Investigating the interactions between data assimilation and post-processing in hydrological ensemble forecasting

François Bourgin, Maria-Helena Ramos, Guillaume Thirel, Vazken
Andréassian

Accepted Manuscript

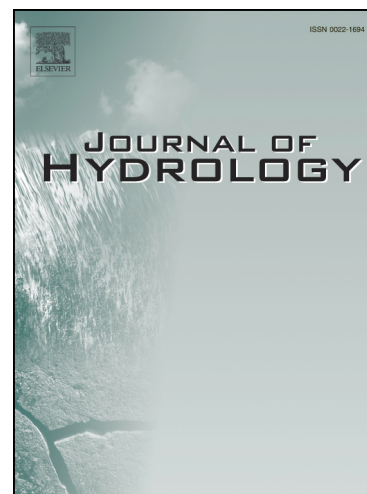Investigating the interactions between data assimilation and post-processing in hydrological ensemble forecasting

F. Bourgin, M-H. Ramos, G. Thirel, V. Andréassian

Investigating the interactions between data assimilation and post-processing in hydrological ensemble forecasting

F. Bourgin[*]

M-H. Ramos

G. Thirel

V. Andréassian

Irstea, UR HBAN, 1 rue Pierre-Gilles de Gennes, CS 10030, F-92761 Antony Cedex, France

Abstract

We investigate how data assimilation and post-processing contribute, either separately or together, to the skill of a hydrological ensemble forecasting system. Based on a large catchment set, we compare four forecasting options: without data assimilation and post-processing, without data assimilation but with post-processing, with data assimilation but without post-processing, and with both data assimilation and post-processing. Our results clearly indicate that both strategies have complementary effects. Data assimilation has mainly a very positive effect on forecast accuracy. Its impact however decreases with increasing lead time. Post-processing, by accounting specifically for hydrological uncertainty, has a very positive and longer lasting effect on forecast reliability. As a consequence, the use of both techniques is

---

[*]Fax number: +33 1 40 96 62 58

Email addresses: francois.bourgin@irstea.fr (F. Bourgin)

recommended in hydrological ensemble forecasting.

1. Introduction

1.1. Addressing uncertainties in hydrological ensemble forecasting

Developing and improving operational hydrological ensemble forecasting systems is a critical step toward better decision-making and risk management. The skill of operational hydrological ensemble forecasting systems is limited by two main sources of uncertainty (Krzysztofowicz, 1999): meteorological uncertainty and hydrological uncertainty. From a pragmatic point of view, the need to properly account for these two main sources of uncertainty arises because (i) a hydrological forecaster has no choice but to rely on uncertain meteorological forecasts; (ii) even with accurate inputs, hydrological forecasts will remain uncertain due to our limited knowledge of initial conditions and the inherent limitations of the forecast model used.

Meteorological uncertainty is commonly addressed by propagating an ensemble (or multi-scenario) input of weather forecasts. For instance, several operational and pre-operational flood forecasting systems across the globe have been set up to be forced by ensemble numerical weather predictions (see Cloke and Pappenberger, 2009, for a review). Addressing the hydrological uncertainty issue is less common, although a general framework of probabilistic forecasting that includes a hydrological post-processing method has been introduced fifteen years ago by Krzysztofowicz (1999). Since then, a number of other hydrological uncertainty processors have been proposed

2

22 (Montanari and Brath, 2004; Montanari and Grossi, 2008; Solomatine and

23 Shrestha, 2009; Coccia and Todini, 2011; Morawietz et al., 2011; Weerts et al.,

24 2011; Ewen and O'Donnell, 2012; Pianosi and Raso, 2012; Smith et al., 2012;

25 Van Steenbergen et al., 2012; Yan et al., 2012), but their use is not widespread

26 for operational ensemble forecasting.

27   Although generally dealt with separately, statistical post-processing and

28 data assimilation (also called real-time model updating in the engineering

29 community) can be intrinsically related in the hydrological forecasting frame-

30 work. Both represent techniques that may be used in a forecasting system

31 to improve the quality of the forecasts (i.e., to provide more accurate and

32 reliable forecasts) and to, ultimately, enhance the usefulness of the forecasts

33 in decision-making. Since forecasting deals with an uncertain future, these

34 techniques aim to bring additional information to the forecast procedure and

35 take into account the various uncertainty sources (or at least the major un-

36 certainty sources) affecting the forecasting chain. This is usually achieved by

37 merging information from model and observations.

38   While data assimilation and post-processing share a general goal, the

39 techniques applied may differ in the practice of hydrological forecasting.

40 These differences usually draw the separation between what is defined as data

41 assimilation and what is defined as post-processing in a modelling framework.

42 The definitions used in this study are the following: we use the term "post-

43 processing" when using the hydrological uncertainty processor (Section 2.4),

44 whose primary purpose is to dress deterministic forecasts with uncertainty

45 based on distributions of past model errors and, this way, build probabilis-

46 tic forecasts. "Data assimilation" refers to techniques applied to perform

3

⁴⁷ the updating of the system before it issues a deterministic forecast. Here

⁴⁸ it concerns the state updating of the hydrological model and a model error

⁴⁹ correction applied to its output (Section 2.3).

⁵⁰   The fact that data assimilation has the potential to improve real-time

⁵¹ streamflow forecasting is widely accepted (see Liu et al., 2012, for a review).

⁵² In contrast to probabilistic and ensemble-based data assimilation methods

⁵³ (e.g., Weerts and El Serafy, 2006; Salamon and Feyen, 2010; Moradkhani

⁵⁴ et al., 2012; Vrugt et al., 2013), deterministic updating schemes are designed

⁵⁵ to improve forecasts without producing probabilistic outputs. They may be

⁵⁶ easier to implement, mainly operationally, but at the price of leaving the

⁵⁷ uncertainty quantification issue unanswered. In these cases, the use of sta-

⁵⁸ tistical post-processing methods together with data assimilation procedures

⁵⁹ provides a way to reduce and quantify the predictive uncertainty in the hy-

⁶⁰ drological forecasts.

⁶¹ *1.2. Integrating uncertainties in hydrological ensemble forecasting*

⁶²   "Ensemble dressing" is an intuitive and operationally-appealing method

⁶³ that allows integration of uncertainties from hydrological modelling and me-

⁶⁴ teorological (ensemble) forcing. The main difference with other ensemble-

⁶⁵ based post-processors (e.g., Wang and Bishop, 2005; Fortin et al., 2006;

⁶⁶ Brown and Seo, 2010; Boucher et al., 2012; Brown and Seo, 2013) is that, for

⁶⁷ ensemble dressing, hydrological modelling errors are assessed separately, and

⁶⁸ later combined with ensemble forecasts. Distributions of modelling errors are

⁶⁹ obtained from long time series of simulated and observed data (i.e., learning

⁷⁰ from the past), and then applied to ensemble forecasts to obtain the total

⁷¹ predictive distribution.

4

[72] In recent studies, the use of ensemble dressing has been implemented [73] and tested to improve the skill of hydrological ensemble forecasting systems. [74] For instance, Reggiani et al. (2009) present a Bayesian ensemble uncertainty [75] processor for medium-range ensemble flow forecasts in the Rhine river basin. [76] Hopson and Webster (2010) use an uncertainty processor based on the k- [77] nearest neighbours (k-NN) resampling method to dress probabilistic medium- [78] range forecasts for two large basins in Bangladesh. Zalachori et al. (2012) [79] compare different strategies based on pre-and post-processing methods to re- [80] move biases in a streamflow ensemble prediction system developed for reser- [81] voir inflow management in French catchments, while Pagano et al. (2013) [82] present a hydrological application of ensemble dressing for 128 catchments [83] in Australia.

[84] The studies mentioned above have in common the fact that they focus [85] on post-processors for operational applications and on the overall evaluation [86] of the quality of post-processed forecasts. Like in the studies that develop [87] and test data assimilation techniques, most of the forecast assessment is on [88] the benefits (in terms of quality) that post-processors or data assimilation [89] may bring to forecast quality (accuracy, reliability, sharpness, etc.) at fixed [90] forecast lead times. Little is known about the interactions between these two [91] components of a forecasting system and the impacts of implementing both [92] post-processing and data assimilation on the performance of the forecasts [93] along the forecast lead times.

1.3. Aim and scope of the study

[95] This study aims to shed light on the interactions between data assimi- [96] lation and post-processing in hydrological ensemble forecasting. We address

5

97 the following questions:

98   1. How does data assimilation impact hydrological ensemble forecasts?

99   2. How does post-processing impact hydrological ensemble forecasts?

100   3. How does data assimilation interact with post-processing to improve the

101   quality and skill of hydrological ensemble forecasts over the forecast lead

102   times?

103   We address these questions with the help of a large set of catchments,

104 making it possible to draw more general and robust conclusions.

2. Data and methods

2.1. Data set

107   A set of 202 unregulated catchments spread over France was used (Fig-

108 ure 1). The catchments represent various hydrological conditions, given the

109 variability in climate, topography, and geology in France. This set includes

110 fast responding Mediterranean basins with intense precipitation as well as

111 larger, groundwater-dominated basins. Some characteristics of the data set

112 are given in Table 1. Catchments were selected to have limited snow in-

113 fluence, since no snowmelt module was used in the hydrological modelling

114 (Section 2.3).

115   Potential evapotranspiration (PE), precipitation, and discharge data were

116 available at hourly time steps over the 1997–2006 period. Temperature inputs

117 originate from the SAFRAN reanalysis (Vidal et al., 2010). PE was estimated

118 using a temperature-based formula (Oudin et al., 2005). Precipitation data

119 come from a reanalysis dataset recently produced by Météo-France based on

6
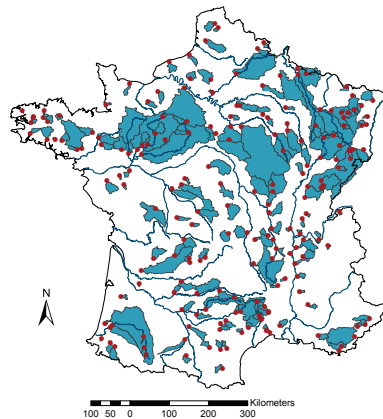
Figure 1: Locations of the 202 French catchments used in this study (dots correspond to the gauging stations, and blue color is catchment areas).

₁₂₀ weather radar and rain gauge network (Tabary et al., 2012). River discharge
₁₂₁ data were extracted from the HYDRO national archive (www.hydro.eaufrance.fr).

₁₂₂ *2.2. PEARP, the Météo-France ensemble forecast*

₁₂₃   A short-range meteorological ensemble prediction system, the Météo-
₁₂₄ France PEARP EPS (Nicolau, 2002), was used to produce hydrological en-
₁₂₅ semble forecasts. The PEARP EPS runs once a day at 18:00 UTC; it has
₁₂₆ 11 members, a 60 h forecast range, and a $0.25°$ (ca. 25 km in France) grid
₁₂₇ resolution. A spatial disaggregation to an 8 km x 8 km grid, which includes
₁₂₈ bias correction, was applied to the PEARP forecasts. Bias correction was
₁₂₉ applied to precipitation forecasts using a multiplying factor obtained from a
₁₃₀ comparison between the mean of the PEARP ensemble and the Météo-France
₁₃₁ SAFRAN reanalysis over a complete year (March 2005 to March 2006). De-
₁₃₂ tails can be found in Thirel et al. (2008). PEARP forecasts were available

7

	Percentiles				
	0.05	0.25	0.50	0.75	0.95
Catchment area (km^2)	31	108	245	653	3761
Mean annual precipitation (mm/y)	725	848	957	1158	1465
Mean annual potential evapotranspiration (mm/y)	645	668	701	745	828
Mean annual runoff (mm/y)	143	232	344	513	964
Q/P ratio	0.18	0.27	0.35	0.47	0.68
P/PE ratio	0.93	1.14	1.36	1.66	2.14
Mean elevation (m)	86	155	306	535	843
Discharge autocorrelation at 48 h	0.28	0.5	0.66	0.81	0.94

Table 1: Characteristics of the 202 catchments. P–precipitation, PE–potential evapotranspiration, Q–discharge.

over the 2005-2009 period, but only the period matching the observed data could be used here, i.e. from August 2005 to December 2006.

PEARP forecasts were already used at the daily time step in recent hydrological studies (Thirel et al., 2008; Randrianasolo et al., 2010). Overall, they showed good quality over France at this time step. The quality for short-term forecasting at hourly time steps (with either raw and post-processed forecasts) is first assessed here.

2.3. The GRP rainfall-runoff forecasting model

The GRP model is a continuous, lumped storage-type model designed for flood forecasting. Its structure was derived from the GR4J model (Perrin et al., 2003) and is composed of a production function and a routing function. The production function consists of a non-linear soil moisture accounting (SMA) reservoir and a volume adjustment coefficient. The routing function

8

146 includes a unit hydrograph and a non-linear routing store. The GRP model

147 uses catchment areal rainfall and PE as inputs; it is parsimonious with three

148 parameters to be calibrated against observed data: one in the production

149 function (the volume adjustment coefficient) and two for the routing function

150 (the base time of the unit hydrograph and the total capacity of the routing

151 store). In this study, the three free parameters were calibrated for each

152 catchment by minimizing the root mean square errors (RMSE) during the

153 first five years of available data (1997– 2001).

154 Importantly, the hourly version of the GRP model uses together two

155 data assimilation procedures for flood forecasting. The first exploits the last

156 available observed discharge to directly update the routing store state, and

157 the second exploits the last relative error to correct the model output with

158 a multiplicative coefficient. More details about the forecasting model GRP

159 and the two assimilation procedures can be found in Berthet et al. (2009).

160 *2.4. Hydrological uncertainty processor*

161 We used a hydrological uncertainty processor (HUP) to evaluate the con-

162 ditional errors of the hydrological model. Only hydrological uncertainty is

163 considered by the HUP here since the model is run with observed weather

164 data. The meteorological uncertainty is subsequently considered through the

165 joint use of the HUP with the PEARP forecasts, as described in Section 2.5.

166 The HUP used here is a data-based and non-parametric method that was

167 applied by Andréassian et al. (2007) to assess model simulation uncertainties

168 and compute empirical uncertainty bounds to flow simulations. Here it is

169 applied to produce probabilistic flow forecasts. The basic idea is to estimate

170 empirical quantiles of relative errors stratified by different flow groups. The

9

171 HUP is trained during the period used for calibrating the parameters of the
172 hydrological model (1997–2001). Note that it is possible that this approach
173 yield optimistic uncertainty estimates, since errors are usually larger on an
174 independent period than during the calibration period. Since forecast error
175 characteristics vary with forecast range when data assimilation is used, the
176 HUP is trained at several lead times separately.

177   For each catchment, the HUP is trained as described below:

Step 1. The hydrological model is run with observed weather data as input and
179   the time series of relative errors is evaluated: Q_{fct}/Q_{obs}, where (Q_{fct},
180   Q_{obs}) are the pairs of discharge forecasts and observations.

Step 2. The time series is stratified into 20 groups according to the magnitude of
182   the Q_{fct}. The limits of each group are fixed so that each group contains
183   the same number of values.

Step 3. Within each group, an empirical distribution of relative errors is defined
185   and 99 quantiles are estimated (corresponding to the percentiles 1%, 2%,
186   ... 98%, 99%).

187   Application of the HUP for another forecast period is described by the
188 last step:

Step 4. Once defined during the training period, the empirical quantiles of rel-
190   ative errors can be applied to any forecast discharge at a certain lead
191   time. The limits of each group are the same as those obtained during
192   the training period. Note that when data assimilation is not used, the
193   empirical quantiles of relative errors are the same whatever the forecast
194   lead time is. Given a discharge forecast Q_{fct}, we first determine the flow

10

₁₉₅ group Q_{fct} belongs to; then Q_{fct} is multiplied by the 99 quantiles of rel-
₁₉₆ ative errors; the 99 values obtained describe the predictive distribution
₁₉₇ at the considered time step and for a given forecast horizon. In cases
₁₉₈ of extrapolation (i.e., when the forecast discharge is out of the range of
₁₉₉ the flow groups defined during the training phase of the HUP), values
₂₀₀ of relative errors from the nearest flow groups (i.e., the lowest or the
₂₀₁ highest flow groups) are used.

₂₀₂ Preliminary studies carried out to compare this approach to other similar
₂₀₃ post-processing approaches suggest that it can yield similar results in terms
₂₀₄ of forecast performance, while being simpler in its application.

₂₀₅ *2.5. Ensemble dressing method: an integrator of the meteorological and hy-*
₂₀₆ *drological uncertainties*

₂₀₇ The ensemble dressing method is used as an integrator of the meteoro-
₂₀₈ logical and hydrological uncertainties. It consists in two steps. Firstly, each
₂₀₉ time an ensemble PEARP forecast is available, the hydrological model is run
₂₁₀ with the ensemble forecast and the HUP is applied, according to Step 4 of
₂₁₁ Section 2.4, to each of the 11 members of the ensemble for each lead time
₂₁₂ considered. Secondly, the 11×99 values obtained at each lead time are
₂₁₃ pooled together and an empirical cumulative distribution is estimated. From
₂₁₄ this distribution, 99 quantiles are retained as the members of the dressed
₂₁₅ ensemble.

₂₁₆ Application and evaluation of the ensemble dressing method for the en-
₂₁₇ semble forecasts is done over an independent period, the 17-month period
₂₁₈ from August 2005 to December 2006.

11

219 *2.6. Experiments*

220 The hydrological ensemble forecast system combines meteorological and *221* streamflow data from observation networks, the Météo-France PEARP en-*222* semble forecast, the GRP rainfall-runoff model with its two data assimilation *223* functions, the hydrological uncertainty processor (HUP) and the ensemble *224* dressing method. Hereafter we will use the term "post-processing" to de-*225* scribe the joint use of the HUP and the ensemble dressing method, while the *226* term "data assimilation" will refer to the two updating techniques used in *227* the GRP model.

228 In order to assess the benefits of data assimilation and post-processing, *229* considered together or separately, different configurations of the forecasting *230* chain were analysed. Our experiments comprise a chain without data assim-*231* ilation and post-processing (NoDA-NoPP), without data assimilation but *232* with post-processing (NoDA-PP), with data assimilation but without post-*233* processing (DA-NoPP), and with both data assimilation and post-processing *234* (DA-PP). The characteristics of the experiments and the acronyms used are *235* given in Table 2.

236 In particular, the NoDA-NoPP experiment corresponds to the situation *237* where the hydrological model is run in simulation mode, i.e., without using *238* recent streamflow observations for data assimilation, and is then driven by *239* the PEARP ensemble forecast when the forecast is issued. When data assim-*240* ilation is used, the state of the routing reservoir of the hydrological model is *241* first updated based on the last observed discharge, and the second procedure *242* is then applied separately at each streamflow ensemble member. This struc-*243* tured analysis allows us to identify the influence of data assimilation and

12

₂₄₄ post-processing separately to assess the benefits of both components when

₂₄₅ used together in the forecasting chain.

	Without data assimilation	With data assimilation
Without post-processing	NoDA-NoPP	DA-NoPP
With post-processing	NoDA-PP	DA-PP

Table 2: Acronyms used for the different experiments used in this study.

₂₄₆ *2.7. Forecast evaluation methods*

₂₄₇ The evaluation of the performance of probabilistic forecasts should reflect

₂₄₈ the different facets of probabilistic forecasts. In this study, the forecasts

₂₄₉ obtained from the four experiments set up (Table 2) were evaluated with

₂₅₀ both deterministic and probabilistic scores. We aimed to assess the influence

₂₅₁ of data assimilation and post-processing on the following characteristics of

₂₅₂ ensemble forecasts: accuracy of the ensemble mean, overall sharpness and

₂₅₃ reliability of the whole ensemble, and overall forecast quality of the ensemble.

₂₅₄ More specifically, we evaluated the accuracy of the ensemble mean val-

₂₅₅ ues with the relative bias (BIAS) and the normalized root-mean-square error

₂₅₆ (NRMSE). To assess the overall reliability of the forecasts, we used the Prob-

₂₅₇ ability Integral Transform (PIT) diagram (see e.g., Laio and Tamea, 2007;

₂₅₈ Thyer et al., 2009) and an index that quantifies deviation from the ideal case,

₂₅₉ the alpha score (Renard et al., 2010). The overall sharpness of the forecasts

₂₆₀ was measured with an index based on the interquartile range that we called

₂₆₁ normalized mean interquartile range (NMIQR). Finally, we assessed the over-

₂₆₂ all forecast quality of the whole ensemble with the mean Continuous Rank

13

263 Probability Skill Score (mean CRPSS). The mean CRPSS is computed with

264 the unconditional streamflow climatology as the reference. These scores are

265 presented in more details in A.

3. Results and discussion

3.1. Forecast accuracy

268 Figure 2 shows the distributions of the two deterministic scores used to

269 assess forecast accuracy: the relative bias (BIAS) and the normalized root-

270 mean-square error (NRMSE). Each score is computed for lead times 6 h,

271 12 h, 24 h and 48 h and for all 202 catchments. The distribution of the 202

272 values is summarized with boxplots.

273 We note that forecast accuracy decreases with increasing lead time for

274 the four experiments. For NoDA experiments (NoDA-NoPP and NoDA-PP),

275 the loss of performance is quite limited: it is only related to the decreasing

276 performance of the PEARP ensemble precipitation forecasts. For DA experi-

277 ments (DA-NoPP and Da-PP), the decrease is stronger and the performances

278 converge toward those of NoDA experiments: the effects of the two DA pro-

279 cedures used in the GRP forecasting model vanish with larger horizons; the

280 decrease in performance of the hydrological model is then added to the losses

281 in performance of the PEARP ensemble precipitation forecasts. Figure 2 also

282 reveals that post-processing does not significantly impact forecast accuracy,

283 whether or not DA is used. DA has a much stronger impact on the ensemble

284 mean values than post-processing, especially for shorter lead times and, to a

285 lower extent, for larger lead times. The two DA procedures used in the GRP

286 forecasting model have been designed to improve the performance of deter-

14

₂₈₇ ministic forecasts and, as it can been seen, they clearly help improving the

₂₈₈ mean of the ensemble forecasts. Post-processing on the other hand primarily

₂₈₉ aims to account for hydrological uncertainty. Its capability to reduce overall

₂₉₀ bias and squared errors in the mean of the ensemble forecasts is limited here.

₂₉₁ Nonetheless, for all lead times, forecast accuracy is best when DA and PP

₂₉₂ are used together, which indicates the benefits of the combined use of data

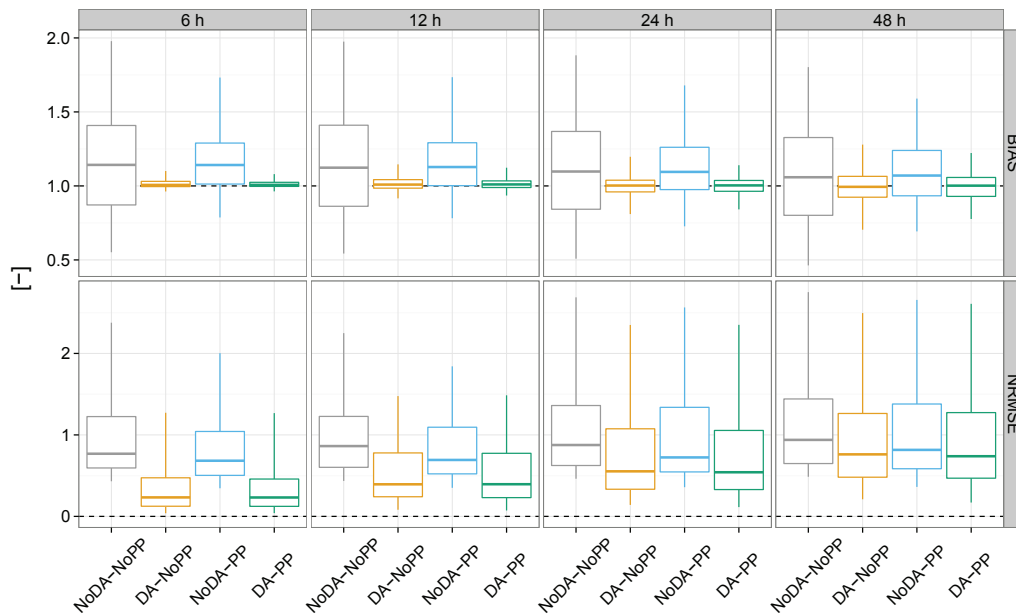₂₉₃ assimilation and post-processing.

Figure 2: Distributions of two deterministic scores, the relative bias (BIAS) and the normalized root-mean-square error (NRMSE), for ensemble streamflow forecasts from the four experiments (see Table 2) and lead times 6 h, 12 h, 24 h and 48 h. Boxplots (5th, 25th, 50th, 75th and 95th percentiles) synthesize the variety of scores over the 202 catchments of the data set.

15

3.2. Reliability

Figure 3 presents the PIT diagrams obtained for each of the 202 catch-
ments, when considering 24 h ahead ensemble forecasts. Since similar figures
were obtained for the other lead times (not shown).

From Figures 3a and 3b, it can be seen that most of the curves are almost
horizontal straight lines, while they would follow the bisector (black lines in
the graphs) in the ideal case of reliable ensemble predictions. Figures 3a
and 3b clearly reveal that the raw ensembles are lacking reliability for all
of the catchments. The impact of post-processing on reliability is apparent
when looking at the results on Figures 3c and 3d: the curves of the ensem-
ble streamflow forecasts with post-processing follow the ideal situation much
more closely than the curves shown in Figures 3a and 3b (ensemble stream-
flow forecasts without post-processing). It means that the overall reliability
of the ensembles is clearly improved with post-processing and this for both
cases, with and without DA. A comparison of solely Figures 3c and 3d con-
firms also the positive impact of data assimilation on the reliability of the
ensembles: the PIT curves of the dressed ensembles are substantially closer
to the diagonal (perfect reliability) when DA is applied.

The PIT diagrams convey a visual evaluation of the overall reliability of
probabilistic forecasts. To quantify it, we used the alpha score, a reliability
index that measures the deviation of the PIT curves from the ideal situation.
Figure 4 presents the distributions of the alpha scores obtained for each
experiment over the 202 catchments. Results in Figure 4 confirm the visual
evaluation obtained with the PIT diagrams: the two experiments that do
not account for hydrological uncertainty (NoDA-NoPP and DA-NoPP) lack
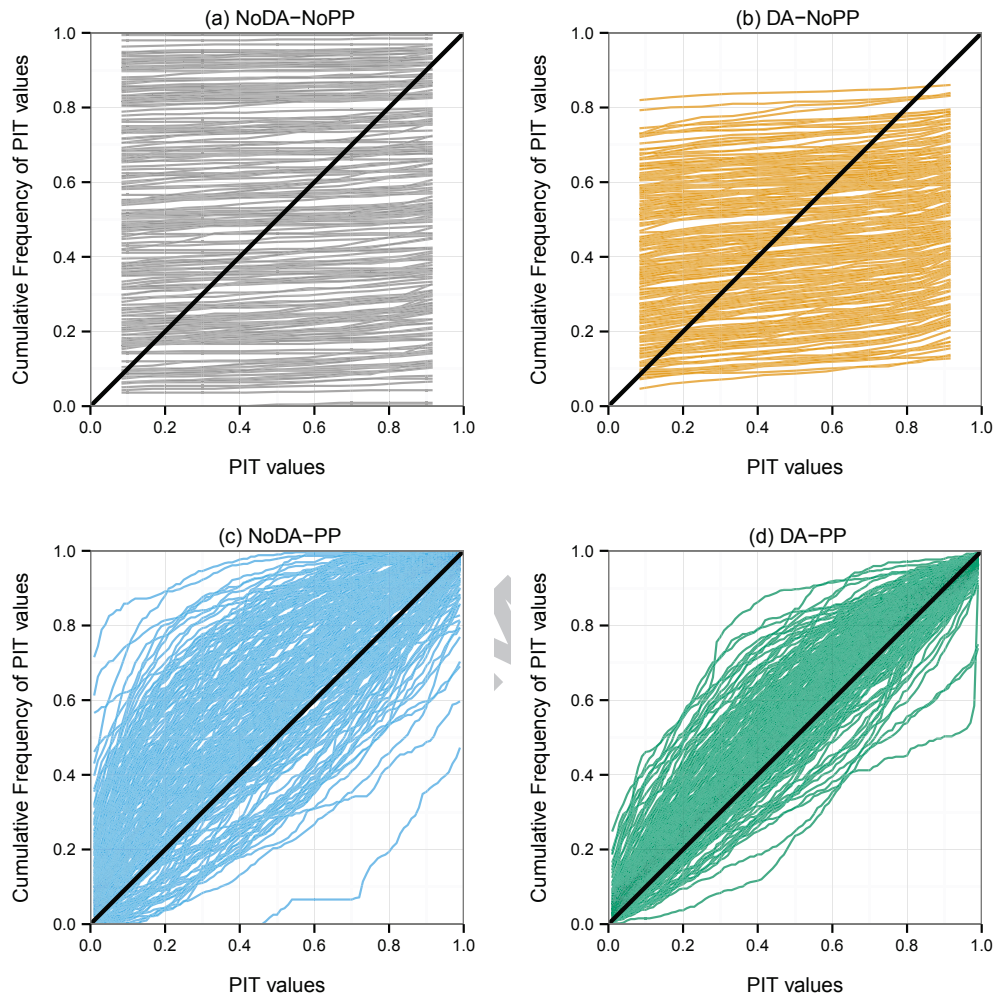
16

Figure 3: PIT diagrams of the 24 h ahead streamflow ensemble forecasts from the four experiments (see Table 2). Each line represents one of the 202 catchments of the data set.

319 reliability. Their alpha values are almost always below 0.5, while the alpha
320 values obtained when hydrological uncertainty is taken into account (NoDA-
321 PP and DA-PP) are almost always higher than 0.5. The benefits of DA is
322 also apparent when comparing, on one hand NoDA-NoPP and DA-NoPP,
323 and on the other hand NoDA-PP and DA-PP, although it can be also seen
324 that DA alone (comparing NoDA-NoPP to DA-NoPP) cannot correct under
325 dispersion of the ensemble forecasts. Post-processing is then a necessary step
326 to achieve reliable forecasts in the forecasting chain analysed.

327 These results suggest that for the 202 catchments studied the spread
328 obtained by propagating solely the precipitation ensembles into the hydro-
329 logical model is too small to properly reflect the range of forecast errors. The
330 deterministic data assimilation strategy used here is effective in improving
331 the reliability of the ensemble forecasts, but it is not self-sufficient to correct
332 the under dispersion of the streamflow ensemble forecasts as revealed by the
333 PIT diagrams in Figure 3 and the alpha scores in Figure 4. This is a strong
334 indication that the hydrological uncertainty issue should be specifically ad-
335 dressed in order to improve the overall reliability of hydrological ensemble
336 forecasts.

337 *3.3. Sharpness*

338 Sharpness is a desirable characteristic of any probabilistic forecast. The
339 sharper the forecast, the less uncertain it is, and thus the more information is
340 conveyed. The four experiments we used made it possible to investigate how
341 meteorological and hydrological uncertainties interact and affect sharpness.
342 Figure 5 shows the distributions of a sharpness index, the normalized mean
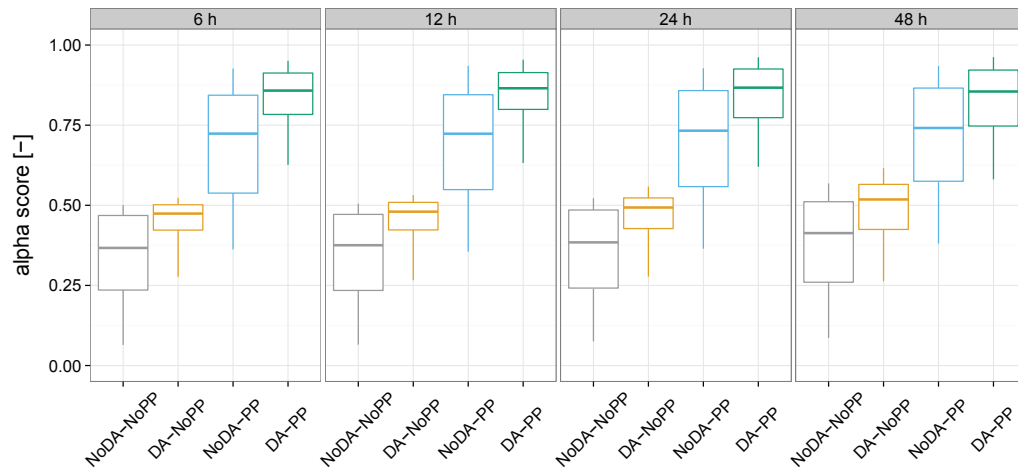343 interquartile range (NMIQR), over 202 catchments.

18

Figure 4: Distributions of the alpha score reliability index for streamflow ensemble forecasts from the four experiments (see Table 2) and for lead times 6 h, 12 h, 24 h and 48 h. Boxplots (5th, 25th, 50th, 75th and 95th percentiles) synthesize the variety of scores over the 202 catchments of the data set. Perfect score is 1.0.

It can be seen that the ensemble spreads of three experiments, NoDA-NoPP, DA-NoPP and DA-PP, increase significantly with increasing lead time, while it is more stable over lead times for the experiment NoDA-PP. For NoDa-NoPP and DA-NoPP, the median value of NMIQR over the 202 catchments raises in a very close behaviour for both experiments, from around 0.05 for 6 h ahead forecasts to 0.13 for 48 h ahead forecasts. For the experiment DA-PP, the increase in the median values is much more important: from 0.07 at 6 h to 0.32 at 48 h. These results indicate that forecast uncertainty increases with increasing lead time as the result of increasing meteorological uncertainty alone (NoDA-NoPP and DA-NoPP) or as the result of increasing meteorological and hydrological uncertainties considered together and with DA (DA-PP). Comparing DA-NoPP and DA-PP reveals the impact of post-processing: taking into account hydrological uncertainty leads to more spread and less sharpness in ensemble forecasts. Comparing NoDA-NoPP and DA-NoPP shows that the propagation of meteorological uncertainty has a rather similar impact on ensemble sharpness whether or not DA is used to update the states of the forecasting model. Remarkably, the ensemble spreads obtained without DA but with post-processing (NoDA-PP) is stable across the lead times with a median value over the 202 catchments around 0.52. This is because statistical post-processing reflects the large errors obtained when the forecasting model does not use DA (see Figure 2). In this case, the spread obtained when taking hydrological uncertainty into account is so large that the increasing spread of the PEARP ensemble forecasts with increasing lead time has no visible impact on the spread of the post-processed ensemble: hydrological uncertainty dominates meteorological uncertainty.

20

369   Not surprisingly, sharper forecasts are obtained when only meteorological
370 uncertainty is taken into account (NoPP experiments). This is to the detri-
371 ment of reliability: ensemble forecasts with only meteorological uncertainty
372 are sharper but not reliable, reflecting the presence of under dispersion (as
373 shown in Section 3.2). The use of post-processing (PP experiments) leads
374 to ensembles that are more spread out because they attempt to handle hy-
375 drological uncertainty and reflect hydrological forecast errors. Ensembles are
376 thus less sharp but, on the other hand, achieve reliability. At this point, it
377 should be remembered that sharp but unreliable forecasts should be consid-
378 ered with caution. Unreliable forecasts can convey a wrong impression of
379 certainty that results from having neglected one or several important sources
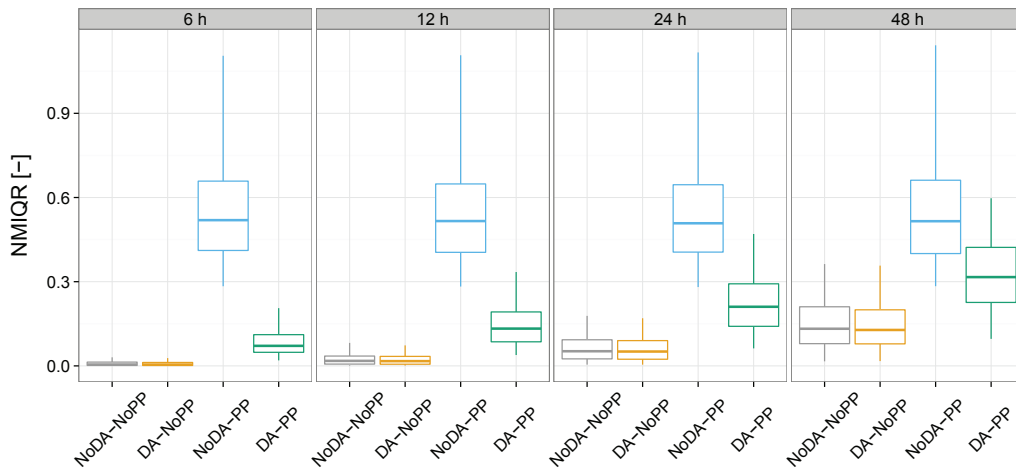380 of uncertainty.

Figure 5: Distributions of the normalized mean interquartile range (NMIQR) for stream-
flow ensemble forecasts from the four experiments (see Table 2) and for lead times 6 h,
12 h, 24 h and 48 h. Boxplots (5th, 25th, 50th, 75th and 95th percentiles) synthesize the
variety of scores over the 202 catchments of the data set. Perfect score is 0.

21

381 *3.4. Mean CRPSS*

382   The analysis of the impacts of data assimilation and post-processing on
383 two important characteristics of probabilistic forecasts, reliability and sharp-
384 ness, showed that post-processing was necessary to improve reliability, but
385 at the cost of lower sharpness, i.e., greater ensemble spread and uncertainty,
386 even if sharpness could be improved with the application of a data assimi-
387 lation procedure. We now turn our attention to the mean CRPSS, a proba-
388 bilistic score that provides an assessment of the overall quality of ensemble
389 forecasts.

390   Figure 6 shows the distributions of the mean CRPSS over 202 catchments.
391 We note that performance decreases with increasing lead time for the two ex-
392 periments with data assimilation: median values of the CRPSS are equal to
393 0.84 (DA-NoPP) and 0.87 (DA-PP) for 6 h range forecasts, and equal to 0.45
394 (DA-NoPP) and 0.57 (DA-PP) for 48 h range forecasts. Mean CRPSS values
395 of the two experiments without data assimilation decrease only slightly but
396 are much lower than values obtained with data assimilation (median values
397 around 0.10 for NoDA-NoPP and around 0.45 for NoDA-PP). This is espe-
398 cially true for shorter lead times and, to a lower extent, for larger lead times.
399 Furthermore, the comparison with the reference climatology shows that data
400 assimilation alone is sufficient to generate skillfull forecasts for more than
401 95% of the catchments for lead times up to 24 h, but post-processing (DA-
402 PP) is necessary to achieve forecasts that have better overall performance
403 than climatology at 48 h.

404   These results show the general added value of data assimilation and post-
405 processing to the overall quality of ensemble forecasts. When evaluating the

22

overall quality of ensemble forecasts with the CRPSS, the benefits in terms of reliability overcome the loss of sharpness that results from accounting for hydrological uncertainty. The streamflow ensemble forecasts that explicitly account for both sources of uncertainty, meteorological and hydrological uncertainties, through post-processing, while reducing as much as possible hydrological uncertainty, here through data assimilation, are the most skillfull forecasts.
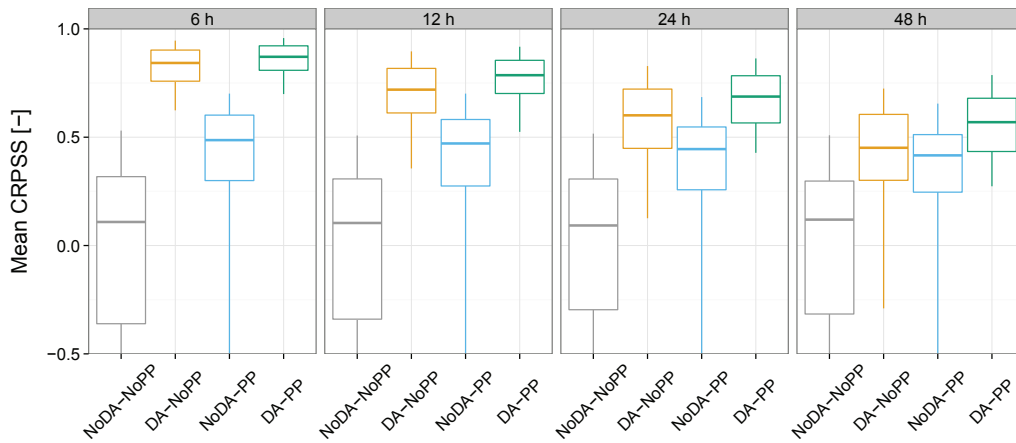

Figure 6: Distributions of the mean CRPSS for streamflow ensemble forecasts from the four experiments (see Table 2) and for lead times 6 h, 12 h, 24 h and 48 h. Boxplots (5th, 25th, 50th, 75th and 95th percentiles) synthesize the variety of scores over the 202 catchments of the data set. Perfect score is 1.0.

4. Summary and conclusions

We investigated the relative contributions of data assimilation and post-processing to the skill of hydrological ensemble forecasts. The study assessed the benefits of data assimilation and post-processing with the help of

23

four configurations of a short-range hydrological ensemble forecasting system: without data assimilation and post-processing (NoDA-NoPP), without data assimilation but with post-processing (NoDA-PP), with data assimilation but without post-processing (DA-NoPP), and with both data assimilation and post-processing (DA-PP).

We applied deterministic and probabilistic scores to streamflow forecasts of a large catchment set which brought into light the main general conclusions listed below:

- We verify the well-known fact that short-range hydrological forecasts benefit from data assimilation. Data assimilation has a strong impact on improving the quality of the ensemble mean, and a much lesser effect on the variability of the ensemble members (i.e., their spread).

- The benefits of a simple yet efficient hydrological uncertainty processor to improve the reliability and the overall quality of the short-range hydrological ensemble forecasts were demonstrated. Post-processing has a strong impact on forecast reliability.

- The benefits of the combined use of data assimilation and post-processing were demonstrated: both contribute to achieve reliable and sharp forecasts, with impacts acting differently according to the target lead time. The stronger impact on forecast reliability comes from the use of post-processing. Adding data assimilation to the system helps in improving sharpness and reliability at all lead times, with higher gains in performance at shorter lead times.

We acknowledge some limitations. It was only possible to evaluate the

24

forecasting chain over a 17-month period of ensemble forecasts, since this was the common period between observations and forecasts we had available. Furthermore, PEARP ensembles are ran only once a day, which limits the number of hourly evaluation pairs. For these reasons, it was not possible to evaluate flows over specific flooding thresholds . However, with increasing data archives, we expect that such an issue will be treated in future work.

Our study considered only one data assimilation technique (state up-dating with error output correction) and one post-processing method (ensemble dressing with hydrological errors) together with one rainfall-runoff model forecasting (GRP model). There are several other techniques and models in the literature that could also be tested using the methodology presented here. For instance, a comparison between different configurations of the method used, or different hydrological uncertainty processors, including methods that take into account the autocorrelation of errors (e.g., Schoups and Vrugt, 2010) could be investigated. Besides, while a bias correction was applied to the PEARP forecasts, a more sophisticated pre-processor (see e.g., Verkade et al., 2013) could be used to further investigate how meteorological and hydrological biases interact and contribute to the quality of the final hydrological ensemble.

Also, the effectiveness of a data assimilation technique or a post- processing method (and hence the choice of the procedures to operate in a forecasting system) is affected by different sources of uncertainties present in a flow forecasting system, including the forcing data, initial conditions, parameter uncertainty and model structural uncertainty. In our study, we followed the works of Krzysztofowicz (1999) and focused on a decomposition of the total

25

uncertainty into meteorological and hydrological uncertainty. Observational or parameter uncertainties were thus not explicitly considered. Additional sources of uncertainty may however affect the performance of data assimilation techniques and post-processors, as well as the way they interact in the forecasting system. Further investigations would be necessary to better assess the extent to which this may affect forecast quality.

Although our findings may be related to the configuration used, they are based on common techniques and on the study of a large set of catchments, which helps in giving robustness and generality to the results obtained. The study also shows that, for a given system configuration, it is interesting to analyse how data assimilation and/or post-processing techniques set up to improve forecast quality affect the attributes of the forecasts and interact to provide overall good forecasts. The aim of a forecaster may then be to achieve a good combination of hydrological model, data assimilation and post-processing procedures that results in an overall good quality of his/her operational system (eventually over specific space and time scales of interest), rather than to search for the best data assimilation technique or post-processor available, without taking into account how they will interact between them and with the probabilistic forecasting system as a whole.

Despite those limitations, our results strongly suggest that data assimilation and post-processing techniques based on hydrological uncertainty processors should be more widely tested to foster their implementation in pre-operational and operational hydrological ensemble forecasting systems and their use in real-time probabilistic forecasting. The use of both strategies is highly recommended since they have complementary effects: data assimila-

26

491 tion has a very positive effect on forecast accuracy, and thus helps reduce
492 hydrological uncertainty, but its impact diminishes with lead time, while
493 post-processing, by accounting for hydrological uncertainty, has a very posi-
494 tive and longer lasting effect on forecast reliability.

27

$_{495}$ A. Evaluation scores

$_{496}$ The evaluation scores used in this article are defined and briefly described
$_{497}$ below. For more details, the reader may refer to Wilks (2011).

$_{498}$ *A.1. Relative Bias*

$_{499}$ The relative bias (BIAS) is defined as the ratio between the mean of
$_{500}$ deterministic forecasts and the mean of observations,

$$\text{BIAS} = \frac{\sum_{k=1}^{N} Q_{fct}(k)}{\sum_{k=1}^{N} Q_{obs}(k)} \tag{1}$$

$_{501}$ where $\left(Q_{fct}(k), Q_{obs}(k)\right)$ is the kth of N pairs of deterministic forecasts and
$_{502}$ observations.

$_{503}$ Values higher (lower) than 1 indicate an overall overestimation (underes-
$_{504}$ timation) of the observed values.

$_{505}$ *A.2. Normalized root-mean-square error*

$_{506}$ The root-mean-square error (RMSE) is a widely used measure of accuracy
$_{507}$ for point forecasts,

$$\text{RMSE} = \left[\frac{1}{N} \sum_{k=1}^{N} \left(Q_{fct}(k) - Q_{obs}(k)\right)^2\right]^{1/2} \tag{2}$$

$_{508}$ where $\left(Q_{fct}(k), Q_{obs}(k)\right)$ is the kth of N pairs of forecasts and observations.

$_{509}$ The lower the RMSE, the better. For a perfect deterministic forecast,
$_{510}$ RMSE=0.

$_{511}$ The normalized root-mean-square error (NRMSE) is obtained by dividing
$_{512}$ the RMSE by the mean runoff. The use of a non-dimensional score facilitates
$_{513}$ the comparison of the results obtained over different catchments.

28

514 *A.3. PIT diagram and alpha score*

515 The Probability Integral Transform (PIT) diagram is a graphical tool
516 used to assess the reliability of probabilistic forecasts (Gneiting et al., 2007;
517 Laio and Tamea, 2007). The PIT diagram corresponds to the empirical
518 cumulative distribution of the PIT values, which are defined for each pair
519 of forecasts and observations as the value that the cumulative predictive
520 distribution F reaches at the observation, $p^{obs} = F(Q_{obs})$. It is analogous to
521 a cumulated version of the rank histogram. If the forecasts are reliable, the
522 PIT values follow a uniform distribution on the interval $[0, 1]$ and the PIT
523 curve is close to the 1:1 line. Reliability of the probabilistic forecasts implies
524 that the observations should not be preferentially located in specific parts
525 of the predictive distributions, but instead should uniformly span the whole
526 predictive range.

527 The alpha score is an index proposed by Renard et al. (2010) to reflect
528 the overall reliability of probabilistic forecasts. The alpha score is directly
529 related to the PIT diagram. It is defined as $1 - 2A$, where A is the area
530 between the bisector and the PIT curve,

$$A = \frac{1}{N} \sum_{k=1}^{N} \left| p^{obs}(k) - p^{th}(k) \right| \tag{3}$$

531 and where $\left(p^{obs}(k), p^{th}(k)\right)$ is the kth of N pairs of observed and theoretical
532 PIT values.

533 The alpha score ranges from 0 to 1. 0 indicates poor reliability while
534 values close to 1 indicate perfect reliability.

29

A.4. Normalized mean interquartile range

To assess the sharpness of probabilistic forecasts, we defined the mean interquartile range (MIQR) as the mean of the interquartile range of forecasts over the evaluation data. The interquartile range, defined as the range between the upper quartile (75th percentile) and the lower quartile (25th percentile) of a distribution, is a robust measure of the spread of a distribution. MIQR is computed as

$$\text{MIQR} = \frac{1}{N} \sum_{k=1}^{N} \left(Q_{fct}^{75}(k) - Q_{fct}^{25}(k) \right) \tag{4}$$

where $\left(Q_{fct}^{25}(k), Q_{fct}^{75}(k) \right)$ is the kth of N pairs of quartiles of the forecasts.

Similarly to the NRMSE, we divided the MIQR by the mean runoff to obtain a non-dimensional score.

A.5. Mean CRPS and mean CRPSS

For a forecast-observation evaluation pair, the Continuous Rank Probability Score (CRPS) (e.g., Matheson and Winkler, 1976; Gneiting et al., 2007) measures the quadratic distance between two cumulative distribution functions, the cumulative predictive distribution $F(x)$ and a Heaviside function based on the observed value $\mathbb{1}\{Q_{obs} \leqslant x\}$):

$$\text{CRPS}(F, Q_{obs}) = \int_{-\infty}^{\infty} \left(F(x) - \mathbb{1}\{Q_{obs} \leqslant x\} \right)^2 \, dx \tag{5}$$

The mean CRPS, $\overline{\text{CRPS}}$, is the average value of the CRPS over the N pairs of evaluation data:

$$\overline{\text{CRPS}} = \frac{1}{N} \sum_{k=1}^{N} \text{CRPS(k)} \tag{6}$$

30

553 The mean Continuous Rank Probability Skill Score (CRPSS) is a skill
554 score based on the CRPS. Skill scores (SS) are used to assess the relative
555 quality of two forecasting systems. They are generally defined as:

$$SS = 1 - \frac{\text{Score}^A}{\text{Score}^B} \tag{7}$$

556 where Score^A and Score^B are the scores of the forecasting system A and
557 B respectively. The forecasting system B is usually termed the reference
558 forecast.

559 Climatology is commonly used as a reference. To compute the mean
560 CRPSS with the unconditional climatology as the reference, an unconditional
561 streamflow ensemble forecast is first obtained from the empirical distribution
562 of all observed discharges over the evaluation period, and then used for all
563 forecast occasions.

571 References

572 Andréassian, V., Lerat, J., Loumagne, C., Mathevet, T., Michel, C., Oudin,
573 L., Perrin, C., 2007. What is really undermining hydrologic science today?
574 Hydrological Processes 21, 2819–2822. doi:10.1002/hyp.6854.

31

Berthet, L., Andreassian, V., Perrin, C., Javelle, P., 2009. How crucial is it to account for the antecedent moisture conditions in flood forecasting? Comparison of event-based and continuous approaches on 178 catchments. Hydrology and Earth System Sciences 13, 819–831.

Boucher, M.A., Tremblay, D., Delorme, L., Perreault, L., Anctil, F., 2012. Hydro-economic assessment of hydrological forecasting systems. Journal of Hydrology 416, 133–144. doi:10.1016/j.jhydrol.2011.11.042.

Brown, J.D., Seo, D.J., 2010. A Nonparametric Postprocessor for Bias Correction of Hydrometeorological and Hydrologic Ensemble Forecasts. Journal of Hydrometeorology 11, 642–665. doi:10.1175/2009JHM1188.1.

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

Cloke, H.L., Pappenberger, F., 2009. Ensemble flood forecasting: A review. Journal of Hydrology 375, 613–626. doi:10.1016/j.jhydrol.2009.06.005.

Coccia, G., Todini, E., 2011. Recent developments in predictive uncertainty assessment based on the model conditional processor approach. Hydrology and Earth System Sciences 15, 3253–3274. doi:10.5194/hess-15-3253-2011.

Ewen, J., O'Donnell, G., 2012. Prediction intervals for rainfall-runoff models: raw error method and split-sample validation. Hydrology Research 43, 637–648. doi:10.2166/nh.2012.038.

Fortin, V., Favre, A.C., Said, M., 2006. Probabilistic forecasting from ensemble prediction systems: Improving upon the best-member method by using

32

a different weight and dressing kernel for each member. Quarterly Journal of the Royal Meteorological Society 132, 1349–1369. doi:10.1256/qj.05.167.

Gneiting, T., Balabdaoui, F., Raftery, A.E., 2007. Probabilistic forecasts, calibration and sharpness. Journal of the Royal Statistical Society Series B-Statistical Methodology 69, 243–268. doi:10.1111/j.1467-9868.2007.00587.x.

Hopson, T.M., Webster, P.J., 2010. A 1-10-Day Ensemble Forecasting Scheme for the Major River Basins of Bangladesh: Forecasting Severe Floods of 2003-07. Journal of Hydrometeorology 11, 618–641. doi:10.1175/2009jhm1006.1.

Krzysztofowicz, R., 1999. Bayesian theory of probabilistic forecasting via deterministic hydrologic model. Water Resources Research 35, 2739–2750.

Laio, F., Tamea, S., 2007. Verification tools for probabilistic forecasts of continuous hydrological variables. Hydrology and Earth System Sciences 11, 1267–1277.

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

Matheson, J.E., Winkler, R.L., 1976. Scoring rules for continu-

33

₆₂₀ ous probability distributions. Management Science 22, 1087–1096.
₆₂₁ doi:10.1287/mnsc.22.10.1087.

₆₂₂ Montanari, A., Brath, A., 2004. A stochastic approach for assessing the
₆₂₃ uncertainty of rainfall-runoff simulations. Water Resources Research 40,
₆₂₄ W01106. doi:10.1029/2003wr002540.

₆₂₅ Montanari, A., Grossi, G., 2008. Estimating the uncertainty of hydrological
₆₂₆ forecasts: A statistical approach. Water Resources Research 44, W00B08.
₆₂₇ doi:10.1029/2008wr006897.

₆₂₈ Moradkhani, H., DeChant, C.M., Sorooshian, S., 2012. Evolution of en-
₆₂₉ semble data assimilation for uncertainty quantification using the particle
₆₃₀ filter-Markov chain Monte Carlo method. Water Resources Research 48,
₆₃₁ W12520. doi:10.1029/2012wr012144.

₆₃₂ Morawietz, M., Xu, C.Y., Gottschalk, L., Tallaksen, L.M., 2011. Systematic
₆₃₃ evaluation of autoregressive error models as post-processors for a prob-
₆₃₄ abilistic streamflow forecast system. Journal of Hydrology 407, 58–72.
₆₃₅ doi:10.1016/j.jhydrol.2011.07.007.

₆₃₆ Nicolau, J., 2002. Short-range ensemble forecasting, in: Proceedings
₆₃₇ WMO/CBS Technical Conferences On Data Processing and Forecasting
₆₃₈ Systems, Cairns, Australia, pp. 2–3.

₆₃₉ Oudin, L., Hervieu, F., Michel, C., Perrin, C., Andreassian, V., Anctil,
₆₄₀ F., Loumagne, C., 2005. Which potential evapotranspiration input for
₆₄₁ a lumped rainfall-runoff model? Part 2 - Towards a simple and efficient

34

642 potential evapotranspiration model for rainfall-runoff modelling. Journal
643 of Hydrology 303, 290–306. doi:10.1016/j.jhydrol.2004.08.026.

644 Pagano, T.C., Shrestha, D.L., Wang, Q.J., Robertson, D., Hapuarachchi,
645 P., 2013. Ensemble dressing for hydrological applications. Hydrological
646 Processes 27, 106–116. doi:10.1002/hyp.9313.

647 Perrin, C., Michel, C., Andreassian, V., 2003. Improvement of a parsimo-
648 nious model for streamflow simulation. Journal of Hydrology 279, 275–289.
649 doi:10.1016/s0022-1694(03)00225-7.

650 Pianosi, F., Raso, L., 2012. Dynamic modeling of predictive uncertainty
651 by regression on absolute errors. Water Resources Research 48, W03516.
652 doi:10.1029/2011wr010603.

653 Randrianasolo, A., Ramos, M.H., Thirel, G., Andreassian, V., Martin, E.,
654 2010. Comparing the scores of hydrological ensemble forecasts issued by
655 two different hydrological models. Atmospheric Science Letters 11, 100–
656 107. doi:10.1002/asl.259.

657 Reggiani, P., Renner, M., Weerts, A.H., van Gelder, P., 2009. Uncertainty
658 assessment via Bayesian revision of ensemble streamflow predictions in the
659 operational river Rhine forecasting system. Water Resources Research 45,
660 W02428. doi:10.1029/2007wr006758.

661 Renard, B., Kavetski, D., Kuczera, G., Thyer, M., Franks, S.W., 2010. Un-
662 derstanding predictive uncertainty in hydrologic modeling: The challenge
663 of identifying input and structural errors. Water Resources Research 46,
664 W05521. doi:10.1029/2009wr008328.

35

[665] Salamon, P., Feyen, L., 2010. Disentangling uncertainties in dis-
[666] tributed hydrological modeling using multiplicative error models and
[667] sequential data assimilation. Water Resources Research 46, W12501.
[668] doi:10.1029/2009wr009022.

[669] Schoups, G., Vrugt, J.A., 2010. A formal likelihood function for param-
[670] eter and predictive inference of hydrologic models with correlated, het-
[671] eroscedastic, and non-Gaussian errors. Water Resources Research 46,
[672] W10531. doi:10.1029/2009wr008933.

[673] Smith, P.J., Beven, K.J., Weerts, A.H., Leedal, D., 2012. Adaptive correction
[674] of deterministic models to produce probabilistic forecasts. Hydrology and
[675] Earth System Sciences 16, 2783–2799. doi:10.5194/hess-16-2783-2012.

[676] Solomatine, D.P., Shrestha, D.L., 2009. A novel method to estimate model
[677] uncertainty using machine learning techniques. Water Resources Research
[678] 45, W00B11. doi:10.1029/2008wr006839.

[679] Tabary, P., Dupuy, P., L'Henaff, G., Gueguen, C., Moulin, L., Laurantin, O.,
[680] Merlier, C., Soubeyroux, J.M., 2012. A 10-year (1997-2006) reanalysis of
[681] Quantitative Precipitation Estimation over France: methodology and first
[682] results, in: Weather Radar and Hydrology, IAHS. pp. 255–260.

[683] Thirel, G., Rousset-Regimbeau, F., Martin, E., Habets, F., 2008.
[684] On the Impact of Short-Range Meteorological Forecasts for Ensemble
[685] Streamflow Predictions. Journal of Hydrometeorology 9, 1301–1317.
[686] doi:10.1175/2008jhm959.1.

36

687 Thyer, M., Renard, B., Kavetski, D., Kuczera, G., Franks, S.W., Srikan-
688   than, S., 2009. Critical evaluation of parameter consistency and
689   predictive uncertainty in hydrological modeling: A case study using
690   Bayesian total error analysis. Water Resources Research 45, W00B14.
691   doi:10.1029/2008wr006825.

692 Van Steenbergen, N., Ronsyn, J., Willems, P., 2012. A non-parametric
693   data-based approach for probabilistic flood forecasting in support of
694   uncertainty communication. Environmental Modelling & Software 33.
695   doi:10.1016/j.envsoft.2012.01.013.

696 Verkade, J.S., Brown, J.D., Reggiani, P., Weerts, A.H., 2013. Post-processing
697   ecmwf precipitation and temperature ensemble reforecasts for operational
698   hydrologic forecasting at various spatial scales. Journal of Hydrology 501,
699   73–91. Times Cited: 00.

700 Vidal, J.P., Martin, E., Franchisteguy, L., Baillon, M., Soubeyroux, J.M.,
701   2010. A 50-year high-resolution atmospheric reanalysis over France with
702   the Safran system. International Journal of Climatology 30, 1627–1644.
703   doi:10.1002/joc.2003.

704 Vrugt, J.A., ter Braak, C.J.F., Diks, C.G.H., Schoups, G., 2013. Hydrologic
705   data assimilation using particle Markov chain Monte Carlo simulation:
706   Theory, concepts and applications. Advances in Water Resources 51, 457–
707   478. doi:10.1016/j.advwatres.2012.04.002.

708 Wang, X., Bishop, C., 2005. Improvement of ensemble reliability with a new

37

[709] dressing kernel. Quarterly Journal of the Royal Meteorological Society 131,

[710] 965–986. doi:10.1256/qj.04.120.

[711] Weerts, A.H., El Serafy, G.Y.H., 2006. Particle filtering and en-

[712] semble Kalman filtering for state updating with hydrological concep-

[713] tual rainfall-runoff models. Water Resources Research 42, W09403.

[714] doi:10.1029/2005wr004093.

[715] Weerts, A.H., Winsemius, H.C., Verkade, J.S., 2011. Estimation of predic-

[716] tive hydrological uncertainty using quantile regression: examples from the

[717] National Flood Forecasting System (England and Wales). Hydrology and

[718] Earth System Sciences 15, 255–265. doi:10.5194/hess-15-255-2011.

[719] Wilks, D.S., 2011. Statistical methods in the atmospheric sciences. 3rd ed.,

[720] Academic, Oxford.

[721] Yan, J., Liao, G.Y., Gebremichael, M., Shedd, R., Vallee, D.R.,

[722] 2012. Characterizing the uncertainty in river stage forecasts condi-

[723] tional on point forecast values. Water Resources Research 48, W12509.

[724] doi:10.1029/2012wr011818.

[725] Zalachori, I., Ramos, M.H., Garçon, R., Mathevet, T., Gailhard, J., 2012.

[726] Statistical processing of forecasts for hydrological ensemble prediction: a

[727] comparative study of different bias correction strategies. Advances in Sci-

[728] ence and Research 8, 135–141. doi:10.5194/asr-8-135-2012.

38

Figure 3. PIT diagrams of the 24 h ahead streamflow ensemble forecasts from the four a-d) experiments (see Table 2). Each line represents one of the 202 catchments of the data set.

Highlights

- Data assimilation and post-processing impact hydrologic ensemble forecasts' skill.

- Data assimilation has a strong impact on forecast accuracy.

- Post-processing has a strong impact on forecast reliability.

- The combined benefits of data assimilation and post-processing were demonstrated.

- We recommend the use of both data assimilation and post-processing in forecasting.

1