Data Assimilation with an Ensemble Kalman Filter Algorithm on an Operational Hydraulic Network

Sébastien Barthélémy, Sophie Ricci, Nicole Goutal, Olivier Thual, Etienne Le Pape

Data Assimilation with an Ensemble Kalman Filter Algorithm on an Operational Hydraulic Network

Sébastien Barthélémy, Sophie Ricci, Nicole Goutal, Olivier Thual, Etienne Le Pape

Abstract

This study describes the implementation and the merits of an Ensemble Kalman Filter algorithm (EnKF) on the 1Dshallow water model MASCARET for the representation of the hydrodynamics of the " Adour maritime" river in south west France. The first part of this work is dedicated to a detailed analysis of the background error covariance functions that are stochastically estimated on an ensemble of MASCARET integrations forced by perturbed upstream boundary conditions. It is shown that the geometric characteristics of the network have a significant impact on the shape of these functions and thus on the data assimilation correction. 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 to improve 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.

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

Sébastien Barthélémy[1]
URA 1875/CERFACS, Toulouse, France
barthelemy@cerfacs.fr

Sophie Ricci
URA 1875/CERFACS, Toulouse, France
ricci@cerfacs.fr

Nicole Goutal
EDF R&D, LNHE, Chatou, France
nicole.goutal@edf.fr

Olivier Thual
URA 1875/CERFACS and INPT, CNRS, IMFT, Toulouse, France
thual@imft.fr

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

KEY WORDS

Flood forecasting, “Adour maritime” hydraulic network, background and observation error covariance estimation, inflation, sequential filtering.

ABSTRACT

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 first part of this work is dedicated to a detailed analysis of the background error covariance functions that are stochastically estimated on an ensemble of MASCARET integrations forced by perturbed upstream boundary conditions. It is shown that the geometric characteristics of the network have a significant impact on the shape of these functions and thus on the data assimilation correction. 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 to improve 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.

1. INTRODUCTION

The hydraulic observing network in France provides water level data at relatively high frequency (usually hourly and up to 5 minutes) but with a 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

[1] barthelemy@cerfacs.fr

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 the model outputs.

It has already been demonstrated that DA can make an important contribution in the context of hydraulic (Durand *et al*. [6], Jean-Baptiste *et al*. [9]), 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 [15], Pannekoucke *et al*. [13]). It was also discussed in the context of hydraulic by (Ricci *et al*. [14]) and (Madsen and Skotner [12]).

While the classical Kalman Filter algorithm (Kalman and Bucy [10]) 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*. [5]). 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 with the dynamic couling software OPALM. Section 4 first illustrates how the network properties impact the shape of the covariance functions. Then it presents the data assimilation results for OSSE and real data experiments. Some conclusive remarks are finally given in Section 5.

2. MODELING OF THE ADOUR MARITIM CATCHMENT

2.1 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 [8]), it solves the conservative form of 1D shallow-water equations:

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

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).

Generally speaking, uncertainties in the model formulation itself due to simplified physics and also in the input fields to the model such as the boundary conditions, initial conditions and hydraulic parameters, turbulence model, numerical schemes translate into errors in the MASCARET simulated hydraulic variables. In spite of constant advances in numerical methods, computational resources and input data acquisition (especially topography and bathymetry), a significant part of these uncertainties remain and DA appears as a complementary way of improving simulations.

2.2 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. 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 cost. 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 modelled. This model was developed by the Service de Prévision des Crues (SPC) of Gironde and ADour (GAD) in collaboration with SCHAPI. SPC GAD and SCHAPI provide public and daily color-scaled risk maps available on-line (from green for low conditions to red for high risk conditions). According to the statistics computed by SCHAPI, the Adour catchmet is ranked amongst the most challenging catchments with a large number of orange and red alerts.

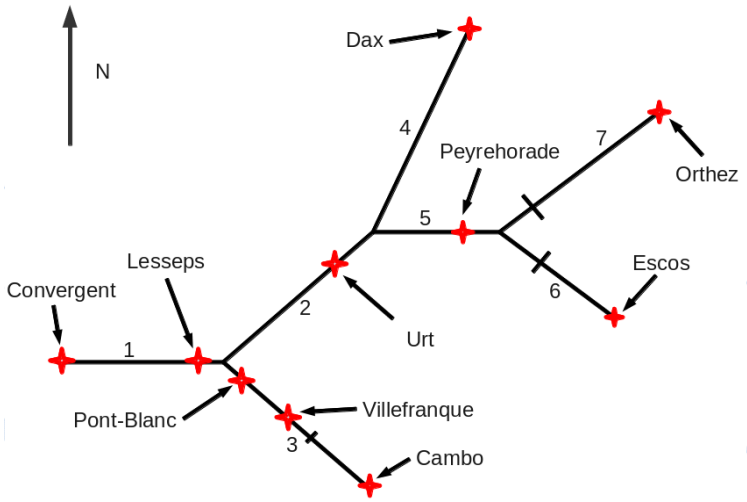


Figure 1: Schematic representation of the "Adour maritime" hydraulic network. The black lines on the reaches 3, 6 and 7 represent the dams on the "Adour maritime" network.

2.3 Implementation

In practice, combining MASCARET with a data assimilation algorithm is managed by the OpenPalm dynmic coupling software (Buis *et al.* [2]) developped by CERFACS and ONERA. It is used as a task parallelism manager to handle communications and data exchanges between MASCARET and the different mathematical units required to sequentially perform Bayesian prediction and update steps. Since each member of the ensemble can be integrated independently the Parasol functionality of OpenPalm is used to efficiently launch MASCARET model integrations, in parallel, on the available processors. The Master processor of Parasol spawns multiple copies of the same computer program (the slaves), each on one or several processors with a different set of input parameters (upstream forcing in the present case), while each slave processor is in charge of executing one MASCARET instance and producing the associated hydraulic state.

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 hydrualic 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

3. ENSEMBLE BASED DATA ASSIMILATION ALGORITHM

3.1 Data assimilation methods

The objective of data assimilation is to estimate the state vector X_i using the observation $\mathbf{y}_i^o$ made at time t_i. In our case X_i represents the discretisation of the water level and discharge in each grid point of the domain. In the prediction step, the probability density function (denoted hereafter PDF) of the state vector is evolved from time t_{i-1} until time t_i. We note $p_b(X_i)$ this PDF (also called the background error PDF) at time t_i. At the analysis time t_i, also called the update step, this background PDF is corrected in order to be more consistent with the observations $\mathbf{y}_i^o$. The new PDF, called the analysis and noted $p_a(X_i)$, is given by Bayes'theorem:

$$p_a(X_i) \propto p(\mathbf{y}_i^o|X_\mathrm{i})p_b(X_\mathrm{i}), \tag{2}$$

where the symbol $\propto$ means "proportional to" and where $p(\mathbf{y}_i^o|X_\mathrm{i})$ represents the data likelihood, i-e the conditional PDF of having the observations $\mathbf{y}_i^o$ given the state X_i.

The EnKF algorithm assumes that both the model state X_i and the observations $\boldsymbol{y}_i^o$ are random variables defined by Gaussian PDF with a zero mean value and an error covariance model. Under these assumptions, the background PDF may be written as:

$$p_b(X_i) \propto \exp\left\{-\frac{1}{2}(X_i - X_i^b)\mathbf{B}^{-1}\left(X_i - X_i^b\right)^T\right\}, \tag{3}$$

where $\mathbf{x}_t^b$ is the background estimate of the true state vector and where **B** is the background error covariance matrix representing modeling errors. The data likelihood may be similarly expressed as:

$$p(\mathbf{y}_i^o|X_i) \propto \exp\left\{-\frac{1}{2}(\mathbf{y}_i^o - H(X_i))^T\mathbf{R}^{-1}(\mathbf{y}_i^o - H(X_i))\right\}, \tag{4}$$

with **R** the observation error covariance matrix representing observation errors. Within this framework, the analysis PDF is also Gaussian and is written as:

$$p_a(X_i) \propto exp\left\{-\frac{1}{2}(X_i - X_\mathrm{i}^a)^T\mathbf{A}^{-1}(X_i - X_\mathrm{i}^a)\right\}, \tag{5}$$

where X_i^a is the analysis estimate of the true state vector and where **A** is the analysis error covariance matrix.

The classical Kalman filter algorithm assumes that the observation *H* is linear (denoted by **H**); in that case, it may be shown that the analysis update in Equation (4) leads to the following equations :

$$X_i^a = X_i^b + \mathbf{K}(\mathbf{y}_i^o - \mathbf{H}(X_i^b)),\ \mathbf{K} = \mathbf{B}\mathbf{H}^T\left(\mathbf{H}\mathbf{B}\mathbf{H}^T + \mathbf{R}\right)^{-1},\ \mathbf{A} = (\mathbf{I} - \mathbf{K}\mathbf{H})\mathbf{B}, \tag{6}$$

where **K** is called the gain matrix.

In contrast, the Ensemble Kalman Filter (EnKF) algorithm illustrated in Figure (2) approximates the forecasted PDF of the state vector $p_b(X_\mathrm{i})$ by performing a series (an ensemble) of *N* independent forward model integrations up to the analysis time t_i, thereby providing *N* forecast estimates of the state vector, called the ensemble members, $X_{i,e}^b = \left[X_i^{b,1}, X_i^{b,2}, \ldots, X_i^{b,N}\right]$. The EnKF algorithm approximates the mean and the covariance of the ensemble, while still making the assumption that all PDF are Gaussian. During the analysis, each ensemble member is updated using the classical Kalman filter formulation in Equation (6) with the difference that the gain matrix $\mathbf{K}_i$ is now calculated from an estimate of the background error covariance matrix, noted $\mathbf{B}_i$:

$$\mathbf{B}_i = \frac{1}{N-1}\sum_{k=1}^{N}\left(\mathbf{X}_i^{b,k} - \overline{\mathbf{X}_\iota}^b\right)\left(\mathbf{X}_i^{b,k} - \overline{\mathbf{X}_\iota}^b\right)^T, \tag{7}$$

and using the ensemble-based stochastic representation of the relationship between state space and observation space. Note that in the present study, we use the EnKF version proposed by Burgers *et al*. [3] where the observation is perturbed for each member: $y_i^o + \varepsilon_i^{o,k}$ where y_i^o is the observation at time t_i, $\varepsilon_i^{o,k}$ is a Gaussian noise with zero mean and variance equal to the prescribed variance of the observation σ_o. We also assume that the observation errors are uncorrelated, i-e the observation error covariance matrix **R** is treated as a diagonal matrix in which each diagonal term is the error variance σ_o^2 associated with water level measurements.

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.

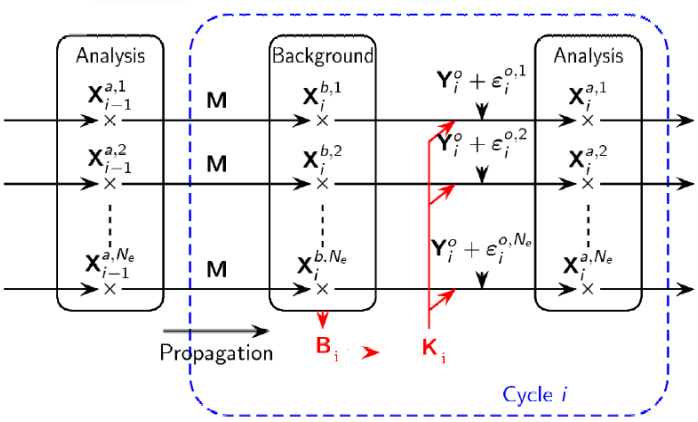


Figure 2: Schematic representation of the EnKF algorithm at assimilation i.

3.2 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.* [5]) and applied in an ensemblist case in (Li *et al.* [11]). While in (Li *et al.* [11]) those diagnostics are used to tune the values of the variance of the background and the observation at the observation points we show in this study how this information can impact the members of the ensemble at the observation points and on the domain consistently with the model thus allowing for the improvement of the results of the assimilation. The method presented here differs if there is one or several observation points, thus for the sake of simplicity we first present the case with only one observation point.

3.2.1 Case with only one observation point

Let us denote by :

- $\mathbf{d}_{o-b}$ the vector of the perturbed observations minus the corresponding background at the observation point ;
- $\mathbf{d}_{a-b}$ the vector of the analyzed states minus the corresponding background at the observation point ;
- $<.>$ 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 :

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

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

Because the true background and observation error variance are not always known, the relation (7) is not always verified. Hence, following (Anderson [1]), when it's feasible, we introduce a factor $\lambda > 0$ and we assume that $\lambda\sigma_b^2$ and σ_o^2 are the true background and observation error variances and then the following relation holds :

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

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

$$\widetilde{\mathbf{X}}^{b,k}(l)=\begin{cases}\sqrt{1+(\lambda-1)C(l)}.\left(\mathbf{X}^{b,k}(l)-\overline{\mathbf{X}}^{b}(l)\right)+\mathbf{X}^{b,k}(l), & 1+(\lambda-1)C(l)\geq 0\\ -\sqrt{|1+(\lambda-1)C(l)|}.\left(\mathbf{X}^{b,k}(l)-\overline{\mathbf{X}}^{b}(l)\right)+\mathbf{X}^{b,k}(l), & 1+(\lambda-1)C(l)<0\end{cases}, \tag{11}$$

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

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 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*.

3.2.2 Case with several observation points

In the case with several observation points, if the matrices **B** and **R** are the true background and observation error covariance matrices then the following relationships are verified :

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

$$<\mathbf{d}_{a-b}\mathbf{d}_{o-b}^{T}>=\mathbf{R}, \quad (13)$$

Just as before we introduce a factor $\lambda > 0$ and we suppose that $\lambda\mathbf{B}$ is the true background error covariance matrix, then the following relation holds :

$$<\mathbf{d}_{o-b}\mathbf{d}_{o-b}^{T}>=\lambda\mathbf{B}+\mathbf{R}, \quad (14)$$

Following Dee [4] we consider the trace of those matrices and λ verifies :

$$\mathrm{Tr}(<\mathbf{d}_{o-b}\mathbf{d}_{o-b}^{T}>)=\lambda\mathrm{Tr}(\mathbf{B})+\mathrm{Tr}(\mathbf{R}), \quad (15)$$

Then we can compute the value of λ. The background vectors are tuned using λ consistently with the model and such that their variances at the observation point verifies (12) :

$$\widetilde{\mathbf{X}}^{b,k}(l)=\begin{cases}\sqrt{1+(\lambda-1)C'(l)}.\left(\mathbf{X}^{b,k}(l)-\overline{\mathbf{X}}^{b}(l)\right)+\mathbf{X}^{b,k}(l), & 1+(\lambda-1)C'(l)\geq 0\\ -\sqrt{|1+(\lambda-1)C'(l)|}.\left(\mathbf{X}^{b,k}(l)-\overline{\mathbf{X}}^{b}(l)\right)+\mathbf{X}^{b,k}(l), & 1+(\lambda-1)C'(l)<0\end{cases}, \quad (16)$$

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 (15) 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.*

4. RESULTS

4.1 Data assimilation results : OSSE experiments

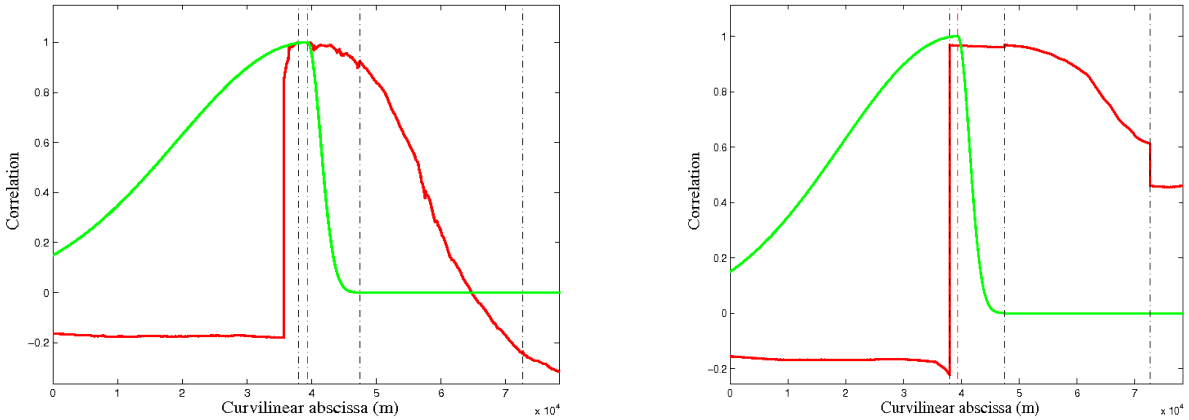
In this section, we apply the DA method in the framework of OSSE experiments. The synthetical water level observations are generated with MASCARET using an additional lateral inflow downstream of the dam on reach number 6 with a maximum of $375\ m^3.s^{-1}$ at the flood peak, considered as the true run. 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 synthetical observations are used for validation purpose of the IEnKF algorithm only.

We first focus on the correlation functions and the associated analysis increment. Then the results of the OSSE are presented.

4.1.1 Background error correlation functions

For the "Adour maritime" hydraulic network, the stochastically computed correlation functions are characterized by an important spatial extent and do not present a gaussian shape. These functions are significanty impacted by the geometry of the network, especially the presence of dams, confluences, slope, tide as well as the correlation time scale of the perturbation applied to the forcings.

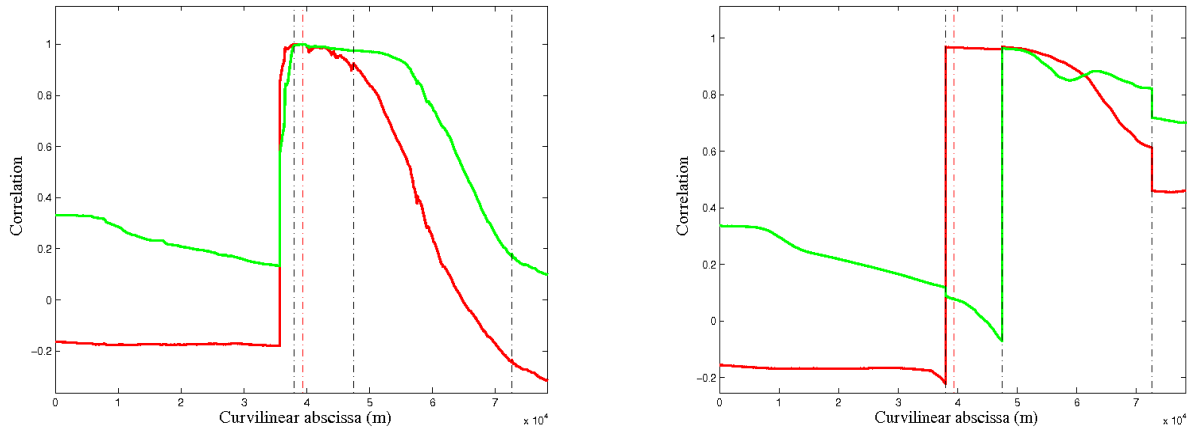
In previous studies by (Ricci *et al*. [14], Madsen and Stokner [12]), an invariant formulation of the background error covariance matrix (the Invariant Kalman Filter is denoted by IKF in the following) is proposed following an estimation on an idealized network. While these statistics lead to the improvement of the simulated water level at a reduced computational cost, they are not consistent with the realistic geometry of the Adour network and hence do not provide an optimal correction. The correlation functions from the IKF are represented in Figures 3-(a) and 3-(b) by green lines and display a significanly shorter correlation length-scale downstrean of the observation than upstream. Figure 3-(a) displays the water level correlation function related to Peyrehorade (red line) computed with the IEnKF along reaches 7-5-2-1. This function shows an important discontinuity at the dam of the reach 7. Upstream of this dam the function present a small anti-correlation while downstream of this dam it exhibits larger correlation length scale.



(a) (b)

Figure 3: Representation of (a) the water level correlation functions and (b) the discharge/water level correlation functions related to Peyrehorade computed with the IEnKF (red lines) and prescribed in an IKF in Ricci *et al*. [14] (green lines). The vertical red dashed line represent the position of the Peyrehorade station along the river. The vertical black dashed lines represent the separation between two following reaches.

Figure 3-(b) displays the discharge/waterlevel correlation function related to Peyrehorade (red line) and computed with the IEnKF at a given time along reaches 7-5-2-1. This function presents discontinuities at the confluences between reaches since discharge is an additive variable.

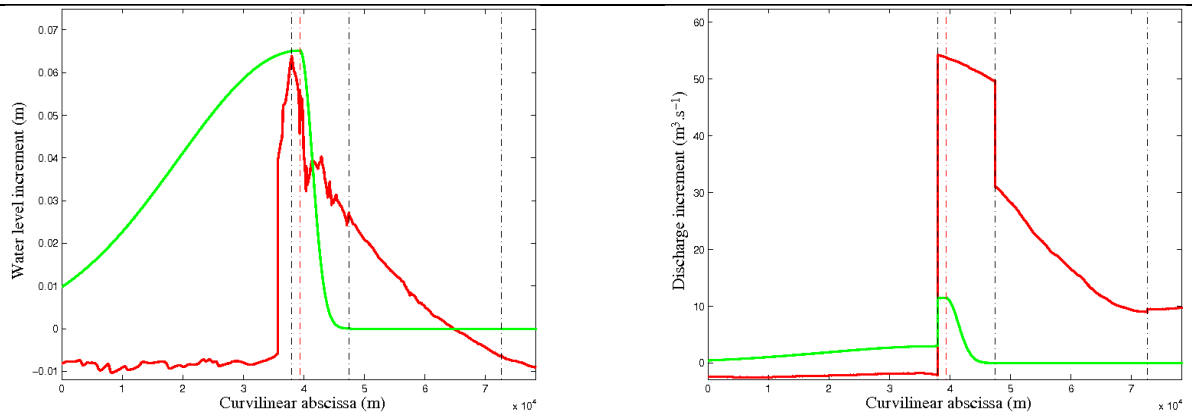


(a) (b)

Figure 4: (a) Water level correlation functions and (b) discharge/water level correlation functions related to Peyrehorade computed at two different times (red first and green second) by the IEnKF.

Figure 4 displays the water level (4-(a)) and discharge/water level (4-(b)) correlation functions computed at two different time steps with the IEnKF showing the evolution of the correlation functions over time and more specifically the influence of tide.

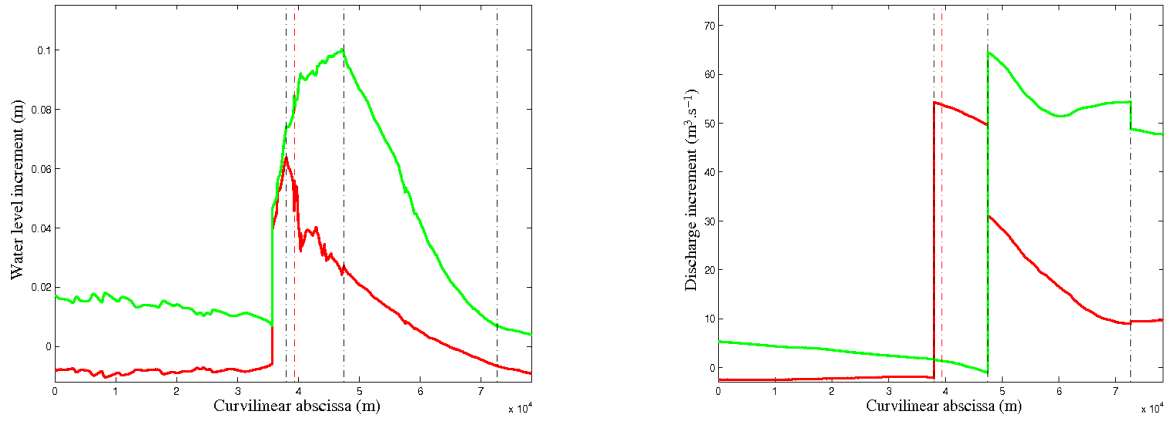
The DA increment in water level and discharge resulting from the analysis with previsously described correlation functions are shown in Figures 5 and 6 for both IKF and IEnKF and at different time steps. The correction from IEnKF is influenced by the presence of the dam on reach number 7 (Figure 5-(a), red line) as well as by the presence of the confluence between reaches (Figure 5-(b), red line) while the corrections from IKF are not influenced by the geometric properties of the network. It shoudl be noted that the discharge increment exhibits a discontinuity at the confluence between the reaches 7 and 5 so that the discharge computed at the confluence between those two reaches is equal to the sum of the discharges computed on reaches 6 and 7. As the univariate (water level/water level) and multivariate (discharge/water level) covariance functions are in coherence with the real hydraulic network and its temporal behavior, the DA increments from IEnKF are optimal. They provide a spatial correction to water level and discharge that allows to describe a corrected hydraulic state over the entire hydraulic network.



(a) (b)

Figure 5: Representation of (a) the analysis water level increment and (b) the analysis discharge increment computed with the IEnKF (red lines) and computed within an IKF with a steady correlation function in Ricci *et al*. [14] (green line).

Figure 6 shows the analysis increment of water level and discharges computed with the IEnKF at two different times showing the evolution of the analyzed increments from one time to an other.



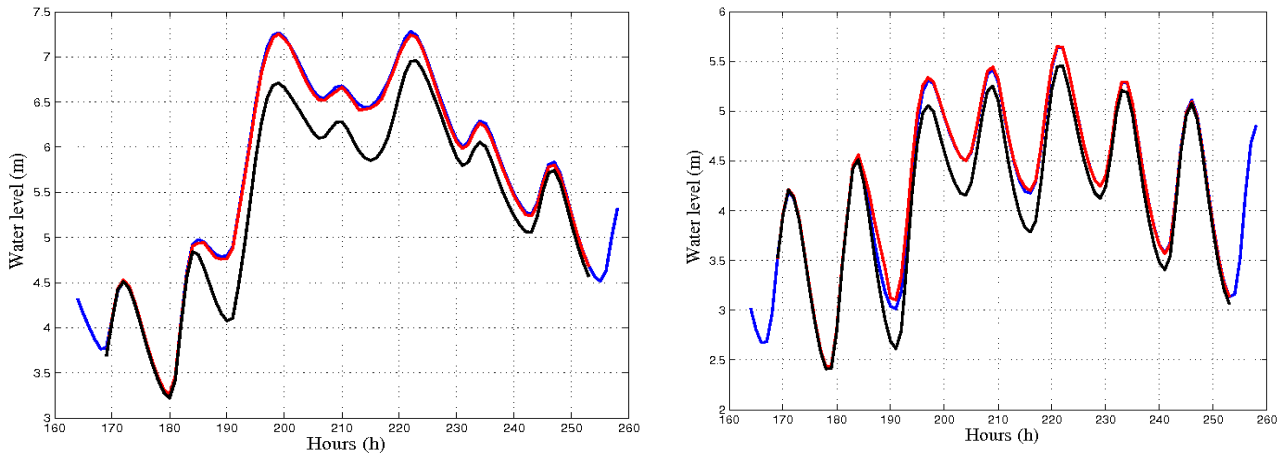
(a) (b)

Figure 6: (a) Analysis water level increment and (b) analysis discharge increment computed with the IEnKF at two different times (red first green next).

Those results show that the correlation functions have an important spatial extent and the related analysis increment are in agreement with the numerical model and the geometry of the "Adour maritime" network. As a consequence the more important the spatial extent of the analysis increment isthe more reliable the forecast. On the other hand the large extent of the analysis increment allows for providing flood maps far from observation points or inputs for local 2D models in ungauged areas with important issues such as cities or nuclear plants.

4.1.2 Data assimilation results for the correction of the hydraulic state

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

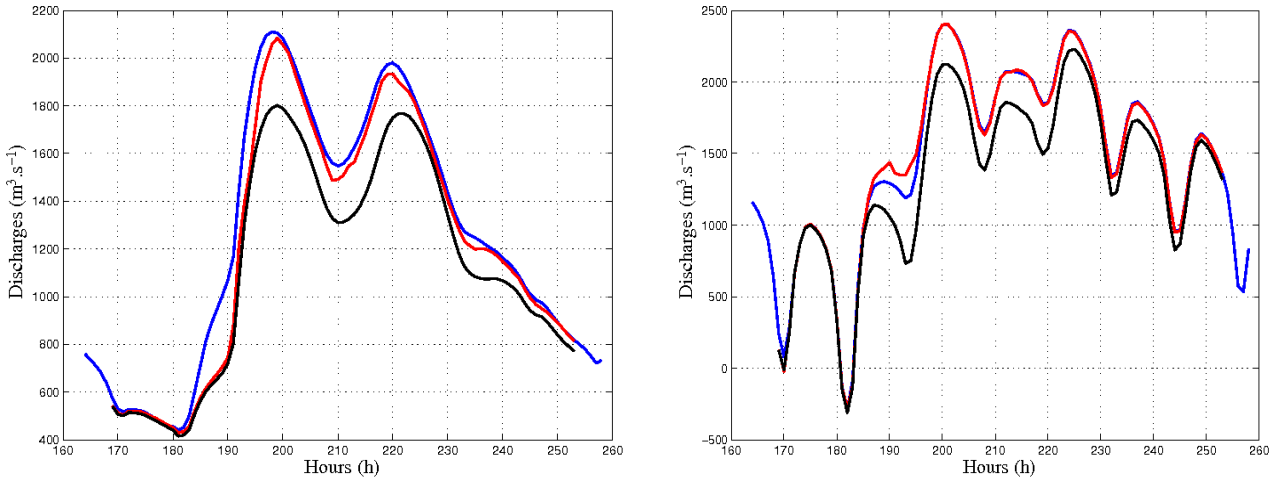


(a) (b)

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).

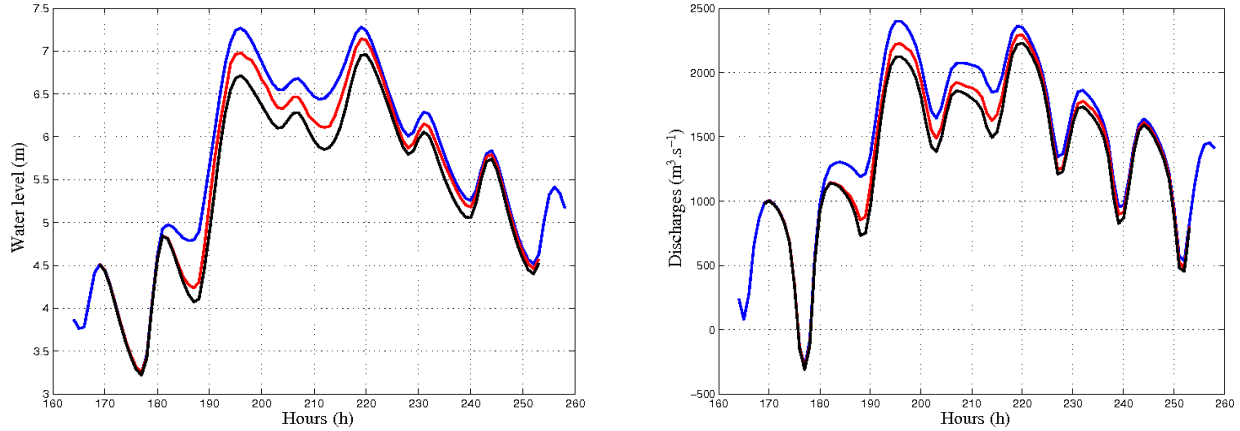


(a) (b)

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 synthetically 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.



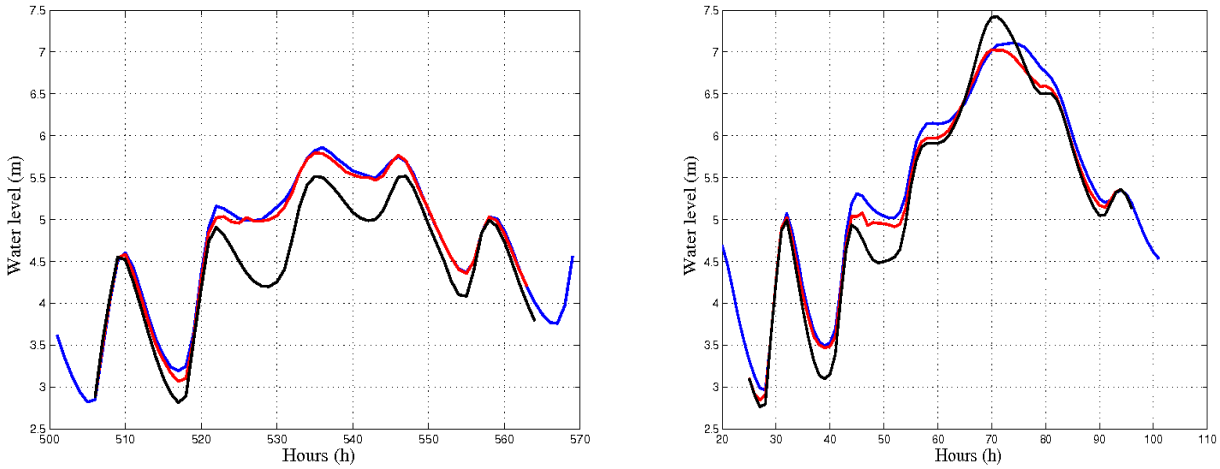
(a) (b)

Figure 9: Forecasted water level computed with the IEnKF (red line) at (a) Peyrehorade at 3h forecast and (b) Urt at 5h forecast. The observations are plotted in blue and the free run without assimilation is plotted in black.

After each assimilation cycle a 12 hours forecast is performed. Figure 9-(a) displays the 3h forecast at Peyrehorade. We notice 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.

4.3 Data assimilation results : real data experiments

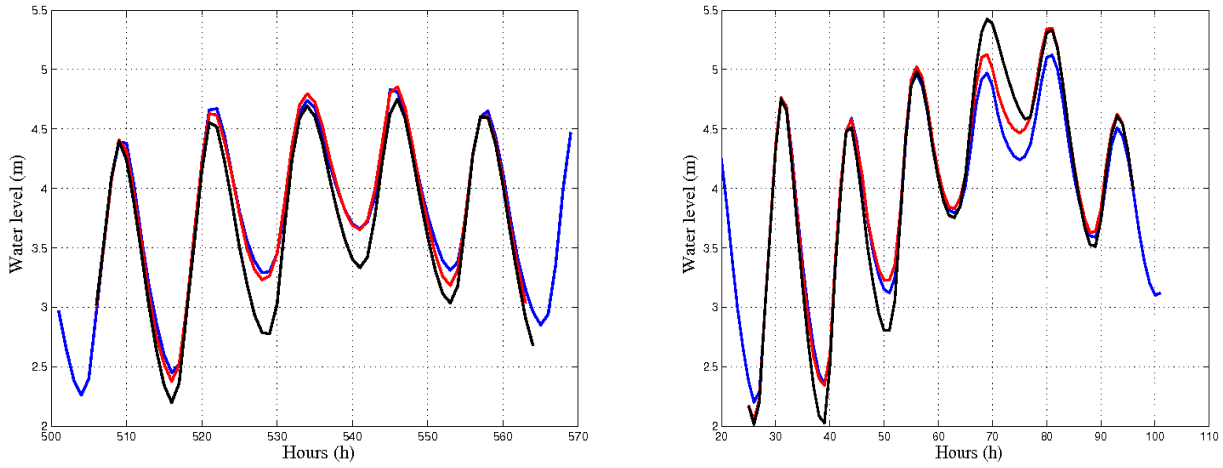
In the case of real data experiments discharge observations are not available, thus only water level results are shown here. The IEnKF is applied to five flood events on the Adour catchment.



(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 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 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.

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 over estimate water level at Urt more than at Peyrehorade when major flood events occur.

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.

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.

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 level at Lesseps are under a strong tidal influence. Hence during the forecast period where the forcings ar set constant the computed water level can not match with the observations.

5.CONCLUSION

This study describes the applicationof 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.

ACKNOWLEDGEMENTS

The authors would like to thank the OpenPALM team for their help and support through this work.

REFERENCES AND CITATIONS

[1] Anderson, J. L. (2007). An adaptive covariance inflation error correlation algorithm for ensemble filters. *Tellus A*, **59**, 210-224.

[2] Buis, S., Piacentini, A., & Déclat, D. (2006). PALM : a computational framework for assembling high performance computing applications. *Concurrency and Computation: Practice and experience*, **18**, 231-245.

[3] Burgers, G., Van Leeuwen, P. J., & Evensen, G. (1998). Analysis scheme in the ensemble Kalman filter. *Monthly weather review*, **126**.

[4] Dee, D. P. (1995). On-line estimation of error covariance and parameters for atmospheric data assimilation. *Quarterly Journal of the Royal Meteorological Society,* **123**, 1128-1145.

[5] Desroziers, G., Berre, L., Chapnik, B., & Poli, P. (2005). Diagnosis of observation, background and analysis errorstatistics in observation space. *Quarterly Journal of the Royal Meteorological Society*, **131**, 3385-3396.

[6] Durand, M., Andreadis, K. M., Alsdorf, D. E., Lettenmaier, D. P., Moller, M., & Wilson, M (2008).Estimation of bathymetric depth and slope from data assimilation of swath altimetry into a hydrodynamic model. *Geophyical Research Letter*, **20**.

[7] Evensen, G. (1994). Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *Journal of Geophysical Research: Oceans*, **99**, 10143-10162.

[8] Goutal, N., & Maurel, F. (2002). A finite volume solver for 1D shallow-water equations applied to an actual river. *International Journal for Numerical Methods in Fluids*, **38**, 1-19.

[9] Jean-Baptiste, N., Malaterre, P. O., Dorée, C., & Sau, J. (2011). Data assimilation for real-time estimation of hydraulic states and unmeasured perturbations of 1D hydrodynamic model. *Mathematics and Computers in Simulation*, **81**, 2201-2214.

[10] Kalman, R. E., & Bucy, R. S. (1961). New results in linear filtering prediction theory. *Journal of basic engineering,* **83**, 95-108.

[11] Li, H., Kalnay, E., & Miyoshi, T. (2009). *Simultaneous estimation of covariance inflation and observations errors within an Ensemble Kalman Filter. Quarterly Journal of the Royal Meteorological Society*, **135**, 523-533.

[12] Madsen, H., & Skotner, C. (2005). Adaptive state updating in real-time river flow forecasting – A combined filtering and error forecasting procedure. *Jounral of hydrology*, **308**, 302-312.

[13] Pannekoucke, O., & Massart, S. (2008). Estimation of the local tensor and normalization for heterogeneous correlation modeling using a diffusion equation. *Quarterly Journal of the Royal Meteorological Society*, **134**, 1425-1438.

[14] Ricci, S., Piacentini, A., Thual, O., Le Pape, E., & Jonville, G. (2011). Correction of upstream flow and hydraulic state with data assimilation in the context of flood forecasting. *Hydrology & EarthSystem Sciences,* **15**.

[15] Weaver, A., & Courtier, P. (2001). Correlation modeling on the sphere using a generalized diffusion equation. *Quarterly Journal of the Royal Meteorological Society*, **127**, 1815-1846.