Advances in Water Resources 83 (2015) 323–339

Contents lists available at ScienceDirect

Advances in Water Resources

journal homepage: www.elsevier.com/locate/advwatres

Assimilating uncertain, dynamic and intermittent streamflow observations in hydrological models

Maurizio Mazzoleni [a,*], Leonardo Alfonso [a], Juan Chacon-Hurtado [a], Dimitri Solomatine [a,b]

[a] *UNESCO-IHE Institute for Water Education, Hydroinformatics Chair, Westvest 7, 2611 AX Delft, The Netherlands*
[b] *Water Resources Section, Delft University of Technology, The Netherlands*

ARTICLE INFO

Article history:
Received 2 July 2014
Revised 6 July 2015
Accepted 7 July 2015
Available online 11 July 2015

Keywords:
Data assimilation
Uncertain observation
Dynamic observations
Intermittent observations
Hydrological modeling

ABSTRACT

Catastrophic floods cause significant socio-economical losses. Non-structural measures, such as real-time flood forecasting, can potentially reduce flood risk. To this end, data assimilation methods have been used to improve flood forecasts by integrating static ground observations, and in some cases also remote sensing observations, within water models. Current hydrologic and hydraulic research works consider assimilation of observations coming from traditional, static sensors. At the same time, low-cost, mobile sensors and mobile communication devices are becoming also increasingly available. The main goal and innovation of this study is to demonstrate the usefulness of assimilating uncertain streamflow observations that are dynamic in space and intermittent in time in the context of two different semi-distributed hydrological model structures. The developed method is applied to the Brue basin, where the dynamic observations are imitated by the synthetic observations of discharge. The results of this study show how model structures and sensors locations affect in different ways the assimilation of streamflow observations. In addition, it proves how assimilation of such uncertain observations from dynamic sensors can provide model improvements similar to those of streamflow observations coming from a non-optimal network of static physical sensors. This can be a potential application of recent efforts to build citizen observatories of water, which can make the citizens an active part in information capturing, evaluation and communication, helping simultaneously to improvement of model-based flood forecasting.

1. Introduction

Despite continuous construction of flood retaining structures such as levees or reservoirs, the floods impact in many countries is still an acute problem [1,2]. For this reason, the demands for flood forecasting systems, which allow decision makers to take the most effective decisions based on the forecasted water levels in rivers, have significantly increased [3].

These systems are supposed to provide forecasts timely, sufficiently accurately and preferably with the estimates of the associated uncertainty [4,5]. Such uncertainty is due to errors in observations, input, model parameters and model structure [6,7].

One technique allowing for reducing predictive uncertainty is a model updating procedure, in which model input, states, parameters or outputs are updated as a response to real time observations coming into the model [8,9].

There are two related groups of updating methods, namely error correction methods (e.g. [10,11]) and data assimilation (DA) [5,12–16]. In this paper we are dealing with the DA methods.

DA is an important and widely used technique, applied in hydrology to efficiently integrate observations into hydrological and hydrodynamic models to obtain improved model predictions and to reduce modeling uncertainty. One of the first filtering technique used to assimilate observed noisy data into models is the Kalman filter (KF) [17]. Different variants of the Kalman filter, such as the extended Kalman filter (EKF) [18], unscented Kalman filter and ensemble Kalman filter (EnKF) [19,20], have been proposed and used in hydrology improving the main limitations of the KF [21]. Recently, Liu et al. [21] provided a detailed review of the status, progresses, challenges and opportunities in advancing DA, stressing an increasing need for implementing reliable data assimilation methods in operational forecasting.

Traditionally, monitoring networks based on in-situ observations and, recently, remote sensing observations, are used to assimilate important hydrological variables as soil moisture [22–24], streamflow [25,26], latent heat flux [27] or water level from remote sensing [28], into water system models.

* Corresponding author. Tel.: +31617971191
E-mail address: m.mazzoleni@unesco-ihe.org (M. Mazzoleni).

http://dx.doi.org/10.1016/j.advwatres.2015.07.004

However, over the last couple of decades the technological improvements provided a growing availability of real-time observations coming from heterogeneous network of sensors [29]. In particular, current interest is towards the dynamic sensors that change their location at random moments and that have variable space and temporal coverage. A drawback of using measurements from dynamic sensors is related to the intrinsic variable accuracy, due to the lack of confidence in the data, and the variable life-span of each individual sensor with the consequent intermittent nature of the observations. In fact, the information coming from a specific sensor might be sent just once, occasionally, or maybe consecutively but not at equidistant time moments. *Solomatine* [30] presented a research plan towards adaptive modeling in heterogeneous time-varying (dynamic) data environments under uncertainty (AMODEU). These ideas are currently being developed in the EU-FP7 WeSenseIt project (2012–2016).

In the last years, a number of attempts to consider distributed observations in hydrological and hydrodynamic models have been made. The following most significant achievements can be mentioned. Clark et al. [31] assessed the impact of using streamflow data (with records equidistant in time) from spatially distributed gauges on flood forecasting. The authors proposed to transform streamflow into log space before computing the covariance matrix in order to improve the standard implementation of the EnKF. They also stressed the importance of considering the time lags between model states and streamflow. Xie and Zhang [32] investigated the performance of a distributed model (SWAT) by means of EnKF to assimilate different hydrological variables and update both model states and parameters. The assimilation of synthetic observations of streamflow at 10 stations was performed at every time step. The authors found out that the assimilation of runoff observations can significantly update the model states and parameters. A similar study was carried out by Chen et al. [33], who assimilated streamflow observations in a semi-distributed model using EnKF. A more detailed study related with assimilation of distributed observations was carried out by Rakovec et al. [34], who used the distributed streamflow observations to update the state of a distributed model built on the PCRaster platform [35]. Different sets of sensors locations along the main river channel and three updating filtering frequencies were used to compare the results of experiments with synthetic and real-life data. The results pointed out that assimilation of streamflow at an interior point can improve the model performance in terms of Root Mean Squared Error (RMSE). A similar study is presented by Lee et al. [36], in which different fixed spatiotemporal adjustment scales were used to update the states of a lumped, semi-distributed and distributed hydrological model using a variational assimilation method. It was found out that assimilation performances were more sensitive to the spatial distribution of sensors rather than to the updating frequency. McMillan et al. [37] applied a Recursive Ensemble Kalman filter (REnKF) in operational flood forecasting in New Zealand in order to overcome the problem of the time lag between upstream and downstream catchments in the assimilation process. They found a significant improvement in the flood prediction when using REnKF, rather than EnKF, in case of streamflow assimilation. It should be noted that the mentioned applications normally deal with stable, regular and homogeneous streams of data (e.g. rainfall, water level, discharge) to be assimilated.

In recent years there is also a growing interest in the assimilation of distributed values of water level from remote sensing in flood-forecasting systems (e.g., [28,38]). A detailed review of assimilating this type of data is presented by Schumann et al. [39].

To the best of our knowledge, however, none of the previous flood-related studies deal with the specific of the dynamic sensor networks and citizen observatories: they consider neither the variable accuracy (uncertainty) of sensors within the basin dynamic at each time step, nor the intermittent nature of such observations. However, in oceanic and meteorological modelling, assimilation of distributed intermittent observations is quite standard. Due to the irregular sampling times of oceanographic observations, most of the ocean data assimilation (ODA) systems use continuous approaches, as 3D-Var or 4D-Var methods, in order to assimilate these intermittent observations at their corresponding times. Huang et al. [40] proposed an improved continuous data assimilation scheme in which an incremental analysis update strategy is combined with a continuous ODA model. In case of observations randomly distributed in time and space, MacPherson [41] compared the repeated insertion (RI) method with an alternative intermittent analysis-forecast cycle (AF). Sinopoli et al. [42] proposed a robust Kalman filtering formulation able to model the arrival of observations as a random process. In addition, Cipra and Romera [43] developed a discrete Kalman Filter which allows the assimilation of incomplete date series. Despite the approaches previously described, in this study we decided to use a more straightforward and pragmatic method, often used in real-time early warning systems, similar to the approach proposed by Cipra and Romera [43] in order to assimilate the intermittent observations into the hydrological model.

The main innovation of this study consists of assimilating distributed uncertain, dynamic and intermittent synthetic observations of discharge to improve the results of hydrological models. The final goal is to demonstrate how hydrological models can be improved when taking into account dynamic and intermittent data coming, for example, from citizens participating in information capture along with (or instead of) using the traditional gauging stations.

At the time of finalizing this paper we did not have the real data from the sensors with the dynamic behavior, so they were simulated by adopting a number of realistic scenarios of discharge observations, with variable uncertainty in time and space.

The paper is organized as follows. First we describe the study area and the hydrological models. Then, the application of the standard Ensemble Kalman Filter used to assimilate the streamflow observations in each sub-basin is presented. Next, the set-up of the various assimilation experiments, considering various scenarios of sensors locations and intermittency, is described. Finally, the results of the assimilation experiment and the main conclusions are presented.

2. Site location and data

2.1. Study area

We consider the well-known and well-gauged Brue catchment, located in Somerset, South West of England, with predominantly rural use and modest slope [44]. The drainage area of the catchment is about 135 km^2, with time response of about 10–12 h at the basin outlet, Lovington. Hourly precipitation data are available at 49 automatic rain stations (Fig. 1); average annual rainfall of 867 mm is measured in the period between 1961 and 1990.

Discharge is measured at the basin outlet by one station at a 15 min time step resolution, having an average value of 1.92 m^3/s. For both precipitation and discharge data, a 3-years complete data set, between 1994 and 1996, is available. Discharge observations used come from rating curves that are typically not very accurate [45]. However, in order to evaluate model performances, observed values of discharge at the basin outlet are assumed to be error-free when compared to model results.

The topography of the area is represented by means of a SRTM 90 m resolution DEM which is used to derive the river (streamflow) network. By knowing the river network, it is possible to divide the basin into 68 sub-basins and estimate the main topographic characteristics (maximum drainage length, drainage area and average elevation) for each of them. The river network is classified using the approach proposed by Horton [46].