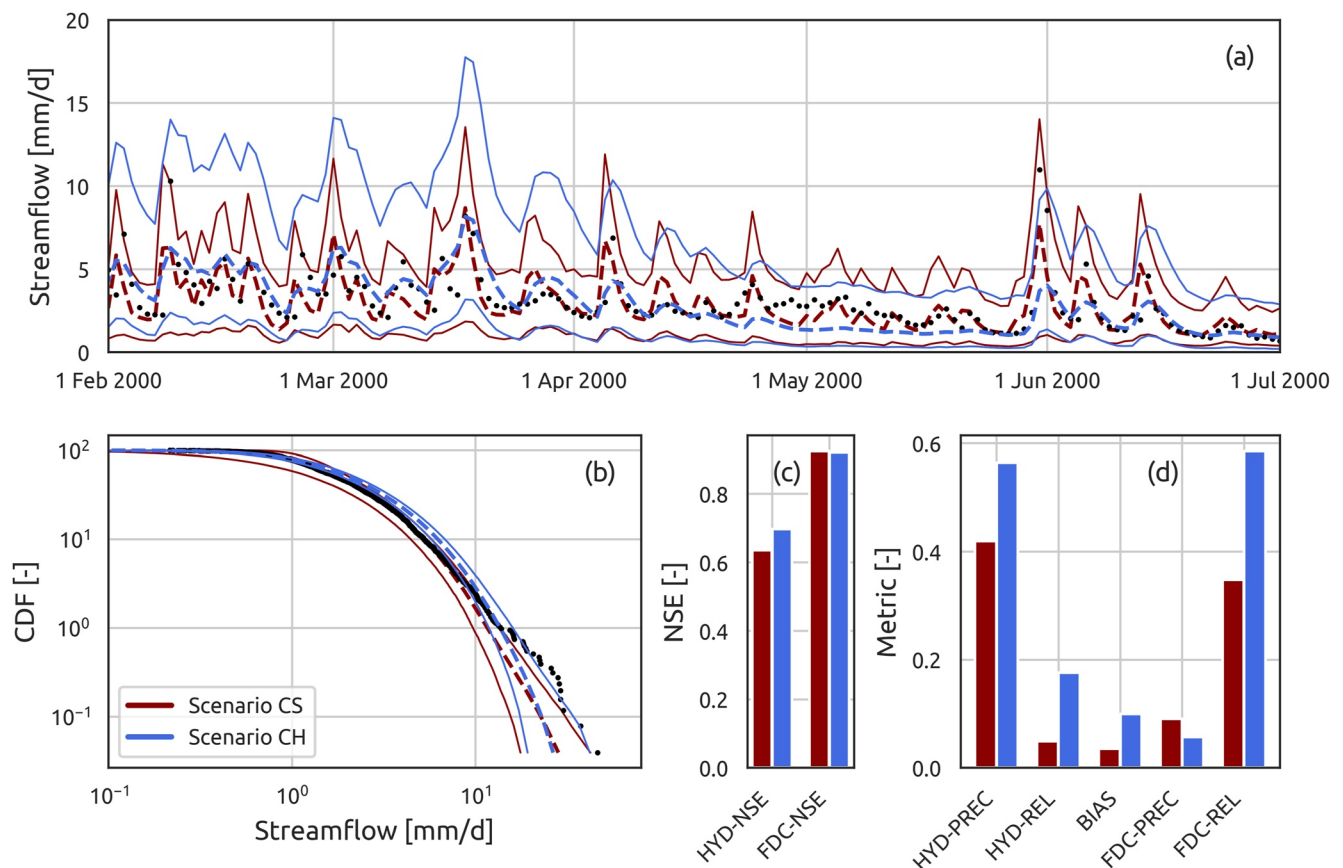


Figure 8. Results of Experiment 2 for a representative catchment (Andelfingen) and time period (P3). Panel (a) shows the hydrograph, panel (b) the annual flow duration curve, panels (c and d) the metrics used for evaluating the simulations. In panels (a and b), the dashed line represents the median simulation; the solid lines represent 95% uncertainty bounds; the black dots represent the observed data.

5.2. Experiment 2: Comparison With Time-Domain Calibration

5.2.1. Representative Catchment: Hydrograph and FDC Representation

Figure 8 compares the results of two gauged scenarios in catchment Andelfingen, in period P3. The two gauged scenarios are model calibration to observed signatures (scenario CS) and model calibration to the observed time series (scenario CH).

Panel (a) shows the simulated hydrographs. Reliability is better in scenario CS than in scenario CH (metric values of 0.049 vs. 0.18). Scenario CS also achieves the tightest (best precision) hydrograph, as confirmed by the precision metric shown in panel (d). However, a closer inspection of the hydrograph indicates that while scenario CH yields a larger uncertainty in winter, it has lower uncertainty in summer. This behavior is not specific to the selected catchment and period, but is generally common across all catchments, especially in years when precipitation is scarce during late spring.

The median hydrograph (dashed line) is more dynamic in scenario CS and captures better the magnitude of the peaks. Nevertheless, panel c shows a slightly better model performance in scenario CH (NSE 0.69 vs. 0.64 for scenario CS).

Panel (b) shows the simulated FDCs. The predictive distribution generated in scenario CH tends to be more precise, especially for low streamflow. On the other hand, the median of scenario CS is closer to the observed data than scenario CH, which tends to overestimate the observed values, particularly for low streamflow. Both simulations tend to underestimate high streamflow. These visual results are confirmed by the performance metrics (panels c and d): the FDC precision is better in scenario CH (0.06 compared to 0.09), the (hydrograph) volumetric

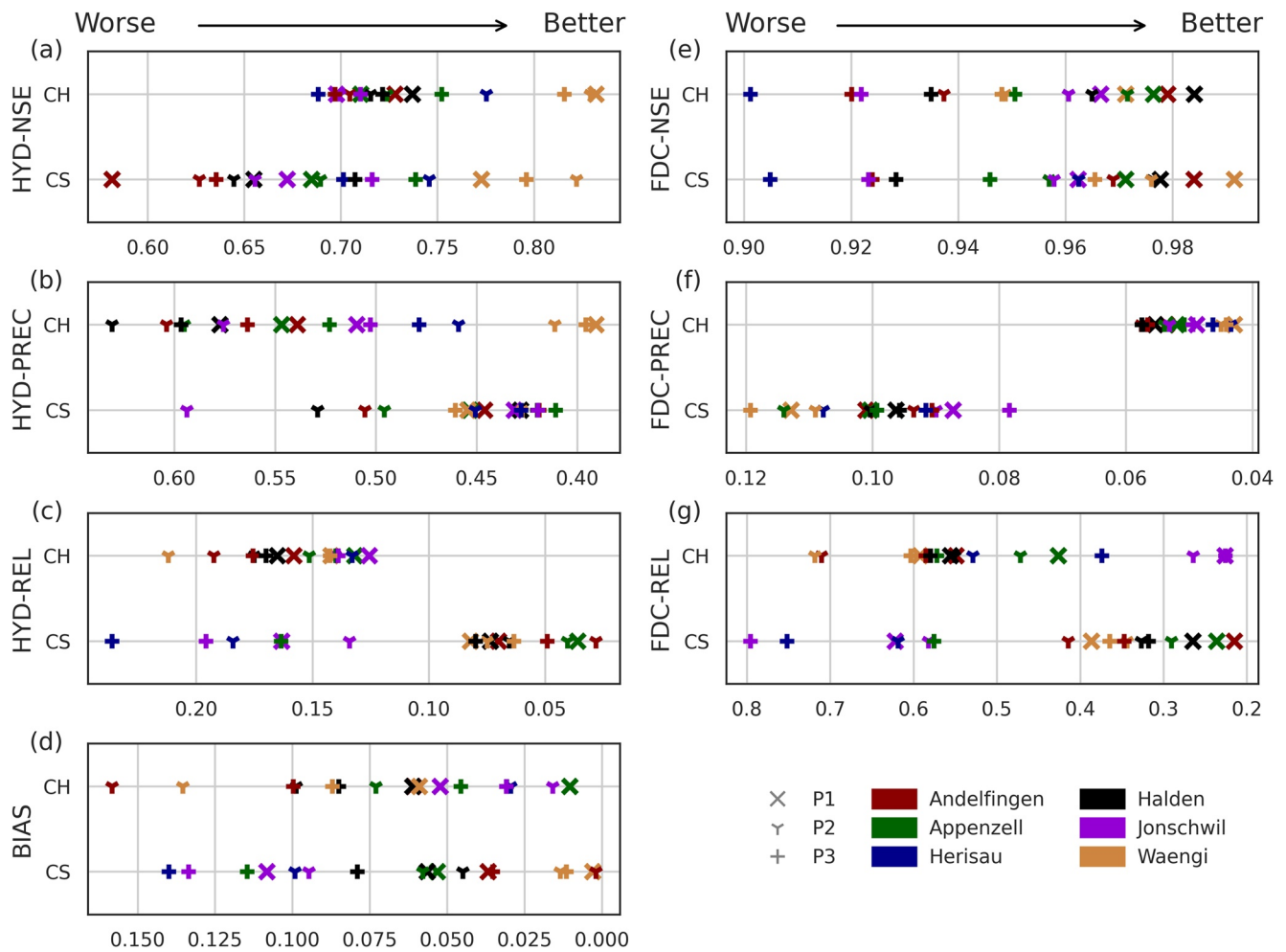


Figure 9. Performance metrics achieved by the simulations in Experiment 2. The left column reports the metrics calculated on the hydrograph; the right column reports the metrics calculated on the annual flow duration curve. Catchments are distinguished by color and periods are distinguished by symbol. Note that some horizontal axes are reversed so that “better” metric values appear consistently on the right in all plots.

bias is worse in scenario CH (0.10 compared to 0.04), and the NSE of the FDCs are similar. Finally, FDC reliability is higher in scenario CS than in scenario CH (0.35 vs. 0.58).

5.2.2. Performance Over All Catchments

Figure 9 reports the performance metrics for scenarios CS and CH for all catchments and periods considered in Experiment 2. From these plots, it is difficult to determine which scenario produces the best “overall” results. In particular:

- Volumetric bias, FDC Nash-Sutcliffe and reliability (panels d, e, and g): scenarios achieve a comparable performance, both in terms of spread and mean value.
- Hydrograph NSE and FDC precision (panels a and f): the simulations in scenario CH outperform those in scenario CS, with a better mean value and a lower spread for both metrics.
- Hydrograph precision and reliability (panels b and c): the simulations in scenario CS outperform those in scenario CH. In particular, in terms of hydrograph precision, scenario CS shows better mean and spread. In terms of hydrograph reliability, scenario CS has a better mean but the presence of some outliers increases broadly the spread.

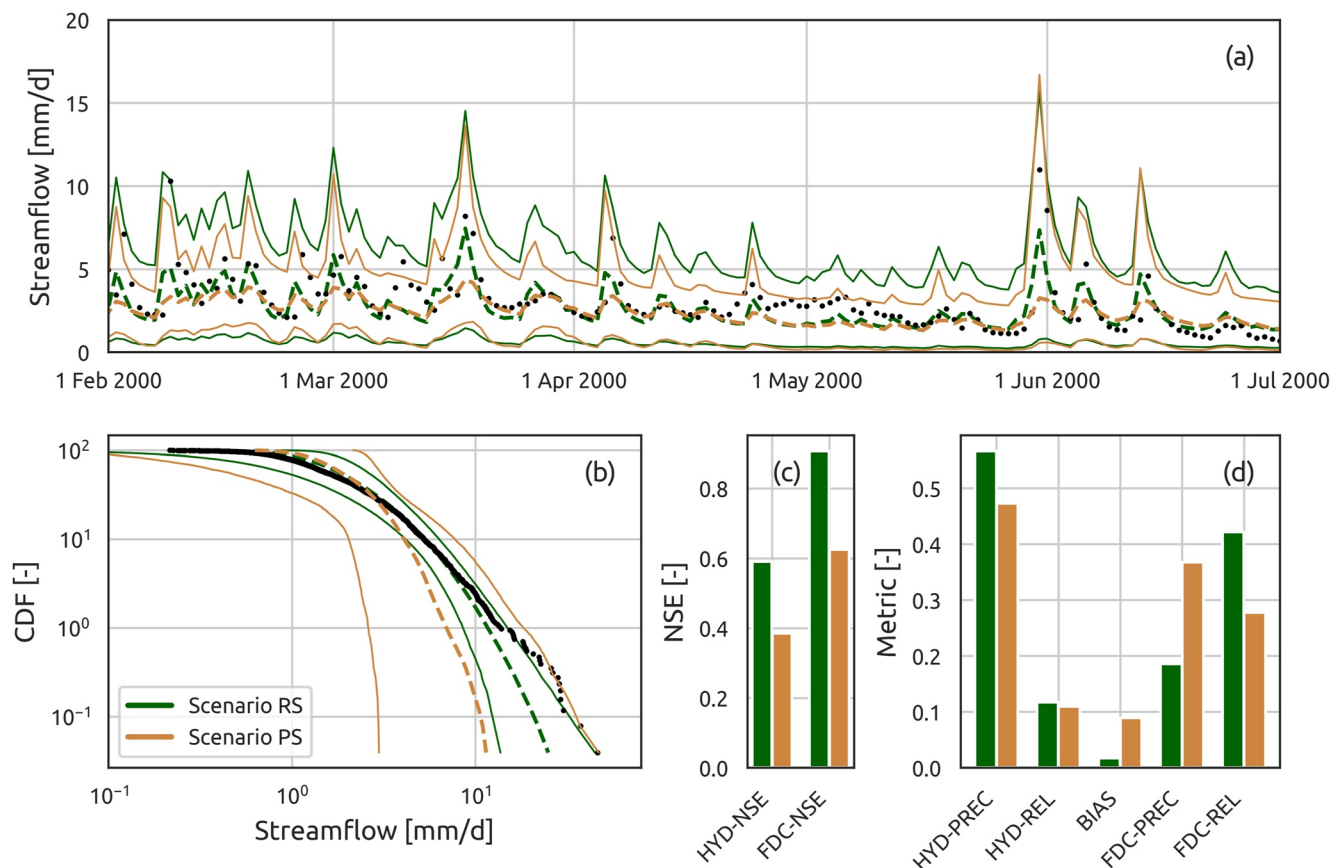


Figure 10. Results of Experiment 3 for a representative catchment (Andelfingen) and time period (P3). Panel (a) shows the hydrograph, panel (b) the annual flow duration curve, panels (c and d) the metrics used for evaluating the simulations. In panels (a and b), the dashed line represents the median simulation; the solid lines represent 95% uncertainty bounds; the black dots represent the observed data.

5.3. Experiment 3: Goodness of the Regionalization Study

5.3.1. Representative Catchment: Hydrograph and FDC Representation

Figure 10 compares the results of scenario RS with scenario PS (prior simulations) in catchment Andelfingen, period P3.

Panel (a) shows the simulated hydrographs. Although scenario PS has narrower uncertainty bounds, we emphasize, as noted in Section 2.1, that in this scenario the standard deviation of the residual error model (σ_{Ω}) is set to zero and, therefore, streamflow uncertainty is represented solely through parametric uncertainty.

The median hydrograph (dashed line) is more dynamic and closer to the observed data in the scenario RS. This behavior is confirmed by the large difference in NSE (panel c) between the two simulations (0.59 for scenario RS, 0.38 for scenario PS).

The biggest difference between the two scenarios can be observed in panel (b), which shows the simulated FDCs. The predictive distribution of the scenario RS, as already analyzed in Section 5.1, is consistent with the observations, in terms of the fit of the median as well as the tightness and distribution of the uncertainty bounds. In contrast, in scenario PS, the median of the predictive distribution is far from the observed values and the uncertainty bounds are extremely wide.

5.3.2. Performance Over All Catchments

Figure 11 presents the performance metrics for scenarios PS and RS. Excluding hydrograph precision (panel b) and FDC reliability (panel g), the scenario RS achieves performance metrics that are always better, or at least similar, to those in scenario PS. This finding is particularly evident in panels (a and e), which show that scenario

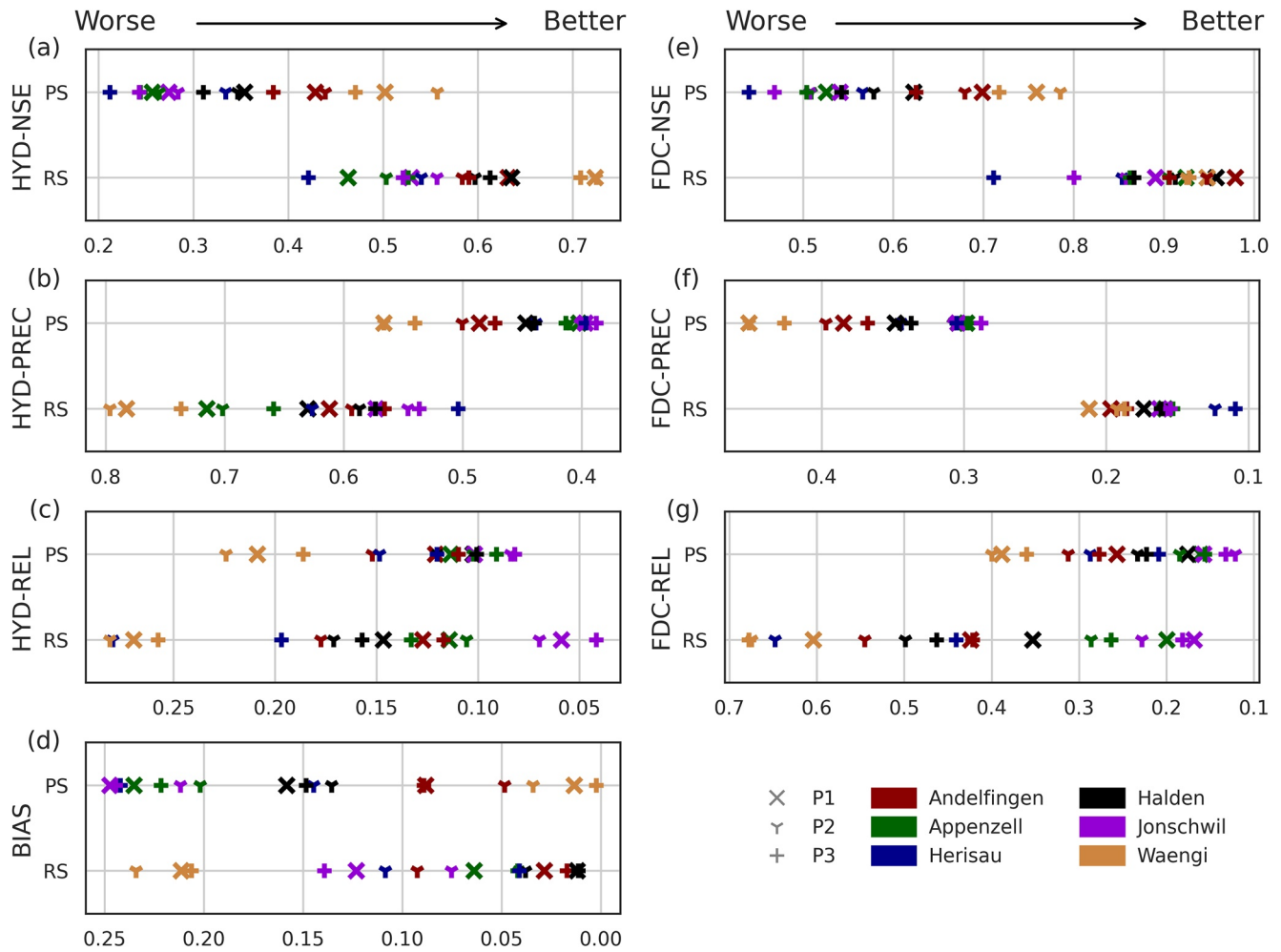


Figure 11. Performance metrics achieved by the simulations in Experiment 3. The left column reports the metrics calculated on the hydrograph; the right column reports the metrics calculated on the annual flow duration curve. Catchments are distinguished by color and periods are distinguished by symbol. Note that some horizontal axes are reversed so that “better” metric values appear consistently on the right in all plots.

RS has higher Nash-Sutcliffe values for both hydrograph and FDC, and in panel (f), which shows that scenario RS achieves a more precise FDC. Hydrograph reliability and volumetric bias (panels c and d), on the other hand, show similar performance for scenarios PS and RS. In particular, hydrograph reliability has the same mean in these two cases, and less spread in scenario PS than in scenario RS; volumetric bias has a lower mean and larger spread for scenario PS than for scenario RS. Finally, panels (b and g) show that scenario PS outperforms scenario RS in terms of a better mean and narrower spread of metric values.

6. Discussion

6.1. Experiment 1: Signature Calibration in Different Scenarios

This experiment compares the results of the three signature-calibration scenarios: concomitant (CS), non-concomitant (NCS), and regionalized (RS). We first discuss aspects associated with the quality of model predictions, then the impact of errors in the signatures, and, finally, parameter identifiability.

6.1.1. Quality of Model Predictions

Model predictions are expected to become increasingly uncertain when moving from gauged to progressively more challenging ungauged conditions. Hence, scenario CS is expected to have the smallest uncertainty, because it uses signatures derived from streamflow observed directly at the catchment and period of interest. Scenario

NCS is expected to yield higher uncertainties than scenario CS, because transferring the signatures in time requires the use of a signature transfer model, which is subject to uncertainty. Scenario RS is expected to yield the highest uncertainty, because it uses signatures estimated from streamflow measurements in donor catchments using a stochastic FDC model, and transfer of signatures in space is arguably more challenging than transfer in time.

This pattern of increasing uncertainty emerges clearly in Figure 6, and is consistent both for hydrographs (panels a and b) and FDCs (panels e and f). In terms of NSE and precision, both for hydrographs and FDCs, performance is highest for scenario CS, followed by scenario NCS, and then scenario RS. On average, the progression from concomitant to RS increases uncertainties in reproducing time series and FDCs and worsens the model fit to the observations. In particular, taking Scenario CS as the (baseline) reference, Scenario NS incurs 36% more uncertainty (i.e., worse precision), and a 5.6% decrease in streamflow NSE, while Scenario RS incurs 38% more uncertainty and a 17% decrease in streamflow NSE. These values are calculated by taking the average of the hydrograph performance metrics shown in Figure 6.

However, from our perspective in this study, a key objective is to provide a reliable description of uncertainty, essentially regardless of its magnitude. As our focus is on streamflow prediction, it is important that uncertainties in reproducing streamflow time series are not underestimated or overestimated, which would lead to overconfidence or underconfidence respectively.

The results in Figure 6 are reassuring, showing that the Bayesian framework is able to achieve comparable reliability and bias in all scenarios, both in terms of hydrographs (panel c) and FDCs (panel g). Reliability performance in an ungauged catchment is sensitive to the correct quantification of uncertainty in the signature transfer model. Such correct quantification is challenging, as for example, extreme events can make the transfer of signatures in time unreliable. This is the case of the catchment Herisau, which has poor reliability in scenario NCS (panels c and g) and will be discussed more in detail in the following section.

Comparison of our results with previous work is limited by differences in methodologies as well by the relatively small number of studies that analyzed PUB under different data availability scenarios. The study of Montanari and Toth (2007) is directly related to our work as it compared the results of time-domain calibration, concomitant (scenario CS, in this study) and non-concomitant (scenario NCS) signature calibration, and prior simulations. Although our inference procedure is different and the case study is based on different signatures, the findings are broadly consistent. In particular, we show an average decrease of streamflow NSE of approximately 5.6% when moving from concomitant to NCS, and that the magnitude of this decrease is highly dependent on the conditions of the period considered.

Biondi and De Luca (2016) compared calibration to observed signatures (our scenario CS) with calibration to RS (our scenario RS), with the objective of estimating design floods for an assigned return period. Their results showed that, for their objective, the use of RS does not deteriorate model performance. Our study, instead, shows a better precision in scenario CS than in scenario RS, but also comparable performance in other metrics (e.g., FDC reliability and volumetric bias).

6.1.2. Impact of Errors in the Signatures

When signatures are derived using data from a different time period or catchment, their numerical value can be very different from the value computed from observed hydrographs in the catchment and time period of interest. This difference is due to the uncertainty associated with the signature model. Time transfer of signatures incurs uncertainty due to time variability of hydrological behavior, whereas signature regionalization uncertainty is affected by unaccounted spatial variability of the signatures.

This study shows the effect of errors associated with the transfer of signatures in time or space:

- *Transfer in time.* In scenario NCS some of the catchments had model performance in one period that was far worse than in the other periods. For example, this is the case for the catchment Herisau in period P3 (blue crosses in Figure 6). Poor model performance in this catchment may be due to the large flood in 2002: such extreme streamflow has a large impact on the FDCs, especially considering that they usually last more than a single day. Therefore, when calibrated using NCS, the model is forced with meteorological inputs that generated the flood but calibrated to signatures that do not contain information about this flood. This mismatch could lead to unrealistic parameter values and explain the mediocre fit.

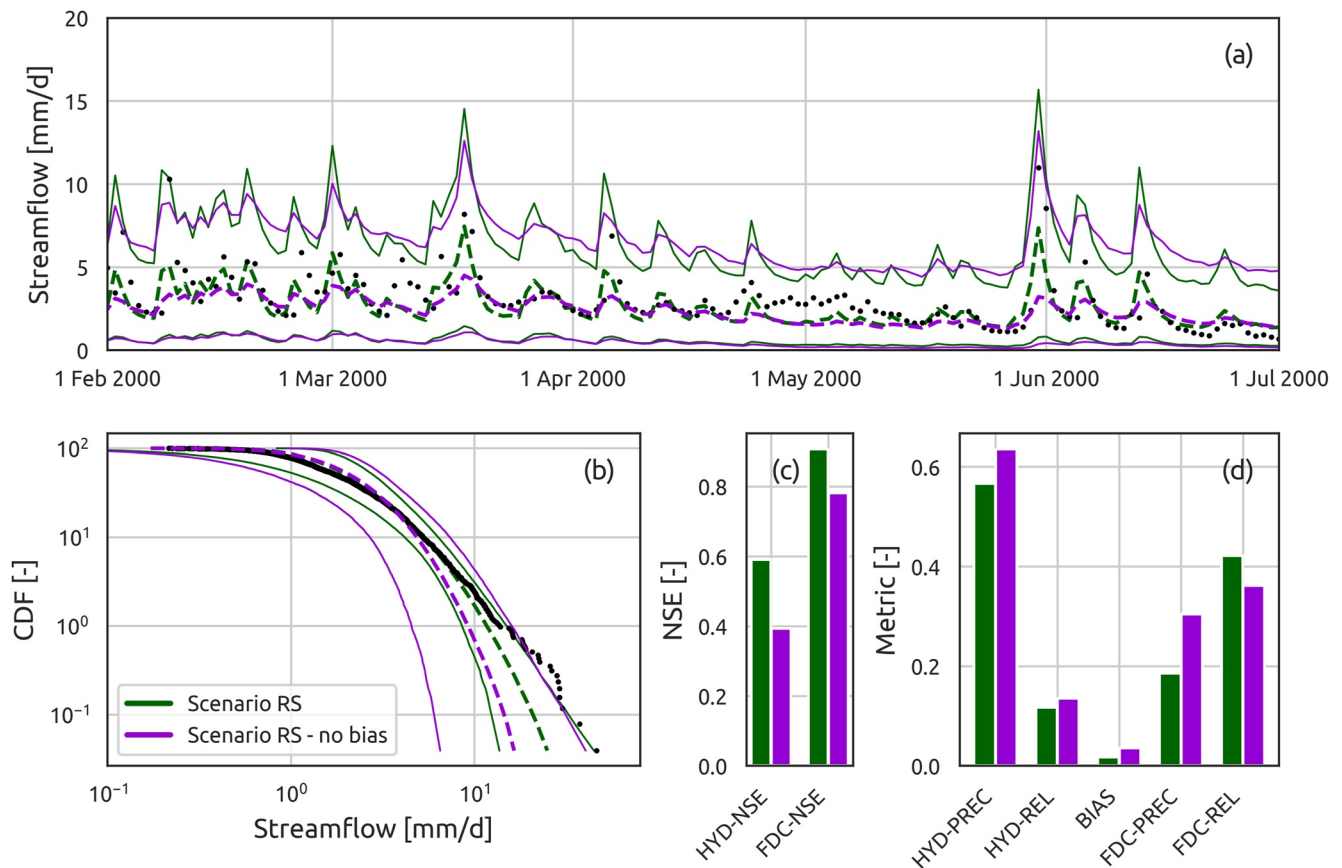


Figure 12. Results of the scenario regionalized signatures with and without bias correction for a representative catchment (Andelfingen) and time period (P3). Panel (a) shows the hydrograph, panel (b) the annual flow duration curve, panels (c) and (d) the metrics used for evaluating the simulations. In panels (a) and (b), the dashed line represents the median simulation; the solid lines represent 95% uncertainty bounds; the black dots represent the observed data.

- *Transfer in space.* The FDCs model has some apparent deficiencies in capturing the observed FDCs, already discussed in Section 4.4.3. This deficiency has led to our decision to augment the FDC model with a bias correction term. Figure 12 illustrates the effect of not implementing bias correction, and compares the results of scenario RS with and without bias correction ($a_e^{(T)} = 1$ in Equation 20). Overall, results are worse if bias correction is omitted, manifesting in increased uncertainty in the hydrograph and FDC, reducing the fit of their medians to observed data, and suffering worse performance in almost all the metrics.

The two examples above highlight that a small difference in estimated signatures can lead to a large difference in streamflow predictions. This sensitivity has also been noted in previous studies. For example, Westerberg et al. (2011) discussed the concept of “disinformative” signatures, Castiglioni et al. (2010) showed the impact of errors in the RS due to deficiencies in the regionalization model, and Fenicia et al. (2018) demonstrated the effect of time shifts in the time series that are not captured by the signatures. All these studies reported that the presence of small errors in the signatures may lead to large differences in the model predictions, as shown in this study.

6.1.3. Parameter Identifiability

Parameter identifiability generally depends on the calibration scenario. In particular, as the signatures on which the model is calibrated become more uncertain, model parameters become less identifiable. However, such behavior is not uniform for all model parameters, but strongly depends on the specific parameter, on the processes that it is intended to represent, and on whether such processes are captured by the signatures used for model calibration. This behavior is shown in Figure 7, which illustrates how some parameters lose identifiability when calibrating to estimated signatures. The following interpretation can be provided for specific cases:

- *Parameters that remain identifiable in all scenarios.* These are parameters that have a strong influence on the model output. For example, parameter D which controls the splitting of the flow between the “fast” and the

“slow” part of the model. Changes in this parameter would strongly affect the FDCs in all scenarios, since the slope of the FDC can be related to the proportion of quick and slow response in the catchment.

- *Parameters that lose their identifiability when moving to an ungauged scenario.* These are parameters that exert an influence on the signatures, but this influence may be blurred by higher uncertainty in the target signatures. For example, parameter k_{WR} , which controls the outflow of the snow reservoir, exhibits this behavior. The transfer error may contribute in obscuring this behavior, but the omission of snow dynamics in the regionalization model can also be responsible; this is the case of the FDCs model, which does not distinguish explicitly the separation between rainfall and snowfall.
- *Parameters that are weakly identifiable, regardless of the scenario.* These are parameters that have an effect on the model output that is not captured by the signatures. For example, consider parameter k_{SR} , which controls the release rate of the “slow” part of the model. This parameter has a strong effect on the hydrograph, by affecting the baseflow component, but its effect on the FDCs may be less evident. This change in behavior compromises its identifiability when the model is calibrated to FDCs. A different choice of signatures (e.g., including the baseflow index) would have probably helped in this specific case, but this analysis is not in the scope of this paper.

Similar results regarding parameters identifiability have been highlighted by Biondi and De Luca (2016), who showed that using observed signatures makes the posterior distributions sharper than when using modeled signatures.

6.2. Experiment 2: Comparison With Time-Domain Calibration

The aim of Experiment 2 is to assess the overall performance of signature calibration, in its best scenario (CS), using the classical time-domain calibration as a benchmark.

The results in Figure 9 (Section 5.2.2) show that scenarios CS and CH present a tradeoff in terms of fit to the observed data, uncertainty, and performance metrics. For example, scenario CS outperforms scenario CH in terms of hydrograph precision and reliability, but rank lower in terms of hydrograph NSE. This finding indicates that calibration to seasonal FDCs is not equivalent to calibration to streamflow time series, and that the choice of signatures needs a careful assessment. Note that a comprehensive comparison between signature and time domain calibration is beyond the scope of this study (e.g., see Fenicia et al., 2018 for an earlier dedicated investigation).

Our findings are in line with previous studies (e.g., Castiglioni et al., 2010; Kavetski et al., 2018), which have shown that signature calibration can achieve similar or slightly inferior performance to time-domain calibration and that, in some cases, the FDCs are a viable surrogate for the hydrograph in the context of model calibration. On the other hand, Kim et al. (2017) pointed out how the lack of time information in the FDCs can reduce their utility, and suggested the combination of FDCs with other signatures, such as the flashiness index, that include timing information (see also Fenicia et al. (2018)).

The study of loss of information due to the usage of different signatures is, however, a broad and important topic that goes beyond the scope of this paper and requires much further work.

6.3. Experiment 3: Goodness of the Regionalization Study

The last experiment compares the performance of simulations in the ungauged scenario RS with simulations generated using the prior distribution of model parameters. The objective is to assess the performance of the regionalization approach, comparing it with an “easy” benchmark, in the sense that the effort and the data needed to run prior simulations are limited and can be done in any ungauged catchment as long as meteorological forcing is available.

The results in Figure 10 (Section 5.3.1) show that the simulations in scenario RS generally outperform the simulations in scenario PS, though not without some subtleties. In terms of hydrographs, the two scenarios have similar performance, with scenario RS achieving a better fit to the observed data and scenario PS achieving better precision. However, in terms of FDCs, model performance in scenario PS drops substantially: the simulated FDC looks clearly unrealistic, with huge uncertainty and large deviations from observed data. These results are confirmed by the analysis overall catchments, shown in Figure 11 (Section 5.3.2).

As such, Figure 10 shows that similar uncertainty bounds in the hydrograph space can correspond to very different uncertainty bounds in the signatures space. In particular, the difference between scenarios PS and RS is not apparent when considering the total uncertainty bounds in the hydrograph space, where both scenarios have similar precision. However, it becomes obvious in the signature space, where scenario PS have much wider uncertainty bounds than RS in terms of FDCs, which is exemplified by a much lower precision. The cause of this difference in behavior is as follows. In scenario PS, predictive uncertainty is determined purely by the uncertainty in the precipitation-streamflow model parameters. PS produces hydrographs with a wide range of dynamics (i.e., ranging from relatively constant to very responsive), which translate into FDCs with very different shapes (i.e., ranging from flat to steep) and hence a wide range of variability. In scenario RS, instead, predictive uncertainty is largely determined by the residual error. This scenario produces hydrographs with much more uniform dynamics, which map to similar FDCs.

6.4. Limitations and Future Work

This study has a number of limitations that warrant follow-up research.

First, the case study catchments in this work all belong to the same geographic region (the Thur basin). For this reason, they can be thought to be similar, or at least less varied than catchments in different climatic regions. Nevertheless, the Thur basin does exhibit appreciable variability in physical and local hydroclimate characteristics, which in turn manifests as appreciable differences in mean streamflow, in seasonality, and in baseflow (see Dal Molin, Schirmer, et al. (2020) for further details). Future studies could consider large-sample hydrology datasets (e.g., Addor et al., 2017) with wider catchment variability, including arid conditions. This work would likely require revisiting key modeling choices, including the choice of precipitation-streamflow models, signature transfer models and model performance metrics (e.g., focusing on low flows). Moreover, given the large computational cost of ABC inference (millions of model runs), efficient numerical and software implementations, including parallel computation, would become paramount.

Second, the case study catchments are located in a well-studied area, where previous work could be exploited for the selection of the streamflow model (Dal Molin, Schirmer, et al., 2020) and the FDC model (Doulatyari et al., 2017). The choice of donor catchments also gives an opportunity, in scenario RS, to assume a similar behavior of the regionalization model in the catchments and, therefore, to estimate the parameters of the signatures error model using other catchments of the network. The application of the proposed methodology in less instrumented donor catchments may be more challenging and deserves future attention.

Third, our study focuses primarily on reliable estimation of streamflow predictive uncertainty rather than on its reduction. Further research is needed to understand the choice of signatures, signature estimation models, and precipitation-streamflow models that minimizes streamflow uncertainty. In this endeavor, the use of alternative measurements, such as independent remote sensing observations (e.g., Nijzink et al., 2018; Winsemius et al., 2008) or other types of constraints (e.g., Gharari et al., 2014) are of major interest.

Fourth, the deterministic FDC model does not account for snow dynamics. When precipitation is used in the signature model to calculate its parameters (Section 4.4.3), no distinction is made between rainfall and snowfall. In this work, this limitation does not appear problematic, given the limited amount of snow in the case study area (at most 20% of annual precipitation). Moreover, the stochastic model includes a bias correction that may partly compensate for such deficiency. Importantly, note that the models calibrated in snow affected catchments (e.g., Appenzell) do not perform systematically worse than the ones of catchments with little snow (e.g., Wängi). Nevertheless, future work should incorporate snow dynamics more explicitly into signature regionalization models.

Finally, this work treats the precipitation-streamflow model and the signature model as conceptually independent from each other. However, it may be of interest to consider the conceptual connections between these types of models. For example, both the signature model of Doulatyari et al. (2017) and the HyMod model represent the catchment as conceptual reservoirs. Future work could explore these similarities and their impact on signature estimation.

7. Conclusions

This study presented a new approach for calibrating a precipitation-streamflow model to regionalized streamflow signatures. The approach relies on ABC, which avoids direct calculation of the likelihood function (instead, it employs sampling from the probability model, which in this case is easier to implement). We presume the availability of a signature model, which predicts the signatures at the target catchment. The main intended application of the new method is for streamflow time series prediction in ungauged catchments, where current methods are limited by not explicitly accounting for the uncertainty in the RS. We show how to estimate uncertainty in the RS, and how to incorporate this uncertainty into the calibration of a precipitation-streamflow model.

The proposed framework is evaluated in a case study based on six catchments of the Thur basin, Switzerland. This catchment has high quality data, a long history of hydrological studies, and high operational importance in the water supply of the region. In order to test the quality of predictions in ungauged conditions, we followed a progression from gauged to ungauged, with three scenarios: “concomitant,” where signatures are observed, “non-concomitant,” where signatures are extrapolated in time, and “regionalized,” where signatures are extrapolated in space. For reference, we also calibrated the model in the time domain and run simulations using parameters sampled from the prior distribution.

The results of the experiments suggest that:

1. In line with expectations, the progression from concomitant to RS increases uncertainties in reproducing streamflow time series (38% more uncertainty) and reduces the fit to the observations (17% decrease in NSE). That said, the uncertainty was reliably estimated in all scenarios, which is an important finding that provides confidence in the proposed approach;
2. Poor quality of model predictions could be attributed to cases where RS are corrupted by large errors, such as in the Herisau catchment where the model is forced with meteorological inputs that generated a flood but calibrated to signatures that do not contain information about this flood. In such cases, the error in estimating the signatures results into streamflow predictions with poor performance metrics (e.g., doubling the volumetric bias);
3. The use of RS may reduce the identifiability of some parameters of the streamflow model (e.g., the release rate of the snow reservoir). This behavior is likely explained by the representation of some processes (e.g., snow) being lost when calibrating to regionalized rather than observed signatures;
4. Signatures based on seasonal FDCs appear to be an adequate choice for the calibration of precipitation-streamflow models, as their use yields streamflow predictions comparable in quality to those obtained in time-domain calibration. However this finding may be specific to the case study and requires further corroboration, including in different hydroclimatic conditions;
5. Calibration to RS generates streamflow predictions of clearly better quality than prior simulations, that is, streamflow predictions obtained using prior parameter distributions. This finding shows that RS contain useful information to constrain model parameters and hence predictions.

The proposed methodology represents a step toward improved predictions in data scarce regions. Future work is needed to generalize the approach to more diverse areas, especially where donor catchments are not available for the calibration of the signature regionalization model, to better understand the impact of signature choice on the results, and to undertake comprehensive comparisons of the proposed methodology to other approaches for PUB.

Appendix A: Streamflow Predictive Distribution

The posterior and prior predictive distributions of the streamflow time series are obtained by sampling from the stochastic model with given parameter sets $\theta^{(H)}$.

A replicate streamflow time series is generated as follows:

1. Take the k th sampled parameter set, $\theta_k^{(H)}$;
2. Run the deterministic precipitation-streamflow model $\mathbf{m}^{(H)}(\theta_{m,k}^{(H)}, \tilde{\mathbf{x}})$ to generate a streamflow time series;
3. Sample a time series of residual errors from the residual error model $\epsilon^{(H)}(\theta_{e,k}^{(H)})$;
4. Substitute the results of Steps 1–2 above into Equation 1 to compute a realization of the precipitation-streamflow model;

5. Repeat for $k = 1, \dots, N_{\text{sam}}$.

The posterior predictive distribution of streamflow is obtained using the posterior (ABC) parameter samples. The prior predictive distribution of streamflow is obtained using samples from the prior parameter distribution.

Appendix B: Precipitation-Streamflow Model Used in the Case Study

This appendix lists the precipitation-streamflow model equations (Appendix B1) and the prior parameter distribution (Appendix B2) used in the empirical case study (Section 4 of the main text).

B1. Model Equations

A schematic of the model is given in Figure 4. The reservoir equations are detailed below.

B1.1. Snow Reservoir

The snow reservoir (WR) is a threshold reservoir with mass balance equation

$$\frac{dS_{\text{WR}}}{dt} = P_{\text{WR}} - Q_{\text{WR}} \quad (\text{B1})$$

where S_{WR} is the storage, P_{WR} is the snow input, and Q_{WR} is the outflow.

The fluxes are given by

$$P_{\text{WR}} = \begin{cases} P & \text{if } T \leq 0 \\ 0 & \text{if } T > 0 \end{cases} \quad (\text{B2})$$

$$M_{\text{WR}}^{\text{max}} = \begin{cases} 0 & \text{if } T \leq 0 \\ k_{\text{WR}} T & \text{if } T > 0 \end{cases} \quad (\text{B3})$$

$$Q_{\text{WR}} = M_{\text{WR}}^{\text{max}} \left[1 - \exp\left(-\frac{S_{\text{WR}}}{m_{\text{WR}}}\right) \right] \quad (\text{B4})$$

where P is the precipitation, T is the (catchment-average) temperature, $M_{\text{WR}}^{\text{max}}$ is the maximum melting rate, k_{WR} is the degree day parameter, and m_{WR} is a smoothing parameter (here, fixed to 2) (Kavetski & Kuczera, 2007).

B1.2. Unsaturated Reservoir

The UR is a threshold reservoir with mass balance equation

$$\frac{dS_{\text{UR}}}{dt} = P_{\text{UR}} - Q_{\text{UR}} - E_{\text{UR}} \quad (\text{B5})$$

where S_{UR} is the storage, P_{UR} is the inflow, Q_{UR} is the outflow, and E_{UR} is the evaporation.

The fluxes are given by

$$\overline{S_{\text{UR}}} = \frac{S_{\text{UR}}}{S_{\text{UR}}^{\text{max}}} \quad (\text{B6})$$

$$E_{\text{UR}} = E_{\text{Pot}} \left[\frac{\overline{S_{\text{UR}}}(1 + m_{\text{UR}})}{1 + m_{\text{UR}}} \right] \quad (\text{B7})$$

$$Q_{\text{UR}} = P_{\text{UR}} \left(\overline{S_{\text{UR}}} \right)^{\beta_{\text{UR}}} \quad (\text{B8})$$

where S_{UR}^{\max} is a parameter indicating the maximum storage of the reservoir, E_{pot} is the potential evaporation input, m_{UR} is a smoothing parameter (here, fixed to 0.01), and β_{UR} is an exponent parameter.

B1.3. Fast Reservoirs

The three “fast” reservoirs (FRs) comprising the fast flow path (see Figure 4) are all linear reservoirs with mass balance equation

$$\frac{dS_{FR}}{dt} = P_{FR} - Q_{FR} \quad (\text{B9})$$

where S_{FR} is the storage, P_{FR} is the inflow, and Q_{FR} is the outflow.

The fluxes are given by

$$P_{FR} = (1 - D)Q_{UR} \quad (\text{B10})$$

$$Q_{FR} = k_{FR} S_{FR} \quad (\text{B11})$$

where D and k_{FR} are model parameters.

B1.4. Slow Reservoir

The SR is also a linear reservoir, with mass balance equation

$$\frac{dS_{SR}}{dt} = P_{SR} - Q_{SR} \quad (\text{B12})$$

where S_{SR} is the reservoir storage, P_{SR} is the inflow, and Q_{SR} is the outflow.

The fluxes are given by

$$P_{SR} = DQ_{UR} \quad (\text{B13})$$

$$Q_{SR} = k_{SR} S_{SR} \quad (\text{B14})$$

where k_{SR} is a model parameter.

B2. Prior Distribution of Parameters

Table A1 lists the calibrated parameters and the bounds of their uniform prior distributions. Parameters with plausible values that span multiple orders of magnitude are inferred in log-space, that is, we infer the logarithm of the parameter; in this case, the uniform prior distribution is applied to the log-transformed value.

Table A1 Details of Parameters of the Precipitation-Streamflow Model				
Parameter	Unit	Model component	Bounds (uniform prior)	Logarithmic transformation applied
k_{WR}	$^{\circ}\text{C}^{-1} \text{ mm d}^{-1}$	WR	$10^{-1} - 10$	Yes
S_{UR}^{\max}	mm	UR	$10^{-1} - 10^3$	Yes
β_{UR}	—	UR	$10^{-3} - 8$	Yes
D	—	FR/SR	0–1	No
k_{FR}	d^{-1}	FR	$10^{-1} - 10$	Yes
k_{SR}	d^{-1}	SR	$10^{-4} - 10^1$	Yes

Appendix C: Choice of Distance Metric in SABC Algorithm

The choice of distance metric ρ to measure the distance between the target and simulated signatures is an important algorithmic setting in the ABC inference.

In this work, the distance metric is defined as the maximum value, over the four seasons, of the mean distance between target and simulated FDCs,

$$\rho(\mathbf{y}, \tilde{\mathbf{y}}) = \max_{s=1, \dots, N_{\text{season}}} \left[\text{mean} \left(\frac{|y_{[j,s]} - \tilde{y}_{[j,s]}|}{\rho_{[j,s]}^{\text{norm}}}; j = 1, \dots, N_{\text{FDC}} \right) \right] \quad (\text{C1})$$

where j indexes the signatures within a given season s (see Section 4.3). For convenience, matrix-style indexing is used in the subscripts to refer to the signatures organized as a matrix with N_{FDC} rows and $N_{\text{season}} = 4$ columns, rather than a “long” vector with $N_{\text{FDC}} N_{\text{season}}$ elements.

The term $\rho_{[j,s]}^{\text{norm}}$ is a normalization factor, which is designed to give the same “weight” to all signatures by accounting for their range of variability,

$$\rho_{[j,s]}^{\text{norm}} = \text{sdev} |y_{[j,s]}^{(k)} - \tilde{y}_{[j,s]}^{(k)}; k = 1, \dots, N_{\text{sam}}| \quad (\text{C2})$$

where $\text{sdev } \mathbf{v}$ is the standard deviation of a set of samples $\mathbf{v} = \{v_k; k = 1, \dots, N_{\text{sam}}\}$.

Since the signatures can be stochastic, the normalization term in Equation C1 is calculated as follows: (a) sample the parameters $\boldsymbol{\theta}^{(\text{H})}$ from their prior distribution, (b) generate N_{sam} sets of simulated signatures $\mathbf{y}_{[j,s]}$ using the precipitation-streamflow model, (c) sample N_{sam} sets of target signatures $\tilde{\mathbf{y}}_{[j,s]}$ from the pre-calibrated signature model, and (d) apply Equation C2.

This normalization within the distance metric is different from the normalization used in previous studies, where $\rho_{[j,s]}^{\text{norm}}$ was simply set to the target signature $\tilde{y}_{[j,s]}$ (e.g., Fenicia et al., 2018). That earlier normalization approach is unsuitable for this study, because the (stochastic) target signatures may take the value of 0, in which case ρ becomes infinite or undefined.

The initial value of the ABC tolerance is set to 0.1 in all scenarios. This value is progressively tightened by the SABC algorithm as it converges (Section 3 and Albert et al. (2015)). Note that as the tolerance is tightened, the impact of the distance metric on the ABC inference decreases (see Albert et al. (2015) and Kavetski et al. (2018) for details).

Appendix D: Performance Metrics

This appendix defines the model performance metrics listed in Section 4.7. The metrics are applied to both the streamflow time series and the set of signatures. The general symbols \mathbf{v} and $\tilde{\mathbf{v}}$ are used to refer to simulated and observed data respectively.

Note that $\tilde{\mathbf{v}} = \{\tilde{v}_i; i = 1, \dots, N\}$ is a vector of length N (e.g., representing a streamflow time series or a set of signatures), where \tilde{v}_i is a scalar (e.g., streamflow at the i th time step or the i th signature). In contrast, $\mathbf{v} = \{v_i; i = 1, \dots, N\}$ is a matrix containing an ensemble of model predictions generated using the procedure in Appendix A, where $\mathbf{v}_i = \{v_i^k; k = 1, \dots, N_{\text{sam}}\}$ is a vector of multiple model realizations of the i th element.

D1. Nash-Sutcliffe Efficiency

The Nash-Sutcliffe efficiency, NSE, measures the fit of the simulated data to the observations,

$$\text{NSE}(\mathbf{v}, \tilde{\mathbf{v}}) = 1 - \frac{\sum_{i=1}^N (\tilde{v}_i - v_i^{\text{med}})^2}{\sum_{i=1}^N (\tilde{v}_i - \text{mean } \tilde{\mathbf{v}})^2} \quad (\text{D1})$$

where $v_i^{\text{med}} = \text{median } \mathbf{v}_i$ is the median of the predictive distribution of the i th element of \mathbf{v} , $\text{mean } \tilde{\mathbf{v}}$ is the mean of the observed values $\tilde{\mathbf{v}}$. The use of median instead of mean in the numerator of Equation D1 is intended to reduce the influence of noise/outliers.

The NS varies between $-\infty$ and 1, with 1 denoting a perfect match.

D2. Precision

The precision metric, P , quantifies the spread of a predictive distribution,

$$P(\mathbf{v}, \tilde{\mathbf{v}}) = \frac{\sum_{i=1}^N \text{sdev } \mathbf{v}_i}{\sum_{i=1}^N \tilde{\mathbf{v}}_i} \quad (\text{D2})$$

where $\text{sdev } \mathbf{v}_i$ denotes the standard deviation of the predictive distribution of the i th element of \mathbf{v} .

Note that the precision metric does not depend on the fit between modeled and observed data. The “best” value of this metric is 0, which occurs for a prediction with no uncertainty, that is, where all predictive replicates form a single curve. Higher values denote worse precision.

D3. Reliability

The reliability metric, R , quantifies the consistency of the observations and the predictive distribution. The metric is defined as the area between the predictive quantile-quantile curve (PQQ plots) and the 1:1 line. The PQQ plot is a uniform quantile-quantile plot of the quantities $\Pr(V_i \leq \tilde{v}_i)$ for $i = 1, \dots, N$, where V denotes the random variable underlying the predictive distribution from which the observed value \tilde{v} is assumed to be sampled from. When applied to streamflow time series analysis, V_i is the streamflow at the i th time step, whereas when applied to FDC analysis it is the FDC at the i th quantile. The mathematical derivation of the reliability metric can be found in McInerney et al. (2017). The best value of this metric is 0. Higher values denote worse reliability.

D4. Volumetric Bias

The volumetric bias metric measures the long-term water balance error of the predicted data,

$$\text{VB}(\mathbf{v}, \tilde{\mathbf{v}}) = \left| \frac{\sum_{i=1}^N \mathbf{v}_i^{\text{mean}} - \sum_{i=1}^N \tilde{\mathbf{v}}_i}{\sum_{i=1}^N \tilde{\mathbf{v}}_i} \right| \quad (\text{D3})$$

where $\mathbf{v}_i^{\text{mean}} = \text{mean } \mathbf{v}_i$ is the mean of the predictive distribution of the i th element of \mathbf{v} . The mean in the numerator of Equation D3 follows the convention from earlier studies (e.g., Ehlers et al., 2019; McInerney et al., 2017; Oliveira et al., 2018) and is less susceptible to noise/outliers due to the integration (summation) over a long time series.

The best value of the volumetric bias metric is 0 and higher values denote an increasing water balance discrepancy.

Data Availability Statement

The study data are deposited in Dal Molin (2022).

Acknowledgments

We thank Editor Hamid Moradkhani, Associate Editor Julie Mai, and the four anonymous reviewers, for constructive comments and feedback. Open access funding provided by ETH-Bereich Forschungsanstalten.

References

- Abdulla, F. A., & Lettenmaier, D. P. (1997). Development of regional parameter estimation equations for a macroscale hydrologic model. *Journal of Hydrology*, 197(1), 230–257. [https://doi.org/10.1016/S0022-1694\(96\)03262-3](https://doi.org/10.1016/S0022-1694(96)03262-3)
- Addor, N., Nearing, G., Prieto, C., Newman, A. J., Le Vine, N., & Clark, M. P. (2018). A ranking of hydrological signatures based on their predictability in space. *Water Resources Research*, 54(11), 8792–8812. <https://doi.org/10.1029/2018WR022606>
- Addor, N., Newman, A. J., Mizukami, N., & Clark, M. P. (2017). The CAMELS data set: Catchment attributes and meteorology for large-sample studies. *Hydrology and Earth System Sciences*, 21(10), 5293–5313. <https://doi.org/10.5194/hess-21-5293-2017>
- Albert, C., Künsch, H. R., & Scheidegger, A. (2015). A simulated annealing approach to approximate Bayes computations. *Statistics and Computing*, 25(6), 1217–1232. <https://doi.org/10.1007/s11222-014-9507-8>
- Almeida, S., Le Vine, N., McIntyre, N., Wagener, T., & Buytaert, W. (2016). Accounting for dependencies in regionalized signatures for predictions in ungauged catchments. *Hydrology and Earth System Sciences*, 20(2), 887–901. <https://doi.org/10.5194/hess-20-887-2016>
- Berger, K. P., & Entekhabi, D. (2001). Basin hydrologic response relations to distributed physiographic descriptors and climate. *Journal of Hydrology*, 247(3–4), 169–182. [https://doi.org/10.1016/S0022-1694\(01\)00383-3](https://doi.org/10.1016/S0022-1694(01)00383-3)
- Betterle, A., Radny, D., Schirmer, M., & Botter, G. (2017). What do they have in common? Drivers of streamflow spatial correlation and prediction of flow regimes in ungauged locations. *Water Resources Research*, 53(12), 10354–10373. <https://doi.org/10.1002/2017wr021144>
- Biondi, D., & De Luca, D. L. (2016). Rainfall-runoff model parameter conditioning on regional hydrological signatures: Application to ungauged basins in southern Italy. *Hydrology Research*, 48(3), 714–725. <https://doi.org/10.2166/nh.2016.097>

- Booker, D. J., & Woods, R. A. (2014). Comparing and combining physically-based and empirically-based approaches for estimating the hydrology of ungauged catchments. *Journal of Hydrology*, 508, 227–239. <https://doi.org/10.1016/j.jhydrol.2013.11.007>
- Botter, G., Porporato, A., Rodríguez-Iturbe, I., & Rinaldo, A. (2009). Nonlinear storage-discharge relations and catchment streamflow regimes. *Water Resources Research*, 45(10), W10427. <https://doi.org/10.1029/2008WR007658>
- Box, G. E. P., & Cox, D. R. (1964). An analysis of transformations. *Journal of the Royal Statistical Society B*, 26(2), 211–252. <https://doi.org/10.1111/j.2517-6161.1964.tb00553.x>
- Boyle, D. P. (2003). In *Multicriteria calibration of hydrologic models*. The University of Arizona.
- Bulygina, N., Ballard, C., McIntyre, N., O'Donnell, G., & Wheeler, H. (2012). Integrating different types of information into hydrological model parameter estimation: Application to ungauged catchments and land use scenario analysis. *Water Resources Research*, 48(6), W06519. <https://doi.org/10.1029/2011wr011207>
- Bulygina, N., McIntyre, N., & Wheeler, H. (2009). Conditioning rainfall-runoff model parameters for ungauged catchments and land management impacts analysis. *Hydrology and Earth System Sciences*, 13(6), 893–904. <https://doi.org/10.5194/hess-13-893-2009>
- Bulygina, N., McIntyre, N., & Wheeler, H. (2011). Bayesian conditioning of a rainfall-runoff model for predicting flows in ungauged catchments and under land use changes. *Water Resources Research*, 47(2), W02503. <https://doi.org/10.1029/2010wr009240>
- Castellarin, A., Botter, G., Hughes, D. A., Liu, S., Ouarda, T. B. M. J., Parajka, J., et al. (2013). Prediction of flow duration curves in ungauged basins. In A. Viglione, G. Blöschl, H. Savenije, M. Sivapalan, & T. Wagener (Eds.), *Runoff prediction in ungauged basins: Synthesis across processes, places and scales* (pp. 135–162). Cambridge University Press.
- Castiglioni, S., Castellarin, A., & Montanari, A. (2009). Prediction of low-flow indices in ungauged basins through physiographical space-based interpolation. *Journal of Hydrology*, 378(3), 272–280. <https://doi.org/10.1016/j.jhydrol.2009.09.032>
- Castiglioni, S., Lombardi, L., Toth, E., Castellarin, A., & Montanari, A. (2010). Calibration of rainfall-runoff models in ungauged basins: A regional maximum likelihood approach. *Advances in Water Resources*, 33(10), 1235–1242. <https://doi.org/10.1016/j.advwatres.2010.04.009>
- Cornes, R. C., van der Schrier, G., van den Besselaar, E. J. M., & Jones, P. D. (2018). An ensemble version of the E-OBS temperature and precipitation data sets. *Journal of Geophysical Research: Atmospheres*, 123(17), 9391–9409. <https://doi.org/10.1029/2017JD028200>
- Dal Molin, M. (2022). Data for “exploring signature-based calibration of hydrological models for streamflow prediction in ungauged basins,” (version 1.0) [Dataset]. Zenodo. <https://doi.org/10.5281/zenodo.7264098>
- Dal Molin, M., Kavetski, D., & Fenicia, F. (2020). SuperflexPy 1.2.0: An open source Python framework for building, testing and improving conceptual hydrological models. *Geoscientific Model Development Discussions*, 2020, 1–39. <https://doi.org/10.5194/gmd-2020-409>
- Dal Molin, M., Schirmer, M., Zappa, M., & Fenicia, F. (2020). Understanding dominant controls on streamflow spatial variability to set up a semi-distributed hydrological model: The case study of the Thur catchment. *Hydrology and Earth System Sciences*, 24(3), 1319–1345. <https://doi.org/10.5194/hess-24-1319-2020>
- Dornes, P. F., Tolson, B. A., Davison, B., Pietroniro, A., Pomeroy, J. W., & Marsh, P. (2008). Regionalisation of land surface hydrological model parameters in subarctic and arctic environments. *Physics and Chemistry of the Earth, Parts A/B/C*, 33(17), 1081–1089. <https://doi.org/10.1016/j.pce.2008.07.007>
- Doulatyari, B., Betterle, A., Basso, S., Biswal, B., Schirmer, M., & Botter, G. (2015). Predicting streamflow distributions and flow duration curves from landscape and climate. *Advances in Water Resources*, 83, 285–298. <https://doi.org/10.1016/j.advwatres.2015.06.013>
- Doulatyari, B., Betterle, A., Radny, D., Celegon, E. A., Fanton, P., Schirmer, M., & Botter, G. (2017). Patterns of streamflow regimes along the river network: The case of the Thur river. *Environmental Modelling & Software*, 93, 42–58. <https://doi.org/10.1016/j.envsoft.2017.03.002>
- Ehlers, L. B., Wani, O., Koch, J., Sonnenborg, T. O., & Refsgaard, J. C. (2019). Using a simple post-processor to predict residual uncertainty for multiple hydrological model outputs. *Advances in Water Resources*, 129, 16–30. <https://doi.org/10.1016/j.advwatres.2019.05.003>
- Evin, G., Thyer, M., Kavetski, D., McInerney, D., & Kuczera, G. (2014). Comparison of joint versus postprocessor approaches for hydrological uncertainty estimation accounting for error autocorrelation and heteroscedasticity. *Water Resources Research*, 50(3), 2350–2375. <https://doi.org/10.1002/2013WR014185>
- Fang, X., Pomeroy, J. W., Westbrook, C. J., Guo, X., Minke, A. G., & Brown, T. (2010). Prediction of snowmelt derived streamflow in a wetland dominated prairie basin. *Hydrology and Earth System Sciences*, 14(6), 991–1006. <https://doi.org/10.5194/hess-14-991-2010>
- Fenicia, F., Kavetski, D., Reichert, P., & Albert, C. (2018). Signature-domain calibration of hydrological models using approximate Bayesian computation: Empirical analysis of fundamental properties. *Water Resources Research*, 54(6), 3958–3987. <https://doi.org/10.1002/2017wr021616>
- Fenicia, F., Kavetski, D., & Savenije, H. H. G. (2011). Elements of a flexible approach for conceptual hydrological modeling: 1. Motivation and theoretical development. *Water Resources Research*, 47(11), W11510. <https://doi.org/10.1029/2010wr010174>
- Fenicia, F., Kavetski, D., Savenije, H. H. G., & Pfister, L. (2016). From spatially variable streamflow to distributed hydrological models: Analysis of key modeling decisions. *Water Resources Research*, 52(2), 954–989. <https://doi.org/10.1002/2015WR017398>
- Gharari, S., Hrachowitz, M., Fenicia, F., Gao, H., & Savenije, H. H. G. (2014). Using expert knowledge to increase realism in environmental system models can dramatically reduce the need for calibration. *Hydrology and Earth System Sciences*, 18(12), 4839–4859. <https://doi.org/10.5194/hess-18-4839-2014>
- He, Y., Bárdossy, A., & Zehe, E. (2011). A review of regionalisation for continuous streamflow simulation. *Hydrology and Earth System Sciences*, 15(11), 3539–3553. <https://doi.org/10.5194/hess-15-3539-2011>
- Hingray, B., Schaeffli, B., Mezghani, A., & Hamdi, Y. (2010). Signature-based model calibration for hydrological prediction in mesoscale Alpine catchments. *Hydrological Sciences Journal*, 55(6), 1002–1016. <https://doi.org/10.1080/02626667.2010.505572>
- Hrachowitz, M., Savenije, H., Blöschl, G., McDonnell, J., Sivapalan, M., Pomeroy, J., et al. (2013). A decade of Predictions in Ungauged Basins (PUB)—A review. *Hydrological Sciences Journal*, 58(6), 1198–1255. <https://doi.org/10.1080/02626667.2013.803183>
- Jiang, S., Zheng, Y., & Solomatine, D. (2020). Improving AI system awareness of geoscience knowledge: Symbiotic integration of physical approaches and deep learning. *Geophysical Research Letters*, 47(13), e2020GL088229. <https://doi.org/10.1029/2020GL088229>
- Kavetski, D., Fenicia, F., Reichert, P., & Albert, C. (2018). Signature-domain calibration of hydrological models using approximate Bayesian computation: Theory and comparison to existing applications. *Water Resources Research*, 54(6), 4059–4083. <https://doi.org/10.1002/2017wr020528>
- Kavetski, D., & Kuczera, G. (2007). Model smoothing strategies to remove microscale discontinuities and spurious secondary optima in objective functions in hydrological calibration. *Water Resources Research*, 43(3), W03411. <https://doi.org/10.1029/2006wr005195>
- Kim, D., Jung, I. W., & Chun, J. A. (2017). A comparative assessment of rainfall-runoff modelling against regional flow duration curves for ungauged catchments. *Hydrology and Earth System Sciences*, 21(11), 5647–5661. <https://doi.org/10.5194/hess-21-5647-2017>
- Kratzert, F., Klotz, D., Herrnegger, M., Sampson, A. K., Hochreiter, S., & Nearing, G. S. (2019). Toward improved predictions in ungauged basins: Exploiting the power of machine learning. *Water Resources Research*, 55(12), 11344–11354. <https://doi.org/10.1029/2019WR026065>
- Li, Y., Grimaldi, S., Pauwels, V. R. N., & Walker, J. P. (2018). Hydrologic model calibration using remotely sensed soil moisture and discharge measurements: The impact on predictions at gauged and ungauged locations. *Journal of Hydrology*, 557, 897–909. <https://doi.org/10.1016/j.jhydrol.2018.01.013>

- Lombardi, L., Toth, E., Castellarin, A., Montanari, A., & Brath, A. (2012). Calibration of a rainfall–runoff model at regional scale by optimising river discharge statistics: Performance analysis for the average/low flow regime. *Physics and Chemistry of the Earth, Parts A/B/C*, 42–44, 77–84. <https://doi.org/10.1016/j.pce.2011.05.013>
- Mazzoleni, M., Verlaan, M., Alfonso, L., Monego, M., Norbiato, D., Ferri, M., & Solomatine, D. P. (2017). Can assimilation of crowdsourced data in hydrological modelling improve flood prediction? *Hydrology and Earth System Sciences*, 21(2), 839–861. <https://doi.org/10.5194/hess-21-839-2017>
- McInerney, D., Thyer, M., Kavetski, D., Lerat, J., & Kuczera, G. (2017). Improving probabilistic prediction of daily streamflow by identifying Pareto optimal approaches for modeling heteroscedastic residual errors. *Water Resources Research*, 53(3), 2199–2239. <https://doi.org/10.1002/2016wr019168>
- Merz, R., & Blöschl, G. (2004). Regionalisation of catchment model parameters. *Journal of Hydrology*, 287(1), 95–123. <https://doi.org/10.1016/j.jhydrol.2003.09.028>
- Montanari, A., & Toth, E. (2007). Calibration of hydrological models in the spectral domain: An opportunity for scarcely gauged basins? *Water Resources Research*, 43(5), W05434. <https://doi.org/10.1029/2006wr005184>
- Nijzink, R. C., Almeida, S., Pechlivanidis, I. G., Capell, R., Gustafssons, D., Arheimer, B., et al. (2018). Constraining conceptual hydrological models with multiple information sources. *Water Resources Research*, 54(10), 8332–8362. <https://doi.org/10.1029/2017WR021895>
- Nott, D. J., Fan, Y., Marshall, L., & Sisson, S. A. (2014). Approximate Bayesian computation and Bayes' linear analysis: Toward high-dimensional ABC. *Journal of Computational and Graphical Statistics*, 23(1), 65–86. <https://doi.org/10.1080/10618600.2012.751874>
- Oliveira, D. Y., Chaffe, P. L. B., & Sá, J. H. M. (2018). Extending the applicability of the generalized likelihood function for zero-inflated data series. *Water Resources Research*, 54(3), 2494–2506. <https://doi.org/10.1002/2017WR021560>
- Parajka, J., Viglione, A., Rogger, M., Salinas, J. L., Sivapalan, M., & Blöschl, G. (2013). Comparative assessment of predictions in ungauged basins – Part I: Runoff-hydrograph studies. *Hydrology and Earth System Sciences*, 17(5), 1783–1795. <https://doi.org/10.5194/hess-17-1783-2013>
- Pool, S., Viviroli, D., & Seibert, J. (2017). Prediction of hydrographs and flow-duration curves in almost ungauged catchments: Which runoff measurements are most informative for model calibration? *Journal of Hydrology*, 554, 613–622. <https://doi.org/10.1016/j.jhydrol.2017.09.037>
- Prieto, C., Le Vine, N., Kavetski, D., García, E., & Medina, R. (2019). Flow prediction in ungauged catchments using probabilistic random forests regionalization and new statistical adequacy tests. *Water Resources Research*, 55(5), 4364–4392. <https://doi.org/10.1029/2018wr023254>
- Sadeh, M., & Vrugt, J. A. (2014). Approximate Bayesian Computation using Markov Chain Monte Carlo simulation: DREAM((ABC)). *Water Resources Research*, 50(8), 6767–6787. <https://doi.org/10.1002/2014WR015386>
- Samaniego, L., Kumar, R., & Attinger, S. (2010). Multiscale parameter regionalization of a grid-based hydrologic model at the mesoscale. *Water Resources Research*, 46(5), W05523. <https://doi.org/10.1029/2008WR007327>
- Schoups, G., & Vrugt, J. A. (2010). A formal likelihood function for parameter and predictive inference of hydrologic models with correlated, heteroscedastic, and non-Gaussian errors. *Water Resources Research*, 46(10), W10531. <https://doi.org/10.1029/2009WR008933>
- Shafii, M., & Tolson, B. A. (2015). Optimizing hydrological consistency by incorporating hydrological signatures into model calibration objectives. *Water Resources Research*, 51(5), 3796–3814. <https://doi.org/10.1002/2014WR016520>
- Singh, R., Archfield, S. A., & Wagener, T. (2014). Identifying dominant controls on hydrologic parameter transfer from gauged to ungauged catchments – A comparative hydrology approach. *Journal of Hydrology*, 517, 985–996. <https://doi.org/10.1016/j.jhydrol.2014.06.030>
- Sivapalan, M. (2006). Pattern, process and function: Elements of a unified theory of hydrology at the catchment scale. In *Encyclopedia of Hydrological Sciences*. John Wiley & Sons, Ltd.
- Sivapalan, M., Takeuchi, K., Franks, S. W., Gupta, V. K., Karambiri, H., Lakshmi, V., et al. (2003). IAHS decade on Predictions in Ungauged Basins (PUB), 2003–2012: Shaping an exciting future for the hydrological sciences. *Hydrological Sciences Journal*, 48(6), 857–880. <https://doi.org/10.1623/hysj.48.6.857.51421>
- Sorooshian, S., & Gupta, V. K. (1983). Automatic calibration of conceptual rainfall-runoff models: The question of parameter observability and uniqueness. *Water Resources Research*, 19(1), 260–268. <https://doi.org/10.1029/WR019i001p0260>
- Spence, C., Whitfield, P. H., Pomeroy, J. W., Pietroniro, A., Burn, D. H., Peters, D. L., & St-Hilaire, A. (2013). A review of the Prediction in Ungauged Basins (PUB) decade in Canada. *Canadian Water Resources Journal*, 38(4), 253–262. <https://doi.org/10.1080/07011784.2013.843867>
- Sun, W., Ishidaira, H., & Bastola, S. (2012). Calibration of hydrological models in ungauged basins based on satellite radar altimetry observations of river water level. *Hydrological Processes*, 26(23), 3524–3537. <https://doi.org/10.1002/hyp.8429>
- Vrugt, J. A., & Sadeh, M. (2013). Toward diagnostic model calibration and evaluation: Approximate Bayesian computation. *Water Resources Research*, 49(7), 4335–4345. <https://doi.org/10.1002/wrcr.20354>
- Westerberg, I. K., Guerrero, J. L., Younger, P. M., Beven, K. J., Seibert, J., Halldin, S., et al. (2011). Calibration of hydrological models using flow-duration curves. *Hydrology and Earth System Sciences*, 15(7), 2205–2227. <https://doi.org/10.5194/hess-15-2205-2011>
- Winsemius, H. C., Savenije, H. H. G., & Bastiaanssen, W. G. M. (2008). Constraining model parameters on remotely sensed evaporation: Justification for distribution in ungauged basins? *Hydrology and Earth System Sciences*, 12(6), 1403–1413. <https://doi.org/10.5194/hess-12-1403-2008>
- Winsemius, H. C., Schaeffli, B., Montanari, A., & Savenije, H. H. G. (2009). On the calibration of hydrological models in ungauged basins: A framework for integrating hard and soft hydrological information. *Water Resources Research*, 45(12), W12422. <https://doi.org/10.1029/2009wr007706>
- Xiang, Y., Chen, J., Li, L., Peng, T., & Yin, Z. (2021). Evaluation of eight global precipitation datasets in hydrological modeling. *Remote Sensing*, 13(14), 2831. <https://doi.org/10.3390/rs13142831>
- Yadav, M., Wagener, T., & Gupta, H. (2007). Regionalization of constraints on expected watershed response behavior for improved predictions in ungauged basins. *Advances in Water Resources*, 30(8), 1756–1774. <https://doi.org/10.1016/j.advwatres.2007.01.005>
- Yu, P.-S., & Yang, T.-C. (2000). Using synthetic flow duration curves for rainfall–runoff model calibration at ungauged sites. *Hydrological Processes*, 14(1), 117–133. [https://doi.org/10.1002/\(sici\)1099-1085\(200001\)14:1<117::Aid-hyp914>3.0.Co;2-q](https://doi.org/10.1002/(sici)1099-1085(200001)14:1<117::Aid-hyp914>3.0.Co;2-q)