

Effect of different uncertainty sources on the skill of 10 day ensemble low flow forecasts for two hydrological models

Mehmet C. Demirel,¹ Martijn J. Booij,¹ and Arjen Y. Hoekstra¹

Received 27 November 2012; revised 23 April 2013; accepted 5 May 2013; published 8 July 2013.

[1] This paper aims to investigate the effect of uncertainty originating from model inputs, parameters and initial conditions on 10 day ensemble low flow forecasts. Two hydrological models, GR4J and HBV, are applied to the Moselle River and performance in the calibration, validation and forecast periods, and the effect of different uncertainty sources on the quality of low flow forecasts are compared. The forecasts are generated by using meteorological ensemble forecasts as input to GR4J and HBV. The ensembles provided the uncertainty range for the model inputs. The Generalized Likelihood Uncertainty Estimation (GLUE) approach is used to estimate parameter uncertainty. The quality of the probabilistic low flow forecasts has been assessed by the relative confidence interval, reliability and hit/false alarm rates. The daily observed low flows are mostly captured by the 90% confidence interval for both models. However, GR4J usually overestimates low flows whereas HBV is prone to underestimate them, particularly when the parameter uncertainty is included in the forecasts. The total uncertainty in GR4J outputs is higher than in HBV. The forecasts issued by HBV incorporating input uncertainty resulted in the most reliable forecast distribution. The parameter uncertainty was the main reason reducing the number of hits. The number of false alarms in GR4J is twice the number of false alarms in HBV when considering all uncertainty sources. The results of this study showed that the parameter uncertainty has the largest effect whereas the input uncertainty had the smallest effect on the medium range low flow forecasts.

Citation: Demirel, M. C., M. J. Booij, and A. Y. Hoekstra (2013), Effect of different uncertainty sources on the skill of 10 day ensemble low flow forecasts for two hydrological models, *Water Resour. Res.*, 49, 4035–4053, doi:10.1002/wrcr.20294.

1. Introduction

[2] Rain-fed rivers in Western and Central Europe have a discharge regime with high flows in winter and low flows in late summer due to the temperate climate. The rivers, e.g., the River Rhine, are generally navigable throughout the year, a situation which has contributed to the region's industrial and trade development. The rivers are used for drinking water supply, irrigation, industrial use, power production, and freight shipment and also fulfill ecological and recreational functions [De Wit *et al.*, 2007]. Floods and low flows are seasonal phenomena that may cause several problems to society. Since floods are eye-catching, quick, and violent events risking human life, contingency plans and water management bodies often focus on flood issues. In contrast, low flows are slowly developing events affecting a much larger area than floods. There is a growing conc-

ern that low flows will intensify due to climate change [Arnell, 1999; Grabs *et al.*, 1997; Hagemann *et al.*, 2008; Middelkoop *et al.*, 2001]. Low flows in rivers may negatively affect the above-mentioned river functions. Severe problems, e.g., water scarcity for drinking water supply and power production, hindrance to navigation and deterioration of water quality, have already been experienced during low flow events in the River Rhine in the dry summers such as in 1969, 1976, 1985, and 2003, indicating the importance of considering these events in addition to flood events.

[3] To anticipate possible low flow events it is crucial that 10 day low flow forecasts become available in addition to short-range (1–4 days) forecasts. The forecasted low flow is commonly given as one value, even though it is an uncertain value. There is an increasing interest to account for uncertain information in decision support systems, e.g., how to operate river navigation and power plants during low flow periods to maximize the gain. One challenge is to develop systems that can use uncertain information [Engelard *et al.*, 2010]. We are interested in forecasting low flows with a lead time of 10 days, and in presenting corresponding uncertainty to provide low flow information to major river users. This study focuses on assessing the uncertainty in ensemble 10 day low flow forecasts for two conceptual hydrological models.

¹Water Engineering and Management, Faculty of Engineering Technology, University of Twente, Enschede, Netherlands.

Corresponding author: M. C. Demirel, Water Engineering and Management, Faculty of Engineering Technology, University of Twente, PO Box 217, NL-7500 AE Enschede, Netherlands. (m.c.demirel@rhinelowflows.nl)

[4] Carrying out a systematic uncertainty analysis in hydrological modeling is an important field in hydrology, according to the numerous recent contributions in well-known journals [Cunha et al., 2012; Rossa et al., 2011; Salamon and Feyen, 2009; Tolson and Shoemaker, 2008]. Uncertainty assessment has been one of the main goals of the Prediction in Ungauged Basins initiative promoted by the International Association of Hydrological Sciences [Montanari, 2011]. Similarly, the Hydrological Ensemble Prediction Experiment, another international initiative, published a special issue on the results of the intercomparison experiment for postprocessing techniques for ensemble forecasts [Van Andel et al., 2012]. Systematic uncertainty analysis consists of several steps such as, classification, importance assessment, quantification, uncertainty propagation through model, and finally, communication of the uncertainty to the end users. After the identification of the main sources of uncertainty [Ewen et al., 2006], these sources must be classified. There are many different approaches for source classification. For instance, Walker et al. (2003) classified uncertainty as originating from model context, input, model structure, and parameters. Other studies distinguished the uncertainty in observations, instruments, and the context of the problem, expert judgment, and indicators [Janssen et al., 2005; Van der Sluijs et al., 2005; Warmink et al., 2011]. It has been commonly accepted that model inputs, parameters, initial conditions, and structure are the major sources of uncertainty in conceptual hydrological models [Refsgaard et al., 2006; Zappa et al., 2011]. We focus on these sources for further analysis. Quantification of the uncertainty sources is probably the most difficult step of the uncertainty analysis.

[5] Uncertainty in forecasted input data, e.g., precipitation and temperature, is mainly from the assumptions and simplifications made when describing atmospheric processes in weather forecast models. In particular, future precipitation amounts are assumed to be very uncertain [Cunha et al., 2012; Roulin, 2007]. To quantify the uncertainty in weather forecasts, an ensemble of lower-resolution forecasts (ENS) has been developed by the European Centre for Medium-Range Weather Forecasts (ECMWF) and other national meteorological services [European Centre for Medium-Range Weather Forecasts, 2012]. The system is operational since 1992, and a number of modifications have been implemented to its structure and grid resolution for improving the numerical weather predictions. In this system, there are 50 different perturbed weather forecasts and an unperturbed control forecast. The 50 members, comprising an ensemble, are computed for a lead time of 15 days using perturbed initial conditions and model physics [Pappenberger et al., 2005; Roulin and Vanitsem, 2005]. Each member of an ensemble is assumed to be equally probable and provide useful information to address the uncertainty in future precipitation amounts [Roulin, 2007]. However, in the context of flow forecasting it is important to assess the precipitation uncertainty in terms of the effect on runoff rather than in terms of comparing forecasted precipitation against observed precipitation [Arnaud et al., 2011; Nester et al., 2012]. For example, Pappenberger et al. [2005] and more recently Pappenberger et al. [2011] used meteorological ensembles in hydrological models with dif-

ferent parameter settings to assess the uncertainty in flood forecasts. Similarly, Randrianasolo et al. [2010] coupled weather ensemble prediction system products from Météo-France with two hydrological models for forecasting discharges of 211 catchments in France for a lead time of 2 days.

[6] Obviously, there are other sources of uncertainty in low flow forecasts in addition to the model input [Meißner et al., 2012; Zappa et al., 2011]. Hydrological models, whether using observed [Cunha et al., 2012] or forecasted [Nester et al., 2012] rainfall, are also limited by their capacity to represent the dominant processes in the river basin with appropriate spatial and temporal scales. Effective values of model parameters affected by local spatial heterogeneities and nonstationarities provide usually loose associations with dominant processes [Lawal et al., 1997; Pappenberger et al., 2005; Stravs and Brilly, 2007]. Therefore, the uncertainty due to model parameters will inevitably influence model outputs. There are a range of methods of quantifying model parameter errors including Monte Carlo simulations and analytical approaches [Montanari and Grossi, 2008]. Generalized likelihood uncertainty estimation (GLUE) is a Monte Carlo based technique developed for calibration and estimation of uncertainty of predictive models using equifinality concept [Beven and Freer, 2001; Stedinger et al., 2008; Viola et al., 2009]. Concerning the choice of the likelihood measure, Beven and Binley [1992] pointed out that many likelihood measures in GLUE can be appropriate for a given application. Jin et al. [2010] compared different likelihood measures and the model uncertainty. They found that a less strict likelihood function obviously leads to a wider confidence interval of the output uncertainty. Therefore, neither a too strict nor a too relaxed likelihood is appropriate for the GLUE assessment. The GLUE method has been widely used for flood forecasting [Pappenberger et al., 2004, 2005] and for simulation of both high and low flows [Freer et al., 1996; Tolson and Shoemaker, 2008; Vázquez et al., 2009; Viola et al., 2009; Tian et al., 2012]. In addition, the GLUE method is simple and relatively easy to implement. Therefore, GLUE is used in this study for model calibration and uncertainty analysis.

[7] The drawbacks and advantages of the GLUE method have been enormously discussed in the hydrology literature [Beven and Young, 2003; Beven et al., 2007, 2008; Li et al., 2010; Mantovan and Todini, 2006; Montanari, 2005; Stedinger et al., 2008].

[8] Beven et al. [2007] showed that if a correct formal likelihood was used in the GLUE method, the results would be identical with the formal Bayesian technique. Stedinger et al. [2008] also showed that GLUE can produce meaningful uncertainty and prediction intervals using a correct likelihood function. Beven [2006] argues whether the assumptions used in a formal Bayesian analysis are valid for any nonsynthetic hydrologic system being modeled. In our ensemble low flow forecasting case study, performing a formal Bayesian method would require a very difficult process of deriving a correct description of the residual errors which are correlated in time and space. As a result, our study utilizes an elaborated low flow likelihood function within the GLUE method to assess parameter uncertainty.

[9] Uncertainty in initial condition of state variables can have a significant effect on low flow forecasts. The summer forecasts, using a model initialized with an unsaturated soil state and with a saturated soil, will be very different. During prolonged dry periods the discharge largely originates from the release of groundwater storage [De Wit et al., 2007; Tallaksen and Van Lanen, 2004]. Therefore, uncertainty coming from the model initial conditions and groundwater storage in particular should be treated separately from the model parameter uncertainty, although model parameters and states are part of the model structure [Butts et al., 2004]. The uncertainties may be amplified when cascaded through the hydrological model [Nester et al., 2012]. Komma et al. [2007] showed that small errors in rainfall may result in larger errors in model outputs. They showed that an uncertainty range of 70% in the precipitation ensemble increased to an uncertainty range of 200% in forecasted runoff with a lead time of 48 h. Although they related this to the nonlinearity of the catchment response, this amplification could also be caused by uncertainties from precipitation measurements and model parameters [Nester et al., 2012]. Komma et al. [2007] also showed that the assessment of precipitation uncertainty should be in terms of the effect on model outputs (herein forecasted low flows) instead of comparing only forecasted precipitation and observed precipitation.

[10] Different methods, i.e., particle filter [DeChant and Moradkhani, 2011] and ensemble Kalman filter [Pasetto et al., 2012], have been applied to assess initial condition uncertainty in the framework of ensemble streamflow prediction. The resulting update is similar to the static GLUE application [Pasetto et al., 2012]. These filters are examples of data assimilation techniques that are often used in short-term flood forecasts [Liu et al., 2012; Moradkhani et al., 2012; Parrish et al., 2012].

[11] The aforementioned studies demonstrate the need for a systematic uncertainty analysis framework that isolates uncertainties due to various weather inputs, parameter estimation, and initial conditions. Understanding the relative contributions of these sources to the total low flow forecast uncertainty and to the quality of forecasts can assist in the future development of ensemble forecasting systems. All studies mentioned constrain either to only flood forecast models or to simulation models used for low flows, but no similar application to low flow forecasts is known to the authors which also uses a sound model state updating procedure for assessing effect of initial condition uncertainty. There have been studies using ENS products for flood forecasting [Devineni et al., 2008; Fundel and Zappa, 2011; Jaun and Ahrens, 2009; Muluye, 2011; Nester et al., 2012; Pagano et al., 2012; Renner et al., 2009; Thirel et al., 2008, 2010] and high-resolution precipitation ensemble forecast of a regional climate model, i.e., Limited Area Ensemble Prediction System developed within COSMO consortium (COSMO-LEPS) [Addor et al., 2011; Zappa et al., 2011], but no study is known to the authors which focuses on 10 day low flow forecasts. Only short-range low flow forecasts up to 4 days are issued by different water authorities for the entire Rhine basin [De Bruijn and Passchier, 2006]. There have been different cross-border projects such as “Floods and low flow manage-

ment in the Moselle and Saar Basin (FLOW MS)” focusing on climate impacts on low flows [Görge et al., 2010]. However, low flow forecasts followed by a systematic uncertainty analysis do not exist for the Rhine basin and Moselle River, in particular, although there is a high demand [Meißner et al., 2012; Rutten et al., 2008] from different sectors (e.g., freight shipment, drinking water supply, and energy production).

[12] The objectives of this study are to assess (1) the uncertainty from ECMWF ensemble forecasted precipitation and potential evapotranspiration; (2) the uncertainty from the parameters of two hydrological models by using the GLUE framework; (3) the uncertainty due to the initial conditions; and (4) the effect of these uncertainties on different low flow forecast quality and reliability measures.

[13] Assessing the isolated three major sources of uncertainty, i.e., model input, parameters, and initial conditions, in ensemble low flow forecasts by applying all steps from identification to communication of uncertainty, is an innovative way to understand and explain the effects of different uncertainties on the skill of low flow forecasts. In terms of model storage update, i.e., estimation of the model states at forecast issue day, a new method is proposed using observed discharge. This is superior compared to using only calibrated model run as there can be inevitable errors between simulated and observed discharge affecting the model initial condition. A further interesting aspect of this study is the use of a new hybrid low flow likelihood function for GLUE, which allows evaluation of low flows. Low flows in the Moselle River are investigated to allow the navigation and energy sectors to timely prepare for low flow conditions as they are the most important economic river functions [Li et al., 2008; Rutten et al., 2008; Svensson and Prudhomme, 2005]. Since the River Rhine is a large-scale river, the Moselle River is selected as a case study. We use an exceedence probability of 75% (Q_{75}) as a threshold for the definition of low flows [Demirel et al., 2012]. The number of days with low flows is sufficient to calibrate a forecast model, and low flows at this threshold are still affecting the important river functions. Several types of ensemble weather forecast products from the ENS data set are incorporated in this study to prepare model inputs, i.e., daily precipitation (P) and potential evapotranspiration (PET) for a lead time of 10 days from the ENS data set. We address the model structure uncertainty by comparing two conceptual models with different complexities: the GR4J conceptual model (Génie Rural à 4 paramètres Journalier) with four parameters [Perrin et al., 2003] and the HBV conceptual model (Hydrologiska Byråns Vattenbalansavdelning) with eight parameters [Lindström et al., 1997]. These models are assumed to represent dominant low flow indicators (predictors) with their appropriate temporal scales as identified by Demirel et al. [2012]. The GR4J model is a French conceptual hydrological model with a simple structure [Perrin et al., 2003]. With four parameters, it provides a minimum level of complexity. The HBV model has been calibrated and operationally used for the River Rhine [Renner et al., 2009]. Moreover, this model has been widely used in Rhine studies such as for real-time flow forecasts [Reggiani and Weerts, 2008b], for climate impact assessment [Eberle,

2005; Hurkmans *et al.*, 2010; Te Linde *et al.*, 2010, 2011], and for assessing uncertainties in flood forecasts due to ensemble weather forecasts by using a Bayesian postprocessor [Reggiani and Weerts, 2008a; Reggiani *et al.*, 2009].

[14] This paper is organized as follows. In the next section, the study area and data are presented. The model structures and the uncertainty analysis method are described in section 3. The results are presented in section 4, and the conclusions are drawn in section 5.

2. Study Area and Data

2.1. Study Area

[15] The Moselle River has a surface area of approximately 27,262 km² and a length of 545 km. The source of the river is in the forested slopes of the Vosges massif and meanders before leaving France to form the border between Germany and Luxembourg for a short distance. The river enters Germany and flows past Trier to its confluence with the Rhine at Koblenz. Two major tributaries, the Sauer and Saar rivers, flow into the Moselle before the Trier dam. There are other dams in the Moselle and Saar rivers, whereas the Sauer river has a natural flow [Ackermann *et al.*, 2000; Belz *et al.*, 1999]. Moreover, the river channels in the Moselle and Saar are mostly canalized for water management purposes and available for river navigation, while the Sauer is not navigable [Behrmann-Godel and Eckmann, 2003]. Annual generated discharge in the Moselle basin is about 410 mm ($\sim 130 \text{ m}^3 \text{ s}^{-1}$). The measured discharge at Cochem station fluctuates between 14 $\text{m}^3 \text{ s}^{-1}$ in dry summers and a maximum of 4000 $\text{m}^3 \text{ s}^{-1}$ during winter floods. The altitude ranges from 59 to 1326 m with a mean altitude of 340 m [Demirel *et al.*, 2012].

2.2. Data

2.2.1. Observed Data

[16] Observed daily precipitation (P) and potential evapotranspiration (PET) estimated with the Penman-Wendling equation [Abwassertechnische Vereinigung - Deutscher Verband für Wasserwirtschaft und Kulturbau e. V. (ATV-DVWK), 2002] were obtained from the German Federal Institute of Hydrology (BfG) in Koblenz (Germany). Both variables are spatially averaged, i.e., disaggregated over 26 Moselle subcatchments.

[17] The mean altitude of these subcatchments has been also provided by BfG. The outlet discharge (Q) for the Moselle (station 6336050 at Cochem) has been provided by the Global Runoff Data Centre (GRDC), Koblenz (Germany). The daily P, PET, and Q data series span from 1951 to 2006 (Table 1).

2.2.2. Meteorological Ensemble Forecast Data

[18] Both precipitation and other meteorological forecast data used in this study are originated from the ECMWF-ENS

control and ensemble forecasts. These ensembles are computed for a lead time of 1–10 days using perturbed initial conditions and model physics (Table 2). A grid size of 0.25° ($\sim 28 \text{ km}$) is chosen to retrieve weather forecast products using the ECMWF Mars retrieval system. The PET forecasts are determined by the Penman-Wendling equation requiring only forecasted surface solar radiation and temperature at 2 m data [ATV-DVWK, 2002]. This is consistent with the observed PET estimation carried out by the Federal Institute of Hydrology in Koblenz, Germany. Both grid-based P and PET ensemble forecast data are first interpolated over 26 Moselle subcatchments using areal weights. These subcatchment averaged data are then aggregated to the Moselle basin level.

3. Methodology

3.1. Overview of the Model Structures

[19] The two hydrological models (GR4J and HBV) are briefly described later. Figure 1 shows the simplified model structures.

3.1.1. GR4J

[20] The GR4J conceptual model has a parsimonious structure with only four calibration parameters and has been frequently used over hundreds of catchments worldwide, with a broad range of climatic conditions from tropical to temperate and semiarid catchments [Perrin *et al.*, 2003]. The GR4J model requires only daily time series of precipitation (P) and potential evapotranspiration (PET) as inputs (Figure 1a). The four parameters in GR4J represent the maximum capacity of the production store (X1), the groundwater exchange coefficient (X2), the 1 day ahead capacity of the routing store (X3), and the time base of the unit hydrograph (X4). All four parameters are used to calibrate the model and estimate the parameter uncertainty (Table 3) based on Tian *et al.* [2012] and Thyer *et al.* [2009]. The upper and lower limits are selected based on previous works [Booij, 2005; Eberle, 2005; Perrin *et al.*, 2003; Pushpalatha *et al.*, 2011; Tian *et al.*, 2012].

3.1.2. HBV

[21] The HBV conceptual model was developed by the Swedish Meteorological and Hydrological Institute in the early 1970s [Lindström *et al.*, 1997]. The HBV model consists of four subroutines: a precipitation and snow accumulation and melt routine, a soil moisture accounting routine, and two runoff generation routines. The input data are daily P and PET. Since the Moselle basin is a rain-fed basin, the snow routine and daily temperature data are not used in this study (Figure 1b). The eight most important parameters in the HBV model (Table 3) are used to estimate the

Table 1. Observed Data

Data	Index	Spatial Resolution	Number of Stations/Subbasins	Period	Time Step	Source
Discharge at Cochem station	Q	Point	1	1951–2006	24 h	GRDC-Koblenz
Precipitation	P	Subcatchments	26	1951–2006	24 h	BfG-Koblenz
Evapotranspiration	PET	Subcatchments	26	1951–2006	24 h	BfG-Koblenz
Mean altitude	h	Subcatchments	26			BfG-Koblenz

Table 2. Overview of the Ensemble Forecast Data

Data	Index	Spatial Resolution	Ensemble Size	Period	Time Step	Lead Time	R^2 (240 h)	MAE (240 h)
Forecasted P	P_{forecast}	28 km	50 + 1 control	2002–2005	24 h	240 h	0.07	3.16 mm
Forecasted PET	PET_{forecast}	28 km	50 + 1 control	2002–2005	24 h	240 h	0.77	0.59 mm

parameter uncertainty [Engeland et al., 2010; Tian et al., 2012; Van den Tillaart et al., 2012].

3.2. Calibration and Validation

[22] The GR4J and HBV models are calibrated using the GLUE method and historical Moselle low flows for the period from 1 January 1971 to 31 December 2001. The calibration period is rigorously selected as the first forecast issue date is 1 January 2002 and the number of low flow events (i.e., 567 days with low flows) in the calibration period is reasonably long for hydrological models [Perrin et al., 2007]. The validation period spans from 1 January 1951 to 31 December 1970. The definition of low flows, i.e., discharges below the Q_{75} threshold of $\sim 113 \text{ m}^3 \text{ s}^{-1}$, is based on previous work by Demirel et al. [2012].

[23] The GLUE method [Beven and Binley, 1992] uses the “equifinality” concept rejecting only one optimal parameter set, instead, it uses many parameters sets that provide relatively equal performance [Beven and Freer, 2001]. This method is developed as an extension of the generalized sensitivity analysis (GSA) of Spear and Hornberger [1980] based on Bayesian Monte Carlo simulations. GLUE has been widely used for calibration of hydrological models since it is easy to implement and allows flexible definition of a likelihood function to evaluate the model outputs and to distinguish between behavioral (accepted) and non-behavioral (rejected) parameter sets [Freer et al., 1996; Ratto et al., 2007; Renard et al., 2010; Shen et al., 2012]. Behavioral parameter sets are then those that provide

predicted low flows that fall within the limits of acceptability with regard to a given likelihood measure [Zheng and Keller, 2007]. It should be noted that the selection of the behavioral parameter sets is based on only the calibration period runs. In this study, the GLUE method, consisting of the three steps later, is applied for the selection of behavioral parameter sets. It is assumed that these parameter sets represent the uncertainty in model parameters.

3.2.1. Step 1: Definition of a Hybrid Likelihood Function for Low Flows

[24] The most commonly used likelihood function in GLUE literature is the Nash-Sutcliffe (NS) coefficient [Beven and Freer, 2001; Nash and Sutcliffe, 1970; Shen et al., 2012]. However, other likelihood functions have been used for low flows [Pushpalatha et al., 2012]. In our study, we combined two low flow likelihood functions using subjectively selected weights. The new hybrid likelihood function (NS_{hybrid}) substantially improves the low flow forecasts as it combines NS based on only low flows (NS_a) and NS based on inverse discharge values (NS_b) (see equations (1)–(3)).

$$NS_a = 1 - \frac{\sum_{j=1}^m (Q_{\text{sim}}(j) - Q_{\text{obs}}(j))^2}{\sum_{j=1}^m (Q_{\text{obs}}(j) - \bar{Q}_{\text{obs}})^2}, \quad (1)$$

where Q_{obs} and Q_{sim} are the observed and simulated values

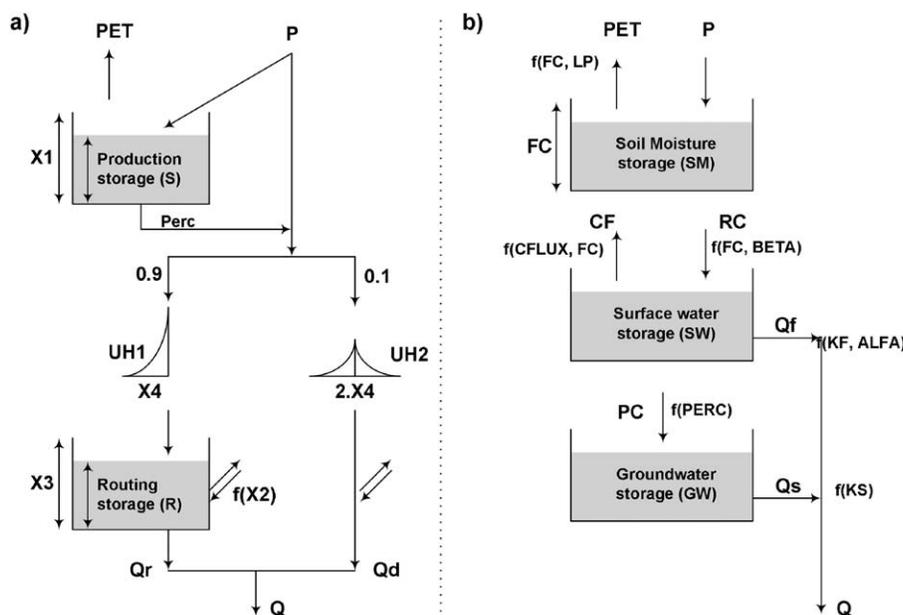

Figure 1. Schematic of (a) GR4J model and (b) HBV model.

Table 3. Parameters of the Models Used and Their Prior Uncertainty Ranges

Parameter	Unit	Range	Description
<i>GR4J Model</i>			
X1	mm	10–2000	Capacity of the production store
X2	mm	–8 to +6	Groundwater exchange coefficient
X3	mm	10–500	One day ahead capacity of the routing store
X4	days	0–4	Time base of the unit hydrograph
<i>HBV Model</i>			
FC	mm	100–800	Maximum soil moisture capacity
LP		0.1–1	Soil moisture threshold for reduction of evapotranspiration
BETA		1–6	Shape coefficient
CFLUX	mm/d	0.1–1	Maximum capillary flow from upper response box to soil moisture zone
ALFA		0.1–2	Measure for nonlinearity of low flow in quick runoff reservoir
KF	d ⁻¹	0.005–0.5	Recession coefficient for quick flow reservoir
KS	d ⁻¹	0.0005–0.2	Recession coefficient for base flow reservoir
PERC	mm/d	0.01–6	Maximum flow from upper to lower response box

for the j th observed low flow day (i.e., $Q_{obs} < Q75$), and m is the total number of low flow days.

$$NS_b = 1 - \frac{\sum_{i=1}^n (1/(Q_{sim}(i) + \epsilon) - 1/(Q_{obs}(i) + \epsilon))^2}{\sum_{i=1}^n (1/(Q_{obs}(i) + \epsilon) - 1/(\bar{Q}_{obs} + \epsilon))^2}, \quad (2)$$

[25] where n is the total number of days (i.e., $m < n$), and ϵ is 1% of the mean observed discharge to avoid infinity during zero discharge days.

$$NS_{hybrid} = \alpha \left(\frac{NS_a}{2 - NS_a} \right) + \beta \left(\frac{NS_b}{2 - NS_b} \right), \quad (3)$$

where both NS_a and NS_b values range from $-\infty$ to 1, with 1 indicating a perfect fit [Pushpalatha et al., 2012]. The weights α and β are selected as 0.3 and 0.7, respectively. These weights have been determined during calibration period. First component of our hybrid likelihood function is strictly developed for the low flows. Therefore, the resultant scores for this component can often be negative. Second component considers the inverse of all discharge values. The weights are determined for making the outcome values of our hybrid likelihood function positive for the calibration runs. In other words, the weights keep the balance between very strict and less strict likelihood functions since in cases with very strict low flow calibration, i.e., high α values, both the GR4J and HBV models show results with very low likelihood values since the NS_a values are negative for both models.

3.2.2. Step 2: Sampling Parameter Sets for Two Conceptual Models

[26] Previous model calibration and sensitivity analysis of the GR4J [Perrin et al., 2003; Pushpalatha et al., 2011] and HBV [Booij, 2005; Eberle, 2005; Tian et al., 2012] in other rain-fed basins have allowed the prior uncertainty ranges of sensitive parameters to be assessed. These studies also indicated significant uncertainties for the sensitive parameters and emphasized the importance of inspecting the upper and lower parameter limits more in detail. Therefore, a sensitivity analysis is pursued using a large parameter space to select the

most important parameters and their appropriate upper and lower limit values.

[27] Independent uniform distributions for each effective parameter are chosen due to the lack of prior knowledge about the true distributions. The typical drawback of the GLUE method is the computational time caused by its random sampling strategy. Therefore, an improved sampling technique, i.e., Latin hypercube sampling (LHS), was used with the GLUE method [McKay et al., 1979]. Compared to a standard GLUE random sampling, LHS substantially reduces the computational burden for sampling and provide a tenfold greater efficiency in parameter space coverage [Shen et al., 2012]. The sampling size should be large enough to ensure a sufficient calibration of the model. In this study, we generated 120,000 parameter sets for each conceptual model using LHS in the range of lower and upper limits given in Table 3. To our knowledge, this is the largest LHS sample size tested in low flow hydrology.

3.2.3. Step 3: Threshold Definition for Behavioral Model Selection

[28] The GR4J and HBV models are run for each of the 120,000 sets in the calibration. The output is evaluated against the observed daily discharge at Cochem station located at the outlet of the Moselle subbasin using the NS_{hybrid} likelihood function to distinguish between behavioral parameter sets (accepted) and nonbehavioral parameter sets (rejected). The parameter sets meeting the predefined threshold criterion ($NS_{hybrid} > 0.40$) are accepted. Although the threshold value is a subjective decision [Jin et al., 2010], we rigorously tested several thresholds based on low flow simulations and the size of the behavioral parameter sets for each model. The selected threshold resulted in two large behavioral parameter sets for parameter uncertainty analysis, i.e., 9770×4 (GR4J) corresponding to $\sim 8\%$ of the sample parameter set and $10,909 \times 8$ (HBV) corresponding to $\sim 9\%$ of the sample parameter set.

3.3. Model Storage Update Procedure

[29] Model storage updating is based on the observed discharge on the forecast issue day (Q_{obs}). This is a crucial step for medium-range and seasonal low flow forecasts

since the model initial state determines the model outputs [Wöhling *et al.*, 2006].

[30] There are two storages in the GR4J model and three storages in the HBV model which are updated during low flow forecasts. The reader is referred to Perrin *et al.* [2003] and Lindström *et al.* [1997] for details of the process formulations in these models. A practical approach is used for both models. First, the two calibrated models are run with their best performing parameter sets, and model states are analyzed. This run is called the “reference run” in other recent works [Fundel and Zappa, 2011; Roulin, 2007; Roulin and Vannitsem, 2005]. The empirical relations between the simulated discharge and the fast runoff for each model are used to divide the observed discharge between the fast and slow runoff components (equations (4) and (5)).

$$k_{\text{GR4J}} = \frac{Q_d}{Q_r + Q_d} \quad (4)$$

$$k_{\text{HBV}} = \frac{Q_f}{Q_f + Q_s}. \quad (5)$$

[31] The Q_r and Q_d in the GR4J model, and Q_f and Q_s in the HBV model are estimated using these fractions together with the observed discharge value at the forecast issue day. Subsequently, the routing storage (R) in the GR4J model is updated for a given value of the $X3$ parameter using equation (6). Further, the surface water (SW) and groundwater (GW) storages in the HBV model can be updated for a given value of KF, ALFA, and KS parameters using equations (7) and (8).

$$Q_r = R \left\{ 1 - \left[1 + \left(\frac{R}{X3} \right)^4 \right]^{-1/4} \right\} \quad (6)$$

$$SW = \left(\frac{Q_f}{KF} \right)^{(1/(1+ALFA))} \quad (7)$$

$$GW = \frac{Q_s}{KS} \quad (8)$$

[32] The other two storages S (in GR4J) and SM (in HBV) are difficult to update using an empirical bottom-up approach. Instead, these two storages are updated using the calibrated model run until the forecast issue day (i.e., top-down approach). It is assumed that the two updated storages (S and SM) represent the reality due to the calibrated model run. However, we are aware that there are inevitable uncertainties associated with this rough estimation of the initial conditions based on observed discharges and calibrated models’ simulations.

3.4. Uncertainty Sources and Quantification

[33] A robust assessment of uncertainties begins with identification of all sources [Refsgaard *et al.*, 2007]. Obviously, not all sources can be quantified. We used available classification schemes to identify and select the most important uncertainty sources in the two conceptual models and for low flow forecasting [Walker *et al.*, 2003; War-

mink *et al.*, 2010]. Three uncertainty sources and their quantification are described in the following sections. It should be noted that errors in model structure are also important for hydrological models [Göttinger and Bárdossy, 2008; Gupta *et al.*, 2012; Renard *et al.*, 2010; Tian *et al.*, 2012]. In this study, the model structure uncertainty is addressed partly by comparing two different model structures.

3.4.1. Input Uncertainty

[34] A rainfall event after the forecast issue day can easily increase flows above low flow threshold in the Moselle River. Therefore, low flow forecasts are highly dependent on the quality of ENS weather forecasts. These forecasts are available at a spatial resolution of approximately $28 \text{ km} \times 28 \text{ km}$ for daily time steps. It has been reported that after several days these forecasts are highly uncertain due to the modeling limitations and complexity of the physical processes involved in the atmosphere [Fundel and Zappa, 2011; Reggiani *et al.*, 2009]. In this study, the 51 ensemble members are used to quantify the uncertainty of future precipitation and potential evapotranspiration amounts. Obviously, new uncertainties are introduced using the empirical PET formula [ATV-DVWK, 2002] and grid data interpolation over 26 Moselle subcatchments.

3.4.2. Parameter Uncertainty

[35] The GLUE method is used for quantification of parameter uncertainties. This method rejects the idea of an optimal system representation and applies the equifinality concept accepting all forecasts using the behavioral parameter sets. This parameter ensemble then allows assessment of the output uncertainty arising from model parameters and partly from model structure [Pappenberger *et al.*, 2005]. All four parameters of GR4J and eight parameters of the HBV are selected to estimate the parameter uncertainty.

3.4.3. Initial Condition Uncertainty

[36] The importance of initial conditions for hydrologic forecasting is well established [Shukla and Lettenmaier, 2011; Wood and Lettenmaier, 2008]. In most of the hydrologic modeling studies, initial conditions refer only to land surface states including soil moisture and snow cover [Li *et al.*, 2009]. In this study, however, all model states, present in the conceptual models used, are included in the uncertainty analysis. This is from the fact that errors in estimated initial slow and fast runoff storages directly affect the low flow forecasts. We demonstrate a dynamic inverse-modeling approach based on observed discharge and uniformly distributed behavioral parameter sets on the forecast issue day for exploring the initial condition uncertainties and characterizing the relative importance of this uncertainty source for low flow forecasts. The $X3$ parameter of GR4J and KF, ALFA, and KS parameters of HBV are uncertain parameters that are directly linked to the model states (i.e., initial conditions). Other parameters are only assumed uncertain in the parameter uncertainty assessment.

3.5. Uncertainty Propagation

[37] The three sources of uncertainty described earlier are propagated through the GR4J and HBV models both

separately and together. The latter case is executed to encapsulate the total uncertainty arising from all three sources together. In other words, this study employs the GLUE, an extended GSA method [Freer et al., 1996; Ratto et al., 2007], to apportion the output uncertainty of a model to different sources of uncertainty.

[38] The 10 day low flow forecasts are issued every day for the test period from 1 January 2002 until 31 December 2005. The posterior distribution of the model outputs (e.g., confidence interval) is based on 10,000 Monte Carlo runs for each day (a total of 1461 days). The size of the Monte Carlo sample is assumed to be reasonable based on the number of behavioral parameter sets and on the relevant literature [Blasone et al., 2008; Franz and Hogue, 2011; Rossa et al., 2011; Shen et al., 2012].

[39] For assessing the effect of the uncertainty in the forecasted input data, we run the models using randomly selected P and PET values from 51 members, while the model parameters and model states are fixed according to the best performing calibrated parameter values (Table 4).

[40] For assessing uncertainty in model parameters, we run the models using randomly selected behavioral parameter sets, while the model inputs are fixed to ECMWF-ENS control forecast P and PET, and the model states are updated using the best performing calibrated parameter values.

[41] For assessing the uncertainty in model states at forecast issue day, the routing storage (R) in the GR4J model is updated using randomly selected behavioral $X3$ parameter values, and the surface water (SW) and groundwater (GW) storages in the HBV model are updated using randomly selected values of KF, ALFA, and KS parameters from behavioral parameter sets for each of the 10,000 Monte Carlo runs for each day. The remaining model parameters are fixed to the best performing calibrated parameter values, and the model inputs are fixed to ECMWF-ENS control forecast P and PET to evaluate the initial condition uncertainty (Table 4).

[42] It should be noted that spatial and temporal consistency of the inputs are preserved to avoid nonphysical outcomes. For assessing the total uncertainty, we run the models using randomly selected model inputs, behavioral parameter sets, and corresponding model states. For example, the storage S can never exceed the $X1$ parameter value in the GR4J model [Perrin et al., 2003]. Similarly, we defined a “saturation rate” as the fraction of SM storage to the calibrated FC parameter.

$$SM(k, t) = FC(k) \times \frac{SM_{opt}(t)}{FC_{opt}}. \quad (9)$$

[43] For each k th Monte Carlo run at each t th forecast issue day, a new parameter set is randomly selected from the behavioral set, and SM is calculated using equation (9). Therefore, the saturation rate is kept constant for a particular day throughout the entire uncertainty propagation framework.

3.6. Uncertainty Presentation

[44] Uncertainty presentation allows the low flow forecasts to be monitored, thus helping to improve forecast quality by analyzing uncertainties in the model outputs and allowing comparison of different models. Obviously, the added value of a low flow forecasts for decision makers depends on its uncertainty characteristics. We employed three forecast quality measures to analyze the results of the uncertainty quantification in 10 day low flow forecasts. These measures have been often used in meteorology [WMO, 2012] and flood hydrology [Renner et al., 2009; Thirel et al., 2008; Velázquez et al., 2010]. In *World Meteorological Organization* [2012], three properties of an accurate probabilistic forecast are defined as reliability, sharpness, and resolution. In this study, three forecast quality measures have been rigorously selected to evaluate these three properties of the forecasts, i.e., reliability diagram—reliability, RCI—sharpness, and contingency table—resolution.

3.6.1. Relative Confidence Interval

[45] The standard 90% confidence interval (90CI) was derived by ordering the 10,000 outputs on every day in the test period and then identifying the 5% and 95% percentiles (i.e., $Q5$ and $Q95$). The 90CI, observed discharge, and 50% percentile (i.e., $Q50$ forecast median) are presented together. The relative confidence interval (RCI) is then calculated for only low flow days j using equation (10) to monitor the evolution of uncertainties with increasing lead time and to compare the effect of different uncertainty sources on the relative confidence interval.

$$RCI = \frac{1}{m} \times \sum_{j=1}^m \frac{Q95(j) - Q5(j)}{Q50(j)}, \quad (10)$$

where m is the total number of low flow days.

Table 4. Overview of the Uncertainty Propagation Test Scheme

Assessed Uncertainty	Forecasted P and PET		Parameters		Initial Conditions	
	1 Control	51 Ensemble	Calibrated Set	GLUE Set	Calibrated Run	GLUE set ($X3$, KF, ALFA, and KS)
Deterministic	✓		✓		✓	
Input		✓	✓		✓	
Parameter	✓			✓	✓	
Initial Condition	✓		✓			✓
Total		✓		✓		✓

Table 5. Contingency Table for the Assessment of Threshold-Based Forecasts

Low Flow Event (Q_{75})	Observed	Not Observed
Forecasted	<i>Hit</i> : the event forecast to occur and did occur	<i>False alarm</i> : event forecast to occur but did not occur
Not forecasted	<i>Miss</i> : the event forecast not to occur but did occur	<i>Correct negative</i> : event forecast not to occur and did not occur

3.6.2. Reliability Diagram

[46] The reliability diagram is an approach used to represent the performance of probabilistic forecasts of selected events, i.e., low flows [Bröcker and Smith, 2007]. A reliability diagram shows the observed relative frequency as a function of forecast probability, and the 1:1 diagonal represents the perfect reliability line [Olsson and Lindström, 2008; Velázquez et al., 2010]. In the present study, nonexceedence probabilities of 50%, 75%, 85%, 95%, and 99% are chosen as thresholds to categorize the discharges from mean flows to extreme low flows. The forecast probability for each forecast day is estimated as the number of ensemble forecasts exceeding these thresholds divided by the total number of ensemble forecasts (i.e., 10,000 ensembles) in that forecast day. The forecasts are then divided into bins of probability categories; here, five bins (categories) are chosen 0%–20%, 20%–40%, 40%–60%, 60%–80%, and 80%–100%. The observed frequency for each day is estimated as one if the observed discharge exceeds the threshold, or zero, if not. The forecast probability and observed frequency can then be drawn.

3.6.3. Contingency Table

[47] We used contingency tables to assess the effect of uncertainty on the performance of low flow forecasts. Contingency tables, particularly used in flood warnings [Martina et al., 2005], can be used to estimate the utility of hydrological forecasts and in their simplest form, indicate the forecast models ability to correctly anticipate the occurrence or nonoccurrence of preselected events (i.e., Q_{75} low flows). The definitions of four cases are given in a two-by-two contingency table (Table 5).

[48] The skill of a forecasting model can be represented on the basis of the hit rate and the false-alarm rate [Cloke and Pappenberger, 2009; Martina et al., 2005]. Both ratios can be easily calculated from the contingency table using equations (11) and (12).

$$\text{hit rate} = \frac{\text{hits}}{(\text{hits} + \text{misses})}, \quad (11)$$

$$\text{false alarm rate} = \frac{\text{false alarms}}{(\text{correct negatives} + \text{false alarms})}. \quad (12)$$

[49] It should be noted that these ratios are also known as the probability of detection and the probability of false detection in other hydrological studies [Velázquez et al., 2010]. The hit and false-alarm rates indicate, respectively, the proportion of events for which a correct warning was issued and the proportion of nonevents for which a false warning was issued by the forecast model.

4. Results and Discussion

4.1. Calibration and Validation

[50] The best performing parameter sets of the two models are shown in Table 6. The corresponding highest NS_{hybrid}

values are 0.62 for the GR4J model and 0.56 for the HBV model. The GR4J model performs better than the HBV model on low flows based on only the best performing simulation in the calibration period. However, the HBV model performed better in the validation period. The highest NS_{hybrid} values did not change using another global optimization technique, i.e., a genetic algorithm [Velázquez et al., 2010], showing that 120,000 parameter sets for each model are enough for calibrating the models. Considering the performance only in the low flow period (i.e., NS_a), the performance of the HBV model is better than the GR4J model in the calibration period. However, the drop in performance of the HBV model in the validation and the forecast periods is much larger than for the GR4J model. Such a drastic drop outside the calibration period is expected from a relatively complex model like HBV with eight parameters since it has more degrees of freedom to adjust to the basin behavior during the calibration period. This characteristic is somewhat concealed by the NS_{hybrid} results due to the subjective weights and due to the insensitive inverse performance index (i.e., NS_b). That is why all three performance indices have been presented in Table 6. The models are calibrated for a relatively wetter climate with ~ 910 mm mean annual precipitation than for the validation period (~ 890 mm) and forecast period (~ 830 mm).

[51] The calibrated models are run for the test period (i.e., 2002–2005) to estimate the fraction of fast runoff to total runoff in the two models (Figure 2) and to update storages S and SM (i.e., top-down approach). The exponential relation between the fraction values and the simulated

Table 6. Calibration, Validation, and Forecast Results

Parameter and Likelihood	1971–2001 Calibration	1951–1970 Validation	2002–2005 Forecast
<i>GR4J Model</i>			
$X1$ (mm)	649.2		
$X2$ (mm)	–1.3		
$X3$ (mm)	47.2		
$X4$ (days)	2.9		
NS_a	–0.68	–1.39	9.29
NS_b	0.99	0.76	0.99
NS_{hybrid}	0.62	0.31	0.45
<i>HBV Model</i>			
FC (mm)	239.6		
LP	0.48		
BETA	2.13		
CFLUX (mm/d)	0.13		
ALFA	1.77		
KF (d^{-1})	0.01		
KS (d^{-1})	0.01		
PERC (mm/d)	0.76		
NS_a	–0.23	–1.13	–70.16
NS_b	0.92	0.90	0.99
NS_{hybrid}	0.56	0.47	0.41

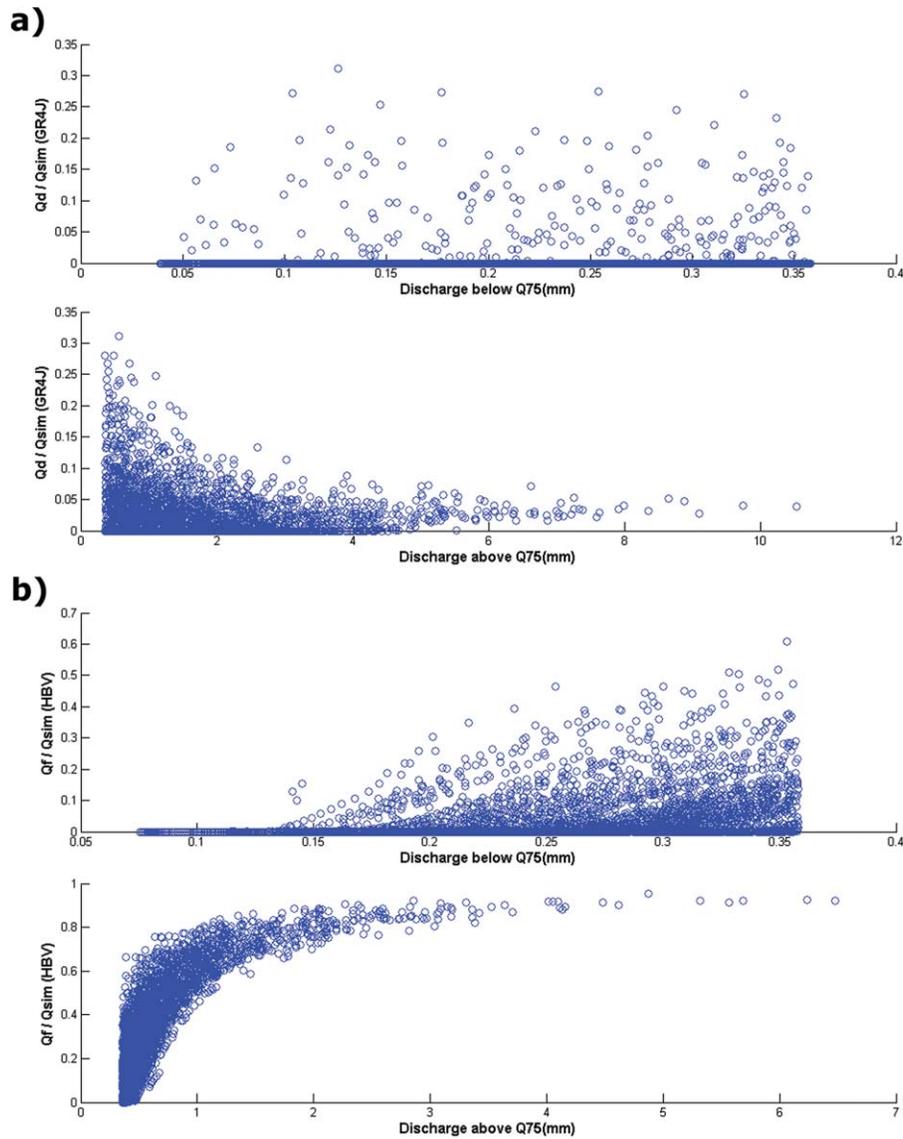


Figure 2. Fraction of fast runoff Q_f or Q_d to total runoff Q_{sim} as a function of simulated discharge, i.e., above and below Q_{75} for the (a) GR4J and (b) HBV models.

discharge shows that the total discharge is dominated by flows from the fast runoff storage during high flows. These categorized values have been used to estimate fast and slow runoff storages in the GR4J and HBV models. The k_{GR4J} fraction is zero for low flows and 0.04 for high flows above 6 mm, whereas the k_{HBV} fraction is about zero for low flows and one for high flows above 6 mm. Table 7 shows the empirical equations fitted to the simulation data presented in Figure 2.

4.2. Effect of Uncertainty on Confidence Intervals of Low Flow Forecasts

[52] For the purpose of determining the extent to which different sources of uncertainty affect low flow forecasts for a lead time of 10 days, the degree of uncertainty in model outputs is expressed by a 90% confidence interval (90CI). The 90CI, the forecast median, and the observed low flows for both models are shown in Figure 3. Daily

Table 7. Empirical Equations to Divide Observed Discharge (Q_{obs}) Into Fast and Slow Runoff

	Q_{obs} Category (mm)		
	$\leq Q_{75}$	$>Q_{75}$ and ≤ 6	>6
k_{GR4J}	0	$-0.00007 \times Q_{obs}^3 + 0.001 \times Q_{obs}^2 + 0.0003 \times Q_{obs} + 0.004$	0.04
k_{HBV}	$4.9 \times Q_{obs}^{3.7} + 0.002$	$0.81 \times e^{0.02 \times Q_{obs}} - 1.4 \times e^{-1.9 \times Q_{obs}}$	1

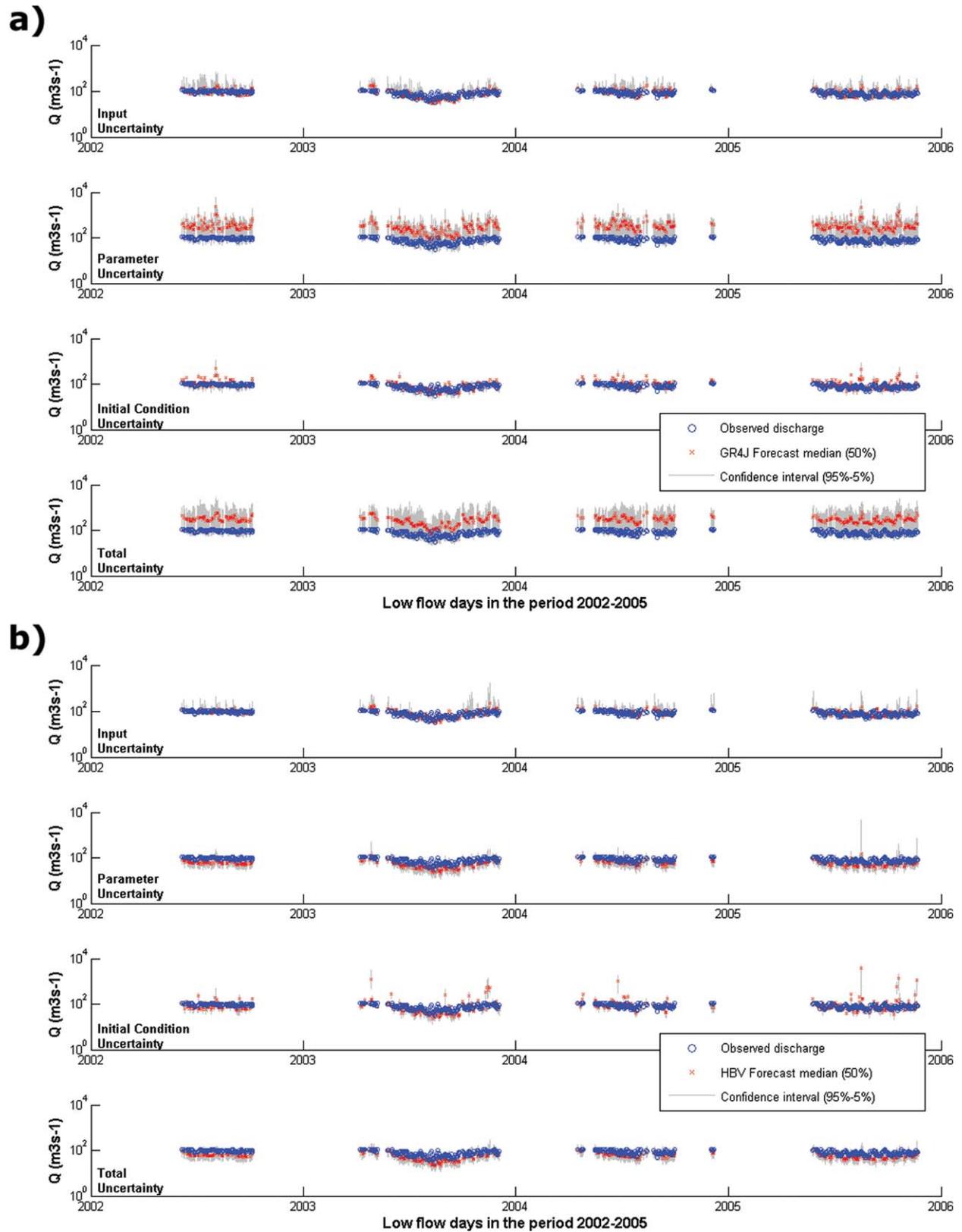


Figure 3. Different uncertainty sources and confidence intervals of low flow forecasts for a lead time of 10 days using the (a) GR4J and (b) HBV models.

discharge values ($\text{m}^3 \text{s}^{-1}$) have been presented on a logarithmic y axis.

[53] There are significant differences between the two model results as 10 day ahead low flows are mostly overestimated by the GR4J model under uncertain conditions. As can be seen from Figure 3 the overestimation is more pronounced for the parameter uncertainty case than for other cases. First thing to be considered are the dependencies and interactions between groundwater storages and model parameters since the fraction of fast runoff to total runoff is about zero showing that the discharge, during low flows, is mainly produced in the groundwater storage (Figure 2). The more pronounced overestimation of GR4J compared for the underestimation of HBV may indicate that the slow responding groundwater storage of the HBV is less sensi-

tive to different behavioral parameter sets. The more complex soil moisture and percolation components of the HBV model can also be a reason for the successful low flow forecasts of the HBV model under uncertain conditions. The systematic overestimation of forecasted precipitation is, therefore, well handled by the HBV model. However, it should be noted that the GR4J model is slightly better than HBV in deterministic 10 day forecasts (Table 6). Further, the low flows are usually underestimated by the HBV model, as shown in the last plot in Figure 3b.

[54] Surprisingly, there have been excessive rainfall forecasts for several days in summer months (e.g., in August 2005) causing high forecasted discharges (Figure 3). These days were carefully examined to determine if they significantly change the overall RCI results. However, they

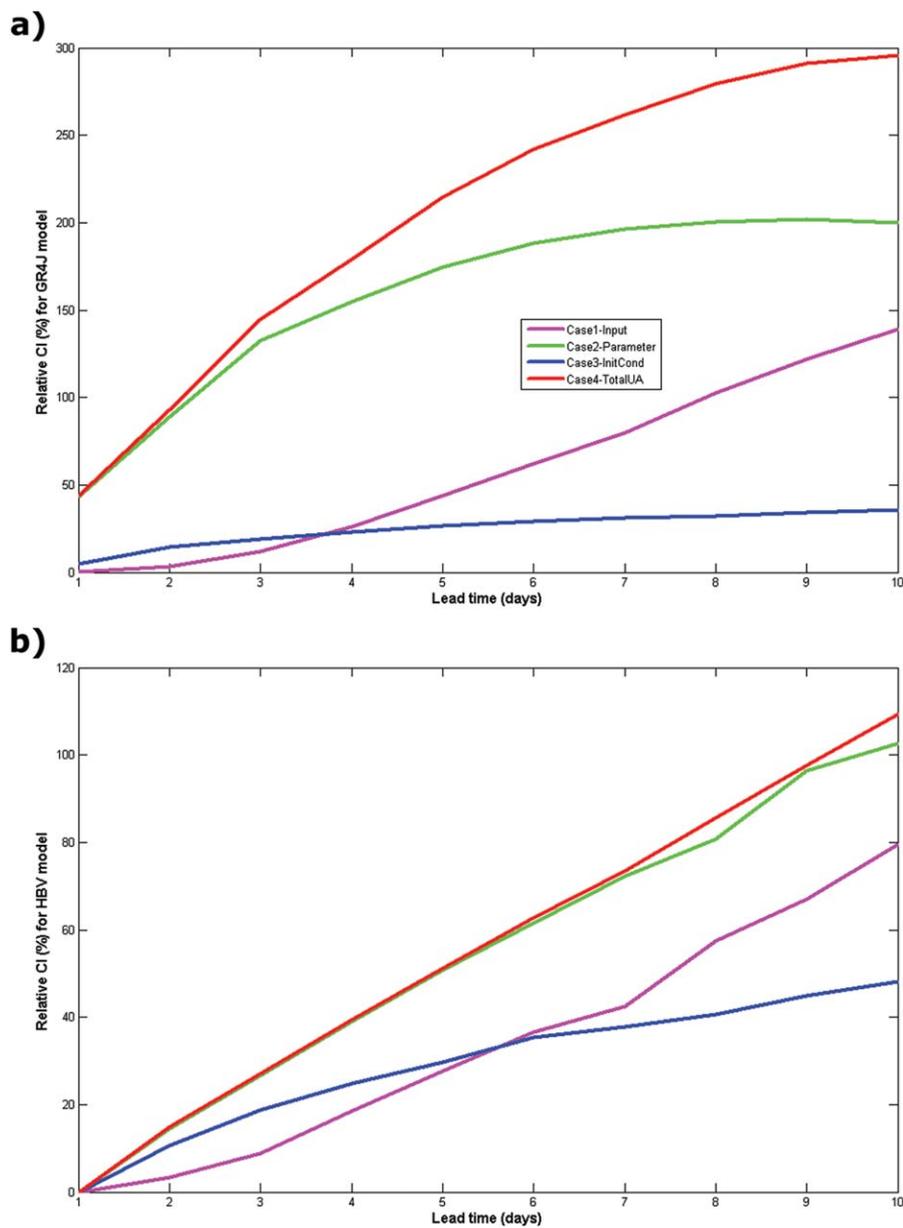


Figure 4. Effect of different uncertainty sources on relative confidence intervals of low flow forecasts with the (a) GR4J and (b) HBV, as a function of lead time.

caused very minor effects on the RCI results that are based on the mean of the confidence interval statistics from 567 low flow days. Thus, the uncertainty in 10 day low flow forecasts is larger in the GR4J model compared to the HBV model. The GR4J model overestimates low flows for all sources of uncertainty and for parameter uncertainty, in particular, whereas the HBV model tends to underestimate low flows.

[55] Figure 4 compares two models and the effect of different uncertainty sources on the RCI of low flow forecasts with increasing lead time. From Figure 4, we can clearly see that the total uncertainty in the GR4J outputs is much higher than in the HBV outputs. This is similar to the results that we have seen in Figure 3. Comparing only 10 day forecasts issued by the two models, the RCI is $\sim 110\%$ for the HBV model and $\sim 300\%$ for the GR4J model (i.e., nearly tripled). One anticipated finding is that the RCI tends to increase with increasing lead time for both models and for all evaluated uncertainty cases. The increase of RCI for the initial condition uncertainty is slowest, showing that the initial condition uncertainty is less sensitive to increasing lead time. This is expected since our storage update procedure only depends on observed discharge and some of the model parameters. However, the uncertainty due to the model inputs (forecast P and PET) increases considerably with increasing lead time. This is from the fact that the error in the en-

semble meteorological forecasts increases for longer lead times due to the atmospheric model limitations. It is interesting to note that the 10 day forecasts are even better using zero precipitation as model input, i.e., NS_{hybrid} for the GR4J results is increased from 0.45 to 0.54 in the test period from 2002 to 2005.

[56] The total uncertainty for the GR4J model is sum of the three sources of uncertainty assessed in this study. Moreover, only half of the RCI comes from uncertain GR4J parameters. The most striking result to emerge from the RCI results is that the parameter uncertainty is dominating the total uncertainty in the HBV model outputs. It is somewhat surprising that nearly all uncertainty comes from the HBV parameters. Parameter interactions in the HBV model can be the main reason for the unexpected total uncertainty which is not the sum of the three sources of uncertainty. Interestingly, the results, as shown in Figure 4, indicate that input uncertainty is even smaller than initial condition uncertainty. On the one hand, this is not expected as ensemble meteorological forecasts are assumed to be one of the most important uncertainty sources in stream-flow forecasts [Engeland et al., 2010; Thirel et al., 2008; Vrugt et al., 2008; Zappa et al., 2011]. On the other hand, the large range of soil moisture related parameters randomly selected from the GLUE behavioral parameter set could enhance the impact of initial condition uncertainty compared to input uncertainty from only 51 P and PET

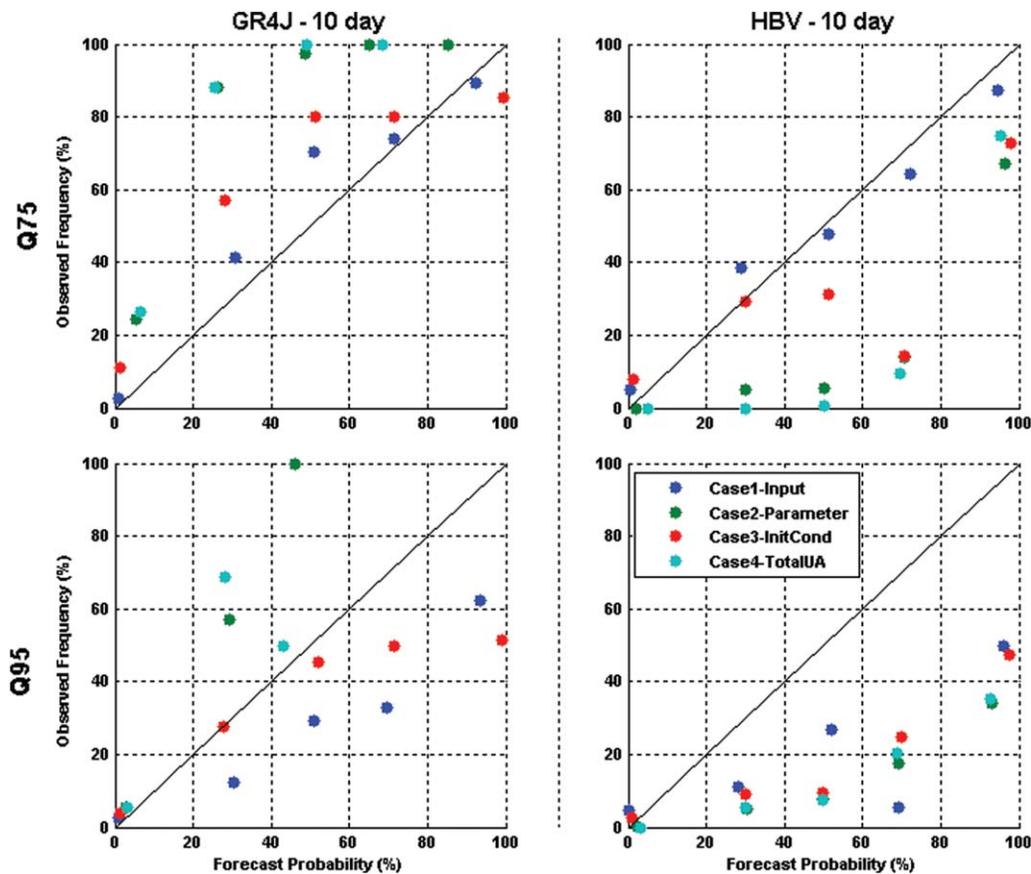


Figure 5. Effect of different uncertainty sources on reliability of 10 day low flow forecasts with GR4J and HBV as a function of two low flow thresholds (i.e., Q_{75} and Q_{95}).

forecast ensemble members. In other words, the dominating effect of parameter uncertainty certainly determines the impact of the initial condition uncertainty due to the parameters used in the storage update procedure. Moreover, during low flow periods slow responding processes like groundwater are more dominant than precipitation as low flows usually occur after prolonged dry periods. This can explain the smaller effect of uncertainty from precipitation compared to the uncertainty from initial conditions. The total uncertainty in 10 day lead time low flow forecasts is not a linear sum of three uncertainty sources in the HBV model due to parameter interactions. This finding is in agreement with the findings of Zappa *et al.* [2011] who showed the full spread obtained from uncertainty superposition of three sources is growing nonlinearly for a hydrological model (i.e., PREVAH (PREcipitation-Runoff-EVApotranspiration)) similar to HBV.

4.3. Effect of Uncertainty on Reliability of Low Flow Forecasts

[57] Figure 5 compares the reliability of 10 day ensemble forecasts of low flows for below Q_{75} and Q_{95} thresholds using the GR4J and HBV models. Figure 5 exhibits the portion of observed data inside predefined forecast intervals. The reliability plots based on forecasts associated with different uncertainties show that GR4J and HBV overestimate or underestimate middle forecast intervals, but the narrowest (i.e., 0%–20%) and the 90% intervals are correctly estimated. From Figure 5 we see clearly that the ensemble Q_{75} forecasts issued by HBV including only the input uncertainty are the most reliable forecasts, and it is confirmed that GR4J provides too wide forecast intervals if all sources of uncertainty are included. The plot for the evaluation of the Q_{75}

forecasts using the HBV model in Figure 5 shows that the average overestimation for the total uncertainty is $\sim 25\%$ inside the highest interval bin (i.e., 80%–100%). The Q_{75} low flow event will then occur in only $\sim 75\%$ of the cases when it is forecasted to almost certainly happen, indicating that every fourth low flow warning will be a false alarm. Initial condition uncertainty has less effect than parameter uncertainty on the reliability of Q_{75} forecasts by the two models. For the Q_{95} low flows, all intervals except for the narrowest interval (i.e., 0%–20%) are overestimated by the HBV model. A comparison of the four subplots in Figure 5 reveals that the parameter uncertainty has a negative effect on the reliability of the forecasts. Moreover, the overestimation of the GR4J model and underestimation of the HBV model are also visible in Figure 5. Finally, the forecasts of extreme low flows (Q_{95}) issued by the GR4J model are more reliable than the forecasts by the HBV model.

4.4. Effect of Uncertainty on Contingency Table of Low Flow Forecasts: Hits and False Alarms

[58] From operational point of view, the main purpose of investigating uncertainty from 10 day ensemble low flow forecasts is to improve the forecasts (e.g., hits) and to reduce false alarms and missed targets in the low flow contingency measures. Figure 6 shows the comparison between the GR4J and HBV models, based on the number of hits, false alarms, misses, and correct rejections for the preselected Q_{75} low flow events. It should be noted that a threshold probability of 0.5 is used to issue a low flow forecast alarm from 10,000 forecasts each day in the test period. Subsequently, a low flow event is assumed to occur if more than half of the 10,000 forecasts are low flows. Subsequently, the $10,000 \times 1461$ forecast ensemble matrix is transformed to a 1×1461

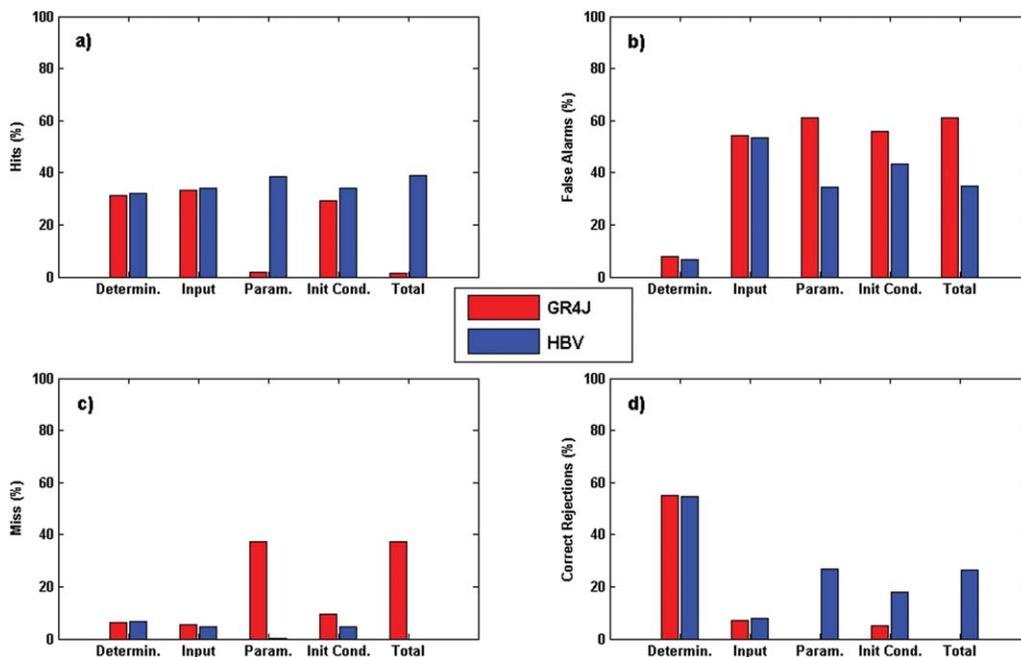


Figure 6. Effect of different uncertainty sources on contingency table. Percentage of (a) hits, (b) false alarms, (c) misses, and (d) correct rejections, to total warnings for 10 day ahead forecasts issued in the test period (i.e., 1461 warnings in the test period).

binary vector consists of zero (no low flow) and one (low flow) values by applying the aforementioned warning threshold of 0.5. This corresponds to a total of 1461 forecasts divided into four subplots in Figure 6. The threshold approach was not necessary for the deterministic run as we run the models only one time with calibrated parameters and control forecasts. The y axes of the subplots show the percentage of different contingency measures for each evaluated uncertainty source, i.e., input, parameter, initial condition, and total. We are aware that the contingency table is very sensitive to the preselected threshold and the number of forecasts [Devineni *et al.*, 2008]. From Figure 6a, we can clearly see that the two models perform similar for the deterministic run, whereas the number of hits declines for the GR4J model forecasts for the total uncertainty. This suggests that adding parameter uncertainty to the model certainly reduces the number of hits. This is what we have seen also in Figure 3a. In case of the HBV model, there is no drop in the number of hits indicating that most of the low flow events (a total of 567 events in the test period) are correctly forewarned by the ensemble forecasts. This is a significant success for a hydrological model calibrated for low flows. Moreover, nonoccurrence of low flow events was also correctly indicated by the HBV model. This could be inferred from the correct rejections plot in Figure 6d.

[59] The most striking result to emerge from Figure 6b is that the percentage of false alarms is highest ($\sim 50\%$) for the forecasts issued by HBV including only input uncer-

tainty. This may seem contradictory with the results presented in Figure 5 as the same ensemble forecasts have been indicated as the most reliable forecasts. However, it should be noted that these two quality measures evaluate the forecasts from totally different aspects, namely, the reliability diagram for the reliability and the contingency table for the sharpness of the forecasts [WMO, 2012].

[60] Figure 7 shows the utility of the low flow forecasts as a function of lead time using the hit rate and the false-alarm rate derived from contingency tables. What is surprising is that the hit rate of GR4J drops significantly from 0.9 to 0.1 by increasing the lead time, whereas the hit rate of HBV is slightly increased from 0.9 to 1. Another important finding was that the hit rate and false-alarm rate of the GR4J and HBV models do not vary significantly as a function of lead time for the deterministic forecasts. Moreover, the false-alarm rate of GR4J does not change considerably by increasing the lead time and for different uncertainty sources. The drop in false-alarm rate is higher for the HBV model. The importance of parameter uncertainty for both models can be clearly seen in Figure 7. The effect of the storage updating procedure and input uncertainty on both models outputs is much smaller.

[61] The findings of the current study (see Figure 7) are consistent with those of Zappa *et al.* [2011] who found slight decreases in the hit rates of low flow forecasts for a lead time of 1 day and 5 days.

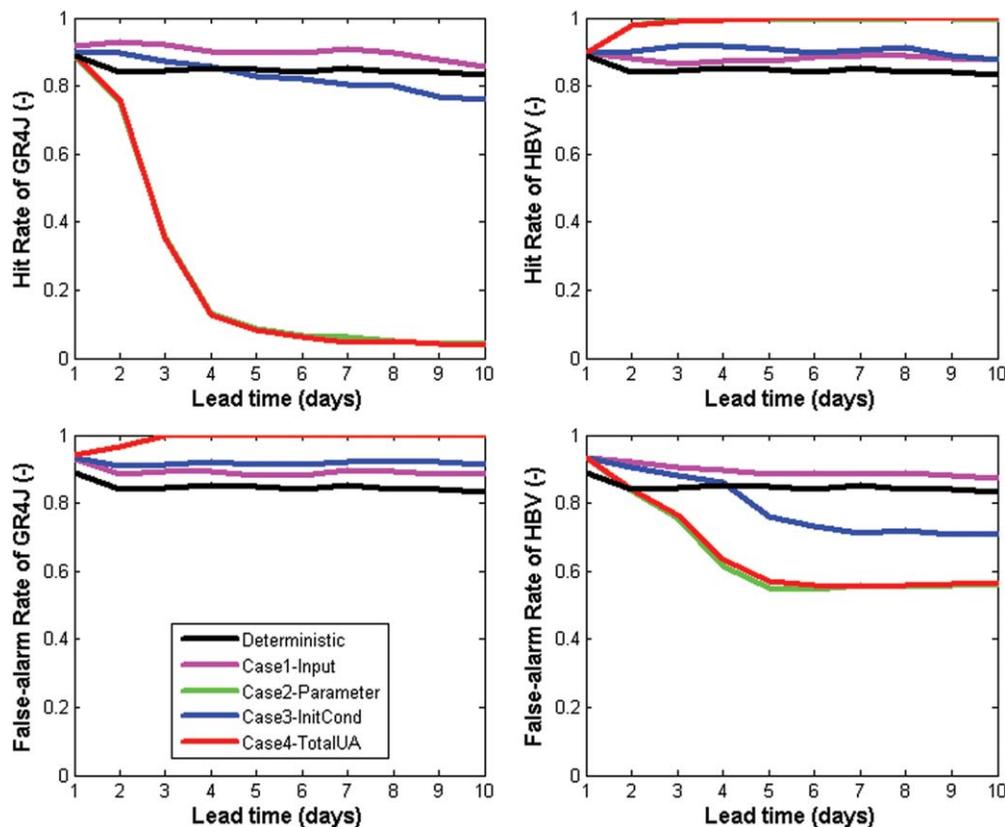


Figure 7. Effect of different uncertainty sources on hit/false-alarm rate of two models as a function of different lead times.

5. Conclusions

[62] The performance of two hydrological models in the calibration, validation, and forecast periods have been compared, and the effect of different uncertainty sources on the quality of 10 day low flow forecasts has been assessed. We applied a systematic uncertainty analysis to identify where the uncertainty comes from and to provide quantified model output uncertainty information to make a robust model comparison. A hybrid performance metric is used for evaluating low flow simulations, whereas the quality of the probabilistic low flow forecasts has been assessed based on relative confidence intervals, reliability, and hit/false-alarm rates. Based on the results presented in this study we can draw the following conclusions.

[63] (1) The 10 day ensemble forecast results show that the daily observed low flows are captured by the 90% confidence interval for both models most of the time. However, the GR4J model usually overestimates low flows, whereas HBV is prone to underestimate low flows. This is particularly the case if the parameter uncertainty is included into the forecasts.

[64] (2) The total uncertainty in the GR4J model outputs is higher than in the HBV model.

[65] (3) The parameter uncertainty has the highest effect, and the input uncertainty has the smallest effect on the low flow forecasts.

[66] (4) A direct relation is found between the number of parameters and the parameter uncertainty according to the RCI results.

[67] (5) The parameter uncertainty for 10 day low flow forecasts issued by the HBV model with eight parameters is almost half of the parameter uncertainty coming from the GR4J with four parameters. This is because the rainfall-runoff process resulting in low flows in the study area is better described by the HBV model.

[68] (6) The forecast distribution based on 10 day low flow forecasts (i.e., Q_{75}) issued by the HBV model was the most reliable forecast distribution if only input uncertainty is considered.

[69] (7) The number of hits is about equal for two models only if the input uncertainty is considered. The parameter uncertainty was the main reason reducing the number of hits.

[70] (8) The deterministic forecasts using the GR4J and HBV resulted in similar performance indices and also similar hit false-alarm rates.

[71] (9) The performance of the HBV model for correct rejections is remarkable, indicating that the model is not only successful for low flows, but also correctly indicates other flows above the Q_{75} threshold while being calibrated on low flows below Q_{75} .

[72] (10) The number of false alarms is almost doubled for the GR4J model considering total uncertainty. The importance of parameter uncertainty on the quality of forecast is emphasized by all forecast quality measures used in this study.

[73] In essence, this paper has shown that the output from two conceptual hydrological models, calibrated for a medium sized $\sim 27.000 \text{ km}^2$ river basin, fed by raw ECMWF meteorological forecasts, is characterized by substantial uncertainty from model parameters. This source of uncertainty effects both the reliability and the sharpness of the

forecasts. This finding is new for low flow forecasts as the significance of the rainfall prediction error is well known and documented for high flows [Pappenberger *et al.*, 2005]. This study has taken a step in the direction of assessing major sources of uncertainties in medium-range low flow forecasts, in addition to flood forecasts for the Moselle River. However, further research has to be conducted on the effect of uncertainties on seasonal low flow forecasts using coarse seasonal weather products. Different types of models, especially data-driven models, may be considered to include in the uncertainty analysis framework for assessing model structure uncertainty explicitly.

[74] **Acknowledgments.** We acknowledge the financial support of the Ir. Cornelis Lely Stichting (CLS), project 20957310. The research is part of the program of the Water Engineering and Management in the University of Twente, and it supports the work of the UNESCO-IHP VII FRIEND-Water program. Discharge data for the River Rhine were provided by the Global Runoff Data Centre (GRDC) in Koblenz (Germany). Areal precipitation and evapotranspiration data were supplied by the Federal Institute of Hydrology (BfG), Koblenz (Germany). REGNIE grid data were extracted from the archive of the Deutscher Wetterdienst (DWD: German Weather Service), Offenbach (Germany). ECMWF-ENS data used in this study have been obtained from the ECMWF data server. We thank Dominique Lucas from ECMWF who kindly guided through the data retrieval process. The GIS base maps with delineated 134 catchments of the Rhine basin were provided by Eric Sprokkereef, the secretary general of the Rhine Commission (CHR). The GR4J and HBV model codes were provided by Tian Ye. We are grateful to the members of the Referat M2-Mitarbeiter/innen group in BfG, Koblenz, in particular, Peter Krahe, Dennis Meißner, Bastian Klein, Robert Pinzinger, Silke Rademacher, and Imke Lingemann for discussions on the value of low flow forecasts, biases in the models, and statistical tests adapted to ensemble forecasts. The constructive review comments of Hamid Moradkhani (Associate Editor), Florian Pappenberger, Thomas E. Adams, and one anonymous reviewer significantly improved this paper.

References

- Ackermann, T., D. Loucks, D. Schwanenberg, and M. Detering (2000), Real-time modeling for navigation and hydropower in the River Mosel, *J. Water Resour. Plann. Manage.*, 126(5), 298–303, doi:10.1061/(ASCE)0733-9496(2000)126:5(298).
- Addor, N., S. Jaun, F. Fundel, and M. Zappa (2011), An operational hydrological ensemble prediction system for the city of Zurich (Switzerland): skill, case studies and scenarios, *Hydrol. Earth Syst. Sci.*, 15(7), 2327–2347, doi:10.5194/hess-15-2327-2011.
- Arnaud, P., J. Lavabre, C. Fouchier, S. Diss, and P. Javelle (2011), Sensitivity of hydrological models to uncertainty in rainfall input, *Hydrol. Sci. J.*, 56(3), 397–410, doi:10.1080/02626667.2011.563742.
- Arnell, N. W. (1999), The effect of climate change on hydrological regimes in Europe: A continental perspective, *Global Environ. Change*, 9(1), 5–23.
- ATV-DVWK (2002), Verdunstung in Bezug zu Landnutzung, Bewuchs und Boden, Merkblatt ATV-DVWK-M Rep. 504, Hennef.
- Behrmann-Godel, J., and R. Eckmann (2003), A preliminary telemetry study of the migration of silver European eel (*Anguilla anguilla* L.) in the River Mosel, Germany, *Ecol. Freshwater Fish*, 12(3), 196–202, doi:10.1034/j.1600-0633.2003.00015.x.
- Belz, J. U., N. Busch, H. Engel, and G. Gasber (1999), Comparative survey of measures on river training at Upper Rhine, Moselle and Saar and their effects on flood behaviour, *Vergleichende Darstellung der Ausbaumassnahmen an Oberrhein, Mosel und Saar und ihre Auswirkungen auf Hochwasser*, 43(6), 283–292.
- Beven, K. J. (2006), On undermining the science?, *Hydrol. Processes*, 20(14), 3141–3146, doi:10.1002/hyp.6396.
- Beven, K. J., and A. Binley (1992), The future of distributed models: Model calibration and uncertainty prediction, *Hydrol. Processes*, 6(3), 279–298.
- Beven, K. J., and J. Freer (2001), Equifinality, data assimilation, and uncertainty estimation in mechanistic modelling of complex environmental systems using the GLUE methodology, *J. Hydrol.*, 249(1–4), 11–29.
- Beven, K. J., and P. Young (2003), Comment on “Bayesian recursive parameter estimation for hydrologic models” by M. Thieman, M. Trosset,

- H. Gupta, and S. Sorooshian, *Water Resour. Res.*, 39(5), 1116, doi:10.1029/2001wr001183.
- Beven, K. J., P. Smith, and J. Freer (2007), Comment on "Hydrological forecasting uncertainty assessment: Incoherence of the GLUE methodology" by Pietro Mantovan and Ezio Todini, *J. Hydrol.*, 338(3–4), 315–318, doi:10.1016/j.jhydrol.2007.02.023.
- Beven, K. J., P. J. Smith, and J. E. Freer (2008), So just why would a modeler choose to be incoherent?, *J. Hydrol.*, 354(1–4), 15–32, doi:10.1016/j.jhydrol.2008.02.007.
- Blasone, R.-S., J. A. Vrugt, H. Madsen, D. Rosbjerg, B. A. Robinson, and G. A. Zyvoloski (2008), Generalized likelihood uncertainty estimation (GLUE) using adaptive Markov Chain Monte Carlo sampling, *Adv. Water Resour.*, 31(4), 630–648, doi:10.1016/j.advwatres.2007.12.003.
- Booij, M. J. (2005), Impact of climate change on river flooding assessed with different spatial model resolutions, *J. Hydrol.*, 303(1–4), 176–198.
- Bröcker, J., and L. A. Smith (2007), Increasing the reliability of reliability diagrams, *Weather Forecasting*, 22(3), 651–661.
- Butts, M. B., J. T. Payne, M. Kristensen, and H. Madsen (2004), An evaluation of the impact of model structure on hydrological modelling uncertainty for streamflow simulation, *J. Hydrol.*, 298(1–4), 242–266.
- Cloke, H. L., and F. Pappenberger (2009), Ensemble flood forecasting: A review, *J. Hydrol.*, 375(3–4), 613–626, doi:10.1016/j.jhydrol.2009.06.005.
- Cunha, L. K., P. V. Mandapaka, W. F. Krajewski, R. Mantilla, and A. A. Bradley (2012), Impact of radar-rainfall error structure on estimated flood magnitude across scales: An investigation based on a parsimonious distributed hydrological model, *Water Resour. Res.*, 48, W10515, doi:10.1029/2012wr012138.
- De Bruijn, K. M., and R. Passchier (2006), Low-flow forecasts for the Rhine, *Rep. Q3427, WL 1 Delft Hydraul.*, Delft, Netherlands.
- De Wit, M. J. M., B. van den Hurk, P. M. M. Warmerdam, P. Torfs, E. Roulin, and W. P. A. van Deursen (2007), Impact of climate change on low-flows in the river Meuse, *Clim. Change*, 82(3), 351–372.
- DeChant, C. M., and H. Moradkhani (2011), Improving the characterization of initial condition for ensemble streamflow prediction using data assimilation, *Hydrol. Earth Syst. Sci.*, 15(11), 3399–3410, doi:10.5194/hess-15-3399-2011.
- Demirel, M. C., M. J. Booij, and A. Y. Hoekstra (2012), Identification of appropriate lags and temporal resolutions for low flow indicators in the River Rhine to forecast low flows with different lead times, *Hydrol. Processes*, doi:10.1002/hyp.9402.
- Devineni, N., A. Sankarasubramanian, and S. Ghosh (2008), Multimodel ensembles of streamflow forecasts: Role of predictor state in developing optimal combinations, *Water Resour. Res.*, 44, W09404, doi:10.1029/2006wr005855.
- Eberle, M. (2005), BFG-1451 hydrological modelling in the River Rhine Basin. Part III: Daily HBV model for the Rhine Basin, *Rep.*, 227 pp., Inst. for Inland Water Manage. and Waste Water Treatment (RIZA), Fed. Inst. of Hydrol. (BfG), Koblenz.
- European Centre for Medium-Range Weather Forecasts (2012), Describing ECMWF's forecasts and forecasting system, in ECMWF Newsl. 133, edited by B. Riddaway.
- Engeland, K., B. Renard, I. Steinsland, and S. Kolberg (2010), Evaluation of statistical models for forecast errors from the HBV model, *J. Hydrol.*, 384(1–2), 142–155.
- Ewen, J., G. O'Donnell, A. Burton, and E. O'Connell (2006), Errors and uncertainty in physically-based rainfall-runoff modelling of catchment change effects, *J. Hydrol.*, 330(3–4), 641–650.
- Franz, K. J., and T. S. Hogue (2011), Evaluating uncertainty estimates in hydrologic models: Borrowing measures from the forecast verification community, *Hydrol. Earth Syst. Sci.*, 15(11), 3367–3382, doi:10.5194/hess-15-3367-2011.
- Freer, J., K. J. Beven, and B. Ambrose (1996), Bayesian estimation of uncertainty in runoff prediction and the value of data: An application of the GLUE approach, *Water Resour. Res.*, 32(7), 2161–2173, doi:10.1029/95wr03723.
- Fundel, F., and M. Zappa (2011), Hydrological ensemble forecasting in mesoscale catchments: Sensitivity to initial conditions and value of reforecasts, *Water Resour. Res.*, 47, W09520, doi:10.1029/2010wr009996.
- Görgen, K., et al. (2010), Assessment of climate change impacts on discharge in the Rhine River Basin: Results of the RheinBlick 2050 project, Lelystad, CHR, 211 pp. [Report Available at http://www.chr-khr.org/files/CHR_I-23.pdf].
- Göttinger, J., and A. Bárdossy (2008), Generic error model for calibration and uncertainty estimation of hydrological models, *Water Resour. Res.*, 44, W00B07, doi:10.1029/2007wr006691.
- Grabs, W., K. Daamen, and F. de Montmollin (1997), *Impact of Climate Change on Hydrological Regimes and Water Resources Management in the Rhine Basin*, CHR/KHR: Lelystad.
- Gupta, H. V., M. P. Clark, J. A. Vrugt, G. Abramowitz, and M. Ye (2012), Towards a comprehensive assessment of model structural adequacy, *Water Resour. Res.*, 48, W08301, doi:10.1029/2011wr011044.
- Hagemann, S., H. Göttel, D. Jacob, P. Lorenz, and E. Roeckner (2008), Improved regional scale processes reflected in projected hydrological changes over large European catchments, *Clim. Dyn.*, 32, 767–781.
- Hurkmans, R., W. Terink, R. Uijlenhoet, P. Torfs, D. Jacob, and P. A. Troch (2010), Changes in streamflow dynamics in the Rhine basin under three high-resolution regional climate scenarios, *J. Clim.*, 23(3), 679–699.
- Janssen, P. H. M., A. C. Petersen, J. P. van der Sluijs, J. S. Risbey, J. R. Ravetz, and K. von Krauss (2005), A guidance for assessing and communicating uncertainties, *Water Sci. Technol.*, 52(6), 125–134.
- Jaun, S., and B. Ahrens (2009), Evaluation of a probabilistic hydrometeorological forecast system, *Hydrol. Earth Syst. Sci.*, 13(7), 1031–1043.
- Jin, X., C.-Y. Xu, Q. Zhang, and V. P. Singh (2010), Parameter and modeling uncertainty simulated by GLUE and a formal Bayesian method for a conceptual hydrological model, *J. Hydrol.*, 383(3–4), 147–155, doi:10.1016/j.jhydrol.2009.12.028.
- Komma, J., C. Reszler, G. Blöschl, and T. Haiden (2007), Ensemble prediction of floods—Catchment non-linearity and forecast probabilities, *Nat. Hazards Earth Syst. Sci.*, 7(4), 431–444, doi:10.5194/nhess-7-431-2007.
- Lawal, S., W. Watt, and D. Watts (1997), A stochastic model of low flows, *Stochastic Hydrol. Hydraul.*, 11(4), 303–321.
- Li, H., L. Luo, and E. F. Wood (2008), Seasonal hydrologic predictions of low-flow conditions over eastern USA during the 2007 drought, *Atmos. Sci. Lett.*, 9(2), 61–66.
- Li, H., L. Luo, E. F. Wood, and J. Schaake (2009), The role of initial conditions and forcing uncertainties in seasonal hydrologic forecasting, *J. Geophys. Res.*, 114, D04114, doi:10.1029/2008jd010969.
- Li, L., J. Xia, C.-Y. Xu, and V. P. Singh (2010), Evaluation of the subjective factors of the GLUE method and comparison with the formal Bayesian method in uncertainty assessment of hydrological models, *J. Hydrol.*, 390(3–4), 210–221.
- Lindström, G., B. Johansson, M. Persson, M. Gardelin, and S. Bergström (1997), Development and test of the distributed HBV-96 hydrological model, *J. Hydrol.*, 201(1–4), 272–288.
- Liu, Y., et al. (2012), Advancing data assimilation in operational hydrologic forecasting: Progresses, challenges, and emerging opportunities, *Hydrol. Earth Syst. Sci.*, 16(10), 3863–3887, doi:10.5194/hess-16-3863-2012.
- Mantovan, P., and E. Todini (2006), Hydrological forecasting uncertainty assessment: Incoherence of the GLUE methodology, *J. Hydrol.*, 330(1–2), 368–381, doi:10.1016/j.jhydrol.2006.04.046.
- Martina, M. L. V., E. Todini, and A. Libralon (2005), A Bayesian decision approach to rainfall thresholds based flood warning, *Hydrol. Earth Syst. Sci. Discuss.*, 2(6), 2663–2706, doi:10.5194/hessd-2-2663-2005.
- McKay, M. D., R. J. Beckman, and W. J. Conover (1979), A comparison of three methods for selecting values of input variables in the analysis of output from a computer code, *Technometrics*, 21(2), 239–245.
- Meißner, D., S. Gebauer, A. H. Schumann, M. Pahlow, and S. Rademacher (2012), Analysis of radar-based precipitation data products as input data to improve water-level forecasting for navigation on the River Rhine, *Hydrologie und Wasserbewirtschaftung*, 56(1), 16–28.
- Middelkoop, H., K. Daamen, D. Gellens, W. Grabs, J. C. J. Kwadijk, H. Lang, B. Parmet, B. Schadler, J. Schulla, and K. Wilke (2001), Impact of climate change on hydrological regimes and water resources management in the Rhine Basin, *Clim. Change*, 49(1–2), 105–128.
- Montanari, A. (2005), Large sample behaviors of the generalized likelihood uncertainty estimation (GLUE) in assessing the uncertainty of rainfall-runoff simulations, *Water Resour. Res.*, 41, W08406, doi:10.1029/2004wr003826.
- Montanari, A. (2011), Uncertainty of hydrological predictions, in *Treatise on Water Science*, edited by P. Wilderer, pp. 459–478, Academic, Oxford.
- Montanari, A., and G. Grossi (2008), Estimating the uncertainty of hydrological forecasts: A statistical approach, *Water Resour. Res.*, 44, W00B08, doi:10.1029/2008wr006897.
- Moradkhani, H., C. M. DeChant, and S. Sorooshian (2012), Evolution of ensemble data assimilation for uncertainty quantification using the particle filter-Markov chain Monte Carlo method, *Water Resour. Res.*, 48, W12520, doi:10.1029/2012wr012144.

- Muluye, G. Y. (2011), Implications of medium-range numerical weather model output in hydrologic applications: Assessment of skill and economic value, *J. Hydrol.*, *400*(3–4), 448–464.
- Nash, J. E., and J. V. Sutcliffe (1970), River flow forecasting through conceptual models. Part I: A discussion of principles, *J. Hydrol.*, *10*(3), 282–290.
- Nester, T., J. Komma, A. Viglione, and G. Blöschl (2012), Flood forecast errors and ensemble spread—A case study, *Water Resour. Res.*, *48*, W10502, doi:10.1029/2011wr011649.
- Olsson, J., and G. Lindström (2008), Evaluation and calibration of operational hydrological ensemble forecasts in Sweden, *J. Hydrol.*, *350*(1–2), 14–24.
- Pagano, T. C., D. L. Shrestha, Q. J. Wang, D. Robertson, and P. Hapuarachi (2012), Ensemble dressing for hydrological applications, *Hydrol. Processes*, *27*, 106–116, doi:10.1002/hyp.9313.
- Pappenberger, F., K. J. Beven, A. De Roo, J. Thielen, and B. Gouweleeuw (2004), Uncertainty analysis of the rainfall runoff model LisFlood within the generalized likelihood uncertainty estimation (GLUE), *Int. J. River Basin Manage.*, *2*(2), 123–133.
- Pappenberger, F., K. J. Beven, N. M. Hunter, P. D. Bates, B. T. Gouweleeuw, J. Thielen, and A. P. J. de Roo (2005), Cascading model uncertainty from medium range weather forecasts (10 days) through a rainfall-runoff model to flood inundation predictions within the European Flood Forecasting System (EFFS), *Hydrol. Earth Syst. Sci.*, *9*(4), 381–393, doi:10.5194/hess-9-381-2005.
- Pappenberger, F., J. Thielen, and M. Del Medico (2011), The impact of weather forecast improvements on large scale hydrology: Analysing a decade of forecasts of the European Flood Alert System, *Hydrol. Processes*, *25*(7), 1091–1113, doi:10.1002/hyp.7772.
- Parrish, M. A., H. Moradkhani, and C. M. DeChant (2012), Toward reduction of model uncertainty: Integration of Bayesian model averaging and data assimilation, *Water Resour. Res.*, *48*, W03519, doi:10.1029/2011wr011116.
- Pasetto, D., M. Camporese, and M. Putti (2012), Ensemble Kalman filter versus particle filter for a physically-based coupled surface–subsurface model, *Adv. Water Resour.*, *47*, 1–13, doi:10.1016/j.advwatres.2012.06.009.
- Perrin, C., C. Michel, and V. Andréassian (2003), Improvement of a parsimonious model for streamflow simulation, *J. Hydrol.*, *279*(1–4), 275–289.
- Perrin, C., L. Oudin, V. Andréassian, C. Rojas-Serna, C. Michel, and T. Mathevet (2007), Impact of limited streamflow data on the efficiency and the parameters of rainfall-runoff models, *Hydrol. Sci. J.*, *52*(1), 131–151, doi:10.1623/hysj.52.1.131.
- Pushpalatha, R., C. Perrin, N. L. Moine, T. Mathevet, and V. Andréassian (2011), A downward structural sensitivity analysis of hydrological models to improve low-flow simulation, *J. Hydrol.*, *411*(1–2), 66–76.
- Pushpalatha, R., C. Perrin, N. L. Moine, and V. Andréassian (2012), A review of efficiency criteria suitable for evaluating low-flow simulations, *J. Hydrol.*, *420–421*, 171–182, doi:10.1016/j.jhydrol.2011.11.055.
- Randrianasolo, A., M. H. Ramos, G. Thirel, V. Andréassian, and E. Martin (2010), Comparing the scores of hydrological ensemble forecasts issued by two different hydrological models, *Atmos. Sci. Lett.*, *11*(2), 100–107, doi:10.1002/asl.259.
- Ratto, M., P. C. Young, R. Romanowicz, F. Pappenberger, A. Saltelli, and A. Pagano (2007), Uncertainty, sensitivity analysis and the role of data based mechanistic modeling in hydrology, *Hydrol. Earth Syst. Sci.*, *11*(4), 1249–1266, doi:10.5194/hess-11-1249-2007.
- Refsgaard, J. C., J. P. van der Sluijs, J. Brown, and P. van der Keur (2006), A framework for dealing with uncertainty due to model structure error, *Adv. Water Resour.*, *29*(11), 1586–1597.
- Refsgaard, J. C., J. P. van der Sluijs, A. L. Højberg, and P. A. Vanrolleghem (2007), Uncertainty in the environmental modelling process—A framework and guidance, *Environ. Modell. Software*, *22*(11), 1543–1556.
- Reggiani, P., and A. H. Weerts (2008a), Probabilistic quantitative precipitation forecast for flood prediction: An application, *J. Hydrometeorol.*, *9*, 76–95.
- Reggiani, P., and A. H. Weerts (2008b), A Bayesian approach to decision-making under uncertainty: An application to real-time forecasting in the river Rhine, *J. Hydrol.*, *356*(1–2), 56–69, doi:10.1016/j.jhydrol.2008.03.027.
- Reggiani, P., M. Renner, A. H. Weerts, and P. van Gelder (2009), Uncertainty assessment via Bayesian revision of ensemble streamflow predictions in the operational river Rhine forecasting system, *Water Resour. Res.*, *45*, W02428, doi:10.1029/2007WR006758.
- Renard, B., D. Kavetski, G. Kuczera, M. Thyer, and S. W. Franks (2010), Understanding predictive uncertainty in hydrologic modeling: The challenge of identifying input and structural errors, *Water Resour. Res.*, *46*, W05521, doi:10.1029/2009wr008328.
- Renner, M., M. G. F. Werner, S. Rademacher, and E. Sprokkereef (2009), Verification of ensemble flow forecasts for the River Rhine, *J. Hydrol.*, *376*(3–4), 463–475.
- Rossa, A., K. Liechti, M. Zappa, M. Bruen, U. Germann, G. Haase, C. Keil, and P. Krahe (2011), The COST 731 Action: A review on uncertainty propagation in advanced hydro-meteorological forecast systems, *Atmos. Res.*, *100*(2–3), 150–167.
- Roulin, E. (2007), Skill and relative economic value of medium-range hydrological ensemble predictions, *Hydrol. Earth Syst. Sci.*, *11*(2), 725–737.
- Roulin, E., and S. Vannitsem (2005), Skill of medium-range hydrological ensemble predictions, *J. Hydrometeorol.*, *6*(5), 729–744.
- Rutten, M., N. van de Giesen, M. Baptist, J. Icke, and W. Uijttewaal (2008), Seasonal forecast of cooling water problems in the River Rhine, *Hydrol. Processes*, *22*(7), 1037–1045.
- Salamon, P., and L. Feyen (2009), Assessing parameter, precipitation, and predictive uncertainty in a distributed hydrological model using sequential data assimilation with the particle filter, *J. Hydrol.*, *376*(3–4), 428–442, doi:10.1016/j.jhydrol.2009.07.051.
- Shen, Z. Y., L. Chen, and T. Chen (2012), Analysis of parameter uncertainty in hydrological and sediment modeling using GLUE method: A case study of SWAT model applied to Three Gorges Reservoir Region, China, *Hydrol. Earth Syst. Sci.*, *16*(1), 121–132, doi:10.5194/hess-16-121-2012.
- Shukla, S., and D. P. Lettenmaier (2011), Seasonal hydrologic prediction in the United States: Understanding the role of initial hydrologic conditions and seasonal climate forecast skill, *Hydrol. Earth Syst. Sci.*, *15*(11), 3529–3538, doi:10.5194/hess-15-3529-2011.
- Spear, R. C., and G. M. Hornberger (1980), Eutrophication in peel inlet: II. Identification of critical uncertainties via generalized sensitivity analysis, *Water Res.*, *14*(1), 43–49, doi:10.1016/0043-1354(80)90040-8.
- Stedinger, J. R., R. M. Vogel, S. U. Lee, and R. Batchelder (2008), Appraisal of the generalized likelihood uncertainty estimation (GLUE) method, *Water Resour. Res.*, *44*, W00B06, doi:10.1029/2008wr006822.
- Stravs, L., and M. Brilly (2007), Development of a low-flow forecasting model using the M5 machine learning method, *Hydrol. Sci. J.*, *52*(3), 466–477.
- Svensson, C., and C. Prudhomme (2005), Prediction of British summer river flows using winter predictors, *Theor. Appl. Climatol.*, *82*(1), 1–15.
- Tallaksen, L., and H. A. J. Van Lanen (2004), *Hydrological Drought: Processes and Estimation Methods for Streamflow and Groundwater*, 579 pp., Elsevier, Amsterdam, Netherlands.
- Te Linde, A. H., J. C. J. H. Aerts, A. M. R. Bakker, and J. C. J. Kwadijk (2010), Simulating low-probability peak discharges for the Rhine basin using resampled climate modeling data, *Water Resour. Res.*, *46*, W03512, doi:10.1029/2009wr007707.
- Te Linde, A. H., P. Bubeck, J. E. C. Dekkers, H. de Moel, and J. C. J. H. Aerts (2011), Future flood risk estimates along the river Rhine, *Nat. Hazards Earth Syst. Sci.*, *11*(2), 459–473, doi:10.5194/nhess-11-459-2011.
- Thirel, G., F. Rousset-Regimbeau, E. Martin, and F. Habets (2008), On the impact of short-range meteorological forecasts for ensemble streamflow predictions, *J. Hydrometeorol.*, *9*(6), 1301–1317, doi:10.1175/2008jhm959.1.
- Thirel, G., F. Regimbeau, E. Martin, J. Noilhan, and F. Habets (2010), Short- and medium-range hydrological ensemble forecasts over France, *Atmos. Sci. Lett.*, doi:10.1002/asl.254.
- Thyer, M., B. Renard, D. Kavetski, G. Kuczera, S. W. Franks, and S. Srikanthan (2009), Critical evaluation of parameter consistency and predictive uncertainty in hydrological modeling: A case study using Bayesian total error analysis, *Water Resour. Res.*, *45*, W00B14, doi:10.1029/2008wr006825.
- Tian, Y., M. J. Booij, and Y. P. Xu (2012), Uncertainty in high and low flows due to model structure and parameter errors, *Stochastic Environmental Research and Risk Assessment*, doi:10.1007/s00477-013-0751-9.
- Tolson, B. A., and C. A. Shoemaker (2008), Efficient prediction uncertainty approximation in the calibration of environmental simulation models, *Water Resour. Res.*, *44*, W04411, doi:10.1029/2007wr005869.
- Van Andel, S. J., A. Weerts, J. Schaake, and K. Bogner (2012), Post-processing hydrological ensemble predictions intercomparison experiment, *Hydrol. Processes*, *27*, 158–161, doi:10.1002/hyp.9595.
- Van den Tillaart, S. P. M., M. J. Booij, and M. S. Krol (2012), Impact of uncertainties in discharge determination on the parameter estimation and

- performance of a hydrological model, *Hydrology Research*, 44(3), 454–466, doi: 10.2166/nh.2012.147.
- Van der Sluijs, J., M. Craye, S. Funtowicz, P. Kloprogge, J. Ravetz, and J. Risbey (2005), Experiences with the NUSAP system for multidimensional uncertainty assessment, *Water Sci. Technol.: J. Int. Assoc. Water Pollut. Res.*, 52(6), 133–144.
- Vázquez, R. F., K. J. Beven, and J. Feyen (2009), GLUE based assessment on the overall predictions of a MIKE SHE application, *Water Resour. Manage.*, 23, 1325–1349, doi:10.1007/s11269-008-9329-6.
- Velázquez, J. A., F. Anctil, and C. Perrin (2010), Performance and reliability of multimodel hydrological ensemble simulations based on seventeen lumped models and a thousand catchments, *Hydrol. Earth Syst. Sci.*, 14(11), 2303–2317, doi:10.5194/hess-14-2303-2010.
- Viola, F., L. V. Noto, M. Cannarozzo, and G. La Loggia (2009), Daily streamflow prediction with uncertainty in ephemeral catchments using the GLUE methodology, *Phys. Chem. Earth Parts A/B/C*, 34(10–12), 701–706.
- Vrugt, J. A., C. J. F. t. Braak, M. P. C., J. M. Hyman, and B. A. Robinson (2008), Treatment of input uncertainty in hydrologic modeling: Doing hydrology backward with Markov chain Monte Carlo simulation, *Water Resour. Res.*, 44, W00B09, doi:10.1029/2007WR006720.
- Walker, W. E., P. Harremoës, J. Rotmans, J. P. van der Sluijs, M. B. A. van Asselt, P. Janssen, and M. P. K. von Krauss (2003), Defining uncertainty: A conceptual basis for uncertainty management in model-based decision support, *Integr. Assess.*, 4(1), 5–17.
- Warmink, J. J., J. A. E. B. Janssen, M. J. Booij, and M. S. Krol (2010), Identification and classification of uncertainties in the application of environmental models, *Environ. Modell. Software*, 25(12), 1518–1527.
- Warmink, J. J., H. Klis, M. J. Booij, and S. J. M. H. Hulscher (2011), Identification and quantification of uncertainties in a hydrodynamic river model using expert opinions, *Water Resour. Manage.*, 25(2), 601–622, doi:10.1007/s11269-010-9716-7.
- World Meteorological Organization (2012), Forecast Verification—Issues, Methods and FAQ, WWRP/WGNE Joint Working Group on Verification. [Available at <http://www.cawcr.gov.au/projects/verification/>.]
- Wöhling, T., F. Lennartz, and M. Zappa (2006), Technical note: Updating procedure for flood forecasting with conceptual HBV-type models, *Hydrol. Earth Syst. Sci.*, 10(6), 783–788, doi:10.5194/hess-10-783-2006.
- Wood, A. W., and D. P. Lettenmaier (2008), An ensemble approach for attribution of hydrologic prediction uncertainty, *Geophys. Res. Lett.*, 35, L14401, doi:10.1029/2008gl034648.
- Zappa, M., S. Jaun, U. Germann, A. Walser, and F. Fundel (2011), Superposition of three sources of uncertainties in operational flood forecasting chains, *Atmos. Res.*, 100(2–3), 246–262, doi:10.1016/j.atmosres.2010.12.005.
- Zheng, Y., and A. A. Keller (2007), Uncertainty assessment in watershed-scale water quality modeling and management: 1. Framework and application of generalized likelihood uncertainty estimation (GLUE) approach, *Water Resour. Res.*, 43, W08407, doi:10.1029/2006wr005345.