BULETINUL INSTITUTULUI POLITEHNIC DIN IAȘI
Publicat de
Universitatea Tehnică „Gheorghe Asachi” din Iași
Tomul LX (LXIV), Fasc. 2, 2014
Secția
CONSTRUCȚII. ARHITECTURĂ

DATA ASSIMILATION APPLICATIONS IN HYDROLOGY

BY

CIPRIAN DELIU[1,*] and ION GIURMA[2]

“Gheorghe Asachi” Technical University of Iași
[1]Department of Mathematics and Informatics
[2] Department of Hydraulic Structures Engineering

Received: June 29, 2014
Accepted for publication: July 14, 2014

Abstract. In this paper we review some data assimilation applications in hydrology and we investigate the benefits of applying the ensemble Kalman filter methods on a specific hydrodynamic model of a river network using the MIKE 11 software. We explore the efects of assimilating measurements from different locations (generated by a reference model) into a model forced with erroneous boundary conditions and also examine the effect of using coloured noise for describing uncertainty in the upstream boundary conditions.

Key words: hydrological systems; Kalman filter; data assimilation.

1. Introduction

Data assimilation techniques have become more popular over the last decade in modelling and forecasting large systems due to the ever increasing computational resources. By combining any available measurements of the state of the system with the model dynamics, data assimilation provides a more robust model and improves the knowledge of the system using the Kalman filter framework which adds to the deterministic model a stochastic part both in the model dynamics and in the measurements (Evensen, 2009). In this paper we review some data assimilation applications in hydrology and we investigate the

[*]Corresponding author: *e-mail*: cipriandeliu@gmail.com

benefits of applying the ensemble Kalman filter methods on a specific hydrodynamic model of a river network using the MIKE 11 software.

The ensemble Kalman filter implemented in the data assimilation module of MIKE 11 is based on a stochastic model formulation in which model uncertainties are quantified and propagated through the model, which also allows for general uncertainty assessment, *i.e.* without assimilation of measurements (Drecourt *et al.*, 2006; Madsen & Skotner, 2005; Sorensen & Madsen, 2004). In our hydrodynamic model we use assimilation of two types of measurements (water levels and discharges) and we assume model errors in the infow discharge boundaries. We explore the effects of assimilating measurements from different locations (generated by a reference model) into a model forced with erroneous boundary conditions and also examine the effect of using coloured noise for describing uncertainty in the upstream boundary conditions.

2. Data Assimilation Framework

We consider a model that can be described in a deterministic discrete-time dynamic system setting as

$$x_{k+1} = \Phi\left(x_k, u_k\right), \tag{1}$$

where: x_k is the state variables of the system at time step k, u_k – the forcing terms of the system (including all boundary conditions) and $\Phi(...)$ – the model operator.

We also consider the measurements vector

$$z_k = C_k x_k , \tag{2}$$

where C_k is a matrix that describes the relation between measurements and state variables (*i.e.* mapping of the state space to measurement space).

The data assimilation process is a succession of two steps:

1º The model is employed to issue a forecast

$$x_k^f = \Phi\left(x_{k-1}^a, u_k\right). \tag{3}$$

2º The observed data are meld with the forecast to provide an updated state (analysis step), which is chosen to be a linear combination of the data and the model

$$x_k^a = x_k^f + G_k\left(z_k - C_k x_k^f\right), \tag{4}$$

where: x_k^f is the forecast state vector, x_k^a – the analysed or updated state vector, G_k – weighting matrix (gain matrix) and $z_k - C_k x_k^f$ – the innovation vector, which includes the differences between the measurements and their model forecasted equivalents.

The Kalman filter is based on a stochastic formulation of the model eq. (1) and measurement (eq. (2)):

$$x_{k+1} = \Phi\left(x_k, u_k + \varepsilon_k\right), \tag{5}$$

$$z_k = C_k x_k + \eta_k , \tag{6}$$

where: ε_k is the model error of the forcing terms and η_k – the random measurement error vector with zero mean and covariance matrix R_k.

The uncertainty of the model forecast is described by the covariance matrix P_k^f . The Kalman gain to be used in the update eq. is given by

$$G_k = P_k^f C_k^T \left(C_k P_k^f C_k^T + R_k\right)^{-1}. \tag{7}$$

The model covariance expresses how the innovation vector should be distributed on the state vector, and the relation between the model and measurement covariances expresses how much weight should be put on the measurement. For non-linear and high-dimensional systems the evaluation of the covariance matrix, P_k^f, requires huge computational costs and storage requirements, end hence makes the filtering infeasible for real-time applications. In the EnKF, the covariance matrix is represented by an ensemble of possible state vectors, which are propagated according to the dynamical system subjected to model errors, and the resulting ensemble then provides estimates of the forecast state vector and covariance matrix.

The EnKF algorithm asumes that measurement errors are uncorrelated *i.e.* the covariance matrix, R_k, is a diagonal matrix, $R_k = \mathrm{diag}\left[\sigma_1^2, \ldots, \sigma_p^2\right]$. For a given set of state vectors $x_{jk}^f = 1, \ldots, M$ we have the following steps:

1. Each member of the ensemble is propagated forward in time according to the stochastic dynamical system, the model error being drawn from a Gaussian distribution with zero mean and known covariance

2. The forecast of the state vector is then estimated as the mean of this ensemble:

$$\bar{x}_k^f = \frac{1}{M} \sum_{j=1}^{M} x_{jk}^f . \tag{8}$$

3. An estimate of the covariance matrix is found

$$P_k^f = \frac{1}{M-1} S_k^f \left(S_k^f\right)^T, \quad \text{where} \qquad S_{j,k}^f = x_{j,k}^f - \overline{x}_k^f . \tag{9}$$

4. Each of the ensemble state vectors are updated sequentially. Since the measurements are independent, the sequential update procedure processes one measurement at a time.

5. The updated state vector and covariance matrix are estimated as

$$\overline{x}_k^a = \frac{1}{M}\sum_{j=1}^{M} x_{j,k}^a, \; P_k^a = \frac{1}{M-1} S_k^a \left(S_k^a\right)^T . \tag{10}$$

In the implementation of the EnKF the covariance matrix P_k is never calculated, all calculations are based on S_k. When no data are available for updating, only steps 1...3 are carried out.

3. MIKE 11 Software

MIKE 11 is a 1-D dynamic modelling system for rivers, channels and reservoirs, used for simulating flow and water level, water quality and sediment transport in rivers, floodplains, irrigation canals, reservoirs and other inland water bodies, so it is applicable for simulating rivers and other open surface water bodies which can be approximated as 1-dimensional flow. It has modules for several types of problems, from flood forecasting and dam breaches to water quality and integrated modelling (groundwater and surface runoff).

In the hydrodynamic simulation module, the following assumptions are made:

a) incompressible and homogeneous fluid;

b) flow is one-dimensional (uniform velocity and water level in cross-section);

c) bottom slope is small;

d) small longitudinal variation in geometry;

e) hydrostatic pressure distribution;

f) conservation of mass

$$\frac{\partial Q}{\partial x} - b\frac{\partial h}{\partial t} = 0 \tag{11}$$

g) conservation of momentum

$$\frac{\partial Q}{\partial t}+\frac{\partial\left(\alpha Q^2/A\right)}{\partial x}+gA\frac{\partial h}{\partial x}+\frac{gn^2Q|Q|}{AR^{4/3}}=0\,. \tag{12}$$

For more details see (Giurma *et al.*, 2009a; Giurma *et al.*, 2009a; DHI, 2011).

There are two data assimilation methods implemented in MIKE 11: Ensemble Kalman Filter and uncertainty prediction, and different types of measurements that can be assimilated: water levels and discharges (in hydrodynamic modelling), or concentration, temperature, salinity (in advection–dispersion modelling). The data assimilation module assumes errors in model forcing (e.g. inflow discharge boundaries, rainfall-runoff boundaries), which are represented as coloured noise using a first order autoregressive process

$$\varepsilon_{k+1}=\phi\varepsilon_k+\eta_k\,, \tag{13}$$

where the autoregressive coefficient is defined in terms of time constant

$$\phi=\exp\left(-\frac{\Delta t\ln 2}{TC}\right). \tag{14}$$

By using an auto-regressive description the model error has some memory that is propagated as part of the model forecast. The model state is augmented with the model errors that are updated as part of the Kalman Filter update scheme.

4. River Model Application

We consider a river model forced by discharge inflow at the upstream boundaries. First we run a reference (true) model with specified boundary conditions, and then, by changing these conditions we obtain a background (false) model with erroneous boundary conditions, on which we will run the data assimilation procedure using as measurements the water levels from reference model at selected locations. We use three different setups:

a) assimilation of water levels from 1 measurement point using white noise model error;

b) assimilation of water levels from 1 measurement point using coloured noise model error;

c) assimilation of water levels from 3 measurement point using coloured noise model error.

The river network consists of a main river named *Mainstrezm* and two tributaries named *Trib1* and *Trib2*, with specified cross-sections, boundary data and hydrodynamic parameters (Fig.1).

The simulation will provide results from all computational grid points (discharges and water levels), and we select 5 locations in which the results are analysed.

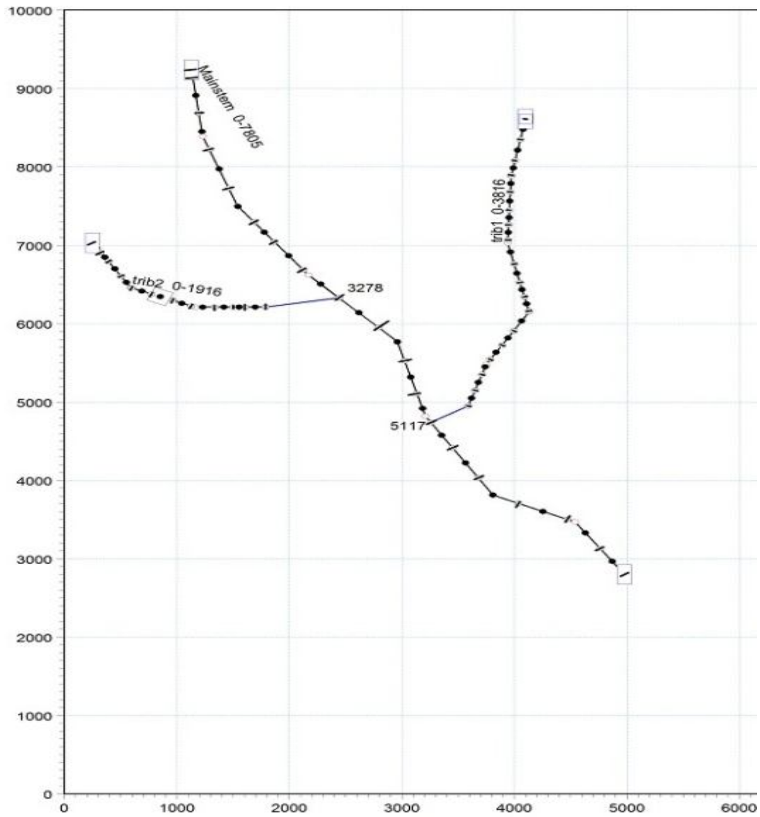

Fig. 1 –River network.

The above true model defines the reference model to be used in the data assimilation run, and the false model has an identical setup except for the upstream discharge boundary conditions.

Water levels from selected points in the river system are extracted from the reference model and used for assimilation in the model forced with erroneous boundary conditions. Three water level locations have been defined; one location at the upstream part of the main stream, one at the downstream part of tributary Trib1, and one at the downstream part of tributary Trib2.

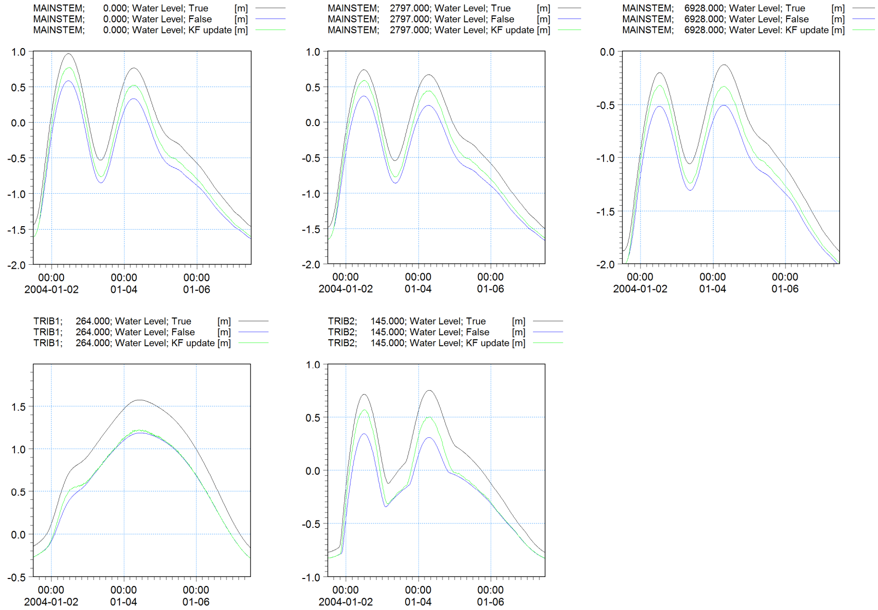

Fig. 2 – Assimilation of water levels from 1 measurement point using white noise model error.

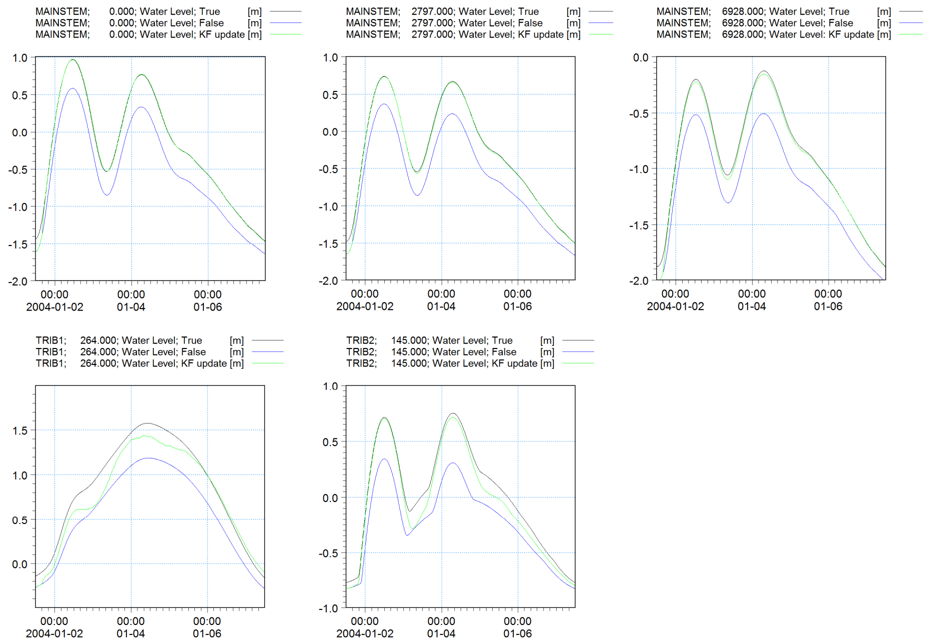

Fig. 3 – Assimilation of water levels from 1 measurement point using coloured noise model error.

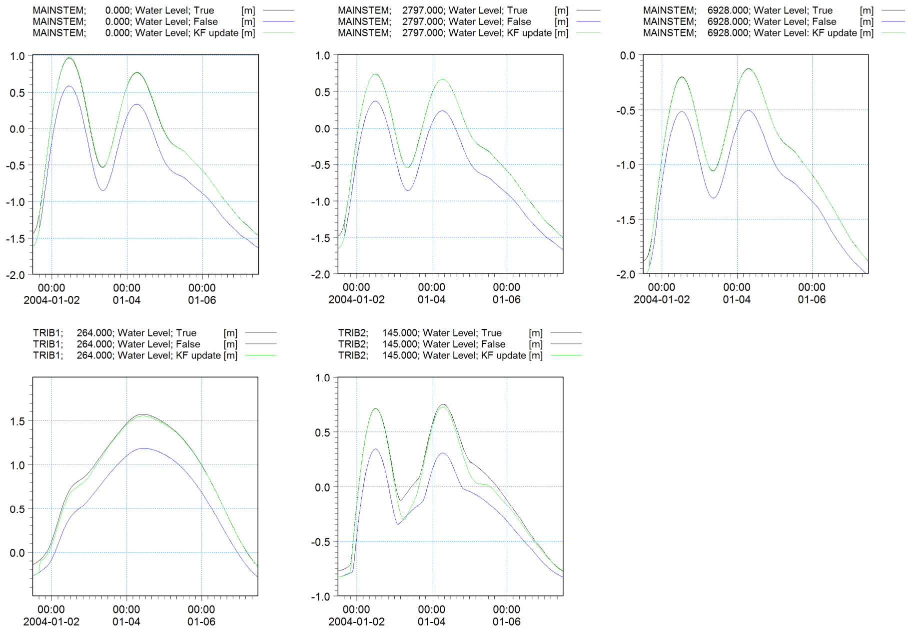

Fig. 4 – Assimilation of water levels from 3 measurement point using coloured noise model error.

In the Data Assimilation module we choose the measurement update mode and the ensemble Kalman filter method. Model uncertainties are defined in the three upstream boundary conditions using in our case a standard deviation of 10% for each discharge boundary. Colored noise can be defined using a first order autoregressive process with a specified time constant.

For the three different setups considered we find the following results in the 5 points we chose:

a) assimilation of water levels from 1 measurement point using white noise model error (Fig. 2);

b) assimilation of water levels from 1 measurement point using colored noise model error (Fig. 3);

c) assimilation of water levels from 3 measurement point using colored noise model error (Fig. 4).

5. Conclusions

Assimilating measurements using the Ensemble Kalman Filter method improves the simulation results in case of erroneous boundary conditions and using coloured noise for describing uncertainty in the upstream boundary conditions provides better results.

The amount of measurements assimilated also improves the quality of the results, and also the location of the assimilated measurements can have an impact on the simulation results, depending on the structure of the river network.

Finally, data assimilation can be used as part of a forecasting system by assimilating the measurements up to the time of forecast, and then propagating the ensemble (without assimilation) in the forecast period.

REFERENCES

Drecourt J.P., Madsen H., Rosbjerg D., *Bias Aware Kalman filters: Comparison and Improvements.* Adv. in Water Res., **29**, 707-718 (2006).

Evensen G., *Data Assimilation - The Ensemble Kalman Filter.* Springer-Verlag, 2009.

Giurma I., Crăciun I., Giurma C.R., *Hidrologie şi hidrogeologie.* Edit. Politehnium, Iaşi, 2009b.

Giurma I., Crăciun I., Giurma C.R., *Hidrologie.* Edit. Politehnium, Iaşi, 2009a.

Madsen H., Skotner C., *Adaptive State Updating in Real-Time River Flow Forecasting – a Combined Filtering and Error Forecasting Procedure.* J. of Hydrol., *308*, 302-312 (2005).

Sorensen J.V.T., Madsen H., *Efficient Kalman Filter Techniques for the Assimilation of Tide Gauge Data in Three-Dimensional Modeling of the North Sea and Baltic Sea System.* J. of Geophys. Res., **109** (2004).

* * * *MIKE 11: A Modeling System for Rivers and Channels, Reference Manual.* Danish Hydr. Inst. (DHI), Denmark, reference Manual, 2014.

APLICAȚII ALE ASIMILĂRII DATELOR ÎN HIDROLOGIE

(Rezumat)

Tehnicile de asimilare a datelor au devenit populare pe parcursul ultimei decade în modelarea şi prognozarea sistemelor de mari dimensiuni datorită dezvoltării capacităților computaționale. Prin combinarea măsurătorilor disponibile ale stării sistemului cu dinamica modelului, asimilarea datelor furnizează modele mai robuste utilizând teoria filtrelor Kalman care adaugă modelului determinist componente stocastice atât la dinamica modelului cât şi la măsurători.

Se trece în revistă câteva aplicații ale asimilării datelor în hidrologie şi se investighează beneficiile aplicării metodelor furnizate de EnKF (Ensemble Kalman Filter) pe modelul hidrodinamic al bazinului unui râu folosind programul MIKE 11. În modelul hidrodinamic studiat se folosesc două tipuri de măsurători (niveluri şi debite) şi se presupune că modelul are erori la debitele de intrare. Sunt explorate efectele asimilării măsurătorilor din diferite locații (furnizate de un model de referință) atunci când condițiile la limită sunt eronate şi de asemenea este examinat efectul folosirii unui zgomot colorat pentru incertitudinile din condițiile la limită.