

This discussion paper is/has been under review for the journal Hydrology and Earth System Sciences (HESS). Please refer to the corresponding final paper in HESS if available.

Can assimilation of crowdsourced streamflow observations in hydrological modelling improve flood prediction?

M. Mazzoleni¹, M. Verlaan², L. Alfonso¹, M. Monego³, D. Norbiato³, M. Ferri³, and D. P. Solomatine^{1,4}

¹UNESCO-IHE Institute for Water Education, Delft, the Netherlands

²Deltares, Delft, the Netherlands

³Alto Adriatico Water Authority, Venice, Italy

⁴Delft University of Technology, Water Resources Section, Delft, the Netherlands

Received: 28 September 2015 – Accepted: 13 October 2015 – Published: 3 November 2015

Correspondence to: M. Mazzoleni (m.mazzoleni@unesco-ihe.org)

Published by Copernicus Publications on behalf of the European Geosciences Union.

HESSD

12, 11371–11419, 2015

Can assimilation of crowdsourced observations improve flood prediction

M. Mazzoleni et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

⏪

⏩

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



Abstract

Monitoring stations have been used for decades to properly measure hydrological variables and better predict floods. To this end, methods to incorporate such observations into mathematical water models have also been developed, including data assimilation. Besides, in recent years, the continued technological improvement has stimulated the spread of low-cost sensors that allow for employing crowdsourced and obtain observations of hydrological variables in a more distributed way than the classic static physical sensors allow. However, such measurements have the main disadvantage to have asynchronous arrival frequency and variable accuracy. For this reason, this study aims to demonstrate how the crowdsourced streamflow observations can improve flood prediction if integrated in hydrological models. Two different types of hydrological models, applied to two case studies, are considered. Realistic (albeit synthetic) streamflow observations are used to represent crowdsourced streamflow observations in both case studies. Overall, assimilation of such observations within the hydrological model results in a significant improvement, up to 21 % (flood event 1) and 67 % (flood event 2) of the Nash–Sutcliffe efficiency index, for different lead times. It is found that the accuracy of the observations influences the model results more than the actual (irregular) moments in which the streamflow observations are assimilated into the hydrological models. This study demonstrates how networks of low-cost sensors can complement traditional networks of physical sensors and improve the accuracy of flood forecasting.

1 Introduction

Observations of hydrological variables measured by physical sensors have been increasingly integrated into mathematical models by means of model updating methods. The use of these techniques allows for the reduction of intrinsic model uncertainty and improves the flood forecasting accuracy (Todini et al., 2005; McLaughlin, 2002).

HESSD

12, 11371–11419, 2015

Can assimilation of crowdsourced observations improve flood prediction

M. Mazzoleni et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures



Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



HESSD

12, 11371–11419, 2015

Can assimilation of crowdsourced observations improve flood prediction

M. Mazzoleni et al.

[Title Page](#)

[Abstract](#)

[Introduction](#)

[Conclusions](#)

[References](#)

[Tables](#)

[Figures](#)



[Back](#)

[Close](#)

[Full Screen / Esc](#)

[Printer-friendly Version](#)

[Interactive Discussion](#)



The main idea behind model updating techniques is to either update model input, states, parameters or outputs as new observations become available (WMO, 1992; Refsgaard, 1997). Input update is the classical method used in operational forecasting as uncertainties of the input data can be considered as the main source of uncertainty (Bergstrom et al., 1980; Krouse, 1979; Canizares et al., 1998; Todini et al., 2005). Regarding the state updating, Kalman filtering theory (Kalman, 1960) is one of the most used approaches when new water observations are available. This method is optimal in case of linear systems. In case of non-linear systems the extended Kalman filter (Kalman, 1960; Verlaan, 1998; Madsen and Canizares, 1999; Aubert et al., 2003) or Ensemble Kalman filter (EnKF, Evensen, 2006) are used in order to overcome the limitation of the linearity assumption. The output updating method (or error prediction) is based on the fact that the error between the model predictions and the new observations are usually found to be serially correlated. One example of output updating method is the regression model, such as an auto-regressive moving average (ARMA) model (Box and Jenkins, 1970). Some of the earlier examples of error prediction include Jamieson et al. (1972); Lundberg (1982); Szollosi-Nagy et al. (1983) and Abebe and Price (2003). Updating of model parameters, originally assessed through optimal calibration, is less common than the other three types of updating in flood forecasting (Young, 1994; Xie and Zhang, 2010; Lu et al., 2013; Yuning et al., 2014). Examples of combined updating of model states and parameters are reported in Moradkhani et al. (2005). Recently, Liu et al. (2012) presented a comprehensive literature review about the latest advanced of model updating procedures in operational flood forecasting by means of water observations coming from physical sensors or remote sensing in a distributed fashion.

Due to the complex nature of the hydrological processes, spatially and temporally distributed measurements are needed in the model updating procedures to ensure a proper flood prediction (Clark et al., 2008; Rakovec et al., 2012). Aubert et al. (2003) integrated distributed values of soil moisture and streamflow from physical sensors in a lumped conceptual hydrological model by means of an extended Kalman Filter.

Streamflow prediction is improved by the assimilation of both soil moisture and streamflow individually and by coupled assimilation. Aubert et al. (2003) found out that the combined assimilation of streamflow and soil moisture observations improve the model performances in all the forecasting periods and noted that the soil moisture assimilation helps to better predict flood events while streamflow – the low flow situations. Cao et al. (2006) proposed a calibration approach based on integration of multiple internal variables with multi-site location, resulting in a more realistic parameterization of the hydrological process. De Lannoy et al. (2007) assimilated distributed values of soil moisture in agricultural field assessing the influence of the biased or the bias-corrected state estimates into a biased model. They pointed out how the results are dependent to the nature of the model itself. In fact, in case of a model that is only biased for soil moisture it is better to post-process the soil moisture with the bias analysis than updating the model states since the large increment of updated soil moisture might results in wrong water balance. Mendoza et al. (2012) evaluated the performances of a flood forecasting scheme assimilating the sparse streamflow observations using an Ensemble Kalman Filter. They found that, for the considered case study, the hydrologic process representation for the upper part of the basin is the major source of uncertainty. In Rakovec et al. (2012); Lee et al. (2012) and Chen et al. (2012) assimilation of distributed streamflow observations were assimilated in the hydrological models with different structures. Overall, the authors found that assimilation of observation in an inner point of the basin helps to further improve the hydrograph estimation. Moreover, it is demonstrated that assimilation performances are more sensitive to the spatial distribution of sensors rather than to the updating frequency. We have conducted experiments on assimilating the synthetic distributed observations of streamflow from physical sensors into two different structures of a semi-distributed hydrological model in order to assess the effect of sensor location and accuracy on the assimilation performances (Mazzoleni et al., 2015a). It has been found that different model structures react in different way to the assimilation of streamflow observations coming from different sensors

Can assimilation of crowdsourced observations improve flood prediction

M. Mazzoleni et al.

[Title Page](#)[Abstract](#)[Introduction](#)[Conclusions](#)[References](#)[Tables](#)[Figures](#)[Back](#)[Close](#)[Full Screen / Esc](#)[Printer-friendly Version](#)[Interactive Discussion](#)

location and that the inaccurate rating curve is the main source of uncertainty in the assimilation process.

It is worth noting that the traditional physical sensors require proper maintenance and personnel which can be very expensive in case of a vast network. Recently, the technological improvement led to the spread of low-cost sensors used to measure hydrological variables such as water level or precipitation. An example of such sensors, defined in the following as “social sensor”, is a smart-phone camera used to measure the water level at a staff gauge with an associate QR code used to infer the spatial location of the measurement (see Fig. 1). The main advance of using these type of sensors is that they can be used not only by technicians but also by regular citizens, and that due to their reduced cost a more spatially distributed coverage can be achieved.

The idea of designing such alternative networks of low-cost social sensors and using the obtained crowdsourced observations is the base of the EU-FP7 WeSenseIt project (2012–2016), which also sponsors this research. Various other projects have also been initiated in order to assess the usefulness of crowdsourced observations inferred by low-cost sensors owned by citizens.

For instance, in the project CrowdHydrology (Lowry and Fienen, 2013), a method to monitor stream stage at designated gauging staffs using crowd source-based text messages of water levels is developed using untrained observers. Cifelli et al. (2005) described a community-based network of volunteers (CoCoRaHS), engaged in collecting precipitation measurements of rain, hail and snow. An example of hydrological monitoring, established in 2009, of rainfall and streamflow values within the Andean ecosystems of Piura, Peru, based on citizen observations is reported in Célleri et al. (2009). Degrossi et al. (2013) used a network of wireless sensors in order to map the water level in two rivers passing by Sao Carlos, Brazil. Recently, the iSPUW Project is aims to integrate data from advanced weather radar systems, innovative wireless sensors and crowdsourcing of data via mobile applications in order to better predict flood events in the urban water systems of the Dallas–Fort Worth Metroplex (Seo et al., 2014; iSPUW 2015). Other examples of crowdsourced the water-related information include the so-

HESSD

12, 11371–11419, 2015

Can assimilation of crowdsourced observations improve flood prediction

M. Mazzoleni et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures



Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



and variable accuracy in a semi-distributed hydrological model. It has been shown that the integration of distributed uncertain intermittent observations with single measurements coming from physical sensors would allow for the further improvements in model accuracy. However, we have not considered the possibility that the asynchronous observations might be coming at the moments not coordinated with the model time steps.

A possible solution to handle asynchronous observations in time with EnKF is to assimilate them at the moments coinciding with the model time steps (Sarkov et al., 2010). However, as these authors mention, this approach requires the disruption of the ensemble integration, the ensemble update and a restart, which may not be feasible for large-scale forecasting applications.

In case of oceanographic studies the continuous approaches, such as 3D-Var or 4D-Var methods, are implemented in order to integrate asynchronous observations at their corresponding arrival moments (Derber and Rosati, 1989; MacPherson, 1991; Huang et al., 2002; Ragnoli et al., 2012). In variational data assimilation, the past observations are used simultaneously to minimize the cost function that measures the weighted difference between background states and observations over the time interval, and identify the best estimate of the initial state condition (Ide et al., 1997; Li and Navon, 2001; Drecourt, 2004).

In addition to the 3D-Var and 4D-Var methods, Hunt et al. (2004) propose a Four Dimensional Ensemble Kalman Filter (4DEnKF) which adapts EnKF to handle observations that have occurred at non-assimilation times. In this method the linear combinations of the ensemble trajectories are used to quantify how well a model state at the assimilation time fits the observations at the appropriate time. Furthermore, in case of linear dynamics 4DEnKF is equivalent to instantaneous assimilation of the measured data (Hunt et al. 2004). Similarly to 4DEnKF, Sarkov et al. (2010) proposed the Asynchronous Ensemble Kalman Filter (AEnKF), a modification of the EnKF, mainly equivalent to 4DEnKF, used to assimilate asynchronous observations (Rakovec et al., 2015). Contrary to the EnKF, in the AEnKF current and past observations are simultaneously assimilated at a single analysis step without the use of adjoint model. Yet another ap-

Can assimilation of crowdsourced observations improve flood prediction

M. Mazzoleni et al.

[Title Page](#)

[Abstract](#)

[Introduction](#)

[Conclusions](#)

[References](#)

[Tables](#)

[Figures](#)



[Back](#)

[Close](#)

[Full Screen / Esc](#)

[Printer-friendly Version](#)

[Interactive Discussion](#)



proach to assimilate asynchronous observations in models is the so-called First-Guess at the Appropriate Time (FGAT) method. Like in 4D-Var, the FGAT compares the observations with the model at the observation time. However, in FGAT the innovations are assumed constant in time and remain the same within the assimilation window (Massart et al., 2010).

Having reviewed all the described approaches, in this study we have decided to use a straightforward and pragmatic method, due to the linearity of the hydrological models implemented in this study, similar to the AEnKF to assimilate the asynchronous crowdsourced observations.

The aims of this study are to (a) assess the influence of different arrival frequency of the crowdsourced observations and their related accuracy on the assimilation performances in case of a single social sensor, (b) to integrate the distributed low-cost social sensors with a single physical sensor to assess the improvement in the flood prediction performances in an early warning system. The methodology is applied in the Brue (UK) and Bacchiglione (Italy) catchments, considering lumped and semi-distributed hydrological models respectively. Due to the fact that streamflow observations from social sensors are not available in the Brue catchment while in the Bacchiglione basin the sensors are being recently installed, the synthetic time series, asynchronous in time and with random accuracy, that imitate the crowdsourced observations, are generated and used.

The study is organized as follows. Firstly, the case studies and the datasets used are presented. Secondly, the hydrological models used are described. Then, the procedure used to integrate the crowdsourced observations is reported. Finally, the results, discussion and conclusions are presented.

HESSD

12, 11371–11419, 2015

Can assimilation of crowdsourced observations improve flood prediction

M. Mazzoleni et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures



Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



2 Case studies and datasets

2.1 Brue basin

The first case study is located in the Brue catchment (Fig. 2), in Somerset, with a drainage area of about 135 km^2 at the basin outlet in Lovington. Using the SRTM DEM with the 90 m resolution it is possible to derive the streamflow network and the consequent time of concentration, by means of the Giandotti equations (1934), which is about 10 h. The hourly precipitation (49 rainfall stations) and streamflow data used in this study are supplied by the British Atmospheric Data Centre from the HYREX (Hydrological Radar Experiment) project (Moore et al., 2000; Wood et al., 2000). The average precipitation value in the basin is estimated using the Ordinary Kriging (Matheron, 1963).

2.2 Bacchiglione basin

The second case study is the upstream part of the Bacchiglione River basin, located in the North-East of Italy, and tributary of the River Brenta which flows into the Adriatic Sea at the South of the Venetian Lagoon and at the North of the River Po delta. The study area has an overall extent and river length of about 400 km^2 and 50 km (Ferri et al., 2012). The main urban area located in the downstream part of the study area is Vicenza. The analyzed part of the Bacchiglione River has four main tributaries. On the Western side the confluences with the Bacchiglione are the Leogra, the Orolò and the Retrone River, whose junction is located in the urban area itself. In Fig. 2 the Retrone River it is not shown since it does not influence the water level measured at the gauged station of Vicenza (Ponte degli Angeli). On the Eastern side there is the Timonchio River (see Fig. 3). The Alto Adriatico Water Authority (AAWA) has implemented an Early Warning System to properly forecast the possible future flood events. Recently, within the activities of the WeSenseIt Project (Ciravegna et al., 2013), one physical sensor and three staff gauges complemented by a QR code (social sensor,

Can assimilation of crowdsourced observations improve flood prediction

M. Mazzoleni et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures



Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



as represented in Fig. 1) were installed in the Bacchiglione River to measure the water level. In particular, the physical sensor is located at the outlet of the Leogra basin while the three social sensors are located at the Timonchio, Leogra and Orolo basin outlet respectively (see Fig. 3).

2.3 Datasets

In the Brue basin two different flood events which occurred between 28 October to 16 November 1994 (flood event 1) and from 14 January to 4 February 1995 (flood event 2) are considered. The observed precipitation values are treated as the “perfect forecasts” and are fed into the hydrological model. The observed streamflow data for the considered flood event are available as well.

In case of Bacchiglione basin, the flood event which occurred in May 2013 is considered; it had the high intensity and resulted in several traffic disruptions at various locations upstream Vicenza. For flood forecasting, AAWA uses the 3 day weather forecast as the input to the hydrological model. The observed values of streamflow and water level at Ponte degli Angeli are used to assess the performance of the hydrological model.

3 Hydrological modelling

3.1 Brue basin

A lumped conceptual hydrological model is implemented to estimate the flood hydrograph at the outlet section of the Brue catchment. Direct runoff is used as input in the conceptual model and assessed by means of the Soil Conservation Service Curve Number (SCS-CN) method. The average value of CN within the catchment is calibrated by minimizing the difference between the simulated volume and observed quickflow, using the method proposed by Eckhardt (2005), at the outlet section.

HESSD

12, 11371–11419, 2015

Can assimilation of crowdsourced observations improve flood prediction

M. Mazzoleni et al.

[Title Page](#)

[Abstract](#)

[Introduction](#)

[Conclusions](#)

[References](#)

[Tables](#)

[Figures](#)

[⏪](#)

[⏩](#)

[◀](#)

[▶](#)

[Back](#)

[Close](#)

[Full Screen / Esc](#)

[Printer-friendly Version](#)

[Interactive Discussion](#)



Can assimilation of crowdsourced observations improve flood prediction

M. Mazzoleni et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

⏪

⏩

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



The input for the hydrological model consists of precipitation only. The hydrological response of the catchment is estimated using a hydrological model that considers the routines for runoff generation and a simple routing procedure. The processes related to runoff generation (surface, sub-surface and deep flow) are modelled mathematically by applying the water balance to a control volume representative of the active soil at the sub-basin scale. The water content $S(t)$ in the soil is updated at each calculation step dt using the following balance equation:

$$S(t + dt) = S(t) + P(t) - R_{\text{sur}}(t) - R_{\text{sub}}(t) - L(t) - ET(t). \quad (2)$$

where P and ET are the components of precipitation and evapotranspiration, while R_{sur} , R_{sub} and L are the surface runoff, sub-surface runoff and deep percolation model states respectively (see Fig. 3). The surface runoff is expressed by the equation based on specifying the critical threshold beyond which the mechanism of dunnian flow (saturation excess mechanism) prevails:

$$R_{\text{sur}}(t) = \begin{cases} C \left(\frac{S(t)}{S_{\text{max}}} \right) P(t) \Rightarrow P(t) \leq f = \frac{S_{\text{max}}(S_{\text{max}} - S(t))}{(S_{\text{max}} - CS(t))} \\ P(t) - (S_{\text{max}} - S(t)) \Rightarrow P(t) > f \end{cases} \quad (3)$$

where C is a coefficient of soil saturation obtained by calibration, and S_{max} is the content of water at saturation point which depends on the nature of the soil and on its use.

The sub-surface flow is considered proportional to the difference between the water content $S(t)$ at time t and that at soil capacity

$$R_{\text{sub}}(t) = c \cdot (S(t) - S_c). \quad (4)$$

while the estimated deep flow is evaluated according to the expression proposed by Laio et al. (2001):

$$L(t) = \frac{K_s}{e^{\beta \left(1 - \frac{S_c}{S_{\text{max}}}\right)} - 1} \left(e^{\beta \left(\frac{S(t) - S_c}{S_{\text{max}}}\right)} - 1 \right). \quad (5)$$

Can assimilation of crowdsourced observations improve flood prediction

M. Mazzoleni et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

⏪

⏩

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



where, K_s is the hydraulic conductivity of the soil in saturation conditions, β is a dimensionless exponent characteristic of the size and distribution of pores in the soil. The evaluation of the real evapotranspiration is performed assuming it as a function of the water content in the soil and potential evapotranspiration, calculated using the formulation of Hargreaves and Samani (1985).

Knowing the values of R_{sur} , R_{sub} and L , it is possible to model the surface Q_{sur} , sub-surface Q_{sub} and deep flow Q_g routed contributes according to the conceptual framework of the linear reservoir at the closing section of the single sub-basin. In particular, in case of Q_{sur} the value of k , which is the inverse of the residence time in the basin slopes, can be calculated as $k = v/L$ where v is a characteristic slopes velocity spatially uniform at sub-basin scale (calibration parameter), which differs according to the type of the considered runoff (surface and sub-surface) and to the different transport processes, while L is the average slope length. On the other hand, in case of Q_{sub} and Q_g the value of k is calibrated comparing the observed and simulated discharge at Vicenza as previously described.

The flood propagation along the main river channel is represented one-dimensional hydrodynamic model, MIKE 11 (DHI, 2005). This model solves the Saint Venant Equations in case of unsteady flow based on an implicit finite difference scheme proposed by Abbott and Ionescu (1967). However, in order to reduce the computational time required by the analysis performed in this study MIKE11 is replaced by a hydrological routing Muskingum–Cunge model (see, e.g. Todini, 2007), considering river cross-sections as rectangular for the estimation of hydraulic radii, wave celerity and the other hydraulic variables.

Calibration of the hydrological and hydrodynamic model parameters is performed by AAWA considering the time series of precipitation from 2000 to 2010 in order to minimize the root mean square error between observed and simulated values of water level at Ponte degli Angeli gauged station.

4 Data assimilation procedure

4.1 Kalman filter

In DA it is typically assumed that the dynamic system can be represented in the state-space as follows:

$$5 \quad \mathbf{x}_t = M(\mathbf{x}_{t-1}, \vartheta, I_t) + \mathbf{w}_t \quad \mathbf{w}_t \sim N(0, S_t). \quad (6)$$

$$\mathbf{z}_t = H(\mathbf{x}_t, \vartheta) + \mathbf{v}_t \quad \mathbf{v}_t \sim N(0, R_t). \quad (7)$$

where, \mathbf{x}_t and \mathbf{x}_{t-1} are state vectors at time t and $t - 1$, M is the model operator that propagates the states \mathbf{x} from its previous condition to the new one as a response to the inputs I_t , while H is the operator which maps the model states into output \mathbf{z}_t . The system and measurements errors \mathbf{w}_t and \mathbf{v}_t are assumed to be normally distributed with zero mean and covariance S and R . In a hydrological modelling system, these states can represent the water stored in the soil (soil moisture, groundwater) or on the earth surface (snow pack). These states are one of the governing factors that determine the hydrograph response to the inputs into the catchment.

15 In case of the linear systems used in this study, the discrete state-space system of Eq. (1) can be represented as follows (Szilagyi and Szollosi-Nagy, 2010):

$$\mathbf{x}_t = \Phi \mathbf{x}_{t-1} + \Gamma I_t + \mathbf{w}_t. \quad (8)$$

$$Q_t = \mathbf{H} \mathbf{x}_t + \mathbf{v}_t. \quad (9)$$

where t is the time step, \mathbf{x} is vector of the model states (stored water volume in m^3), Φ is the state-transition matrix (function of the model parameters n and k), Γ is the input-transition matrix, \mathbf{H} is the output matrix, and I and Q are the input (forcing) and model output (discharge in this case). For example, for $n = 3$ the matrix \mathbf{H} is expressed as $\mathbf{H} = [0 \ 0 \ k]$. Expressions for matrices Φ and Γ can be found in Szilagyi and Szollosi-Nagy (2010).

Can assimilation of crowdsourced observations improve flood prediction

M. Mazzoleni et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

⏪

⏩

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



For the Bacchiglione model, the preliminary sensitivity analysis on the model states (soil content S and the storage water x_{sur} , x_{sub} and x_L related to Q_{sur} , Q_{sub} and Q_g) is performed in order to decide on which of the states to update. The results of this analysis (shown in the next section) pointed out that the stored water volume x_{sur} (estimated using Eq. (7) with $n = 1$, $H = k$ and I_t replaced by R_{sur}) is the most sensitive state and for this reason we decided to update only this state.

The Kalman Filter (KF, Kalman, 1960) is a mathematical tool which allows estimating, in an efficient computational (recursive) way, the state of a process which is governed by a linear stochastic difference equation. KF is optimal under the assumption that the error in the process is Gaussian; in this case KF is derived by minimizing the variance of the system error (error in state) assuming that the model state estimate is unbiased. In an attempt to overcome these limitations, various variants of the Kalman filter, such as the extended Kalman filter (EKF), unscented Kalman filter and ensemble Kalman filter (EnKF) have been proposed.

Kalman filter procedure can be divided in two steps, namely forecast equations, (Eqs. 10 and 11), and update (or analysis) equations (Eqs. 12–14):

$$\mathbf{x}_t^- = \mathbf{\Phi} \mathbf{x}_{t-1}^+ + \mathbf{\Gamma} I_t. \quad (10)$$

$$\mathbf{P}_t^- = \mathbf{\Phi} \mathbf{P}_{t-1}^+ \mathbf{\Phi}^T + \mathcal{S}. \quad (11)$$

$$\mathbf{K}_t = \mathbf{P}_t^- \mathbf{H}^T (\mathbf{H} \mathbf{P}_t^- \mathbf{H}^T + \mathbf{R})^{-1}. \quad (12)$$

$$\mathbf{x}_t^+ = \mathbf{x}_t^- + \mathbf{K}_t (Q_t^0 - \mathbf{H} \mathbf{x}_t^-). \quad (13)$$

$$\mathbf{P}_t^+ = (\mathbf{I} - \mathbf{K}_t \mathbf{H}) \mathbf{P}_t^-. \quad (14)$$

where \mathbf{K}_t is the Kalman gain matrix, \mathbf{P} is the error covariance matrix, Q^0 is the new observation and \mathbf{M}_Q is the model error matrix. The prior model states \mathbf{x} at time t are updated, as the response to the new available observations, using the analysis equations (Eqs. 12 to 14). This allows for estimation of the updated states values (with superscript +) and then assessing the background estimates (with superscript –) for the next time step using the time update equations (Eqs. 10 and 11).

4.2 Assimilation of asynchronous streamflow observations with irregular accuracy

In most of the hydrological applications of DA, observations from physical sensors are integrated into water models at a regular, synchronous, time step. However, as showed in Fig. 1, a social sensor can be used by different operators, having different accuracy, to measure water level at a specific point. For this reason, social sensors provide crowdsourced observations which are asynchronous in time and with a higher degree of uncertainty than the one of observations from physical sensors. In particular, crowdsourced observations have three main characteristics: (a) irregular arrival frequency (asynchronicity), (b) random accuracy, (c) random number of observations received by the static device within two model time steps.

As described in the Introduction, various methods have been proposed in order to include asynchronous observations in models. Having reviewed them, in this study we are proposing a somewhat simpler DA approach for integrating Crowdsourced Observations into hydrological models (DACO). This method is based on the assumption that the change in the model states and in the error covariance matrices within the two consecutive model time steps t_0 and t (observation window) is linear, while the inputs are assumed constant. All the observations received during the observation window are assimilated in order to update the model states and output at time t . Therefore, assuming that one observation would be available at time t_0^* , the first step of such a filter (a in Fig. 4) is the definition of the model states and error covariance matrix at t_0^* as:

$$\mathbf{x}^-(t_0^*) = \mathbf{x}(t^*) \frac{t_0^* - t_0}{t - t_0} - \mathbf{x}(t_0^*) \frac{t_0^* - t}{t - t_0}. \quad (15)$$

$$\mathbf{P}^-(t_0^*) = \mathbf{P}(t^*) \frac{t_0^* - t_0}{t - t_0} - \mathbf{P}(t_0^*) \frac{t_0^* - t}{t - t_0}. \quad (16)$$

The second step (b in Fig. 4) is the estimation of the updated model states and error covariance matrix, as the response to the streamflow observation $\mathbf{Q}_{\text{obs}}(t_0^*)$. The esti-

HESSD

12, 11371–11419, 2015

Can assimilation of crowdsourced observations improve flood prediction

M. Mazzoleni et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

⏪

⏩

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



mation of the posterior values of $\mathbf{x}^+(t_0^*)$ and $\mathbf{P}^+(t_0^*)$ is performed by Eqs. (13) and (14) respectively. The Kalman gain, estimated by Eq. (12), where the prior values of model states and error covariance matrix at t_0^* are used. Knowing the posterior value of $\mathbf{x}(t_0^*)$ and $\mathbf{P}(t_0^*)$ it is possible to predict the value of states and covariance matrix at one model step ahead, t^* (c in Fig. 4) using the model forecast equations Eqs. (10) and (11).

The last step (d in Fig. 4) is the estimation of the updated value of \mathbf{x} and \mathbf{P} at time step t^* . This is performed by means of a linear interpolation between the current values of \mathbf{x} and \mathbf{P} at t_0^* and t^* :

$$\mathbf{x}^+(t) = \mathbf{x}(t^*) \frac{t - t_0^*}{t^* - t_0^*} - \mathbf{x}(t_0^*) \frac{t - t^*}{t^* - t_0^*}. \quad (17)$$

$$\mathbf{P}^+(t) = \mathbf{P}(t^*) \frac{t - t_0^*}{t^* - t_0^*} - \mathbf{P}(t_0^*) \frac{t - t^*}{t^* - t_0^*}. \quad (18)$$

Assuming that the new streamflow observation is available at an intermediate time t_1^* (between t_0^* and t), the procedure is repeated considering the values at t_0^* and t as for the linear interpolation. Then, in case when no more observations are available, the updated value of $\mathbf{x}^+(t)$ is used to predict the model states and output at $t + 1$ (Eqs. 10 and 11). Finally, in order to account for the intermittent behaviour of such observations, the approach proposed by Mazzoleni et al. (2015b) is applied. In this method, the model states matrix \mathbf{x} is updated and forecasted when observations are available, while without observations the model is run using Eq. (10) and covariance matrix \mathbf{P} propagated at the next time step using Eq. (11)

4.3 Observation accuracy

In this section, the uncertainty related to the streamflow observations is characterised. The observational error is assumed to be the normally distributed noise with zero mean and given standard deviation (Weerts and El Serafy, 2006):

$$\sigma_Q(t) = \alpha(t) \cdot Q_{\text{true}}(t). \quad (19)$$

Can assimilation of crowdsourced observations improve flood prediction

M. Mazzoleni et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

⏪

⏩

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



where the coefficient α is a variable related to the degree of uncertainty of the measurement. Due to the unpredictable accuracy in the crowdsourced observations the coefficient α is assumed to be a random stochastic variable between 0.1 and 0.3 (Mazzoleni et al., 2015b). Cortes et al., 2014 argue (and this is a reasonable suggestion) that the uncertainty of a measurement provided by a well-trained technician is smaller than the one coming from a normal citizen. For this reason we assumed that the maximum value of α is three times higher than the uncertainty coming from the physical sensors. The value of Q_{true} is the streamflow value measured at a asynchronous time step and it will be described in the next section.

5 Experimental setup

In this section, two sets of experiments are performed in order to test the proposed method and assess the benefit to integrate crowdsourced observations, asynchronous in time and with variable accuracy, in real-time flood forecasting.

In the first set of experiments, called “Experiments 1”, assimilation of streamflow observations at one social sensor location is carried out to understand the sensitivity of the employed hydrological model (KMN) under various scenarios of such observations.

In the second set of experiments, called “Experiments 2”, the distributed observations coming from social and physical sensors, at four locations within the Bacchiglione basin, are considered, with the aim of assessing the improvement in the flood forecasting accuracy. The social sensors, showed in Figs. 1 and 3, were installed in the summer of 2014 within the framework of the WeSenseIt project.

5.1 Experiments 1: assimilation of crowdsourced observations from one social sensor

The focus of Experiments 1 is to study the performance of the hydrological model (KMN) assimilating crowdsourced observations, having lower arrival frequencies than

Can assimilation of crowdsourced observations improve flood prediction

M. Mazzoleni et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

⏪

⏩

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



ment, in this case at Ponte degli Angeli. For this reason, five different experimental settings are introduced (Fig. 6) corresponding to different types of sensors used.

Firstly, only the observations coming from the physical sensor at the Leogra sub-basin are used to update the hydrological model of basin B (setting A). Secondly, in setting B the model improvement in case of assimilation of crowdsourced observations at the same location of setting A is analyzed. In setting C only the distributed crowdsourced observations within the basin are assimilated into the hydrological model. Then, setting D accounts for the integration of crowdsourced and physical observations, contrary to the setting C where the physical sensors is dropped in favour of the crowdsourced sensor at Leogra. Finally, setting E consider the complete integration between physical and social sensor in Leogra, Timonchio and Orolo sub-basins.

6 Results and discussions

6.1 Experiments 1: influence of crowdsourced observations on flood forecasting

The observed and simulated hydrographs at the outlet section of the Brue basin with and without the model update (considering hourly streamflow observations) are reported in Fig. 7 for two different flood events, characterised by the number and separation of flood peaks. As expected, it can be seen that the updated model tends to better represent the flood events than model without updating.

The results of scenario 1 for flood event 1, having from 1 to 30 observations within the observation window, are represented in Fig. 8. As it can be seen, increasing the number of observations within the observation window results in the improvement of the NSE for different lead time values, but it becomes negligible for more than five observations. This means that the additional observations do not add information useful for improving the model performance. The same type of analysis is performed with the scenarios 2 to 9 (Fig. 9). The results obtained in Fig. 9 show that in case of irregular ar-

Can assimilation of crowdsourced observations improve flood prediction

M. Mazzoleni et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures



Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



Can assimilation of crowdsourced observations improve flood prediction

M. Mazzoleni et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

⏪

⏩

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



rival frequency (scenarios 2 and 3) the NSE is higher than in scenarios 4 and 5, where observations vary in accuracy. These results point out that the model performance is more sensitive to the accuracy of the observations than to the moment in time at which the streamflow observations become available. However, it can be observed how from scenarios 2 to 5 the trend it is not as smooth as the one obtained with scenario 1. This can be related to the fact that NSE may vary with varying arrival frequency and observations accuracy. It is worth noting that the irregular moment in which the observation becomes available within the observation window is randomly selected in the synthetic experiments. This means that for a given number of observations (for example 5), the five observations will arrive at different moments, for different model runs, and this results in different values of NSE. The reason why scenario 1 has a smooth trend is in the fact that since the arrival frequency is set as regular for different model runs, the moments in which the observations became available will always be the same for a given number of observations. A smooth trend is also obtained for scenarios 6, 7, 8 and 9 but this is related to the periodic behaviour of the observations as explained below.

In order to remove the random behaviour related to the irregular arrival frequency and observation accuracy, different model runs (100 in this case) are carried out, assuming different random values of arrival and accuracy (coefficient α) during each model run, for a given number of observations and lead time, for the two flood events (see Fig. 10 for the flood event 1). Overall, the standard deviation of NSE, σ (NSE), tend to decrease for the high number of observations. Scenario 2 has the lower standard deviation for low values of discharge observations due to the fact that the arrival frequency has to coincide with the model time step and this tends to stabilize the NSE. In addition, the irregular arrival frequency (scenarios 2 and 3) has a higher impact on the σ (NSE) than on the mean NSE value μ (NSE). Besides, the variable observations accuracy (scenario 4) influences more μ (NSE) than σ (NSE), as described before. The combined effect of irregular frequency and uncertainty is reflected in scenario 5 which has the lower mean and higher standard deviation of NSE if compared to the first four scenarios.

An interesting fact is that passing from periodic (Fig. 10a and b) to non-periodic (Fig. 10c and d) behaviour of the observations the standard deviation is significantly reduced, while the mean remains the same. A non-periodic behaviour of the observations, common in real life, helps to reduce the fluctuation of the NSE generated by the random behaviour of streamflow observations.

Table 1 shows the NSE values and model improvement obtained for the different experimental scenarios during flood event 1 and 2.

Finally, the results obtained for scenarios 10 and 11 are showed in Fig. 11. Also in this case, the NSE obtained for the flood event 1 is higher than the one obtained for the flood event 2. The irregular number of observations in scenario 10 in each observation window seems to provide the same model performances μ (NSE) if compared with scenario 9. One the main outcome is that the intermittent nature of the observations (scenario 11) induces a drastic reduction of the NSE and an increase in its noise in both considered flood events.

6.2 Experiments 2: influence of distributed physical and social sensors

In order to find out what model states leads to a maximum increase of the model performance, a preliminary sensitivity analysis is performed. The four model states, x_S , x_{sur} , x_{sub} and x_L are perturbed by $\pm 20\%$ around the true state value using the uniform distribution, every time step from the initial time step up to the perturbation time (PT). No correlation between time steps is considered. After PT the model realizations are run without perturbation in order to assess the perturbation effect on the system memory. From the results reported in Fig. 12, the model state x_{sur} is the most sensitive states if compared to the other five. In addition, the perturbations of all the states seem to affect the model output even after the PT (high system memory).

The synthetic physical and crowdsourced observations are assimilated in order to improve the poor model prediction in Vicenza affected by the underestimated estimation of the 3-days rainfall forecast used as normal input in flood forecasting practice in this area. Scenarios 10 and 11, described in the previous sections, are used in this ex-

Can assimilation of crowdsourced observations improve flood prediction

M. Mazzoleni et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

⏪

⏩

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



periment in order to represent an irregular and random behaviour of the crowdsourced observations.

The results of this analysis are shown in Fig. 13. Different model runs (100) are performed for the Leogra sub-basin, Fig. 13, to reduce the effect induced by the random arrival frequency and accuracy of the crowdsourced observations within the observation window as described above. It can be seen that the assimilation of physical observations provides a better flood prediction at the Leogra basin if compared with the assimilation of a small number of crowdsourced observations. In particular, Fig. 13a and b show that the same NSE values achieved with assimilation of physical observations (hourly frequency and high accuracy) can be obtained by assimilating between 10 and 20 crowdsourced observations per hour. However, the overall reduction of NSE in case of intermittent observations is such that even with a high number of observations (even higher than 50 per hour) the NSE is always lower than the one obtained assimilating physical observations for any lead time. Figure 13c and d show analogous results expressed in terms of different lead times.

Figures 14 and 15 show the results obtained from the experiments settings represented in Fig. 6 in case of physical and crowdsourced observations. Also in this case, different simulation runs (100) of random values of arrival frequency and uncertainty are performed.

One of the main outcomes of these analyses is that the replacement of a physical sensor for a social sensor at only one location (settings B) does not improve the model performance in terms of discharge (Fig. 14) for different lead time values. Distributed locations of social sensors (setting C) can provide higher value of NSE than a single physical sensor, even for low number of observations in both regular and intermittent crowdsourced observations. It is interesting to note that in case of integration between social and physical sensors (setting D) the NSE is higher than in case of setting C for low number of observations. However with the higher number of observations, setting C is the one providing the best model improvement for low lead time values. This can be due to the fact that the physical sensor at Leogra provides constant improvement, for

HESSD

12, 11371–11419, 2015

Can assimilation of crowdsourced observations improve flood prediction

M. Mazzoleni et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures



Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



Can assimilation of crowdsourced observations improve flood prediction

M. Mazzoleni et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures



Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



a given lead time, while the social sensor tends to achieve better results with a higher number of observations. This dominant effect of the social sensor, in case of high number of observations, tends to increase for the higher lead times. The best model improvement is achieved in case of setting E, i.e. fully integrating physical sensor with distributed social sensors. In case of intermittent observations (Fig. 14d–f), it can be noticed that the setting D provides higher improvement than setting C. In case of high lead time value (12 h) results of setting C tend to be similar to the ones obtained with setting B. As in case of scenario 10, also in case of scenario 11 the best results are achieved in case of setting E.

Figure 15 shows the standard deviation of the NSE obtained for the different settings in case of lead time 4 h. Higher σ (NSE) values are obtained in case of setting B, while including different sources of crowdsourced observations tend to decrease the value of σ (NSE). It can be observed how σ (NSE) decreases for high value of crowdsourced observations. As expected, the lowest values of σ (NSE) are achieved including the physical sensor in the DA procedure. Similar considerations can be drawn in case of scenario 11, where an higher and more perturbed σ (NSE) values are obtained.

7 Conclusions

This study demonstrates how crowdsourced observations, asynchronous in time and with variable accuracy, can improve flood prediction if integrated in hydrological models. Such observations are assumed to be inferred using low-cost social sensors as, for example, staff gauge connected to a QR code on which people can read the water level indication and send the observations via a mobile phone application. This type of social sensor is tested within the framework of the WeSenseIt FP7 Project. Two different case studies, the Brue (UK) and Bacchiglione (Italy) basins, are considered, and the two types of hydrological models are used. In the Experiments 1 (Brue basin) the sensitivity of the model results to the different frequencies and accuracies of the streamflow observations coming from a hypothetical social sensor at the basin outlet

HESSD

12, 11371–11419, 2015

Can assimilation of crowdsourced observations improve flood prediction

M. Mazzoleni et al.

[Title Page](#)[Abstract](#)[Introduction](#)[Conclusions](#)[References](#)[Tables](#)[Figures](#)[Back](#)[Close](#)[Full Screen / Esc](#)[Printer-friendly Version](#)[Interactive Discussion](#)

are assessed. On the other hand, in the Experiments 2 (Bacchiglione basin), the influence of the combined assimilation of crowdsourced observations, coming from a distributed network of social sensors, and existing streamflow observations from physical sensors, used in the Early Warning System implemented by AAWA, is evaluated. Due to the fact that crowdsourced streamflow observations are not yet available in both case studies, realistic synthetic observations with various characteristics of arrival frequency and accuracy are introduced.

In Experiments 1 it is found that the increasing the number of crowdsourced observations within the observation window increases the model performance even if these observations have irregular arrival frequency and accuracy. Therefore, observations accuracy affects the average value of NSE more than the moment when these observations are assimilated into the model. However, the arrival frequency of the observations results in a significant noise in the NSE estimation. This noise is reduced when the assimilated observations are considered having non-periodic behaviour. In addition, the intermittent nature of the observations tends to drastically reduce the NSE of the model for different values of lead times. In fact, if the intervals between the observations are too large then the abundance of crowdsourced data at other times and places is no longer able to compensate their intermittency.

Experiments 2 showed that, in the Bacchiglione basin, the integration of observations from social sensors and single physical sensor can improve the flood prediction even in case of a small number of intermittent crowdsourced observations. In case of both physical and social sensors located at the same place the assimilation of crowdsourced observations give the same model improvement than the assimilation of physical observations only in case of high number and non-intermittent behaviour.

It can be concluded that for the considered case studies and hydrological models, the assimilation of the crowdsourced observations can improve model performances. In particular, the integration of a network of social sensors with an existing physical sensor can improve the model predictions as shown in the Bacchiglione case study.

Can assimilation of crowdsourced observations improve flood prediction

M. Mazzoleni et al.

Title Page	
Abstract	Introduction
Conclusions	References
Tables	Figures
⏪	▶
◀	▶
Back	Close
Full Screen / Esc	
Printer-friendly Version	
Interactive Discussion	

In our study we have obtained interesting results, however, this work has still certain limitations. Firstly, the proposed method used to assimilate crowdsourced observation is applied in case of linear hydrological model. For this reason, the next steps in this research will be the implementation of the proposed methodology to non-linear hydrological models. Then, realistic synthetic streamflow observations are used in the experiments of this study, while to further validate the results obtained in this study real-life observations coming from social sensors should be considered. Finally, advancing methods for a more accurate assessment of the data quality and accuracy of streamflow observations coming from social sensors need to be considered.

Acknowledgements. This research was funded in the framework of the European FP7 Project WeSenseIt: Citizen Observatory of Water, grant agreement No. 308429. Data used were supplied by the British Atmospheric Data Centre from the NERC Hydrological Radar Experiment Dataset <http://www.badc.rl.ac.uk/data/hyrex/> and by the Alto Adriatico Water Authority.

References

Abbott, M. B. and Ionescu, F.: On the numerical computation of nearly-horizontal flows, *J. Hydraul. Res.*, 5, 97–117, 1967.

Alfonso, J. L. and Price, R.: Coupling hydrodynamic models and value of information for designing stage monitoring networks, *Water Resour. Res.*, 48, 2012.

Arnold Jr., C. P. and Dey, C. H.: Observing-systems simulation experiments: past, present and future, *B. Am. Meteorol. Soc.*, 67, 687–695, 1986.

Au, J., Bagchi, P., Chen, B., Martinez, R., Dudley, S. A., and Sorger, G. J.: Methodology for public monitoring of total coliforms, *Escherichia coli* and toxicity in waterways by Canadian high school students, *J. Environ. Manage.*, 58, 213–230, 2000.

Aubert, D., Loumagne, C., and Oudin, L.: Sequential assimilation of soil moisture and streamflow data in a conceptual rainfall–runoff model, *J. Hydrol.*, 280, 145–161, 2003.

Babovic, V., Canizares, R., Jensen, H. R., and Klinting, A.: Neural networks as routine for error updating of numerical models, *J. Hydraul. Eng. ASCE*, 127, 181–193, 2001.



Can assimilation of crowdsourced observations improve flood prediction

M. Mazzoleni et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures



Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



- Degrossi, L. C., do Amaral, G. G., de Vasconcelos, E. S. M., de Albuquerque, J. P., and Ueyama, J.: Using Wireless Sensor Networks in the Sensor Web for Flood Monitoring in Brazil, in: Proceedings of the 10th International ISCRAM Conference, Baden-Baden, Germany, 12–15 May 2013, 2013.
- 5 Derber, J. and Rosati, A.: A global oceanic data assimilation system, *J. Phys. Oceanogr.*, 19, 1333–1347, 1989.
- Drecourt, J. P.: Data assimilation in hydrological modelling, Ph.D. thesis dissertation, Technical University of Denmark, Hørsholm, 176 pp., 15 June, 2004.
- Eckhardt, K.: How to construct recursive digital filters for baseflow separation, *Hydrol. Process.*, 19, 507–515, 2005.
- 10 Engel, S. R. and Voshell, J. R.: Volunteer biological monitoring: can it accurately assess the ecological condition of streams?, *American Entomologist*, 48, 164–177, 2002.
- Errico, R. M. and Privé, N. C.: An estimate of some analysis-error statistics using the Global Modeling and Assimilation Office observing-system simulation framework, *Q. J. Roy. Meteor. Soc.*, 140, 1005–1012, 2014.
- 15 Errico, R. M., Yang, R., Privé, N. C., Tai, K. S., Todling, R., Sienkiewicz, M. E., and Guo, J.: Development and validation of observing-system simulation experiments at NASA's Global Modeling and Assimilation Office, *Q. J. Roy. Meteor. Soc.*, 139, 1162–1178, doi:10.1002/qj.2027, 2013.
- 20 Evensen, G.: *Data Assimilation: the Ensemble Kalman Filter*, Springer, Germany, 2006.
- Ferri, M., Monego, M., Norbiato, D., Baruffi, F., Toffolon, C., and Casarin, R.: AMICO: la piattaforma previsionale per i bacini idrografici del Nord Est Adriatico (I), in: Proc. XXXIII Conference of Hydraulics and Hydraulic Engineering, Brescia (Italy), 10–14 September 2012, 2012.
- 25 Fienen, M. N. and Lowry, C. S.: Social.Water – a crowdsourcing tool for environmental data acquisition, *Comput. Geosci.*, 49, 164–169, doi:10.1016/j.cageo.2012.06.015, 2012.
- Giandotti, M.: Previsione delle piene e delle magre dei corsi d'acqua, Ministero LL.PP., Memorie e studi idrografici, 8, Servizio Idrografico Italiano, Rome, 1934 (in Italian).
- Hargreaves, G. H. and Samani, Z. A.: Estimating potential evapotranspiration, *J. Irrig. Drain. E.-ASCE*, 108, 223–230, 1982.
- 30 Huang, B., Kinter, J. L., and Schopf, P. S.: Ocean data assimilation using intermittent analyses and continuous model error correction, *Adv. Atmos. Sci.*, 19, 965–992, 2002.

**Can assimilation of
crowdsourced
observations improve
flood prediction**

M. Mazzoleni et al.

[Title Page](#)[Abstract](#)[Introduction](#)[Conclusions](#)[References](#)[Tables](#)[Figures](#)[Back](#)[Close](#)[Full Screen / Esc](#)[Printer-friendly Version](#)[Interactive Discussion](#)

- Hunt, B. R., Kalnay, E., Kostelich, E. J., Ott, E., Patil, D. J., Sauer, T., Szunyogh, I., Yorke, J. A., and Zimin, A. V.: Four-dimensional ensemble Kalman filtering, *Tellus A*, 56, 273–277, 2004.
- Ciravegna, F., Huwald, H., Lanfranchi, V., Wehn, and de Montalvo, U.: Citizen observatories: theWeSenseIt vision, INSPRIRE 2013, Infrastructure for spatial information in the European Community, Florence, Italy, 23–27 June 2013, 2013.
- Ide, K., Courtier, P., Ghil, M., and Lorenc, A. C.: Unified notation for data assimilation: operational, sequential and variational, *J. Meteorol. Soc. Jpn.*, 75, 181–189, 1997.
- Integrated Sensing and Prediction of Urban Water for Sustainable Cities (ISPUW), available at: <http://ispuw.uta.edu/nsf/> (last access: 26 September 2015), 2014.
- Jamieson, D. G., Wilkinson, J. C., and Ibbitt, R. P.: Hydrologic forecasting with sequential deterministic and stochastic stages, in: Proceedings from the International Symposium on Uncertainties in Hydrology and Water Resources Systems, Tucson, Arizona, 11–14 December, 1, 177–187, 1972.
- Kalman, R. E.: A new approach to linear filtering and prediction problems, *J. Basic Eng.-T. ASME*, 82, 35–45, 1960.
- Laio, F., Porporato, A., Ridolfi, L., and Rodriguez-Iturbe, I.: Plants in water-controlled ecosystems: active role in hydrologic processes and response to water stress II: probabilistic soil moisture dynamics, *Adv. Water Resour.*, 24, 707–723, 2001.
- Lee, H., Seo, D.-J., Liu, Y., Koren, V., McKee, P., and Corby, R.: Variational assimilation of streamflow into operational distributed hydrologic models: effect of spatiotemporal scale of adjustment, *Hydrol. Earth Syst. Sci.*, 16, 2233–2251, doi:10.5194/hess-16-2233-2012, 2012.
- Li, Z. J. and Navon, I. M.: Optimality of variational data assimilation and its relationship with the Kalman filter and smoother, *Q. J. Roy. Meteor. Soc.*, 127, 661–683, 2001.
- Liu, Y., Weerts, A. H., Clark, M., Hendricks Franssen, H.-J., Kumar, S., Moradkhani, H., Seo, D.-J., Schwanenberg, D., Smith, P., van Dijk, A. I. J. M., van Velzen, N., He, M., Lee, H., Noh, S. J., Rakovec, O., and Restrepo, P.: Advancing data assimilation in operational hydrologic forecasting: progresses, challenges, and emerging opportunities, *Hydrol. Earth Syst. Sci.*, 16, 3863–3887, doi:10.5194/hess-16-3863-2012, 2012.
- Lowry, C. S. and Fioren, M. N.: CrowdHydrology: crowdsourcing hydrologic data and engaging citizen scientists, *Ground Water*, 51, 151–156, 2013.

Can assimilation of crowdsourced observations improve flood prediction

M. Mazzoleni et al.

[Title Page](#)

[Abstract](#)

[Introduction](#)

[Conclusions](#)

[References](#)

[Tables](#)

[Figures](#)

[⏪](#)

[⏩](#)

[◀](#)

[▶](#)

[Back](#)

[Close](#)

[Full Screen / Esc](#)

[Printer-friendly Version](#)

[Interactive Discussion](#)



- Lu, H., Hou, T., Horton, R., Zhu, Y., Chen, X., Jia, Y., Wang, W., and Fu, X.: The streamflow estimation using the Xinanjiang rainfall runoff model and dual state-parameter estimation method, *J. Hydrol.*, 480, 102–114, 2013.
- Lundberg, A.: Combination of a conceptual model and an autoregressive error model for improving short time forecasting, *Nord. Hydrol.*, 13, 233–246, 1982.
- MacPherson, B.: Dynamic initialization by repeated insertion of data, *Q. J. Roy. Meteor. Soc.*, 117, 965–991, 1991.
- Massart, S., Pajot, B., Piacentini, A., and Pannekoucke, O.: On the merits of using a 3D-FGAT assimilation scheme with an outer loop for atmospheric situations governed by transport, *Mon. Weather Rev.*, 138, 4509–4522, 2010.
- Matheron, G.: Principles of geostatistics, *Econ. Geol.*, 58, 1246–1266, 1963.
- Mazzoleni, M., Alfonso, L., and Solomatine, D. P.: Effect of spatial distribution and quality of sensors on the assimilation of distributed streamflow observations in hydrological modeling, *Hydrolog. Sci. J.*, in review, 2015a.
- Mazzoleni, M., Alfonso, L., Chacon-Hurtado, J., and Solomatine, D. P.: Assimilating uncertain, dynamic and intermittent streamflow observations in hydrological models, *Adv. Water Res.*, 83, 323–339, 2015b.
- McLaughlin, D.: An integrated approach to hydrologic data assimilation: interpolation, smoothing and filtering, *Adv. Water Resour.*, 25, 1275–1286, 2002.
- Mendoza, P. A., McPhee, J., and Vargas, X.: Uncertainty in flood forecasting: a distributed modeling approach in a sparse data catchment, *Water Resour. Res.*, 28, 1944–7973, 2012.
- Moore, R. J., Jones, D. A., Cox, D. R., and Isham, V. S.: Design of the HYREX raingauge network, *Hydrol. Earth Syst. Sci.*, 4, 521–530, doi:10.5194/hess-4-521-2000, 2000.
- Moradkhani, H., Sorooshian, S., Gupta, H. V., and Houser, P.: Dual state-parameter estimation of hydrological models using Ensemble Kalman filter, *Adv. Water Resour.*, 28, 135–147, 2005.
- Ragnoli, E., Zhuk, S., Donncha, F. O., Suits, F., and Hartnett, M.: An optimal interpolation scheme for assimilation of HF radar current data into a numerical ocean model, *Oceans*, 1–5, doi:10.1109/OCEANS.2012.6404901, 2012.
- Rakovec, O., Weerts, A. H., Hazenberg, P., Torfs, P. J. J. F., and Uijlenhoet, R.: State updating of a distributed hydrological model with Ensemble Kalman Filtering: effects of updating frequency and observation network density on forecast accuracy, *Hydrol. Earth Syst. Sci.*, 16, 3435–3449, doi:10.5194/hess-16-3435-2012, 2012.

**Can assimilation of
crowdsourced
observations improve
flood prediction**

M. Mazzoleni et al.

[Title Page](#)[Abstract](#)[Introduction](#)[Conclusions](#)[References](#)[Tables](#)[Figures](#)[⏪](#)[⏩](#)[◀](#)[▶](#)[Back](#)[Close](#)[Full Screen / Esc](#)[Printer-friendly Version](#)[Interactive Discussion](#)

Rakovec, O., Weerts, A. H., Sumihar, J., and Uijlenhoet, R.: Operational aspects of asynchronous filtering for flood forecasting, *Hydrol. Earth Syst. Sci.*, 19, 2911–2924, doi:10.5194/hess-19-2911-2015, 2015.

Refsgard, J. C.: Validation and intercomparison of different updating procedures for real-time forecasting, *Nord. Hydrol.*, 28, 65–84, 1997.

Roy, H. E., Pocock, M. J. O., Preston, C. D., Roy, D. B., and Savage, J.: Understanding Citizen Science and Environmental Monitoring, Final Report of UK Environmental Observation Framework, NERC Centre for Ecology & Hydrology and Natural History Museum, Wallingford, 2012.

Sakov, P., Evensen, G., and Bertino, L.: Asynchronous data assimilation with the EnKF, *Tellus A*, 62, 24–29, 2010.

Seo, D. J., Kerke, B., Zink, M., Fang, N., Gao, J., and Yu, X.: iSPUW: a vision for integrated sensing and prediction of urban water for sustainable cities, in: Proceedings of Dynamic Data-driven Environmental Systems Science Conference, Cambridge, MA, 5–7 November, 2014.

Sittner, W. T. and Krouse, K. M.: Improvement of hydrologic simulation by utilizing observed discharge as an indirect input, in: NOAA Technical Memorandum NWS HYDRO-38, Silver Spring, Md. Dept. of Commerce, National Oceanic and Atmospheric Administration, National Weather Service, February, 1979.

Szilagyi, J. and Szollosi-Nagy, A.: *Recurvive Streamflow Forecasting: a State Space Approach*, CRC Press, London, UK, 2010.

Szollosi-Nagy, A., Bartha, P., and Harkayi, K.: Microcomputer based operational hydrological forecasting system for river Danube, Technical Conference on Mitigation of Natural Hazards through Real-Time Data Collection Systems and Hydrological Forecasting, Sacramento, California, 1983.

Todini, E.: A mass conservative and water storage consistent variable parameter Muskingum-Cunge approach, *Hydrol. Earth Syst. Sci.*, 11, 1645–1659, doi:10.5194/hess-11-1645-2007, 2007.

Todini, E., Alberoni, P., Butts, M., Collier, C., and Khatibi, R.: ACTIF best practice paper – understanding and reducing uncertainty in flood forecasting, in: Proceeding of ACTIF International conference on Innovation, advances and implementation of flood forecasting technology, 17–19 October, Tromsø, Norway, 2005.

Verlaan, M.: Efficient Kalman filtering algorithms for hydrodynamic models, Ph.D. thesis, Delft University of Technology, the Netherlands, 1998.

Weerts, A. H., El Serafy, G. Y. E.: Particle filtering and ensemble Kalman filtering for state updating with hydrological conceptual rainfall–runoff models, *Water Resour. Res.*, 2, W09403, doi:10.1029/2005WR004093, 2006.

WeSenseIt: Citizen Water Observatories, available at: <http://wesenseit.eu/> (last access: 24 September 2015), 2012.

Wood, S. J., Jones, D. A., and Moore, R. J.: Accuracy of rainfall measurement for scales of hydrological interest, *Hydrol. Earth Syst. Sci.*, 4, 531–543, doi:10.5194/hess-4-531-2000, 2000.

World Meteorological Organization (WMO): Simulated real-time intercomparison of hydrological models, *Operational Hydrology Rep.* 38, Geneva, 1992.

Xie, X. and Zhang, D.: Data assimilation for distributed hydrological catchment modeling via ensemble Kalman filter, *Adv. Water Resour.*, 33, 678–690, doi:10.1016/j.advwatres.2010.03.012, 2010.

Young, P. C.: *Recursive Estimation and Time Series Analysis*, Springer-Verlag, Berlin, 1984.

HESSD

12, 11371–11419, 2015

Can assimilation of crowdsourced observations improve flood prediction

M. Mazzoleni et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

⏪

⏩

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



HESSD

12, 11371–11419, 2015

Can assimilation of crowdsourced observations improve flood prediction

M. Mazzoleni et al.

Table 1. NSE values in case of different experimental scenarios during flood event 1 and 2.

Scenario	Flood event 1			Flood event 2		
	1 obs	100 obs	Improvement	1 obs	100 obs	Improvement
1	0.775	0.896	0.135	0.537	0.879	0.390
2	0.775	0.895	0.134	0.537	0.876	0.388
3	0.760	0.895	0.151	0.501	0.875	0.428
4	0.699	0.888	0.212	0.318	0.856	0.629
5	0.692	0.885	0.218	0.304	0.850	0.642
6	0.775	0.895	0.134	0.537	0.877	0.388
7	0.758	0.895	0.153	0.486	0.876	0.445
8	0.708	0.888	0.203	0.338	0.857	0.605
9	0.696	0.885	0.214	0.283	0.852	0.667

[Title Page](#)[Abstract](#)[Introduction](#)[Conclusions](#)[References](#)[Tables](#)[Figures](#)[◀](#)[▶](#)[◀](#)[▶](#)[Back](#)[Close](#)[Full Screen / Esc](#)[Printer-friendly Version](#)[Interactive Discussion](#)



Figure 1. Example of a low-cost social sensor, and crowdsourced observations, implemented in the Bacchiglione river, Italy, under the WeSenseIt project.

HESSD

12, 11371–11419, 2015

Can assimilation of crowdsourced observations improve flood prediction

M. Mazzoleni et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

⏪

⏩

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



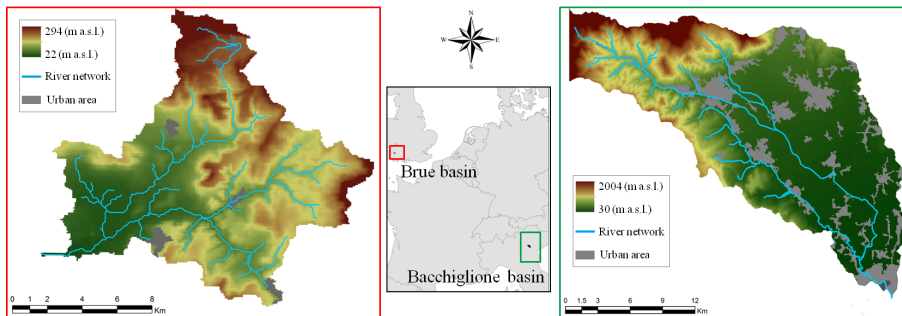


Figure 2. The two case studies considered in this study.

Can assimilation of crowdsourced observations improve flood prediction

M. Mazzoleni et al.

Title Page	
Abstract	Introduction
Conclusions	References
Tables	Figures
◀	▶
◀	▶
Back	Close
Full Screen / Esc	
Printer-friendly Version	
Interactive Discussion	



Can assimilation of crowdsourced observations improve flood prediction

M. Mazzoleni et al.

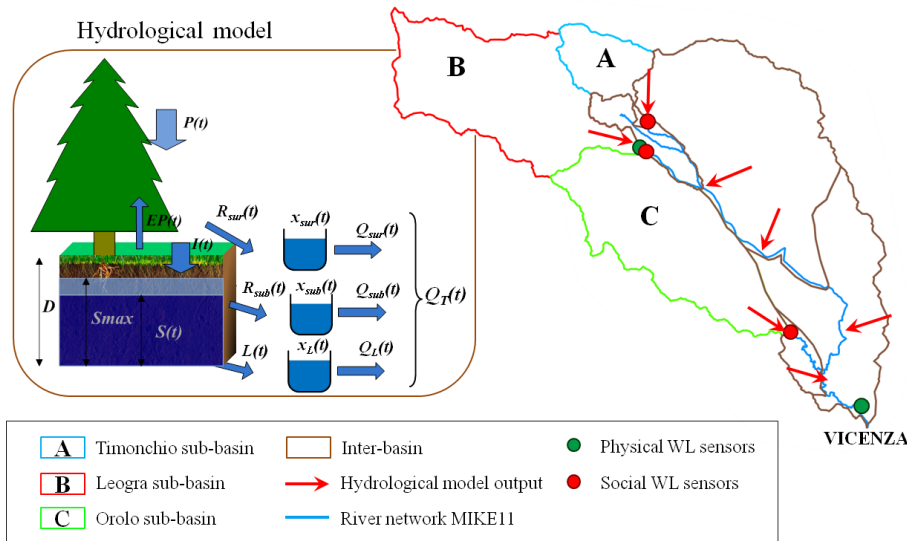


Figure 3. Structure of the early warning system AMICO and location of the physical and social sensors implemented in the Bacchiglione basin by the Alto Adriatico Water Authority.

Title Page

Abstract Introduction

Conclusions References

Tables Figures

◀ ▶

◀ ▶

Back Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



Can assimilation of crowdsourced observations improve flood prediction

M. Mazzoleni et al.

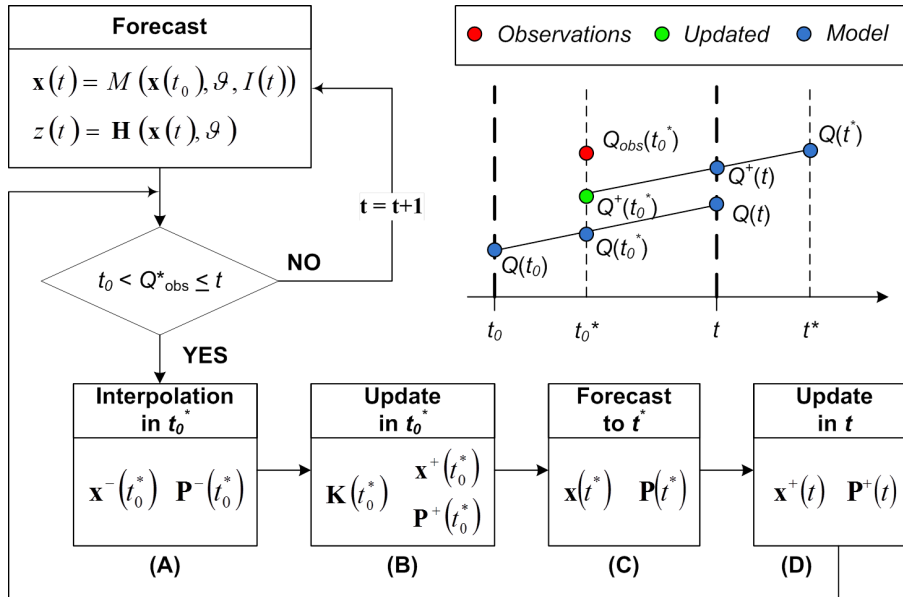


Figure 4. Graphical representation of the DACO method proposed in this study to assimilate crowdsourced observations.

[Title Page](#)

[Abstract](#)

[Introduction](#)

[Conclusions](#)

[References](#)

[Tables](#)

[Figures](#)

[⏪](#)

[⏩](#)

[◀](#)

[▶](#)

[Back](#)

[Close](#)

[Full Screen / Esc](#)

[Printer-friendly Version](#)

[Interactive Discussion](#)

Can assimilation of crowdsourced observations improve flood prediction

M. Mazzoleni et al.

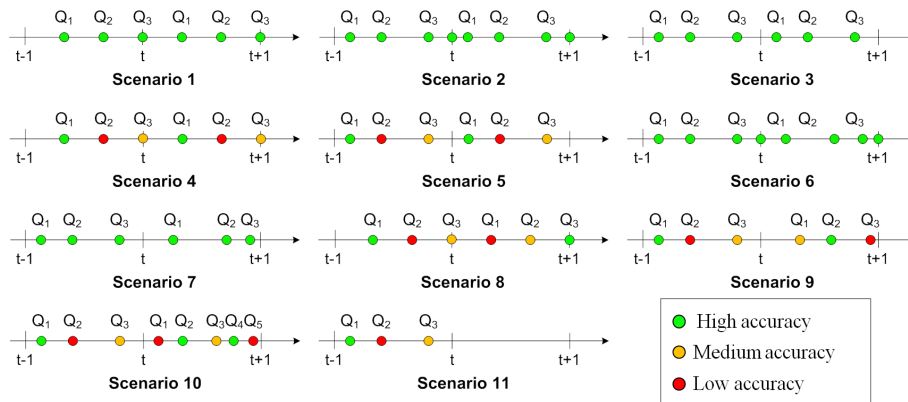


Figure 5. The experimental scenarios representing different configurations of arrival frequency, number and accuracy of the streamflow observations.

Title Page

Abstract Introduction

Conclusions References

Tables Figures

⏪ ⏩

◀ ▶

Back Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



Can assimilation of crowdsourced observations improve flood prediction

M. Mazzoleni et al.

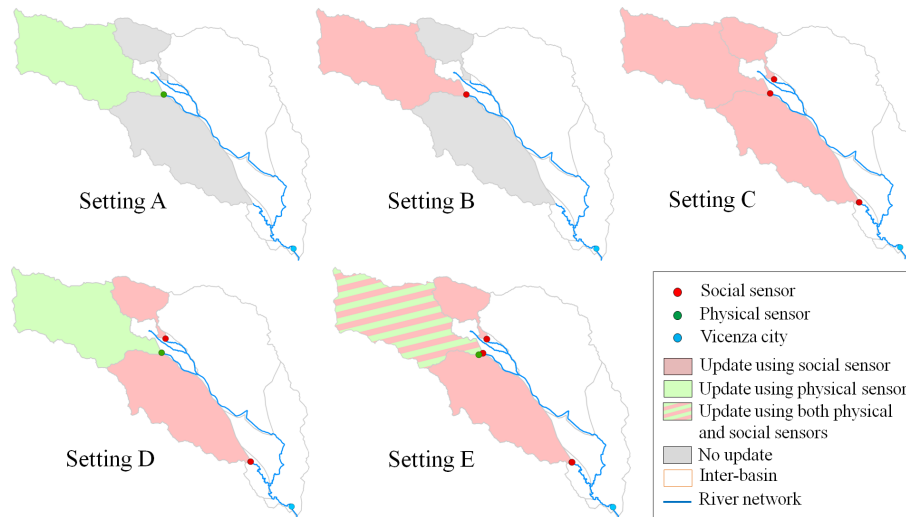


Figure 6. Different experimental settings implemented within the Bacchiglione basin during Experiments 2.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

⏪

⏩

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



Can assimilation of crowdsourced observations improve flood prediction

M. Mazzoleni et al.

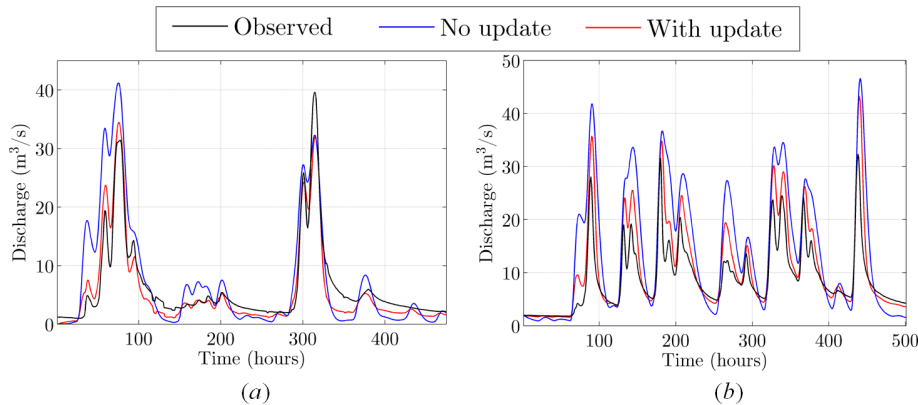


Figure 7. The observed and simulated hydrographs, with and without assimilation, for flood event 1 **(a)** and 2 **(b)** in the Brue basin.

[Title Page](#)
[Abstract](#) [Introduction](#)
[Conclusions](#) [References](#)
[Tables](#) [Figures](#)
⏪ ⏩
◀ ▶
[Back](#) [Close](#)
[Full Screen / Esc](#)
[Printer-friendly Version](#)
[Interactive Discussion](#)



Can assimilation of crowdsourced observations improve flood prediction

M. Mazzoleni et al.

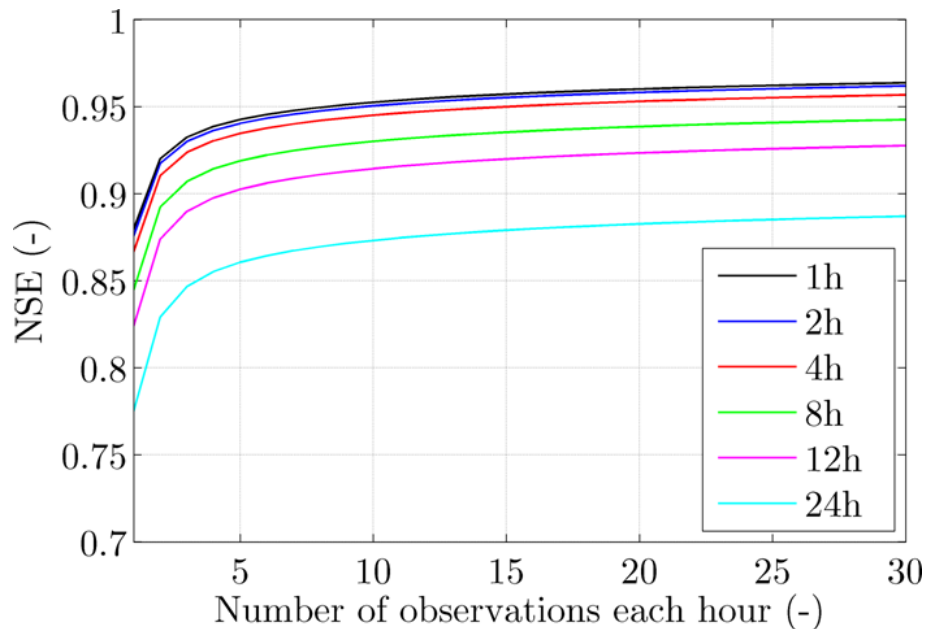


Figure 8. Model improvement during flood event 1, in case of different lead times, assimilating streamflow observations according to scenario 1.

[Title Page](#)[Abstract](#)[Introduction](#)[Conclusions](#)[References](#)[Tables](#)[Figures](#)[◀](#)[▶](#)[◀](#)[▶](#)[Back](#)[Close](#)[Full Screen / Esc](#)[Printer-friendly Version](#)[Interactive Discussion](#)

Can assimilation of crowdsourced observations improve flood prediction

M. Mazzoleni et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

⏪

⏩

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

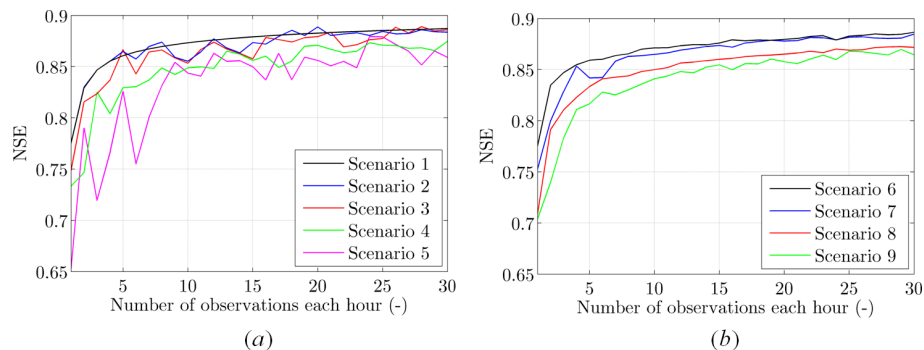


Figure 9. Model improvement during flood event 1 (lead time of 24 h), assimilating diverse values of streamflow observations according to the experimental scenarios from 1 to 9 with **(a)** observations with periodic behaviour, **(b)** observations with non-periodic behaviour.

Can assimilation of crowdsourced observations improve flood prediction

M. Mazzoleni et al.

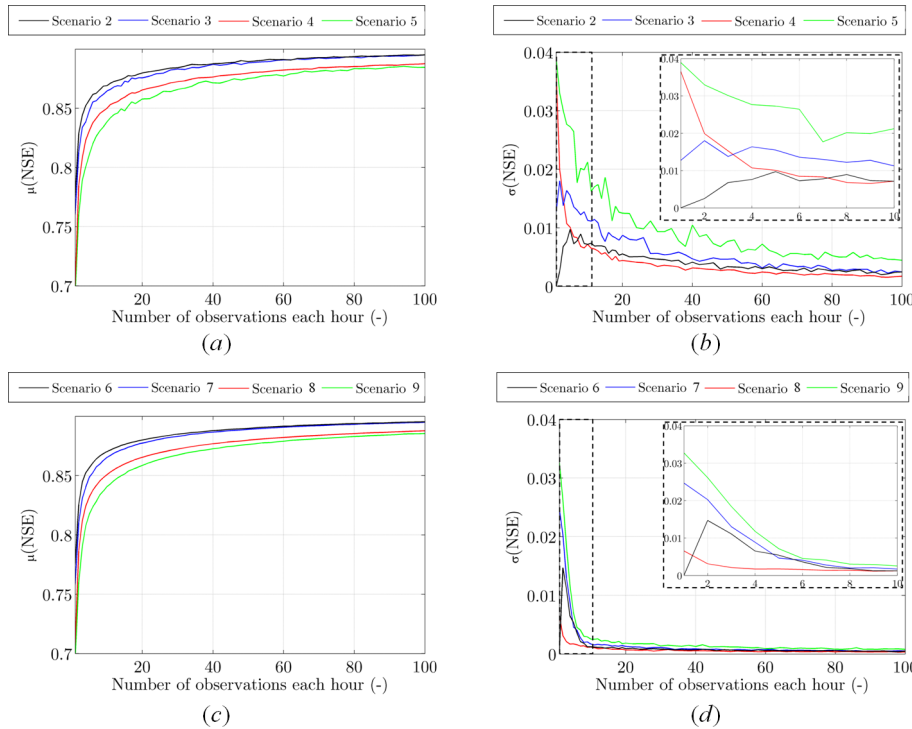


Figure 10. Dependency of $\mu(\text{NSE})$ and $\sigma(\text{NSE})$ on the number of observations, for the scenarios 2, 3, 4, 5, 6, 7, 8 and 9 in case of flood event 1.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

◀

▶

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



Can assimilation of crowdsourced observations improve flood prediction

M. Mazzoleni et al.

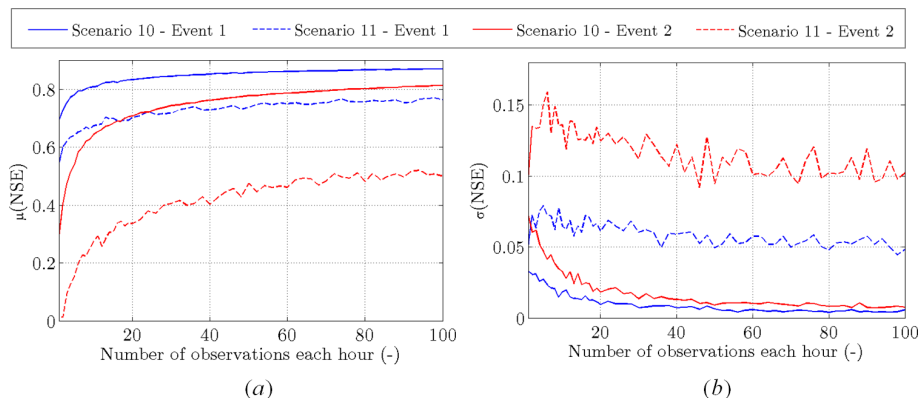


Figure 11. Dependency of the $\mu(\text{NSE})$ and $\sigma(\text{NSE})$ on the number of observations, for the scenarios 10 and 11 in case of flood events 1 (a) and 2 (b).

[Title Page](#)[Abstract](#)[Introduction](#)[Conclusions](#)[References](#)[Tables](#)[Figures](#)[⏪](#)[⏩](#)[◀](#)[▶](#)[Back](#)[Close](#)[Full Screen / Esc](#)[Printer-friendly Version](#)[Interactive Discussion](#)

Can assimilation of crowdsourced observations improve flood prediction

M. Mazzoleni et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

◀

▶

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

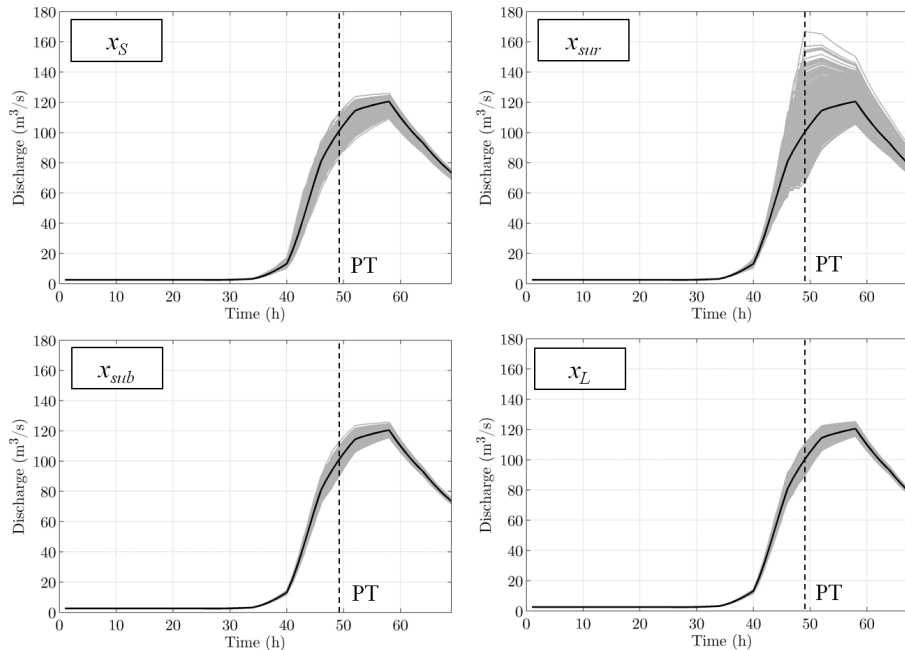


Figure 12. Effect of perturbing the model states on the model output, Bacchiglione case study. PT = Perturbation Time.

Can assimilation of crowdsourced observations improve flood prediction

M. Mazzoleni et al.

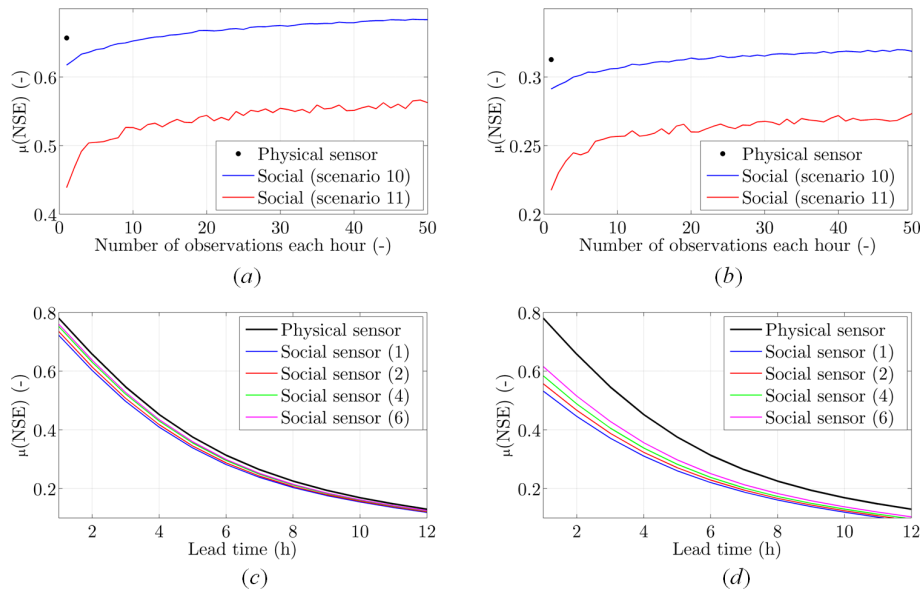


Figure 13. Model performance expressed as $\mu(\text{NSE})$ values – assimilating observations from physical and social sensors at Leogra gauged station having characteristic described in scenarios 10 (c) and 11 (d).

Title Page

Abstract Introduction

Conclusions References

Tables Figures

◀ ▶

◀ ▶

Back Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



Can assimilation of crowdsourced observations improve flood prediction

M. Mazzoleni et al.

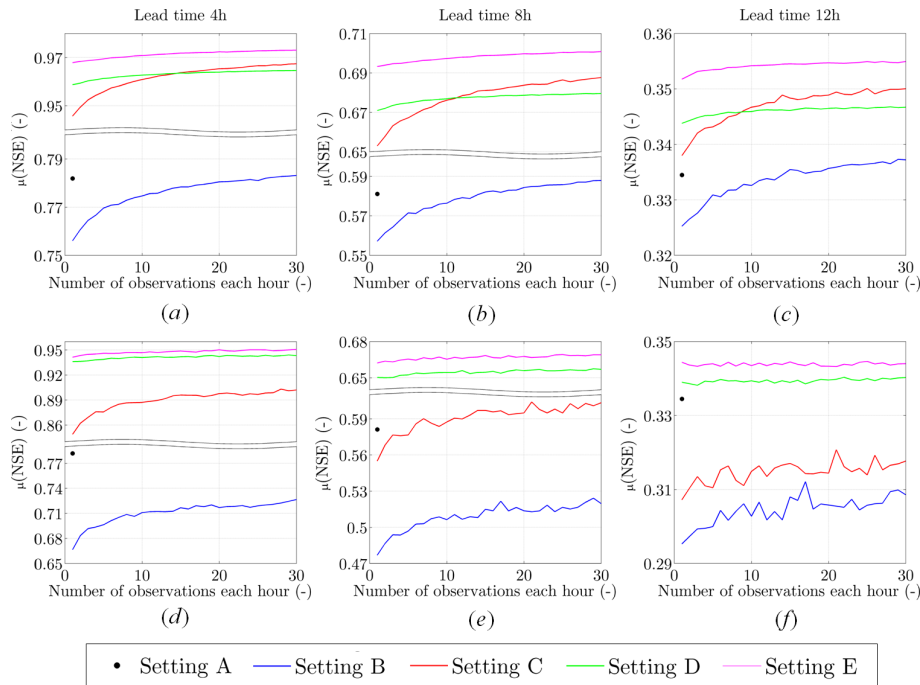


Figure 14. Model performance expressed as $\mu(\text{NSE})$ – assimilating different number of crowdsourced observations, for the three lead time values, having characteristic of scenario 10 (first row) and 11 (second row).

[Title Page](#)

[Abstract](#) | [Introduction](#)

[Conclusions](#) | [References](#)

[Tables](#) | [Figures](#)

[⏪](#) | [⏩](#)

[◀](#) | [▶](#)

[Back](#) | [Close](#)

[Full Screen / Esc](#)

[Printer-friendly Version](#)

[Interactive Discussion](#)



Can assimilation of crowdsourced observations improve flood prediction

M. Mazzoleni et al.

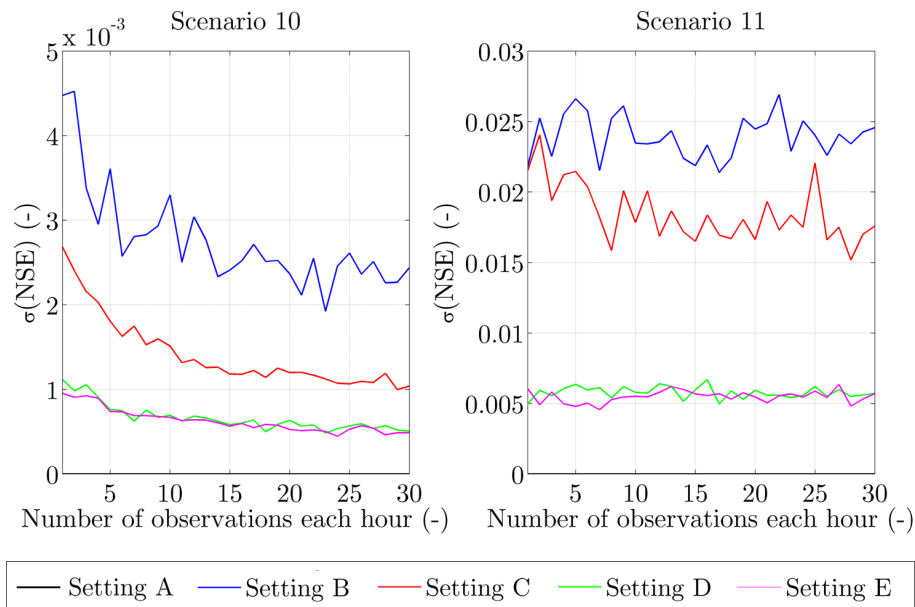


Figure 15. Variability of performance expressed as $\sigma(\text{NSE})$ – assimilating crowdsourced observations within setting A, B, C and D, assuming the lead time of 4 h, for scenarios 10 and 11.