Recursive Bayesian Estimation of Conceptual Rainfall-Runoff Model Errors in Real-Time Prediction of Streamflow

Tajiki, M.; Schoups, G.; Franssen, H.J. Hendricks; Najafinejad, A.; Bahremand, A.

DOI
10.1029/2019WR025237

Publication date
2020

Document Version
Final published version

Published in
Water Resources Research

Citation (APA)
Tajiki, M., Schoups, G., Franssen, H. J. H., Najafinejad, A., & Bahremand, A. (2020). Recursive Bayesian Estimation of Conceptual Rainfall-Runoff Model Errors in Real-Time Prediction of Streamflow. *Water Resources Research*, *56*(2), 1-25. Article e2019WR025237. https://doi.org/10.1029/2019WR025237

Water Resources Research

RESEARCH ARTICLE
10.1029/2019WR025237

Key Points:
- A novel Bayesian approach for recursive estimation of hydrologic model errors in data assimilation applications
- Application to a case study identifies the slow-flow component of the Hymod rainfall-runoff model as the main source of model errors
- Short-term probabilistic streamflow predictions improve when accounting for both model errors and uncertainty in Hymod model parameters

Supporting Information:
- Supporting Information S1

Correspondence to:
M. Tajiki,
maryam.tajiki@gmail.com

Citation:
Tajiki, M., Schoups, G., Hendricks Franssen, H. J., Najafinejad, A., & Bahremand, A. (2020). Recursive Bayesian estimation of conceptual rainfall-runoff model errors in real-time prediction of streamflow. *Water Resources Research*, *56*, e2019WR025237. https://doi.org/10.1029/2019WR025237

Received 26 MAR 2019
Accepted 26 JAN 2020
Accepted article online 29 JAN 2020

Recursive Bayesian Estimation of Conceptual Rainfall-Runoff Model Errors in Real-Time Prediction of Streamflow

M. Tajiki[1], G. Schoups[2], H. J. Hendricks Franssen[3], A. Najafinejad[1], and A. Bahremand[1]

[1]Department of Watershed Management, Gorgan University of Agricultural Science and Natural Resources, Gorgan, Iran, [2]Department of Water Management, Delft University of Technology, Delft, The Netherlands, [3]Research Centre Jülich, Institute of Bio- and Geosciences: Agrosphere, Jülich, Germany

Abstract Conceptual rainfall-runoff models account for the spatial dynamics of hydrological processes in a basin using simple spatially lumped storage-flow relations. Such rough approximations introduce model errors that are often difficult to characterize. Here, we develop and apply a methodology that recursively estimates and accounts for model errors in real-time streamflow prediction settings by adding time-dependent random noise to the internal states (storages) of the hydrological model. Magnitude of the added noise depends on a precision (inverse variance) parameter that is estimated from rainfall-runoff data. A recursive Bayesian technique is used for estimation: posteriors of hydrological parameters and states are updated through time with an ensemble Kalman filter, whereas the posterior of the precision parameter is updated recursively using a novel gamma density approximation technique. Applying this algorithm to different model error scenarios allows identification of the main source of model errors. The methodology is applied to short-term streamflow prediction with the Hymod rainfall-runoff model in a semi-cold, semi-humid basin in Iran. Results show that (i) streamflow prediction in this snow-dominated basin is more affected by model errors in the slow flow than the quick flow component of the model, (ii) accounting for model errors in the slow flow component improves both low and high flow predictions, and (iii) predictive performance further improves by accounting for Hymod parameter uncertainty in addition to model errors. Overall, accounting for model errors increased Nash-Sutcliffe efficiency (by 1–5%), reduced mean absolute error (by 2–43%), and improved probabilistic predictive performance (by 50–80%).

1. Introduction

While water is the essence of life, it may also threaten human life in conditions of excess (floods) or scarcity (droughts). Damages can potentially be mitigated using a warning system based on flow prediction. To predict streamflow, both deterministic and probabilistic modeling have been applied in the literature. These two approaches have been discussed by Beven (2011), although the application of a purely deterministic approach is not recommended (Renard et al., 2010) as deterministic models do not estimate uncertainty. Estimating uncertainties in measured data (e.g., discharge), model inputs (e.g., rainfall and parameters), and the model itself is important because of its impact on predicted streamflow through complex error propagation. To solve this problem data assimilation techniques, which can consider and quantify all sources of uncertainty (Liu et al., 2012; Liu & Gupta, 2007), have been used to provide better hydrologic predictions given their strength in combining model predictions and measured data.

Data assimilation allows not only to estimate hydrological model states (Pauwels & De Lannoy, 2009; Weerts & El Serafy, 2006) but also jointly states and parameters (Hendricks Franssen & Kinzelbach, 2008; Moradkhani et al., 2005; Wang et al., 2009) from available data such as discharge (e.g., Clark et al., 2008) and soil moisture (e.g., Lievens et al., 2016). Among data assimilation methods, Kalman Filters provide a sequential time-stepping procedure, in which a prior prediction is combined with new observations as they become available, resulting in updated model states and/or parameters. Several studies have used the ensemble Kalman filter (EnKF) for probabilistic streamflow prediction with either spatially lumped (e.g., Moradkhani et al., 2005), semi-distributed (e.g., Li et al., 2015), or distributed (e.g., Clark et al., 2008) rainfall-runoff models.

While data assimilation provides a proper technique for streamflow predictions, its success crucially depends on an accurate characterization of the various sources of uncertainty in model and data (Pathiraja et al., 2018; Weerts & El Serafy, 2006). This essentially entails two challenges: (i) how to mathematically model and parameterize the different sources of uncertainty and (ii) how to estimate the resulting error parameters (e.g. variance of random noise) from available data and prior information.

With respect to the first challenge, sources of uncertainty include errors in data (e.g., river discharge), in model inputs (parameters, initial conditions, and boundary conditions) and in the model itself, as discussed by Vrugt (2016). If not all sources of uncertainty are considered, the predictive uncertainty will be underestimated (Pathiraja et al., 2018; Salamon & Feyen, 2009). Various probabilistic models have been proposed for representing these errors, ranging from pragmatic approaches that lump all errors together into a residual error term (Bates & Campbell, 2001; Schoups & Vrugt, 2010) to attempts at modeling each error source separately (Kavetski et al., 2006; Kuczera et al., 2006; Li et al., 2014). While the second approach is conceptually appealing, it requires the use of prior knowledge to avoid issues of nonuniqueness in estimated error parameters (Renard et al., 2010). Data assimilation applications typically use the second approach, with model errors accounted for by adding random noise to the model states (e.g., Moradkhani et al., 2005), and error parameters such as noise variances fixed a priori, thereby avoiding the non-uniqueness issue. Alternative approaches for quantifying model error in conceptual rainfall-runoff models include using constant (Huard & Mailhot, 2008; Kavetski et al., 2006) and time-varying random model parameters (Anderson, 2007; Kuczera et al., 2006; Li et al., 2014; Stroud & Bengtsson, 2007), as well as spatiotemporal varying model error (Anderson, 2009; Bauser et al., 2018), time-varying random bias (Drécourt et al., 2006; Pauwels et al., 2013) or implicitly accounting for model errors using an inflated observation-error covariance (Gejadze et al., 2017). Irrespective of the error model used, it is useful to take a Bayesian approach and account for uncertainty in the model error parameters, since model errors are typically difficult to quantify a priori, that is, without comparing model simulations to actual data.

With respect to the second challenge, once a suitable error model and parameterization is formulated, its parameters (e.g., variance of random noise) need to be estimated from available data. In data assimilation this is sometimes referred to as adaptive Kalman filtering and filter/inflation tuning. For a real-world system with restricted knowledge about sources of uncertainty, tuning a filter is complicated (Anderson & Anderson, 1999) but an essential part in assimilating data (Reichle et al., 2008). Estimating model error parameters could be based on maximum likelihood estimation (Mitchell & Houtekamer, 2000), maximizing the posterior (Kuczera et al., 2006; Li et al., 2014; Renard et al., 2011), approximating the entire posterior (Bauser et al., 2018; Stroud & Bengtsson, 2007), or estimation of an inflation factor (Anderson & Anderson, 1999; Rasmussen et al., 2016). It can be done offline by batch estimation of the model error parameters (Tandeo et al., 2015) or online by sequentially updating the posterior of the model error parameters (Bauser et al., 2018; Crow & Van den Berg, 2010; Reichle et al., 2008). Additionally, it can be done manually using trial and error via multiple EnKF runs (Anderson & Anderson, 1999; Xie et al., 2014), or it can be automatically executed (Clark et al., 2008; Moradkhani et al., 2005). In contrast to automatic tuning, the manual approach is time consuming and not as transparent and reproducible.

Although a Gaussian distribution can be used to approximate the posterior of model error parameters (Miyoshi, 2011), it is acknowledged in previous researches (Anderson, 2009; Pathiraja et al., 2018) that a Gaussian distribution might not be the best choice for representing uncertainty in model error parameters that are nonnegative (e.g., noise variance) Alternative distributions which have been proposed in the atmospheric literature for nonnegative model error parameters are the gamma distribution (Stroud et al., 2018; Stroud & Bengtsson, 2007) and the Wishart distribution (Raanes et al., 2018; Ueno & Nakamura, 2016).

The goal of this paper is to develop and apply a methodology that accounts for model errors of a conceptual rainfall-runoff model in a real-time streamflow prediction setting. By accounting for model errors we aim to improve short-term predictive performance under both low-flow and high-flow conditions. In line with the two challenges outlined above, our methodology consists of two parts:

1. Identifying the main source of model error: That is, which model component is the main source of streamflow prediction error? To answer this question, we evaluate different hypotheses about where in the hydrological model the errors originate by adding noise to different model components and

comparing the resulting predictive performance. The underlying idea is that a correct accounting of model errors should translate in better probabilistic predictions.

2. Quantifying model error: Each hypothesis in (i) requires online estimation of the magnitude (variance or precision) of model errors, along with model states (and parameters); this is done with a Bayesian algorithm that recursively updates (an approximation of) the posterior of the precision and alternates with state updates by the EnKF. This can be seen as an automatic tuning of the EnKF, which otherwise would require manual tuning (multiple EnKF runs) to find a satisfactory level of model noise for each model error scenario in (i).

Application on a case study in Iran showcases both aspects of our approach: first, it illustrates how recursive model noise estimation automatically tunes the EnKF ensemble spread; second, it reveals which model component contains the main source of error for the specific basin and model used here.

The remainder of the paper is structured as follows. The study area and hydrologic model are introduced in the section 2, including the addition of random noise to the model. The applied and proposed approach for recursive Bayesian estimation is presented in section 3, with experimental setup and evaluation criteria presented in section 4. Section 5 contains results, followed by discussion and conclusions.

2. Study Area and Model Description

The proposed method was applied to the Roudak catchment with an area of 437 km^2 located in the north of Tehran province, Iran (Figure 1). The catchment is located within a semi-cold and semi-humid climate and has 80 frost days per year. It is typically snow covered between November and March, and the hydrologic regime of the catchment is affected by snow melt and accumulation. Yearly mean air temperature is 9 °C and varies throughout the year from −2 °C (January) to 22 °C (July). Average annual precipitation (rainfall and snow) estimated by the Thiessen method is 757 mm; the driest month is July (19 mm) and the wettest month is November (115 mm).

Precipitation and temperature show a strong spatial variability related to the large elevation difference in the catchment of 2578 m, with average slopes of 50%. Land use is predominantly rangeland (70%) and shrub land (20%), with the remaining 10% consisting of urban areas, planted forest and agriculture. The main river of the Roudak catchment is a 30 km perennial river with average observed discharge ranging from 2.9 (August) to 20.1 (April) m^3/s, and overall average discharge of 7.2 m^3/s during 2008–2016.

Data used in this study includes daily discharge from one hydrometric station (Roudak station), daily precipitation from six rain gauges in the catchment (Shemshak, Garmabdar, Ahar, Fasham, Amameh, and Galookan), and daily temperature from one meteorological station (Latian) located 13 km southeast of the catchment (Figure 1). All data are collected for the period 2008–2016 by the regional water organization of Tehran.

The rest of this section describes our approach for modeling rainfall-runoff processes in the Roudak catchment, which consists of deterministic and stochastic components.

2.1. Deterministic Component: Hymod Model

Hymod is applied at the daily time scale to simulate rainfall-runoff in the catchment. Due to the role of snow and the large elevation differences in the Roudak catchment, two snow reservoirs have been added to the original Hymod model to appropriately account for snowmelt. A schematic of the modified Hymod model is shown in Figure 2 (Black components). Model inputs consist of

1. *Forcing (boundary conditions)*: daily precipitation (P) and temperature (T).
2. *Initial storage levels*: storages consist of two snow storage reservoirs (SW_1, SW_2), storage in the nonlinear reservoir representing the watershed soil water storage (SMc), three quick-flow reservoirs (S_{q1},S_{q2}, and S_{q3}), and a slow-flow reservoir (S_s). All seven storages or states are contained in vector $\mathbf{S}$.
3. *Parameters*: maximum storage capacity (C_{max}), degree of spatial variability of the soil moisture capacity (β), flow distributing factor (α), and inverse residence times of the quick reservoirs (R_q), and the slow reservoir (R_s). All five parameters are contained in vector θ.

Referring to Figure 2, model calculations proceed as follows. First, precipitation is split into rainfall and snow using the degree day method considering a threshold temperature (Tt) and degree day factor to

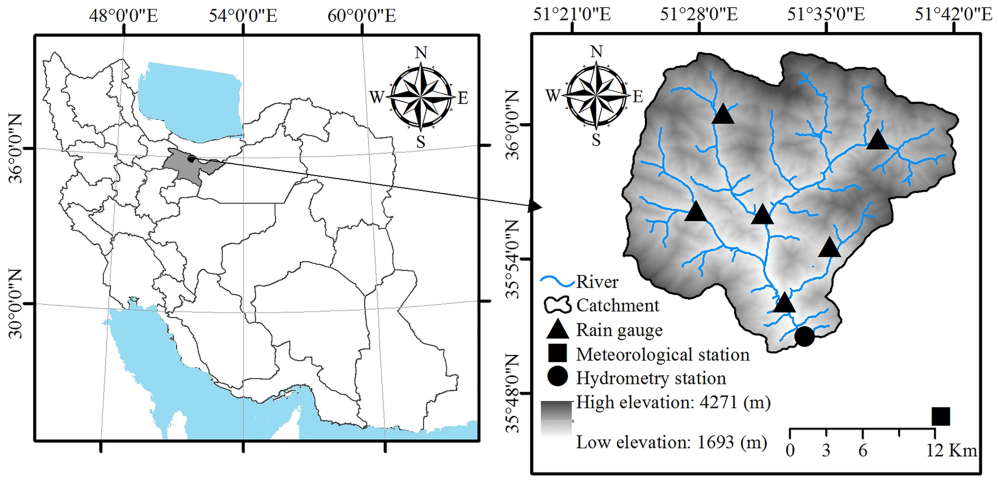

Figure 1. Location of the study area in Iran (left) and location of rain gauges, meteorological station, and hydrometric station in the Roudak catchment (right).

calculate snow accumulation and melting (Gao et al., 2014) for two different elevation zones within the catchment, that is, above and below 2,200 m elevation. Potential evapotranspiration is estimated using the Hargreaves model (Gavilán et al., 2006) for the two elevation zones. Estimated rainfall, snowmelt, and PET are then used as input for the soil water balance, which uses a nonlinear reservoir to compute actual ET and total (surface and subsurface) runoff (referred to as excess rainfall or ER). Runoff depends on soil water storage C, which is assumed to vary across the basin according to equation (1) developed by Moore (1985) to estimate the saturated fraction of the catchment:

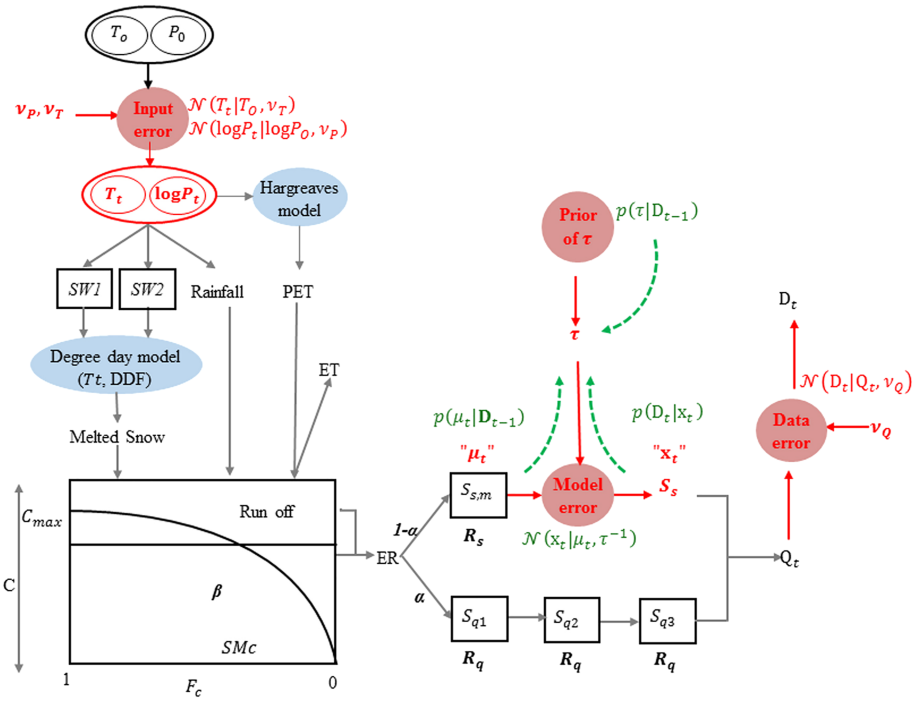

Figure 2. Schematic diagram of the deterministic (black color) and stochastic (red color) components of the modified Hymod model. Rectangles indicate reservoirs (storages or states) for which a water balance is computed. As an example, model error is added to the slow-reservoir state S_s. The different terms that make up the posterior of τ are shown in green (see section 3 and equation (10)).

$$F_c = 1 - \left(1 - \frac{C}{C_{max}}\right)^{\beta} \quad (1)$$

Runoff is split between quick and slow runoff components using parameter α, routed to the basin outlet using parameters R_q and R_s, and finally summed to get simulated river discharge.

For a single time step t, the Hymod model can be represented by a deterministic function that converts inputs into outputs:

$$\left[Q_{m,t}, \mathbf{S}_{m,t}\right] = f_{Hymod}(\mathbf{S}_{t-1}, P_t, T_t, \mathrm{PET}_t, \theta) \quad (2)$$

where $\mathbf{S}_{t-1}$ is the vector of seven states at time $t-1$, P_t,T_t, and PET_t are precipitation, temperature and potential evapotranspiration at time t respectively, and θ is the vector of five parameters.

The model in equation (2) is driven by "true," i.e., error-corrected, inputs, while model outputs in equation (2) (storages and discharge at time t) include subscript m ("m" for model) to differentiate them from their observed and true counterparts. Model and observed outputs are not equal because of various error sources, including errors in the specified model inputs (forcing or boundary conditions, parameters, initial states), measurement (e.g., due to imperfect measurement devices, and temporal or spatial incompatibility between the measurement and the corresponding model variable), and model conceptualization and parameterization (model error). These sources of uncertainty are accounted for by including various stochastic components, as described next and shown in Figure 2 (red components).

2.2. Stochastic Component 1: Input Errors (Forcing, Parameters, and Initial States)

Daily values for air temperature and log-transformed precipitation are assumed to have Gaussian random errors:

$$T_t \sim \mathcal{N}\left(T_{o,t}, \nu_T\right) \quad (3)$$

$$\log P_t \sim \mathcal{N}\left(\log P_{o,t}, \nu_P\right) \quad (4)$$

where $T_{o,t}$ is daily observed temperature and $P_{o,t}$ is daily observed precipitation, while T_t and P_t are the corresponding "true," i.e., error-corrected, values. These equations treat the "true" values as random samples from a Gaussian distribution centered on the observed values. The standard deviation of observed temperature is assumed to be 2°K, resulting in a value of the variance ν_T equal to 4°K. A relative error of 50% is assumed for observed precipitation, so that ν_P is equal to 0.25. The assumption that precipitation error is log-normally distributed (Li et al., 2015; Lü et al., 2016) and temperature is normally distributed (Lü et al., 2016) is commonly made. Uncertainty in initial model states is specified as a normal distribution with means equal to the states $\overline{\mathbf{S}}_{t^*}$ at the end of a 1 year spin-up (period 2012–2013) with variances ν_S set according to a 10% relative error on each state:

$$\mathbf{S}_{t^*} \sim \mathcal{N}\left(\overline{\mathbf{S}}_{t^*}, \nu_{\mathbf{S}}\right) \quad (5)$$

where t^* is the last time step of the spin-up period.

Finally, prior uncertainty of the Hymod parameters (C_{max},β,α,R_q,R_s) is defined as a uniform distribution for each parameter with prior lower (lb) and upper (ub) bounds shown in Table 1.

2.3. Stochastic Component 2: Observation Errors

Error in observed discharge is described by a normal distribution with standard deviation increasing as a function of discharge to account for the greater uncertainty in the rating curve at higher flows (Clark et al., 2008):

$$D_t \sim \mathcal{N}(Q_t, \nu_Q) \quad (6)$$

where D_t is observed discharge, Q_t is true discharge, and $\nu_Q = (0.1D_t)^2$, that is, a 10% observation error is assumed.

Table 1
Prior Ranges of the Hymod Parameters and Calibrated Values With SCE-UA Using Data for the Period 2008–2012

Parameter	Definition	Prior uncertainty		Calibrated parameter
		Lower band	Upper band	
C_{max}	maximum soil water storage capacity (mm)	0	1000	290
B	Degree of spatial variability of the soil water storage capacity	0	5	4.5
A	Flow distributing factor	0.01	1.0	0.2
R_q	Inverse residence time of each quick reservoir (day^{-1})	0.5	0.8	0.75
R_s	Inverse residence time of the slow reservoir (day^{-1})	0.01	0.1	0.03

2.4. Stochastic Component 3: Model Errors

Given true values for all model inputs, computed model outputs in equation (2) may still differ from the true outputs due to various model assumptions and simplifications. We assume that these errors can be accounted for by adding time-dependent random noise to the hydrological model states and/or flows in Hymod. Ideally, we would want to add noise to that part of the model that contains or is expected to contain the largest errors. As this is in principle not known a priori, the strategy here is to test and compare various scenarios, where each scenario makes different assumptions about the source of the model errors by adding noise to a different part of the model. Specifically, referring to the Hymod diagram in Figure 2, we wish to determine whether significant errors originate from (i) the slow flow component of the model, (ii) the quick flow component, or (iii) both. As such, errors are considered separately on each of the following flows and states: Q, ER, S_s, S_{q1}, S_{q2}, and S_{q3}. Adding noise on discharge Q in principle captures errors in all "upstream" components of the model but may not be the most effective as it does not consider the source of error within the model. For example, if the source of the discharge errors originates from conceptual or parameterization errors in the nonlinear reservoir of the model, then it may be better to add noise to ER (see Figure 2). This noise then propagates through both slow and quick flow components of the model, eventually affecting discharge predictions.

Formally, let x_t represent one of the variables (Q, ER, S_s, S_{q1}, S_{q2}, or S_{q3}) at time t. Then adding Gaussian model errors can be achieved via

$$x_t \tilde{} \mathcal{N}\left(\mu_t, \tau^{-1}\right) \tag{7}$$

where x_t represents the Hymod variable of interest, μ_t represents the modeled value of the selected variable (before adding noise), and τ is the precision (inverse variance) of the added noise. With the exception of adding noise to Q, this approach adds noise "inside" the model, and thus turns Hymod into a probabilistic model:

$$[Q_t, \mathbf{S}_t] \tilde{} p_{\text{Hymod}}(Q_t, \mathbf{S}_t | \mathbf{S}_{t-1}, P_t, T_t, \text{PET}_t, \tau, \Theta) \tag{8}$$

where τ is now an additional input, and outputs on the left-hand side represent modeled values including random model errors, as opposed to modeled values in equation (2). The use of symbol ~ in equation (8) makes clear that outputs are now sampled values from a probabilistic model as opposed to outputs from a deterministic model ("=" in equation (2)). In this setting, noise added to an internal state or flow propagates to all "downstream" variables in the model (see Figure 2).

An important decision relates to the value of τ. Whereas standard errors of measurements can often be estimated a priori, magnitudes of model error are difficult to pinpoint in advance, especially when it relates to errors of an unobserved state in the model. Therefore, we treat τ as a (time-invariant) random variable and account for its prior uncertainty using as prior distribution a gamma distribution with two parameters (α_0,β_0):

$$\tau \tilde{} Ga(\alpha_0, \beta_0) \tag{9}$$

where α_0 and β_0 are shape and rate parameters of the gamma distribution. Their values are chosen to reflect initial uncertainty about τ and were selected according to magnitude of the state or flow to which model

error was added (Table S1 in the supporting information). An example for the prior distribution of τ is shown in Figure S1. This prior distribution strikes a balance between accommodating the possibility of large model errors (for values of τ close to zero) as well as small model errors (for large values of τ in the long tail of the distribution). Reasons for choosing a gamma distribution are (i) it has positive support and is thus appropriate for modeling a nonnegative variable such as τ and (ii) it is computationally attractive as it leads to an efficient recursive estimation scheme, as shown in the next section.

3. Recursive Bayesian Estimation

Our goal is now to recursively estimate posteriors for the model error precision as well as for the Hymod states and parameters. Every time a new discharge observation becomes available it is assimilated by (1) updating the posterior of τ (precision of model errors), and (2) updating the states and parameters of Hymod. We first present a stepwise overview of the algorithm and then give mathematical details for updating the posteriors. The posterior update procedure can be summarized as follows:

Randomly draw N initial states and Hymod parameter sets from their prior distributions (equation (5) and Table 1).

For each day

1. Predict

Randomly draw N precipitation and temperature values from their prior distributions (equations (3) and (4)).

Randomly draw N model error values by repeating N times the following two steps: (i) sample a value for model error precision (τ) (inverse variance) from its current posterior distribution and (ii) sample an error value from $\mathcal{N}(0, \tau^{-1})$.

Run the probabilistic version of Hymod (Figure 2 and equation (8)): This propagates input ensembles through the model in Figure 2, thereby creating ensembles for all intermediate variables in this figure, including ensembles of predicted states and discharge.

2. Update posterior of model error precision

See section 3.1 for details: The posterior of τ is represented by a gamma density whose shape (α) and rate (β) parameters are updated with the current discharge observation using equations (14)–(16).

3. Update posterior of states and parameters

See section 3.1. for details: Posteriors for states and parameters are represented by ensembles that are updated with the current discharge observation using the EnKF (equation (17)).

End

Note that the posterior of model error precision τ is updated separately from the posterior of states and parameters. The alternative of adding the precision as an additional unknown to the state vector, that is, the state-augmentation approach, was not considered here, since previous studies have shown that this does not work well for variance parameters (DelSole & Yang, 2010; Stroud & Bengtsson, 2007). Stroud et al. (2018) point out that state-augmentation generally does not work well for parameters with small correlation with the states. The following two sections provide mathematical details for updating the posteriors.

3.1. Sequential Updating of the Posterior of τ

Our method sequentially updates the posterior of precision or inverse variance τ of the model errors. This is achieved by maintaining a gamma density approximation of the posterior: as new data become available the approximation is updated in real-time to reflect new information on model errors. Our method was inspired by Stroud et al. (2018) and is also closely related to Stroud and Bengtsson (2007). Stroud and Bengtsson (2007) present an exact closed-form (analytical) updating scheme for the posterior of the precision of model errors (represented by a gamma density), but their scheme assumes that the variance of model and observation errors are related by a common scaling factor. When this assumption is relaxed, posterior approximations are required. For example, Anderson (2007) and Stroud et al. (2018) used Gaussian approximations. Since our interest is in the estimation of the precision (or variance) of model errors—a positive quantity—

we propose to use gamma approximations for its posterior, as in Stroud and Bengtsson (2007). The gamma density is a common choice for modeling precision variables in Bayesian analysis, partly because of its computational properties; for example a gamma prior is conjugate to a linear-Gaussian likelihood, leading to closed-form expressions for the posterior, as used in Stroud and Bengtsson (2007). In this paper, we deal with nonlinear models, so we lose conjugacy and the posterior is not available analytically, although it can be efficiently approximated, as detailed below. This results in an approximate but efficient recursive method for estimating the posterior of τ that is fully Bayesian and extends point (posterior mode) estimation of the model noise, as in, for example, Li et al. (2014).

Mathematically, the posterior of τ at time t is written recursively in terms of the posterior at time $t-1$ (which becomes the prior for time t) and the likelihood at time t. For model noise on a selected flow or state variable x (see previous section), the formula is

$$p(\tau|\boldsymbol{D}_t) \propto p(\tau|\boldsymbol{D}_{t-1}) \int_{x_t} \int_{\mu_t} p(\mu_t|\boldsymbol{D}_{t-1}) \mathcal{N}\left(x_t|\mu_t, \tau^{-1}\right) p(D_t|x_t) d\mu_t dx_t \tag{10}$$

where D_t is the current discharge measurement and $\mathbf{D}_t = (\mathbf{D}_{t-1}, D_t)$ holds all discharge data up to and including time t. The prior of τ at time t is written as $p(\tau|\,\mathbf{D}_{t-1})$; for the first time step, this distribution is equal to $Ga(\tau|\,\alpha_0, \beta_0)$ (equation (9)). The double integral represents the likelihood of τ at time t: $p(\mu_t|\,\mathbf{D}_{t-1})$ is the predictive distribution of the selected flow or state variable before adding noise, $\mathcal{N}(x_t|\mu_t, \tau^{-1})$ is the model noise term (equation (7)), and $p(D_t|\,x_t)$ is likelihood at time t of the true flow or state (after adding noise).

Equation (10) can be understood by reference to the model error depicted in Figure 2, which shows the case of model error added to the slow-reservoir state S_s. For this case μ_t corresponds to the modeled state $S_{s,\mathrm{m}}$ ("m" indicates modeled) before adding model error, and x_t corresponds to the true state S_s after adding model error. In order to update the posterior for τ at time t we want to combine three sources of information (arrows in Figure 2): (i) the prior for τ (i.e., the posterior for τ at time $t-1$), (ii) the predictive distribution of $S_{s,m}$ ($p(\mu_t|\,\mathbf{D}_{t-1})$), and (iii) probabilistic information on S_s provided by the current discharge observation ($p(D_t|\,x_t)$). As Figure 2 shows, the latter two are combined with $\mathcal{N}(x_t|\mu_t, \tau^{-1})$ to provide a probabilistic information stream to τ; this constitutes the likelihood term for τ and it corresponds to the double integral in equation (10).

To make the double integral in equation (10) tractable, both the predictive distribution $p(\mu_t|\,\mathbf{D}_{t-1})$ and likelihood $p(D_t|\,x_t)$ are approximated by Gaussian distributions. Specifically, following Stroud et al. (2018), predictive distribution $p(\mu_t|\,\mathbf{D}_{t-1})$ is approximated by $\mathcal{N}\left(\mu_t|\mu_\mu, v_\mu\right)$, where μ_μ and v_μ are mean and variance, respectively, of an ensemble of predicted values of μ_t, which is the selected flow or state variable before adding model noise. Likewise, likelihood $p(D_t|\,x_t)$ is approximated by $\mathcal{N}(x_t|\mu_x, v_x)$, where μ_x and v_x are estimated by linearizing via linear regression the relation between predicted Q_t and x_t: $Q_t = \mu_{Q,pred} + \psi(x_t - \mu_{x,pred})$ or $x_t = \frac{Q_t - \mu_{Q,pred}}{\psi} + \mu_{x,pred}$. This last equation is then used to translate mean D_t and variance v_Q for the likelihood of Q_t into mean and variance of x_t:

$$\mu_x = \frac{D_t - \mu_{Q,pred}}{\psi} + \mu_{x,pred} \tag{11}$$

$$v_x = \frac{v_Q}{\psi^2} \tag{12}$$

where $\mu_{Q,pred}$ and $\mu_{x,pred}$ are (ensemble) means of predicted Q_t and x_t, respectively, and ψ is the linear regression coefficient between the two ensembles.

With these Gaussian approximations, we can rewrite equation (10) in the following form:

$$p(\tau|\boldsymbol{D}_t) \propto p(\tau|\boldsymbol{D}_{t-1}) \int_{x_t} \int_{\mu_t} \mathcal{N}\left(\mu_t|\mu_\mu|v_\mu\right) \mathcal{N}\left(x_t|\mu_t, \tau^{-1}\right) \mathcal{N}(x_t|\mu_x, v_x) d\mu_t dx_t \tag{13}$$

As all terms under the integral are now Gaussian, the double integral can be easily computed (see Appendix A). While this yields a closed-form formula for the posterior, the resulting density does not correspond to a known parametric density. To get an online recursive estimation algorithm, that is, one that allows easy

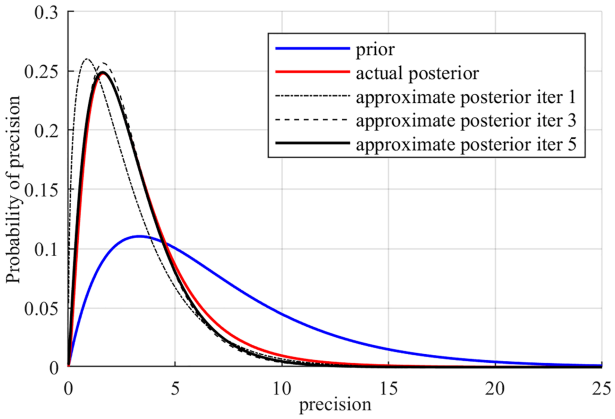

Figure 3. Approximating the actual posterior of τ (in black) by a gamma density by iterating equations (14)–(16), starting from the prior.

posterior sampling and avoids the need to reprocess data and likelihoods from previous time steps, we propose to approximate the posterior by a new gamma density (updated from the gamma prior). Without such an approximation, the posterior would grow in complexity every time step, because each time step an additional likelihood term is added that turns the posterior into a nonstandard density. Hence, even though the posterior is one-dimensional, the mathematical expression for the posterior grows in complexity. The gamma approximation prevents this and maintains a compact representation of the posterior and thus allows for recursive updating without the need to go back to earlier data, while also being easy to sample from.

Hence, at each time step, the posterior of τ is represented by a gamma distribution: $p(\tau|\mathbf{D}_{t-1}) \approx Ga(\tau|\alpha_{t-1}, \beta_{t-1})$ and $p(\tau|\mathbf{D}_t) \approx Ga(\tau|\alpha_t, \beta_t)$. This allows us to derive closed-form recursive update equations for α_t and β_t (see Appendix A):

$$\beta_t = \alpha_{t-1} - \tau^2 \frac{d^2 logf\ (\tau)}{d\,\tau^2} \tag{14}$$

$$\beta_t = \beta_{t-1} - \frac{d\ logf\ (\tau)}{d\,\tau} - (\alpha_t - \alpha_{t-1})\tau^{-1} \tag{15}$$

$$\tau = \frac{\alpha_t - 0.5}{\beta_t} \tag{16}$$

where expressions for $\log f$ and its derivatives are given in the appendix. Since α_t and β_t in equations (14) and (15) depend on τ, we update τ in equation (16) and recompute α_t and β_t again with equations (14) and (15). The update in equation (16) sets τ to a value that lies in between the mean ($\frac{\alpha_t}{\beta_t}$) and the mode ($\frac{\alpha_t - 1}{\beta_t}$) of the new gamma density approximation. At each time t, these three equations are iterated 10 times (with initial $\tau = \frac{\alpha_{t-1} - 0.5}{\beta_{t-1}}$) to yield a gamma approximation of the posterior of τ for that time step. We iterate 10 times but typically observe convergence after two to five iterations. The iteration could be made more sophisticated by monitoring convergence of the iterations and using a stopping criterion, but the computational cost of 10 iterations is small. As new data come in, these update equations result in an automatic tuning of the amount of model noise (τ) so that the resulting probabilistic prediction is neither too wide nor too narrow.

An alternative approach to the gamma approximation would be to maintain a particle approximation of the posterior and update it using sequential importance sampling, as in a particle filter. This approach is not tested here, as the parametric gamma approximation worked quite well in this case, as illustrated in Figure 3. This figure compares the actual posterior for τ at a specific time step, as given by equation (A7), together with the gamma approximation obtained by iteratively applying equations (14)–(16): The iteration starts from the prior for that time step (blue in Figure 3) and iteratively moves the posterior approximation closer to the actual posterior (red). Typically, only a few iterations are required to get a good posterior approximation (Figure 3). Note that the approximation is a good but not perfect fit, since the actual posterior is not exactly gamma.

3.2. Sequential Updating of the Hymod State and Parameter Posteriors

The states and parameter posteriors are represented by ensembles, which are generated in the prediction step outlined above. The EnKF is used to update each ensemble member i with the current discharge observation:

$$(\mathbf{S}_t, \theta|\mathbf{D}_t)_i = (\mathbf{S}_t, \theta|\mathbf{D}_{t-1})_i + \mathbf{K}_t\left(\widetilde{D}_t - Q_t\right)_i \qquad i = 1, 2, ..., N \tag{17}$$

where $\widetilde{D}_t$ represents randomly perturbed observed discharge (with perturbation variance v_Q). On the right-hand side of equation (17), predicted values for states $\mathbf{S}_t$ and discharge Q_t are generated by the sampling

Table 2
Overview of Data Assimilation Experiments

No.	Name of the experiment	Sources of uncertainty	
		Parameters	Model error
1	StrNoNoise	No	No
2	Str$Q$	No	Q
3	StrS_{q3}	No	S_{q3}
4	StrS_{q2}	No	S_{q2}
5	StrS_{q1}	No	S_{q1}
6	StrS_s	No	S_s
7	StrER	No	ER
8	Str$S_{q1}S_s$	No	S_{q1} & S_s
9	Par- StrNoNoise	Yes*	No
10	Par-Str$Q$	Yes*	Q
11	Par-StrS_{q3}	Yes*	S_{q3}
12	Par-StrS_{q2}	Yes*	S_{q2}
13	Par-StrS_{q1}	Yes*	S_{q1}
14	Par-StrS_s	Yes*	S_s
15	Par-StrER	Yes*	ER
16	Par-Str$S_{q1}S_s$	Yes*	S_{q1} & S_s

Note. All experiments include uncertainty in initial states and boundary conditions (precipitation and temperature).
[a]Yes*: four parameters out of five parameters including β, α, R_q, and R_s.

process described in the predict step at the start of section 3. Finally, $\mathbf{K}_t$ in equation (17) is the Kalman gain vector and is written as

$$\mathbf{K}_t = \mathbf{V}_{(\mathbf{S}_t,\theta),Q_t}\left(v_{Q,pred} + v_Q\right)^{-1} \tag{18}$$

where $\mathbf{V}_{(\mathbf{S}_t,\theta),Q_t}$ is a vector of ensemble cross-covariances between states-parameters ($\mathbf{S}_t$, θ) and predicted discharge Q_t and $v_{Q,pred}$ is the ensemble variance of predicted discharge Q_t.

The EnKF update in equation (17) may result in values for states and parameters outside their physical range. For this reason, the following constraints are applied to states and parameters after the EnKF update:

$$[\boldsymbol{S}_t] > 0 \tag{19}$$

$$lb < [\theta] < ub \tag{20}$$

$$SMc < C_{max}/(\beta + 1) \tag{21}$$

Similar constraints were also used in previous work (Wang et al., 2009).

4. Experimental Setup and Evaluation Metrics

The recursive estimation methodology in the previous section is applied to short-term (1-, 2-, or 3-day ahead) streamflow prediction in the Roudak catchment during the period September 2012 to August 2016. Since it is not known a priori which model components contain significant model errors, a series of data assimilation experiments is set up, where each experiment considers different sources of uncertainty, as summarized in Table 2. Each experiment considers uncertainty in initial conditions (initial model storages) and model forcings (precipitation and temperature). Experiments 1–9 use fixed values for the Hymod parameters for the period of 2012–2016, obtained via batch calibration with the Shuffled Complex Evolution (SCE-UA) global optimization algorithm (Duan et al., 1992) for the period 2008–2012. The calibrated parameters obtained with SCE-UA are presented in Table 1. Since parameters are fixed in these experiments, only states are updated in equation (17). Experiments 1–8 differ in terms of how model errors are accounted for, including an experiment without model error (Experiment 1). The last column in Table 2 lists the state or flow variable (Q, S_s, S_{q1}, S_{q2}, S_{q3}, or ER) for which model error is included with equation (7). In each case, precision τ (inverse variance) of the model error is estimated recursively using the method described in section 3.1. In Table 2, each experiment is named according to the selected state or flow variable that has model error. Note that the estimated values for τ possibly interact with other errors. For example, even though rainfall errors are accounted for, a misspecification in the rainfall error model can affect the estimated values of τ. In general, as demonstrated by Renard et al. (2010), completely separating model input and structural errors is very challenging.

In contrast to experiments 1–8, Experiments 9–16 include parameter uncertainty and both states and parameters are updated with the EnKF via the joint state-parameter approach according equation (17). However, due to strong interaction between C_{max} and β (Moradkhani et al., 2005) in Hymod, C_{max} is not estimated but is set equal to the value estimated by calibration with SCE-UA. Experiments 9–16 follow the same naming convention as Experiments 1–8 to distinguish each model error scenario, except that now "Par" is prepended to the name to indicate that parameter uncertainty is included. For example, experiment Par-StrS_{q1} accounts for parameter uncertainty and considers model error on internal Hymod storage S_{q1}.

Note that in Experiments 8 and 16, model errors are added on two internal states, namely, S_{q1} and S_s. The model in those experiments thus includes two τ values ($\tau_{S_{q1}}$ and τ_{S_s}), with the posterior of $\tau_{S_{q1}}$ updated when mean predicted flow is more than 4 m^3/s (the median flow), and the posterior of τ_{S_s} updated when mean predicted flow is less than 4 m^3/s. This model thus assumes that high and low flows have different error levels, with the median flow chosen as a reasonable cut-off value. This results in uncorrelated errors added to the slow and fast flow storages. Contrast this with Experiments 7 and 15, where noise added on ER affects both slow and fast flow and thus results in perfectly correlated errors for the slow and fast flow storages. As such,

the experiments considered cover the two corner cases (no correlation and perfect correlation). Experiments where (partially) correlated random errors are added to two or more model components are not considered here; this extension is discussed in section 3.2.

All experiments assume that the Hymod parameters are static. The experiments that include parameter uncertainty sequentially update the parameter posterior with each new discharge value. As a result, the parameter posterior may change through time, because data are processed sequentially instead of in one single batch. If the assumption of static parameters is correct, then after processing a sufficient number of data points the parameter posterior should stabilize.

Finally, note that none of the experiments adds noise to the soil moisture (*SMc*) and snow (*SW1*, *SW2*) storages. We use Experiments 7 and 15 to account for errors in the soil moisture and runoff generation calculations; these experiments add noise to variable ER (Figure 2), which is the main output from the soil moisture calculations. Although not included in the paper, experiments that added noise to the snow storages were also done, but these led to poor and inconsistent results. The probable reason for this is the poor correlation between the snow storage levels and river discharge. This poor correlation leads to large posterior uncertainty on the snow storages (see Figure 11, to be discussed later) and is due to snow melt in the model being dependent on air temperature and not on snow storage (at least, as long as there is enough snow to melt).

All presented experiments used 5,000 ensemble members; this number was determined by trial and error to achieve robust estimates of the error variance.

To compare predictive performance of the data assimilation experiments, three evaluation criteria are considered, that is, Nash-Sutcliffe Efficiency (NSE), mean absolute error (MAE), and relative logarithmic score (RLS):

$$NSE = 1 - \frac{\Sigma(m_{obs} - m_{Predict})^2}{\Sigma(m_{obs} - \overline{m_{obs}})^2} \tag{22}$$

$$MAE = \frac{1}{n}\,\Sigma(|m_{obs} - m_{Predict}|) \tag{23}$$

$$RLS = \text{logscore} - \text{perfect logscore} \tag{24a}$$

$$\text{logscore} = -\frac{1}{2}log2\pi - \frac{1}{2}log(v_{obs} + v_{Predict}) - \frac{(m_{obs} - m_{Predict})^2}{2(v_{obs} + v_{predict})} \tag{24b}$$

where n is the number of observations, m and v are mean and variance, respectively, of either the predicted ("predict") or the observed ("obs") discharge.

Equation (24b)b is a modification of the logarithmic score of Good (1952) suitable for evaluating probabilistic predictions against noisy observations. It corresponds to the error-convolved logarithmic score of Ferro (2017, equation (5)) and was recently also used by Pathiraja et al. (2018).

Larger values for NSE and RLS, and smaller values for MAE, indicate better performance, with a maximum of one for NSE, a maximum of zero for RLS, and a minimum of zero for MAE. The logarithmic score in equation (24b)b calculates log-density value of a Gaussian observation given a (approximate) Gaussian predictive density for new observations. The RLS measures how close a probabilistic prediction is to the "perfect" prediction, that is, the prediction with maximum value for the logarithmic score. The latter corresponds to a deterministic prediction that coincides with the observed value and is obtained by setting $m_{predict} = m_{obs}$ and $v_{Predict} = 0$ in equation (24b)b. The logscore in equation (24b) is a special case of the logarithm of the Bayesian model evidence (or marginal likelihood) $p(x)p(D|x)dx$, where $p(x)$ is predictive (prior) distribution for discharge x and $p(D|x)$ is likelihood of x given an observed discharge D. Indeed, if $p(x) = N(x|m_{predict}, v_{predict})$ and $p(D|x) = N(m_{obs}|x, v_{obs})$, then logarithm of the model evidence reduces to equation (24b)b. As such, the logscore in equation (24b) inherits properties of the model evidence, specifically it penalizes overfitted models (Volpi et al., 2017). For example, an overfitted, overly complex

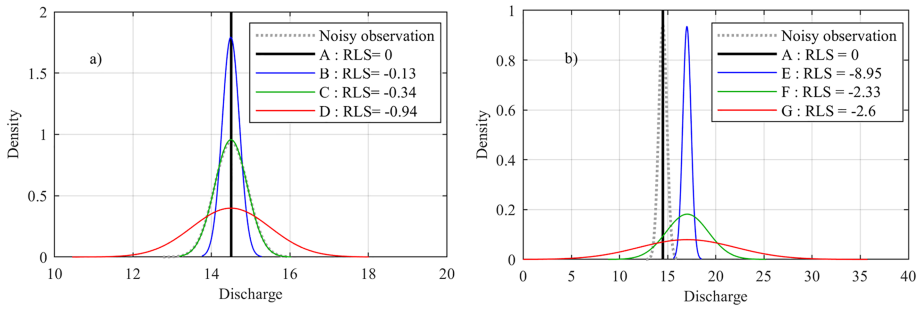

Figure 4. The RLS (equation (24a)) for various unbiased (a) and biased (b) probabilistic predictions and a noisy observation. Since observations are assumed unbiased, predictions that do not align with the observations are biased.

model will have a large variance $v_{Predict}$ that decreases the value of the logscore. In general, the model evidence (and logscore) prefers models with small $v_{Predict}$ as long as $m_{predict}$ is close enough to m_{obs}; the limiting case is a "perfect" model with a deterministic prediction that coincides with the observed value. This behavior is further illustrated in Figure 4, which compares seven different probabilistic predictions (A–G) with a noisy observation. For unbiased predictions (Figure 4a), the RLS increases as variance of the prediction decreases and the largest (perfect) score is obtained for a deterministic prediction (A). For biased predictions (Figure 4b), the RLS rewards predictive distributions that overlap with the observation distribution. Low scores result when the predictive variance is either too small (E) or too large (G). The lack of overlap is measured by the third term of equation (24b)b. Note that NSE and MAE do not take predictive and observation variance into account. The overall ranking of predictions in Figure 4 based on the RLS is A > B > C > D > F > G > E.

For reference, perfect log-scores (equation (24b)b) for the entire study period are shown in Figure S2 in the supporting information. The average value for the entire time series (the perfect log-score) is −0.16, with smaller values during high flows and larger values during low flows, because of our assumption that the observation error increases with discharge. In the results section below, RLS values will be reported, which are expressed relative to the numbers in Figure S2 (based on equation (24a)).

5. Results

The results section is divided into two parts. First, the benefits of including model errors are demonstrated by comparing results for Experiment 2 (Str*Q*) with those for experiment 1 (StrNoNoise). Second, the source of model errors is investigated by comparing results from experiments where model error is added "inside" of the model, with and without considering parameter uncertainty (Experiments 2–16). Throughout, the focus is on 1-, 2-, and 3-day ahead streamflow predictions.

5.1. Model Error on Discharge (*Q*)

We begin by illustrating the presence of model errors, that is, errors not accounted for by assumed uncertainties in initial storage levels, precipitation, temperature, and Hymod parameters. Figure 5a shows that 1-day ahead streamflow predictions without including model errors (experiment StrNoNoise) cannot capture observed data properly. There are significant deviations especially during low flow periods, with narrow predictive uncertainty bands (ensemble spread too small), for example, in August–November 2014.

The simplest approach of accounting for model errors is to lump them all together as random noise added to the simulated discharge *Q* (experiment Str*Q*) with noise level quantified by precision parameter τ and estimated recursively from the data. The probabilistic streamflow predictions for this experiment are shown in Figure 5b. Here, we see much better coverage of the observations, with wider predictive uncertainty bands, especially during low flows. This becomes very obvious when plotting the RLS (equation (24a)) as a function of time (see green plots in Figure 5): higher values during low flow indicate predictive distributions that do not overlap with the observation distribution, similar to Case E in Figure 4b. The inclusion of model errors in experiment Str*Q* significantly improves (1-day ahead) predictive performance of the model, with an average RLS of −1.39 for Str*Q* versus −2.71 for StrNoNoise. The low RLS for the latter (Figure 5a) mostly resulted

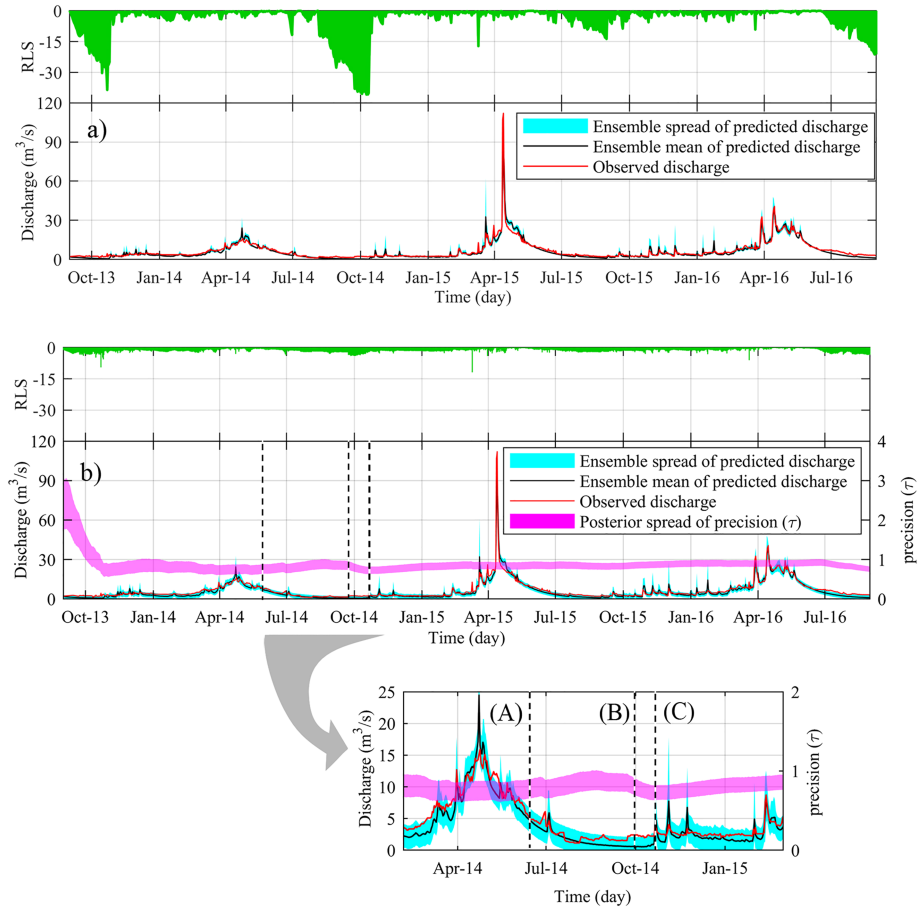

Figure 5. One-day ahead probabilistic discharge prediction with corresponding RLS values for (a) experiment StrNoNoise and (b) experiment Str*Q*.

from underprediction of low flows, for example, September to November 2013 and August to October 2014, which is improved by adding model errors (Figure 5b). Interestingly, the corresponding NSE and MAE values are quite similar for the two cases (0.91 versus 0.91 for NSE, 1.22 versus 1.31 for MAE, respectively), suggesting that adding model errors to Q increases the ensemble spread but leaves the ensemble mean largely the same.

Discharge ensemble spread in experiment Str*Q* is controlled by precision parameter τ. Figure 5 shows evolution of its estimated posterior, which is updated recursively with each new measurement. The posterior of τ fluctuates through time in response to the mismatch between predicted and observed discharge: The value of τ increases during periods of small error and decreases during periods of large error. Since precision is the inverse of variance, this behavior corresponds to automatic ensemble deflation (increase in τ) or inflation (decrease in τ) to ensure adequate overlap between predictive and observation distributions. Figure 5b illustrates this for three time steps (A–C). At Point A, predicted discharge closely follows observed discharge, leading to an increase in precision, i.e. a reduction in the noise level. At Point B, predicted discharge diverges from observed discharge, triggering a subsequent decrease in precision to account for the larger errors. According to the EnKF update equations (equation (17)), the resulting increase in ensemble spread in principle leads to a state update that moves predicted discharge closer to the data, which then again triggers an increase in τ, as illustrated in Point C. As predicted discharge converges to the observed discharge in the third time section (C) the precision starts increasing again and keeps on increasing slightly until the end of the prediction period.

Figure 6 provides further details by plotting the posterior of τ at Points A–C shown in Figure 5. At time A (left column in Figure 6), the predictive distribution without model error (top figure) is already quite good, and

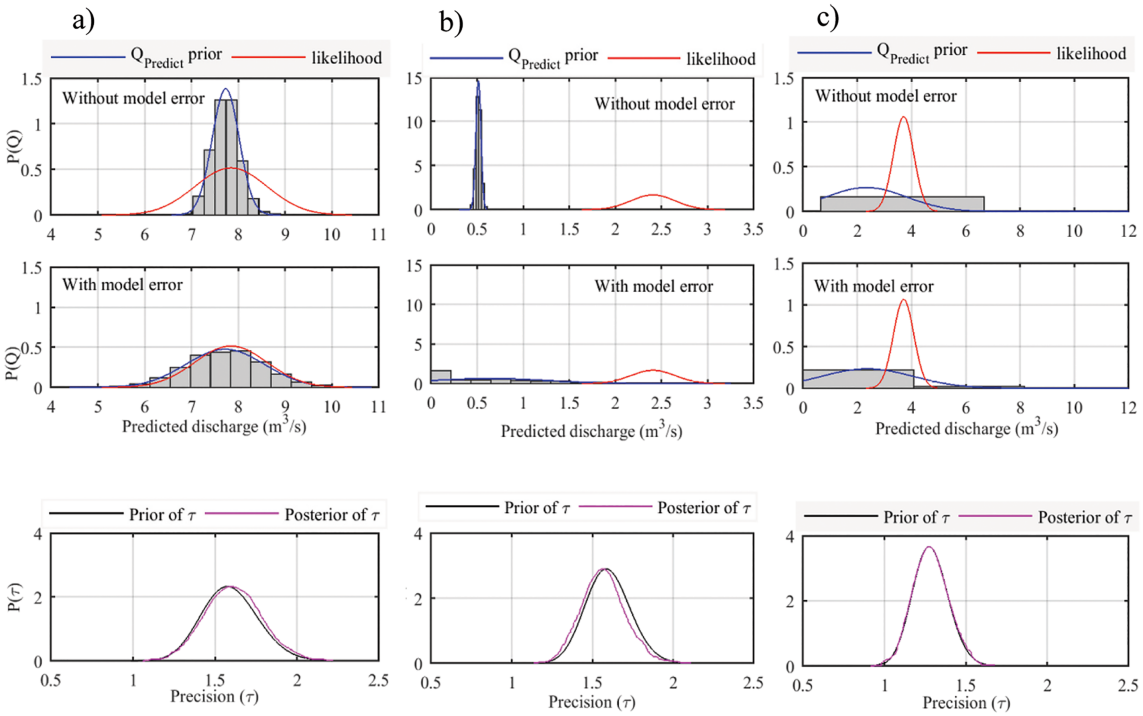

Figure 6. Details on the update of the posterior of τ at the specified time in Figure 5 including (a) A, (b) B, and (c) C.

the predictive distribution with model error (middle figure) becomes too wide when compared to the likelihood. This results in a shift of the posterior of τ to the right (bottom figure), that is, toward smaller model noise. At time B (middle column in Figure 6), the predictive distribution without model error (top figure) is biased and too narrow. Adding model errors widens the predictive distribution (middle figure), and the posterior of τ shifts to the left (bottom figure) to further increase model errors and overlap with the likelihood function. Finally, at time C (right column in Figure 6), there is adequate overlap between the likelihood and the predictive distribution and thus the posterior of τ barely changes (its posterior does not shift).

While the results in this section show that it is important to account for model error to account for misspecified uncertainty by the assumed uncertainties in initial storage levels, precipitation, and temperature, the approach of adding a constant random noise to the predicted discharge (Str$Q$) can however be improved. A first option is to use a more sophisticated model for the noise on Q (e.g., Ammann et al., 2018; Schoups & Vrugt, 2010). However, this does not provide insight into the source of the model errors. A second option, which is followed in this paper, is to improve error quantification by tracing down source of the errors, as discussed next.

5.2. Model Error on States

This section reports results for Experiments 1–16 in terms of predictive performance and estimated posteriors. The hypothesis is that we should get better predictive performance by adding model errors to that part of the model that is responsible for errors in simulated discharge. As such, we use predictive performance of the different experiments to draw conclusions about where the main sources of error are in the Hymod model for the basin studied here.

5.2.1. Predictive Performance

In this section we compare predictive performance of the model when adding model errors to (i) the quick flow storages of the model, (ii) the slow flow storage of the model, and (iii) both quick and slow flow storages. Performance is evaluated in terms of the criteria from section 4, that is, NSE, MAE, and RLS.

First, Figure 7 shows that adding noise on one of the quick flow reservoirs (S_{q1}, S_{q2}, and S_{q3}) results in better predictive performance compared to adding noise on the discharge (Q) in terms of NSE, MAE, and RLS. This

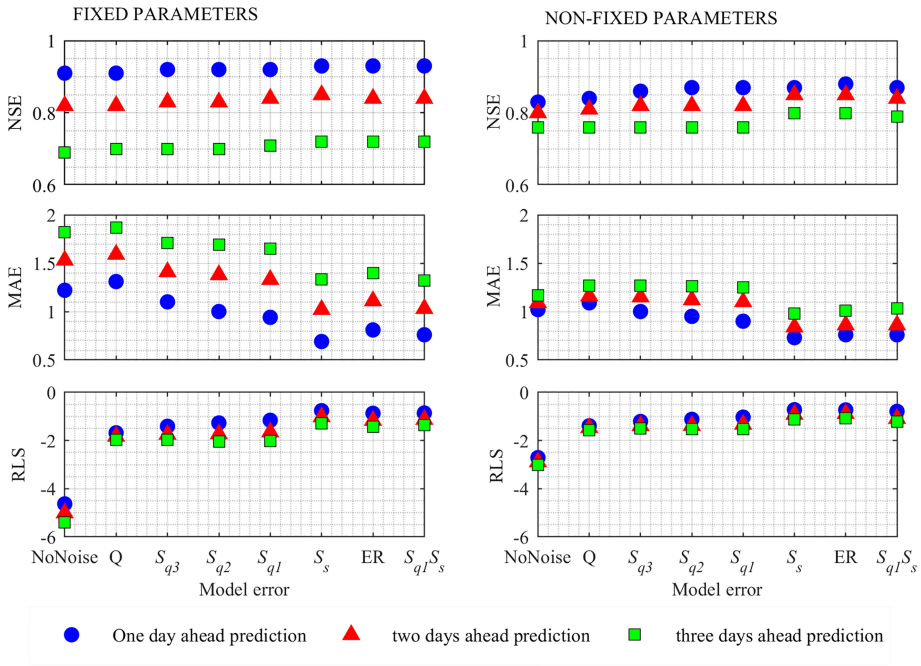

Figure 7. Predictive performance (for 1-, 2-, and 3-day ahead predictions) when adding model error to different model components, considering three criteria (NSE: first row, MAE: second row, and RLS: third row) and keeping parameters either fixed (left column) or updating them using the EnKF (right column).

improvement gets slightly better as noise is added further "upstream" in the model (from S_{q3} to S_{q2} to S_{q1}) (Figure 7).

Second, adding noise to the slow flow reservoir (S_s) gives better predictive performance than adding noise to any of the quick flow reservoirs, suggesting that model errors are more significant in the slow flow storage of the model than in the quick flow storage. Adding noise to S_s not only improves performance for low flows, but also leads to better high flow predictions (see Figure 8).

The next question is whether adding noise on both fast and slow flow storages of the model can further improve performance. First, consider the case of adding noise on the ER, which according to the model structure in Figure 2 leads to correlated errors on both fast and slow flow storages. Adding error on ER leads to good predictive performance (Figure 7), similar to adding noise on the slow flow storage (S_s) alone. As an alternative uncorrelated noise is added both on fast (S_{q1}) and slow (S_s) storages. This experiment is denoted as $S_{q1}S_s$. Figure 7 shows that this experiment does not improve on the experiment where noise is added only on S_s. Hence, the S_s experiment comes out as giving the overall best predictive performance (shown in Figure S3). Finally, note in Figure 7 that updating Hymod parameters results in better performance compared to keeping the Hymod parameters fixed. An exception is the higher NSE scores when using fixed versus updated parameters, which is likely due to the fact that the fixed parameters were obtained by prior calibration with SCE-UA, which maximizes NSE. Figure 7 further shows that fixing the parameters gives a more rapid decrease in predictive performance when the prediction horizon increases from 1 to 3 days, than the experiments where parameters were updated. For example, with noise on S_s, the NSE score decreases from ~0.9 to ~0.8 when going from 1- to 3-day ahead predictions with updated parameters, whereas the NSE score decreases from >0.9 to ~0.7 when going from 1- to 3-day ahead predictions with fixed parameters (Figure 7).

5.2.2. Estimated Posteriors

This section presents the estimated posteriors of the Hymod parameters, as well as posterior of the model noise precision. Results of adding error to the different quick flow reservoirs (S_{q1}, S_{q2}, and S_{q3}) are similar; therefore, we only present results for model error on S_{q1}.

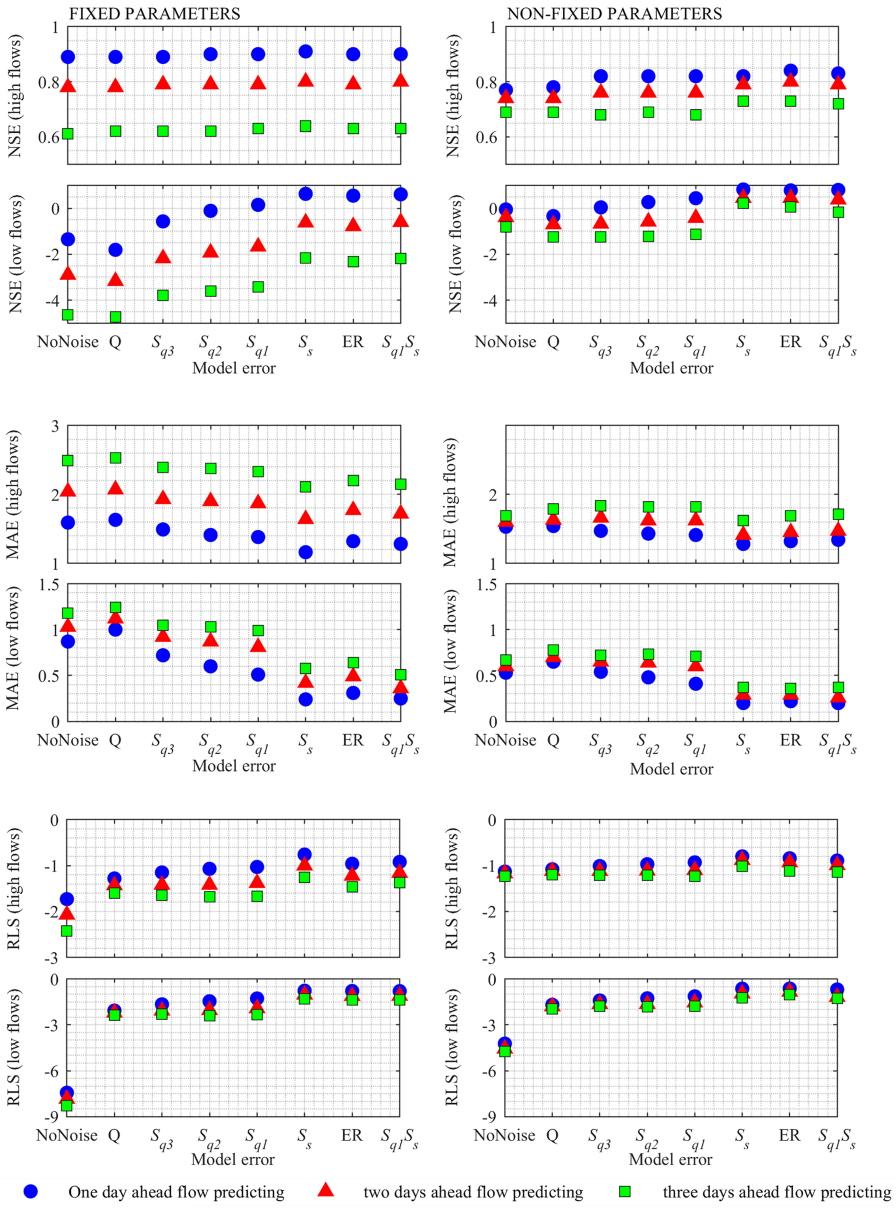

Figure 8. Predictive performances (for 1-, 2-, and 3-day ahead predictions) when adding model error to different model components, separately, for high and low flows.

We start with the posterior of precision (τ), which provides an estimate of the amount of model error that is added on each flow or state variable. The temporal evolution of the posterior of precision τ for different model error experiments is depicted in Figure 9, for both fixed and nonfixed Hymod parameters. Figure 9 shows that the precision fluctuates through time due to changes in the error of predicted discharge in each time step. This time variation of precision reduces after the first half of the study period and it becomes relatively stable in the second half of the simulation period.

Smaller precision values result in larger noise and greater potential for improvement by data assimilation. The estimated precision for the case of adding error on the slow flow reservoir is low, up to 1.5, reflecting large model errors, while estimated precision for the quick flow reservoir is high, up to 500, reflecting small model errors. This agrees with the overall better predictive performance of experiment Par-StrS_s compared to experiment Par-StrS_{q1} (Figure 7). The amount of noise is also affected by the magnitude of each state; the quick flow storages are generally quite small in magnitude, translating in small model errors (Figure 11).

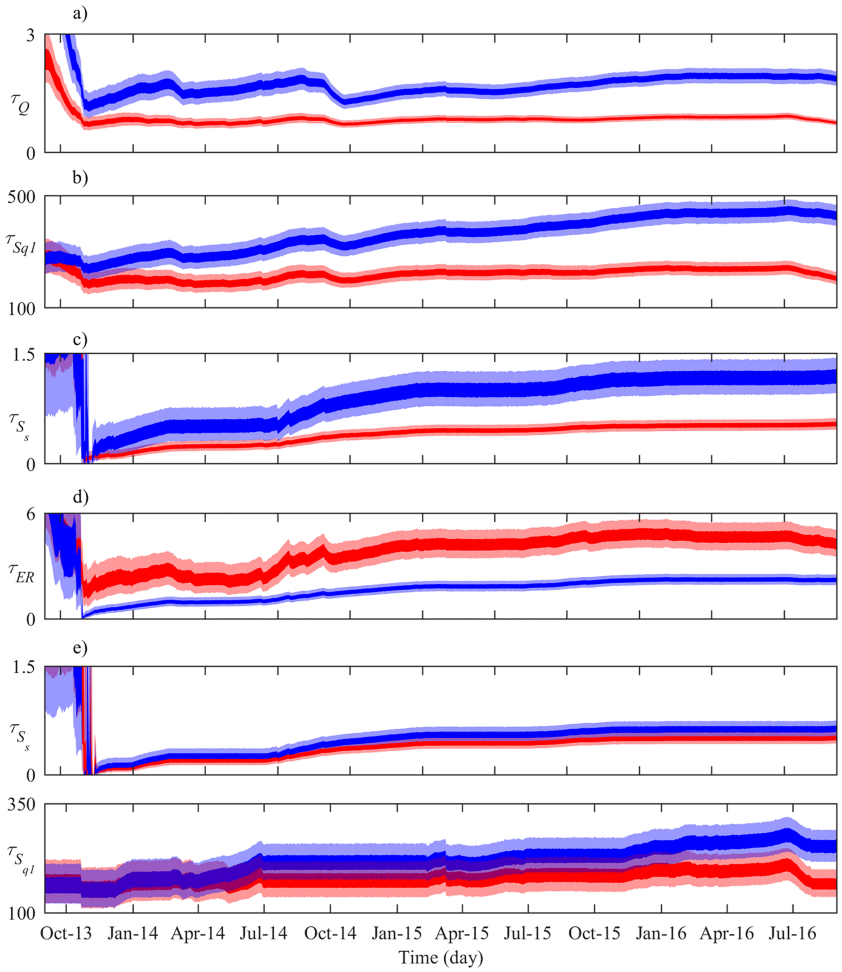

Figure 9. Temporal evolution of the posterior of precision τ with 90% and 50% uncertainty bands for (a) Str$Q$ and Par-Str$Q$, (b) StrS_{q1} and Par-StrS_{q1}, (c) StrS_s and Par-StrS_s, (d) StrER and Par-StrER, and (e) Str$S_{q1}S_s$ and Par-Str$S_{q1}S_s$; blue plots: nonfixed Hymod parameters; red plots: fixed Hymod parameters.

Comparing experiment Par-Str$S_{q1}S_s$ (last two plots in Figure 9) with experiments Par-StrS_s and Par-StrS_{q1} shows that the estimated precisions by adding error on quick and slow flow reservoirs (S_{q1} and S_s) simultaneously are smaller than adding error on just one state (S_{q1} or S_s). For example, the Par-Str$S_{q1}S_s$ experiment results in larger model errors added to S_s (smaller values for τ_{S_s}) compared to the Par-StrS_s experiment. In the former experiment, τ_{S_s} only applies to flows lower than 4 m^3/s, whereas in the latter τ_{S_s} applies to all flows. Hence, by focusing only on lower flows, we find that the model error added to S_s increases.

Results in Figure 9 further show that the posterior values of precision for experiments with updated Hymod parameters are higher than those with fixed Hymod parameters. This shows that uncertainty associated with model error decreases by considering parameter uncertainty. This expected behavior is however reversed in the experiment where model error is added on ER: Par-StrER results in lower τ values (larger model errors) compared to StrER. An explanation can be found in the value of parameter α, which is greater for StrER (0.2, see Table 1) than for Par-StrER (around 0.1, see Figure 10). Referring to Figure 2, a larger value for α translates into not only more flow going through the quick flow storages but also larger errors propagating from ER to the quick flow storages (for a given error on ER). Conversely, for a given error on the quick flow storages (and on Q), the noise on ER should be smaller when parameter α increases.

Next, temporal evolution of the Hymod parameter posteriors is illustrated in Figure 10. Parameter posteriors for experiments Par-Str$Q$ (Figure 10a) and Par-StrS_{q1} (Figure 10b) are similar. The only difference occurs for parameter R_q, which in experiment Par-StrS_{q1} is slightly higher than for Par-Str$Q$. When adding error to the

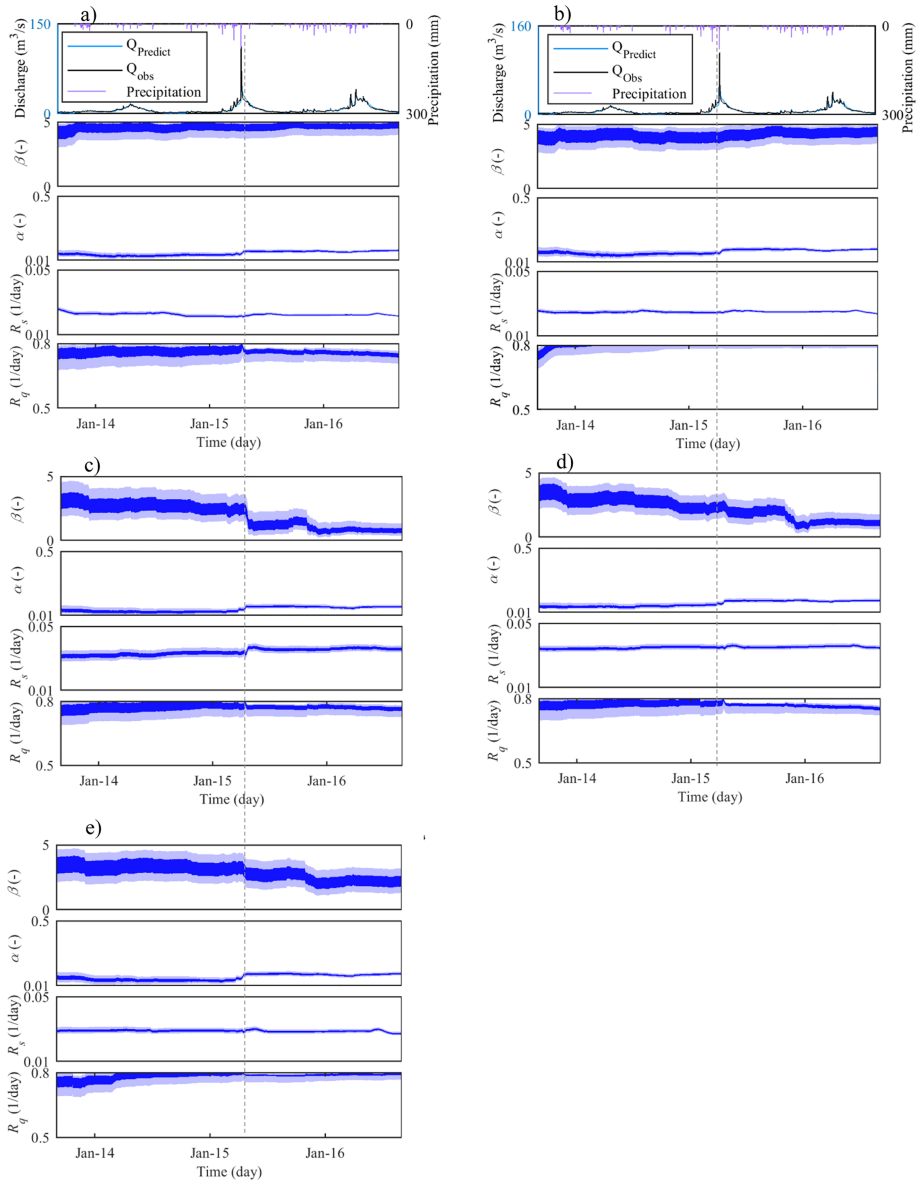

Figure 10. Temporal evolution of the Hymod parameter posteriors (β, α, R_s, and R_q) with 90% and 50% uncertainty bands for (a) Par-Str$Q$, (b) Par-StrS_{q1}, (c) Par-StrS_s, (d) Par-StrER, and (e) Par-Str$S_{q1}S_s$. Precipitation and predicted and observed discharge time series are also shown.

slow flow reservoir (Figure 10c), a noticeable jump occurs in the posterior of parameter β, as well as a minor jump in the posterior of parameter α, both coinciding with a significant peak in the observed discharge record. Similar behavior is observed when adding model error to ER (Figure 10d). The small inferred value for parameter α indicates that most of the ER is routed through the slow reservoir (see Figure 2), which explains the similar results for the ER and S_s experiments in Figure 10. Finally, comparing correlated (i.e., Par-StrER, Figure 10d) and uncorrelated, (i.e., Par-Str$S_{q1}S_s$, Figure 10e) model errors on both quick and slow flow reservoirs again results in similar parameter posteriors, with slightly larger values for parameter β in experiment Par-Str$S_{q1}S_s$.

For completeness, temporal evolution of the Hymod state posteriors is given in Figure 11. These results show that (1) posterior uncertainty is smallest for the states that are strongly related to observed discharge, that is, S_s and S_{q1}, and is largest for states further "upstream" in the model, that is, SMc, SW_1, and SW_2 (see Figure 2). The first snow reservoir (SW_1) especially, which corresponds to the low elevation zone, is

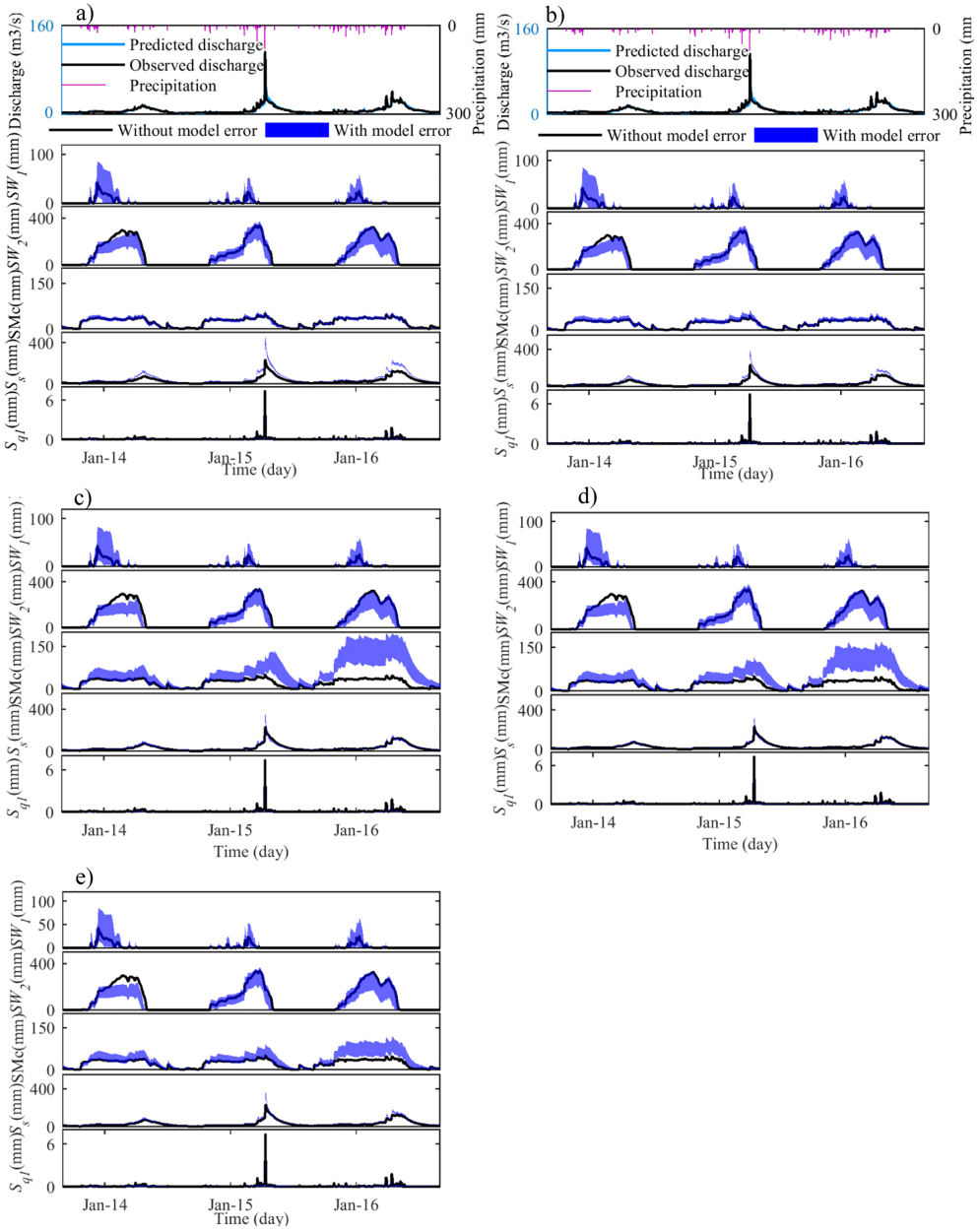

Figure 11. Temporal evolution of the Hymod state posteriors (SW_1, SW_2, SMc, S_s, and S_{q1}) with 90% uncertainty bands for different experiments: (a) Par-Str$Q$, (b) Par-StrS_{q1}, (c) Par-StrS_s, (d) Par-StrER, and (e) Par-Str$S_{q1}S_s$.

associated with large posterior uncertainty, suggesting it has weak correlation with observed river discharge; (2) the experiments that have a jump in the posterior for parameter β (see Figure 10) have a corresponding jump in the posterior for *SMc*. This is due to a strong posterior correlation between *SMc* and parameter β (given that parameter C_{max} is kept fixed), as shown in Figure S4, with smaller values for β corresponding to larger values of *SMc*.

6. Discussion

This section discusses (i) what the results tell us about model errors of Hymod for short-term streamflow prediction in the Roudak catchment and (ii) significance and possible extensions of the proposed approach for sequential estimation of model errors (i.e., parameter τ).

6.1. Hymod Model Errors

The Hymod performance for Roudak catchment improved by accounting for model errors assigned to the different states (Figure 7) to account for misspecified uncertainty in inputs. It was found that the largest improvement is obtained by accounting for model errors in the slow flow component (S_s) of Hymod, resulting in improved streamflow predictions for both low flow and high flow conditions (Figure 8). This suggests that the main errors for simulating streamflow in this basin originate in the slow-flow (groundwater) component of Hymod. Considering uncertainty in the Hymod parameters together with model errors further improved predictive performance for two and three days ahead predictions.

Comparison between the worst performing experiment (Str$Q$) and the best performing experiment (Par-StrSs) (see Figure 7), shows that RLS and MAE for Par-StrS_s are much improved compared to the Str$Q$ experiment. Also temporal variation of RLS for these two experiments can be compared by seeing Figures 5 and S3. The improvements are significant during low flow periods, although the predicted high flows are also improved. Nevertheless, there are remaining deviations between observed and predicted flows, especially for peak flows in the period April–May 2015. Note that this is also the period during which a jump can be observed in the posterior of parameter β, suggesting there are additional errors at play that are not accounted for with the current approach.

To understand the reason for temporal variation of β, the formulation of Hymod should be considered. After the peak flow in April–May 2015, parameter β starts decreasing (Figure 10), causing an increase in SMc (Figure 11) following the posterior correlation between these two variables (Figure S4). The posterior correlation between β and SMc (for fixed C_{max}) follows from the Hymod equations relating SMc and ER: The plots in Figure S5 illustrate that the same ER (and thus the same simulated discharge) results by simultaneously decreasing β and increasing SMc. This suggests that updating β is less effective than updating other Hymod parameters as its variation is compensated by variation in SMc. This assumption is confirmed by running an experiment similar to Par-StrS_s but now keeping β fixed. The two experiments, that is, Par-StrS_s with fixed and nonfixed β, have indeed similar predictive performance: NSE of 0.86 and 0.87, MAE of 0.74 and 0.73, RLS of −0.71 and −0.72 with fixed β and nonfixed β respectively. In spite of difficulties in estimating β, given its tight correlation with SMc, streamflow prediction is improved by considering uncertainty of Hymod parameters together with model errors, and the estimated Hymod parameters become nearly stable in the second half of the study period (except for β).

The predicted high flow in April–May 2015 for the experiments with error added on S_s, S_{q1}, S_{q1} S_s, and ER is illustrated in Figure S6. The predicted discharge improved from Par-Str$Q$ to Par-StrS_s, with no further improvement for experiment Par-StrS_{q1}. A potential reason for not achieving improvement in predicted high flow in experiment Par-StrS_{q1} is that the streamflow record is dominated by low flow periods, during which storage S_{q1} is close to 0, which requires very little noise. This results in posterior estimates of $\tau_{S_{q1}}$ that are quite large (i.e., up to 400 in Figure 9), corresponding to small model errors being added to S_{q1} throughout the flow record. To filter out low-flow impacts on the estimation of $\tau_{S_{q1}}$, experiments Str$S_{q1}S_s$ and Par-Str$S_{q1}S_s$ estimated two separate τ values, one for model errors on S_s during low flow conditions (i.e., $Q < 4$ m^3/s), and another for model errors on S_{q1} during high flow conditions (i.e., $Q > 4$ m^3/s). However, the Par-Str$S_{q1}S_s$ experiment did not give any significant improvement for peak streamflow prediction compared to Par-StrS_s, (Figure S6) although it does somewhat attenuate temporal changes in parameter β during April–May 2015 (see Figure 10).

Besides model errors, other errors may also contribute to underestimating the peak flow in April-May 2015. For example, measured streamflow typically is less accurate at high flows, especially when discharge exceeds values used to determine the rating curve. As such, the 10% observation error assumed here (equation (6)) could be refined by careful rating curve error analysis. Likewise, rainfall errors in this mountainous catchment are likely much more complex than assumed here (50% error, equation (4)); more refined error models could be considered to evaluate their impact on peak flow predictions.

6.2. Recursive Estimation of Model Errors

Our approach for sequentially estimating and updating model noise parameter τ via a gamma density approximation of its posterior is a new contribution to the hydrological literature. The approach is closely

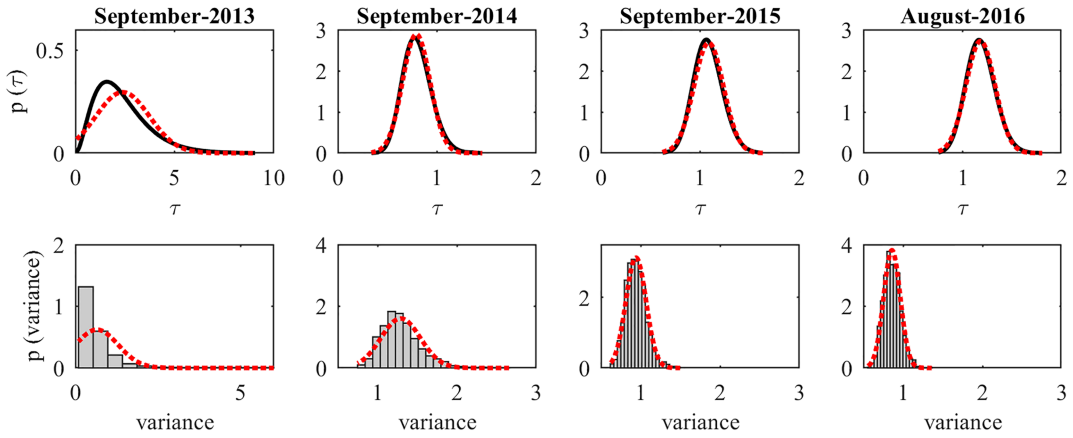

Figure 12. (top row) Gamma (black) and Gaussian (red) approximations of the posterior of noise precision τ at different times. (bottom row) Corresponding posteriors of the noise variance and their Gaussian approximations for experiment Par-StrSs.

related to Anderson (2007); the latter can be understood as maintaining a Gaussian density approximation of the posterior of the variance through time. Hence, a comparison between the two approaches comes down to the choice of a gamma approximation for the noise precision versus a Gaussian approximation for the noise variance. The computational effort for either approach is small, so that preference should be given to the highest quality approximation. A priori, noise parameter τ cannot be negative, making a gamma approximation more suitable a priori. However, operationally a Gaussian approximation may still work, as long as the posterior is not close to 0. Figure 12 illustrates how non-Gaussian the approximated posteriors for precision (τ) and the corresponding variance are in our case study. The gamma approximations in these plots were obtained with the approach proposed in this paper, whereas the Gaussian approximations have the same mean and variance as the gamma approximations. These results show that the posterior is initially non-Gaussian and asymmetric, but tends to become more Gaussian as more data are assimilated. This is a common tendency in large-data settings (Walker, 1969), although application of the EnKF also tends to make posteriors more Gaussian (Zhou et al., 2011). This suggests that a Gaussian approximation may be fine in this case, except for early assimilation times. The gamma approximation has the advantage of also applying to early times without adding any significant computational cost. For example, comparing computing time for experiments with the same number of predicted days, that is, 1, 2, or 3 days, shows Experiment 2 (with updating the precision posterior) is 8% slower than Experiment 1 (without updating the precision posterior).

When noise is added to more than one state, as was done in the ParStr$S_{q1}S_s$ experiment, the proposed approach a priori assumes independent noise on the different states. It may hence be useful to extend the approach to allow for correlated noise. This could for example be done using the scaling factor approach (Brankart et al., 2010), where the correlation structure is fixed a priori and a single scaling factor is estimated recursively. The posterior for this scaling factor could again be approximated by a gamma distribution, as long as efficient update equations can be derived. Going beyond the scaling factor approach, the entire state noise correlation or covariance matrix could be estimated recursively from the data. The advantage of this approach is that it makes minimal a priori assumptions about the noise. Extending our sequential estimation approach to this case would require a parametric approximation to the posterior covariance matrix; the Wishart distribution as used in Ueno and Nakamura (2016) could be a natural candidate for this purpose.

7. Conclusions

This paper introduced and applied a novel recursive Bayesian approach for estimating model errors of a conceptual rainfall-runoff model in a data assimilation context. The approach accounts for model errors by adding time-dependent random noise to the internal states (storages) of the hydrological model. The level of noise is controlled by a precision parameter, which is estimated recursively (online) from rainfall-runoff data using a novel gamma density approximation technique. This results in automatic tuning of the noise in the

EnKF equation for updating hydrological model states and parameters, and as such can be thought of as a form of automatic ensemble inflation/deflation that avoids the need for manual specification of noise levels in an EnKF. The recursive approach for estimating model errors could be further extended to the joint estimation of correlated noise on multiple model states.

Application of this technique to short-term rainfall-runoff prediction for a specific case study basin in Iran shows that it is effective in improving probabilistic predictions with the Hymod conceptual rainfall-runoff model. To gain insights into the source of model errors, noise was added to different model states, while also accounting for other sources of uncertainty (i.e., precipitation, temperature, initial states, and parameters), and the predictive performance was assessed for each case. The best performance was achieved by adding error on the slow flow component of the model, which improved both slow and quick flow predictions. Furthermore, accounting for uncertainty of hydrologic model parameters improved model performance, especially for three days ahead predictions. Although results showed improvement in both low and high flow predictions, not all high flows could be completely captured, suggesting there are other errors at play that are not accounted for by the methodology used here. Extreme peak flows may have a different error structure compared to other parts of the discharge record, and as such may require special treatment. Finally, we note that while the methodology is general, the conclusions about model error are specific for the basin (snow-dominated) and model (Hymod) used here. For example, in rain-dominated and humid basins one would expect models errors in the quick flow model component to be important.

Acknowledgments
Data associated with this study can be downloaded from GitHub (https://github.com/MaryamTajiki/ModelError).

References

Ammann, L., Reichert, P., & Fenicia, F. (2018). A framework for likelihood functions of deterministic hydrological models. *Hydrology and earth system sciences discussions*, 1–39. https://doi.org/10.5194/hess-2018-406

Anderson, J. (2009). Spatially and temporally varying adaptive covariance inflation for ensemble filters. *Tellus A: Dynamic meteorology and oceanography*, *61*(1), 72–83. https://doi.org/10.1111/j.1600-0870.2008.00361.x

Anderson, J. L. (2007). An adaptive covariance inflation error correction algorithm for ensemble filters. *Tellus A: Dynamic meteorology and oceanography*, *59*(2), 210–224. https://doi.org/10.1111/j.1600-0870.2006.00216.x

Anderson, J. L., & Anderson, S. L. (1999). A Monte Carlo implementation of the nonlinear filtering problem to produce ensemble assimilations and forecasts. *Monthly weather review*, *127*, 2741–2758. https://doi.org/10.1175/1520-0493(1999)127<2741:AMCIOT>2.0.CO;2

Bates, B., & Campbell, E. (2001). A Markov chain Monte Carlo scheme for parameter estimation and inference in conceptual rainfall-runoff modeling. *Water Resources Research*, *37*(4), 937–947. https://doi.org/10.1029/2000WR900363

Bauser, H., Berg, D., Klein, O., & Roth, K. (2018). Inflation method for Ensemble Kalman Filter in soil Hydrology. *Hydrology and earth system sciences discussions*, *22*(9), 4921–4934. https://doi.org/10.5194/hess-22-4921-2018

Beven, K. (2011). *Rainfall-runoff modelling: The primer*. Hoboken, New Jersey: John Wiley.

Brankart, J. M., Cosme, E., Testut, C. E., Brasseur, P., & Verron, J. (2010). Efficient adaptive error parameterizations for square root or Ensemble Kalman Filters: Application to the control of ocean mesoscale signals. *Monthly weather review*, *138*(3), 932–950. https://doi.org/10.1175/2009MWR3085.1

Clark, M. P., Rupp, D. E., Woods, R. A., Zheng, X., Ibbitt, R. P., Slater, A. G., et al. (2008). Hydrological data assimilation with the Ensemble Kalman Filter: Use of streamflow observations to update states in a distributed hydrological model. *Advances in water resources*, *31*(10), 1309–1324. https://doi.org/10.1016/j.advwatres.2008.06.005

Crow, W., & Van den Berg, M. (2010). An improved approach for estimating observation and model error parameters in soil moisture data assimilation. *Water Resources Research*, *46*, W12519. https://doi.org/10.1029/2010WR009402

DelSole, T., & Yang, X. (2010). State and parameter estimation in stochastic dynamical models. *Physica D*, *239*, 1781–1788. https://doi.org/10.1016/j.physd.2010.06.001

Drécourt, J. P., Madsen, H., & Rosbjerg, D. (2006). Bias aware Kalman Filters: Comparison and improvements. *Advances in water resources*, *29*(5), 707–718. https://doi.org/10.1016/j.advwatres.2005.07.006

Duan, Q., Sorooshian, S., & Gupta, V. (1992). Effective and efficient global optimization for conceptual rainfall-runoff models. *Water Resources Research*, *28*(4), 1015–1031. https://doi.org/10.1029/91WR02985

Ferro, C. A. (2017). Measuring forecast performance in the presence of observation error. *Quarterly Journal of the Royal Meteorological Society*, *143*, 2665–2676. https://doi.org/10.1002/qj.3115

Gao, H., Hrachowitz, M., Fenicia, F., Gharari, S., & Savenije, H. H. G. (2014). Testing the realism of a topography driven model (FLEX-Topo) in the nested catchments of the Upper Heihe, China. *Hydrology and Earth System Sciences*, *18*, 1895–1915. https://doi.org/10.5194/hess-18-1895-2014

Gavilán, P., Lorite, I. J., Tornero, S., & Berengena, J. (2006). Regional calibration of Hargreaves equation for estimating reference ET in a semiarid environment. *Agricultural water management*, *81*(3), 257–281. https://doi.org/10.1016/j.agwat.2005.05.001

Gejadze, I., Oubanas, H., & Shutyaev, V. (2017). Implicit treatment of model error using inflated observation-error covariance. *Quarterly Journal of the Royal Meteorological Society*, *143*(707), 2496–2508. https://doi.org/10.1002/qj.3102

Good, I. J. (1952) Rational Decisions. *Journal of the Royal Statistical Society, Series B*, *14*, 107–114.

Hendricks Franssen, H. J., & Kinzelbach, W. (2008). Real-time groundwater flow modeling with the ensemble Kalman filter: Joint estimation of states and parameters and the filter inbreeding problem. *Water Resources Research*, *44*, W09408. https://doi.org/10.1029/2007WR006505

Huard, D., & Mailhot, A. (2008). Calibration of hydrological model GR2M using Bayesian uncertainty analysis. *Water Resources Research*, *44*, W02424. https://doi.org/10.1029/2007WR005949

Kavetski, D., Kuczera, G., & Franks, S. (2006). Bayesian analysis of input uncertainty in hydrological modeling: 1. Theory. *Water Resources Research*, *42*, W03407. https://doi.org/10.1029/2005WR004368

Kuczera, G., Kavetski, D., Franks, S., & Thyer, M. (2006). Towards a Bayesian total error analysis of conceptual rainfall-runoff models: Characterising model error using storm-dependent parameters. *Journal of hydrology*, *331*, 161–177. https://doi.org/10.1016/j.jhydrol.2006.05.010

Li, Y., Ryu, D., Western, A., & Wang, Q. (2015). Assimilation of stream discharge for flood forecasting: Updating a semidistributed model with an integrated data assimilation scheme. *Water resources research*, *51*, 3238–3258. https://doi.org/10.1002/2014WR016667

Li, Y., Ryu, D., Western, A. W., Wang, Q. J., Robertson, D. E., & Crow, W. T. (2014). An integrated error parameter estimation and lag-aware data assimilation scheme for real-time flood forecasting. *Journal of hydrology*, *519*, 2722–2736. https://doi.org/10.1016/j.jhydrol.2014.08.009

Lievens, H., De Lannoy, G. J. M., Al Bitar, A., Drusch, M., Dumedah, G., Franssen, H. J. H., et al. (2016). Assimilation of SMOS soil moisture and brightness temperature products into a land surface model. *Remote sensing of environment*, *180*, 292–304. https://doi.org/10.1016/j.rse.2015.10.033

Liu, Y., & Gupta, H. (2007). Uncertainty in hydrologic modeling: Toward an integrated data assimilation framework. *Water resources research*, *43*, W07401. https://doi.org/10.1029/2006WR005756

Liu, Y., Weerts, A., Clark, M., Hendricks Franssen, H. J., Kumar, S., Moradkhani, H., et al. (2012). Advancing data assimilation in operational hydrologic forecasting: Progresses, challenges, and emerging opportunities. *Hydrology and earth system sciences*, *16*(10), 3863–3887. https://doi.org/10.5194/hess-16-3863-201

Lü, H., Crow, W. T., Zhu, Y., Ouyang, F., & Su, J. (2016). Improving streamflow prediction using remotely-sensed soil moisture and snow depth. *Remote sensing*, 8. https://doi.org/10.3390/rs8060503

Mitchell, H., & Houtekamer, P. (2000). An adaptive Ensemble Kalman Filter. *Monthly weather review.*, *128*, 416–433.

Miyoshi, T. (2011). The Gaussian approach to adaptive covariance inflation and its implementation with the local Ensemble Transform Kalman Filter. *Monthly weather review*, *139*, 1519–1535. https://doi.org/10.1175/2010MWR3570.1

Moore, R. J. (1985). The probability-distributed principle and runoff production at point and basin scales. *Hydrological Sciences Journal*, *30*(2), 273–297. https://doi.org/10.1080/02626668509490989

Moradkhani, H., Sorooshian, S., Gupta, H., & Houser, P. (2005). Dual state–parameter estimation of hydrological models using Ensemble Kalman Filter. *Advances in water resources*, *28*(2), 135–147. https://doi.org/10.1016/j.advwatres.2004.09.002

Pathiraja, S., Moradkhani, H., Marshall, L., Sharma, A., & Geenens, G. (2018). Data-driven model uncertainty estimation in hydrologic data assimilation. *Water resources research*, *54*, 1252–1280. https://doi.org/10.1002/2018WR022627

Pauwels, V., & De Lannoy, G. J. (2009). Ensemble-based assimilation of discharge into rainfall-runoff models: A comparison of approaches to mapping observational information to state space. *Water Resources Research*, *45*, W08428. https://doi.org/10.1029/2008WR007590

Pauwels, V. R. N., de Lannoy, G. J. M., Hendricks Franssen, H. J., & Vereecken, H. (2013). Simultaneous estimation of model state variables and observation and forecast biases using a two-stage hybrid Kalman Filter. *Hydrology and earth system sciences*, *17*, 3499–3521. https://doi.org/10.5194/hess-17-3499-2013

Raanes, P.N., Bocquet, M., & Carrassi, A. (2018). Adaptive covariance inflation in the Ensemble Kalman Filter by Gaussian scale mixtures. arXiv preprint arXiv:1801.08474

Rasmussen, J., Madsen, H., Jensen, K., & Refsgaard, J. (2016). Data assimilation in integrated hydrological modelling in the presence of observation bias. *Hydrology and earth system sciences*, *20*, 2103–2118. https://doi.org/10.5194/hess-20-2103-2016

Reichle, R., Crow, W., & Keppenne, C. (2008). An adaptive Ensemble Kalman Filter for soil moisture data assimilation. *Water Resources Research*, *44*, W03423. https://doi.org/10.1029/2007WR006357

Renard, B., Kavetski, D., Kuczera, G., Thyer, M., & Franks, S. W. (2010). Understanding predictive uncertainty in hydrologic modeling: The challenge of identifying input and structural errors. *Water Resources Research*, *46*, W05521. https://doi.org/10.1029/2009WR008328

Renard, B., Kavetski, D., Leblois, E., Thyer, M., Kuczera, G., & Franks, S. W. (2011). Toward a reliable decomposition of predictive uncertainty in hydrological modeling: Characterizing rainfall errors using conditional simulation. *Water Resources Research*, *47*, W11516. https://doi.org/10.1029/2011WR010643

Salamon, P., & Feyen, L. (2009). Assessing parameter, precipitation, and predictive uncertainty in a distributed hydrological model using sequential data assimilation with the particle filter. *Journal of hydrology*, *376*(3-4), 428–442. https://doi.org/10.1016/j.jhydrol.2009.07.051

Schoups, G., & Vrugt, J. (2010). A formal likelihood function for parameter and predictive inference of hydrologic models with correlated, heteroscedastic, and non-Gaussian errors. *Water Resources Research*, *46*, W10531. https://doi.org/10.1029/2009WR008933

Stroud, J., & Bengtsson, T. (2007). Sequential state and variance estimation within the Ensemble Kalman Filter. *Monthly weather review*, *135*, 3194–3208. https://doi.org/10.1175/MWR3460.1

Stroud, J., Katzfuss, M., & Wikle, C. (2018). A Bayesian adaptive Ensemble Kalman Filter for sequential state and parameter estimation. *Monthly weather review*, *146*(1), 373–386. https://doi.org/10.1175/MWR-D-16-0427.1

Tandeo, P., Pulido, M., & Lott, F. (2015). Offline parameter estimation using EnKF and maximum likelihood error covariance estimates: Application to a subgrid-scale orography parameterization. *Quarterly journal of the royal meteorological society*, *141*, 383–395. https://doi.org/10.1002/qj.2357

Ueno, G., & Nakamura, N. (2016). Bayesian estimation of the observation-error covariance matrix in ensemble-based filters. *Quarterly Journal of the Royal Meteorological Society*, *142*, 2055–2080. https://doi.org/10.1002/qj.2803

Volpi, E., Schoups, G., Firmani, G., & Vrugt, J. A. (2017). Sworn testimony of the model evidence: Gaussian Mixture Importance (GAME) sampling. *Water Resources Research*, *53*(7), 6133–6158. https://doi.org/10.1002/2016WR020167

Vrugt, J. A. (2016). Markov chain Monte Carlo simulation using the DREAM software package: Theory, concepts, and MATLAB implementation. *Environmental Modelling & Software*, *75*, 273–316. https://doi.org/10.1016/j.envsoft.2015.08.013

Walker, A. M. (1969). On the asymptotic behavior of posterior distributions. *Journal of the Royal Statistical Society: Series B (Methodological)*, *31*(1), 80–88. https://doi.org/10.1111/j.2517-6161.1969.tb00767.x

Wang, D., Chen, Y., & Cai, X. (2009). State and parameter estimation of hydrologic models using the constrained Ensemble Kalman Filter. *Water Resources Research*, *45*, W11416. https://doi.org/10.1029/2008WR007401

Weerts, A. H., & El Serafy, G. Y. (2006). Particle filtering and Ensemble Kalman Filtering for state updating with hydrological conceptual rainfall-runoff models. *Water Resources Research*, *42*, W09403. https://doi.org/10.1029/2005WR004093

Xie, X., Meng, S., Liang, S., & Yao, Y. (2014). Improving streamflow predictions at ungauged locations with real-time updating: Application of an EnKF-based state-parameter estimation strategy. *Hydrology and Earth System Sciences*, *18*, 3923–3936. https://doi.org/10.5194/hess-18-3923-2014

Zhou, H., Gomez-Hernandez, J., Hendricks Franssen, H. J., & Li, L. (2011). An approach to handling non-Gaussianity of parameters and state variables in ensemble Kalman filtering. *Advances in water Resources*, *34*(7), 844–864. https://doi.org/10.1016/j.advwatres.2011.04.014

Appendix A.

This appendix derives the update equations, equations (14)–(16) in the text, for recursively updating the posterior of precision τ of the model errors. For simplicity, we rewrite equations (13) in the text by omitting the time index and the explicit dependence on data D, and by writing the gamma prior as $Ga(\tau|\alpha_0,\beta_0)$, where α_0 is the shape parameter and β_0 is the rate parameter. From equation (13), the posterior for τ can then be written as

$$p(\tau) \propto Ga(\tau|\alpha_0,\beta_0)\int_x\int_\mu \mathcal{N}\left(\mu|\mu_\mu,v_\mu\right)\mathcal{N}\left(x|\mu,\tau^{-1}\right)\,(\mathcal{N}|x\mu_x,v_x)d\mu dx \tag{A1}$$

Integrating out x yields

$$p(\tau) \propto Ga(\tau|\alpha_0,\beta_0)\int_\mu \mathcal{N}\left(\mu|\mu_\mu,v_\mu\right)\mathcal{N}\left(\mu_x|\mu,\tau^{-1}+v_x\right)d\mu \tag{A2}$$

And integrating out μ yields

$$p(\tau)\propto Ga(\tau|\alpha_0,\beta_0)\,f(\tau) \tag{A3}$$

where $f(\tau)=\mathcal{N}\left(\mu_x|\mu_\mu,\tau^{-1}+v_x+v_\mu\right)$. This gives $p(\tau)$ as a nonstandard distribution, but we can approximate it with a gamma distribution $q(\tau)$ by derivative matching, that is, by matching first and second derivatives of $p(\tau)$ and $q(\tau)$ (in log scale) at a given point. Taking the gamma density $q(\tau)$ and its derivatives:

$$logq(\tau)\propto(\alpha-1)\,log\,\tau-\beta\tau \tag{A4}$$

$$\frac{d}{d\tau}logq(\tau)=\frac{\alpha-1}{\tau}-\beta \tag{A5}$$

$$\tau^2\frac{d^2}{d\tau^2}logq(\tau)=-(\alpha-1) \tag{A6}$$

The corresponding expressions for $\log p(\tau)$ and its derivatives are

$$\log p(\tau)\propto(\alpha_0-1)\log\tau-\beta_0\tau+logf(\tau) \tag{A7}$$

$$\frac{d}{d\tau}\log p(\tau)=\frac{\alpha_0-1}{\tau}-\beta_0+\frac{d}{d\tau}\log f(\tau) \tag{A8}$$

$$\tau^2\frac{d^2}{d\tau^2}\log p(\tau)=-(\alpha_0-1)+\tau^2\frac{d^2}{d\tau^2}\log f(\tau) \tag{A9}$$

where

$$\log f(\tau)=-\frac{1}{2}\log a-\frac{b^2}{2a} \tag{A10}$$

$$\frac{d}{d\tau}\log f(\tau)=\frac{1}{2\tau^2 a}-\frac{b^2}{2\tau^2a^2} \tag{A11}$$

$$\tau^2\frac{d^2}{d\tau^2}\log f(\tau)=-\frac{1}{\tau\,a}+\frac{1}{2\tau^2a^2}+\frac{b^2}{\tau a^2}-\frac{b^2}{\tau^2a^3} \tag{A12}$$

with $a=\tau^{-1}+v_x+v_{mu}$ and $b=\mu_x-\mu_\mu$. Equating derivatives of $\log p(\tau)$ and $\log q(\tau)$ and solving for α and β yields the following update equations:

$$\alpha = \alpha_0 - \tau^2 \frac{d^2 \log f\ (\tau)}{d\tau^2} \tag{A13}$$

$$\beta = \beta_0 - \frac{d\ \log f\ (\tau)}{d\tau} - (\alpha - \alpha_0)\tau^{-1} \tag{A14}$$

$$\tau = \frac{\alpha - 1 + w}{\beta} \tag{A15}$$

where w is considered equal to 0.5 to match derivatives at a point in between mode ($w = 0$) and mean ($w = 1$) of the gamma distribution. Equation (A15) updates τ to a point in between mode and mean of the new approximation. This can be thought of as a modified Newton update. Iteratively applying these equations results in τ eventually becoming equal to the mean of p and yields a gamma approximation q with the same derivatives (in log-scale) as p at the mean of p. Alternatively, the last equation can be replaced by $\tau = \frac{\alpha - 1}{\beta}$, which leads to τ becoming equal to the mode of p, and a resulting gamma approximation q with the same derivatives as p at the mode of p.