8-1-2014

Data Assimilation And Multidimensional Model Coupling On The Adour Catchment, South-West Of France

Sébastien Barthélémy

Sophie Ricci

Nicole Goutal

Etienne Le Pape

Olivier Thual

11th International Conference on Hydroinformatics
HIC 2014, New York City, USA

DATA ASSIMILATION WITH AN ENSEMBLE KALMAN FILTER ALGORITHM ON AN OPERATIONAL HYDRAULIC NETWORK

S. BARTHELEMY (1), S. RICCI (1), N. GOUTAL (2), O. THUAL (3), E. LEPAPE (4)

(1): URA 1875 CNRS/CERFACS, Toulouse, France

(2): EDF R&D, LNHE, Toulouse, France

(3):URA 1875 CNRS/CERFACS and INPT, CNRS, IMFT, Toulouse France

(4) : SCHAPI, Toulouse, France

This study describes the implementation and the merits of an Ensemble Kalman Filter algorithm (EnKF) on the 1D-shallow water model MASCARET for the representation of the hydrodynamics of the "Adour maritime" river in south west France. The data assimilation algorithm is validated in the framework of Observating System Simulation Experiment; it is shown that the assimilation of in-situ water level observations allows improving water level and discharge over the entire hydraulic network, where no data are available. Finally, the method is applied in the context of real data experiments for recent major flood events of the Adour catchment. The algorithm provides a corrected hydraulic state that can be used as an initial condition for further forecast as well as an input for 1D/2D model coupling.

INTRODUCTION

The hydraulic observing network in France provides water level data at relatively high frequency (usually hourly and up to 5 minutes) but with an inhomogenous spatial repartition at 1500 stations over the 21 000 km that are under the Service Central d'Hydrométéorologie et d'appui à la Prévision des Inondations (SCHAPI) supervision. These data are used for calibration and validation of hydraulic models that are then used for real-time forecast and flood risk evaluation. In spite of recent advances in numerical model and observing system developments, hydraulic simulation and forecast are limited by uncertainties in the description of topography, bathymetry, friction coefficients, rating curves and measurements. In order to overcome these limitations and allows for a reliable description of flood plains, Data Assimilation (DA) combines information from the numerical model with observations, taking into account errors on both model and observation, thus reducing the range of uncertainty of the model outputs.

It has already been demonstrated that DA can make an important contribution in the context of hydraulic (Jean-Baptiste *et al*. [6]), still the application of DA for flood forecasting is not operational yet and should be further investigated. In the present study, the key point is to demonstrate how DA allows to correct the entire hydraulic state (water level and discharge) using only water level observations located at a limited number of observing stations on the network. Consequently, the corrected hydraulic state can be used for flood plain description, initial condition for further forecast and also as boundary condition for 1D/2D coupling.

Within the DA algorithm, the information available at the observing stations is spread to the entire hydraulic network as well as to the unobserved variables thanks to the background error covariance functions. The description of these statistics is a major field of research in meteorology or oceanography (Weaver and Courtier [10]). It was also discussed in the context of hydraulic by (Ricci *et al*. [9]).

While the classical Kalman Filter algorithm (Kalman and Bucy [7]) requires the formulation of the tangent-linear and adjoint codes for the hydraulic model to properly evolve the background error covariance matrix, the EnKF algorithm stochastically estimates these statistics among an ensemble of members, i.e, a sample of integrations of the hydraulic model that represents the uncertainty in the model state. Here, the assumption is made that the uncertainty is mostly due to errors in the upstream forcing. In order to enlarge the spread of uncertainty within the ensemble that tends to be under-dispersive, an inflation method based on an *a posteriori* diagnostic is implemented (Desroziers *et al*. [3]). The resulting algorithm is denoted by IEnKF (for Inflated Ensemble Kalman Filter).

The IEnKF is applied to the Adour Maritime catchment. The river width, bathymetry and slope vary along the river that is approximated by a 1D flow. The outline of the paper is as follows: Section 2 describes the numerical model MASCARET and the Adour Maritime hydraulic network. Section 3 presents the IEnKF algorithm and its implementation. Section 4 presents the data assimilation results for OSSE and real data experiments. Some conclusive remarks are finally given in Section 5.

MODELING OF THE ADOUR MARITIM CATCHMENT

1D Hydraulic model MASCARET

MASCARET is a component of the open-source integrated suite of solvers TELEMAC-MASCARET for use in the field of free-surface flow that solves the Reynolds Averaged Navier-Stokes equations. TELEMAC-MASCARET is managed by a consortium of core organizations and used for dimensioning and impact studies. MASCARET is mainly developped by EDF and CETMEF (Goutal and Maurel [5]), it solves the conservative form of 1D shallow-water equations:

$$\begin{cases} \frac{\partial S}{\partial t} + \frac{\partial Q}{\partial x} = q_a \\ \frac{\partial Q}{\partial t} + \frac{\partial (Q^2/S)}{\partial x} + gS\frac{\partial Z}{\partial x} = -\frac{gQ^2}{SK_s^2 R_H^{4/3}}, \end{cases} \quad (1)$$

where S is the wetted area $(m^2.s^{-1})$, Q is the discharge $(m^3.s^{-1})$, q_a is the lateral inflow per meter $(m^2.s^{-1})$, g stands for the gravity $(m.s^{-2})$, Z is the free surface height (m), K_s is the strickler coefficient $(m^{1/3}.s^{-1})$ and R_H is the hydraulic radius (m).

The "Adour maritime" hydraulic network

The Adour maritime hydraulic network covers 160,98 km, it is composed of 7 reaches with 3 confluences and 3 dams located on reaches 3, 6 and 7 (see Figure 1). The entire network is under tidal influence except upstream of the dams. The upstream forcings are described by observed water level converted into discharges by rating curves at the stations of Dax, Orthez, Escos and Cambo. Since the rating curves are build from a limited number of water level and discharge measurements and are usually extrapolated for higher flows, there are significant uncertainties related to these upstream boundaries. The downstream forcing is given by water level at the observing station of Convergent on the Atlantic Ocean coast. Water level observations are available hourly at Lesseps, Urt, Pont-Blanc, Villefranque and Peyrehorade. The flow is represented within the riverbed and infinite banks except in the neighboring of Peyrehorade where floodplains are locally modeled. This model was developed by the SPC GAD[1] in collaboration with SCHAPI.

[1] Service de Prévision des Crues (SPC) of Gironde and ADour (GAD)

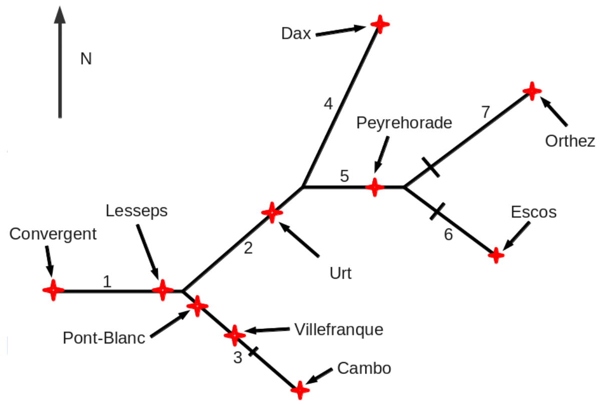


Figure 1, schematic representation of the “Adour maritime” hydraulic network. The black lines on the reaches 3, 6 and 7 represent the dams on the “Adour maritime” network.

Implementation

As illustrated in Figure 2 all the members are propagated from one assimilation cycle to the other by the hydraulic model. Every hour an assimilation is performed using water level observations only. Still, due to the multivariate (between water level and discharge) covariances stochastically estimated in **B**, the DA provides a correction of both water level and discharge hourly. A 12-hour forecast is integrated after each DA up-date in order to quantify the impact of the hydraulic state correction at different lead-times. It should be noted that there is a time delay of respectively 6h, 10h and 12h between the upstream boundaries and the observing stations of Peyrehorade, Urt and Lesseps. Beyond this period the upstream forcings are prescribed as constant values since no hydrological model is available upstream of the hydraulic network.

ENSEMBLE BASED DATA ASSIMILATION ALGORITHM

Data assimilation methods

The objective of data assimilation is to estimate the state vector $\mathbf{X}_i$ of a model using an observation $\mathbf{y}_i^o$ made at time t_i. In our case the state vector is the discretisation of the water level and discharges over the domain. In this study an EnKF algorithm (Evensen, [4]) was used to perform assimilation over the “Adour maritime” network. In the EnKF the state vector is estimated as the mean of an ensemble of perturbed integrations of the model $\mathbf{X}_i^{b,1}, \dots, \mathbf{X}_i^{b,N}$ at time t_i called background vectors. As illustrated on Figure (2) for the assimilation cycle i, the ensemble of analyzed states $\mathbf{X}_{i-1}^{a,1}, \dots, \mathbf{X}_{i-1}^{a,N}$ are propagated by the model from the assimilation cycle $i-1$ to the assimilation cycle i to provide the background states $\mathbf{X}_i^{b,k} = M\left(\mathbf{X}_i^{a,k}\right)$ over which the background error covariance matrix $\mathbf{B}_i$ is computed. The assimilation cycle i consists in assimilating a perturbed observation vector $\mathbf{y}_i^o + \varepsilon_i^{o,k}$ to correct the background vectors $\mathbf{X}_i^{b,k}$, using the Kalman Filter gain matrix $\mathbf{K}_i$:

$$\mathbf{X}_i^{a,k} = \mathbf{X}_i^{b,k} + \mathbf{K}(\mathbf{y}_i^{o,k} + \varepsilon_i^{o,k} - \mathbf{H}(\mathbf{X}_i^{b,k})), \quad (6)$$

$$\mathbf{K}_i = \mathbf{B}_i\mathbf{H}^T(\mathbf{H}\mathbf{B}_i\mathbf{H}^T + \mathbf{R})^{-1}, \quad (7)$$

The background error covariance functions in $\mathbf{B}_i$ allow to spread the information from the observing stations (only 6 locations) to the rest of the simulated domain as well as to the unobserved variables (discharge in the present case). These covariances are implicitly evolved in time over the assimilation cycles and stochastically estimated amongst the hydraulic states within the ensemble that are generated using different forcing conditions.

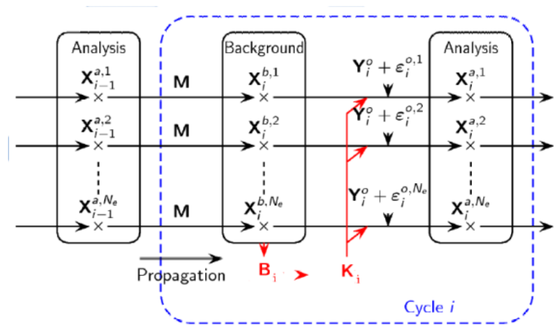


Figure 2, schematic representation of the EnKF algorithm at assimilation cycle i.

Online estimation of the matrices B and R

The online estimation of the matrices **B** and **R** is derived from the consistency diagnostics of **B** and **R** presented in Desroziers *et al.* [3] and applied in an ensemblist case in Li *et al.* [8]. While in (Li *et al.* [8]) those diagnostics are used to tune the values of the error variance of the background and the observation at the observing stations, we show in this study how this information can impact the members of the ensemble on the whole domain consistently with the model thus allowing for the improvement of the results of the assimilation. The method presented here is different whether there is one or several observing station, thus for the sake of simplicity we first present the case with only one observing station.

Case with only one observing station

Let us denote by:

- $\mathbf{d}_{o-b}$ the vector of the perturbed observations minus the corresponding background at the observing station ;
- $\mathbf{d}_{a-b}$ the vector of the analyzed states minus the corresponding background at the observing station ;
- $<.>$ the mean of a given value over the whole members ;

At a given time, if σ_b^2 and σ_o^2 are the true background and observation error variance then the following relationships are verified:

$$<\mathbf{d}_{o-b}\mathbf{d}_{o-b}^T>=\sigma_b^2+\sigma_o^2, \quad (8)$$

$$<\mathbf{d}_{a-b}\mathbf{d}_{o-b}^T>=\sigma_o^2, \quad (9)$$

In practice the relation (7) is not always verified. Following (Anderson [1]), when it's feasible, we introduce a factor $\lambda^2>0$ such that:

$$<\mathbf{d}_{o-b}\mathbf{d}_{o-b}^T>=\lambda^2\sigma_b^2+\sigma_o^2, \quad (10)$$

From (10) we can compute the value of λ. The background vectors have to be tuned using λ, consistently with the model and such that relation (8) holds. Let us denote by $\widetilde{\mathbf{X}}^{b,k}$ the k-th member tuned, we have:

$$\widetilde{\mathbf{X}}^{b,k}(l)=\big(1+(\lambda-1)C(l)\big).(\mathbf{X}^{b,k}(l)-\overline{\mathbf{X}}^b(l))+\overline{\mathbf{X}}^b(l), \quad (11)$$

where l stands for the grid point number and C is the correlation function related to the observing station.

Nevertheless the later estimation of λ depends on the prior estimation of σ_o which may not be the true observation standard deviation. By performing an assimilation we can compute $\mathbf{d}_{a-b}$ and give an estimate of σ_o by using (9).

We can now give a new estimate of λ using (10) and σ_o and so on. This leads to an iterative process where the $\widetilde{\mathbf{X}}^{b,k}$, σ_b and σ_o are successively estimated using (9), (10) and (11). We consider that this process has converged when (10) olds for a value of λ close enough to 1 according to an criterion fixed *a priori*.

Case with several observing stations

In the case with several observing stations we introduce a factor $\lambda^2 > 0$ such that:

$$<\mathbf{d}_{o-b}\mathbf{d}_{o-b}^T>=\lambda^2\mathbf{B}+\mathbf{R}, \quad (12)$$

Following Dee [3] we consider the trace of those matrices and λ^2 verifies:

$$\mathrm{Tr}(<\mathbf{d}_{o-b}\mathbf{d}_{o-b}^T>)=\lambda^2\mathrm{Tr}(\mathbf{B})+\mathrm{Tr}(\mathbf{R}), \quad (13)$$

We compute the value of λ using (13). The background vectors are tuned using λ consistently with the model using (11) where $C(l) = \mathbf{Max}\{C_1(l); C_2(l); C_3(l)\}$, where C_1, C_2 and C_3 are the correlation functions related to the stations of Peyrehorade, Urt and Lesseps respectively. By performing an assimilation we compute $\mathbf{d}_{a-b}$ and give an estimate of **R.**

We now give a new estimate of λ using (13) and **R,** and so on. Just as before this leads to an iterative process and we consider that it has converged when λ is close enough to 1 according to a criterion fixed *a priori*.

RESULTS

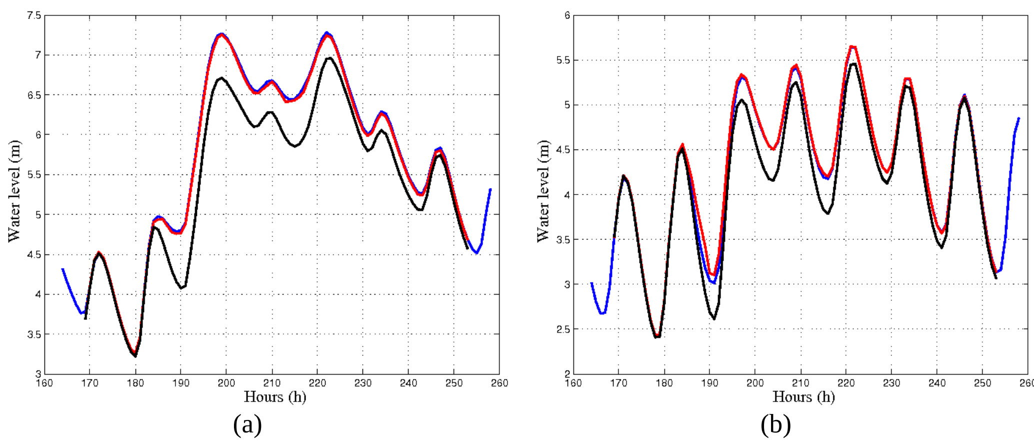
Data assimilation results: OSSE experiments

In this section, we apply the DA method in the framework of OSSE experiments. The synthetic water level observations are generated with MASCARET using an additional lateral inflow downstream of the dam on reach number 6 with a maximum of 375 $m^3.s^{-1}$ at the flood peak, considered as the true run. This forcing is correlated in time with the upstream forcings of reaches 6 and 7. The true water level are perturbed with a measurement error, then assimilated hourly while the discharge synthetic observations are used for validation purpose of the IEnKF algorithm only.

We first focus on the results of DA in the framework of the OSSE. Then the results with real data experiments are presented.

Data assimilation results for the correction of the hydraulic state

In this experiment, the DA algorithm is sequentially applied using synthetic observations of water level at Peyrehorade from the first time step of the flood event to the end of the flood peak.

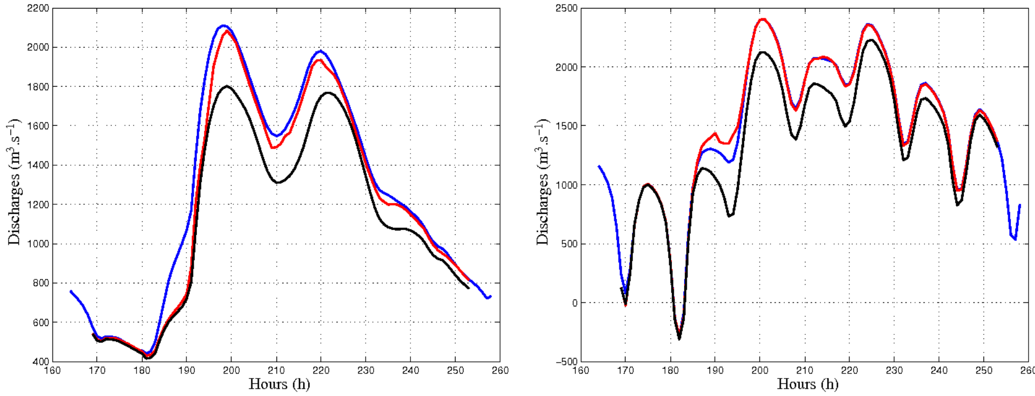


(a) (b)

Figure 7, analyzed water level (red line) computed with the IEnKF at (a) Peyrehorade and (b) Urt. The observations are plotted in blue and the free run without assimilation in black.

Figure 7-(a) shows that the analyzed water level at the station of Peyrehorade is brought closer to the observations than the free run (no assimilation). As the inflation factor from equation (8) is greater than 1, the background error variance is increased, resulting in a proportionally smaller uncertainty in the observations when the inflation is used.

The correction of water level computed at Peyrehorade is spread over the domain consistently with the model state error statistics through the covariance functions. Hence the observation assimilated at Peyrehorade translates into a correction at Urt and the water level is also improved at Urt as displayed in Figure 7-(b).



(a) (b)

Figure 8, analyzed discharges computed with the IEnKF (red line) at (a) Peyrehorade and (b) Urt. The observations are plotted in blue and the free run without assimilation in black.

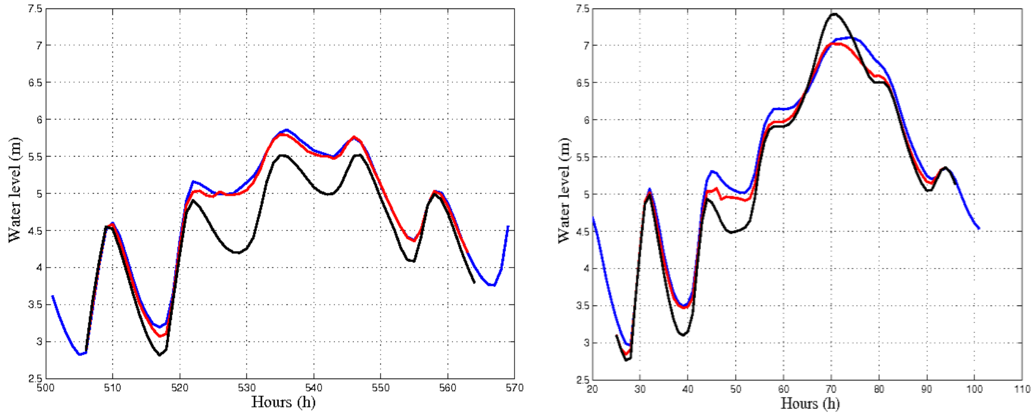
Figure 8-(a) displays the analyzed discharges at Peyrehorade and Urt. Here again the analyzed discharges are brought closer to the observations by the DA algorithm which means that the correlation between the errors in discharges and the water levels are correctly estimated in the IEnKF algorithm. This correction is then spread over the domain to unobserved locations. Still, it should be noted that in the framework of OSSE, the relation between water level and discharge prescribed by the model are consistent with the relation between water level and discharge prescibed by the observation (as the observations are syntheticly generated). This assumption may no longer hold in real case study as the model usually suffers from errors in the bathymetry and friction coefficients that locally impact the relation between water level and discharge. The same remark is valid for the spatial correction: the model state error at different locations is coherent with the observation error at different observing stations, thus the assimilation of observations from one location leads to an improvement at other locations, but this result may hot hold for real case experiments.

After each assimilation cycle a 12 hours forecast is performed. Numerical experiments not reported here show that as the lead time increases, the analyzed water level drifts towards the free run results since the impact of a correction to the hydraulic state (i.e the initial condition for further forecast) is limited in time. Nevertheless the forecast is improved for short-range forecast compared to the run without assimilation. This demonstrates the need to extent the control vector of the DA to more than the instantaneous hydraulic state, for instance correcting upstream forcings or hydraulic parameters in order to improve medium and long range forecasts.

Data assimilation results: real data experiments

In the case of real data experiments discharge observations are not available, thus only water level results are shown here. The IEnKF is applied to five flood events on the Adour catchment.

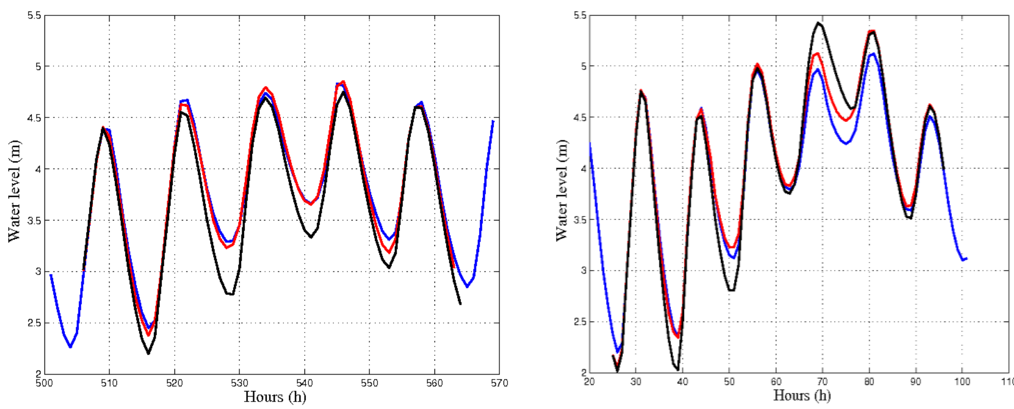
Figure 10 displays analyzed water level for two different flood events. Figure 10-(a) represents a small flood event whereas Figure 10-(b) represents an important flood event.



(a) (b)

Figure 10, analyzed water level (red lines) at Peyrehorade for two different flood events. The observations are plotted in blue and the free run without assimilation is plotted in black.

Figure (11) displays the analyzed water level at Urt for the two events presented in Figure (10). As expected in the framework of real data experiments, the improvement of water level at Urt is smaller than what was obtained in the framework of OSSE. Indeed, due to the imperfect modeling of the “Adour maritime” hydraulic network in the neighbourhood of Urt (with infinite banks and missing flood plains), the model tends to overestimate water level at Urt more than at Peyrehorade when major flood events occur.



(a) (b)

Figure 11, analyzed water level (red lines) at Urt for the two same events as Figure 6. The observations are plotted in blue and the free run without assimilation in black.

Forecast range	Peyrehorade	Forecast range	Urt	Forecast range	Lesseps
0h	77%	0h	58%	0h	38%
1h	61%	2h	39%	2h	6%
3h	34%	5h	13%	6h	-2%
6h	11%	10h	3%	12h	-2%

Table 1, mean improvement of the RMS at the stations of Peyrehorade, Urt and Lesseps for different forecast range.

Nonetheless in every case the data assimilation results in an improvement of the root-mean-square (RMS) of the analyzed water at the stations where the assimilation is performed. Table (1) summarizes the mean improvement on the five flood events considered here of the RMS at each station for different forecast range.

The mean improvement of the RMS of the analyzed water level at the stations is correlated to the ability of the model to compute correct water level without assimilation. Hence at Peyrehorade where the model is usually not good during flood peaks the mean improvement of the RMS with the IEnKF is about 77%. For each station the mean improvement of the analyzed water level decreases during the forecast period. In the case of Lesseps the mean improvement of the RMS during the forecast period is negative. This is because the computed water levels at Lesseps are under a strong tidal influence. Hence during the forecast period where the forcings ar set constant the computed water level cannot match with the observations.

CONCLUSION

This study describes the application of an EnKF algorithm on “Adour maritime” network using the hydraulic code MASCARET. This allows for the study of the model error correlation functions. Those functions are closely related to the geometry of the model and the forcings and are characterized by important spatial extent. An additional algorithm that estimates the background and observation error covariances at the observing stations was implemented and shows good results over the entire network in terms of analyzed and forecasted water level for short range forecast compared to a free run without assimilation. A direction application of this work is to provide inputs for local 2D models.

REFERENCES

[1] Anderson, J. L., “An adaptive covariance inflation error correlation algorithm for ensemble filters.”, *Tellus A*, Vol. 59, No.2, (2007), pp 210-224.

[2] Dee, D. P., “On-line estimation of error covariance and parameters for atmospheric data assimilation.”, *Quarterly Journal of the Royal Meteorological Society,* Vol. 123, No. 4, (1995), pp 1128-1145.

[3] Desroziers, G., Berre, L., Chapnik, B. and Poli, P., “Diagnosis of observation, background and analysis error statistics in observation space.”, *Quarterly Journal of the Royal Meteorological Society*, Vol. 131, No. 613, (2005), pp 3385-3396.

[4] Evensen, G., “Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics.”, *Journal of Geophysical Research: Oceans*, Vol. 99, No. C5, (1994), pp 10143-10162.

[5] Goutal, N. and Maurel, F., “A finite volume solver for 1D shallow-water equations applied to an actual river.”, *International Journal for Numerical Methods in Fluids*, Vol. 38, No. 1, (2002), pp 1-19.

[6] Jean-Baptiste, N., Malaterre, P. O., Dorée, C. and Sau, J., “Data assimilation for real-time estimation of hydraulic states and unmeasured perturbations of 1D hydrodynamic model.”, *Mathematics and Computers in Simulation*, Vol. 81, No. 10, (2001), pp 2201-2214.

[7] Kalman, R. E. and Bucy, R. S., “New results in linear filtering prediction theory.”, *Journal of basic engineering*, Vol. 83, No. 1, (1961), pp 95-108.

[8] Li, H., Kalnay, E. and Miyoshi, T., “Simultaneous estimation of covariance inflation and observations errors within an Ensemble Kalman Filter.*”, Quarterly Journal of the Royal Meteorological Society*, Vol. 135, No. 639, (2009), pp 523-533.

[9] Ricci, S., Piacentini, A., Thual, O., Le Pape, E. and Jonville, G., “Correction of upstream flow and hydraulic state with data assimilation in the context of flood forecasting.”, *Hydrology & EarthSystem Sciences*, Vol. 15, No. 11, (2011).

[10] Weaver, A. and Courtier, P., “Correlation modeling on the sphere using a generalized diffusion equation.”, *Quarterly Journal of the Royal Meteorological Society*, Vol. 127, No. 575, (2001), pp 1815-1846.