HYDROLOGICAL PROCESSES
Hydrol. Process. **29**, 1434–1449 (2015)
Published online 25 July 2014 in Wiley Online Library
(wileyonlinelibrary.com) DOI: 10.1002/hyp.10259

Post-processing of medium-range probabilistic hydrological forecasting: impact of forcing, initial conditions and model errors

Emmanuel Roulin* and Stéphane Vannitsem
Institut Royal Météorologique de Belgique, Brussels, Belgium

Abstract:

The impact of errors in the forcing, errors in the model structure and parameters, and errors in the initial conditions is investigated in a simple hydrological ensemble prediction system. The hydrological model is based on an input nonlinearity connected with a linear transfer function and forced by precipitation forecasts from the European Centre for Medium-Range Weather Forecast (ECMWF) Ensemble Prediction System (EPS). The post-processing of the precipitation and/or the streamflow using information from the reforecasts performed by ECMWF is tested. For this purpose, hydrological reforecasts are obtained by forcing the hydrological model with the precipitation from the reforecast data.

In the present case study, it is found that the post-processing of the hydrological ensembles with a statistical model fitted on the hydrological reforecasts improves the verification scores better than the use of post-processed precipitation ensembles. In the case of large biases in the precipitation, combining the post-processing of both precipitation and streamflow allows for further improvements. During the winter, errors in the initial conditions have a larger impact on the scores than errors in the model structure as designed in the experiments. Errors in the parameter values are largely corrected with the post-processing. Copyright © 2014 John Wiley & Sons, Ltd.

KEY WORDS hydrological ensemble predictions; post-processing; model errors; predictability

Received 13 March 2014; Accepted 27 May 2014

INTRODUCTION

Three main types of errors affect hydrological forecasts: errors as a result of modelling (parameters and structure), errors as a result of initial conditions, and errors as a result of external forcings like precipitations.

Calibrating a hydrological model usually consists in optimizing an objective function that compares model output and observations. Methods have been developed that find a global optimum in a robust and efficient way, e.g. the shuffled complex evolution (Duan *et al.*, 1992). Uncertainties in the parameters may be assessed in a Bayesian framework using the generalized likelihood uncertainty estimation (refer to, e.g. Beven, 2001) or the shuffled complex evolution Metropolis (Vrugt *et al.*, 2003). Implicitly assuming that all types of errors are included in parameter uncertainty results in biases in parameter values and underestimation of the confidence interval of predictions (Ajami *et al.*, 2007).

The structure of hydrological models should be supported by the observations. In the Data-Based Mechanistic approach (DBM, Young, 1998, 2012), different transfer functions are compared on the basis of several criteria, and an efficient and parsimonious model structure is selected if it has physical sense. Uncertainty as a result of the model structure has been addressed by combining the output of multiple models (e.g. Georgakakos *et al.*, 2004). In the Bayesian Model Averaging (BMA) method, each model is weighted according to the likelihood that it is the correct model (Ajami *et al.*, 2007).

Uncertainties in initial conditions are dealt with by data assimilation that aims at updating model state variables or parameters using observations. Data assimilation is used in hydrological modelling and forecasting with a variety of methods and observations (refer to a review by Liu *et al.*, 2012). For instance, DBM methodology is suited to incorporate a recursive filter and an adaptive gain for real-time flood forecasting (Beven *et al.*, 2012). The observed streamflow has been assimilated to update the soil moisture using variational assimilation (Seo *et al.*, 2003) and ensemble Kalman filter (Komma *et al.*, 2008). Assimilation of soil moisture products from remote sensing in rainfall-runoff models has been shown to improve the modelling performances (e.g. Brocca *et al.*, 2012).

When forecasting the streamflow beyond the response time of the catchment, uncertainties about the quantitative precipitation forecast are increasingly outweighing other

*Correspondence to: Emmanuel Roulin, Institut Royal Météorologique de Belgique, Avenue Circulaire, 3, B-1180, Brussels, Belgium.
E-mail: emmanuel.roulin@meteo.be

Copyright © 2014 John Wiley & Sons, Ltd.

sources of uncertainties. For seasonal forecasts, Ensemble Streamflow Predictions (ESPs) are obtained by using historical weather data to simulate an ensemble of streamflow traces conditioned on the current hydroclimatic state (Hashino *et al.*, 2007). For several years, forecast centres issue ensemble predictions by integrating a Numerical Weather Prediction (NWP) model from different initial conditions and with perturbed physics in order to quantify uncertainties about the forecasts (e.g. Molteni *et al.*, 1996). It has become widespread to use these ensembles as input to hydrological models (e.g. Roulin and Vannitsem, 2005 and the review by Cloke and Pappenberger, 2009). However, meteorological forecasts suffer from the same types of errors as mentioned above (model and initial conditions), and ensembles are biased and generally underdispersed. Post-processing techniques have been developed to correct model outputs on the basis of past forecasts. The following methods have been applied for precipitation: logistic regression (Applequist *et al.*, 2002; Hamill and Whitaker, 2006), Extended Logistic Regression (ELR; Wilks, 2009; Schmeits and Kok, 2010; Roulin and Vannitsem, 2012), BMA (Sloughter *et al.*, 2007; Schmeits and Kok, 2010) and quantile regression (Bremnes, 2004; Friederichs and Hense, 2008).

Operational NWP models are frequently modified. It was therefore proposed to use state-of-the-art actual models to reforecast past situations in order to provide samples covering enough different meteorological situations to build robust statistical models for post-processing subsequent forecasts (Hamill *et al.*, 2004). Since March 2008, a reforecast data set is available at the ECMWF that is produced with the same model system used to perform the actual forecasts (Hagedorn, 2008). This data set is updated every week with five member reforecasts of the last 18 years at the same date (since June 2012, the last 20 years). Roulin and Vannitsem (2012) used this data set in various settings for the post-processing of precipitation predictions. Improvement in the Continuous Ranked Probability Score (CRPS) is found for a window of 5 weeks that is compatible with an operational setting.

Biases in hydrological predictions can be corrected on the base of reference simulations and the corresponding observations. In the context of ESP, Hashino *et al.* (2007) have compared three post-processing methods – namely, the event bias correction, the regression, and the quantile mapping to adjust each model-simulated trace independently. Pagano *et al.* (2012) have proposed an ensemble dressing method also based on the difference between historical simulations and observations. Wood and Schaake (2008) used the correlation of forecast ensemble means with observations to generate a conditional forecast mean and spread. The last two approaches do not preserve ensemble traces. Madadgar *et al.* (2012) show the improvement of the post-processing based on a copula function over the quantile mapping technique. Other nonparametric (Van Steenbergen *et al.*, 2012; Brown and Seo, 2013) and parametric techniques (Reggiani and Weerts, 2008; Engeland and Steinsland, 2014) have been recently proposed. The use of post-processed precipitation for ensemble streamflow forecasts has been tested (Zalachori *et al.*, 2012; Verkade *et al.*, 2013); the results show that the errors linked to hydrological modelling remain a key component to the total predictive uncertainty. In this context, post-processing is recognized as an important step to achieve forecasts that are unbiased and reliable and have the highest skill and accuracy possible and that are coherent with the climatology (van Andel *et al.*, 2013).

As discussed above, the methodologies used to tackle the different sources of uncertainties are numerous and interconnected, but there is no clear picture of the impact of post-processing of either the precipitation forcings or the hydrological forecasts in the presence of the different potential sources of errors. The present study aims at clarifying this impact. This will be tackled by introducing errors in the model structure as the difference between the outputs of a first model taken as the 'truth' and of a second simpler model. Sensitivity to errors in the model parameter and in the initial conditions is also analysed. Forcing uncertainties are taken from operational ensemble predictions. The section on Methodology presents the hydrological ensemble prediction system components – the hydrological model and the ensemble predictions of precipitation, the post-processing methods applied, the scores used for verification, and the experiments performed. The results of these experiments are described in the Results Section, and conclusions are then drawn.

METHODOLOGY

Test basin

The test basin is the Ourthe Orientale at Mabompré in the Ardennes region in Belgium (area 317 km^2, elevation 280–650 m, and mean annual rainfall 1029 mm, Figure 1). Data from October 1990 to September 1995 are used for the model development (Figure 2). Discharge data were measured at the stream gauge of Mabompré. Precipitation data consist of daily pluviometric observations averaged over the catchment using the Thiessen method, and temperature data were measured at the Nadrin station at the edge of this basin.

The hydrological model

In this study two hydrological models are used. Their structures have been identified on observed precipitation, temperature and discharge time series following the DBM

Copyright © 2014 John Wiley & Sons, Ltd.

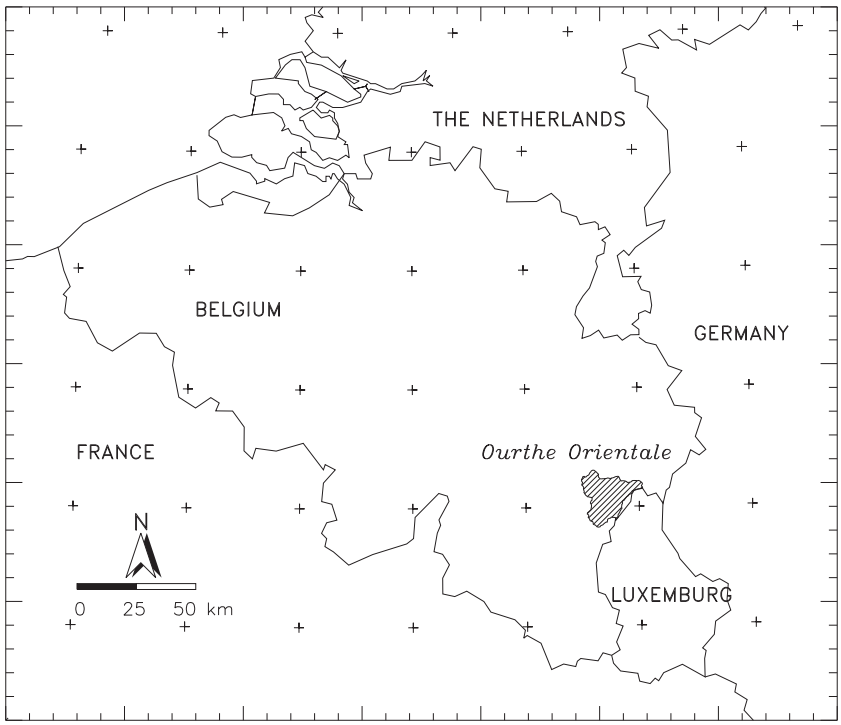

Figure 1. Location map of the Ourthe Orientale catchment in Belgium. Grid points (+) of EPS archives corresponding to the T399 horizontal resolution in use at ECMWF from 2006 to 2010

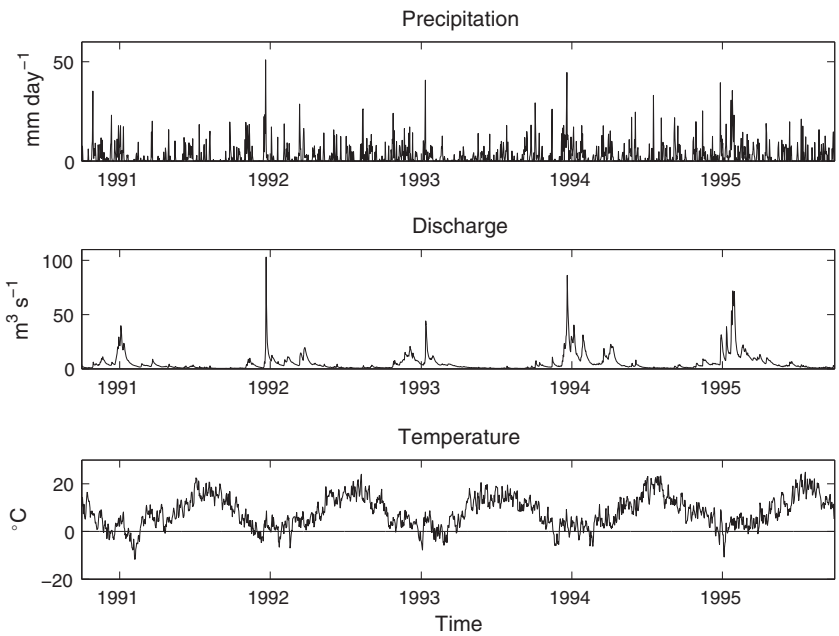

Figure 2. Time series of areal precipitation (mm day^{-1}), streamflow (m^3 s^{-1}), and temperature (°C) from October 1990 to September 1995, Ourthe Orientale catchment

approach (Young, 1998, 2012). This inductive procedure results in a parsimonious model that explains the dominant modes contained in the data and can be interpreted in physically meaningful terms. Both models are of the 'Hammerstein' type connecting an input nonlinearity and a linear transfer function (TF). The two models differ by the nonlinearity between the input rainfall and 'effective rainfall'. The first model will be used to build a reference discharge series that will serve for identifying and estimating the second model and for

Copyright © 2014 John Wiley & Sons, Ltd.

evaluating the forecasts issued from the different experiments. The second model will be used only in 'forecasting mode'; its structure is simpler as it does not take temperature into account as input data nor simulates a soil component. Instead, the nonlinearity for effective precipitation depends on the value of the flow.

The first model is depicted in Figure 3. The input data are the observed rainfall amount r and temperature t. This model will be referred to as DBMS in the following. This model was inspired from Young (2001) except for the structure of the storage model and the use of temperature. The rectangular boxes represent first-order linear TFs. The first one simulates the water status of the basin s. The TF from effective precipitation to discharge has been identified as a second-order process and can be split into two first-order parallel linear TFs that simulate the fast and slow components of the flow, respectively x_1 and x_2. When used in forecast mode, the decomposed parallel pathway form permits to avoid instability occurring with the second-order TF form (Young and Tomlin, 2000). Two squares with thick borders sketch the nonlinear functions used for the evaluation of effective precipitation: the first one, $f(t)$, is used for the input v of the water storage model and the second one, $g(s)$, for the input u of the routing TF. The full stochastic model comprises an error model ξ that is assumed to be 0. The model is described by the following set of equations at time step k:

$$f_k = \begin{cases} f_1 & \text{if } t_k < \theta_1, \\ f_2 & \text{if } \theta_1 \leq t_k < \theta_2, \\ f_3 & \text{if } \theta_2 \leq t_k \end{cases} \tag{1a}$$

$$v_k = f_k \cdot r_k \tag{1b}$$

$$s_k = c \cdot s_{k-1} + d \cdot v_{k-\delta} \tag{1c}$$

$$g_k = \begin{cases} g_1 + g_2 s_k & \text{if } s_k < \sigma, \\ g_3 + g_4 s_k & \text{if } \sigma \leq s_k, \end{cases} \tag{1d}$$

$$u_k = g_k \cdot r_k \tag{1e}$$

$$x_{1,k} = \alpha_1 \cdot x_{1,k-1} + \beta_1 \cdot u_{k-\delta} \tag{1f}$$

$$x_{2,k} = \alpha_2 \cdot x_{2,k-1} + \beta_2 \cdot u_{k-\delta} \tag{1g}$$

$$x_k = x_{1,k} + x_{2,k} \tag{1h}$$

The parameters of the deterministic part of the model are $\{f_1, f_2, f_3, \theta_1, \theta_2, c, d, \delta, g_1, g_2, g_3, g_4, \sigma, \alpha_1, \alpha_2, \beta_1, \beta_2\}$. Note that the delay δ is equal to 1 day for the three TFs. Let us remind the readers that this model is designed for comparative experiments in an idealized framework; therefore, the potential overfitting relative to measured series has not been considered, and parsimony refers rather to the second model. For this development, algorithms from the CAPTAIN toolbox for Matlab (Taylor *et al.*, 2007) have been used. The State-Dependent Parameter (SDP) algorithms allow for the evaluation of nonparametric dependence (f(t), g(s) and h(q)) and for choosing functions for the nonlinearity (Equations (1a), (1d) and (2a)). The Refined Instrumental Variable algorithms (RIVBJID) are used to identify the order and evaluate the parameters of a TF between the effective precipitation and the streamflow.

The second model is sketched in Figure 4, and the detailed equations are provided in Equation (2). During the first forecasting step, the nonlinearity for the effective rainfall is estimated using the discharge observed on the previous day as a surrogate of the water status of the basin. For the following steps, the deterministic output of the TF, x_{k-1}, is used. This model will be called DBMQ.

$$h_k = \begin{cases} h_1 + h_2 x_k & \text{if } x_k < \sigma, \\ h_3 & \text{if } \sigma \leq x_k \end{cases} \tag{2a}$$

$$u_k = h_k \cdot r_k \tag{2b}$$

$$x_{1,k} = \alpha_1 \cdot x_{1,k-1} + \beta_1 \cdot u_{k-\delta} \tag{2c}$$

$$x_{2,k} = \alpha_2 \cdot x_{2,k-1} + \beta_2 \cdot u_{k-\delta} \tag{2d}$$

$$x_k = x_{1,k} + x_{2,k} \tag{2e}$$

As initial conditions of this second model, we take x_0 from the reference simulation and decompose it in two flows according to

$$x_{1,0} = G_1/(G_1 + G_2)x_0 \tag{3a}$$

$$x_{2,0} = G_2/(G_1 + G_2)x_0 \tag{3b}$$

with the steady-state gains $G_i = \beta_i/(1 - \alpha_i)$ (e.g. Young and Tomlin, 2000). Another possibility is to use the initial conditions from the flows of the two pathways in the reference simulation as if they could be observed. This second approach will be used to test the impact of the choice of the initial conditions. In this latter setting, the model will be named DBMQI.

ECMWF Ensemble Prediction System and reforecasts

Ensemble predictions from ECMWF-EPS issued daily at 00 UTC (Universal Time Coordinated) during the period from 12 March 2008 to 31 December 2012 have been used. This period starts with a change to the Integrated Forecast System cycle c32r3, also corresponding to the beginning of the availability of reforecasts. There are 11 different cycles during this period of almost 5 years including a horizontal resolution change from about 50 to 32 km and the implementation of the ensemble data assimilation scheme. Each ensemble is composed of 1 control and 50 perturbed members. The forecasted precipitation has been cumulated over 24 h from 06 UTC to 06 UTC to match observations as

Copyright © 2014 John Wiley & Sons, Ltd.

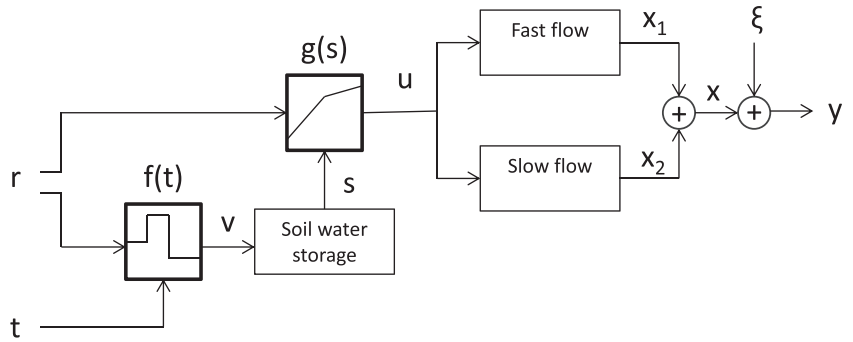

Figure 3. Block diagram of the simulation model (DBMS); the squares represent nonlinearity and the rectangles first-order linear transfer functions

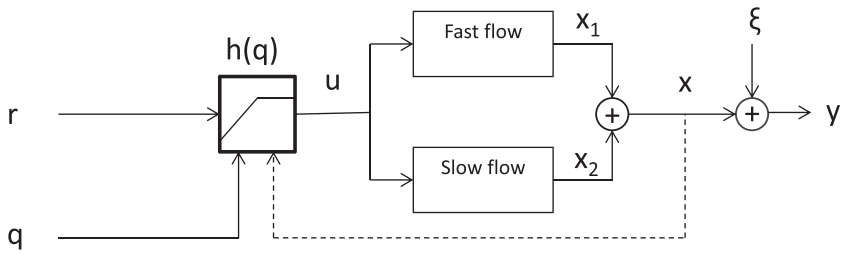

Figure 4. Block diagram of the forecast model (DBMQ); the square represents nonlinearity, and the rectangles represent first-order linear transfer functions

closely as possible. The cumulated precipitation has been averaged over the test catchment. As we focus on uncertainties in precipitation, observed temperature is used even in forecast situations.

The reforecasts are issued once a week. They are made available on Monday for situations starting 17 days later (Thursday) but for the past 20 years. This schedule allows for defining windows extending to 2 weeks ahead of the current week. A calibration window of 5 weeks (2 weeks before and 2 weeks after) is chosen, and 18 years of past situations are used. This means that for every week, a data set of 90 ensemble forecasts is available. A few weeks before a forecast system change, pre-operational ensembles computed with the new cycle are made available. Reforecasts corresponding to this period are also available so that the 5-week window of reforecasts is still available after the change in the operational system without mixing reforecasts from different cycles. Reforecasts are based on five member ensembles, one control and four perturbed members. Like for EPS, precipitation is cumulated over 24 h and is spatially averaged.

Post-processing

Three post-processing methods, one for precipitation and two for the streamflow, are used. For precipitation, ELR (Wilks, 2009) is implemented as Roulin and Vannitsem (2012), preserving the temporal structure of individual members in such a way that it can be used as input in the hydrological model. The parameters are calibrated by maximizing a likelihood function. A correction is applied for the bias in the regression parameters because of the small size of the ensembles in the reforecast data set (Roulin and Vannitsem, 2012). To post-process a raw precipitation ensemble, the 51 members are first ranked and assigned a probability. The corresponding quantiles are then extracted from the post-processed marginal predictive cumulative distribution function based on the reforecasts and finally reordered with respect to the raw ensemble values.

For the streamflow, the variance inflation (INFL) method is adopted (Wood and Schaake, 2008; Johnson and Bowler, 2009). This method is based on the conditions that the climatological variance of the ensemble members should be the same as the variance of the truth and that the correlation of the ensemble members with the ensemble mean should be the same as the correlation of the truth with the ensemble mean. It also satisfies that the mean-square error (MSE) of the ensemble mean is minimized and that the ensemble spread is, on average, representative of the uncertainty in the mean (Johnson and Bowler, 2009). A generalization to many predictors is also proposed (Van Schaeybroeck and Vannitsem, 2014). At time t, the ith ensemble member $f_t^i = \overline{f}_t + e_t^i$ is modified as follows:

$$g_t^i = \mu_x + \alpha\left(\overline{f}_t - \mu_f\right) + \beta e_t^i, \tag{4}$$

where μ_x and μ_f are the mean of the observations and of the forecasts in the calibration data set and

Copyright © 2014 John Wiley & Sons, Ltd.

$$\alpha = \rho_{x\bar{f}} \frac{\sigma_x}{\sigma_{\bar{f}}} \tag{5a}$$

$$\beta = \sqrt{\left(1 - \rho_{x\bar{f}}^2\right)} \frac{\sigma_x}{\sigma_e} \tag{5b}$$

where $\rho_{x\bar{f}}$ is the correlation of the observation with the ensemble mean, σ_x is the standard deviation of the observations, $\sigma_{\bar{f}}$ is the standard deviation of the ensemble mean, and σ_e is the square root of the average ensemble variance. The parameters of the calibration are estimated based on the hydrological reforecasts h_t^i and corresponding true values x_t. Like for the ELR, care has to be taken for the small size of the reforecasts (K = five members). Among the statistics to be estimated, only the variance of the ensemble mean has to be corrected for the ensemble size:

$$\sigma_{\bar{f}}^2 = \sigma_{\bar{h}}^2 - \langle \sigma_h^2 \rangle / K \tag{6}$$

where the angle bracket denotes the average over a set of realizations. The correction of the variance of the ensemble mean has to be taken into account in the estimation of the correlation coefficient $\rho_{x\bar{f}}$; Note that the covariance $\mathrm{cov}(x, \bar{h})$ is not affected by the ensemble size and that the average of the variances of the reforecasts over the calibration data set is unbiased $\left(\sigma_e^2 = \langle \sigma_h^2 \rangle\right)$. The corrected parameters are then

$$\alpha = \frac{\mathrm{cov}(x, \bar{h})}{\sigma_{\bar{h}}^2 - \langle \sigma_h^2 \rangle / K} \tag{7a}$$

$$\beta = \sqrt{\left(1 - \frac{\mathrm{cov}^2(x, \bar{h})}{\sigma_x^2 \left(\sigma_{\bar{h}}^2 - \langle \sigma_h^2 \rangle / K\right)}\right) \frac{\sigma_x^2}{\langle \sigma_h^2 \rangle}} \tag{7b}$$

The parameters are estimated on the logarithm of observations and of the reforecasts so that the distributions are closer to a Gaussian. A post-processed logarithm of the streamflow is exponentiated, and the verification is performed in discharge units. Separate in-sample tests have shown that the INFL performs well at making the mean of the corrected ensemble spread match the mean squared error in the log scale, but some distortion arises when the corrected discharges are transformed back.

Error in Variable Model Output Statistics (EVMOS) was introduced by Vannitsem (2009) recognizing the presence of errors in both observations and model observables. This scheme allows for the maintenance of an appropriate variability of the corrected forecasts even for long lead times and provides a framework in which both the deterministic and the ensemble forecasts can be corrected. For one predictor, hydrological ensembles are post-processed for every member i following

$$g_t^i = \delta + \epsilon f_t^i \tag{8}$$

and two parameters are estimated:

$$\delta = \mu_x - \epsilon \mu_f \tag{9a}$$

$$\epsilon = \sqrt{\frac{\sigma_x^2}{\sigma_{\bar{f}}^2}} \tag{9b}$$

Preliminary tests have shown that the method performs better if all five members of the hydrological reforecasts are taken into account. Therefore, the training data set corresponding to one window of 5 weeks and one lead time comprises $5 \times 5 \times 18 = 450$ couples of past forecasts and observations. Again, the regression is computed after a logarithmic transformation of the streamflow data in order to get approximately Gaussian forecast error.

Verification

The ensembles are evaluated with a set of summary measures including the bias or mean error (ME)

$$ME = \langle \bar{f} - x \rangle \tag{10}$$

the root-mean-square error of the ensemble mean (RMSE)

$$RMSE = \sqrt{\left\langle \left(\bar{f} - x\right)^2 \right\rangle} \tag{11}$$

and the average spread

$$SPREAD = \sqrt{\left\langle \sigma_{\bar{f}}^2 \right\rangle} \tag{12}$$

where $\bar{f}$ is the ensemble mean, $\sigma_{\bar{f}}^2$ is the ensemble variance, x is the observation. To be statistically consistent, the average ensemble variance should match the MSE of the ensemble mean (Talagrand *et al.*, 1997), or, equivalently, the *SPREAD* should be equal to the *RMSE*.

Further insight into the probabilistic skill of the ensemble is given by the CRPS, evaluating the distance between two distributions:

$$CRPS = \int_{-\infty}^{\infty} [P(x) - P_o(x)]^2 dx \tag{13a}$$

$$P_o(x) = \begin{cases} 0, & x < x_o \\ 1, & x \geq x_o \end{cases} \tag{13b}$$

$$P(x) = p_i \equiv \frac{i}{K}, \quad \text{for } x_i < x < x_{i+1} \tag{13c}$$

where $\{x_1, \cdots, x_K\}$ are the ordered members of the ensemble of size K and x_o is the observation.

Copyright © 2014 John Wiley & Sons, Ltd.

We use the methodology developed by Hersbach (2000) to compute the CRPS for an ensemble system and to decompose the average value into three components:

$$\langle CRPS \rangle = UNC + REL - RES \tag{14}$$

where UNC is the uncertainty and is the best achievable value of CRPS in case only climatological information is available, REL is the reliability testing whether, on average, the frequency that the verifying analysis was found to be below the middle of the interval number i is proportional to i/K, and RES is the resolution that measures the superiority of the forecast system with respect to a forecast system only based on climatology. In the following, CRPS will simply denote its average value.

Experiments

The different model versions have been combined with the three post-processing methods in a set of experiments to investigate the potential improvement of the ensembles using the information contained in the reforecasts. Acronyms are summarized on Table I. The hydrological ensembles obtained with DBMS are used to study the impact of errors in the forcing. The sensitivity to the values of the model parameters is tackled by adding errors in the DBMS model. First, in the DBMSPF model version, the function (Equation (1d)) used to estimate the effective precipitation as a fraction of the input precipitation is shifted by an amount of −0.04 and +0.08. The same change is applied for the whole range of the model state variable that represents the soil water status. In the DBMSPS model version, the coefficient α_1 of the slow flow (Equation (1f)) is changed from the value of 0.978 in the reference model to either 0.95 or 0.99. In the DBMSPQ model version, the coefficient α_2 of the quick flow (Equation (1g)) is set to 0.4 and to 0.7 instead of the value of 0.55 as in the reference model. As already indicated in Section on The Hydrological Model, the DBMQ model has a simpler structure than DBMS. In order to investigate the impact of the initial conditions, the DBMQI version of this model takes the values of initial flows of the slow and quick paths from the reference simulation of DBMS from which streamflow 'observations' are supplied.

Table I. Models and post-processing methods associated with the experiment's acronyms

Model	
DBMS	Reference simulation model
DBMSPF	Sensitivity to effective precipitation function
DBMSPS	Sensitivity to slow flow parameter
DBMSPQ	Sensitivity to quick flow parameter
DBMQ	Forecast model
DBMQI	Initial conditions from reference simulation
Post-processing	
RAW	Forcing with raw precipitation ensembles
ELR	Precipitation ensembles corrected with extended logistic regression
INFL	Raw hydrological ensembles corrected with the inflation method
EV1	Raw hydrological ensembles corrected with the EVMOS with one predictor
ELR/INFL	ELR followed by INFL post-processing
ELR/EV1	ELR followed by EV1 post-processing

The basic experiments consist in the forcing of the models with the raw precipitation ensembles, e.g. DBMS-RAW (i.e. no post-processing). Then, the precipitation ensembles are modified according to the logistic distributions and are used as input to the hydrological models, e.g. DBMQ-ELR. The raw reforecasts are also used in the hydrological models and provide the hydrological reforecasts. The parameters of post-processing, INFL and EVMOS with one predictor (EV1), are estimated on the basis of these hydrological reforecasts and the reference simulation results at the past dates. These parameters are used to post-process the raw hydrological ensembles, e.g. DBMQI-INFL and DBMQI-EV1. The last set of experiments consists in applying successively the post-processing of precipitation and the post-processing of the hydrological ensembles. The information used for both steps is coming from a 5-week window of reforecasts: First, the parameters of the ELR are taken to post-process the same five-member reforecasts that were used for calibration; next, these modified five-member precipitation ensembles serve as input to a hydrological model to provide improved hydrological reforecasts; then, the hydrological reforecasts together with the corresponding discharge values from the reference simulation allow to obtain new parameter values for the INFL or the EVMOS method. Finally, these new parameter sets are used to post-process hydrological ensembles that are resulting already from the forcing with ELR-corrected precipitation ensembles. For instance, DBMS-ELR/INFL corresponds to the post-processing of DBMS-ELR ensembles. The same correction factors for the ensemble size that were applied in the optimization of the parameters of the ELR (refer to Roulin and Vannitsem, 2012) have to be used to modify the five-member precipitation ensembles.

RESULTS

Ensemble predictions of precipitation

The raw ensemble predictions of areal precipitation are validated against the corresponding values obtained by spatially averaging rain-gauge data. Results are presented

Copyright © 2014 John Wiley & Sons, Ltd.

separately for the winter (October to March, Figure 5) and for the summer (April to September, Figure 6). The raw ensemble predictions of areal precipitation are characterized by small values of the ME during the winter ranging from -0.07 to $0.2\,\mathrm{mm\,day^{-1}}$ and by positive values during the summer decreasing from 0.8 on the first forecast day ($D+1$) to $0.5\,\mathrm{mm\,day^{-1}}$ at $D+9$. During the forecast range in the winter, the RMSE grows from 2.3 to $4.2\,\mathrm{mm\,day^{-1}}$.The RMSE is larger during the summer as it increases from 3.7 to $4.9\,\mathrm{mm\,day^{-1}}$. The spread of the raw ensembles increases with the lead time. For the two seasons, the average spread is lower than the corresponding RMSE values, which is an indication that the raw ensembles are underdispersed. CRPS also increases with lead time. At the beginning of the forecasts, CRPS values during the winter are lower than CRPS values during the summer. These differences correspond to lower values of the reliability (REL) and higher values of the resolution (RES). Remembering Equation (14), these two components contribute to better CRPS during the winter despite that the uncertainty component is higher for the winter ($1.96\,\mathrm{mm\,day^{-1}}$) than for the summer ($1.84\,\mathrm{mm\,day^{-1}}$). At the end of the forecast, the differences are less important except for the uncertainty components depending only on the validation samples.

The post-processing of the ensembles using the ELR results for the winter in a positive bias. The RMSE is almost unchanged; the spread is increased, leading to overdispersion, and the CRPS is decreased. This improvement of the CRPS is mainly because of a decrease of the reliability component. For the summer, a decrease of the bias is obtained, which remains positive around $0.3\,\mathrm{mm\,day^{-1}}$ throughout the forecast range. The RMSE is increased, and the CRPS is decreased the first six forecast days. Like for the winter, this CRPS improvement is due to the decrease of the reliability

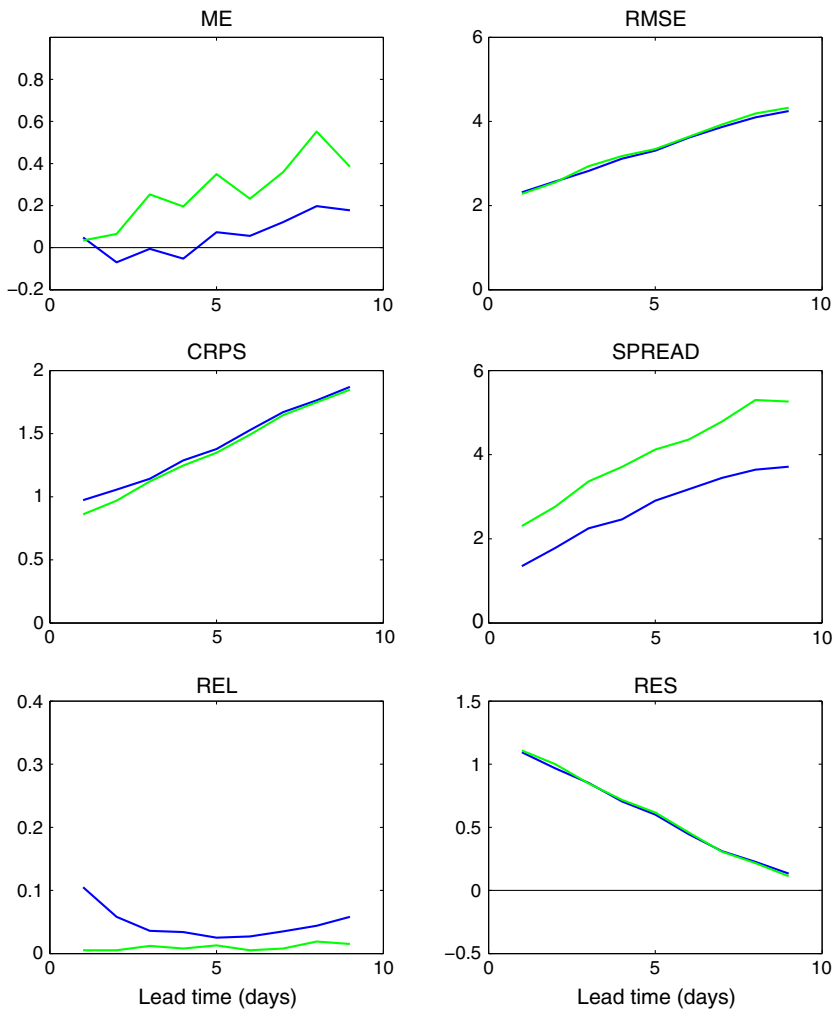

Figure 5. Verification of the ensemble predictions for precipitation during the winter: (blue) raw ensembles, (green) post-processed using the extended logistic regression with parameters fitted on reforecasts. Mean error (ME), root-mean-square error (RMSE), continuous ranked probability score (CRPS), square root of the average ensemble variance (SPREAD), reliability (REL) and resolution (RES) terms of the CRPS. All is expressed in $\mathrm{mm\,day^{-1}}$

Copyright © 2014 John Wiley & Sons, Ltd.

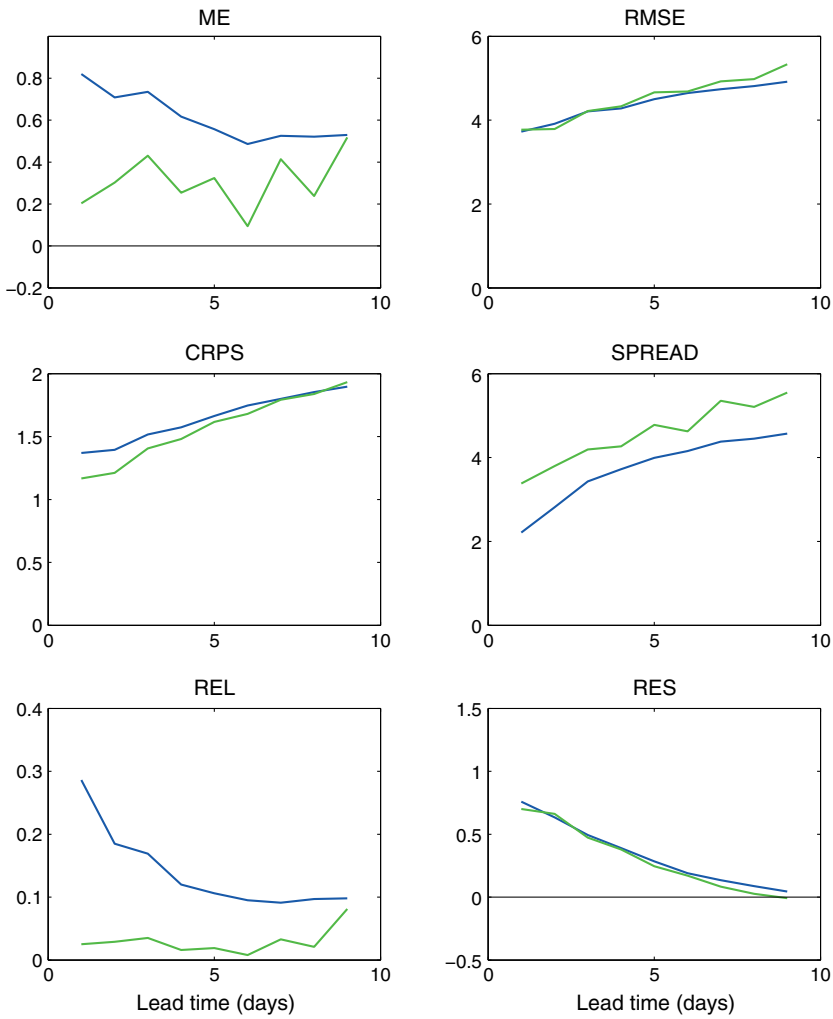

Figure 6. As in Figure 5 but during the summer

term, whereas the resolution term is degraded mostly from $D+7$, cancelling out the improvement of the reliability at the end of the forecast.

Overall, the post-processing of precipitation based on reforecasts is correcting the ensemble reliability, but an additional bias is introduced and is believed to be associated with the number of reforecast ensemble members even if considerable correction of the parameter has already been obtained as in Roulin and Vannitsem (2012). This aspect needs further investigations based on additional reforecasts.

Hydrological ensemble predictions

Winter. The DBMQ model is taken as the hydrological model used to forecast the streamflow simulated with the DBMS model. The first experiment consists in forcing the model with the raw precipitation ensembles (DBMQ-RAW). The verification results for the winter are presented in Figure 7. The bias is negative starting from $D+3$ and stays above $-0.2\,\mathrm{m^3\,s^{-1}}$. The RMSE grows with the forecast horizon faster than the average spread of the ensembles does. This indicates that the hydrological ensembles are underdispersed. The same diagnostic was already derived for the raw precipitation ensembles. Because of the delay of 1 day in the model, the forecast at $D+1$ is deterministic with the input of the precipitation observed the day before. Hence, the first value of the CRPS corresponds to the mean absolute error, and the reliability and resolution components depend on how the errors distribute into positive and negative values. The CRPS grows with the forecast horizon and can be decomposed into three terms: uncertainty that stays almost constant during the forecast range because it only depends on the verification sample ($2.6\,\mathrm{m^3\,s^{-1}}$ for the winter), the reliability starting from about $0.3\,\mathrm{m^3\,s^{-1}}$ at the first forecast day and decreasing afterwards, and the resolution also decreasing.

We now analyse the verification results of the five post-processing methods applied to the DBMQ model to

Copyright © 2014 John Wiley & Sons, Ltd.

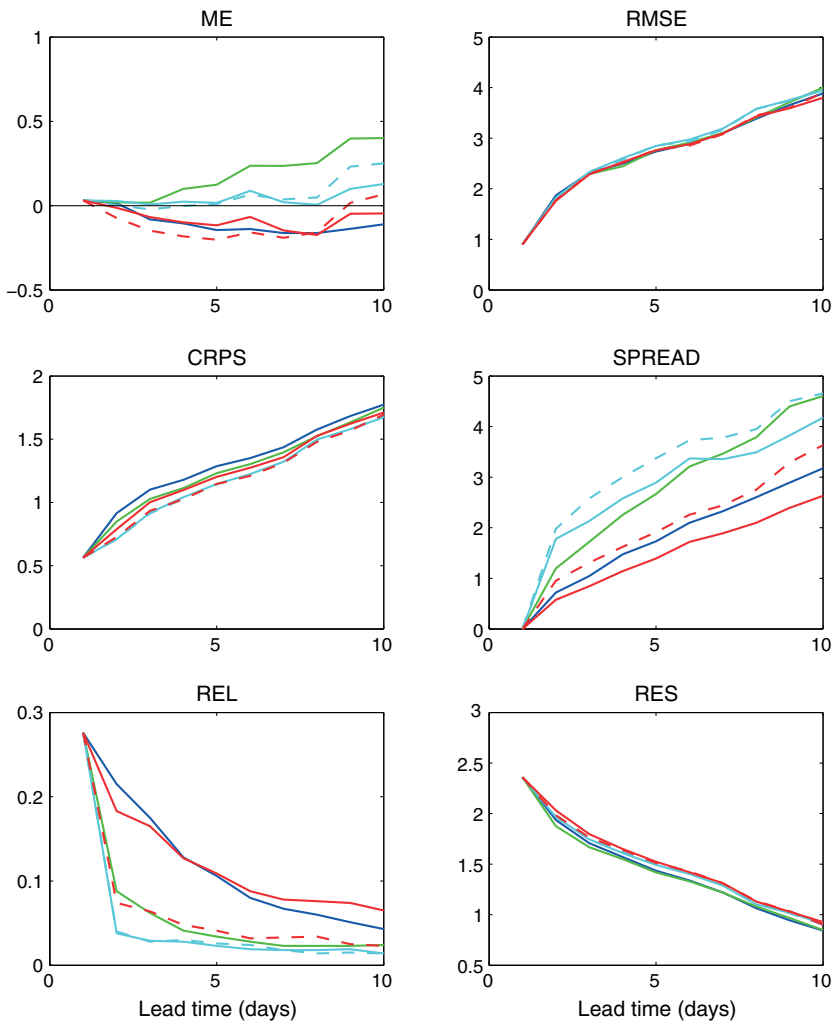

Figure 7. Verification of hydrological ensemble predictions with DBMQ during winter: (blue) DBMQ-RAW, (green) DBMQ-ELR, (light blue) DBMQ-INFL, (light blue, dashed) DBMQ-ELR/INFL, (red) DBMQ-EV1, and (red, dashed) DBMQ-ELR/EV1. All is expressed in $m^3 s^{-1}$

mitigate errors in the forcing and in the model structure. Forcing DBMQ with post-processed precipitation ensembles (DBMQ-ELR) results in a positive bias increasing with lead time. This follows the trend of the bias of the post-processed precipitation ensembles. The RMSE is almost unchanged. The spread is increased so that the new hydrological ensembles are overdispersed beyond $D+5$. The CRPS is decreased because of a lowering of the reliability component and despite a decrease of the resolution component at $D+2$ and $D+3$.

The variance inflation method is applied on the hydrological ensembles obtained using the DBMQ model (DBMQ-INFL). The bias is well corrected, but the RMSE is slightly increased. The spread is increased to values closer to RMSE than for DBMQ-ELR ensembles. The CRPS is improved because of both the reduction of the reliability term and the increase of the resolution term. The post-processing of precipitation ensembles followed by the post-processing of streamflow ensembles (DBMQ-ELR/INFL) does not provide further improvements as compared with the INFL alone because the bias is degrading the last three forecast days.

The EVMOS post-processing (DBMQ-EV1) results in almost no change in ME and RMSE, and the spread is decreased away from values reached by RMSE. The CRPS is improved because of an increase of the resolution component. Unlike the two previous methods (ELR and INFL), the reliability component is only slightly changed. The post-processing of precipitation ensembles with ELR followed by the post-processing of DBMQ hydrological ensembles (DBMQ-ELR/EV1) results in a negative bias, no change in RMSE and a spread similar to the spread of the raw hydrological ensembles. The benefit of combining both methods consists in the improvement of the reliability component. Together with an increased resolution component, it contributes to a decrease of CRPS.

The hydrological ensembles (DBMQ-RAW) are underdispersed. For a perfectly reliable system, we would

Copyright © 2014 John Wiley & Sons, Ltd.

expect that the observation is an outlier for a percentage of situations equal to 2/(K + 1) where K is the size of the ensemble, that is, 4% for $K = 51$. During the winter, it is as high as 63% at $D + 2$ and decreases to 16% at $D + 10$. There are more outliers larger than any ensemble member than lower (respectively 40 and 23% at $D + 2$). The number of outliers changes with the post-processing method in a way that is consistent with the changes in reliability. For the winter, the EVMOS post-processing does not improve the number of outliers, whereas the four other methods do. The ELR/INFL performs best. At $D + 10$, the methods using the corrected precipitation (ELR, ELR/INFL and ELR/EV1) result in overdispersion, and the number of outliers is 0 or 1%.

Among the five post-processing methods tested, the inflation performs best for winter situations with respect to the bias, the CRPS, the reliability and to the average spread being closer to the RMSE.

The validation results of the hydrological ensembles obtained by forcing the DBMQ model with the raw precipitation ensembles (DBMQ-RAW) are compared with the results of the perfect model situation (DBMS-RAW) as well as with those obtained with the experiment using correct initial conditions (DBMQI-RAW) in Figure 8. By design, DBMS-RAW has no error at $D + 1$, and the resolution component of the CRPS is equal to the uncertainty component. The RMSE grows with the lead time only because of the errors in the forcing. The difference between CRPS values of DBMS-RAW and DBMQ-RAW decreases with lead time because of the sharp decrease of the reliability component for DBMQ-RAW, whereas the difference between the values of the resolution components of DBMS-RAW and DBMQ-RAW does not vary much throughout the forecast. When taking the initial conditions from the reference simulation (DBMQI-RAW), the RMSE is restored to the values of the perfect model situation. The

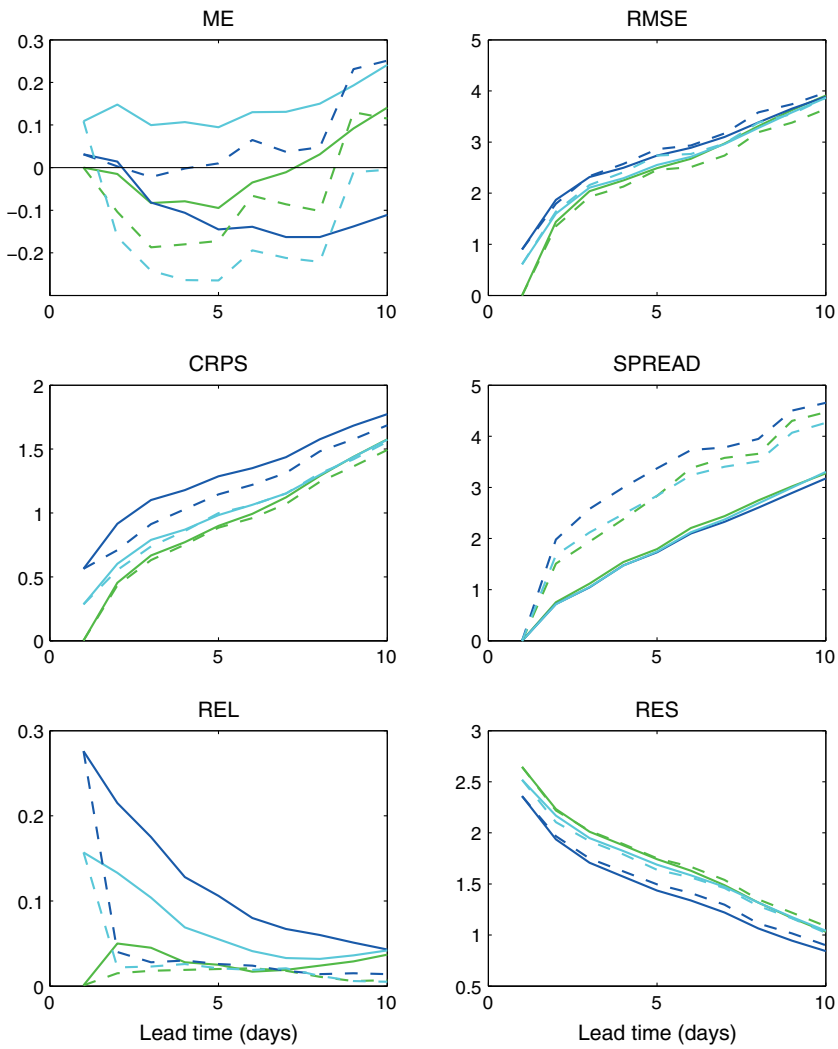

Figure 8. Verification of hydrological ensemble predictions during winter: (blue) DBMQ, (light blue) DBMQI, (green) DBMS, (continuous) RAW, and (dashed) ELR/INFL. All is expressed in $m^3 s^{-1}$

Copyright © 2014 John Wiley & Sons, Ltd.

values of the CRPS with DBMQI-RAW are also closer to those of DBMS-RAW, and the difference between both vanishes from $D+7$ onwards. This is because of a lowering of the reliability component compared to DBMQ-RAW and because of the increase of the resolution component close to the values of DBMS-RAW. During the winter, the impact of errors in the initial conditions seems to dominate. The effect of post-processing of the different hydrological ensembles is also shown in Figure 8. Only the combined pre-processing and post-processing with ELR/INFL is displayed. The post-processing improves the RMSE of the DBMS ensembles, whereas it degrades the RMSE of DBMQ and DBMQI. The post-processing improves the CRPS of DBMQ ensembles better than the CRPS of DBMS and of DBMQI. The reliability of the post-processed ensembles is similar for the three systems. The resolution is improved for DBMQ mitigating partly the lack of good initial conditions.

The different post-processing methods applied to DBMS and DBMQI ensembles are compared in figures included in the Supporting Information. It can be noted from these results that the pre-processing (ELR) degrades the bias that becomes increasingly positive with forecast lead time like for DBMQ. The post-processing with EVMOS does not modify the average spread, and it does not improve the reliability of DBMS-EV1 and partly improves the reliability of DBMQI-EV1. The combination of ELR/EV1 allows for increasing the spread and correcting the reliability.

Summer. The hydrological ensembles obtained with DBMQ during the summer are now analysed (Figure 9). The forcing with the raw precipitation ensembles (DBMQ-RAW) results in a positive bias increasing with the lead time associated with the large positive bias of the precipitation ensembles. The RMSE values also increase,

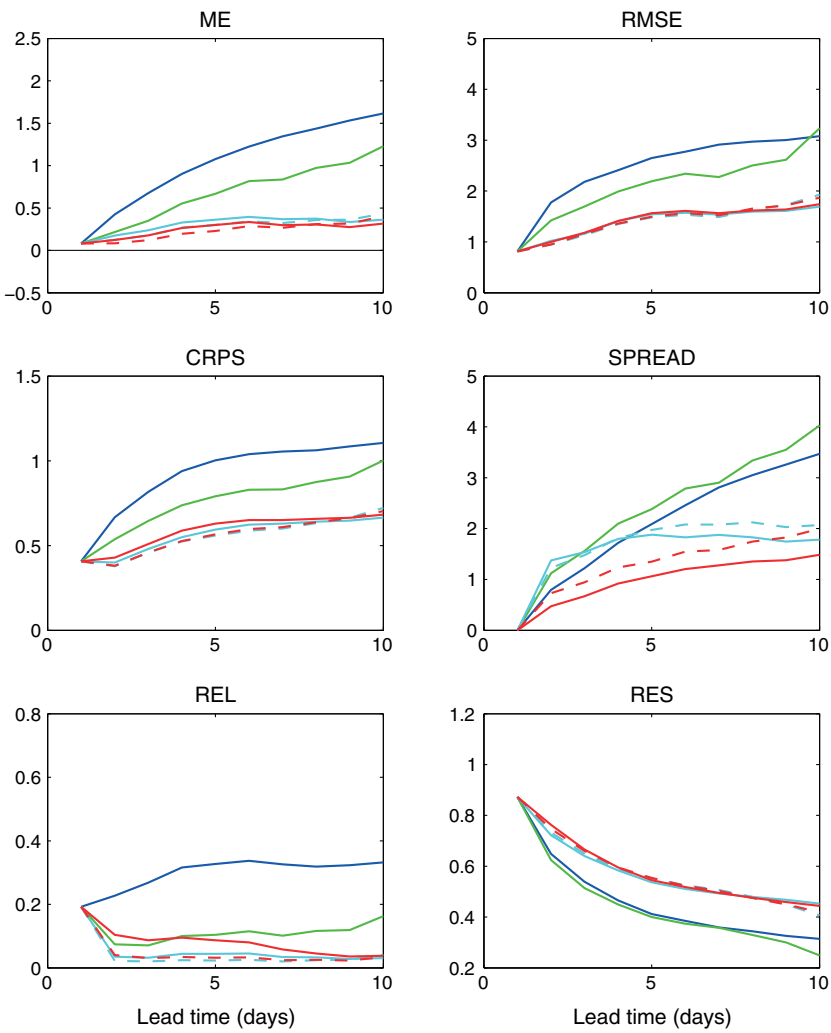

Figure 9. Verification of hydrological ensemble predictions with DBMQ during summer: (blue) DBMQ-RAW, (green) DBMQ-ELR, (light blue) DBMQ-INFL, (light blue, dashed) DBMQ-ELR/INFL, (red) DBMQ-EV1, and (red, dashed) DBMQ-ELR/EV1. All is expressed in $m^3 s^{-1}$

Copyright © 2014 John Wiley & Sons, Ltd.

exceeding the corresponding values of the average spread except the last 3 days. The average CRPS also grows with the forecast days.

During the summer, the lowering of the positive bias of the precipitation ensembles with the ELR pre-processing results in the lowering of the bias of the DBMQ-ELR ensembles. This, in turn, induces a decrease of the RMSE (except for $D+10$). The average spread is overdispersed. The CRPS is improved because of the reduction of the reliability term, whereas the resolution is slightly degraded.

The inflation method (DBMQ-INFL) allows reducing the bias more significantly than with the forcing with ELR-corrected precipitation ensembles. The RMSE is also reduced substantially, and the average spread is reduced (from $D+5$ onwards) consistently. The CRPS is improved thanks to a large decrease of the reliability and a significant increase of the resolution. Unlike the winter, the post-processing of precipitation ensembles followed by the post-processing of hydrological ensembles (DBMQ-ELR/INFL) results in further improvements of the bias and of the CRPS as the corrected hydrological ensembles are almost perfectly reliable with a resolution similar to DBMQ-INFL. The results with EVMOS (DBMQ-EV1 and DBMQ-ELR/EV1) are similar to those with the inflation method.

During the summer, when a large positive bias is present in the forcing precipitation ensembles, all methods improve the ME, RMSE and CRPS, and the best scores are achieved when applying successively the pre-processing of the precipitation ensembles and the post-processing of the hydrological ensembles.

We now compare the DBMS-RAW and DBMQ-RAW ensembles during the summer (Figure 10). In the first forecast day, there is no error in DBMS-RAW, but the scores, ME, RMSE and CRPS grow with lead time faster

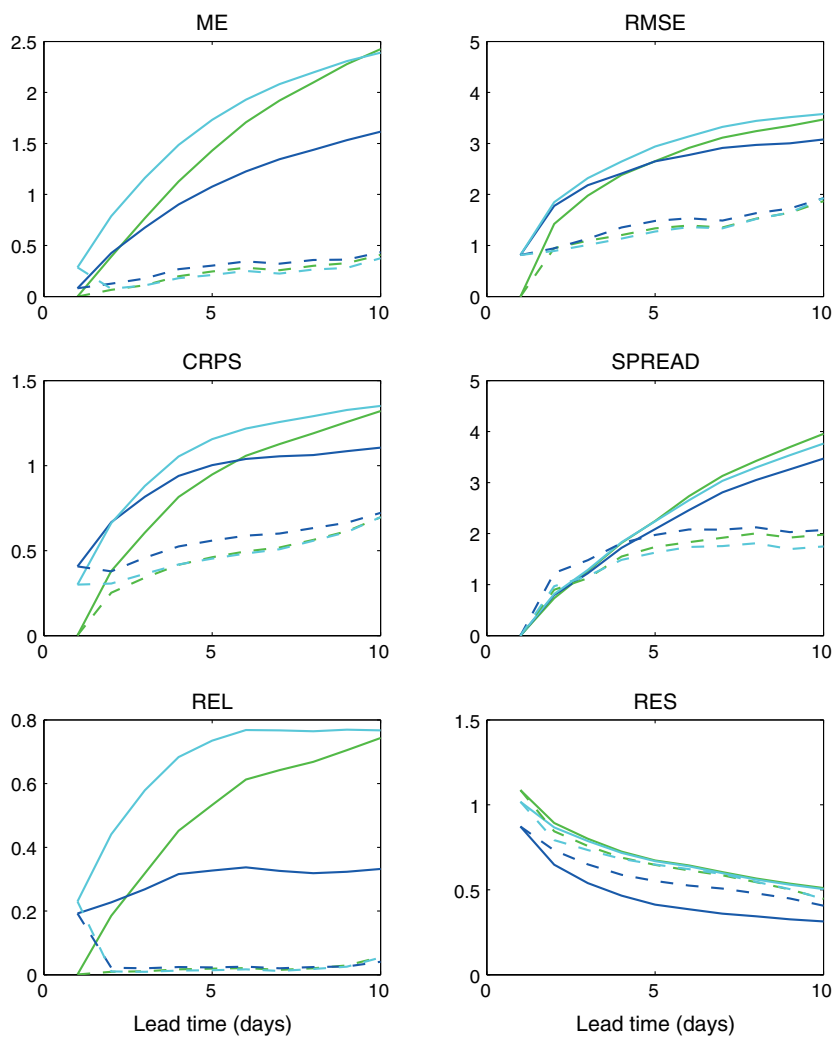

Figure 10. Verification of hydrological ensemble predictions during summer: (blue) DBMQ, (light blue) DBMQI, (green) DBMS, (continuous) RAW, and (dashed) ELR/INFL. All is expressed in $m^3 s^{-1}$

Copyright © 2014 John Wiley & Sons, Ltd.

than that for DBMQ-RAW. Like for the winter, the DBMS-RAW ensembles have a better resolution during the forecast range that is offset by the rapid loss of reliability after the sixth forecast day. This should be related to the fact that DBMQ is not using information about temperature, and the variability allowed by this model is lower than the variability in DBMS output.

With the correct initial conditions taken from the reference simulation, the ensembles (DBMQI-RAW) have a positive bias at $D+1$ that has a larger value than that for DBMQ-RAW. For longer lead times, the ME, RMSE and CRPS have the worst values of the three systems. Only the resolution term of the CRPS with DBMQI-RAW is improved during the forecast. To understand this feature, one must note that the calibration of the model parameters is performed once for all seasons and is more sensitive to large events occurring during the winter. During the summer, the difference between the two models is such that the initial conditions taken from DBMS are not suitable for DBMQ, and a data assimilation scheme would probably improve the hydrological ensembles.

The results of the post-processing (ELR/INFL) for the three models (DBMS, DBMQ and DBMQI) are compared in Figure 10. Results with the different post-processing methods applied to DBMS and DBMQI ensembles during the summer are included in the Supporting Information. With the post-processing, almost all the scores are improved to similar values. The resolution for DBMQ is clearly improved, whereas it is slightly degraded for DBMS and DBMQI. The post-processed ensembles are almost perfectly reliable for the three models. The CRPS follows the trend of the resolution, with DBMQ very close to DBMS and

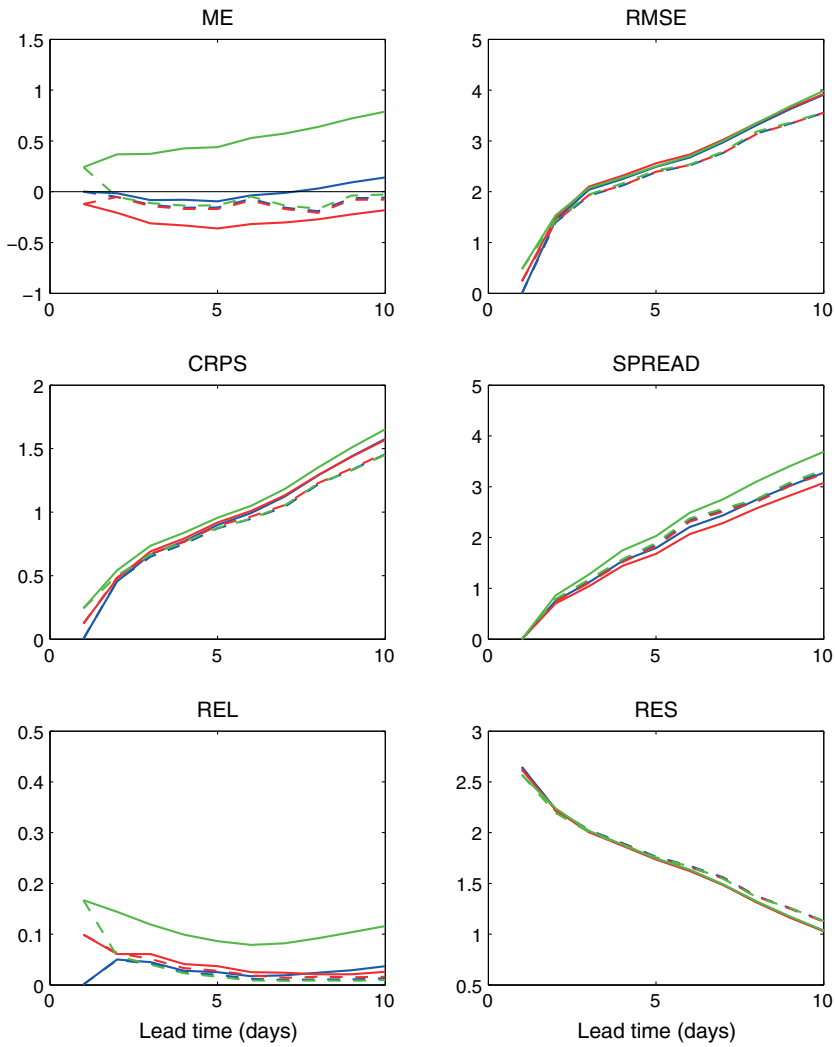

Figure 11. Verification of hydrological ensemble predictions with DBMSPF (error in the effective precipitation function) during winter: (blue) optimal parameter values for the function g_k (refer to Section on Experiments), (red) $g_k = g_k - 0.04$, (green) $g_k = g_k + 0.08$, (continuous) DBMSPF-RAW, (dashed) DBMSPF-EV1. All is expressed in $m^3 s^{-1}$

Copyright © 2014 John Wiley & Sons, Ltd.

DBMQI. For the summer, there is a clear benefit to post-process the hydrological ensembles.

Errors in model parameters

Finally, the sensitivity to DBMS model parameters is analysed. Three parameters have been tested (refer to Section on Experiments). The results are presented for the parameter of the effective precipitation during the winter only (Figure 11), and the results for the other parameters and for the summer may be found in the Supporting Information. The range of the tested parameter values is such that the ME is significantly changed. For instance, decreasing the value of the parameter of the effective precipitation induces a negative bias throughout the forecast range, while on the other hand, increasing this value results in a positive bias. The post-processing has been applied in this context, and only EV1 has been plotted in Figure 11. The hydrological ensembles obtained with errors in the parameters are considerably corrected using EV1 with values (ME, RMSE and CRPS) close to the ones obtained with the optimal parameters.

The same analysis has been performed with errors added to the values of the parameter of the fast flow and of the slow flow (Supporting Information) with the same qualitative conclusions. The other post-processing methods have also been applied in the context of the DBMS model with errors in the parameters (not shown). The differences between the methods are similar to the differences already highlighted. In summary, the post-processing based on reforecasts effectively corrects errors due to the presence of suboptimal parameters.

CONCLUSIONS

In this study, the impacts of errors in the forcing, the model structure and parameters, and the role of initial conditions have been analysed using a system consisting of precipitation forecasts from the operational ensemble prediction system of ECMWF and two hydrological models based on a linear TF between effective rainfall and streamflow. The structure of the two models differs in the way that the nonlinearity between input precipitation and effective rainfall is controlled: the first model using temperature and a simple soil storage representation and the second using the observed or forecasted streamflow as the surrogate of the catchment wetness status. The impact of errors was studied by means of a set of verification scores of the resulting hydrological ensemble predictions against a reference simulation performed with the first hydrological model forced with observed precipitation and temperature.

To improve these ensemble forecasts, the use of post-processing techniques was investigated. To this aim, hydrological reforecasts were constructed by forcing the hydrological models with meteorological reforecasts produced operationally using the same NWP model version as that used for the ensemble predictions. As clearly demonstrated, the post-processing based on the information of the hydrological reforecasts improves the reliability and the resolution of the ensembles with the second model. It also improves the reliability of the reference model and of the forecast model with known initial conditions but not their resolution. Post-processing the precipitation only does not improve the resolution of the hydrological ensembles indicating the necessity of post-processing the hydrological ensembles. This, in turn, advocates for the development of hydrological reforecasts as developed in the present work.

Good initial conditions for the forecast model are necessary to obtain the best resolution, but the post-processing mitigates the lack of this information. During the summer, the post-processing must be applied together with the use of better initial conditions in order to achieve the best scores. Finally, altering the parameters of the reference model degrades the scores of the resulting ensembles, but post-processing improves these scores that converge to the values of post-processed ensembles resulting from the model version with optimal parameters.

ACKNOWLEDGEMENTS

The authors would like to thank Peter Young and James Taylor for their support to the use of the CAPTAIN toolbox and Bert Van Schaeybroeck for useful discussions. Thanks are also due to Thomas Pagano for his valuable suggestions. This work was partly supported by the Belgian Federal Science Policy Office under contracts SD/CA/04A and BR/121/A2/STOCHCLIM.

REFERENCES

Ajami NK, Duan Q, Sorooshian S. 2007. An integrated hydrologic bayesian multimodel combination framework: Confronting input, parameter, and model structural uncertainty in hydrologic prediction. *Water Resources Research* **43**(1). DOI: 10.1029/2005WR004745

van Andel SJ, Weerts A, Schaake J, Bogner K. 2013. Post-processing hydrological ensemble predictions intercomparison experiment. *Hydrological Processes* **27**(1): 158–161. DOI: 10.1002/hyp.9595

Applequist S, Gahrs GE, Pfeffer RL, Niu X-F. 2002. Comparison of methodologies for probabilistic quantitative precipitation forecasting. *Weather and Forecasting* **17**: 783–799. DOI: 10.1175/1520-0434(2002)017<0783:COMFPQ>2.0.CO;2

Beven KJ. 2001. *Rainfall-runoff modelling: the primer*. John Wiley & Sons Ltd: Chichester.

Beven KJ, Leedal DT, Smith PJ, Young PC. 2012. Identification and representation of state dependent non-linearities in flood forecasting using DBM methodology. In *System identification, environmental modelling, and control system design*, Wang L, Garnier H (eds). Springer: London; 341–366.

Copyright © 2014 John Wiley & Sons, Ltd.

Bremnes JB. 2004. Probabilistic forecasts of precipitation in terms of quantiles using nwp model output. *Monthly Weather Review* **132**: 338–347. DOI: 10.1175/1520-0493(2004)132<0338:PFOPIT>2.0.CO;2

Brocca L, Moramarco T, Melone F, Wagner W, Hasenauer S, Hahn S. 2012. Assimilation of surface-and root-zone ascat soil moisture products into rainfall–runoff modeling. *IEEE Transactions on Geoscience and Remote Sensing* **50**(7): 2542–2555.

Brown JD, Seo D-J. 2013. Evaluation of a nonparametric post-processor for bias correction and uncertainty estimation of hydrologic predictions. *Hydrological Processes* **27**(1): 83–105. DOI: 10.1002/hyp.9263

Cloke H, Pappenberger F. 2009. Ensemble flood forecasting: A review. *Journal of Hydrology* **375**(34): 613–626. DOI: 10.1016/j.jhydrol.2009.06.005

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

Engeland K, Steinsland I. 2014. Probabilistic postprocessing models for flow forecasts for a system of catchments and several lead times. *Water Resources Research* **50**(1): 182–197. DOI: 10.1002/2012WR012757

Friederichs P, Hense A. 2008. A probabilistic forecast approach for daily precipitation totals. *Weather and Forecasting* **23**: 659–673. DOI: 10.1175/2007WAF2007051.1

Georgakakos KP, Seo D-J, Gupta H, Schaake J, Butts MB. 2004. Towards the characterization of streamflow simulation uncertainty through multimodel ensembles. *Journal of Hydrology* **298**(14): 222–241. DOI: 10.1016/j.jhydrol.2004.03.037

Hagedorn R. 2008. Using the ECMWF reforecast dataset to calibrate EPS forecasts. *ECMWF Newsletter* **117**: 9–13.

Hamill TM, Whitaker JS. 2006. Probabilistic quantitative precipitation forecasts based on reforecast analogs: Theory and application. *Monthly Weather Review* **134**: 3209–3229. DOI: 10.1175/MWR3237.1

Hamill TM, Whitaker JS, Wei X. 2004. Ensemble reforecasting: Improving medium-range forecast skill using retrospective forecasts. *Monthly Weather Review* **132**: 1434–1447. DOI: 10.1175/1520-0493(2004)132<1434:ERIMFS>2.0.CO;2

Hashino T, Bradley AA, Schwartz SS. 2007. Evaluation of bias-correction methods for ensemble streamflow volume forecasts. *Hydrology and Earth System Sciences* **11**(2): 939–950. DOI: 10.5194/hess-11-939-2007

Hersbach H. 2000. Decomposition of the continuous ranked probability score for ensemble prediction systems. *Weather and Forecasting* **15**: 559–570. DOI: 10.1175/1520-0434(2000)015<0559:DOTCRP>2.0.CO;2

Johnson C, Bowler N. 2009. On the reliability and calibration of ensemble forecasts. *Monthly Weather Review* **137**: 1717–1720. DOI: 10.1175/2009MWR2715.1

Komma J, Blöschl G, Reszler C. 2008. Soil moisture updating by ensemble kalman filtering in real-time flood forecasting. *Journal of Hydrology* **357**(34): 228–242. DOI: 10.1016/j.jhydrol.2008.05.020

Liu Y, Weerts AH, Clark M, Hendricks Franssen H-J, Kumar S, Moradkhani H, Seo D-J, Schwanenberg D, Smith P, van Dijk AIJM, van Velzen N, He M, Lee H, Noh SJ, Rakovec O, Restrepo P. 2012. Advancing data assimilation in operational hydrologic forecasting: progresses, challenges, and emerging opportunities. *Hydrology and Earth System Sciences* **16**(10): 3863–3887. DOI: 10.5194/hess-16-3863-2012

Madadgar S, Moradkhani H, Garen D. 2012. Towards improved post-processing of hydrologic forecast ensembles. *Hydrological Processes*. DOI: 10.1002/hyp.9562

Molteni F, Buizza R, Palmer TN, Petroliagis T. 1996. The ecmwf ensemble prediction system: Methodology and validation. *Quarterly Journal of the Royal Meteorological Society* **122**(529): 73–119. DOI: 10.1002/qj.49712252905

Pagano TC, Shrestha DL, Wang QJ, Robertson D, Hapuarachchi P. 2012. Ensemble dressing for hydrological applications. *Hydrological Processes*. DOI: 10.1002/hyp.9313

Reggiani P, Weerts A. 2008. A bayesian approach to decision-making under uncertainty: An application to real-time forecasting in the river rhine. *Journal of Hydrology* **356**(12): 56–69. DOI: 10.1016/j.jhydrol.2008.03.027

Roulin E, Vannitsem S. 2005. Skill of medium-range hydrological predictions. *Journal of Hydrometeorology* **6**: 729–744.

Roulin E, Vannitsem S. 2012. Postprocessing of ensemble precipitation predictions with extended logistic regression based on hindcasts. *Monthly Weather Review* **140**: 874–888. DOI: 10.1175/MWR-D-11-00062.1

Schmeits MJ, Kok KJ. 2010. A comparison between raw ensemble output, (modified) bayesian model averaging, and extended logistic regression using ecmwf ensemble precipitation reforecasts. *Monthly Weather Review* **138**: 4199–4211. DOI: 10.1175/2010MWR3285.1

Seo D-J, Koren V, Cajina N. 2003. Real-time variational assimilation of hydrologic and hydrometeorological data into operational hydrologic forecasting. *Journal of Hydrometeorology* **4**: 627–641. DOI: 10.1175/1525-7541(2003)004<0627:RVAOHA>2.0.CO;2

Sloughter JML, Raftery AE, Gneiting T, Fraley C. 2007. Probabilistic quantitative precipitation forecasting using bayesian model averaging. *Monthly Weather Review* **135**: 3209–3220. DOI: 10.1175/MWR3441.1

Talagrand O, Vautard R, Strauss B. 1997. Evaluation of probabilistic prediction systems. In *Workshop on Predictability, 20-22 October 1997,* European Centre for Medium-Range Weather Forecasts, 1–25.

Taylor CJ, Pedregal DJ, Young PC, Tych W. 2007. Environmental time series analysis and forecasting with the captain toolbox. *Environmental Modelling & Software* **22**(6): 797–814.

Van Schaeybroeck B, Vannitsem S. 2014. Ensemble post-processing using member-by-member approaches: theoretical aspects. *Quarterly Journal of the Royal Meteorological Society*. Accepted manuscript online. DOI: 10.1002/qj.2397

Van Steenbergen N, Ronsyn J, Willems P. 2012. A non-parametric data-based approach for probabilistic flood forecasting in support of uncertainty communication. *Environmental Modelling & Software* **33**: 92–105. DOI: 10.1016/j.envsoft.2012.01.013

Vannitsem S. 2009. A unified linear model output statistics scheme for both deterministic and ensemble forecasts. *Quartely Journal of the Royal Meteorological Society* **135**(644): 1801–1815. DOI: 10.1002/qj.491

Verkade J, Brown J, Reggiani P, Weerts A. 2013. Post-processing ECMWF precipitation and temperature ensemble reforecasts for operational hydrologic forecasting at various spatial scales. *Journal of Hydrology* **501**: 73–91. DOI: 10.1016/j.jhydrol.2013.07.039

Vrugt JA, Gupta HV, Bouten W, Sorooshian S. 2003. A shuffled complex evolution metropolis algorithm for optimization and uncertainty assessment of hydrologic model parameters. *Water Resources Research* **39**(8). DOI: 10.1029/2002WR001642

Wilks DS. 2009. Extending logistic regression to provide full-probability-distribution mos forecasts. *Meteorological Applications* **16**(3): 361–368. DOI: 10.1002/met.134

Wood AW, Schaake JC. 2008. Correcting errors in streamflow forecast ensemble mean and spread. *Journal of Hydrometeorology* **9**: 132–148. DOI: 10.1175/2007JHM862.1

Young PC. 1998. Data-based mechanistic modelling of environmental, ecological, economic and engineering systems. *Environmental Modelling and Software* **13**: 105–122.

Young PC. 2001. Data-based mechanistic modelling and validation of rainfall-flow. In *Model Validation: Perspectives in Hydrological Science*, Anderson M, Bates P (eds). J. Wiley: Chichester; 117–161.

Young PC. 2012. Data-based mechanistic modelling: natural philosophy revisited? In *System identification, environmental modelling, and control system design*, Wang L, Garnier H (eds). Springer: London; 321–340.

Young PC, Tomlin CM. 2000. Data-based mechanistic modelling and adaptative flow forecasting. In *Flood forecasting: what does current research offer the practitioner?*, Lees M, Walsh P (eds). British Hydrological Society (BHS) Occasional Paper no. 12. Centre for Ecology and Hydrology on behalf of BHS: Wallingford; 26–40.

Zalachori I, Ramos M-H, Garçon R, Mathevet T, Gailhard J. 2012. Statistical processing of forecasts for hydrological ensemble prediction: a comparative study of different bias correction strategies. *Advances in Science and Research* **8**: 135–141. DOI: 10.5194/asr-8-135-2012

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of this article at the publisher's web site.

Copyright © 2014 John Wiley & Sons, Ltd.