Quantification and Reduction of Predictive Uncertainty for Sustainable Water Resources Management (Proceedings of Symposium HS2004 at IUGG2007, Perugia, July 2007). IAHS Publ. 313, 2007.

Data assimilation in a large-scale distributed hydrological model for medium-range flow forecasts

ADRIANO ROLIM DA PAZ, WALTER COLLISCHONN, CARLOS E. M. TUCCI, ROBIN T. CLARKE & DANIEL ALLASIA

Instituto de Pesquisas Hidráulicas, Universidade Federal do Rio Grande do Sul (IPH/UFRGS), Av. Bento Gonçalves, 9500, CEP 91501-970, Porto Alegre (RS), Brazil

adrianorpaz@yahoo.com.br

Abstract As part of a research project aimed to improve medium-range streamflow forecasts used in the operational planning of Brazilian hydro-electric power systems, a large-scale distributed hydrological model has been used to obtain streamflow forecasts for up to 12 days in advance using observed and predicted precipitation data. Observed streamflow data up to the time of forecast start were used to update state variables calculated by the model. This data assimilation was performed by applying an empirical procedure. Several configurations of this empirical procedure have been tested and this paper presents results obtained in applying it to the Rio Grande basin, one of the test beds of the Hydrologic Ensemble Prediction Experiment (HEPEX), focusing on forecasts at Furnas sub-basin (51 900 km^2). Results show that flow forecasts were not improved by updating state variables related to river flow, while significant improvements were obtained by updating state variables related to groundwater storage.

Key words flow forecasting; data assimilation; distributed hydrological model

INTRODUCTION

A key aspect in the application of any hydrological model to obtain real time streamflow forecasts is the updating or data assimilation procedure (Refsgaard, 1997). Hydrological models can be operated in simulation mode or in adaptive mode. In simulation mode, the model output is based on previous model inputs, such as rainfall. In adaptive mode, the model output is based on previous model inputs as well as on previous observed outputs, which are used to update the model before a new forecast is issued. For real-time forecasting it is necessary to have a model operating in adaptive mode (Moore *et al.*, 2005) to account for uncertainties in input data and inadequacies of model structure, its parameter values and initial conditions.

Model updating or data assimilation procedures can be classified according to the variables that are modified: input variables; model states; model parameters and output variables (Madsen & Skotner, 2005). The updating procedures that are used most widely change the state variables or the output variables. A very common approach to model updating focuses on the prediction of future model errors, based on past model errors. Toth *et al.* (1999), for instance, used ARMA models to predict forecasting errors of a deterministic rainfall–runoff model, and Goswani *et al.* (2005) assessed the performance of eight real-time updating procedures mostly based on error prediction.

The advantage of this approach is that it can be easily applied to complex models such as full hydrodynamic flood propagation models (Madsen & Skotner, 2005).

State variables updating can be based on observed errors in river flow, and can use empirical methods or more formal Kalman filtering (Moore *et al.*, 2005; Romanowicz *et al.*, 2006). For more complex distributed and nonlinear models, Kalman filtering usually lead to highly complex computations (O'Connell & Clarke, 1981), but recent studies have developed and applied computationally cost-effective approaches for such methods (e.g. Madsen & Stokner, 2005; Canizares *et al.*, 2001). We developed an empirical data assimilation procedure to be used in conjunction with the large-scale MGB-IPH model and applied it to obtain medium range streamflow forecasts (up to 12 days) for the Rio Grande, in Brazil, using observed and predicted rainfall as input. This paper presents a description of the updating procedure, and results of tests that were performed by varying the configuration and parameter values of this procedure.

METHODS

Configuration of forecasting tests

Streamflow forecasts were obtained by running the large-scale hydrological model MGB-IPH (Collischonn & Tucci, 2001; Allasia *et al.*, 2006) using observed and forecast input data. The hydrological model was run in continuous simulation mode, in daily time steps, using observed rainfall data up to the time of forecast issue. From this time up to the end of the next 10 days, Quantitative Precipitation Forecasts (QPF) were used. Successive forecasts were made every week, starting on Wednesday, and extending for 12 days up to Friday of the following week, considering that no rainfall would occur at the last two days, when no QPFs were available. For each forecast the hydrological model was run with observed rainfall data during a warm-up period that lasted a few months, extending up to the preceding Tuesday. Calculated and observed discharge values were compared at several sites, and the updating procedure was applied each day during this warming up period.

Hydrological model

The hydrological model used was the MGB-IPH large scale model, which is composed of modules for calculating the soil water budget, evapotranspiration, flow propagation inside a cell, and flow routing through the drainage network. The drainage basin is divided into elements of area (normally on a square grid) connected by channels. The Grouped Response Unit (GRU) (Kouwen *et al.*, 1993) approach is used for hydrological classification of all areas with a similar combination of soil and land cover without consideration of their exact locality within the grid (or cell). A cell contains a limited number of distinct GRUs. The soil water budget is computed for each GRU, and runoff generated from the different GRUs in the cell is then summed and routed through the river network. Flow generated within each cell is routed to the stream network using three linear reservoirs (baseflow, subsurface flow and surface flow). Streamflow is propagated through the river network using the Muskingum-Cunge method.

The model was calibrated by changing values of parameters while maintaining relations between land use and parameter values (Collischonn *et al.*, 2007). The multi-objective MOCOM-UA optimization algorithm (Yapo *et al.*, 1998) was employed, considering three objective-functions: volume bias (ΔV); Nash-Sutcliffe model efficiency for streamflow (NS); and Nash-Sutcliffe for the logarithms of streamflow ($NS_{\log}$). The NS coefficient is given by:

$$NS = 1 - \frac{\sum \left[Q_{obs}(t) - Q_{calc}(t) \right]^2}{\sum \left[Q_{obs}(t) - \overline{Q_{obs}} \right]^2} \tag{1}$$

where $Q_{\text{obs}}(t)$, $Q_{\text{calc}}(t)$ are the observed and calculated discharges at time step t, and $\overline{Q_{obs}}$ is the average observed discharge.

Precipitation forecasts

Quantitative precipitation forecasts were obtained 10 days in advance, with a horizontal resolution of about 40 km, from the regional ETA model which is being run operationally by the Brazilian Center for Weather Prediction (Chou, 1996; Chou *et al.*, 2000).

Data assimilation procedure

The data assimilation or updating procedure described here is an improvement to the method described by Collischonn *et al.* (2005). Updated variables are streamflow values calculated along the river network, and water content in the groundwater reservoir in each model cell. The updating method consists of continuously comparing observed and calculated flows during a warming up period prior to forecast issue. An updating correction factor (FCA) is calculated for each gauging station p where observed streamflow is available, through the expression:

$$FCA_p = \sum_{t=t_0-t_a}^{t_0} Q_{obs}^t \Bigg/ \sum_{t=t_0-t_a}^{t_0} Q_{calc}^t \tag{2}$$

where Q_{obs} and Q_{calc} are observed and calculated streamflow, respectively; t is the time step; t_0 is the time of forecast issue; t_a is the averaging time.

The correction factors are applied to correct streamflow variables for each cell located upstream of each gauging station, using as a weighting factor the area drained by each cell. At the cell where the streamgauge is located, observed flows were used in place of calculated ones. For cells close to the streamgauge, this scheme assumes that flow recorded at the streamgauge is virtually correct. For cells far upstream from the gauge, calculated flows are assumed to be more reliable, and corrections are damped out according to equation (3):

$$Qup_{i,p} = FCA_p \cdot Qcalc_i \cdot \left(A_i / A_p \right)^{ebac} + Qcalc_i \cdot \left[1 - \left(A_i / A_p \right)^{ebac} \right] \tag{3}$$

where $Qup_{i,p}$ is the updated value of discharge at cell i, located upstream of p, A_i and

A_p are the drainage areas upstream of cell i and gauging station p, respectively; *ebac* is an updating parameter with values between 0 and 1.

The updating procedure described above refers to the river discharge variable at several cells. A similar updating procedure was adopted to correct volumes in groundwater storage. Each cell of the model has three linear reservoirs that represent the retention and delay of water subsequently released as surface, subsurface and groundwater flow. Outflow from these reservoirs in each cell becomes inflow to the river network where it is routed using the Muskingun-Cunge method. During long, dry periods, the greater part of flow comes from groundwater storage. The model maintains a continuous record of the fraction of flow in the drainage network that comes from surface, subsurface and groundwater. Groundwater storage in each cell upstream of streamgauge p is updated using the same correction factor (*FCA*) used for river flow. Unlike river discharge, groundwater storage updating is not weighted by drainage area relations between cell and gauging point, but by the fraction of river flow that is of groundwater origin (PB_i), according to equation (4):

$$VBup_{i,p} = \left(FCA_p\right)^{bx} \cdot VB_i \cdot \left(PB_i\right) + VB_i \cdot \left(1 - PB_i\right) \quad (4)$$

where $VBup_{i,p}$ is the updated storage in the groundwater reservoir of cell i; VB_i is the calculated storage at cell i; PB_i is the fraction of river flow at cell i originated from groundwater and *bx* is an updating parameter with values between 0 and 1. When *bx* is close to 1, groundwater updating is relatively quick; when *bx* is close to 0, the correction is somewhat smoothed, therefore taking more time steps to make the necessary corrections. On the other hand, smaller values of this parameter lead to more stable results, since real-time observed streamflow values may have random errors that would, otherwise, result in overcorrection. Another parameter called *PBlim* is introduced in order to denote the minimum fraction necessary to apply such correction. Several configurations of this empirical updating procedure were tested, varying the values of corresponding parameters, in order to provide a better understanding of the effect of each parameter over the quality of flow forecasts.

Study area

The Rio Grande is the main tributary of the River Paraná in its upper basin and drains an area of about 145 000 km^2 (Fig. 1). Mean annual rainfall over the basin is approximately 1400 mm and is highly concentrated during summer. In order to test the data assimilation procedure, medium-range flow forecasts were obtained at Furnas reservoir (drainage area 51 900 km^2), which is one of the main hydropower reservoirs of the basin.

RESULTS AND DISCUSSION

Calibration and verification of hydrological model

Hydrological model parameters were calibrated for each sub-basin for the period 1970–1980 while the period 1981–2001 was used for model validation. As a result of the multi-objective optimization, several Pareto optimal solutions were found, and a

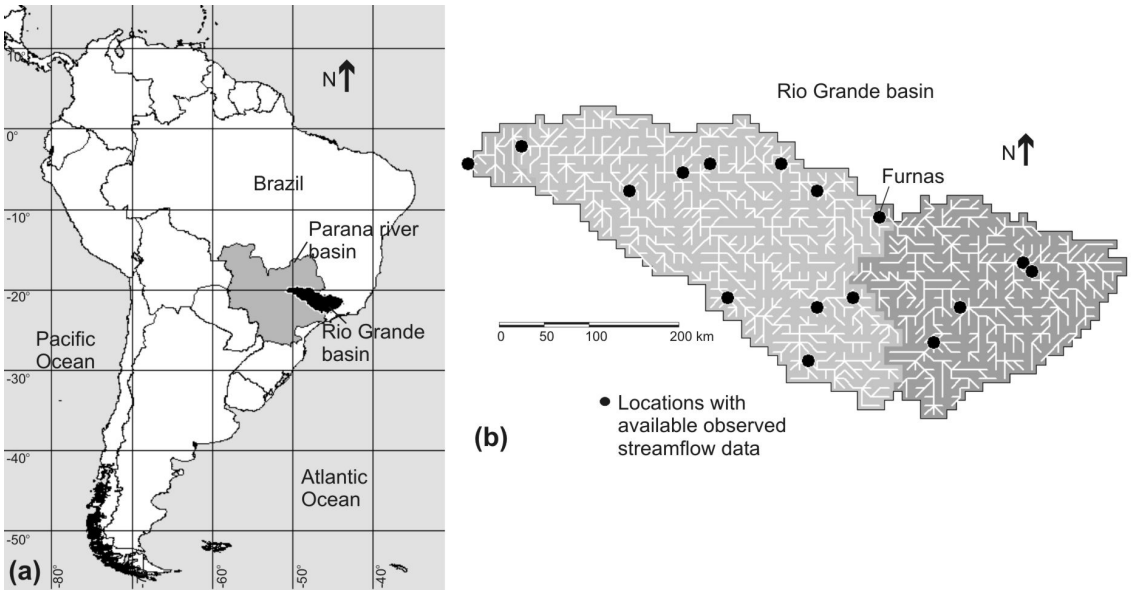

Fig. 1 (a) Location of the Rio Grande basin and (b) its representation in the hydrological model (0.1° square grid cells) with location of Furnas Reservoir.

single solution was chosen among them, aiming to provide an acceptable trade-off in fitting of the different parts of the hydrograph, or the different objective-functions, as suggested by Bastidas *et al*. (2002). In both calibration and verification, the values obtained for *NS* and NS_{log} coefficients were about 0.9 in all but one of the sub-basins. Values of volume bias were also acceptable, with values less than 0.05% during calibration and less than 7% at validation.

Flow forecasts

A total of 25 different configurations of the empirical updating procedure were tested with varying parameters values, but for the sake of simplicity just five configurations were selected for presentation here (Table 1). The *NS* coefficients related to flow forecasts at Furnas reservoir for different lead times are shown in Fig. 2. Results are shown by using, respectively, QPF of the ETA model and perfect rainfall forecasts. One can note that that *NS* values varied slightly with respect to parameter *ebac*, as the curves related to configurations 1, 2 and 3 are quite similar. This result means that the parameter related to correction of the river flow variable in each cell did not influence the flow forecasts. This result could be expected since the daily time step of the data is inadequate in relation to the hydrological characteristics of the basin. Updating of the streamflow variables may have an impact on forecasts for only a few hours.

Table 1 Configurations of the empirical updating procedure.

Parameter	Configuration:				
	1	2	3	4	5
ebac	1.0	0.2	0.0	0.2	0.2
bx	0.2	0.2	0.2	1.0	0.2
PBlim	0.3	0.3	0.3	0.3	0.1

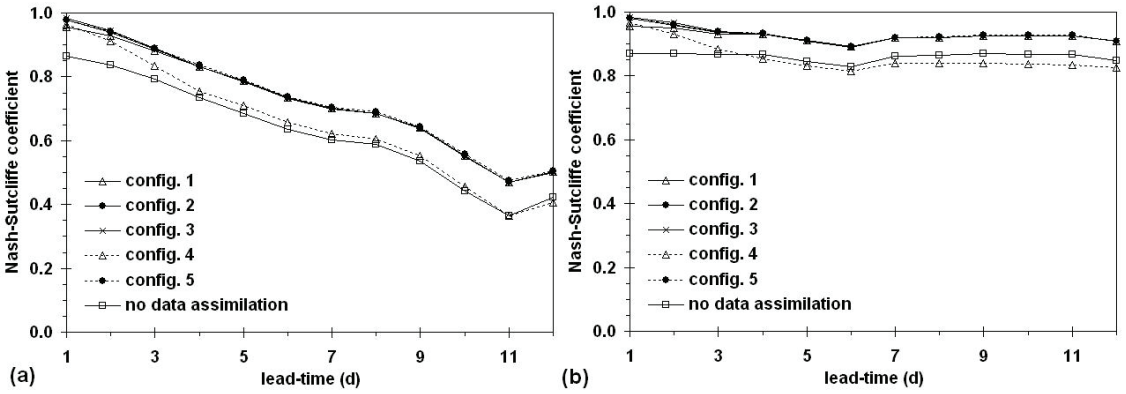

Fig. 2 Nash-Sutcliffe coefficients as a function of lead time and of the updating procedure configuration using: (a) QPF of ETA model, and (b) the perfect rainfall forecast.

The quality of forecasts in terms of *NS* values was more sensitive to variations in the exponent *bx* that is used to control the velocity of correction in the groundwater linear reservoir of each cell. Using *bx* equal to 1 (configuration 4) resulted in *NS* values almost 10% smaller than when using *bx* equal to 0.2, for lead times greater than 4 days and using QPF. These results suggest that groundwater storage variables should be updated relatively slowly, in order to avoid the excessive correction that could be related to noise or errors in streamflow observations.

The values of 0.1 and 0.3 were tested for parameter *PBlim*, which represent a minimum fraction of 10% and 30% respectively, of water in the river flow that originated from groundwater (PB_i) necessary to start the updating of groundwater storage. No meaningful differences in the forecasts were obtained varying the parameter *PBlim* in this range.

The same influence of the parameter values over the *NS* coefficients was observed when using perfect rainfall forecasts. However, in the latter, *NS* values ranged from 0.82 to 0.98 for all lead-times. Using QPF of the ETA model, *NS* decreased as a function of lead-time reaching a minimum of 0.47 at an 11 day lead-time. If no data assimilation procedure is adopted, *NS* values decay approximately by 0.1 when using QPF of the ETA model and by 0.06 when using perfect rainfall forecasts. To illustrate the performance of the flow forecasts, results are presented for some periods for the outlet of the Furnas sub-basin (Fig. 3). The traces on each graph show the forecasts of hydrographs issued weekly (on Wednesdays) up to the 12 day time horizon, so that there are several such traces shown, one for each run of the hydrological model, and shown by different symbol styles. In general, the agreement is good, principally at forecasting the time of hydrograph rise and the magnitude of the peak corresponding to each rainfall event. One can note the importance of data assimilation for flow forecasting at the beginning of each 12 day forecast, by the small difference between observed and forecast values. For comparison, flow forecasts were also made using the perfect rainfall forecast (Fig. 3b). As expected, the quality of flow forecasts is highly improved, showing that relatively large errors are originated by erroneous rainfall forecasts.

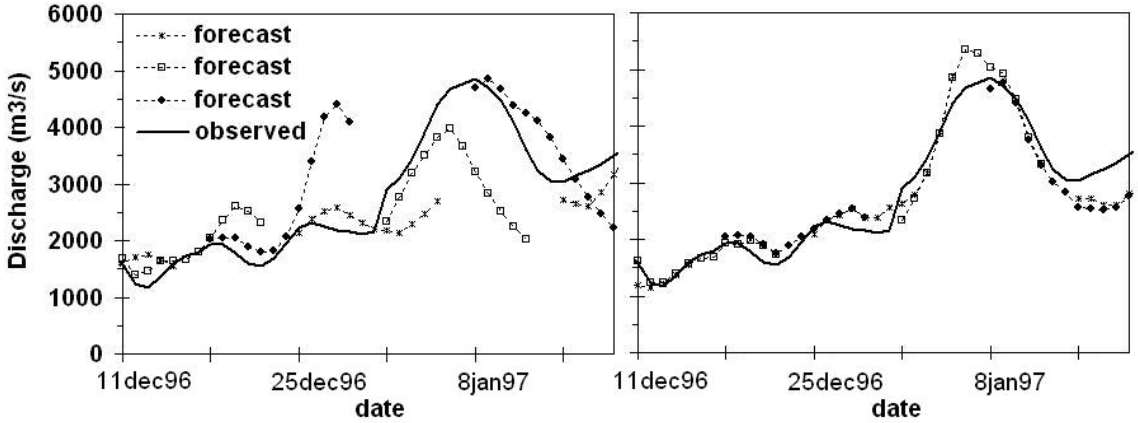

Fig. 3 Medium-range flow forecasts at the outlet of the Furnas sub-basin: (a) using rainfall forecasts from ETA model, (b) using perfect rainfall forecasts. Symbol styles are used only to distinguish between different runs of the hydrological model, giving the 12-day-ahead forecasts initiated every Wednesday.

SUMMARY AND CONCLUSIONS

This paper describes a data assimilation procedure which is used with a large-scale hydrological model in order to update the values of several variables in the model using observed streamflow at only a few gauging stations in a basin. The updating procedure largely improves the quality of flow forecasts of the model in the prediction range of 1–12 days. Updated parameters were tested and results suggest that observed data should be used only to update variables related to baseflow (or groundwater storage), due to the relatively rapid response of the basin and low frequency of observations (one daily).

Due to possible random errors in observations, no excessive weight or confidence should be given to any individual streamflow information. Therefore, corrections should be smoothed out by selecting relatively low values of the empirical parameter *bx*, leading to slow corrections. It should be noted that the same updating procedure could be used in cases with hourly data availability, and slower basin response, but other optimal updating parameters will probably be found.

Acknowledgements Financial assistance for this research was provided by FINEP/CT-Hidro (Financiadora de Estudos e Projetos) from the Brazilian Ministry of Science and Technology (MCT).

REFERENCES

Allasia, D., Collischonn, W., Silva, B. C. & Tucci, C. E. M. (2006) Large basin simulation experience in South America. In: *Predictions in Ungauged Basins: Promises and Progress* (ed. by M. Sivapalan, T. Wagener, S. Uhlenbrook, E. Zehe, V. Lashmi, X. Liang, Y. Tachikawa & P. Kumar), 360–370. IAHS Publ. 303. IAHS Press, Wallingford, UK.

Bastidas, L. A., Gupta, H. V. & Sorooshian, S. (2002) Emerging paradigms in the calibration of hydrologic models. In: *Mathematical Models of Large Watershed Hydrology* (ed. by V. Singh & D. K. Frevert), 25–66. Water Resources Publications, Littleton, Colorado, USA.

Canizares, R., Madsen, H., Jensen, H. R. & Vested, H. J. (2001) Developments in operational shelf sea modelling in Danish waters. *Estuarine, Coastal and Shelf Science* **53**, 595–605.

Chou, S. (1996) The regional model ETA (in Portuguese). *Climanálise*. Special Edition, INPE.

Chou, S., Nunes, A. & Cavalcanti, I. (2000) Extended range forecasts over South America using the regional ETA model. *J. Geophys. Res.* **105**(D8), 10147–10160.

Collischonn, W. & Tucci, C. (2001) Hydrological simulation of large drainage basins (in Portuguese). *Brazilian J. Water Resour.* **6**(1), 15–35.

Collischonn, W., Haas, R., Andreolli, I. & Tucci, C. (2005) Forecasting River Uruguay flow using rainfall forecasts from a regional weather-prediction model. *J. Hydrol.* **205**, 87–98.

Collischonn, W., Allasia, D., Silva, B. C. & Tucci, C. E. M. (2007) The MGB-IPH model for large scale rainfall runoff modeling. *Hydrol. Sci. J.* (in press).

Goswani, M., O'Connor, K., Bhattarai, K. P. & Shamsedlin, A. Y. (2005) Assessing the performance of eight real-time updating models and procedures for the Brosna River. *Hydrol. Earth System Sci.* **9**(4), 394–411.

Kouwen, N., Soulis, E., Pietroniro, A., Donald, J. & Harrington, R. (1993) Grouped response units for distributed hydrologic modeling. *J. Water Resour. Plan. Manage.* **119**(3), 289–305.

Madsen, H. & Skotner, C. (2005) Adaptive state updating in real-time river flow forecasting – a combined filtering and error forecasting procedure. *J. Hydrol.* **308**, 302–312.

Moore, R. J., Bell, V. A. & Jones, D. A. (2005) Forecasting for flood warning. *Comptes Rendus Geosci.* **337**, 203–217.

O'Connell, P. E. & Clarke, R. T. (1981) Adaptive hydrological forecasting – a review. *Hydrol. Sci. Bull.* **26**(2), 179–205.

Refsgaard, J. C. (1997) Validation and intercomparison of different updating procedures for real-time forecasting. *Nordic Hydrol.* **28**, 65–84.

Romanowicz, R. J., Young, P. C. & Beven, K. J. (2006) Data assimilation and adaptive forecasting of water levels in the river Severn catchment, United Kingdom. *Water Resour. Res.* **42**(W06407), doi:10.1029/2005WR004373.

Toth, E., Montanari, A. & Brath, A. (1999) Real-time flood forecasting via combined use of conceptual and stochastic models. *Phys. Chem. Earth B* **24**(7), 793–798.

Yapo, P. O., Gupta, H. V. & Sorooshian, S. (1998) Multi-objective global optimization for hydrologic models. *J. Hydrol.* **204**, 83–97.