8-1-2014

Assimilation Of Heterogeneous Uncertain Data, Having Different Observational Errors, In Hydrological Models

Maurizio Mazzoleni

Leonardo Alfonso

Dimitri P. Solomatine

ASSIMILATION OF HETEROGENEOUS UNCERTAIN DATA, HAVING DIFFERENT OBSERVATIONAL ERRORS, IN HYDROLOGICAL MODELS

MAURIZIO MAZZOLENI (1), LEONARDO ALFONSO (1), DIMITRI SOLOMATINE (1)(2)
(1): Integrated Water Systems and Governance, UNESCO-IHE Institute for Water Education, Westvest 7, Delft, The Netherlands
(2): Water Resources Section, Delft University of Technology, Delft, The Netherlands

Accurate real-time forecasting of river water level is an important issue that has to be addressed in order to prevent and mitigate water-related risk. To this end, data assimilation methods have been used to improve the forecasts ability of water model merging observations coming from stations and model simulations. As a consequence of the increasing availability of dynamic and cheap sensors, having variable life-span, space and temporal coverage, the citizens are becoming an active part in information capturing, evaluation and communication. On the other hand, it is difficult to assess the uncertain related to the observation coming from such sensors.
The main objective of this work is to evaluate the influence of the observational error in the proposed assimilation methodologies used to update the hydrological model as response of distributed observations of water discharge. We tested the developed approaches on a test study area - the Brue catchment, located in the South West of England, UK. The Ensemble Kalman filter is applied to the semi-distributed hydrological model. Distributed observations of discharge are synthetically generated. Different types of observational error are introduced assuming diverse sets of probability distributions, first and second order moments. The results of this work show how the assimilation of distributed observations, can improve the hydrologic model performance with a better forecast of flood events. It is found that different observational error types can affects the model accuracy.

INTRODUCTION

A common problem in data assimilation methods is the correct evaluation of the uncertainty related to the accuracy of the real-time measurements. In case of streamflow observations, errors can be related to an inappropriate water level (WL) measurement or to the wrong assessment of the rating curve used to transform the value of WL into discharge. Di Baldassarre and Montanari [2] gave a detailed description of errors in stage and velocity gaugings, assumption of a particular form of rating curve, and errors related to any cross-section changes. Usually, in data assimilation procedure, errors in hydrological variables are assumed normally distributed. Clark et al. [1] proposed an improved version of the EnKF by transforming observed and modelled streamflow to log space before computing the Kalman gain. The main goal of this work is to analyse the effects of synthetic distributed observations of streamflow having different types of observational errors on the assimilation process, in case of using a semi-distributed hydrological model. The Ensemble Kalman Filter (EnKF) (Evensen [3]) is implemented in order to integrate the synthetic observations with the hydrologic model. The uncertainty coming from the sensors measurements is divided in two different sources of error (as described by Clark et al. [1]), (A) errors in the rating curve estimate, and, (B) errors in the proper measurement of WL. Two different probability distributions, normal and uniform, are considered in this analysis. The results of this work show how the appropriate definition of the observational error affects the model accuracy. Nonetheless, overall the assimilation of distributed streamflow observations can improve the hydrologic model performance with a better flood prediction.

STUDY AREA AND DATA SET

The methodology proposed in this study is applied to the Brue basin, UK. The rainfall, mean annual value about 867 mm, and discharge, average value of 1.92 m^3/s, data value are measured at 49 automatic rain stations and at the basin outlet by one station, respectively, at a 15min time step resolution. In order to estimate the average precipitation value in each sub-basin, an Ordinary Kriging is used to optimally interpolate the point observations of precipitation from the rainfall station (Matheron [4]). A flood event occurred from the 16/12/1995 to 01/01/1996 is used to calibrated the hydrological model, while, for the validation analysis, a flood event occurred between 08/11/1994 to 16/11/1994 is considered.

SEMI-DISTRIBUTED HYDROLOGICAL MODELLING

A semi-distributed hydrological model is used to take into account the spatial variability of the uncertain observations of discharge and estimate the flow hydrograph at the outlet section of the basin. For this reason, the Brue basin is divided into different sub-basins having a small drainage area (in average around 2 km^2). In order to estimate the discharge hydrograph at the outlet point of each sub-basin a conceptual lumped model is implemented. The Soil Conservation Service Curve Number (SCS-CN) method is used to assess the direct runoff, i.e. the input of each conceptual model. Then, a lumped conceptual model based on the Continuous Kalinin-Milyukov-Nash (KMN) equation is implemented to estimate the outflow discharge as convolution of the input I with the impulse-response function. Szilagyi and Szollosi-Nagi [6] adapted the KMN equation in order to express it as a discrete state-pace system. Finally, the output each upstream sub-basins is propagated, using the Muskingum channel routing method, at the downstream section and sum up with the discharge estimated for the downstream sub-basin.

METHODOLOGY

Data assimilation method

Synthetic distributed observations of discharge are used to update the states of the semi-distributed hydrological model by means of the EnKF (Evensen [3]). In the EnKF the efficiency of the filter is closely dependent by the ensemble size (Pauwels and De Lannoy [5]). For this reason, the ensemble is estimated perturbing the forcing data and the parameters using a uniform distribution with standard distribution function of two parameters ε_I and ε_p. The values of ε_p and model realization n_{ens} are estimated using the approach proposed by Pauwels and De Lannoy [5]. As a result of this analysis a value of NRR (Pauwels and De Lannoy [5]), which corresponds to n_{ens} and ε_p equal 65 and 0.5, about 0.89 is obtained considering the calibration flood event.

Synthetic distributed observations of discharge are used in this study in order to update the model states and consequent outflow hydrograph. These synthetic observations are assumed normally or uniformly distributed, around $\mathbf{Q}_{true}$, with 0 mean and standard deviation σ_Q. The value of $\mathbf{Q}_{true}$ is estimated using the approach described in Weerts and El Serafy [7] for each given sub-basin k. Then, the observational error σ_Q, at the sub-basin k, is estimated as function of a coefficient α:

$$\sigma_Q^k(t) = \alpha \cdot \mu_Q^{k\,2}(t) \tag{1}$$

Experimental setup:

In order to represent the dynamic behaviour of the sensors, 10 different spatial patterns of sensors location are considered. In particular, the observation are assumed to come in pattern 1 from all the sub-basins, in pattern 2 from the outlet of the basin, pattern 3, 4, 5, and 6 from specific blocks in the basin, pattern 7 from sub-basin of order 1 (Horton classification), pattern 8 from the main river reach while pattern 9 and 10 from specific location along the main river reach. Different error scenarios of observation uncertainty are generated assuming diverse probability distributions and values of coefficient α:

A. In this first scenario the error is coming from an inadequate estimation of the rating curve used to transform WL in discharge assuming α equal to 0.1 (Weerts and El Serafy [7]).
B. The observational error, in the second scenario, is independent from the rating curve estimation and it depends only to the error in the measurement of WL. In this case, two different probability distributions, normal (B1) and uniform (B2), are used to characterize the vector of observations using a α_B equal to 0.05.
C. In the last scenario both the error in the rating curve and measurement are considered. The value of the coefficient α_C is assumed as the sum of the two previous coefficients.

RESULTS

In the following, model results are summarized according to the three methodological steps described before. From Table 1 it can be observed how only few spatial patterns of sensors location provide a significant model improvement in both the type of observational errors.

Table 1. Model performances, expressed in terms of NSI, in case of different spatial patterns and observational errors.

	Spatial patterns									
Error	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
A	0.82	0.32	0.53	0.46	0.51	0.66	0.39	0.80	0.82	0.64
B1	0.87	0.65	0.64	0.69	0.61	0.79	0.53	0.89	0.89	0.88
B2	0.87	0.63	0.63	0.68	0.62	0.77	0.55	0.88	0.88	0.88
C	0.75	0.33	0.48	0.45	0.36	0.55	0.33	0.69	0.68	0.55

Model performances are strictly related to the type of observational error. From Table 1, it can be seen how the best model improvement is achieved assuming the error coming only from the measurement process, i.e. error B with normal or uniform distribution of the error in case of α equal to 0.05. Table 1 showed how the difference between the hydrograph obtained with error A and error B1 and B2 is significant and comparable.

CONCLUSIONS

The main goal of this study is to evaluate the influence of the observational error in the assimilation of distributed uncertainty observations of discharge to update the states of a semi-distributed hydrological model. Different spatial patterns of sensors location are implemented. An EnKF is used to assimilate the synthetic real-time observations of discharge coming from distributed locations within the basin. It is shown how the assimilation of uncertain distributed observations of streamflow can improve the flood prediction. In particular, it is demonstrated that uniform and normal distributions of the observational error lead to similar results in terms of NSI. To conclude, different observational error types can affect in different way the model accuracy and the consequent flood prediction.

ACKNOLEDGEMENTS

This research is funded in the framework of the European FP7 Project WeSenseIt: Citizen Observatory of Water, grant agreement no 308429. Data used are supplied by the British Atmospheric Data Centre from the NERC Hydrological Radar Experiment Dataset http://www.badc.rl.ac.uk/data/hyrex/.

REFERENCES

[1] Clark M.P., Rupp D.E., Woods R.A., Zheng X., Ibbit R.P., Slater A.G., Schmidt J. and Uddstrom M.L., “Hydrological data assimilation with the ensemble Kalman filter: Use of streamflow observation to update states in a distributed hydrological model”, *Advances in Water Resources*, Vol. 31, (2008), pp 1309-1324.

[2] Di Baldassarre G. and Montanari A., “Uncertainty in river discharge observations: a quantitative analysis”, Hydrol. Earth Syst. Sci., Vol. 13, (2009), pp 913-921.

[3] Evensen G., “The Ensemble Kalman Filter: Theoretical Formulation and Practical Implementation”, *Ocean Dynamics*, Vol.53, (2003), pp 343-367.

[4] Matheron G., “*Principles of Geostatistics*”, Economic Geology, Vol. 58, (1963), pp 1246-1266.

[5] Pauwels V.R.N. and De Lannoy G.J.M., “Ensemble-based assimilation of discharge into rainfall-runoff models: A comparison of approaches to mapping observational information to state space”, *Water Resour. Res.*, Vol. 45, (2009), pp 1-17.

[6] Szilagyi J. and Szollozi-Nagi A., “*Recurvive Streamflow Forecasing: A State Space Approach*”, CRC Press, London, UK, (2010).

[7] Weerts, A.H. and El Serafy G.Y.E., “Particle filtering and ensemble Kalman filtering for state updating with hydrological conceptual rainfall-runoff models”, *Water Resources Research*, Vol. 2, (2006), doi:10.1029/2005WR004093.