Hydrology and Earth System Sciences

A past discharge assimilation system for ensemble streamflow forecasts over France – Part 2: Impact on the ensemble streamflow forecasts

G. Thirel[1,2], E. Martin[1], J.-F. Mahfouf[1], S. Massart[3], S. Ricci[3], F. Regimbeau[4], and F. Habets[5]

[1]CNRM-GAME/GMME, Météo-France, URA 1357, Toulouse, France
[2]IES, Joint Research Centre, European Commission, Ispra, Italy
[3]URA CNRS/CERFACS No. 1875, Toulouse, France
[4]Direction de la climatologie, Météo-France, Toulouse, France
[5]UMR Sisyphe, UPMC, ENSMP, CNRS, Paris, France

Abstract. The use of ensemble streamflow forecasts is developing in the international flood forecasting services. Ensemble streamflow forecast systems can provide more accurate forecasts and useful information about the uncertainty of the forecasts, thus improving the assessment of risks. Nevertheless, these systems, like all hydrological forecasts, suffer from errors on initialization or on meteorological data, which lead to hydrological prediction errors. This article, which is the second part of a 2-part article, concerns the impacts of initial states, improved by a streamflow assimilation system, on an ensemble streamflow prediction system over France. An assimilation system was implemented to improve the streamflow analysis of the SAFRAN-ISBA-MODCOU (SIM) hydro-meteorological suite, which initializes the ensemble streamflow forecasts at Météo-France. This assimilation system, using the Best Linear Unbiased Estimator (BLUE) and modifying the initial soil moisture states, showed an improvement of the streamflow analysis with low soil moisture increments. The final states of this suite were used to initialize the ensemble streamflow forecasts of Météo-France, which are based on the SIM model and use the European Centre for Medium-range Weather Forecasts (ECMWF) 10-day Ensemble Prediction System (EPS). Two different configurations of the assimilation system were used in this study: the first with the classical SIM model and the second using improved soil physics in ISBA. The effects of the assimilation system on the ensem-

ble streamflow forecasts were assessed for these two configurations, and a comparison was made with the original (i.e. without data assimilation and without the improved physics) ensemble streamflow forecasts. It is shown that the assimilation system improved most of the statistical scores usually computed for the validation of ensemble predictions (RMSE, Brier Skill Score and its decomposition, Ranked Probability Skill Score, False Alarm Rate, etc.), especially for the first few days of the time range. The assimilation was slightly more efficient for small basins than for large ones.

1 Introduction

The development of meteorological ensemble prediction systems (EPSs) during recent years has allowed their use to spread into many related topics. This is especially the case in hydrometeorology, where EPSs are increasingly used to produce ensemble streamflow forecasts. The assessment of uncertainty is a key point for hydrological forecasters, and enables them to take risk-based decisions. Many projects have therefore been launched on this topic, like the Hydrologic Ensemble Prediction EXperiment (HEPEX, Schaake et al., 2006). The HEPEX project aims to bring together meteorologists and hydrologists to address the issue of hydrological forecast uncertainty, including uncertainty in the meteorological forcing, the hydrological modelling, and the final user needs. In Europe, the European Flood Alert System (EFAS) provides flood alerts to several European partner countries (Ramos et al., 2007). These alerts are based on ensemble streamflow predictions using the European Centre for

Medium-range Weather Forecasts (ECMWF) EPS and permit action to be taken several days before an event. Other operational or non-operational hydrological ensemble suites are also being developed in the USA (Wood et al., 2005), The Netherlands (Van Andel et al., 2008), Belgium (Van den Bergh and Roulin, 2010), Switzerland (Zappa et al., 2008), England (He et al., 2009), and France (Rousset-Regimbeau et al., 2007; Thirel et al., 2008). Many operational ensemble streamflow forecast systems are reviewed in Cloke and Pappenberger (2009).

However, such systems are rarely updated with discharge observations that would allow them to better fit the actual situation. This may result in poor performance of the system, especially for the short-range forecast horizons, and forecasters have difficulties in using the predictions adequately as they find it hard to understand and trust a prediction whose starting point is far from the observations. For this reason, the assimilation of streamflow observations is, along with other possibilities like satellite soil moisture data, a very promising way to improve the quality of streamflow predictions. The use of satellite observations of soil moisture is quite common in the hydrological scientific community, its aim being to improve the simulation of soil moisture states in the model or even streamflow simulations (Lakshmi, 2004; Zaitchik et al., 2008; Crow et al., 2009). However, the scales and frequencies of availability of such data are not necessarily suited to hydrological predictions on a large scale and the use of observed discharges can provide a promising way to improve such forecasts. Streamflow observations are regularly available on fixed points and, furthermore, they can be used without any post-processing. Several studies are thus developing on this topic, and aim to use past discharges in order to improve hydrological states of models. For example, Komma et al. (2008) used an updating method based on the Ensemble Kalman Filter in a rainfall-runoff model. This operational system used streamflow observations and updated soil moisture states. A case study on a catchment showed a significant decrease in the error, and an increase in the Nash criterion. Pauwels and De Lannoy (2009) analysed different methods for updating the soil moisture states of a conceptual rainfall-runoff model by using discharges observations for a small-scale catchment. Other studies have been performed by Aubert et al. (2003) and Seo et al. (2009) but it still remains difficult to find ensemble streamflow prediction systems that use streamflow assimilation, more especially on a large scale.

The aim of this study was to assess the effects of using initial states improved by a past discharge assimilation system as described in Thirel et al. (2010a), on ensemble streamflow forecasts over France. Météo-France performs such forecasts in operational mode for the whole of France with the hydro-meteorological model SAFRAN-ISBA-MODCOU (SIM). These forecasts use a real-time SIM-analysis chain which is forced by an atmospheric analysis, and provide a daily hydro-meteorological analysis. The final states of this SIM-analysis chain initialize the ensemble streamflow forecasts. However, no updating or assimilation of streamflow observations is used to keep the hydrological analysis close to observations. Thus results may be subject to drift or events may be missed. That is why a data assimilation system using past discharges and incrementing the soil moisture states of the model was implemented in testing mode, in order to improve the hydro-meteorological analysis (Thirel et al., 2010a). The impacts of the data assimilation system on mid-term range ensemble streamflow forecasts are assessed in this study by means of a set of statistical scores. Thus, the relevancy of our approach for improving the ensemble streamflow forecasts can be assessed. The impacts on the ensemble aspects are studied here as well as on the errors relative to observations.

The first part of this study describes the SIM hydrometeorological model used and the way ensemble streamflow predictions are set up from this system with the ECMWF (European Centre for Medium-range Weather Forecasts) EPS. Then, the streamflow assimilation system will be described, and a summary of its validation will be given. Finally, a large set of statistical scores will be used to quantify the impacts of the assimilation system on the 10-day SIM-ECMWF ensemble streamflow system, first for 148 assimilated stations and then for 49 independent stations.

2 SIM ensemble streamflow forecasts

2.1 The SIM model

SAFRAN-ISBA-MODCOU (SIM) is a hydro-meteorological suite developed at Météo-France. This distributed model simulates the evolution of soil moisture over France, and models streamflows for a total of 881 stations. SIM is based on the ISBA (Interactions between Soil, Biosphere and Atmosphere) Land Surface Model (Noilhan and Planton, 1989), which simulates water and energy fluxes between the soil and the atmosphere for 9892 8-km grid meshes distributed over France. The MODCOU (MODèle COUplé, Coupled Model, Ledoux et al., 1989) hydrological model simulates the spatial and temporal evolution of aquifers on the Seine and Rhône basins. For the other basins, the amount of water going from soil to rivers was set to a constant based on low flows (Quintana Seguí et al., 2009). An optional exponential profile of the hydraulic conductivity of the soil can be used (Quintana Seguí et al., 2009) and it has been shown to improve the dynamics of floods. SIM has been validated over a 10-year period for 881 French stations (Habets et al., 2008) and gave realistic water and energy budgets, streamflow, aquifer levels and snowpack simulations.

2.2 The meteorological EPS used

The medium-range ECMWF (European Centre for Medium-range Weather Forecasts) EPS was used to produce ensemble streamflow forecasts for this study. The version used was the 51-member 10-day meteorological EPS (Buizza et al., 1999). Both temperature and precipitation members were used to produce the hydrological members. In order to provide these data on the 8-km ISBA grid, temperature and precipitation were downscaled (Rousset-Regimbeau et al., 2007). First, a spatial downscaling was performed. Then, the classical OACI gradient was used for the temperature, and an altitude gradient was calibrated for the precipitation. The downscaling resulted in a good spatial distribution and mean of precipitation when compared with observations, but showed a weak spread of the precipitation ensemble (method and validation in Rousset-Regimbeau et al., 2007).

2.3 The SIM ensemble streamflow predictions

The SIM ensemble streamflow predictions were implemented and validated against SIM-analysis streamflows (i.e. streamflows produced by using meteorological observations) by Rousset-Regimbeau et al. (2007). It showed good overall results for both high and low flows. The performance of these operational streamflows forecasts relative to observed discharges is assessed in Thirel et al. (2010b) and, despite the small spread of the ensemble, shows encouraging results for medium discharges, but poor performance for high flows. The SIM ensemble streamflow prediction system was constructed in two steps. First, the SIM-analysis suite was run in order to produce the initial states of the ensemble chain. This SIM-analysis suite used the SAFRAN-analysis parameters to produce a hydrological analysis. The suite was run every day in real-time in continuous mode, producing hydric states of the soil, the rivers and the aquifers. These final states were used to initialize the SIM ensemble streamflow predictions. Thus, initialized by these states, ensemble streamflow forecasts were produced every day by forcing SIM with the ECMWF EPS temperature and precipitation members.

3 The past discharge assimilation system

A streamflow assimilation system was implemented in the SIM-analysis suite and validated (see the first part of this paper: Thirel et al., 2010a). Its role is to improve the streamflow simulation of this chain, thus providing better initial states to the SIM ensemble streamflow system. Based on the Best Linear Unbiased Estimator (BLUE), this assimilation system uses streamflow observations and updates the ISBA soil moisture states in order to improve the SIM streamflow simulations. The use or not of improved physics in ISBA, the exponential profile of the hydraulic conductivity (Quintana Seguí et al., 2009), was tested for each state variable choice.

The Jacobian matrix used by the BLUE represents the dependence of the MODCOU discharges on variations in the ISBA soil moisture initial states. This matrix was estimated for every daily assimilation by making runs of SIM with small perturbations on its soil moisture initial states (background state) and deducing the variation of the streamflow simulation for this soil moisture variation. The perturbed runs had to be performed separately on each of the sub-basins of a large basin in order to be able to deduce the impacts of the soil moisture in each of the upstream sub-basins on a downstream discharge simulation. The variance of observation error was simply estimated by a function of the square of the observed discharge. The variance of background error was estimated by studying the effects of SAFRAN temperature and precipitation errors on SIM soil moisture.

A set of 148 assimilated stations was studied for the period from 11 March 2005 to 30 September 2006 in order to validate the assimilation system for 6 configurations (3 different variable states, and for each one, 2 physics of the model). The assimilation was performed every day, with daily discharge observations. The assimilation system showed a significant improvement in streamflow simulations on average for the 148 assimilated stations, with an increased Nash criterion, and decreased RMSE and bias for each configuration. Moreover, the increments imposed by the system remained low, showing that the model's fluxes were only slightly modified. The assimilation proved to be more efficient on wet soils, which is perfectly consistent with the fact that soil moisture is not the most important factor of discharge production during dry periods. This was confirmed by the fact that, during dry periods, only rare and tiny adjustments were made to the soil moisture. Moreover, the effects of assimilation were not significant on sub-basins where an aquifer layer was simulated in MODCOU because, for such basins, these aquifer layers have more impact on streamflow simulations than the soil moisture has.

The use of the exponential profile of hydraulic conductivity showed an improvement in the data assimilation effects, with lower increments, RMSE, and bias. Moreover, except for the experiment with the two separate layers in the variable state, the Nash criterion was improved. The experiment combining layer 2 and layer 3 soil moistures and the exponential profile of the hydraulic conductivity (called IS_2) had the best performance, seen in its good Nash criterion, best RMSE and lowest increments. This conclusion was confirmed by the study of scores for a selection of 49 independent stations. When compared to the same experiment without the improved physics (IS_1), it was seen that, although the Nash criterion was equivalent, RMSE and increments were lower when the exponential profile was used, showing the interest of using this option. It is important not to change the ISBA fluxes too much. The improvement in the performance of the model for non-assimilated stations shows the interest of using a distributed model. Indeed, for lumped models, it is impossible to improve discharge simulations with an

asssimilation system if no observation are available for these stations. A more complete description and validation of the system is available in the first part of this paper (Thirel et al., 2010a).

4   Impacts of the assimilation system on the ensemble streamflow forecasts

The ensemble streamflow forecasts initialized by IS_1 did not use the exponential profile but those initialized by IS_2 did use this profile for the sake of consistency with the data assimilation system physics. In this study, 3 ensemble streamflow forecast systems are compared: the original streamflow predictions (without assimilation and without the improved physics), the ensemble streamflow predicitons using IS_1 (with assimilation but without the improved physics), and the ensemble streamflow predictions using IS_2 (with assimilation and with the improved physics). The study period was from 11 March 2005 to 30 September 2006 and the scores were averaged for the 148 assimilated stations already used in Thirel et al. (2010a). Scores were computed for the mean of the ensemble (Ratio-Spread and Ratio-RMSE), for the exceeding of thresholds (Brier Skill Score and its decomposition, and False Alarm Rate) and for the whole of the streamflow range (Ranked Probability Skill Score). When needed, discharge observations coming from the French database "Banque Hydro" (website: http://www.hydro.eaufrance.fr) were used as a reference for the scores. These observations had not yet been used by the assimilation system when they were compared to the forecasts. The thresholds used for the computation of the Brier Skill Score and Ranked Probability Skill Score were provided by the long-term climatology quantiles defined in the French streamflow observations database (Banque Hydro). These thresholds were: Q99, Q98, Q95, Q90, Q80, Q70, Q60, Q50 (all used for defining exceeding scores), Q40, Q30, Q20, Q10, Q5, Q2 and Q1 (used for defining non-exceeding scores). Q99, computed over a long period, means that 99% of the observed daily streamflows are under this value (idem for the other thresholds).

4.1   Set-up of the experiments

The experiments are described here for a forecast beginning on day D at 00:00 UTC.

The original system was initialized by the real-time SIM-analysis suite (analysis means meteorological fields created from observations and model outputs) but the two sets of improved initial states used a re-analysed SAFRAN-analysis (means analysis re-run a posteriori with more precipitations observations than available in real-time) and the version of the assimilation system with the variable state using a combination of the soil moisture of the two soil layers. It is important to note that the SAFRAN-analysis of the original ensemble streamflow prediction system is a real-time one, and

thus is not the same as the one used for the two experiments initialized by IS_1 and IS_2. However, for reasons of computing time, it was not possible to re-run the SAFRAN-analysis suite and the ensemble streamflow prediction system with the more recent SAFRAN-analysis. The SIM-analysis suite, with or without assimilation, used the SAFRAN-analysis data of day D-1.

For the assimilation system, discharge observations of day D-1 (averaged over this day) were used. The assimilation system final states (i.e. at D-1 24:00 UTC) were used as the initial states of the forecasts (i.e. at D 00:00 UTC). The forecasts were run from day D at 00:00 UTC to day D+9 24:00 UTC using the ECMWF meteorological EPS of the same dates. The discharges forecasts of day D, D+1, ..., D+9, averaged over each of these days, were used in the following and compared (if necessary) to the observations discharges averaged over the same days.

4.2   Ratio-Spread and Ratio-Root Mean Square Error (Ratio-RMSE)

The Ratio-RMSE (description in Appendix A) is plotted in Fig. 1 for the experiment without assimilation (in black), initialized by the IS_1 experiment (in green) and initialized by the IS_2 experiment (in blue), over the 10-day range. This score logically increased with the time range. The score was the best for the experiment initialized by the IS_2 states, then for the one using the IS_1 states and, finally, the worst for the experiment without any streamflow assimilation. However, the evolution of the Ratio-RMSE for the three experiments was different. The quality of the IS_1 experiment decreased most rapidly with the lead time of the forecast (increase of the Ratio-RMSE). The quality of the IS_2 experiment was more stable than IS_1 but the increase of the Ratio-RMSE from day 1 to day 10 was greater than in the original experiment. It is likely that the IS_2 curve converged (at a medium time-range) towards the Ratio-RMSE of an experiment without assimilation but with a SIM version including the exponential profile of the hydraulic conductivity. Unfortunately, because of shortage of computing time, this experiment could not be run.

The study of the Ratio-RMSE showed that the ensemble streamflow forecasts were closer to the observations than the original forecasts when both assimilation and exponential profile were used (IS_2), even after a 10-day lead time. When only the streamflow assimilation system was implemented (IS_1), the forecast was improved at the beginning of the time range but the improvement was small (though still present) for the last three days. As the reduction of the RMSE was the objective of the data assimilation techniques, it is satisfying to observe that its effects could still be seen after a few days.
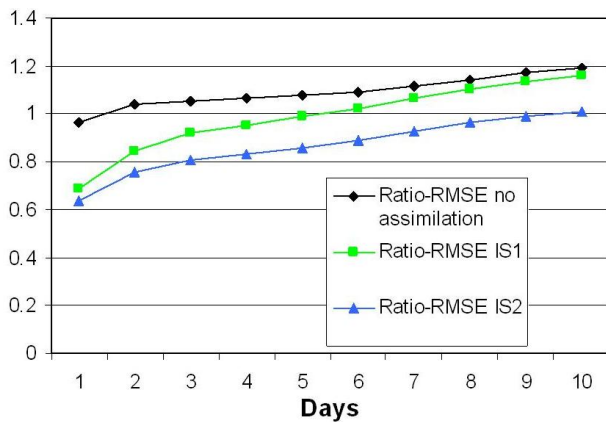
Fig. 1. Evolution of the Ratio-RMSE (no units) with the time range for the SIM ensemble streamflow forecasts, averaged over the 148 selected stations. Experiment without assimilation in black, with IS_1 in green, and with IS_2 in blue.

Fig. 2. Evolution of the Ratio-Spread (no unity) with the time range for the SIM ensemble streamflow forecasts, averaged on the 148 selected stations. Experiment without assimilation in black, with IS_1 in green, and with IS_2 in blue.

The Ratio-Spread (description in Appendix A) is plotted in Fig. 2 for the experiment without assimilation (in black), initialized by the IS_1 states (in green) and initialized by the IS_2 states (in blue). The evolution of this score is represented along the 10-day lead time of the ensemble streamflow forecasts. The Ratio-Spread was quite low, especially for the earliest days, when the spread was (on average) less than a tenth of the mean observed streamflow (Ratio-Spread equal to 0.1 after 3 days). The score increased linearly with the time range, and was around 0.4 for the last day, which means that the spread was, on average, around 40% of the streamflow observations. This figure shows that the ensemble dispersion was much too low in the earliest days, especially when compared to the Ratio-RMSE, and thus did not represent the uncertainty of the prediction accurately. The three experiments had scores very close together along the time range. The IS_2 experiment seemed to have a lower score for the ninth and tenth days, but this was considered as negligible.

The non-modification of the spread of the ensemble by the assimilation system is consistent with the fact that the assimilation is not intended to change this score. Its only impact on the score was that the spread was slightly increased when the soil moisture was increased by the assimilation and, slightly decreased when the soil moisture was reduced. But, as the increments imposed by the BLUE were low, its impact on the spread were low too.

4.3 Brier Skill Score (BSS) and Ranked Probability Skill Score (RPSS)

The Brier Skill Score (description in Appendix A, Brier, 1950) was computed for the three experiments and for the 10-day range. This score gives an assessment of the quality of an ensemble prediction for the exceeding of a thresh-
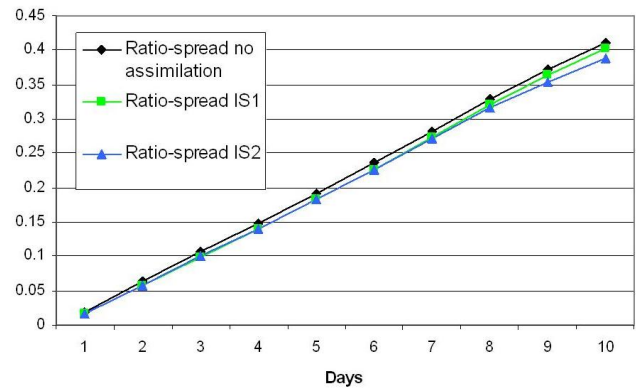
old. The reference Brier Score, necessary for computing the BSS, was calculated for a climatology of streamflow observations. This reference Brier Score was adjusted following the method developed by Weigel et al. (2007) to correct the bias of the BSS caused by the lower number of members in the ensemble.

Figure 3 (top) shows the evolution of the BSS for day 1 of the three experiments studied, with the quantiles used as thresholds. The BSS (like the Ratio-RMSE) showed that the best performance was achieved for IS_2, then for IS_1 (but quite close to IS_2), and the lowest skill was for the experiment without assimilation. A resampling test (Hamill, 1999) showed that IS_1 and IS_2 were significantly different except for the Q95-Q70 range, and that IS_1 and the experiment without assimilation were significantly different for all the thresholds considered. For the three experiments, the BSS was largely positive for the Q90-Q30 range and largely negative for the Q1-Q10 range. The bad scores for low flows (Q1-Q10 range) were due to the small number of cases, which biased the score, but also and mainly to small but continuous biases for these periods. For low flows, discharges are mainly fed by aquifers and rain rarely occurs. Thus, the spread of the ensemble is very low and, even if the simulated streamflows are very close to the observations, the forecasted frequencies are 0 or 1 most of the time. This results in marked differences between observed and forecasted frequencies, which are compared in the Brier Score, and so the BSS for low flows is negative. For the Q99-Q95 range, the BSS was negative when no assimilation was used, but positive when using the two sets of assimilated initial states. It showed the important impact of the assimilation system on the improvement in flood prediction.
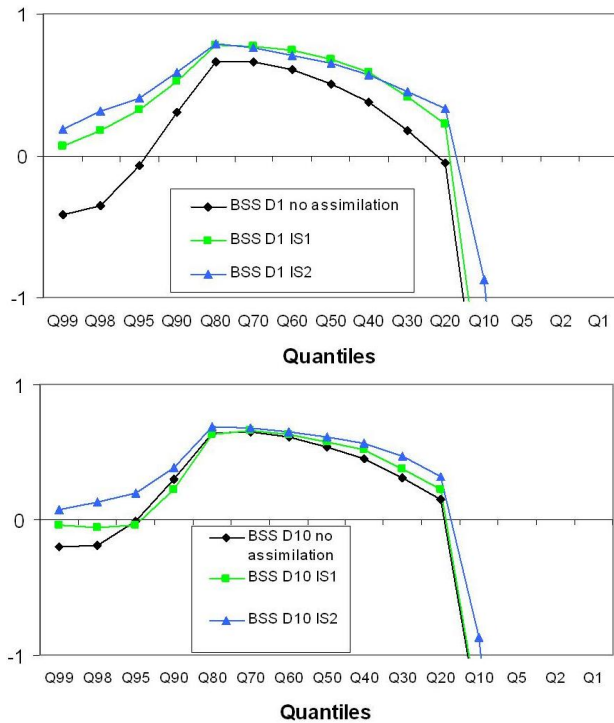
Fig. 3. Evolution of the Brier Skill Score with the quantiles for day 1 (top) and day 10 (bottom) of the SIM ensemble streamflow forecasts, averaged on the 148 selected stations. Experiment without assimilation in black, with IS_1 in green, and with IS_2 in blue.

Fig. 4. Evolution of the Ranked Probability Skill Score with the time range for the SIM ensemble streamflow forecasts, averaged over the 148 selected stations. Experiment without assimilation in black, with IS_1 in green, and with IS_2 in blue.

The assimilated initial states were very efficient for high flows (Q99–Q90), and moderately efficient for medium flows (Q80–Q20). This is consistent with the fact that the assimilation is more efficient for wet soils than for dry soils.

Fig. 3 (bottom) shows the BSS for day 10 of the three experiments studied. Here, although the IS_1 and IS_2 curves are close together for most of the thresholds for day 1, the two curves are more distant for the last day. The IS_1 curve is very close to the no assimilation curve for day 10. This means that, without the exponential profile of the hydraulic conductivity, the assimilation system has a small input after 10 days of predictions. Moreover, the resampling test showed that IS_1 and IS_2 were significantly different for all the thresholds, but that IS_1 and the experiment without assimilation were significantly different only for the extreme thresholds. Thus, the impacts of the assimilation system seemed large for the earliest days of prediction, but negligible afterwards, the exponential profile becoming the main improvement factor for the last days of forecast. Moreover, the BSS for day 10 for the experiment without assimilation was very similar to, or even better than (for the Q95–Q80 range) the BSS for the experiment initialized by IS_1. This behaviour for the Q95–Q80 range could be due to the physics used, without the exponential profile, which resulted in a bad simulation of floods and in an unexpected second flow peak (smaller than the first one)
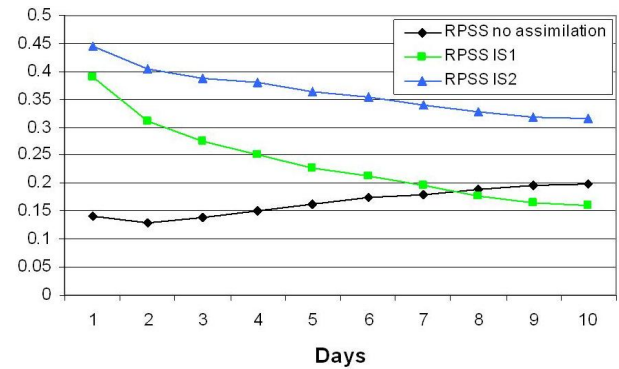
a few days after, provoked by a bad temporal distribution of the drainage fluxes in ISBA (see Fig. 1 in Quintana Seguí et al., 2009). The second peak, which was produced by the drainage of ISBA in these cases, did not exist for streamflow observations.

The BSS (not shown here) was evolving regularly from day 1 to day 10 for each experiment from the value of day 1 to the value of day 10.

The evolution of the RPSS (description in Appendix A) for the 10-day range is shown in Fig. 4 for the three experiments. All the thresholds used for the computation of the BSS were used for this score. IS_1 and IS_2 were significantly different for all the thresholds but IS_1 and the experiment without assimilation were only different for the first days of prediction. The best RPSS was for the IS_2 experiment, going from 0.45 for day 1, down to 0.31 for day 10, which means that the ensemble streamflow prediction system brought more information than the climatology for the whole time range (positive score). The IS_1 experiment had an RPSS decreasing from 0.39 down to 0.16. It showed that the non-use of the exponential profile for this experiment made this score decrease more rapidly with the lead time. Finally, the RPSS for the experiment without assimilation increased from 0.14 for day 1 to 0.2 for day 10. This increase was surprising at first, and made the RPSS for the experiment without assimilation higher than the RPSS for the IS_1 experiment for the last three days.

Although the EPS used remained the same for the three experiments, the SAFRAN-analysis used in the SIM-analysis suite was different. In real time, the SAFRAN-analysis (used by operational forecasts) only uses a limited number of rain gauge measurements for the precipitation analysis, and that leads to an under-estimation of precipitation. The SAFRAN-analysis used for the IS_1 and IS_2 experiments was re-computed afterwards with a larger number of rain gauge data, which provided a better quality analysis. Moreover,

Rousset-Regimbeau et al. (2007) showed that, for the first few days, the ensemble precipitation forecasts were close to the observations but that the EPS tended to over-estimate the precipitation for the last days of prediction. A compensation effect could thus have occurred for the experiment without assimilation improving the RPSS with the lead time. This could also be a reason why the BSS for day 10 is, for some thresholds, better for the experiment without assimilation, than for the experiment initialized by IS_1.

4.4 False Alarm Rate (FAR)

Many scores like the False Alarm Rate (description in Appendix A), the Hit Rate, the non- or good-prediction rates or the ROC curves were computed during this study. We will only present the False Alarms Rate here as this score is the most related to the operational needs. To compute this score, a percentage of 90% of the members (not-)exceeding a threshold was used.

The False Alarm Rate is shown in Fig. 5 for day 1 (top) for all the quantiles studied. This score was quite good, with a rate under 20% for the thresholds lower than the Q70 for the three experiments. For the high flows, the FAR remained under 50% for the experiment without assimilation, under 40% for IS_1, and under 30% for IS_2. The improvement of the FAR was significant when the assimilation system was used for the highest thresholds, especially with the improved physics, whose impact is indeed focused on floods. For the lowest thresholds, the improvement was very small and the IS_2 experiment even seemed to degrade this score very slightly.

For day 10 (Fig. 5, bottom), the FAR increased for high thresholds, but decreased for low thresholds for the three experiments, when compared to the FAR for day 1. The FAR was better for the experiment without assimilation than for the IS_1 experiment for some thresholds (Q95 to Q70), confirming the behaviour already seen for the RPSS. The FAR for the Q99 must be looked at carefully, because the low number of cases for this threshold surely biased the score. For day 10, as for the BSS, the FAR was better for the experiment without assimilation than for the experiment with IS_1 for the exceeding of some thresholds (Q95-Q70). Once again, the bad temporal simulation of the drainage fluxes induced by this old version of the physics in ISBA can explain this. So it caused unexpected false alarms. The use of the improved physics seems particularly important for floods.

The FAR (not shown here) was evolving regularly from day 1 to day 10 for each experiment from the value of day 1 to the value of day 10.

4.5 Decomposition of the Brier Score

The impact of the assimilation system on the terms of the decomposition of the Brier Score (description in Appendix A) will be studied in this part of the paper, mainly concerning the resolution and reliability. These scores are very help-
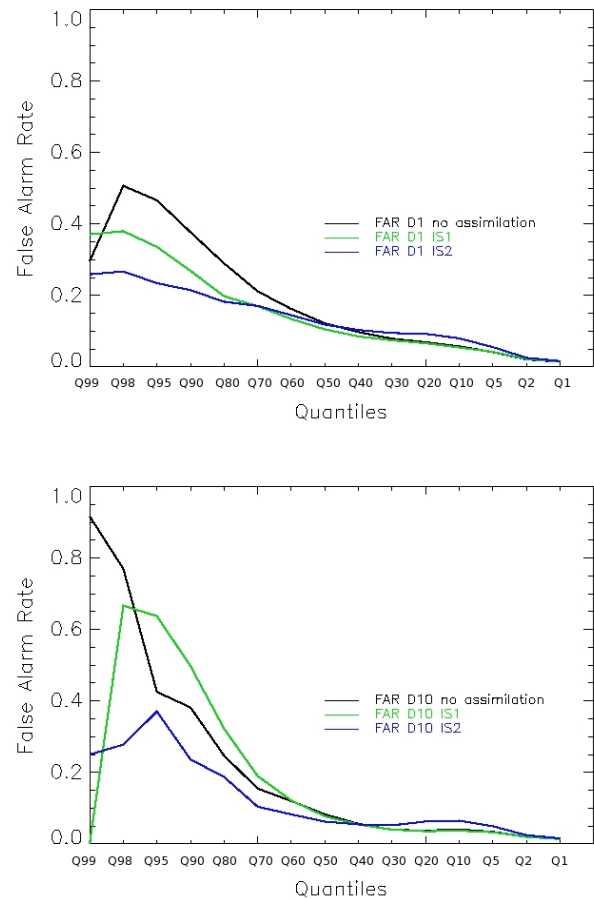

Fig. 5. False Alarm Rate (FAR) for day 1 (top) and day 10 (bottom), averaged for the 148 selected stations. Experiment without assimilation in black, with IS_1 in green, and with IS_2 in blue. The quantiles exceedance is defined when at least 90% of the members exceeds these quantiles.

ful to assess the probabilistic aspect of ensemble predictions. So, it would be good news if the assimilation could improve these scores, even if this is not the primary goal of the data assimilation.

The resolution term describes the capacity of the system to separate the probability classes. Its skill score, which is shown here, is positively oriented, with a perfect prediction if the resolution skill score is equal to 1, and a bad score if the resolution skill score is negative.

The resolution skill score is plotted for day 1 on Fig. 6 (top left) for all the thresholds and for the three experiments. The resolution skill score was quite low for the extreme quantiles but it was improved for all the thresholds by the two assimilated initial states, which brought an equivalent improvement in the score. The improvement was the highest for the high thresholds (Q99-Q30). The resolution skill score was very low for the extreme thresholds. For the low thresholds, this could be explained by the fact that they corresponded to non-
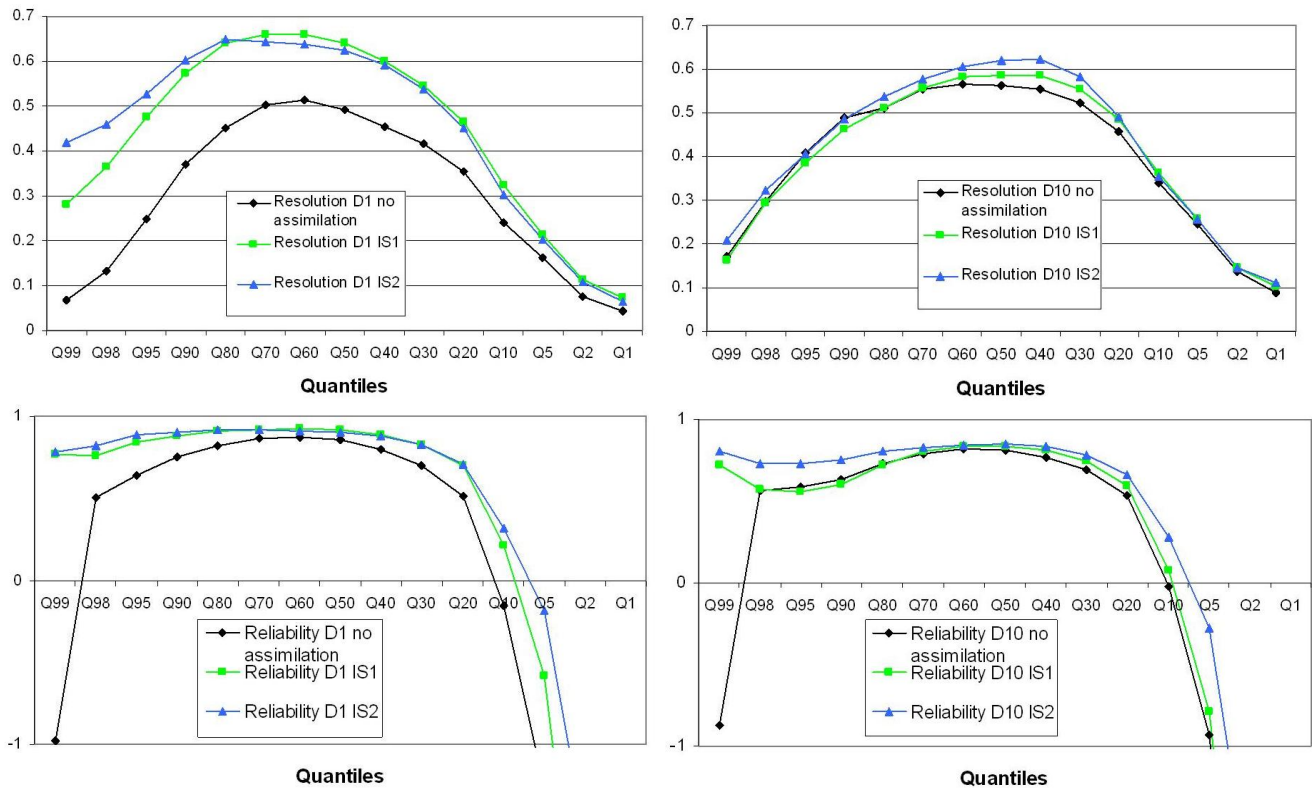
Fig. 6. Resolution Skill Score for day 1 (top left) and day 10 (top right), averaged for the 148 selected stations. Reliability Skill Score for day 1 (bottom left) and day 10 (bottom right), averaged for the 148 selected stations. Experiment without assimilation in black, with IS_1 in green, and with IS_2 in blue.

rain events most of the time, so all the hydrological scenarios were equivalent. For the high thresholds, it could be assumed that, because of the small number of cases, the score was biased. Moreover, these events are quite difficult to predict well (for both meteorological and hydrological forecasts). But, the score was hugely improved by the assimilation system for these high thresholds.

For day 10 (Fig. 6, top right), the shape of the curves was almost the same but only the thresholds from Q20 to Q60 were improved by the assimilation. The IS_1 experiment only brought a weak improvement, whereas the IS_2 states showed a significantly better resolution. The temporal evolution of the resolution for all the thresholds (not given here) showed that IS_1 and IS_2 were more or less equivalent at the beginning of the time range, but then IS_1 had a lower and lower impact and converged towards the score of the experiment without assimilation.

The reliability describes the capacity of the system to predict exact probabilities. Its skill score, shown here, is positively oriented, with a perfect prediction if the reliability skill score is 1, and a bad score if the reliability skill score is lower to 0.

The reliability skill score is plotted for day 1 on Fig. 6 (bottom left) for all the thresholds and for the three experiments. This score was quite high (i.e. good), especially for the highest thresholds (Q99-Q20 range). The lower (i.e. worse) score for the low flows could be explained by a bias for such discharge simulations. The reliability skill score was improved for all the thresholds by the assimilation system, and this improvement was apparently greater for the Q30-Q1 range, for which the score was less good. For the Q99-Q20 range, IS_1 and IS_2 had a similar reliability but, for the lowest thresholds, the IS_2 experiment was the best. The reliability skill score for the experiment without assimilation for the Q99 is probably due to a low number of events.

For day 10 (Fig. 6, bottom right), the reliability skill score of IS_1 and IS_2 was slightly lower than for day 1, but increased for the experiment without assimilation for the Q30-Q1 thresholds (no doubt due to the rainfall over-estimation of the ECMWF EPS, which reduced the hypothetical bias mentioned previously). The reliability skill score of the IS_1 experiment was very close to the reliability skill score of the experiment without assimilation, whereas the IS_2 experiment still brought an improvement for all the thresholds.

These scores (not shown here) were evolving regularly from day 1 to day 10 for each experiment from the value of day 1 to the value of day 10.

The uncertainty, not shown here, represents the difficulty of predicting an event, and is negatively oriented. As it is not dependent on the prediction system, this score was not influenced by the assimilation system. This score was, in our case, low for the extreme thresholds (under 0.1 for the Q99-Q80 and Q5-Q1 ranges), and a little higher for the other thresholds (but it remained under 0.25). This shows that the (non-)exceeding of the medium thresholds was more difficult to predict than that of the extreme thresholds.

4.6 Basin size study

Some scores were computed with regard to the size of the basin. It was interesting to see if the assimilation was efficient for all kinds of basins or only for small or large basins.

The Ratio-RMSE is plotted in Fig. 7 (for day 1 at top, day 10 at bottom) for the three experiments and averaged on 6 categories of basins with regard to their area: under $600\,km^2$, from $600\,km^2$ to $1000\,km^2$, from $1000\,km^2$ to $2000\,km^2$, from $2000\,km^2$ to $4000\,km^2$, from $4000\,km^2$ to $10\,000\,km^2$, and over $10\,000\,km^2$. For day 1, we can see that the assimilation system was efficient for all the categories studied and for both configurations. However, the assimilation seemed a little more efficient for the smallest basins than for the largest. This was possibly due to the fact that the observation errors (which were proportional to the squares of the observations) were high for the large basins, whereas the background error (which was linked to the surface area) was not necessarily greater for large basins than for small ones. The balance between these two error matrices was calibrated for the stations as a whole, not independently for each of the stations (see Thirel et al., 2010a). Thus, the assimilation could trust the background state more than the observation state for large basins, and so limited its impact.

For the last day of prediction, the effects of the assimilation IS_1 were very small, with an improvement for the 600–$1000\,km^2$ category only. For the IS_2 experiment, the impacts were still significant for the whole range, more especially for the smallest basins. The ratio-RMSE (not shown here) was evolving regularly from day 1 to day 10 for each experiment from the value of day 1 to the value of day 10.

The evolution of the RPSS with the basin size was also studied, but is not shown here. This score was a little better for the largest basins than for the smallest basins, which was consistent with the space scales targeted by the SIM model and by the ECMWF EPS. The proportion of improvement in the RPSS was bigger for the small basins than for the largest basins, which reinforced the conclusions of the Ratio-RMSE study.
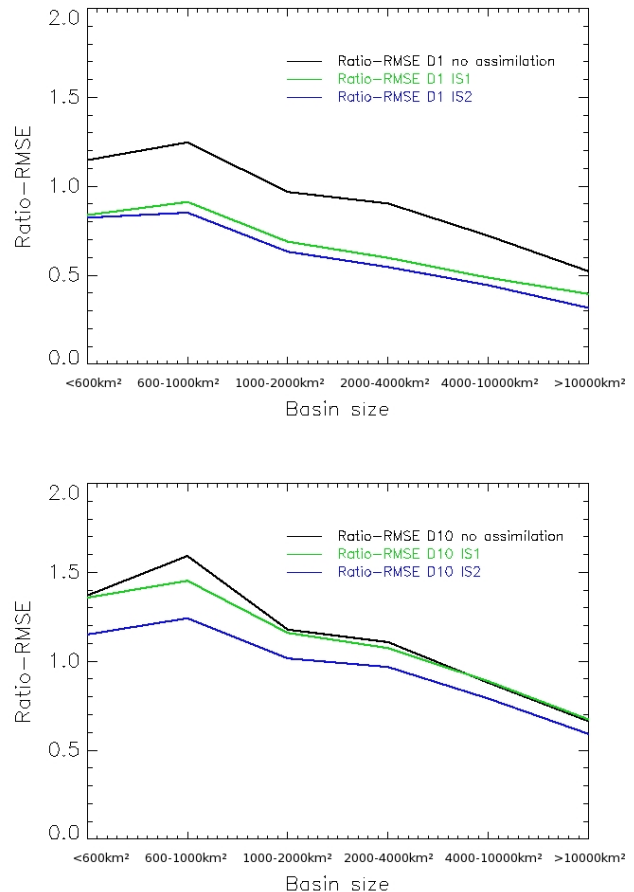

Fig. 7. Ratio-RMSE for day 1 (top) and day 10 (bottom), averaged over the selected stations according to the basin sizes. Experiment without assimilation in black, with IS_1 in green, and with IS_2 in blue.

5 Analysis on independent stations

In this section, some of the scores used previously are computed for a selection of 49 independent stations. These stations are located on upstream or downstream parts of the 148 assimilated stations, and are the ones already used in Thirel et al. (2010a). Scores were averaged over the period from 11 March 2005 to 30 September 2006, and the experiments initialized by the states from the IS_1, IS_2, and no-assimilation experiments will be compared as in the previous section. In fact, the forecasted streamflows studied in this section come from the same forecasts as in the previous section; only the selection of the stations to be studied has been modified. It was shown in Thirel et al. (2010a) that these 49 independent stations had better streamflow simulations after applying the assimilation system and, here, it is shown that this improvement continues for some days after the assimilation.
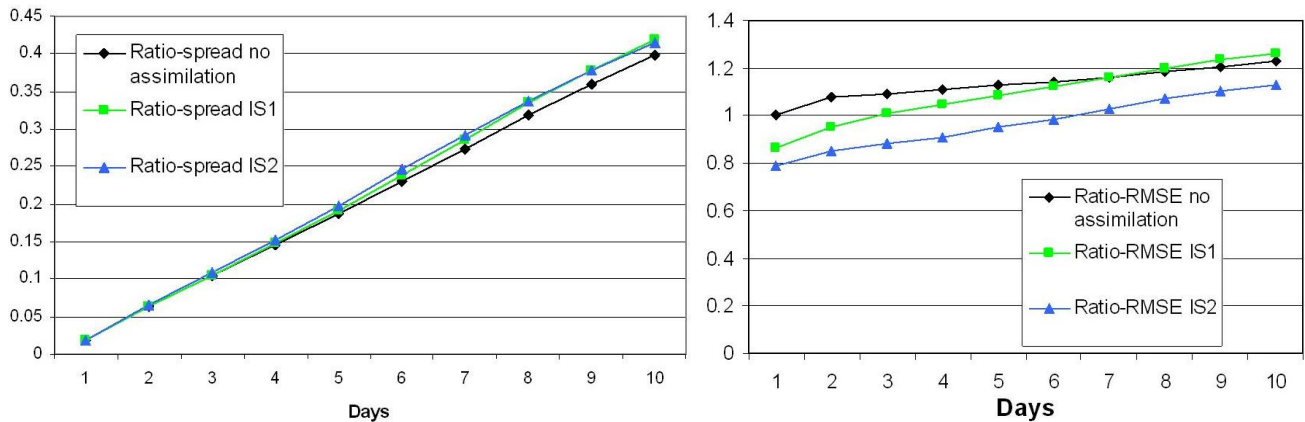
Fig. 8. Evolution of the Ratio-Spread (left, no units) and of the Ratio-RMSE (right, no units) with the time range for the SIM ensemble streamflow forecasts, averaged over the 49 independent stations. Experiment without assimilation in black, with IS_1 in green, and with IS_2 in blue.

5.1 Ratio-Spread and Ratio-Root Mean Square Error (Ratio-RMSE)

The Ratio-Spread for the 49 independent stations was hardly modified by the assimilation system (Fig. 8, left). The Ratio-Spread seems a little lower for the experiment without assimilation, but the difference is very small.

Concerning the Ratio-RMSE (Fig. 8, right), it can be seen that the improvement brought by the assimilation system is significant: the experiment using the IS_1 states (in green) shows a lower Ratio-RMSE than the original forecasts (in black) for the first 6 days. Then, however, this score deteriorates slightly. This shows that the assimilation of observed discharges can significantly improve the forecasts even for stations not assimilated. The better score for the original experiment for days 8 to 10 is probably due to the overestimation of the ECMWF EPS rainfall, which is counterbalanced by the under-estimation of the real-time SAFRAN analysis rainfall.

The improvement in the forecasts using the IS_2 states (assimilation + improved physics) is very clear and persists over the 10-day range of the forecasts. However, as expected, the effects of the assimilation system are lower for these 49 independent stations, than for the 148 assimilated stations studied previously (an improvement around 0.2 for the first day for the 49 independent stations, as against an improvement higher than 0.3 for the first day for the 148 assimilated stations (see Fig. 1)).

5.2 Brier Skill Score (BSS) and Ranked Probability Skill Score (RPSS)

The BSS for these 49 independent stations is shown in Fig. 9 for day 1 (top) and day 10 (bottom). The shapes of these curves are very similar to those of the curves for the 148

assimilated stations. For day 1, the improvement is real from the original experiment to the experiment initialized by IS_1, and from the latter to the experiment initialized by IS_2. The experiment using the IS_2 states has positive scores for floods, showing that the forecasts bring better information than the climatology.

For day 10, the improvement due to the use of the assimilated states is obviously lower, and the IS_1 states seem to decrease the performance of the forecasts for flows bigger than the Q90. This confirms what has already been seen for the Ratio-RMSE. However, the BSS for the forecasts using IS_2 are improved for all the thresholds, and positive scores can be observed even for thresholds higher than the Q95. The evolution of the BSS from day 1 to day 10 (not shown here) was regular for each experiment.

The RPSS for the 3 experiments is displayed in Fig. 10 over the whole 10-day lead-time. Once again, the score of the experiment using IS_1 only improves for the first few days and then decreases. But, for the forecasts combining the assimilation system and the improvement of the physics, the improvement in the RPSS is significant over the whole lead-time, with an increase of more than 0.2 in the RPSS for day 1, and the improvement still close to 0.1 after 10 days. The improvement due to the assimilation is obviously lower for these 49 independent stations than it is for the 148 assimilated stations (see Fig. 4).

Other scores are not displayed here, but they showed similar improvements for the two experiments using assimilated initial states as seen for the 148 assimilated stations (but on a smaller scale).

Figure 11 shows the values of RPSS averaged over the 19-month period, for both the 148 assimilated stations (circles) and the 49 independent stations (triangles). The experiment using no assimilation and the one using the IS_2 states are
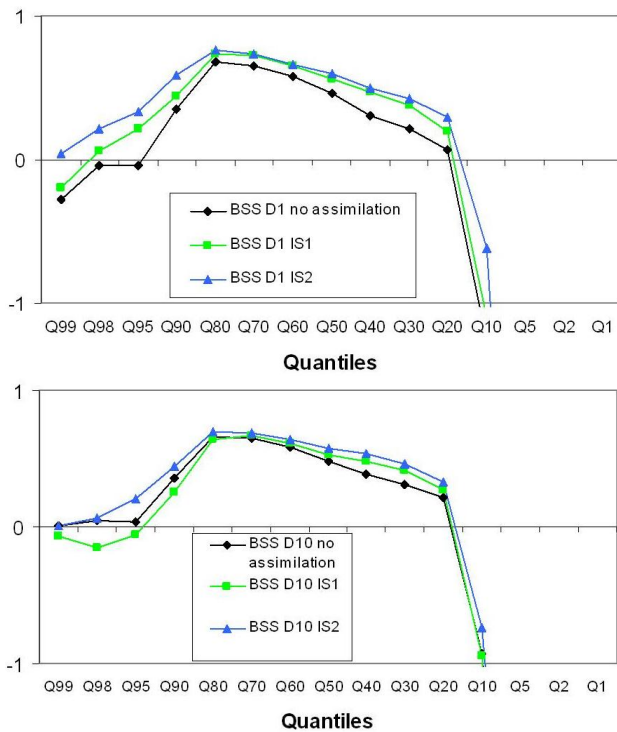
Fig. 9. Evolution of the Brier Skill Score with the quantiles for day 1 (top) and day 10 (bottom) of the SIM ensemble streamflow forecasts, averaged over the 49 independent stations. Experiment without assimilation in black, with IS_1 in green, and with IS_2 in blue.
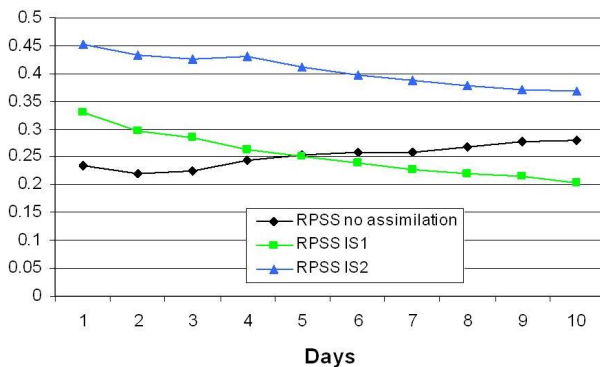

Fig. 10. Evolution of the Ranked Probability Skill Score with the time range for the SIM ensemble streamflow forecasts, averaged over the 49 independent stations. Experiment without assimilation in black, with IS_1 in green, and with IS_2 in blue.

plotted here, for day 1 and day 10 of forecasts. The global improvement of the RPSS over France when using assimilated initial states is very clear. Although a deterioration of the RPSS with lead-time can be seen for the experiment using assimilated states, the original experiment does not show such behaviour, which is consistent with Figs. 4 and 10. It

is quite difficult to find any trend concerning which type of basin is the best forecasted or which region has the best score. However, forecasts for the Loire river basin seem to be the most durably improved by the assimilation system. In contrast, results are poor for some stations of the South-West (Garonne and Dordogne river basins) and in the East (Meuse, Moselle and Saône river basins). The evolution of the RPSS from day 1 to day 10 (not shown here) was regular for each experiment.

6 Conclusions

The impacts of an assimilation system of past discharges on medium-range ensemble streamflow forecasts were assessed in this study. The ensemble streamflow prediction system was constructed on the basis of the SAFRAN-ISBA-MODCOU (SIM) hydrometeorological model. This model was used in two different ways: as a hydrometeorological analysis, and as the model producing the streamflow forecasts. These streamflow forecasts were initialized by the SIM-analysis suite, which used the SAFRAN-analysis meteorological data. A data assimilation system was implemented in order to improve this initialization (Thirel et al., 2010a). For this, observed discharges were used, and the Best Linear Unbiased Estimator (BLUE) determined the increments to be imposed on the ISBA soil moisture in the SIM-analysis suite. The impacts of this streamflow assimilation system on ensemble streamflow forecasts were assessed over a 569-day period for two configurations of the assimilation system using a state variable averaging the two ISBA layers soil moistures: the first one only contained the assimilation system (IS_1), but the second one (IS_2) also contained improved physics of the soil moisture (the exponential profile of the hydraulic conductivity). The use of these two sets of initial states was compared with the use of real-time non-assimilated initial states, when initializing the SIM-ECMWF suite, i.e. the medium-range (10-day) ensemble streamflow prediction system of Météo-France. The physics of the prediction model was coherent with the physics of the assimilation part. The comparison was performed over 148 assimilated stations, and over 49 independent stations.

The study showed that the three experiments had quite good statistical scores for 148 assimilated stations. The assimilation experiments improved the performances of the forecasts in terms of RMSE, BSS, RPSS, and FAR especially. The spread was not changed, as this is not the aim of the assimilation process. When only using the assimilation, but no improved physics (IS_1), the forecasts were mainly improved for the first days, then the improvement was lower. For the experiment using both the assimilation system and the improved physics (IS_2), better scores were observed, for the first days of the leadtime, but also for the last days of the leadtime.
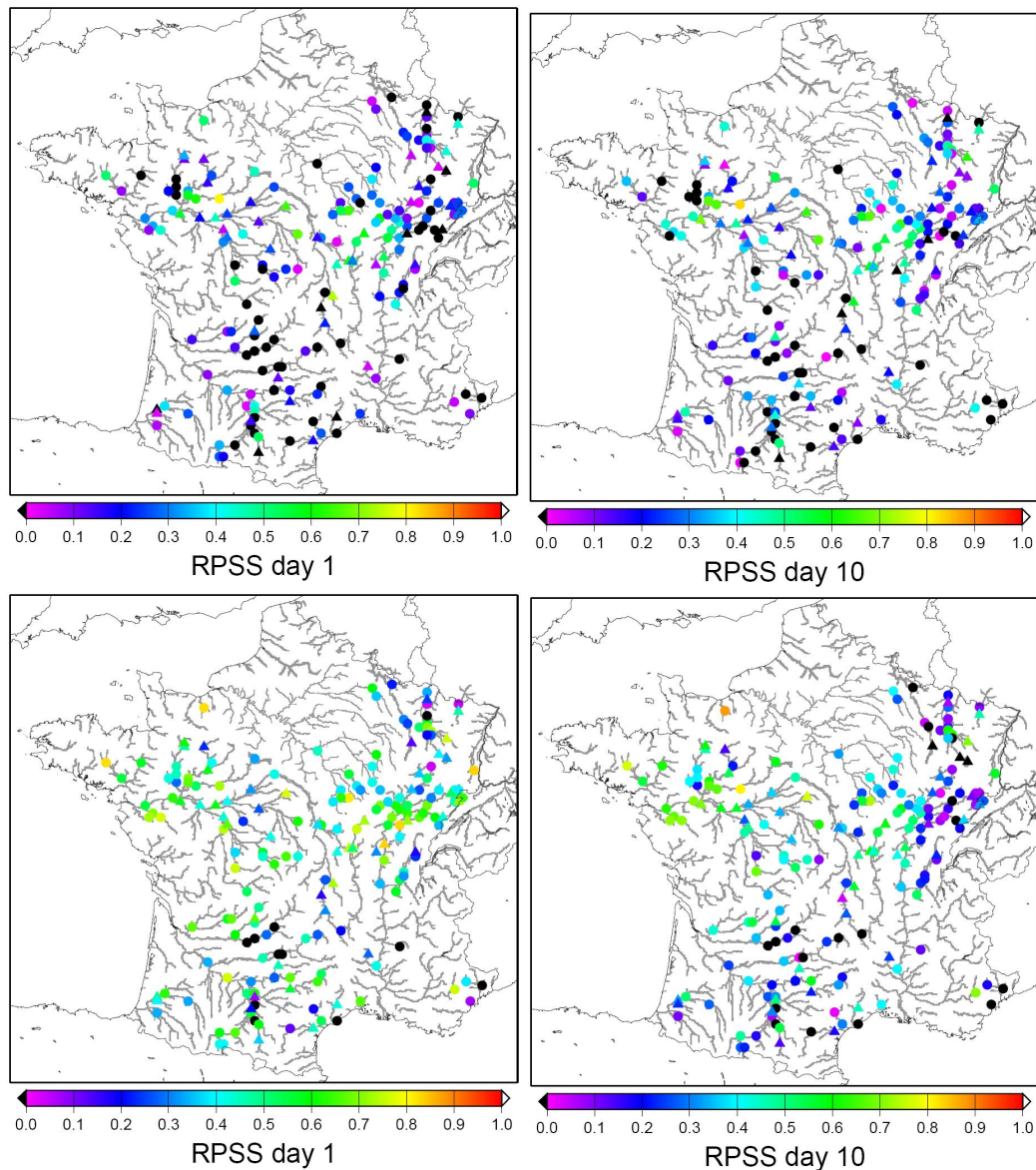
Fig. 11. Ranked Probability Skill Score for day 1 (left) and day 10 (right) without assimilation (top) and with the IS_2 assimilation (bottom) averaged for each station. The 148 assimilated stations are represented by circles, and the 49 independent stations are plotted as triangles. Black points are negative values.

These conclusions were confirmed by the study of the statistical scores for 49 independent stations. This highlights an important advantage of distributed hydrological models over lumped models. However, the improvements given by the assimilation systems (in configuration IS_1 or IS_2) were obviously lower than for the 148 assimilated stations.

This study demonstrated the potential of using a past discharge assimilation system in order to improve the medium-range ensemble streamflow forecasts of a distributed hydrometeorological model, SIM. More precisely, the assimilation system itself showed its greater impact at the begin-ning of the time-range, whereas the last days of predictions were improved by the use of the exponential profile of the hydraulic conductivity in the soil rather than by the assimilation system.

However, some aspects of the assimilation system could be improved, as stated in Thirel et al. (2010a). First, the performance of the system could be improved by a better assessment of the soil moisture and observation errors. The balance between the background error matrix and the observation error matrix should be carried out basin by basin, or at least separately for large and small basins, in order to deal with the

huge observation errors for large basin stations. This should improve the quality of floods prediction, which are the events on the interest of flood alert services.

Moreover, the use of an Ensemble Kalman Filter could improve the non-linearities of the model, resulting in better assimilated states. Although the drainage flux varies linearly with the soil moisture, the runoff does not which may lead to problems concerning the respect of the linearity hypothesis of the BLUE (even if tests on the use of an external loop on the BLUE showed only slight improvements).

Concerning the statistical aspect of the ensemble streamflow predictions, it appeared that much improvement is achievable. The spread remained very low, especially for the first days of forecast, and a score like the resolution was too low. That is why further research is needed in this field. For example, the dispersion of the meteorological EPS could be improved in a dynamical way. Moreover, other ensemble parameters, like radiation or wind, could improve the representation of the meteorological uncertainty if they were used in addition to the precipitation and temperature ensemble data.

The quality of the hydrological forecasts should also be hugely improved when the new Var-EPS of the ECMWF comes into use, because of its better spatial resolution and uncertainty description. The Var-EPS is scheduled to be included into the operational ensemble SIM suite at Météo-France in the near future. This point is very important to better predict localized and extreme events, which are often difficult to predict for flood alert services.

Finally, the streamflow assimilation system will be adapted to initialize the operational ensemble SIM suite by using real-time observed discharges of a selected number of relevant stations. Thus, its use by operational forecasters would become easier, and this would help to improve the French flood alert services.

Appendix A

The statistical scores

Ratio-Spread

The Ratio-Spread is a statistical score derived from the spread, in which each station spread is divided by the mean of the relevant observations on the period studied. This feature allows differences in spread due to the magnitude of the variable to be taken into account. In our case, streamflows are much higher for large basins than for small basins. That is why the spread is often higher for large basins than for small ones. Use of the Ratio-Spread allows the errors to be normalized:

$$\text{Ratio} - \text{Spread} = \frac{1}{N\bar{o}} \sum_{i=1}^{N} \sqrt{\frac{1}{n} \sum_{k=1}^{n} (m_{k,i} - m_i)^2},$$

with n the number of members, N the number of days, m_i the mean of the ensemble, $\bar{o}$ the averaged observations, and $m_{k,i}$ the value of member k for day i.

Ratio-RMSE

The Ratio-RMSE is computed from the RMSE in the same way that the Ratio-Spread is computed from the spread:

$$\text{Ratio} - \text{RMSE} = \frac{1}{\bar{o}} \sqrt{\frac{1}{N} \sum_{i=1}^{N} (m_i - o_i)^2},$$

with o_i the observation for day i.

Brier Skill Score

The Brier Score (BS) and its associated Skill Score (BSS, Brier, 1950) are statistical scores widely used for the study of ensemble predictions. They make it possible to assess the quality of ensemble predictions regarding the (non-) exceeding of thresholds.

$$\text{BS} = \frac{1}{N} \sum_{k=1}^{N} (y_k - o_k)^2 \quad 0 \le BS \le 1,$$

with y_k the probability of the forecasted event, and $o_k = 1$ if the event is observed, $o_k = 0$ if it is not observed. BS$= 0$ for a perfect forecast, and BS is close to 1 for bad forecasts.

In order to make comparisons between two ensemble predictions systems, the BSS is used:

$$\text{BSS} = 1 - \frac{\text{BS}}{\text{BS}_{\text{ref}}} \quad -\infty \le \text{BSS} \le 1,$$

with BS_{ref} the BS of a reference experiment (a climatology of streamflow observations in our case). The BSS is positively oriented, with a value close to 1 for a perfect forecast. A positive value describes an improvement in forecasts compared to the climatology used.

Decomposition of the Brier Skill Score

The BS can be decomposed into a sum of terms (see Murphy, 1973 for the demonstration), called reliability, resolution and uncertainty.

$$\text{BS} = \text{BS}_{\text{rel}} - \text{BS}_{\text{res}} + \text{BS}_{\text{unc}},$$

with:

$$\text{BS}_{\text{rel}} = \frac{1}{N} \sum_{i=1}^{n+1} N_i (y_i - \bar{o}_i)^2,$$

$$\text{BS}_{\text{res}} = \frac{1}{N} \sum_{i=1}^{n+1} N_i (\bar{o}_i - \bar{o})^2,$$

$$\text{BS}_{\text{unc}} = \bar{o}(1 - \bar{o}),$$

with N_i the number of forecasts in the category i and $\bar{o} = \frac{1}{N}\sum_{k=1}^{N}o_k$. Reliability corresponds to the capacity of the system to predict correct probabilities; a value of 0 means perfect reliability. Resolution describes the capacity of the system to separate the probability classes; it is positively oriented. The uncertainty is the variance of observations.

In this article, the Reliability and resolution skill scores were used. These skill scores are positively oriented, with a value close to 1 for a perfect forecast. A positive value describes an improvement in forecasts compared to the climatology used:

$$\text{BSS}_{\text{rel}} = 1 - \frac{\text{BS}_{\text{rel}}}{\text{BS}_{\text{ref}}},$$

$$\text{BSS}_{\text{res}} = \frac{\text{BS}_{\text{res}}}{\text{BS}_{\text{ref}}}$$

Ranked Probability Skill Score

The Ranked Probability Score (RPS) is a score derived from the Brier Score. It assesses the ensemble predictions over the whole range of values of the parameter considered. The forecasts are divided into $J = 16$ classes in the study, which are determined by the values of the 15 quantiles defined by the Banque Hydro. Here y_j is the probability of the forecasted event for the class j. We define:

$$Y_m = \sum_{j=1}^{m}y_j, m = 1,...,J,$$

$$O_m = \sum_{j=1}^{m}o_j, m = 1,...,J.$$

The RPS is then defined as:

$$\text{RPS} = \frac{1}{N}\sum_{k=1}^{N}[\sum_{m=1}^{J}(Y_m - O_m)^2]_k$$

and its Skill Score is:

$$\text{RPSS} = 1 - \frac{\text{RPS}}{\text{RPS}_{\text{ref}}} \qquad -\infty \le \text{RPSS} \le 1.$$

The RPS is perfect when equal to 0, and bad when equal to 1. The RPSS is perfect when equal to 1, and a negative RPSS indicates behaviour worse than the climatology.

False Alarm Rate

False Alarm rate and Hit Rate are scores indicating the quality of a deterministic prediction. These scores also can be defined in case of ensemble predictions. For such predictions, it is considered that an event is (or is not) predicted if $p\%$ of the members predict (or do not predict) it. p can be adjusted to the user's needs and was taken equal to 90 in this study. Table A1 is used to define these scores.

Table A1. Contingency table of possible events.

Event	Observed	Not observed
Forecasted	a	b
Not forecasted	c	d

We can define:

$$\text{False\_Alarm\_Rate} = \frac{b}{a+b}$$

$$\text{Hit\_Rate} = \frac{a}{a+c}$$

References

Aubert, D., Loumagne, C., and Oudin, L.: Sequential assimilation of soil moisture and streamflow data in a conceptual rainfall-runoff model, J. Hydrol., 280, 145–161, 2003.

Brier, G.: Verification of forecasts expressed in terms of probability, Mon. Weather Rev., 78, 1–3, 1950.

Buizza, R., Miller, M., and Palmer, T. N.: Stochastic representation of model uncertainties in the ECMWF ensemble prediction system, Q. J. Roy. Meteor. Soc., 125, 2887–2908, 1999.

H. Cloke and F. Pappenberger, Ensemble Flood Forecasting: A review, J. Hydrol., 375, 613–626, 2009.

Crow, W. T. and Ryu, D.: A new data assimilation approach for improving runoff prediction using remotely-sensed soil moisture retrievals, Hydrol. Earth Syst. Sci., 13, 1–16, doi:10.5194/hess-13-1-2009, 2009.

Habets, F., Boone, A., Champeau, J. L., Etchevers, P., Leblois, E., Ledoux, E., Lemoigne, P., Martin, E., Morel, S., Noilhan, J., Quintana Seguí, P., Rousset-Regimbeau, F., and Viennot, P.: The SAFRAN-ISBA-MODCOU hydrometeorological model applied over France, J. Geophys. Res., 113, D06113, doi:10.1029/2007JD008548, 2008.

Hamill, T. M.: Hypothesis Tests for Evaluating Numerical Precipitation Forecasts, Weather Forecast., 14, 155–167, 1999.

He, Y., Wetterhall, F., Cloke, H. L., Pappenbeger, F., Wilson, M., Freer, J., and McGregor, G.: Tracking the uncertainty in flood

alerts driven by grand ensemble weather predictions, Meteorol Appl., 16(1), 91–101, 2009.

Komma, J., Glöschl, G., and Reszler, C.: Soil moisture updating by Ensemble Kalman Filtering in real-time flood forecasting, J. Hydrol., 357, 228–242, 2008.

Lakshmi, V.: Use of satellite remote sensing in hydrological predictions in ungauged basins, Hydrol. Process., 18(5), 1029–1034, 2004.

Ledoux, E., Girard, G., de Marsilly, G., and Deschenes, J.: Spatially distributed modeling: Conceptual approach, coupling surface water and groundwater, Unsaturated Flow Hydrologic Modeling, Theory and Practice, edited by: Morel-Seytoux, H. J., NATO ASI Series C, 275, Kluwer, 435–454, 1989.

Murphy, A. H.: A new vector partition of the probability score, J. Appl. Meteorol., 12, 595–600, 1973.

Noilhan, J. and Planton, S.: A simple parameterization of land surface processes for meteorological models, Mon. Weather Rev., 117, 536–549, 1989.

Pauwels, V. R. N. and De Lannoy, G. J. M.: Ensemble-based assimilation of discharge into rainfall-runoff models: A comparison of approaches to mapping observational information to state space, Water Resour. Res., 45, W08428, doi:10.1029/2008WR007590, 2009.

Quintana Seguí, P., Martin, E., Habets, F., and Noilhan, J.: Improvement, calibration and validation of a distributed hydrological model over France, Hydrol. Earth Syst. Sci., 13, 163–181, doi:10.5194/hess-13-163-2009, 2009.

Ramos, M.-H., Bartholomes, J., and Thielen del Pozo, J.: Development of decision support products based on ensemble forecasts in the European Flood Alert System, Atmos. Sci. Lett., 8, 113–119, 2007.

Rousset-Regimbeau, F., Habets, F., Martin, E., and Noilhan, J.: Ensemble streamflow forecasts over France, ECMWF Newsletter No. 111, Spring, 2007.

Schaake, J., Franz, K., Bradley, A., and Buizza, R.: The Hydrologic Ensemble Prediction EXperiment (HEPEX), Hydrol. Earth Syst. Sci. Discuss., 3, 3321–3332, doi:10.5194/hessd-3-3321-2006, 2006.

Seo, D.-J., Cajina, L., Corby, R., and Howieson, T.: Automatic State Updating for Operational Streamflow Forecasting via Variational Data Assimilation, J. Hydrol., 9, 367, 255–275, 2009.

Thirel, G., Rousset-Regimbeau, F., Martin, E., and Habets, F.: On the Impact of Short-Range Meteorological Forecasts for Ensemble Streamflow Predictions, J. Hydrometeorol., 9, 1301–1317, 2008.

Thirel, G., Martin, E., Mahfouf, J.-F., Massart, S., Ricci, S., and Habets, F.: A past discharges assimilation system for ensemble streamflow forecast over France – Part 1: Description and validation of the assimilation system, Hydrol. Earth Syst. Sci., 14, 1623-1637, doi:10.5194/hess-14-1623-2010, 2010.

Thirel, G., Regimbeau, F., Martin, E., Noilhan, J., and Habets, F.: Short and medium-range hydrological ensemble forecasts over France, Atmos. Sci. Lett., 11, 72–77, doi:10.1002/asl.254, 2010.

Van Andel, S. J., Price, R. K., Lobbrecht, A. H., van Kruiningen, F., and Mureau, R.: Ensemble Precipitation and Water-Level Forecasts for Anticipatory Water-System Control, J. Hydrometeorol., 9, 776–788, 2008.

Van den Bergh, J. and Roulin, E.: Hydrological ensemble prediction and verification for the Meuse and Scheldt basins, Atmos. Sci. Lett., 2010.

Weigel, A. P., Liniger, M. A., and Appenzeller, C.: The discrete Brier and ranked probability skill scores, Mon. Weather Rev., 135, 118–124, 2007.

Wood, A. W., Kumar, A., and Lettenmaier, D. P.: A retrospective assessment of National Centers for Environmental Prediction climate model-based ensemble hydrologic forecasting in the western United States, J. Geophys. Res., 110, D04105, doi:10.1029/2004JD004508, 2005.

Zaitchik, B. F., Rodell, M., and Reichle, R. H.: Assimilation of GRACE Terrestrial Water Storage Data into a Land Surface Model: Results for the Mississippi River Basin, J. Hydrometeorol., 9, 535–548, 2008.

Zappa, M., Rotach, M. W., Arpagaus, M., Dorninger, M., Hegg, C., Montanari, A., Ranzi, R., Ament, F., Germann, U., Grossi, G., Jaun, S., Rossa, A., Vogt, S., Walser, A., Wehran, J., and Wunram, C.: MAP D-PHASE: Real-time demonstration of hydrological ensemble prediction systems, Atmos. Sci. Lett., 2, 80–87, doi:10.1002/asl.183., 2008.