Journal of Hydrology 519 (2014) 2775–2784

Contents lists available at ScienceDirect

Journal of Hydrology

journal homepage: www.elsevier.com/locate/jhydrol

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

ARTICLE INFO

Article history:
Available online 7 August 2014

Keywords:
Hydrological ensemble forecasting
Data assimilation
Post-processing
Ensemble dressing
Uncertainty propagation

SUMMARY

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 recommended in hydrological ensemble forecasting.

1. Introduction

1.1. Addressing uncertainties in hydrological ensemble forecasting

Developing and improving operational hydrological ensemble forecasting systems is a critical step toward better decision-making and risk management. The skill of operational hydrological ensemble forecasting systems is limited by two main sources of uncertainty (Krzysztofowicz, 1999): meteorological uncertainty and hydrological uncertainty. From a pragmatic point of view, the need to properly account for these two main sources of uncertainty arises because (i) a hydrological forecaster has no choice but to rely on uncertain meteorological forecasts and (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 (Montanari and Brath, 2004; Montanari and Grossi, 2008; Solomatine and Shrestha, 2009; Coccia and Todini, 2011; Morawietz et al., 2011; Weerts et al., 2011; Ewen and O'Donnell, 2012; Pianosi and Raso, 2012; Smith et al., 2012; Van Steenbergen et al., 2012; Yan et al., 2012), but their use is not widespread for operational ensemble forecasting.

Although generally dealt with separately, statistical post-processing and data assimilation (also called real-time model updating in the engineering community) can be intrinsically related in the hydrological forecasting framework. Both represent techniques that may be used in a forecasting system to improve the quality of the forecasts (i.e., to provide more accurate and reliable forecasts) and to, ultimately, enhance the usefulness of the forecasts in decision-making. Since forecasting deals with an uncertain future, these techniques aim to bring additional information to the forecast procedure and take into account the various uncertainty sources (or at least the major uncertainty sources) affecting the forecasting chain. This is usually achieved by merging information from model and observations.

While data assimilation and post-processing share a general goal, the techniques applied may differ in the practice of hydrological forecasting. These differences usually draw the separation between what is defined as data assimilation and what is defined as post-processing in a modelling framework. The definitions used in this study are the following: we use the term "post-processing" when using the hydrological uncertainty processor (Section 2.4), whose primary purpose is to dress deterministic forecasts with

* Corresponding author. Tel.: +33 1 40 96 65 79; fax: +33 1 40 96 62 58.
E-mail address: francois.bourgin@irstea.fr (F. Bourgin).

http://dx.doi.org/10.1016/j.jhydrol.2014.07.054
0022-1694/

uncertainty based on distributions of past model errors and, this way, build probabilistic forecasts. "Data assimilation" refers to techniques applied to perform the updating of the system before it issues a deterministic forecast. Here it concerns the state updating of the hydrological model and a model error correction applied to its output (Section 2.3).

The fact that data assimilation has the potential to improve real-time streamflow forecasting is widely accepted (see Liu et al., 2012, for a review). In contrast to probabilistic and ensemble-based data assimilation methods (e.g., Weerts and El Serafy, 2006; Salamon and Feyen, 2010; Moradkhani et al., 2012; Vrugt et al., 2013), deterministic updating schemes are designed to improve forecasts without producing probabilistic outputs. They may be easier to implement, mainly operationally, but at the price of leaving the uncertainty quantification issue unanswered. In these cases, the use of statistical post-processing methods together with data assimilation procedures provides a way to reduce and quantify the predictive uncertainty in the hydrological forecasts.

1.2. Integrating uncertainties in hydrological ensemble forecasting

"Ensemble dressing" is an intuitive and operationally-appealing method that allows integration of uncertainties from hydrological modelling and meteorological (ensemble) forcing. The main difference with other ensemble-based post-processors (e.g., Wang and Bishop, 2005; Fortin et al., 2006; Brown and Seo, 2010; Boucher et al., 2012; Brown and Seo, 2013) is that, for ensemble dressing, hydrological modelling errors are assessed separately, and later combined with ensemble forecasts. Distributions of modelling errors are obtained from long time series of simulated and observed data (i.e., learning from the past), and then applied to ensemble forecasts to obtain the total predictive distribution.

In recent studies, the use of ensemble dressing has been implemented and tested to improve the skill of hydrological ensemble forecasting systems. For instance, Reggiani et al. (2009) present a Bayesian ensemble uncertainty processor for medium-range ensemble flow forecasts in the Rhine river basin. Hopson and Webster (2010) use an uncertainty processor based on the k-nearest neighbors (k-NN) resampling method to dress probabilistic medium-range forecasts for two large basins in Bangladesh. Zalachori et al. (2012) compare different strategies based on pre- and post-processing methods to remove biases in a streamflow ensemble prediction system developed for reservoir inflow management in French catchments, while Pagano et al. (2013) present a hydrological application of ensemble dressing for 128 catchments in Australia.

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

1.3. Aim and scope of the study

This study aims to shed light on the interactions between data assimilation and post-processing in hydrological ensemble forecasting. We address the following questions:

1. How does data assimilation impact hydrological ensemble forecasts?
2. How does post-processing impact hydrological ensemble forecasts?
3. How does data assimilation interact with post-processing to improve the quality and skill of hydrological ensemble forecasts over the forecast lead times?

We address these questions with the help of a large set of catchments, making it possible to draw more general and robust conclusions.

2. Data and methods

2.1. Data set

A set of 202 unregulated catchments spread over France was used (Fig. 1). The catchments represent various hydrological conditions, given the variability in climate, topography, and geology in France. This set includes fast responding Mediterranean basins with intense precipitation as well as larger, groundwater-dominated basins. Some characteristics of the data set are given in Table 1. Catchments were selected to have limited snow influence, since no snowmelt module was used in the hydrological modelling (Section 2.3).

Potential evapotranspiration (PE), precipitation, and discharge data were available at hourly time steps over the 1997–2006 period. Temperature inputs originate from the SAFRAN reanalysis (Vidal et al., 2010). PE was estimated using a temperature-based formula (Oudin et al., 2005). Precipitation data come from a reanalysis dataset recently produced by Météo-France based on weather radar and rain gauge network (Tabary et al., 2012). River discharge data were extracted from the HYDRO national archive (www.hydro.eaufrance.fr).

2.2. PEARP, the Météo-France ensemble forecast

A short-range meteorological ensemble prediction system, the Météo-France PEARP EPS (Nicolau, 2002), was used to produce hydrological ensemble forecasts. The PEARP EPS runs once a day at 18:00 UTC; it has 11 members, a 60 h forecast range, and a 0.25° (ca. 25 km in France) grid resolution. A spatial disaggregation to an 8 km × 8 km grid, which includes bias correction, was applied to the PEARP forecasts. Bias correction was applied to precipitation forecasts using a multiplying factor obtained from a comparison between the mean of the PEARP ensemble and the Météo-France SAFRAN reanalysis over a complete year (March 2005–March 2006). Details can be found in Thirel et al. (2008). PEARP forecasts were available over the 2005–2009 period, but only the period matching the observed data could be used here, i.e. from August 2005 to December 2006.

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

2.3. The GRP rainfall–runoff forecasting model

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