

Estimation of lateral inflows using data assimilation in the context of real-time flood forecasting for the Marne catchment in France

Estimation des apports latéraux via assimilation de données pour le bassin versant de la Marne dans un contexte opérationnel de prévision des crues.

Johan Habert¹

DREAL Champagne-Ardenne, Chalons-en-Champagne, France
johan.habert@developpement-durable.gouv.fr

Sophie Ricci, Andrea Piacentini, Gabriel Jonville

URA 1875/CERFACS, Toulouse, France
ricci@cerfacs.fr, piacentini@cerfacs.fr, jonville@cerfacs.fr

Etienne Le Pape

SCHAPI, Toulouse, France
Etienne.lepape@developpement-durable.gouv.fr

Olivier Thual

URA1875/CERFACS and INPT, CNRS, IMFT, Toulouse, France
thual@imft.fr

Nicole Goutal, Fabrice Zaoui, Riadh Ata

LNHE EDF R&D, Chatou, France
nicole.goutal@edf.fr, fabrice.zaoui@edf.fr, riadh.ata@edf.fr

KEY WORDS

Flood forecasting, Data assimilation, Kalman filter, Karstic catchment

Abstract

The present study describes the assimilation of discharge in-situ data for operational flood forecasting. The study was carried out on the Marne river (France) catchment where lateral inflows's uncertainty are important due to karstic areas. This source of error was partly accounted for using an Extended Kalman Filter algorithm built on top on a monodimensional hydraulic model. The lateral inflows were sequentially adjusted over a sliding 48-hour time window. The correction leads to a significant improvement in the simulated water level and discharge in re-analysis and forecast modes. These results pave the way for the operational use of the data assimilation procedure for real-time forecasting at the French flood forecasting service.

1. INTRODUCTION

In 2006, 9% of the French population was exposed to flood risk, one of the greatest natural risks causing damage and human loss [12]. The French flood forecasting service (SCHAPI - Service Central d'Hydrométéorologie et d'Appui à la Prévision des Inondations), in collaboration with the 22 local flood forecasting centers (SPC - Service de Prévision des Crues) produces a twice-daily broadcast vigilance map available to governmental authorities and general public (<http://www.vigicrues.gouv.fr>).

The reliability of flood forecasting strongly depends on the quality of the hydraulics model, its boundary conditions (upstream and lateral inflow), hydrological initial conditions and numerical parameters. Much effort has been directed towards the estimation of hydrologic and hydraulics model parameters especially for

¹Corresponding author

the statistical analysis of parameters uncertainties [3][13] usually using a historical batch of data, assuming time-invariant parameters. In practice, in addition to model simulation and batch calibration, the reliable operation of a watershed system requires a continuous correction of the forecast as observational data become available. This entails the critical need to extend the use of data assimilation (DA) in hydraulics and hydrology. DA offers a convenient framework to overcome some of the limits of the calibration processes. Observations and simulation outputs are combined to estimate an optimal set of model parameters [1][9] and consequently reduce uncertainties in the simulation. The Kalman Filter (KF) [5] algorithm is the most commonly used sequential DA algorithm which results in the optimal estimation for linear dynamic models with Gaussian uncertainties. It was extended to non linear problems using a first order approximation of Taylor series, namely Extended Kalman Filter (EKF), or an ensemble approach for the Ensemble Kalman Filter (EnKF) [4]. These algorithms are now widely being used in hydrology and hydraulics for the estimation of model states [7][8] as well as model parameters [10].

In the context of hydrodynamics modeling, Ricci et al (2011) [11] showed that the EKF assimilation of water level observations on the Adour catchment with the 1D hydraulic model MASCARET [6] developed by LNHE (Laboratoire National d'Hydraulique et d'environnement) from EDF R&D (Electricité De France Recherche et Développement) enabled to improve flood forecasting of 60% at a one hour lead time and of 25% at a twelve-hour lead time. In the present study, a similar approach is applied to flood forecasting in the Marne catchment where the presence of karstic areas makes it hard to correctly specify upstream and lateral inflows to the model. In the framework of operational flood forecasting, the SAMA (Seine Amont Marne Amont) SPC has developed two different models on two limited areas of the Marne catchment described in Figure 1. On these limited areas, the batch calibration of the model was possible and the uncertainties due to lateral inflow were accounted for by artificially adjusting the Strickler coefficients. Still, in order to increase the forecast time on the Marne catchment, both models were recently merged into a global model far more difficult to calibrate. The need for a coherent estimation of the so far neglected inflows, which represent the dynamics of the karstic areas, and the unmodeled tributaries, which represent the dynamics of catchment areas, motivates the use of a DA procedure using in situ measurements. A realistic and time varying estimation of the lateral inflows is then achieved using a sequential data assimilation approach on 10 flood events over 2001-2010. It is shown that, in spite of certain limitations described further on, this approach provides a reliable estimate of the lateral inflows and leads to the improvement of the flood forecast at meaningful lead times for operational use.

The structure of the paper is as follows. Section 2 provides a description of the Marne catchment. It also provides the arguments for building a DA scheme to estimate lateral inflows. The DA method is described in Section 3 along with the choices made for the implementation of the algorithm. Section 4 gives an overview of the results, highlighting the merits of the approach for flood forecasting along with its limitations on a representative flood event. Conclusions are given in Section 5.

2. MODELING OF THE MARNE CATCHMENT

The Marne catchment is a karstic basin located East of the Paris basin. The Marne river is a 525 km long tributary of the Seine river, its source is located in the Langres Plateau in the Haute-Marne department (Figure 1). The study is carried out on the upstream part of the river that is strongly sensitive to local precipitation and where flash floods (5 to $120 \text{ m}^3 \cdot \text{s}^{-1}$ within a 24h period in October 2006 at Condes) occur. The landscape of the catchment is defined by forested plateaus, incised valleys and presents numerous limestone outcrops. Therefore the catchment area includes karstic areas and tributaries whose behaviour is highly non linear and thus difficult to forecast.

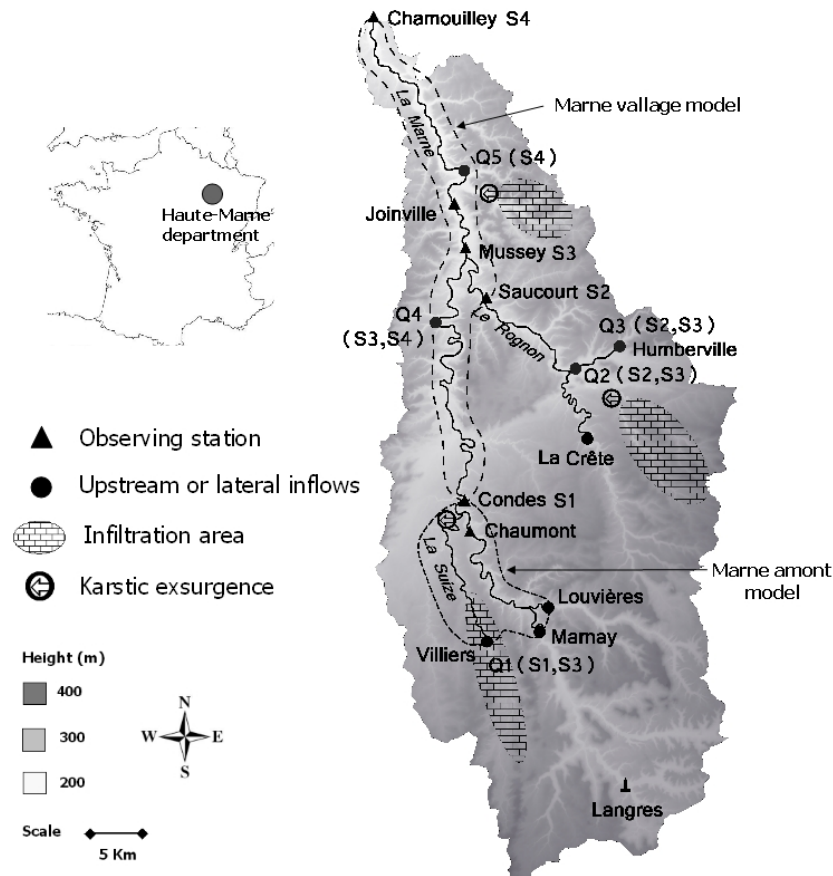


Figure 1: The Marne catchment, the sub-models are circled with dashed lines. The hydrological observing stations S_1 , S_2 , S_3 and S_4 are represented by black triangles. Lateral inflows Q_1 , Q_2 , Q_3 , Q_4 and Q_5 are represented by black dots and the observing stations on which they depend are in parentheses.

As of today, the operational forecast relies on the integration of two sub-models on the Marne Amont and Marne Vallage areas (Figure 1). These models are based on the mono-dimensional numerical code MASCARET for hydraulics describing the Saint-Venant (Shallow Water) equations and developed by EDF and CETMEF (Centre d'Etudes Techniques Maritimes Et Fluviales). They were calibrated independently using water level in-situ measurements from five hydrological observing stations (Chaumont, Condes, Mussey, Joinville and Chamouilley). The underestimation of the lateral inflow that leads to the underestimation of the simulated water levels and discharge was partly corrected by reducing the Strickler coefficients and thus artificially increasing the water level. These sub-models are used for operational forecast at SAMA SPC since 2008. They provide a satisfying water level signal, still discharges are usually underestimated and the maximum lead time for the forecast is 15 hours.

In order to extend the maximum lead time and benefit from measurements at Saucourt, the sub-models were merged into a global model extending from Villiers to Chamouilley and including the karstic areas of the Rognon. The upstream flows are specified at five upstream stations (Marnay, Louvières, Villiers, La Crête and Humberville). This global model underestimates the discharges of 50% on average over ten significant events as presented in Figure 2 for a representative event for Condes and Mussey observing stations (observations are in blue and simulations, denoted by Free Run, in black). Indeed, the global model area is about 2250 km² when the area controlled by the five upstream stations is only about 755 km². Thus it appears that the modeling of lateral inflows, despite the lack of hydrologic rainfall-runoff model on the area, represents a key step towards the use of the global model for the Marne catchment. Five lateral inflows were then added to the model to represent the exsurgences of the Suize (Q_1), the Seurre on the Rognon catchment (Q_2 and Q_3), and tributaries upstream Musey (Q_4) and Chamouilley (Q_5).

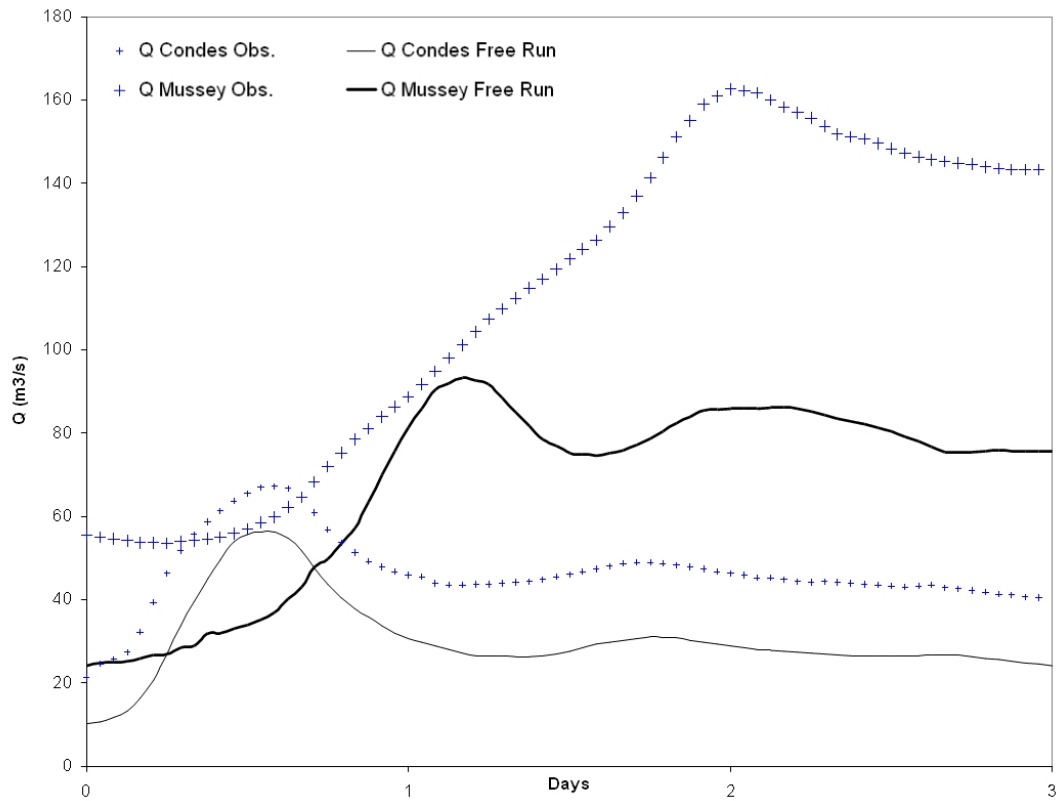


Figure 2: Observed (blue crossed curves) and simulated discharges for Free Run (black solid curves) in Condes (thin curves) and Mussey (thick curves). March 2008 event.

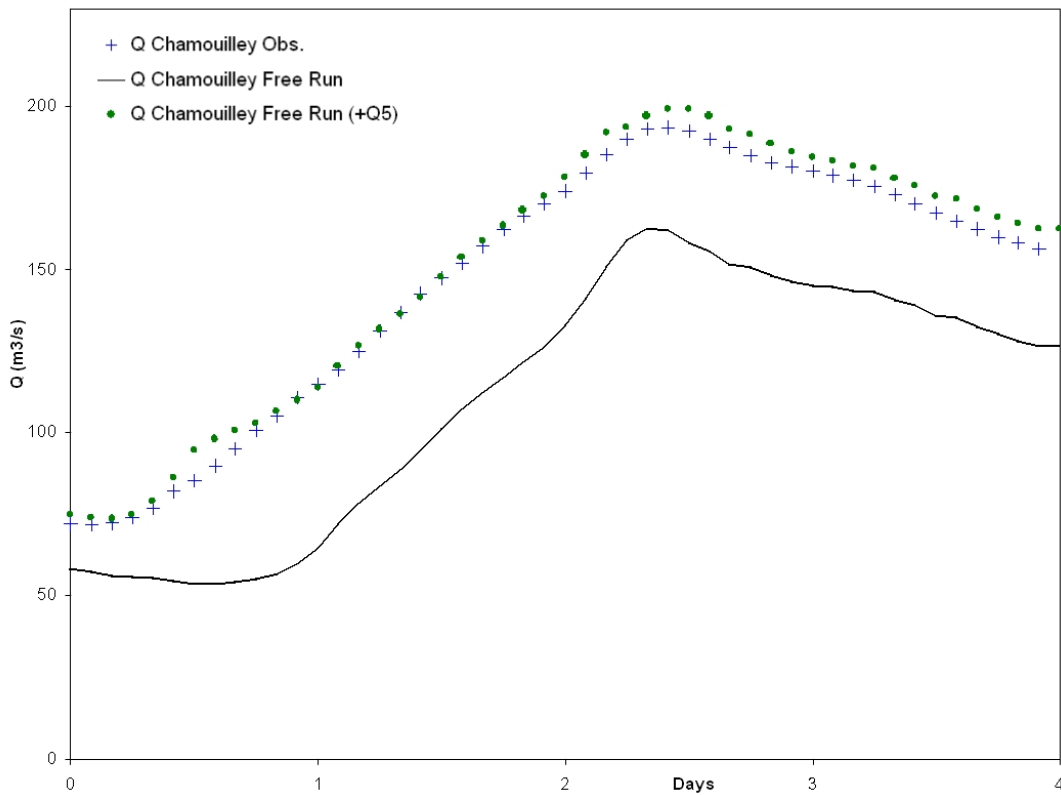


Figure 3: Discharge at Chamouilley, observation (blue crossed curve), Free Run (black solid curve) and Free Run with Q_5 (green dotted curve). March 2008 event.

Given the homogeneous response of the catchment to an oceanic rainfall event, a water budget approach enables to describe a coherent, yet perfectible, behavior of the catchment. For instance, adding the lateral inflow Q_5 estimated as eight times bigger than the hydrograph at Villiers leads to a satisfying simulation of the discharge at Chamouilley as illustrated in Figure 3 where observation is represented by the blue crossed curve, the Free Run simulation is in black and the Free Run simulation with the addition of the lateral inflow Q_5 is represented in green by the dotted curve. Similar approach using Villiers and Humberville as elementary hydrographs to estimate each lateral inflow was used to identify multiplicative coefficients A_i , with $i \in [1, 5]$ and their statistics, from a batch calibration over the 10 flood events. These values presented in Table 1 can be greatly improved with a sequential method that allows for temporal variability of the coefficients, essential for instance for summer and early autumn events when the karst behavior is complex and the Villiers station may not be representative of the entire catchment's dynamics.

Q_i	Elementary hydrograph	A_i	Std
Q_1	Villiers	3	0.8
Q_2	Villiers	4	1.3
Q_3	Humberville	3	1.8
Q_4	Villiers	2.5	1.4
Q_5	Villiers	5.5	2

Table 1: Multiplicative corrective coefficients A_i for lateral inflows and standard deviations with $i \in [1, 5]$.

3. SEQUENTIAL DATA ASSIMILATION METHOD

Data assimilation approaches try to identify the optimal estimate of the true value of an unknown variable \mathbf{x} that includes, in this work, the set of corrective coefficients A_i with $i \in [1, 5]$. The a priori knowledge on these coefficients given in Table 1 describes the background vector \mathbf{x}^b . The observation vector \mathbf{y}^o is composed of hourly discharge measurements, a conservative variable, at Condes, Saucourt, Mussey and Chamouilley (respectively denoted by S_1 , S_2 , S_3 and S_4 in Figure 1). The analysis is performed on a sliding time window, over which the A_i coefficients are assumed to be constant. Assuming that the background, the observation and the analysis are unbiased, the analysis vector \mathbf{x}^a for cycle k can be formulated as a correction to the background parameters:

$$\mathbf{x}_k^a = \mathbf{x}_k^b + \mathbf{K}_k (\mathbf{y}_k^o - H(\mathbf{x}_k^b))$$

where $\mathbf{K}_k = \mathbf{B}_k \mathbf{H}_k^T (\mathbf{H}_k \mathbf{B}_k \mathbf{H}_k^T + \mathbf{R}_k^{-1})$ is the gain matrix, \mathbf{B}_k , \mathbf{R}_k are the background and observation errors, $\mathbf{y}_k = H(\mathbf{x}_k)$ is the model equivalent of the observations, generated by the observation operator H and its Jacobian \mathbf{H}_k is the tangent linear of the hydraulics model computed in the vicinity of \mathbf{x}_k^b , which can be approximated using a finite differences scheme. The observation operator consists of two operations, the more costly of which is the non-linear integration of the hydraulics model given the upstream and lateral flow conditions over the assimilation window. The second operation is the selection of the calculated discharges at the observation points and at the observation times. $H(\mathbf{x}_k^b)$ represents the discharges at the observation points and times computed by MASCARET using the background parameters and its Jacobian. The local estimation of the tangent linear dynamics of the model with respect to the boundary conditions of the domain is a strong hypothesis. This method can be seen as an EKF algorithm without model errors (the model is considered as perfect). Since there is no propagation model for the parameters, the usual propagation steps of the KF algorithm are irrelevant here: the background error covariance matrix is invariant between the cycles.

The analysis is cycled over the period covering the entire flood event, thus allowing A_i to vary between the cycles. For cycle k , the observations over the first 8 hours (the re-analysis period) are used to estimate the optimal coefficients and a 24-hour forecast is carried out. Each of the five lateral inflows Q_1 , Q_2 , Q_3 , Q_4 , Q_5 is controlled using downstream measurements from the hydrological observing station as described in Figure 1; Q_1 is controlled by Condes and Mussey, Q_2 and Q_3 are controlled by Saucourt and Mussey, Q_4 is controlled by Mussey and Chamouilley and Q_5 is controlled by Chamouilley only. A $5 \text{ m}^3 \cdot \text{s}^{-1}$ standard deviation error is assumed on the discharge measurements to account for errors in misadjustment of pressure tube and extrapolation of water level-discharge rating curves. The background error covariance is described

by a diagonal matrix with the same standard deviation error on every A_i estimated as the mean of those described in Table 1. The data assimilation algorithm was implemented using the Open-PALM dynamic coupler developed at CERFACS and designed to couple independent code components with a high level of modularity in the data exchanges while providing a straightforward parallelization environment [2].

4. RESULTS

The benefits from the application of the sequential analysis over the 10 floods events for the Marne catchment are summarized in Table 2 presenting the Nash-Sutcliffe criteria computed with data assimilation (DA) and without (Free Run) at the maximum lead time for the forecast at each observing station. It was shown that the assimilation of discharge measurements allows for a significant improvement of the simulated discharges in re-analysis (not shown) and forecast mode; with an the average improvement of 0.91 at the maximum lead time.

Observing station	Condes	Mussey	Joinville	Chamouilley	Saucourt
Forecast lead time	+6h	+12h	+13h	+21h	+10h
Free Run	0.61	0.01	0.14	-1.38	-0.20
DA	0.87	0.78	0.55	0.47	0.80
Improvement	0.26	0.77	0.69	1.85	1.

Table 2: Nash-Sutcliffe criteria for Free Run and DA simulations averaged over 10 flood events over 2001-2010 at maximum lead time for each observing station

For the major flood event in December 2010, the use of the DA procedure for real-time forecast would have improved both discharge peak forecasts of 15 % as illustrated in Figure 4 (only the first peak is shown). The discharges are presented on the left vertical axis and water levels are on the right vertical axis with solid thin and thick lines, respectively. Observations are represented by blue crossed curves, the Free Run 12 hour forecast by black solid curves and the analysis 12 hour forecast by red dashed curves. For the discharge and water level, the area between observation and Free Run is shaded in red when the model underestimates the signal and in blue when it overestimates the signal. The corrective coefficients for Q_1 , Q_2 , Q_3 and Q_4 represented in Figure 5 are globally smaller than 1 when the Free Run overestimates the observed discharge and bigger than 1 where the Free Run underestimates the observed discharge. These values also depend on the sensitivity of the discharge at the observing station with respect to each A_i ; this information is accounted for within the linearized observation operator \mathbf{H} .

Still, the assumption of a constant correction of the lateral inflows over a DA analysis cycle can lead to an inappropriate correction as displayed in Figure 4 for Day 3. The 12-hour forecasted discharge at Day 3, resulting from the DA procedure ($175 \text{ m}^3 \cdot \text{s}^{-1}$ where the solid vertical line intersects the thick red dotted curve) is computed using the corrected inflows resulting from the assimilation of the difference between the Free Run and the observation during the 8-hour period in Figure 4 represented in grey between the two thin vertical lines at Day 2 + 4h and Day 2 + 12h. Over this period, the average discharge difference is bigger than that at Day 3, thus the DA procedure leads to an over correction of the discharge (an over decrease in this case). To sum it up, if the model-observation error is not monotonous over the re-analysis and forecast period, the DA procedure can lead to an under or over correction. A possible leverage for this problem is to shorten the re-analysis period and thus allow for more temporal variability of the corrective coefficient. Globally in Figure 4, the water level state is also significantly improved, still the correction of the lateral inflows with assimilation of the discharge measurements does not improve the water level peak simulation, it even degrades the first peak. Indeed, when the relation between water level and discharge in the model is not coherent with the relation between water level and discharge in the observation, the sign of model-observation error on discharge is different from the sign of the model-observation error on water level. This is the case from Day 2.5 to Day 3.5; the Free Run simulation overestimates the discharge (blue area in Figure 3) but underestimates the water level (red area in Figure 4). Here, the DA correction tends to decrease the lateral inflow in order to decrease the simulated discharge, thus leading to the decrease of the simulated water level when the latter was already too small. In order to improve the model rating curve at Mussey, a local correction of the K_s coefficient is applied and enables to simultaneously improve water level and

discharge (green dotted curves in Figure 4). Based on these results, ongoing work aims at including the Ks coefficient within the DA control vector.

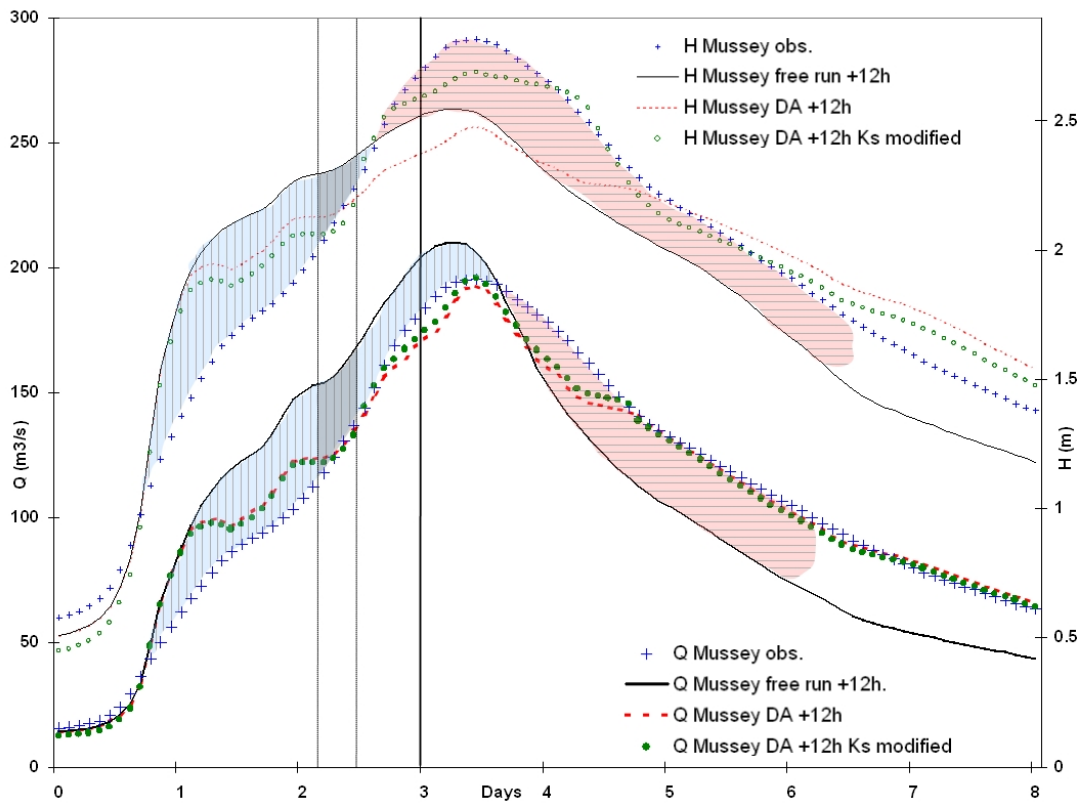


Figure 4: Discharge (thick curves) and water level (thin curves) at +12h, December 2010 (Mussey): observations (blue crossed curves), Free Run (black solid curves), DA analysis (red dashed curves) and DA analysis with modified Strickler coefficient (green dotted curves).

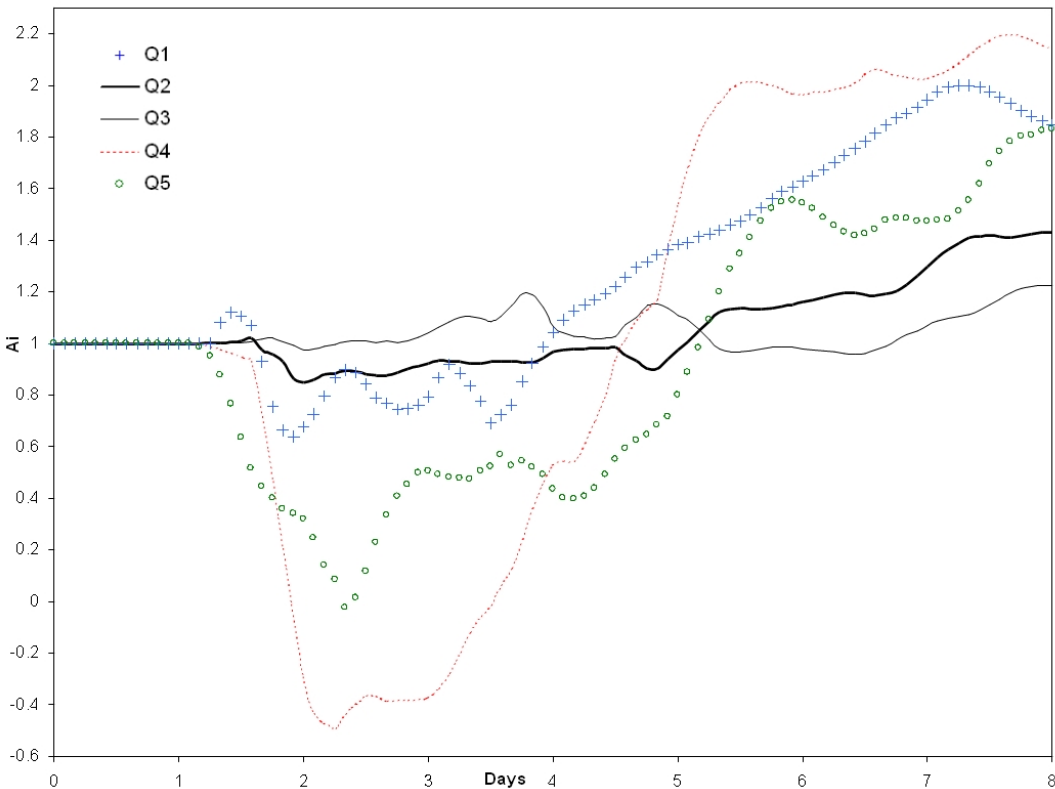


Figure 5: A_i coefficients for Q_1 , Q_2 , Q_3 and Q_4 during December 2012 event.

5. CONCLUSION

The sequential assimilation of discharge measurements in real time mode was presented in this paper. The study focuses on the application of an Extended Kalman Filter (EKF) algorithm for the Marne catchment under the assumption that the relation between the lateral boundary conditions of the domain and the simulated discharge is fairly linear. It was shown that the estimation of the time varying contribution of karstic areas and neglected tributaries can be achieved. This leads to the improvement of the forecast at meaningful lead time for operational use. Since the method developed here is not catchment dependent, it was applied for other French basins and it is currently being integrated in the real-time forecasting platform for operational use at SCHAPI. The reduced computational cost of the procedure is also a strong advantage. The extension of the control vector to model parameters such as the Strickler coefficients or the bathymetry is one of the perspective for further works as it will allow to correct the relation between water level and discharge within the model. Additionally, theoretical work should be pursued to relax the linear hypothesis on the model using, for instance, an ensemble approach.

REFERENCES

- [1] F. Bouttier and P. Courtier: Data assimilation concepts and methods, *ECMWF Lecture Note*, 1999.
- [2] S. Buis, A. Piacentini, and D. Déclat: PALM: A Computational framework for assembling high performance computing applications, *Concurrency Computat.: Pract. Exper.*, 18(2), 247-262, 2006.
- [3] Q. Duan, V. Gupta, S. Sorooshian, A. N. Rousseau, R. Turcotte: Preface, in *Calibration of watershed Models, Water Sci. Appl. Ser.*, vol. 6; edited by Q. Duan et al., p. v, AGU, Washington, D. C, 2003.
- [4] G. Evensen : Sequential data assimilation with a non linear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *J. Geophys. Res.*, 99:10143-62, 1994.
- [5] A. Gelb: Applied optimal estimation, *Cambridge Mass.: MIT Press*, 1974.
- [6] N.Goutal and F.Maurel: A finite volume solver for 1D shallow water equations applied in an actual river, *Int. J. Numer. Meth. Fluids*, 38(2), 1-19, 2002.
- [7] N. Jean-Baptiste, P-O.Malaterre, C. Dorée, J. Sau: Data assimilation for real-time estimation of hydraulics states and unmeasured perturbations in a 1d hydrodynamic model. *Math. and Computers in Simulation*, 81, 2201-2214, 2009.
- [8] P-O. Malaterre, J-P. Baume and N. Jean-Baptiste: Calibration of open channel flow models: a system analysis and control engineering approach, *SimHydro 2010*.
- [9] D. Maclaughlin and L. Townley: A reassessment of the groundwater inverse problem, *Water Resour. Res.*, 32, 1131-1161, 1996.
- [10] H. Moradkhani, S. Sorooshian, H. Gupta, P. Houser: Dual-state parameters estimation of hydrological models using ensemble Kalman filter. *Adv. in Water Res.*, 28, 135-147, 2005.
- [11] S. Ricci, A. Piacentini, O. Thual, E. Le Pape, G. Jonville: Correction of upstream flow and hydraulics state with data assimilation in the context of flood forecasting, *Hydrol. Earth Syst. Sci*, 15, 1-21, 2011.
- [12] CGDD-SOeS (Commissariat Général au Développement Durable - Service de l'Observation et des Statistiques), Croissance du nombre de logements en zones inondables, *Le point sur*, 6, Février 2009.
- [13] J.A. Vrugt, H. V. Gupta, W. Bouten, S. Sorooshian: A Shuffled complex evolution metropolis algorithm for optimization and uncertainty assessment of hydrological parameters. *Water Resour. Res.*, 39(8), 1-14, 2003.