Improving flow forecasting by error correction modelling in altered catchment conditions

Francesca Pianosi,[1]* Andrea Castelletti,[1] Leonardo Mancusi[2] and Elisabetta Garofalo[2]

[1] *Dipartimento di Elettronica, Informazione e Bioingegneria*
[2] *RSE Sp, Ricerca sul Sistema Energetico, Italy*

Received 20 November 2012; Accepted 13 February 2013

INTRODUCTION

Mathematical models are widely used to simulate hydrological processes at different temporal and spatial scales. Operational flow forecasting is one of the most common applications and provides essential support to a range of water resources management activities, including early warning services and the optimal operation of water infrastructures. In recent years, the interest in operational flow forecasting systems has steadily raised as advances in numerical weather prediction have significantly enhanced their predictive capability in terms of both forecast accuracy and lead time (Habets *et al*., 2004).

Hydrological models are often classified into physically based, conceptual and data-driven (empirical) models. For operational forecasting, conceptual models, either lumped or semi-distributed, are often deemed to offer an acceptable compromise between prediction accuracy and ease of use (O'Connor, 1992). In fact, although based on a number of simplifying assumptions, conceptual models usually offer reasonably accurate predictions while requiring much less data than physically based models in both their calibration and application. On the other hand, with respect to data-driven models, they have the advantage of giving an interpretable representation of the hydrologic processes and as such are more easily understood and trusted by users.

Since conceptual models are based on strong simplifications of the complex and heterogenous processes occurring in a basin, they usually cannot reproduce all the catchment responses with equal accuracy. The forecast errors induced by these simplifications are usually referred to as structural errors. Other sources of errors are the uncertainty in the model parameters and the errors in the model inputs arising from measurement noise as well as data pre-processing like, for instance, spatial averaging (Gupta *et al*., 2006). When weather forecasts are used to feed the hydrological model, flow forecast errors may be amplified by the uncertainties that propagate from the atmospheric model through the rainfall–runoff process (Pappenberger *et al*., 2011).

*Correspondence to: Francesca Pianosi, Dipartimento di Elettronica, Informazione e Bioingegneria Politecnico di Milano, Italy.
E-mail: francesca.pianosi@polimi.it

Different approaches have been proposed to reduce this uncertainty and improve the predictive accuracy of conceptual models. One option is to combine several conceptual models into a multi-model approach (Link and Barker, 2006). Models can differ in structure or they can share the same structure and have different parameterizations. At each time step, the flow forecast is a combination of the forecasts provided by the different models. The combination takes into account the accuracy that each model demonstrated in hydrological conditions similar to the current ones, and it can be based on expert judgment or on automatic procedures. Different methods, including simple averaging, weighted averaging, neural networks and fuzzy systems, have been discussed and compared in the literature (e.g. Shamseldin *et al*. (1997); Xiong *et al*. (2001); Fenicia *et al*. (2007)). These works consistently demonstrate that using a combination of several models provides more accurate results than using any individual model. Although these findings refer to one-step-ahead flow estimates, it may be reasonably assumed that they also hold over longer lead time.

Another option explored in the literature is to couple the conceptual hydrological model with a data-driven error correction model. The error model exploits the autocorrelation of forecast errors as well as its correlation with other hydroclimatic variables available in real time to predict the future error. This is then removed from the original flow forecast to produce a 'corrected' flow forecast. The identification of the error correction model follows the conventional steps of data-driven modelling, using the historical time series of forecast errors as target output. The task is made more laborious because multiple correction models must be developed for the errors in the various lead time. Brath *et al*. (2002) and Abebe and Price (2003) compare linear and nonlinear (Neural Networks) approaches to forecast errors with a lead time up to 6 hours, while Yu and Chen (2005) use Fuzzy Rules to predict forecast errors up to 4 h. Xiong and O'Connor (2002) compare four different error correction models, a linear Auto-Regressive model, a piecewise linear model, a piecewise linear model with fuzzy threshold and a Neural Network, to improve model simulations with daily time resolution. The conclusion of this work is that linear models provide equivalent performances to more sophisticated models. Goswami *et al*. (2005) also compare different linear and nonlinear models (including Neural Networks and Nonlinear Auto-Regressive eXogenous-Input Model) and extend the lead time up to 6 days using rainfall observations as representative of ideal rainfall forecasts. The adoption of this ideal input scenario allows for a comparison of different error correction models in a setting not influenced by the uncertainty introduced by imperfect rainfall forecasts. The developed error correction model can then be equivalently applied in real time with imperfect forecasts although the final forecasting accuracy is expected to reduce. As historical weather forecasts become available, simultaneously accounting for all uncertainty sources, including errors introduced by rainfall forecast, is becoming a viable option. For instance, Bogner and Kalas (2008) develop an error correction model based on linear and Wavelet Transform to improve flow forecasts up to 7 days, using different weather forecast products. In Bogner and Pappenberger (2011), the lead time is further extended up to 10 days and the error correction system is coupled with an uncertainty processor that provides an estimate of the final predictive uncertainty of the flow forecasting system after output correction.

In all the above mentioned works, forecast errors reflect the limitations of hydrological models in reproducing natural processes because of noisy observations, uncertain parameters, oversimplification of the rainfall–runoff description and uncertainty in input forecasts. However, in many applications, another reason for the discrepancy between flow forecasts and observations is that hydrological models aim at reproducing only natural processes while actual flows are deeply influenced by human activities. A typical example is the alteration of the hydrological regime caused by the operation of dams and barriers on upstream river reaches or tributary rivers. In some catchments, such an alteration is so deep that the flow regime is almost uninfluenced by natural processes for long periods of the year. When this happens, flow forecasting systems cannot rely on models that only explain natural rainfall–runoff processes but they should also include components that attempts at predicting human-induced effects. The effort is becoming more and more urgent as the integration of environmental and engineering systems and thus human–water interactions are deepening while natural, pristine watershed are becoming the exception rather than the rule (Sivapalan *et al*., 2012)

In this paper, we show that data-driven forecast correction models can be used to simultaneously tackle forecast errors from structural, parameter and input uncertainty in reproducing the natural hydrological cycle, and errors that arise from neglecting human-induced alteration. To account for such different error sources, we developed a forecast correction system composed of two layers: (i) a classification system that detect the current flow condition and thus what is the most likely source of errors for the hydrological model, either the uncertainty in reproducing natural rainfall–runoff processes or the unmodelled human influence on such natural processes and (ii) a set of error correction models that are alternatively activated, each tailored to the specific source of errors. Our approach can thus be

viewed as an attempt at combining the above discussed methods of output error correction and multi-model approach. As it will be demonstrated by application to the case study, the resulting correction system can significantly improve flow forecasting while employing relatively simple, e.g. linear, model structures. The case study is the highly anthropized Aniene river basin in Italy, where a flow forecasting system is being established to support the operation of a hydropower dam located at the Tivoli section. Flow forecasts are obtained by combining numerical weather forecasts and a conceptual, lumped hydrological model and then used to assist in the efficient scheduling of power production. While the forecasting system produces rather accurate predictions in high-flow conditions, in low-flow conditions, it is largely inaccurate, especially at the hourly time scale, because it does not consider the hydrological alteration caused by several upstream barriers that are operated for hydropower production.

The paper is organized as follows. The Aniene river basin and the flow forecasting system are briefly described in the next section. The analysis of time series of observed and forecasted flows clearly demonstrate the hydrological alteration induced by the upstream dams and the inadequacy of the hydrological model in reproducing such impacts. In the following section, the error correction system is described. The description is intentionally maintained at a general level in order to highlight the general validity of the proposed approach and its portability to other case study applications. Next, numerical results for the Aniene River are presented, and the proposed approach is critically discussed. Directions for further research are given in the last section.

THE ANIENE RIVER BASIN

The case study area is the Aniene River in Italy (Figure 1). The watershed here considered is closed at the Tivoli section, where the river is dammed and part of the flow diverted to the S. Giovanni hydropower plant. The catchment surface area covers about 690 km^2, with an altitude ranging from 213 m to 2156 m a.s.l and an average of 800 m a.s.l. In the five years period considered in this study (2006–2010), the average precipitation was 1200 mm/year, of which 83 mm/year was snowfall. Snow accumulation and melting affects only a limited part of the basin, with snow covering about the 25% of the basin at its maximum (January) and snowmelt ending by early April.

Upstream of the Tivoli section, seven run-of-river power plants are located with a total installed generation capacity of 84 MW and a mean annual production of about 300 GWh. Five of these seven barriers create small impoundments for a total storing capacity of 770 000 m^3. Dividing this amount by an average river flow rate of 13.5 m^3/s, the total water retention capacity is almost 16 h. It follows that upstream barriers cannot shift water volumes from one day to another, but they can induce within-day fluctuations, with higher flows in peak hours when the electricity demand is higher and lower flows in low demand hours. Such fluctuations are clearly detectable in the time series of observed flow (Figure 2a).

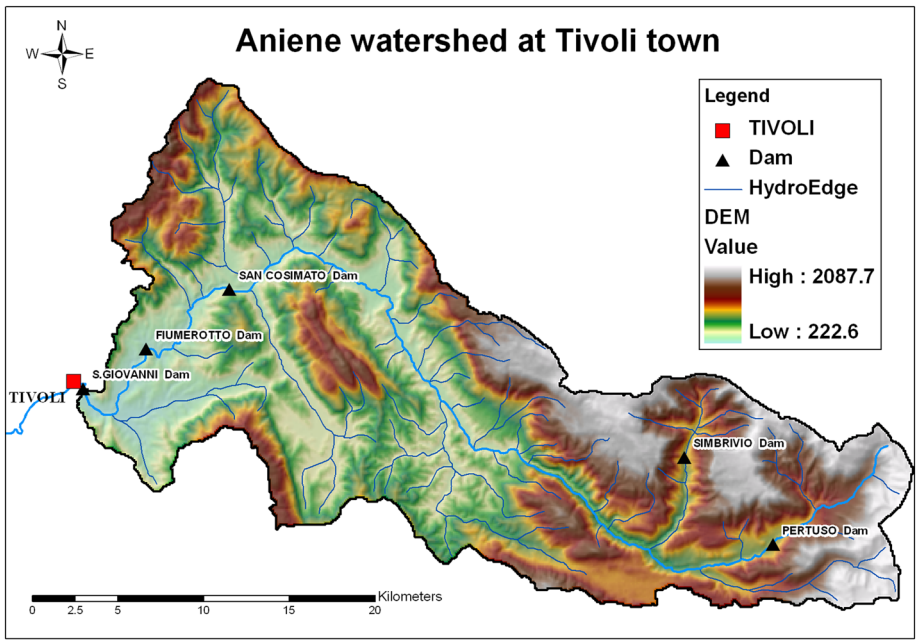

Figure 1. The Aniene basin

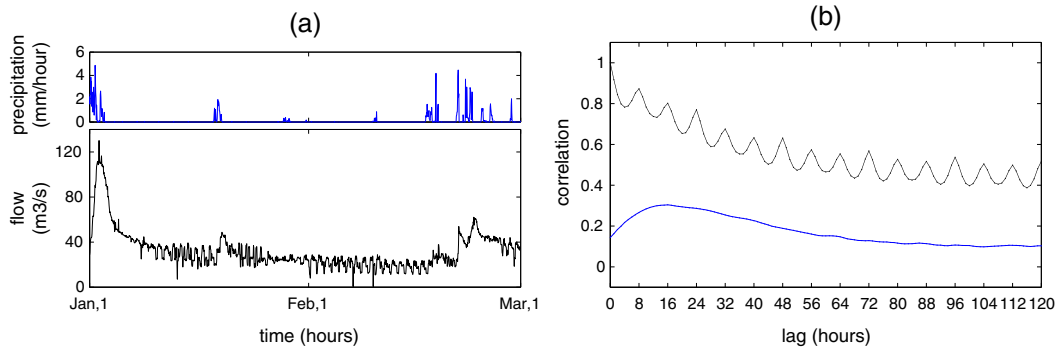

Figure 2. Left: observed rainfall (top panel) and flow (bottom panel) from January to February 2007. Right: cross-correlation function between observed rainfall and flow (blue) and autocorrelation function of the flow (black) over the period 2007–2011

Another evidence of the altered hydrological regime caused by the operation of the upstream barriers is provided by cross-correlation analysis. Figure 2b compares the cross-correlation function between observed rainfall and flow (blue line), and the autocorrelation function of the observed flow (red line). The rainfall–flow cross-correlation has a peak at time lag of 16 (h), which is approximatively the travel time of the catchment. The flow autocorrelation function instead has many local maxima at time lags multiple of 8 h, as a consequence of human-induced fluctuations with a time period of approximately 8 h. For all lag values, the cross-correlation between rainfall and flow is much lower than the flow autocorrelation, which means that the human-induced dynamics of the flow generally overrides the rainfall–runoff process dynamics.

The hydrological model AD2

The hydrological model used to generate flow forecast at the Tivoli section is the lumped conceptual AD2 model (Manfreda and Fiorentino, 2008; Manfreda *et al.*, 2012). The model is composed of several modules that represent different processes at the basin scale: snow accumulation and melting, evapotranspiration, superficial infiltration into the soil, direct overland flow, subsurface flow and deep infiltration into groundwater. The model inputs are average precipitation and temperature in the basin. Separation of rainfall and snowfall uses a simple temperature threshold while snowmelt is estimated based on a degree-day approach. Evapotranspiration is computed as a function of the soil content and the potential evapotranspiration which is estimated using the Hargreaves equation. The soil water balance is described through the use of a bucket scheme taking into account overland flow, interflow and groundwater recharge. Infiltration and overland flow are estimated using a runoff coefficient that depends on the soil water content. Subsurface flow is a fraction of the soil water content when the latter exceeds a given threshold, here set to the 60% of the water storage capacity of the soil. Groundwater recharge is proportional to the soil permeability at saturation and to soil moisture. The total discharge is given by the sum of surface, subsurface and groundwater flow at different time delays.

The AD2 simulation model was calibrated using historical data from the period 2006–2007. MODIS snow cover area data were used to calibrate the degree-day factor. Time series of hourly precipitation, temperature and flow were used to manually calibrate the soil water balance module. Precipitation data from nine rain gauge stations were averaged to produce the precipitation input to the model, while temperature data from seven stations were used to produce two temperature inputs, one for the snow-covered area and one for the uncovered area, assuming a temperature–altitude relationship of 6.5 ° C/km. The comparison of observed and simulated flows shows that the hydrological model is able to effectively reproduce the evolution of floods on a hourly time scale, while low-flow conditions are reproduced with sufficient accuracy only at higher (e.g. daily) time resolution. These limitations will be discussed and explained in the following paragraphs.

Subsequently, the model was set up to be run in forecasting mode. Numerical weather forecasts for the Aniene basin are provided by the numerical model COSMO-LAMI (Limited Area Model Italy) model. The model is run at CINECA, a computing centre managed by a partnership of Italian universities and research institutes. Weather forecasts have a spatial resolution of 7 km, a temporal resolution of 3 h and are produced every morning at 1:00 a.m. for a 72-h lead time. The manager of the S. Giovanni hydropower plant downloads the weather forecast from the CINECA website around 8 o'clock. At that time, the plant manager can also download hourly data of flow, rainfall and temperature collected until 6:00 a.m. These observations are used to estimate the current state of the basin by running the model in simulation mode over the previous 24 h. The state estimate so obtained is then used as the initial condition for another model run, this time forced by weather forecasts, in order to generate the flow forecasts. The procedure is sketched in the upper left part of Figure 3.

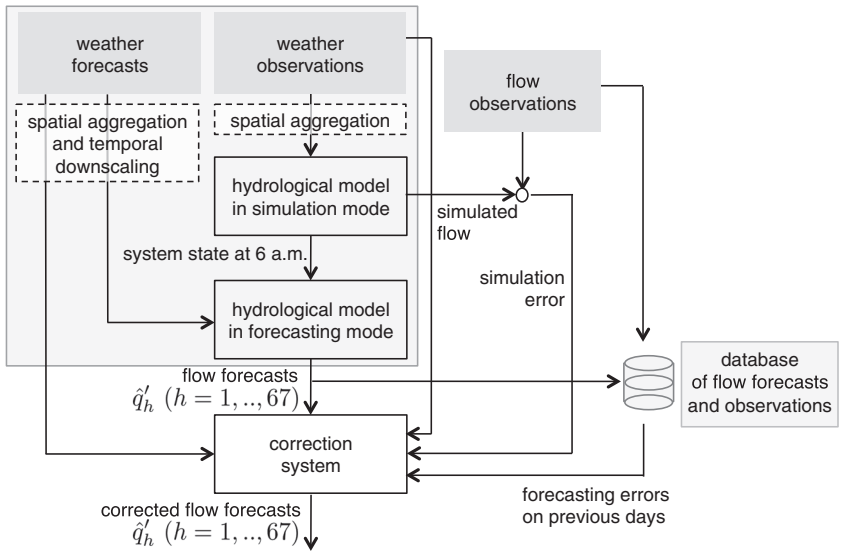

Figure 3. The flow forecasting system, including the hydrological model (grey box) and the forecast correction system

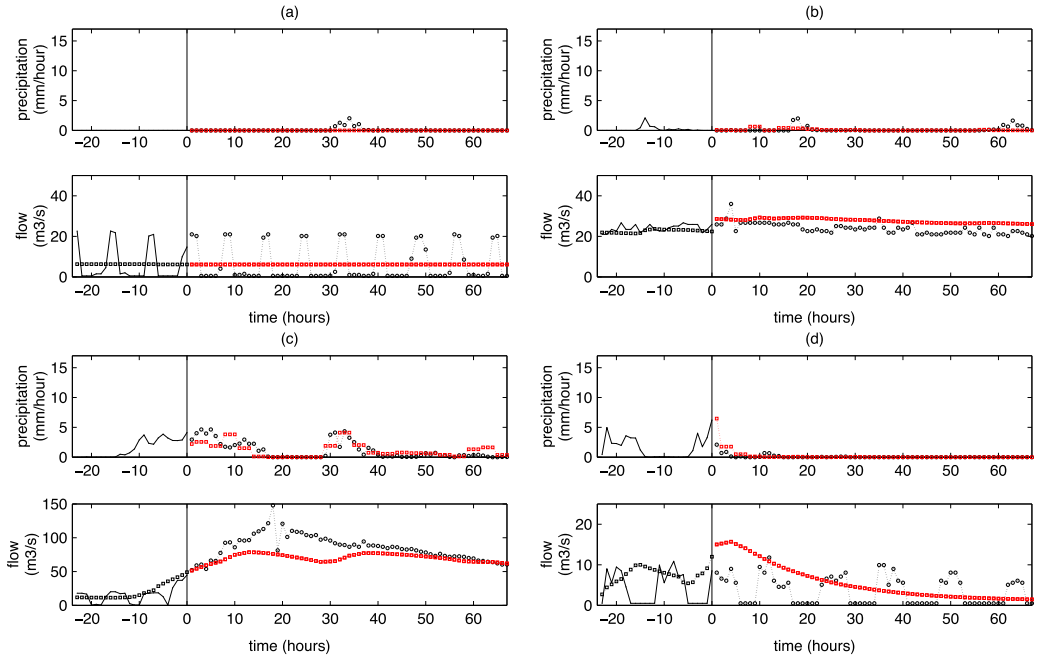

Figure 4. Four typical examples of flow regime. In all plots, black lines and circles are past and future (unknown) observations, red squares are forecast, black squares in bottom panels are model output in simulation mode. Time 0 is the time when the flow forecast is made, i.e. 6 a.m. on a given day (Aug 9, 2007 in (a); Feb 19, 2007 in (b); Mar 22, 2008 in (c); Oct 23, 2009 in (d))

Sources of error in the flow forecasts

The performance of the hydrological model in forecasting mode can be reconstructed over the period 2007–2011 for which numerical weather forecasts were saved and stored. Visual inspection of the time series of observed flows and associated forecasts reveals three main system modes, illustrated by the examples in Figure 4 (panels a–c). In all plots, time 0 refers to the moment when the last hydrometeorological observation is collected, i.e. 6:00 a.m. on a given day. This is also the time for which the state of the hydrological model can be reconstructed via simulation and the initial time of the model run in forecasting mode (see previous paragraph). From now onwards, it will be thus referred to as the *time of forecast*. Note that the forecast horizon is reduced to 67 h because weather forecasts from 1:00 a.m. to 6:00 a.m. are not used in this scheme.

Figure 4a is an example of mode 1 (dry conditions). In mode 1, rainfall events, if any, are small enough to be

intercepted by upstream dams, and the flow pattern is completely determined by the dams' operation. For instance, in Figure 4a, a rainfall event occurred between the 30th and 35th hour in the forecast horizon (black line in the top panel), which however did not reflect into the flow (bottom panel) that follows a fluctuating pattern during the entire forecast horizon. Notice that, in this example, the rainfall event was not predicted by rainfall forecasts (red line in the top panel), and thus flow forecasts (red line in the bottom panel) equal the base flow.

Figure 4b is an example of mode 2 (recession). Just as in mode 1, rainfall is very low, but this time the flow follows a much more natural pattern, which is the recession curve of a flood event that occurred in previous days and, as such, is also quite accurately reproduced by the flow forecast.

Finally, Figure 4c exemplifies mode 3 (flood conditions), where a significant rainfall event occurs in the forecast horizon, increasing both observed and forecasted flows.

The above analysis demonstrates the twofold nature of the forecast errors. In mode 1, the AD2 model cannot accurately forecast flows because they are largely affected by the operation of the upstream barriers, which is not included in the hydrological model. This can be regarded as a kind of structural error in the sense that the hydrological model does not include the human component that also influences the flow formation process. In modes 2 and 3, instead, flow forecasts are closer to observed flows and forecast errors are mainly due to structural, parameter and input uncertainty in the reproduction of the natural rainfall–runoff process.

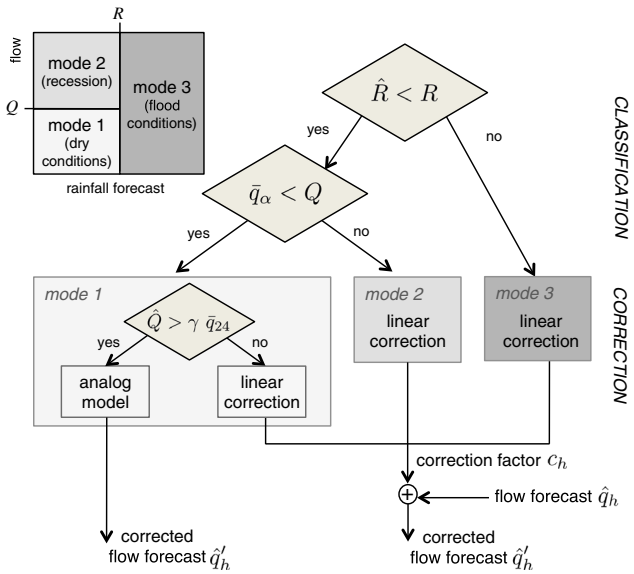

Figure 5. The forecast correction system, including the classification system, the linear correction models and the analog model

THE ERROR CORRECTION MODEL

In order to reduce the errors of the hydrological model, a forecast correction system is developed. The correction system uses the flow forecast provided by the hydrological AD2 model and any other information available at the time of forecast (6 a.m. on a given day) to produce a new flow forecast. This information may include meteorological observations at different time lags but also model errors in simulation mode as well as forecast errors made on previous days (Figure 3). Since the sources of errors of the AD2 model are different depending on the system mode on a given day, the forecast correction should also be different from mode to mode. Therefore, the correction system includes two components: a *classification system* that, based on the meteo-hydrological observation available at the time of forecast, can distinguish the current mode of the system; and a *correction model* that provides an improved flow forecast according to a set of equations specifically tailored for each mode. Figure 5 summarizes the structure of the forecast correction system.

The classification system

In our case study, the classification system should distinguish among three different system modes. Mode 1 is characterized by low rainfall forecast and low-flow conditions. Mode 2 is also characterized by low rainfall forecast but differs from mode 1 because observed flows are higher. Finally, mode 3 is characterized by high rainfall forecast. This classification can be reproduced by the following set of if-then rules:

if $\hat{R} < R$ and $\bar{q}_\alpha < Q$ mode 1
if $\hat{R} < R$ and $\bar{q}_\alpha \geq Q$ mode 2
if $\hat{R} \geq R$ mode 3

where $\bar{q}_\alpha$ is the average observed flow in the α hours before the time of forecast, $\hat{R}$ is the cumulate forecasted rainfall in the next 67 h and R and Q are threshold values to be estimated together with parameter α. In our application, we first gave these parameters a set of tentative values, based on visual inspection of the system trajectories. Later, they were refined through an automatic calibration procedure that will be discussed in Sec. 3.4

The correction models

The simplest correction model takes up the following form

$$\hat{q}'_h = \hat{q}_h + c_h \tag{1}$$

where $\hat{q}_h$ is the flow forecast for the h-th hour in the forecast horizon ($h = 1, \ldots, 67$) produced by the hydrological model on a given day, $\hat{q}'_h$ is the 'corrected' flow forecast and c_h is the

correction factor. The latter is given by an equation of the form

$$c_h = f(\mathbf{z}; \boldsymbol{\theta}_h) \tag{2}$$

where $\mathbf{z}$ is a vector of hydrometeorological information that are available at the time of forecast, and $\boldsymbol{\theta}_h$ is a vector collecting the parameters of the correction function. The choice of the input vector $\mathbf{z}$ and the function f, as well as the calibration of the parameters $\boldsymbol{\theta}_h$, are driven by data analysis. In fact, from the definition of the correction factor c_h in Equation (1), the identification of model (2) boils down to a regression problem over historical time series of forecast errors.

The selection of the model input $\mathbf{z}$ can be based on the analysis of the relation between forecast errors and candidate input variables. Candidate input variables in our case study included: average forecasted rainfall in the prediction horizon; observed flow and rainfall at varying lag times; simulated flows and simulation errors by the AD2 model at different lags and forecast errors made by the AD2 model on previous days. The relation between these variables and forecast errors can be explored by multiple statistical tools, from correlation and average mutual information analysis to more sophisticate input variable selection techniques (Bowden *et al*., 2005). Once the model input have been chosen, regression analysis is applied to choose the proper function $f(\cdot)$ and its parameters. Following a parsimonious approach (Young *et al*., 1996), in our study, we first tested simple relations, i.e. linear, while leaving the option for choosing more complex relations only if the modeling results were considered not satisfactory. Since the correction model is expected to take up different forms depending on the system mode, the above analysis is repeated for each mode using the subset of data relative to the mode under exam.

In our case study, cross-correlation analysis shows that the simulation error made at the time of forecast is a relevant predictor of forecast errors in all modes, dry, flood and recession. The reason is that the simulation error is an indirect estimate of the distance between simulated and actual state of the system, which conditions the model output along the entire forecast horizon.

In mode 1 (dry conditions) and 2 (recession), the forecast errors made on previous days were also selected as input variables, because in these modes similar errors are reproduced for several subsequent days. In mode 3 (flood conditions), instead, forecast errors on previous days are not very informative, because when entering flood conditions, the persistence in the system behaviour is altered by the exogenous meteorological driver. In this mode, the cumulate observed rainfall in the last 24 h proved to be a more informative regressor, as it helps distinguishing whether a rainfall event has already started or not. Following this analysis, in mode 1 and 2, the linear correction model takes up the form

$$c_h = \theta^i_{h1} + \theta^i_{h2} e^{\text{sim}} + \theta^i_{h3} e^{\text{for}}_h \qquad i = 1, 2 \tag{3}$$

where e^{sim} is the simulation error at the time of forecast and e^{for}_h is the last available forecast error for lead time h (i.e. the one produced on the day before for $h < 24$; two days before for $24 \le h < 48$ and three days before for $h \ge 48$).

In mode 3, the model takes up the form

$$c_h = \theta^3_{h1} + \theta^3_{h2} e^{\text{sim}} + \theta^3_{h3} \bar{p}_{24} \tag{4}$$

where $\bar{p}_{24}$ is the cumulate observed rainfall in the last 24 h before the time of forecast. Each correction model thus has three parameters $\left(\theta^i_{h1}, \theta^i_{h2}, \theta^i_{h3}\right)$ for each possible lead time h, for a total of $3 \times 67 = 201$ parameters, which can be easily estimated by applying the linear least-squares formula. As for the linearity assumption, it proved quite satisfactory but for mode 1, where the linear model needed to be further corrected, as explained next.

The analog model

In mode 1 (dry conditions), some events cannot be properly corrected by the simple linear structure described above. Figure 4d is an example. The top panel shows that a rainfall event occurred just around the time of forecast, but observed flows (black line in the bottom panel) are not affected by such rainfall that was probably intercepted by upstream barriers. However, the hydrological model reacts to the rainfall input and generates the flood hydrograph described by the red line. The distance between measured and forecasted flow in this event is so large that adding an empirical correction factor to the latter cannot really close the gap between the two. In this case, the approach of Equation (1) must be abandoned, and the correction model consists in generating a completely new flow forecast $\hat{q}'_h$ independently from the forecast $\hat{q}_h$ produced by the hydrological model.

To generate the new forecast, we could develop a

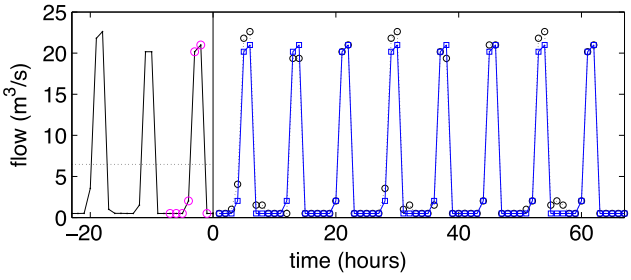

Figure 6. Example of flow forecast produced by the analog model (blue). The analog model replicates the flow pattern (magenta) identified in the observed flows (black) in the 24 h before the time of forecast (time 0). The horizontal dashed line is the threshold value ($\beta \cdot \bar{q}_{24}$) used to distinguish between high and low flows

conceptual reservoir scheme based on profit maximization. However, this is not a viable option in the case study area, since the upstream facilities that induce the alteration of the river regime are operated not only to maximize the profit at the local level but also to account for other needs, for instance load balancing, of the wider pool of plants they belong to. Since reproducing such a complex behaviour by a conceptual model was not straightforward, we opted again for a data-driven approach. Specifically, based on the observation that in dry conditions the hydrological regime often replicates a given pattern for several days, we decided to use a model based on analogy, as sometimes done in weather operational forecasting (Renaud *et al*., 2012). Figure 4d shows an example of such replicated patterns. It is composed of 'high' flows (around 10 m^3/s) for 6–7 h and 'low' flows (around the minimum environmental flow of 1.5 m^3/s) for 5–6 h. The analog model produces the flow forecasts for the next 67 h by replication of the pattern identified in the observed flows of the last 24 h.

Specifically, the analog model performs the following steps:

1. *Test whether the correction model should be replaced by the analog model*. The analog model is used when the flow forecast from the hydrological model is expected to be too far from the actual system conditions. This is tested by comparing the average value of the flow forecasts with the average flow observed in the last 24 h ($\bar{q}_{24}$). Precisely, the analog model is activated when the average flow forecast is higher than $\gamma \cdot \bar{q}_{24}$, where γ is a parameter to be estimated.
2. *Identify the pattern to be replicated*. Hourly flows in the last 24 h are classified in 'high' and 'low' depending on whether they exceed a given threshold or not. The threshold value varies with the event under exam and is computed as $\beta \cdot \bar{q}_{24}$, where β is a parameter to be estimated. The pattern is defined as the flow trajectory between a 'high-to-low' change and a 'low-to-high' change (see Figure 6).
3. *Replicate the pattern*. The corrected flow forecast $q_h^{'}$ is set equal to the flow value in the pattern for the corresponding lead time h.

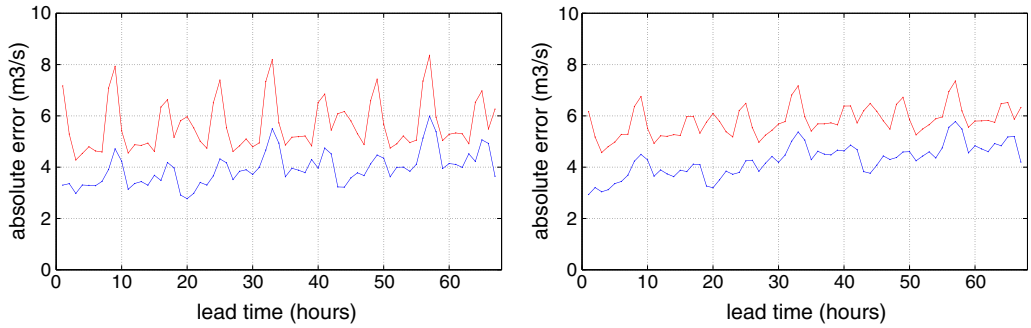

Figure 7. Absolute forecast error before (red) and after (blue) the correction for different lead time h. Left: average over the validation period (2007 and 2009). Right: average over the entire dataset (2007–2011)

Table I. Performance metrics over the calibration (2008, 2010 and 2011) and validation (2007 and 2009) horizon before and after error correction for several lead times (MAE is the mean absolute error; NSE is Nash–Sutcliffe efficiency)

		Lead time (hours)	$h=1$	$h=24$	$h=48$	$h=67$	
Calibration	MAE	before correction	5.45	5.99	6.35	6.37	(m^3/s)
	MAE	after correction (linear)	2.67	3.92	4.55	4.57	(m^3/s)
	MAE	after correction (nonlinear)	2.35	3.76	4.17	4.14	(m^3/s)
	NSE	before correction	0.78	0.67	0.62	0.71	(−)
	NSE	after correction (linear)	0.93	0.76	0.69	0.73	(−)
	NSE	after correction (nonlinear)	0.94	0.78	0.74	0.76	(−)
Validation	MAE	before correction	7.17	6.51	6.59	6.26	(m^3/s)
	MAE	after correction (linear)	3.29	3.68	4.17	3.64	(m^3/s)
	MAE	after correction (nonlinear)	3.29	3.99	4.00	3.26	(m^3/s)
	NSE	before correction	0.71	0.71	0.68	0.83	(−)
	NSE	after correction (linear)	0.90	0.85	0.82	0.84	(−)
	NSE	after correction (nonlinear)	0.89	0.77	0.82	0.85	(−)

The analog model includes the two parameters, γ and β. Just as done for the classification system, these parameters were first manually tuned, and later their value was refined using the parameter estimation procedure described in the next paragraph.

Final parameter refinement

The parameterization of the correction system's components can be obtained using a mix of manual and formal (automatic) calibration procedures. First, the parameters of the classification system are decided and then data split accordingly and the single correction models calibrated.

To improve the parameterization so obtained, all the parameters are re-optimized in a single, automatic calibration process, using a global optimization algorithm suitably modified to our goal. The algorithm explores the space of manually tuned parameters (in our case, the parameters R, Q, α of the classification system, and the parameters β and γ of the analog model). At each iteration, the selected parameter set $(R, Q, \alpha, \beta, \gamma)$ is used to split the calibration dataset, the parameters of each empirical model are estimated by automatic calibration (linear least squares in our case), and the correction model so obtained is evaluated by a performance metric. Based on such metric, the global optimization algorithm selects a new parameter set and iterates the above procedure until a given termination condition is reached.

In this study, a genetic algorithm (Goldberg, 1989) was used (implemented by the *ga* function of the Genetic Algorithm and Direct Search Matlab Toolbox). The performance metric acts as the fitness function of the genetic algorithm. The mean absolute error (MAE) after correction was used in our case study. It is computed as the average across days of the MAE over the forecast horizon made on a given day. The latter is defined as

$$\mathrm{MAE} = 1/67 \sum_{h=1}^{67} |\bar{q}_h - \hat{q}'_h| \qquad (5)$$

MAE is preferred over other more widely used metrics (e.g. mean squared error or Nash–Sutcliffe efficiency (NSE)) because it does not increase the relevance of datapoints of higher magnitude and thus is particularly suitable when calibrating a model that is expected to perform well especially for low errors (Dawson *et al.*, 2007), as in our case. However, other choices may be made depending on the specific purpose of the modelling exercise.

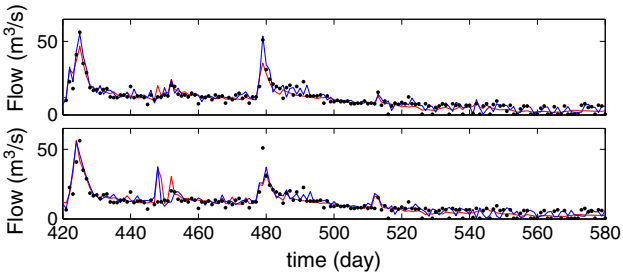

Figure 8. Measured flow (black dots), flow forecast before correction (red) and after correction (blue) with lead time $h = 1$ (top panel) and $h = 24$ h (bottom panel). Zoom of about 5 months in the validation horizon (2007)

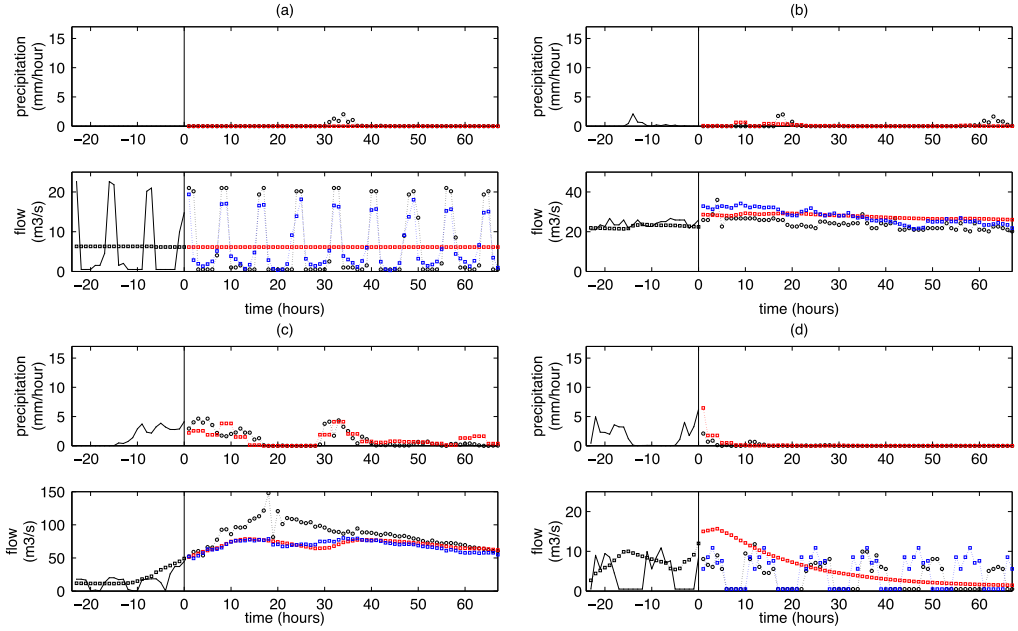

Figure 9. Flow forecast after correction (blue circles) in the four cases of Figure 4. Again, black lines and circles are past and future (unknown) observations, red circles are forecast before correction, black squares are model output in simulation mode

APPLICATION RESULTS AND DISCUSSION

The correction system so obtained was validated by comparing the flow forecasting accuracy before and after the correction, over a time period not used for calibration (years 2007 and 2009) and over the entire available time series (2007–2011). Figure 7a shows the absolute forecast error of the model before (red) and after (blue) the correction, for different lead times (average over the forecasting days). The absolute error obviously increases with the lead time; more interestingly, it is significantly lower after the correction, thus demonstrating the value of the proposed correction system.

The same values are reported in Table I, where besides the MAE of the forecasting system, the NSE is also reported for several lead times. It can be seen that also in terms of the latter metric, performances after correction are significantly improved, especially at lower lead times. As a matter of comparison, the table also reports the same figures for the case of nonlinear correction models, that is, using the same classification system and model inputs but replacing Equations (3) and (4) by nonlinear relations (specifically, two feedforward neural networks with four tangent sigmoid neurons in the hidden layer and a linear output layer). It can be seen that the performances of the nonlinear error correction system are slightly better than those of the linear one over the calibration dataset and almost equivalent over the validation dataset.

A visual inspection of the modelling results is given in Figures 8 and 9. Figure 8 reports the time series of flow forecasts before and after (linear) correction for lead time $h = 1$ and $h = 24$. It shows that overall, the corrected flow forecast follows more closely the measured flows, although the forecasting accuracy is obviously much lower for higher lead time. Figure 9 analyzes the flow forecast after correction for the four typical events already reported in Figure 4. It shows that the corrected flow forecast (blue line) is much closer to observed flows (black) especially when the system is in mode 1, i.e. in case (a) and (d) (both in the validation period).

Validation results show that the correction system can significantly contribute to improve flow forecasting accuracy on the proposed case study, despite the relative simplicity of its single components. On the other hand, the modular structure of the correction system allows for easily improving any of its components whenever increased data availability would permit it or when applying the same methodology to a different case study. The classification system can be easily scaled to more than three modes and/or be based on more complex rules than those used here. A review of the available approaches to building modular models for hydrological applications can be found in Solomatine and Siek (2006). Correspondingly, the correction models for each mode can be tailored to the regime under study, just as we did in mode 1 with the analog model.

However, increasing the complexity of the input–output relationship of Equation (2) may significantly increase the computing time for the parameter refinement procedure. With this respect, using linear relationships as we did in our study is quite effective since at each iteration of the global optimization, linear least squares can be used to calibrate the correction models and thus the refinement procedure is very fast. For instance, the results reported in this paper were obtained using a population of 20 individuals and 51 generations, for a total of $20 \times 51 = 1020$ model evaluations, which requires applying the linear least-squares formula $1020 \times 3 = 3060$times, for a total computing time of less than 3 min (on a laptop with 2.53 GHz Intel Core 2 Duo Processor, 4 GB 1067 MHz memory).

In our case study, limited data availability and thus the risk of overfitting was another reason for preferring simple relations like linear ones. On the other hand, since a different linear relation is identified for each mode of the system, we expect it to be a more reasonable approximation of the input–output relation than if one relation was used for the entire regressor space. In other terms, since we use different model in different domains, we expect the linear approximation to be more acceptable in each domain. Such a modular approach has been previously demonstrated to be effective in rainfall–runoff modelling (Iorgulescu and Beven, 2004; Solomatine and Xue, 2004). The hypothesis is also confirmed by the comparison with a more complicated, neural network model, which does not improve forecasting efficiency significantly on our case study.

Another advantage of the modular approach is that it enhances interpretability of the modelling results in hydrologically sound terms. While building the correction system, the modeller can gain a deeper understanding of the variety of catchment responses and the sources of errors in the hydrological model. Once the model is calibrated and implemented, it provides the user with an indication of what is the current mode of the system, why correction is needed, and even the possibility not to activate the correction model if the original hydrological model is considered sufficiently reliable in the current mode (as it could be for mode 3 in our case).

CONCLUSIONS

In this paper, we proposed a modular data-driven flow forecast correction model composed of a classification system that activates alternative error correction routines based on the current flow conditions. The classification detects whether the source of error is natural or human

induced and correspondingly launches a data-driven model tailored to the identified source of error. This very simple approach exhibits a number of advantages with respect to more sophisticated approaches such as: (i) it is computational efficient and easily implementable; (ii) help in gaining a deeper understanding of the catchment behaviour and the sources of errors in the original hydrological model and (iii) is highly portable and does not require any specific data, e.g. on the human-induced variability, to detect and correct flow forecasting error.

As a case study, we considered the highly anthropized Aniene river basin in Italy, where a flow forecasting system is being established to support the operation of a hydropower dam. Results show that, even by using very basic methods, namely if-then classification rules and linear correction models, the proposed methodology considerably improves the forecasting capability of the original hydrological model under different flow regimes.

Future research will focus on studying the uncertainty associated to the corrected hydrological flow forecast for different lead times and developing suitable methods for quantifying the residual uncertainty remaining after error correction. This is a step of outmost importance to enhance the reliability of the operational flow forecasting system and assist decision makers in model-based operation while recognizing uncertainty.

ACKNOWLEDGEMENTS

This work has been financed by the Research Fund for the Italian Electrical System under the Contract Agreement between RSE S.p.A.(Research for Energetic System) and the Ministry of Economic Development - General Directorate for Nuclear Energy, Renewable Energy and Energy Efficiency, stipulated on July 29, 2009 in compliance with the Decree of March 19, 2009. The data used in this study were provided by the Regional Hydrographic service (Rome) and by ENEL-Green Power.

REFERENCES

Abebe A, Price R. 2003. Managing uncertainty in hydrological models using complementary models. *Hydrological Sciences Journal* **48**(5): 679–692.

Bogner K, Kalas M. 2008. Error-correction methods and evaluation of an ensemble based hydrological forecasting system for the Upper Danube catchment. *Atmospheric Science Letters* **9**(2): 95–102.

Bogner K, Pappenberger F. 2011. Multiscale error analysis, correction, and predictive uncertainty estimation in a flood forecasting system. *Water Resources Research* **47**(W07525).

Bowden G, Dandy G, Maier H. 2005. Input determination for neural network models in water resources applications. part 1 - background and methodology. *Journal of Hydrology* **301**(1–4): 75–92.

Brath A, Montanari A Toth E. 2002. Neural networks and non-parametric methods for improving real-time flood forecasting through conceptual hydrological models. *Hydrology and Earth System Sciences* **6**(4): 627–639.

Dawson C, Abrahart R, See L. 2007. Hydrotest: A web-based toolbox of evaluation metrics for the standardised assessment of hydrological forecasts. *Environmental Modelling & Software* **22**(7): 1034–1052.

Fenicia F, Solomatine Dk, Savenije H, Matgen P. 2007. Soft combination of local models in a multi-objective framework. *Hydrology and Earth System Sciences* **11**(6): 1797–1809.

Goldberg D. 1989. *Genetic Algorithms in Search, Optimization and Machine Learning*. Addison Wesley: Reading, MA.

Goswami M, O'Connor KM, Bhattarai KP, Shamseldin AY. 2005. Assessing the performance of eight real-time updating models and procedures for the Brosna River. *Hydrology and Earth System Sciences* **9**(4): 394–411.

Gupta HV, Beven KJ, Wagener T. 2006. Model Calibration and Uncertainty Estimation, in: *Encyclopedia of Hydrological Sciences* edited by: Anderson M, John Wiley & Sons Ltd, Chichester, 1–17.

Habets F, LeMoigne P, Noilhan J. 2004. On the utility of operational precipitation forecasts to served as input for streamflow forecasting. *Journal of Hydrology* **293**(1–4): 270–288.

Iorgulescu I, Beven K. 2004. Nonparametric direct mapping of rainfall-runoff relationships: An alternative approach to data analysis and modeling? *Water Resources Research* **40**(W08403).

Link WA, Barker RJ. 2006. Model weights and the foundations of multimodel inference. *Ecology* **87**(10): 2626–2635.

Manfreda S, Fiorentino M. 2008. *Mountains: Sources of Water, Sources of Knowledge*, vol. **31**. Springer, Ch. Flood Volume Estimation and Flood Mitigation: Adige river basin; pp. 243–264.

Manfreda S, Funicelli L, Mancusi L. September 10th-15th 2012. Previsione idrologica per la gestione degli impianti idroelettrici. In: Proceedings of the XXXIII Conference of Hydraulics and Hydraulic Engineering. Brescia, Italy.

O'Connor KM. 1992. Special Issue on River Flow Forecasting. *Journal of Hydrology* **133**: 1–178.

Pappenberger F, Thielen J, Del Medico M. 2011. The impact of weather forecast improvements on large scale hydrology: analysing a decade of forecasts of the european flood alert system. *Hydrological Processes* **25**: 1091–1113.

Renaud M, Zin I, Obled C, Bontron G, Djerboua A. 2012. Toward real-time daily PQPF by an analog sorting approach: application to flash-flood catchments. *Journal of Applied Meteorology and Climatology* **51**(3): 505–520.

Shamseldin AY, O'Connor KM, Liang GC. 1997. Methods for combining the output of different rainfall-runoff models. *Journal of Hydrology* **197**: 203–229.

Sivapalan M, Savanjie H, Blöschl G. 2012. Socio-hydrology: A new science of people and water. *Hydrological Processes* **26**(8): 1270–1276.

Solomatine D, Siek M. 2006. Modular learning models in forecasting natural phenomena. *Neural Networks* **19**(2): 215224.

Solomatine D, Xue Y. 2004. M5 model trees compared to neural networks: application to flood forecasting in the upper reach of the Huai River in China. *ASCE Journal of Hydrologic Engineering* **9**(6): 491–501.

Xiong L, O'Connor K. 2002. Comparison of four updating models for real-time river flow forecasting. *Hydrological Sciences Journal* **47**(4): 621–640.

Xiong L, Shamseldin AY, O'Connor KM. 2001. A non-linear combination of the forecasts of rainfall-runoff models by the first-order takagi-sugeno fuzzy system. *Journal of Hydrology* **245**(1–4): 196–217.

Young P, Parkinson S, Lees M. 1996. Simplicity out of complexity in environmental modelling: Occam's razor revisited. *Journal of Applied Statistics* **23**(2–3): 165–210.

Yu P-S, Chen S-T. 2005. Updating real-time flood forecasting using a fuzzy rule-based model. *Hydrological Sciences Journal* **50**(2): 265–278.