



**HAL**  
open science

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

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

## ► To cite this version:

François Bourgin, Maria-Helena Ramos, Guillaume Thirel, Vazken Andréassian. Investigating the interactions between data assimilation and post-processing in hydrological ensemble forecasting. *Journal of Hydrology*, 2014, 519 (Part D), pp.2775-2784. 10.1016/j.jhydrol.2014.07.054 . hal-02600032

**HAL Id: hal-02600032**

**<https://hal.inrae.fr/hal-02600032>**

Submitted on 1 Dec 2020

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

## Accepted Manuscript

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

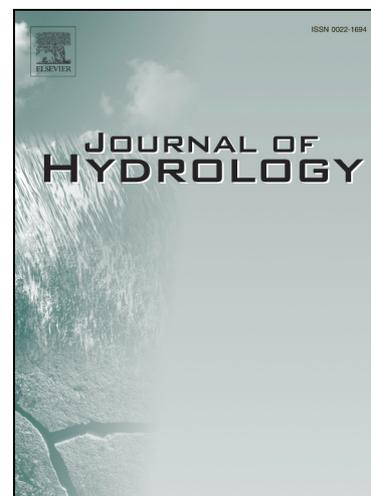
F. Bourgin, M-H. Ramos, G. Thirel, V. Andréassian

PII: S0022-1694(14)00588-5

DOI: <http://dx.doi.org/10.1016/j.jhydrol.2014.07.054>

Reference: HYDROL 19785

To appear in: *Journal of Hydrology*



Please cite this article as: Bourgin, F., Ramos, M-H., Thirel, G., Andréassian, V., Investigating the interactions between data assimilation and post-processing in hydrological ensemble forecasting, *Journal of Hydrology* (2014), doi: <http://dx.doi.org/10.1016/j.jhydrol.2014.07.054>

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

# 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.

*Key words:* hydrological ensemble forecasting, data assimilation, post-processing, ensemble dressing, uncertainty propagation

---

## 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

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

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.

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.

### 94 *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

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.

## 105 **2. Data and methods**

### 106 *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

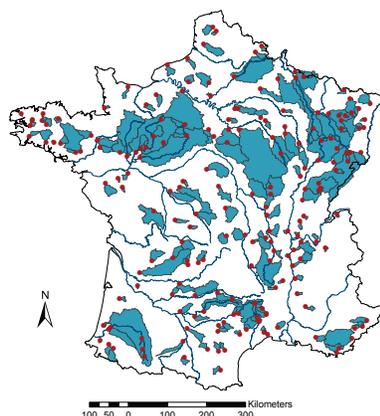


Figure 1: Locations of the 202 French catchments used in this study (dots correspond to the gauging stations, and blue color is catchment areas).

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](http://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  $0.25^\circ$  (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

	Percentiles				
	0.05	0.25	0.50	0.75	0.95
Catchment area (km <sup>2</sup> )	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.

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

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

### 140 2.3. The GRP rainfall-runoff forecasting model

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

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

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:

178 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.

181 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.

184 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:

189 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

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  $11 \times 99$  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.

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

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

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.

### 266 **3. Results and discussion**

#### 267 *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-

287 ministic forecasts and, as it can be 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.

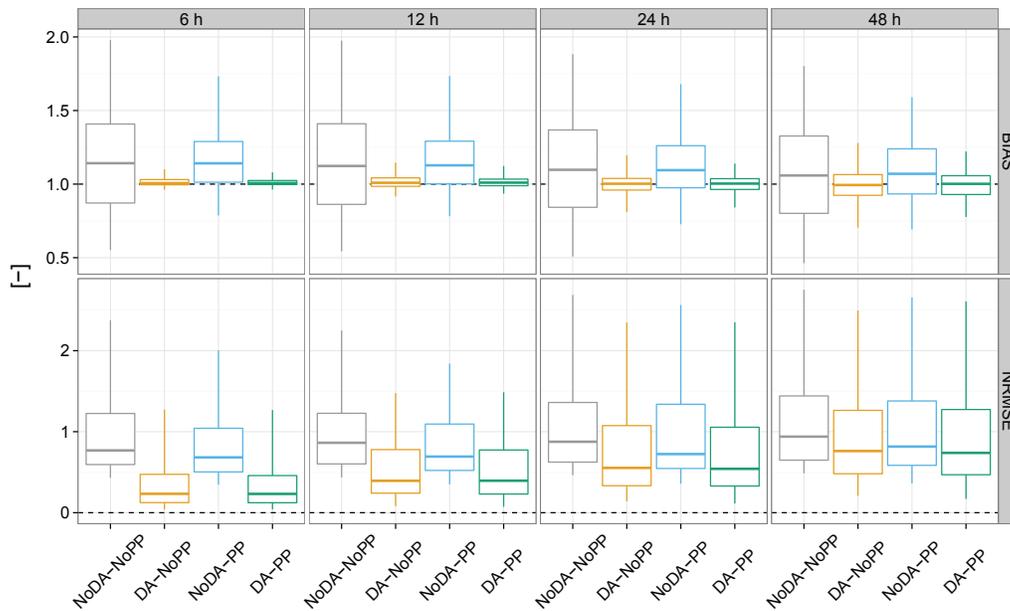


Figure 2: Distributions of two deterministic scores, the relative bias (BIAS) and the normalized root-mean-square error (NRMSE), for ensemble streamflow forecasts from the four experiments (see Table 2) and lead times 6 h, 12 h, 24 h and 48 h. Boxplots (5th, 25th, 50th, 75th and 95th percentiles) synthesize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

294 *3.2. Reliability*

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

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

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

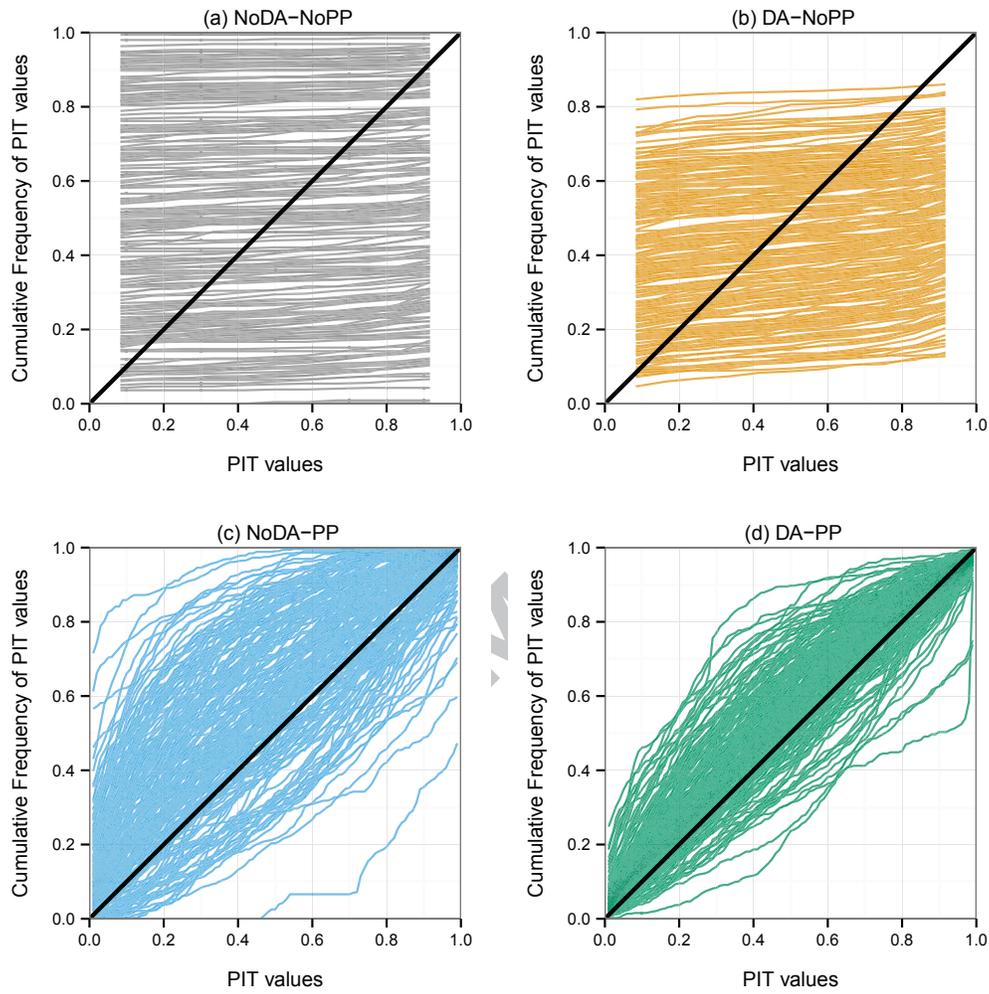


Figure 3: PIT diagrams of the 24 h ahead streamflow ensemble forecasts from the four experiments (see Table 2). Each line represents one of the 202 catchments of the data set.

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.

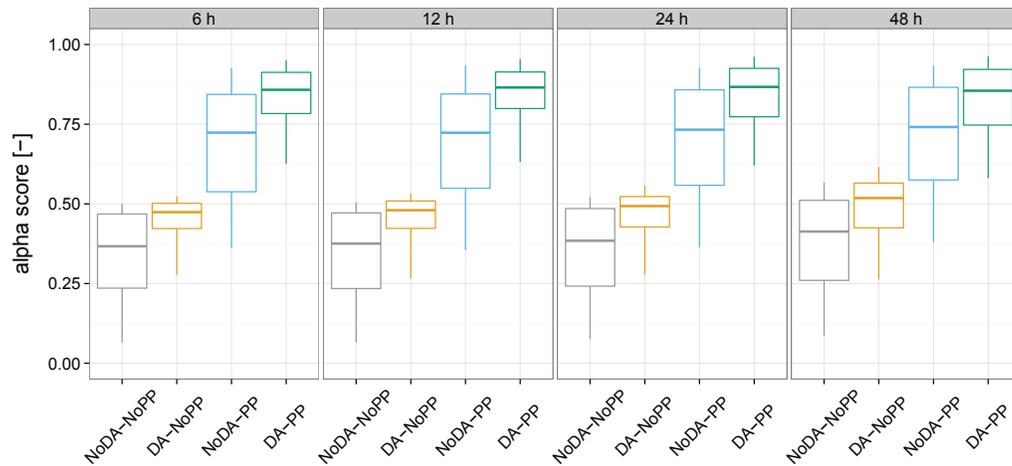


Figure 4: Distributions of the alpha score reliability index for streamflow ensemble forecasts from the four experiments (see Table 2) and for lead times 6 h, 12 h, 24 h and 48 h. Boxplots (5th, 25th, 50th, 75th and 95th percentiles) synthesize the variety of scores over the 202 catchments of the data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

344 It can be seen that the ensemble spreads of three experiments, NoDA-  
345 NoPP, DA-NoPP and DA-PP, increase significantly with increasing lead time,  
346 while it is more stable over lead times for the experiment NoDA-PP. For  
347 NoDa-NoPP and DA-NoPP, the median value of NMIQR over the 202 catch-  
348 ments raises in a very close behaviour for both experiments, from around 0.05  
349 for 6 h ahead forecasts to 0.13 for 48 h ahead forecasts. For the experiment  
350 DA-PP, the increase in the median values is much more important: from  
351 0.07 at 6 h to 0.32 at 48 h. These results indicate that forecast uncertainty  
352 increases with increasing lead time as the result of increasing meteorological  
353 uncertainty alone (NoDA-NoPP and DA-NoPP) or as the result of increas-  
354 ing meteorological and hydrological uncertainties considered together and  
355 with DA (DA-PP). Comparing DA-NoPP and DA-PP reveals the impact of  
356 post-processing: taking into account hydrological uncertainty leads to more  
357 spread and less sharpness in ensemble forecasts. Comparing NoDA-NoPP  
358 and DA-NoPP shows that the propagation of meteorological uncertainty has  
359 a rather similar impact on ensemble sharpness whether or not DA is used  
360 to update the states of the forecasting model. Remarkably, the ensemble  
361 spreads obtained without DA but with post-processing (NoDA-PP) is stable  
362 across the lead times with a median value over the 202 catchments around  
363 0.52. This is because statistical post-processing reflects the large errors ob-  
364 tained when the forecasting model does not use DA (see Figure 2). In this  
365 case, the spread obtained when taking hydrological uncertainty into account  
366 is so large that the increasing spread of the PEARP ensemble forecasts with  
367 increasing lead time has no visible impact on the spread of the post-processed  
368 ensemble: hydrological uncertainty dominates meteorological uncertainty.

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.

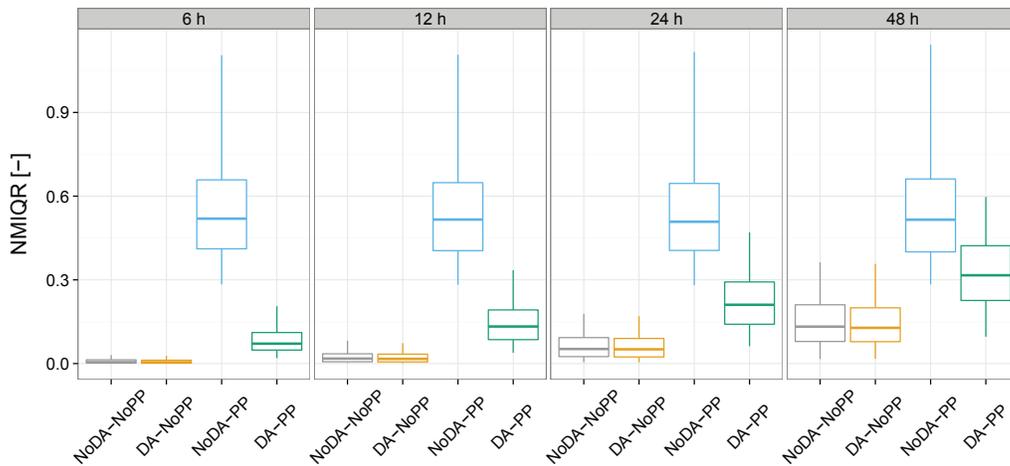


Figure 5: Distributions of the normalized mean interquartile range (NMIQR) for stream-flow ensemble forecasts from the four experiments (see Table 2) and for lead times 6 h, 12 h, 24 h and 48 h. Boxplots (5th, 25th, 50th, 75th and 95th percentiles) synthesize the variety of scores over the 202 catchments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

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

406 overall quality of ensemble forecasts with the CRPSS, the benefits in terms  
 407 of reliability overcome the loss of sharpness that results from accounting for  
 408 hydrological uncertainty. The streamflow ensemble forecasts that explicitly  
 409 account for both sources of uncertainty, meteorological and hydrological un-  
 410 certainties, through post-processing, while reducing as much as possible hy-  
 411 drological uncertainty, here through data assimilation, are the most skillfull  
 412 forecasts.

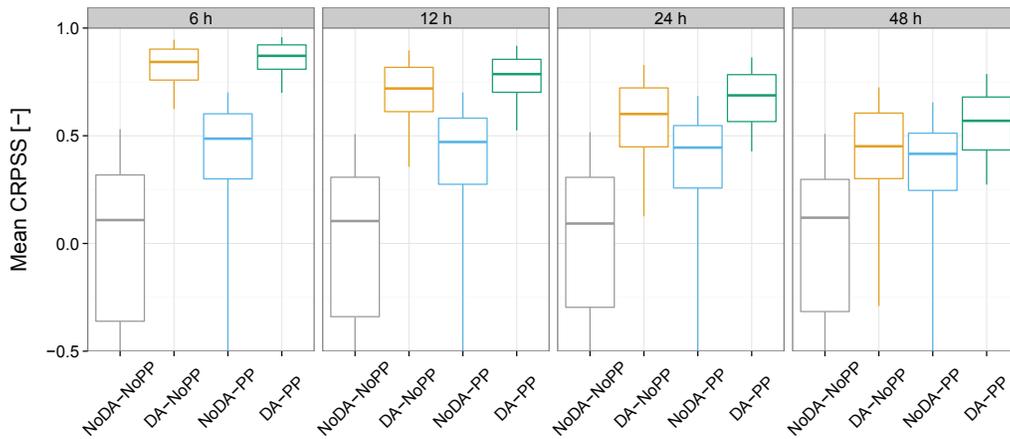


Figure 6: Distributions of the mean CRPSS for streamflow ensemble forecasts from the four experiments (see Table 2) and for lead times 6 h, 12 h, 24 h and 48 h. Boxplots (5th, 25th, 50th, 75th and 95th percentiles) synthesize the variety of scores over the 202 catchments of the data set. Perfect score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

#### 413 4. Summary and conclusions

414 We investigated the relative contributions of data assimilation and post-  
 415 processing to the skill of hydrological ensemble forecasts. The study as-  
 416 sessed the benefits of data assimilation and post-processing with the help of

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

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

- 425 • We verify the well-known fact that short-range hydrological forecasts  
426 benefit from data assimilation. Data assimilation has a strong impact  
427 on improving the quality of the ensemble mean, and a much lesser effect  
428 on the variability of the ensemble members (i.e., their spread).
- 429 • The benefits of a simple yet efficient hydrological uncertainty processor  
430 to improve the reliability and the overall quality of the short-range  
431 hydrological ensemble forecasts were demonstrated. Post-processing  
432 has a strong impact on forecast reliability.
- 433 • The benefits of the combined use of data assimilation and post-processing  
434 were demonstrated: both contribute to achieve reliable and sharp fore-  
435 casts, with impacts acting differently according to the target lead time.  
436 The stronger impact on forecast reliability comes from the use of post-  
437 processing. Adding data assimilation to the system helps in improving  
438 sharpness and reliability at all lead times, with higher gains in perfor-  
439 mance at shorter lead times.

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

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

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

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

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

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

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

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.

ACCEPTED MANUSCRIPT

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,

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

501 where  $(Q_{fct}(k), Q_{obs}(k))$  is the  $k$ th 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,

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

508 where  $(Q_{fct}(k), Q_{obs}(k))$  is the  $k$ th 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.

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,

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

531 and where  $(p^{obs}(k), p^{th}(k))$  is the  $k$ th 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.

535 *A.4. Normalized mean interquartile range*

536 To assess the sharpness of probabilistic forecasts, we defined the mean  
 537 interquartile range (MIQR) as the mean of the interquartile range of fore-  
 538 casts over the evaluation data. The interquartile range, defined as the range  
 539 between the upper quartile (75th percentile) and the lower quartile (25th per-  
 540 centile) of a distribution, is a robust measure of the spread of a distribution.  
 541 MIQR is computed as

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

542 where  $(Q_{fct}^{25}(k), Q_{fct}^{75}(k))$  is the  $k$ th of  $N$  pairs of quartiles of the forecasts.

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

545 *A.5. Mean CRPS and mean CRPSS*

546 For a forecast-observation evaluation pair, the Continuous Rank Proba-  
 547 bility Score (CRPS) (e.g., Matheson and Winkler, 1976; Gneiting et al., 2007)  
 548 measures the quadratic distance between two cumulative distribution func-  
 549 tions, the cumulative predictive distribution  $F(x)$  and a Heaviside function  
 550 based on the observed value  $\mathbf{1}\{Q_{obs} \leq x\}$ :

$$\text{CRPS}(F, Q_{obs}) = \int_{-\infty}^{\infty} (F(x) - \mathbf{1}\{Q_{obs} \leq x\})^2 dx \quad (5)$$

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

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

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:

$$SS = 1 - \frac{\text{Score}^A}{\text{Score}^B} \quad (7)$$

556 where  $\text{Score}^A$  and  $\text{Score}^B$  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.

## 564 Acknowledgments

565 The authors thank Météo-France for providing the meteorological data  
566 and Banque HYDRO for the hydrological data. The financial support of  
567 SCHAPI to the first author is also gratefully acknowledged.

568 The authors thank Dr. Thomas Pagano, two anonymous reviewers, and  
569 the Guest Editor Hamid Moradkhani for their critical and constructive eval-  
570 uation of the manuscript, which helped improving its quality.

## 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.

- 575 Berthet, L., Andreassian, V., Perrin, C., Javelle, P., 2009. How crucial is  
576 it to account for the antecedent moisture conditions in flood forecasting?  
577 Comparison of event-based and continuous approaches on 178 catchments.  
578 *Hydrology and Earth System Sciences* 13, 819–831.
- 579 Boucher, M.A., Tremblay, D., Delorme, L., Perreault, L., Anctil, F., 2012.  
580 Hydro-economic assessment of hydrological forecasting systems. *Journal*  
581 *of Hydrology* 416, 133–144. doi:10.1016/j.jhydrol.2011.11.042.
- 582 Brown, J.D., Seo, D.J., 2010. A Nonparametric Postprocessor for Bias Cor-  
583 rection of Hydrometeorological and Hydrologic Ensemble Forecasts. *Jour-*  
584 *nal of Hydrometeorology* 11, 642–665. doi:10.1175/2009JHM1188.1.
- 585 Brown, J.D., Seo, D.J., 2013. Evaluation of a nonparametric post-processor  
586 for bias correction and uncertainty estimation of hydrologic predictions.  
587 *Hydrological Processes* 27, 83–105. doi:10.1002/hyp.9263.
- 588 Cloke, H.L., Pappenberger, F., 2009. Ensemble flood forecasting: A review.  
589 *Journal of Hydrology* 375, 613–626. doi:10.1016/j.jhydrol.2009.06.005.
- 590 Coccia, G., Todini, E., 2011. Recent developments in predictive uncertainty  
591 assessment based on the model conditional processor approach. *Hydrology*  
592 *and Earth System Sciences* 15, 3253–3274. doi:10.5194/hess-15-3253-2011.
- 593 Ewen, J., O'Donnell, G., 2012. Prediction intervals for rainfall-runoff models:  
594 raw error method and split-sample validation. *Hydrology Research* 43,  
595 637–648. doi:10.2166/nh.2012.038.
- 596 Fortin, V., Favre, A.C., Said, M., 2006. Probabilistic forecasting from ensem-  
597 ble prediction systems: Improving upon the best-member method by using

- 598 a different weight and dressing kernel for each member. *Quarterly Journal*  
599 *of the Royal Meteorological Society* 132, 1349–1369. doi:10.1256/qj.05.167.
- 600 Gneiting, T., Balabdaoui, F., Raftery, A.E., 2007. Probabilistic fore-  
601 casts, calibration and sharpness. *Journal of the Royal Statistical So-*  
602 *ciety Series B-Statistical Methodology* 69, 243–268. doi:10.1111/j.1467-  
603 9868.2007.00587.x.
- 604 Hopson, T.M., Webster, P.J., 2010. A 1-10-Day Ensemble Forecasting  
605 Scheme for the Major River Basins of Bangladesh: Forecasting Se-  
606 vere Floods of 2003-07. *Journal of Hydrometeorology* 11, 618–641.  
607 doi:10.1175/2009jhm1006.1.
- 608 Krzysztofowicz, R., 1999. Bayesian theory of probabilistic forecasting via  
609 deterministic hydrologic model. *Water Resources Research* 35, 2739–2750.
- 610 Laio, F., Tamea, S., 2007. Verification tools for probabilistic forecasts of  
611 continuous hydrological variables. *Hydrology and Earth System Sciences*  
612 11, 1267–1277.
- 613 Liu, Y., Weerts, A.H., Clark, M., Franssen, H.J.H., Kumar, S., Moradkhani,  
614 H., Seo, D.J., Schwanenberg, D., Smith, P., van Dijk, A., van Velzen, N.,  
615 He, M., Lee, H., Noh, S.J., Rakovec, O., Restrepo, P., 2012. Advancing  
616 data assimilation in operational hydrologic forecasting: progresses, chal-  
617 lenges, and emerging opportunities. *Hydrology and Earth System Sciences*  
618 16, 3863–3887. doi:10.5194/hess-16-3863-2012.
- 619 Matheson, J.E., Winkler, R.L., 1976. Scoring rules for continu-

- 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

- 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.

- 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.

- 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

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.

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.

ACCEPTED MANUSCRIPT

**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.