Weather Radar Information and Distributed Hydrological Modelling (Proceedings of symposium HS03 held during IUGG2003 at Sapporo, July 2003). IAHS Publ. no. 282, 2003.

Real-time assimilation of radar-based precipitation data and streamflow observations into a distributed hydrological model

DONG-JUN SEO, VICTOR KOREN & NEFTALI CAJINA
Hydrology Laboratory, Office of Hydrological Development, National Weather Service, 1325 East-West Highway, Silver Spring, Maryland 20910, USA.
dongjun.seo@noaa.gov

Abstract Along with uncertainty in quantitative precipitation forecast (QPF), uncertainties in quantitative precipitation estimate (QPE) and in soil moisture states represent arguably the biggest sources of error in real-time hydrological prediction. To deal with these uncertainties, some form of state updating is employed in practice, that assimilates real-time streamflow observations. The purpose of this work is to explore real-time assimilation of radar-based QPE and streamflow observations into a distributed hydrological model via variational assimilation (VAR). The particular VAR formulation examined in this work deals with the inverse problem of estimating grid cell-specific initial model soil moisture states and spatially-uniform (but time-varying) multiplicative adjustment factors to QPE and potential evaporation (PE) based on real-time observations of streamflow, radar-based QPE and climatological estimates of PE.

Key words data assimilation; distributed hydrological modelling; radar rainfall; real-time forecasting; streamflow

INTRODUCTION

To deal with various sources of error in the initial and boundary conditions, and in model parameters and structure, some form of state updating is necessary in operational forecasting that makes use of real-time streamflow observations. Compared to state updating of lumped models that input raingauge-based precipitation estimates, that of distributed models, driven in particular by radar-based precipitation estimates, presents at least two additional difficulties: (a) the dimensionality of the estimation problem is very large; and (b) the error structure in radar-based precipitation estimates is not very simple. In theory, state updating of distributed models may be carried out using, e.g. variational assimilation (VAR) (Jazwinski, 1970), distributed-parameter Kalman filter (Tzafestas, 1978) and (distributed-parameter) ensemble Kalman filter (Evensen, 1994) techniques. In the least squares sense, they all attempt to essentially solve the same minimization problem and, if the model dynamics is linear, are essentially equivalent (Li & Navon, 2001). As such, we are not so much interested here in the particular choice of the assimilation technique as in formulating the observation models and prescribing the error statistics. These issues are common to both VAR and Kalman filtering, and are of fundamental interest in state updating of distributed models.

DESCRIPTION OF THE ASSIMILATION PROCEDURE

We cast state updating of distributed models as the following static estimation problem:

$$Z = H(X) + V \tag{1}$$

where $\boldsymbol{Z}$, $\boldsymbol{X}$, and $\boldsymbol{V}$ denote the observation, state (or control) and error vectors, $\boldsymbol{Z} = [\boldsymbol{Z}_Q, \boldsymbol{Z}_P, \boldsymbol{Z}_E, \boldsymbol{Z}_B]$, $\boldsymbol{X} = [X_0, X_P, X_E]$ and $\boldsymbol{V} = [V_Q, V_P, V_E]$, respectively, and $H(\ )$ denotes the structure function that relates the states to the observations via the model dynamics. In the observation vector $\boldsymbol{Z}$, $\boldsymbol{Z}_Q$ denotes the vector of hourly streamflow observations in the most recent J hours (referred to herein as the assimilation window): $\boldsymbol{Z}_Q = [z_{Q,k-J}, z_{Q,k-J+1}, \ldots, z_{Q,k}]$, where $z_{Q,k}$ denotes the streamflow observation at hour k; $\boldsymbol{Z}_P$ and $\boldsymbol{Z}_E$ denote the vectors of precipitation and potential evaporation (PE) estimates in the most recent I hours (referred to herein as the control horizon): $\boldsymbol{Z}_P = [z_{P,u,k-I}, z_{P,u,k-I+1}, \ldots, z_{P,u,k}; u = 1,\ldots, n]$ and $\boldsymbol{Z}_E = [z_{E,u,k-I}, z_{E,u,k-I+1}, \ldots, z_{E,u,k}; u = 1,\ldots, n]$, respectively, where $z_{P,u,k}$ and $z_{E,u,k}$ denote the precipitation and PE, respectively, estimates at location u (i.e. the grid cell centred at u) at hour k, and n denotes the number of grid cells in the basin, and $\boldsymbol{Z}_B$ denotes the vector of *a priori* estimates of the model states, $\boldsymbol{Z}_B = [\boldsymbol{Z}_{B,s1}, \boldsymbol{Z}_{B,s2}, \ldots, \boldsymbol{Z}_{B,s6}, \boldsymbol{Z}_{B,r1}, \boldsymbol{Z}_{B,r2}, \boldsymbol{Z}_{B,r3}]$, where the subscripts s and r signify that the variables are associated with soil moisture accounting and routing models, respectively. For example, $\boldsymbol{Z}_{B,s1}$ in $\boldsymbol{Z}_B$ denotes the vector of the first soil moisture state variable, $\boldsymbol{Z}_{B,s1} = [z_{B,s1,u,k-I}; u = 1,\ldots, n]$. The size of the grid cell used in the distributed model is approximately 4×4 km^2. Soil moisture accounting is performed via the Sacramento model (SAC-SMA; Burnash *et al.*, 1973), as applied to each grid cell without lateral exchange. The routing model is based on kinematic wave, and has hillslope and channel flow components (Koren *et al.*, 2002). The above hydrological and hydraulic models constitute the current version of the National Weather Service Hydrology Laboratory's (NWS/HL) Research Modelling System (RMS) (for details see Koren *et al.*, 2002). There are a total of eight model state variables, six in SAC-SMA and two in routing (one for hillslope and one for channel). Though channel routing has only one state variable (i.e. the wetted channel cross-section), we used three control variables per grid cell following the discretization of the channel reach within a grid cell into three segments as implemented in RMS. Hence, for a basin made of 100 grid cells, there are 1000 control variables associated with the model states. In the state vector $\boldsymbol{X}$ in equation (1), X_0 denotes the initial conditions of the model states at the beginning of the control horizon, X_P and X_E denote the hourly multiplicative adjustment factors to the precipitation and PE estimates, respectively, within the control horizon. In the error vector $\boldsymbol{V}$ in equation (1), V_Q denotes the zero-mean error vector associated with observation and model simulation of hourly streamflow, and V_P and V_E denote the zero-mean error vectors associated with estimation of grid cell-specific hourly radar precipitation and PE, respectively.

From equation (1), we have the following least squares minimization problem subject to the model dynamics:

$$\text{Minimize } J = 1/2[\boldsymbol{Z} - H(\boldsymbol{X})]^T \boldsymbol{R}^{-1}[\boldsymbol{Z} - H(\boldsymbol{X})] \tag{2}$$

Noting independence among the zero-mean error vectors, we may rewrite equation (2) as:

$$\text{Minimize}\quad \begin{aligned} J = & 1/2[\boldsymbol{Z}_Q - G(\boldsymbol{X}_0, \boldsymbol{X}_P, \boldsymbol{X}_E)]^T R_Q^{-1}[\boldsymbol{Z}_Q - G(\boldsymbol{X}_0, \boldsymbol{X}_P, \boldsymbol{X}_E)] \\ & + 1/2[\boldsymbol{Z}_P - H_P \boldsymbol{X}_P]^T R_P^{-1}[\boldsymbol{Z}_P - H_P \boldsymbol{X}_P] \\ & + 1/2[\boldsymbol{Z}_E - H_E \boldsymbol{X}_E]^T R_E^{-1}[\boldsymbol{Z}_E - H_E \boldsymbol{X}_E] \\ & + 1/2[\boldsymbol{Z}_B - \boldsymbol{X}_0]^T B^{-1}[\boldsymbol{Z}_B - \boldsymbol{X}_0] \end{aligned} \tag{3}$$

subject to

$$\boldsymbol{X}_{k+1} = g(\boldsymbol{X}_k, \boldsymbol{X}_{Pk}, \boldsymbol{X}_{Ek}) \tag{4}$$

$$\boldsymbol{X}_{\min} \le \boldsymbol{X}_{k+1} \le \boldsymbol{X}_{\max} \tag{5}$$

where $G(\ )$ denotes the (forward) run of the soil moisture accounting and routing models from the beginning of the control horizon to the hours concurrent to the respective streamflow observations in $\boldsymbol{Z}_Q$, $g(\ )$ denotes the model dynamics over one time step, X_k denotes the model (soil moisture accounting and routing) state variables at time step k and $\boldsymbol{R}_Q$, $\boldsymbol{R}_P$, $\boldsymbol{R}_E$, and B denote the error covariance matrices associated with streamflow, precipitation, PE and the *a priori* model states at the beginning of the control horizon, respectively. It is worth noting that the above constrained minimization is essentially the same as the Kalman smoothing problem (see e.g. Schweppe, 1973) if one considers adjustment to precipitation (and to a lesser degree PE) as correcting the model errors in the linearized dynamics of equation (4): $\boldsymbol{R}_P$ may then be interpreted as the model error covariance Q in the usual Kalman filter notation.

There are four error covariance matrices to be specified in equation (3); $\boldsymbol{R}_Q$, $\boldsymbol{R}_P$, $\boldsymbol{R}_E$, and $\boldsymbol{B}$, on which performance of the assimilation procedure depends critically. Here, as a first step in the continuing investigation, we made a simplifying assumption that all the error covariance matrices are diagonal. For precipitation, it is probably too gross an assumption in that uniform multiplicative adjustment of precipitation estimates does not change the spatial correlation structure of the estimation error. Another approximation of significance made here is that because the true states were not known, the error variances (i.e. the diagonal entries) could only be specified very loosely. As a first guess, we used the sample variances of the estimates and model simulations of the states in place of the error variances.

The minimization problem in equations (3)–(5) is solved by the conjugate gradient method of Fletcher-Reeves-Polak-Ribiere (Press *et al.*, 1986), which worked well for state updating of lumped models (Seo *et al.*, 2002). In the minimization, the gradients were evaluated via the adjoint code of the forward model (i.e. the SAC-SMA and routing), generated automatically by the Tangent linear and Adjoint Model Compiler (TAMC; Giering, 1999). Significant modifications to the TAMC-generated code were necessary to improve computational efficiency. To test the procedure, we performed assimilation retrospectively (assuming clairvoyant QPF) for the basin near Watts, Oklahoma, USA. The basin, modelled extensively in the Distributed Model Intercomparison Project (http://www.nws.noaa.gov/oh/hrl/dmip), has an area of 1645 km^2 and a time-to-peak of approximately 20 h. Figure 1 shows an example of the assimilation result, where the vertical line indicates the time of the most recent streamflow observation at the basin outlet. Figure 2 is the same as Fig. 1, but 1 h later (i.e. an additional streamflow observation was available). The beneficial effect of state updating to outlet streamflow simulation, particularly on the rising limb of the hydrograph, is apparent.

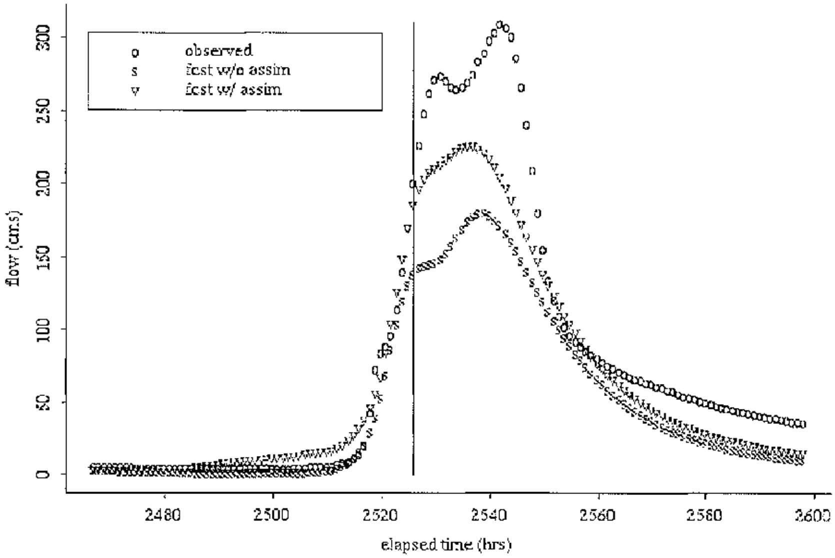

Fig. 1 An example of VAR-aided simulation as compared with simulation without state updating and the observed flow, where the vertical line indicates the time of the most recent streamflow observation at the basin outlet.

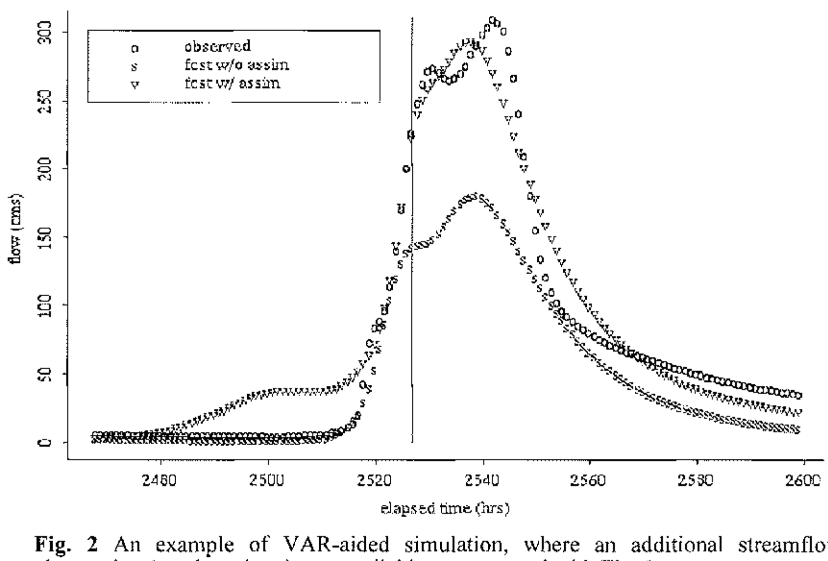

Fig. 2 An example of VAR-aided simulation, where an additional streamflow observation (one hour later) was available as compared with Fig. 1.

ISSUES

A number of questions arise from the above assimilation problem that are pertinent to state updating of distributed models in general. Here we identify several of them:

(a) How does the length of the control horizon affect the assimilation results? Presumably, the longer the control horizon is, the less sensitive the assimilation result is to the initial conditions at the beginning of the control horizon (Chao & Chang, 1992). Then, one may be able to specify the background error covariance matrix $\boldsymbol{B}$ only very approximately if a longer control horizon can be afforded computationally (e.g. always begin assimilation in an inter-storm period).

(b) How does the size of the assimilation window affect the results? Of further interest is the presumed dependence of sizing of the window to the model states: the size may depend on which part of the hydrograph the streamflow observations come from.

(c) How accurately should the error covariance in precipitation estimates be modelled? It is possible to model in $\boldsymbol{R}_P$ the space–time correlation structure of error in precipitation estimates. It is not clear, however, whether the marginal improvement may justify the far greater computational expense.

(d) Is bin-by-bin adjustment, as opposed to multiplicative adjustment over some large area, of model states really necessary? Given that soil moisture observations are not routinely available and that it is too difficult to relate model soil moisture states to (point) observations, one may consider adjusting the model states over some large area based on some predetermined scaling relationships. Such simplification will not only reduce the computational burden but may also improve the robustness of the procedure.

(e) How beneficial is the state updating based on streamflow observations at the basin outlet to streamflow prediction at interior locations? For state updating of distributed models to be operational viable, updating-aided streamflow prediction must consistently be at least as good as the prediction, without updating at all interior locations of the basin. We note here that work is under way to investigate these and other issues, and the results will be reported in the near future.

Acknowledgements This work is supported by the Advanced Hydrological Prediction Service (AHPS) of the National Weather Service.

REFERENCES

Burnash, R. J. C., Ferral, R. L. & McGuire, R. A. (1973) A generalized streamflow simulation system—conceptual modelling for digital computers. National Weather Service, NOAA, and the State of California Department of Water Resources, Technical Report Joint Federal-State River Forecast Center, Sacramento, California, USA.

Chao, W. C. & Chang, L.-P. (1992) Development of a four-dimensional variational assimilation analysis system using the adjoint method at GLA Part 1: Dynamics. *Mon. Weath. Rev.* **120**, 1661–1673.

Evensen, G. (1994) Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *J. Geophys. Res.* **99**(C5), 10143–10162.

Giering, R. (1999) *Tangent Linear and Adjoint Model Compiler User's Manual*, TAMC Version 5.2 (available from http://www.autodiff.com/tamc).

Jazwinski, A. H. (1970) *Stochastic Processes and Filtering Theory*. Academic Press, San Diego, California, USA.

Koren, V., Smith, M., Reed, S., Zhang, Z. & Seo, D. J. (2002) Hydrology Laboratory Research Modelling System of the National Weather Service. *J. Hydrol.* (submitted).

Li, Z. & Navon, I. M. (2001) Optimality of variational data assimilation and its relationship with the Kalman filter and smoother. *Quart. J. Roy. Met. Soc.* **127**, 661–683.

Press, W. H., Flannerty, B. P., Teukolsky, S. A. & Vetterling, W. T. (1986) *Numerical Recipes*. Cambridge University Press, Cambridge, UK.

Schweppe, F. C. (1973) *Uncertain Dynamic Systems*. Prentice-Hall, Englewood Cliffs, New Jersey, USA.

Seo, D.-J., Koren., V. & Cajina, N. (2003) Real-time variational assimilation of hydrological and hydrometeorological data into operational hydrological forecasting. *J. Hydromet.* (submitted).

Tzafestas, S. G. (1978) Distributed parameter in state estimation. In: *Distributed Parameter Systems* (ed. by W. H. Ray & D. G. Lainiotis), Chapter 3. Dekker, New York, USA.