



HAL
open science

Ensemble-based algorithm for error reduction in hydraulics in the context of flood forecasting

Sébastien Barthélémy, Sophie Ricci, Etienne Le Pape, Mélanie Rochoux, Olivier Thual, Nicole Goutal, Johan Habert, Andrea Piacentini, Gabriel Jonville, Fabrice Zaoui, et al.

► To cite this version:

Sébastien Barthélémy, Sophie Ricci, Etienne Le Pape, Mélanie Rochoux, Olivier Thual, et al.. Ensemble-based algorithm for error reduction in hydraulics in the context of flood forecasting. FLOODrisk 2016 - 3rd European Conference on Flood Risk Management, E3S Web of Conferences, 7, pp.18022, 2016, 10.1051/e3sconf/20160718022 . hal-04691417

HAL Id: hal-04691417

<https://hal.science/hal-04691417v1>

Submitted on 8 Sep 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

Ensemble-based algorithm for error reduction in hydraulics in the context of flood forecasting

Sébastien Barthélémy^{1,a}, Sophie Ricci¹, Etienne Le Pape², Mélanie Rochoux¹, Olivier Thual^{1,3}, Nicole Goutal^{4,5}, Johan Habert⁶, Andrea Piacentini¹, Gabriel Jonville¹, Fabrice Zaoui⁴, Philippe Guoin²

¹CECI, CERFACS – CNRS, 42 avenue Gaspard Coriolis, 31820 Toulouse, France

²SCHAPI, 42 avenue Gaspard Coriolis, 31820 Toulouse, France

³INPT, CNRS, IMFT, Toulouse, France

⁴LNHE, EDF R&D, Chatou, France

⁵LNSH, EDF R&D, Chatou, France

⁶DREAL Champagne Ardenne, Châlons-en-Champagne, France

Abstract. Over the last few years, a collaborative work between CERFACS, LNHE (EDF R&D), SCHAPI and CE-REMA resulted in the implementation of a Data Assimilation (DA) method on top of MASCARET in the framework of real-time forecasting. This prototype was based on a simplified Kalman filter where the description of the background error covariances is prescribed based on off-line climatology constant over time. This approach showed promising results on the Adour and Mame catchments as it improves the forecast skills of the hydraulic model using water level and discharge in-situ observations. An ensemble-based DA algorithm has recently been implemented to improve the modelling of the background error covariance matrix used to distribute the correction to the water level and discharge states when observations are assimilated from observation points to the entire state. It was demonstrated that the flow dependent description of the background error covariances with the EnKF algorithm leads to a more realistic correction of the hydraulic state with significant impact of the hydraulic network characteristics.

1 Introduction

The present study was carried out in the framework of a collaboration between CERFACS, SCHAPI and LNHE (EDF). It deals with the simulation of river hydrodynamics, with a focus on environmental risk assessment related to flooding. It relies on the use of the MASCARET software that solves the mono-dimensional shallow water equations and is used by SCHAPI (Service Central d'Hydrométéorologie et d'Appui à la Prévision des Inondations) and SPCs (Services de Prévision des Crues) to build models for catchments of interest in France. In spite of the advances in numerical modelling and the expertise invested in the development and the use of MASCARET, the capacity for real-time anticipation of extreme flood events remains limited because of several sources of uncertainties in hydraulic models. In order to better forecast flood events, these uncertainties should be identified, quantified and reduced. To begin with, forcing data that represent hydrologic boundary conditions for hydraulic models usually result from the transformation of uncertain observed water levels into discharges with an uncertain rating curve or from discharges that are forecasted by uncertain hydrologic models. Another source of uncertainty is the description of the river channel and flood plain geometry. This requires on-site measurements of topographic and bathymetric profiles to provide a spatially

distributed geometry. Additionally, the model equations are based on simplifications and parameterizations of the physics. The parameterization schemes are calibrated to adjust the model behaviour to observed water levels or discharges, typically, through the calibration of friction coefficients. The calibration of the river bed and flood plain friction coefficients is usually achieved once for all using a batch of observations such as water levels from a limited number of flood events, thus providing time-invariant values for the model parameters. It is important to mention that errors in the model inputs and in the model equations are sometimes difficult to discriminate. These uncertainties usually translate into errors in the model representation of the water level-discharge (H-Q) relation that is not coherent with that from the reality and generally translate into an imperfect prediction of the hydraulic state in the river and flood plains. Following the path of deterministic forecast achieved at SCHAPI and SPCs, an error reduction approach was implemented in the framework of the collaboration between CERFACS, SCHAPI and LNHE. Previous work by [1] assumed that major sources of uncertainties relate to hydrologic forcing. The authors proposed to implement a data assimilation (DA) algorithm to reduce these errors and thus improve the flow simulation and forecast. [1] describe the improvement of the simulated hydraulic state assimilating water level and/or discharge observations in the context of operational flood forecasting. These preliminary

^a Corresponding author: barthelemy@cerfacs.fr

developments are gathered in a platform called DAMP (Data Assimilation for Mascaret Prototype). Water level data were sequentially assimilated using an Extended Kalman Filter (EKF) algorithm to control the upstream flow for the hydraulic network and dynamically correct the hydraulic state. The first step of the analysis is based on the assumption that the upstream flow can be adjusted using a simple three-parameter correction, and the second step consists in correcting the hydraulic state every hour (the observation frequency) to provide an improved initial condition for a forecast simulation. This procedure is applied on a sliding time-window over the entire period of each flood event for the Adour and Marne catchments; the results were interpreted for several events over each catchment. It was shown that the simulation with DA is significantly closer to the observation than the free run (no assimilation) over the re-analysis period as well as over the forecast period. It was also found that the sensitivity to an initial condition for the forecast mode is negligible compared to the sensitivity to the upstream flow, except at short forecast lead-time. Recent work by [2] provides a time-dependent correction of the river bed and flood plain friction coefficients to account for errors in the bathymetry that vary as water level reaches different sections of the described geometry. It is worth noting that the system presented in [1] and [2] is used operationally at SPC SAMA since 2014 and is currently being implemented at SCHAPI for operational flood forecasting. The present paper is based on [3] and it investigates the impact of the flow dynamics and the river geometry on the background error covariance functions in an Ensemble Kalman Filter (EnKF) algorithm with MASCARET for the Adour catchment. It was demonstrated that the correlation length-scales for the errors in the hydraulic state are large with a strong impact of the

$$\frac{\partial S}{\partial t} + \frac{\partial Q}{\partial x} = 0$$

$$\frac{\partial Q}{\partial t} + \frac{\partial QV}{\partial x} + gS \frac{\partial z}{\partial x} + J + J_s = 0 \text{ with } J = \frac{Q^2}{S^2 K_s^2 R_H^{4/3}}$$

where $S [m^2]$ is the river cross sectional area, $Q [m^3 s^{-1}]$ is the discharge, V is the flow velocity $[m s^{-1}]$, $q_a(x,t) [m^2 s^{-1}]$ is the lateral lineic discharge, $K_s [m^{1/3} s^{-1}]$ is the friction coefficient, R_H is the hydraulic radius, g is the gravity, J and J_s represent regular and singular head losses respectively. The river cross sectional area S is, for each location x , a function of the water level $H = Z(x,t) - Z_{\text{bottom}}(x,t)$, where $Z(x,t) [m]$ is the free surface height and where $Z_{\text{bottom}} [m]$ corresponds to the riverbed bathymetry. The unsteady kernel of MASCARET was used in this study.

1.2 Hydraulic modelling and test case catchment

The Adour maritime hydraulic network (Fig. 1) is located in South West France, close to the Atlantic Ocean. The section of the river network that is described in this work is 161 km long. 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 (available every 15 minutes) translated into discharges with a local rating curve established at the observing stations of Dax, Orthez, Escos and Cambo. Since

bathymetry and the confluences in the hydraulic network; they also vary in time especially due to the maritime influence. Finally, the EnKF leads to a correction of water level and discharge that is coherent with the model equations and it further improves the DA results when compared to those of the previously described EKF. The implementation of an ensemble-based DA algorithm thus appears as the next step towards a physically consistent DA system for flood forecasting.

The structure of the paper is as follows: Section 2 provides a brief description of the hydraulic model equations and application to the Adour catchment. The ensemble-based approach is presented in Sect. 3 as well as the elements for the implementation of this approach with a coupling software and MASCARET. In Sect. 4, the DA results are presented for a synthetic test case as well as over a set of selected real flood events. Conclusions and perspectives for this work are given in Sect. 5.

2 Hydraulic modelling

2.1 Shallow water equations

Along this hydraulic network, the 1D form of the shallow water equations is solved with the MASCARET [4] software developed by EDF-R&D and CEREMA (Centre d'Etudes et d'Expertise sur les Risques, l'Environnement, la Mobilité et l'Aménagement), widely used for modelling flood events, submersion waves resulting from the failure of hydraulic infrastructures, river control, and channel waves propagation. The 1D shallow equations read (non-conservative form):

the rating curves are built from a limited number of water level and discharge measurements and are usually extrapolated for higher flows, these are significant uncertainties related to these upstream boundaries. The downstream forcing is given by observed water level at the observing station of Convergent on the Atlantic Ocean's coast. Water level observations are available hourly at Lesseps, Urt, Pont-Blanc, Villefranque and Peyrehorade. It takes approximately 5 h for the upstream forcings to be propagated to the observing stations of Peyrehorade and Pont-Blanc, 10 h for Urt and 12 h for Lesseps. It should be noted that in forecast mode, the upstream hydrological forcings are set constant to the last observed value beyond this transfer time. Whereas in forecast the downstream forcing is given by the forecast water level computed by the SHOM (Service Hydrographique et Océanographique de la Marine). The Adour 1D hydraulic model is described with 548 topographic and bathymetric cross sections interpolated over 2,795 grid points. The river is represented as a 1D flow bounded with infinitely high banks except in the neighbouring of Peyrehorade, where a limited number of cross sections give a local description of the flood plains. The Gironde-Adour-Dordogne (GAD) SPC developed this model for operational purposes in collaboration with SCHAPI. The tidal influence of the Atlantic Ocean combined with the influence of the mountainous region of the Pyrenees result in a complex atmospheric and hydraulic dynamic over the Adour catchment. According to SCHAPI's

statistical records, the Adour catchment is ranked amongst the most challenging catchments in France due to a large number of orange and red alerts. The alerts are defined by water level thresholds at the stations of Cambo-les-Bains, Orthez, Dax, Escos and Peyrehorade. For example, the yellow, orange, and red thresholds at Peyrehorade are set to 2.1 m, 4.1 m, and 4.9 m respectively. Flood events on the Adour maritime hydraulic network can be categorized in three types. First, flood peaks occurring on reach 4 with a slow dynamic over 7 to 14 days and maximum discharge between 400 and 1000 $m^3 \cdot s^{-1}$. Second, flood peaks on reaches 6 and 7 resulting from flash floods event over 2 to 3 days with a maximum discharge between 800 and 1650 $m^3 \cdot s^{-1}$ for reach 6 and a maximum discharge between 400 and 1100 $m^3 \cdot s^{-1}$ for reach 7. Third, flood peaks on reach 3 that are eventually correlated in time with flood peaks on reaches 6 and 7 and last 2 to 3 days. These events might occur simultaneously and may be worsened by tidal influence.

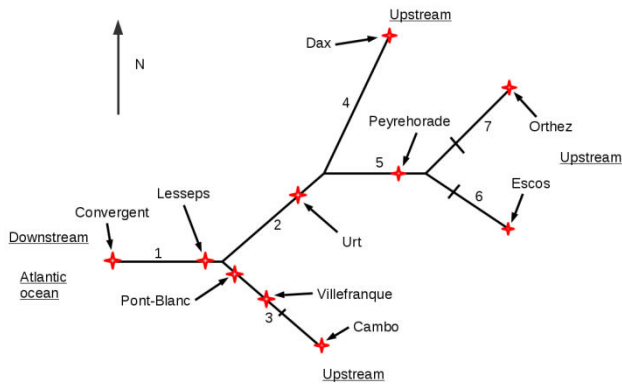


Figure 1. Hydraulic network scheme of the Adour river network simulated with MASCARET with reach numbering. Water level observing stations are represented with red crosses. Dams on reaches 3, 6 and 7 are represented with black markers.

3 Ensemble methods for uncertainty reduction

3.1 Monte-Carlo approach and statistics estimation

The Monte Carlo (MC) approach allows for propagating the uncertainty sources through the physical model M by running the numerical code N_e times independently with different sets of input variables, randomly generated from their assumed probability distributions. This provides a sample of deterministic solutions, which are used to infer the statistical properties of the output distribution. MC methods are straightforward and robust techniques that allow for the treatment of nonlinearities and the computation of the whole pdf of model outputs. As a drawback, MC simulations are computationally demanding, especially when dealing with high quantiles as it is the case in hydraulics for flood events. In order to achieve statistically reliable results, the number of runs should be large enough to reach the convergence; hence the application of classical

MC techniques is limited to non-expensive numerical models. An extension of the MC approach is the generalized likelihood uncertainty estimation (GLUE) procedure, introduced by [5] and widely used in hydrology. The analytical treatment of the problem is also a less expensive approach, but is limited to simple relationships between model inputs and outputs, and thus less applied in real life problems. An alternative approach is to build a surrogate model for the physical model. When Monte-Carlo based methods are used to sample the prior distribution, one idea is to use a surrogate partial differential equations model and thus compute efficiently the statistics of the quantities of interest [6]. Numerical methods to build a surrogate issued from spectral-based representations entitled Polynomial Chaos (PC) and introduced by [7] are very promising in terms of precision and cost reduction [8] as long as the input uncertainty remains reasonably correlated and the numerical model weakly non-linear. While the reduced model issue is beyond the scope of this study, the MC approach is used in the following to estimate the model state error covariance matrix for the DA algorithm.

3.2 EnKF algorithm

The DA algorithm used in this study is the Ensemble Kalman filter algorithm commonly referred to as EnKF [9]. The control vector is composed of the hydraulic state described by the discretized water level and discharge over the domain. The EnKF decomposes in an analysis step and a forecast step that are sequentially applied and presented in Fig. 2 for the analysis cycle at time i .

The EnKF relies on the integration of an ensemble of N_e perturbed members (indexed by the subscript k) and on the assumption that the stochastic estimate of the ensemble statistics is a fair representation of the model state error statistics. For an analysis at time i , the background error covariance matrix B_i is stochastically estimated following Eq. (1)

$$B_i = \frac{1}{N_e - 1} \sum_{k=1}^{N_e} (x_i^{b,k} - \hat{x}_i^b)(x_i^{b,k} - \hat{x}_i^b)^T \quad (1)$$

where $\hat{x}_i^b = \frac{1}{N_e} \sum_{k=1}^{N_e} x_i^{b,k}$ corresponds to the ensemble mean (members are indexed with k) and T stands for the transition operator. Then the Kalman gain K_i is computed with $K_i = B_i H^T (H B_i H^T + R)^{-1}$ where H is the tangent linear of the observation operator H that maps the control vector onto the observation space and R is the observation error covariance matrix. For each member of the ensemble over the i -th assimilation cycle, the analysis step consists in assimilating a perturbed observation vector $y_i^o + \varepsilon_i^{o,k}$ (with $\varepsilon_i^{o,k}$ a Gaussian noise with zero mean) [10], to correct the background vector $x_i^{b,k}$ using the classical Kalman filter update equation:

$$x_i^{a,k} = x_i^{b,k} + K_i (y_i + \varepsilon_i^{o,k} - H(x_i^{b,k})) \quad (2)$$

It should be noted that the perturbation of the observation in Eq. (2), $y_i^o + \varepsilon_i^{o,k}$, allows for maintaining the spread of uncertainty within the ensemble and avoiding the divergence of the filter. As illustrated in Fig. 2 at time $(i-1)$, the analysed states are propagated forward in time by the model $M_{i-1,i}$ to provide an ensemble of background states for the next assimilation cycle (here, the model error is supposed to be negligible):

$$x_i^{b,k} = M(x_{i-1}^{a,k}) \quad (3)$$

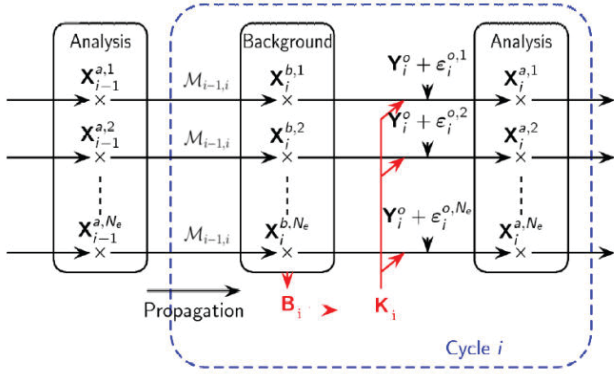


Figure 2. Schematic representation of the EnKF algorithm for the analysis cycle i over the time period $[i-1, i]$.

In the present study, as the control state is composed of the discretized hydraulic state in water level and discharge denoted by (Z, Q) . The matrix \mathbf{B} decomposes into 4 symmetrical sub-matrices that correspond to univariate and multivariate covariances. In the following, we denote \mathbf{B}_{ZZ} , B_{ZZ} and \mathbf{B}_{QQ} , B_{QQ} the univariate water level and discharge background error covariance sub-matrices and \mathbf{B}_{ZQ} and \mathbf{B}_{QZ} , B_{QZ} the multivariate water level/discharge background error covariance sub-matrices with $\mathbf{B}_{ZQ} = \mathbf{B}_{QZ}^T$, $B_{ZQ} = B_{QZ}^T$. Thus, \mathbf{B} can be written as:

$$B = \begin{pmatrix} B_{ZZ} & B_{ZQ} \\ B_{QZ} & B_{QQ} \end{pmatrix} \quad (4)$$

The water levels at the observing stations are translated into a water level correction that is spread over the entire hydraulic network thanks to the univariate covariance function for water level and into a discharge correction thanks to the multivariate covariance function.

The representation of the ensemble is thus a key issue for the EnKF. In this study we consider that the major source of uncertainty in the simulated water level and discharge results from an approximate knowledge of the hydrology of the catchment that provides upstream and lateral boundary conditions to the hydraulic model. As a consequence the ensemble is generated by adding a Gaussian noise to the observed forcings. The river bed friction coefficient could have also been perturbed so as to represent an error in the model hydraulic parameters within the ensemble. A sensitivity analysis on the spatialized friction coefficient for

the Adour river that is under maritime influence is on going in the framework of complementary studies.

Downstream the hydraulic network, at the observing station of Convergent, the water level is perturbed with a Gaussian noise characterized by a zero mean and temporal correlation length scale of $6 h$ that is coherent with the tidal cycle. The observed upstream forcings are perturbed with a Gaussian noise characterized by a temporal auto-correlation length-scale of $4 h$ that was estimated with synthetic experiments. The amplitude of the perturbation is set equal to 15% of the discharge so that the error increases with the discharge value to traduce the uncertainty on the rating curve for high flow. An illustration of the upstream forcing perturbation for 3 members of the ensemble at Escos is given in Fig. 3.

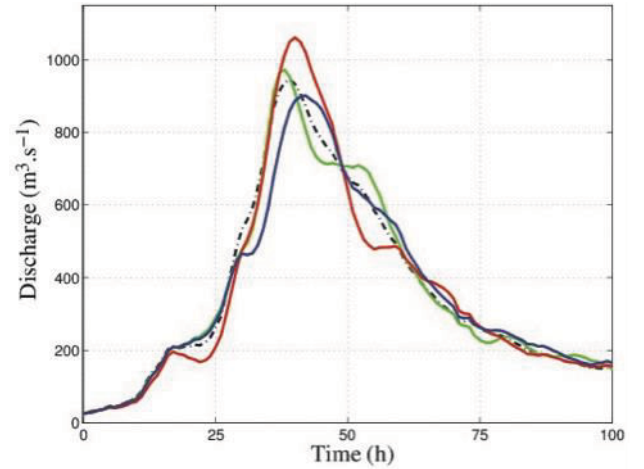


Figure 3. Time evolving profile of the upstream forcing on reach 6 during 2011 flood event: the black dashed line corresponds to the original forcing; green, red and blue lines correspond to perturbed forcings.

A common drawback of the EnKF is that when the size of the ensemble is limited, it tends to be under-dispersive and may diverge over time ignoring the observed information. [11] proposed to artificially increase the dispersion within the ensemble by introducing a multiplicative inflation factor to the ensemble anomalies that increases the model state error variance at the observation point. According to [12], the difference between each member of the ensemble $x^{b,k}$ and its mean can be inflated using a time-varying inflation factor $\lambda > 1$ thus defining a new set of background states $\varkappa^{b,k}$ as well as a new set of anomalies over which the background error covariance matrix can be computed:

$$\varkappa^{b,k} - \hat{x}^b = \lambda(x^{b,k} - \hat{x}^b). \quad (5)$$

The inflation factor is derived from the consistency criterion presented in [13] based on the assumption that the errors in the background and observation error matrices are uncorrelated and properly described so that:

$$d_{o-b} d_{o-b}^T \geq HBH^T + R. \quad (6)$$

Considering only one observation is assimilated, covariances matrices are reduced to scalars; $HBH^T = \sigma_b^2$

and $R = \sigma_o^2$. If Eq. 6 does not hold and $\lambda^2 = d_{o-b}d_{o-b}^T > -\sigma_o^2 > \sigma_b^2$

then λ is specified as follows:

$$\lambda^2 = \frac{d_{o-b}d_{o-b}^T > -\sigma_o^2}{\sigma_b^2} \quad (7)$$

and the inflated background error variance at the observation point is increased to $\sigma_b^2 = \lambda^2 \sigma_b^2$. The inflation is then applied from the observation point to the entire spatial domain (grid points are indexed with j) in coherence with the shape of the univariate correlation function C at the observation point using

$$\begin{aligned} \hat{x}^{b,k}(j) - \hat{x}^b(j) = \\ 1 + (\lambda - 1) \vee C(j) \vee (x^{b,k(j)} - \hat{x}^b(j)). \end{aligned} \quad (8)$$

This formulation ensures that

$1 + (\lambda - 1) \vee C(j) \vee > 1$, implying that the mean of the ensemble over the domain and the background error correlations over the domain are preserved and that the background error covariances are locally increased in the neighboring of the observation point. When N_o observations distributed over the hydraulic network are assimilated, the observation error matrix R is diagonal. The inflation equation now reads:

$$1 + \sum_{n=1}^{N_o} (\lambda_n - 1) \vee C_n(j) \vee (x^{b,k(j)} - \hat{x}^b(j)). \quad (9)$$

$$\hat{x}^{b,k}(j) - \hat{x}^b(j) =$$

where $\sum_{n=1}^{N_o} (\lambda_n - 1) \vee C_n(j) \vee$ is larger than 1 if $\lambda_n > 1$ for all observation points ($n = 1, \dots, N_o$). The resulting algorithm is called the Inflated Ensemble Kalman Filter denoted by IEnKF in the following.

3.3 Implementation with MASCARET

The EnKF DA algorithm is implemented with the OpenPALM [14] software. It is an open-source dynamic code coupler that has been jointly developed at CERFACS and ONERA since 1996. OpenPALM has been originally designed for DA algorithm for operational oceanography forecasting; it has now reached a high-degree of maturity and stability, with applications ranging from operational DA (oceanography, atmospheric chemistry, hydrology) to industrially oriented multi-physics modelling (fluid-structure interaction, combustion-acoustics interaction). OpenPALM provides a straightforward parallel environment based on high performance implementation of the Message Passing Interface standard (i.e., MPICH, OpenMPI, LAM/MPI). This interface is able to perform both data parallelism (i.e. simultaneous execution on multiples cores of the same code component for a unique data set) and task parallelism (i.e. simultaneous execution on multiples cores of multiple tasks across the same or different data sets) which is particularly well adapted for ensemble-based algorithms where each model integration can be achieved

independently of the others. The PALM_PARASOL functionality in OpenPALM particularly addresses this need. It is based on the master/slaves principle, and allows for an efficient management of memory and processor allocation issues according to available resources. The different units are organised thanks to the IHM interface named PrePALM. The flowchart for the EnKF algorithm is presented in Fig. 4. In the present work, OpenPALM is used as a task parallelism manager to handle communications and data exchanges between MASCARET. The different mathematical units (here coded in Fortran) required to perform the EnKF are sequentially achieved with a PALM Branch. The first unit *read_damocles_EnKF* is in charge of setting-up the ensemble, in the present case, the hydrological forcing. This information is then sent to the *master_appel_mascaret* unit by a PALM communication and the ensemble members of MASCARET are run simultaneously on the available processors. The Master processor of PALM_PARASOL spawns multiple copies of MASCARET (the slaves), each on one or several processors with a different set of input parameters (hydrological forcing in the present case). The ensemble outputs are thus post-processed by the *make_cov* unit that provides a stochastic estimate of the background error covariances in B . The synthetic observations are generated by the unit *generate_obs*. Both the observation vector and the B matrix are communicated to the analysis unit *analyse_BLUE* that achieves the DA analysis step. This sequence is iterated over the time when observations are available.

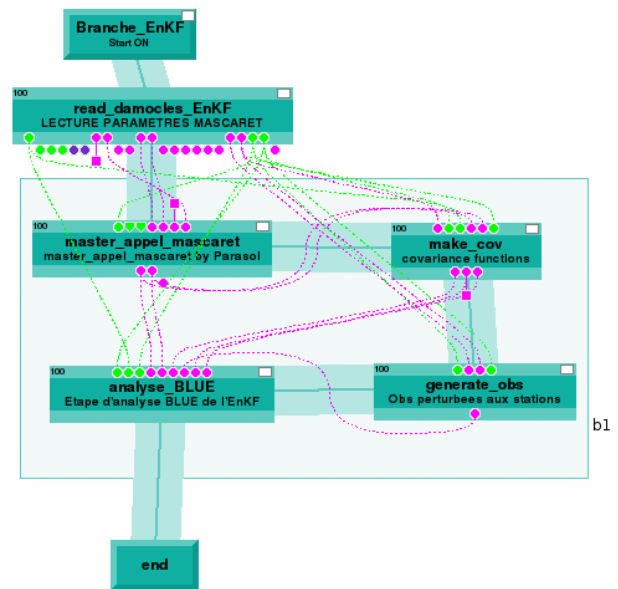


Figure 4. OpenPALM scheme for the EnKF. Each rectangle (called unit) corresponds to a Fortran subroutine associated to a part of the EnKF algorithm (*master_appel_mascaret* propagates the water line, *analyse_BLUE* corresponds to the analysis step...). The lines that link the units are data exchanged between the subroutines.

4 Results

In the following, the merits of the EnKF DA algorithm are first highlighted in the framework of an Observing System Simulation Experiment (OSSE) denoted by SE. Then, it is applied on a set of real experiment (denoted by RE), where in-situ water level observations are assimilated. In SE, the water level observations are generated from the integration of a reference run using the 2011 flood event hydrological forcing, in addition with an artificial lateral inflow injected downstream of the dam on reach number 6. This inflow is correlated in time with the upstream forcings on reaches 6 and 7, i.e. this is a random inflow with the same correlation length-scale as the upstream forcing (4 h) and a maximum discharge of $575 \text{ m}^3 \text{ s}^{-1}$ at the flood peak. Synthetic data result from the extraction of simulated water level at observing time and location on top of which a Gaussian noise is added to represent the observation error. In both SE and RE cases, the ensemble members are integrated using perturbed upstream forcings only (no lateral inflow).

4.1 Observing System Simulation Experiments Results

In the following the ensemble mean for water level and discharge are displayed. Synthetic water level observations are sequentially assimilated at Peyrehorade. The water level results of the sequential application of the IEnKF for the 2011 flood event in the framework of SE are shown in Figs. 5-6 for Peyrehorade and Urt respectively. The discharge results are displayed in Figs. 7-8.

It should be noted that at the time of analysis, the water level at both observing stations is brought closer to the observations than the free run (without assimilation). The correction from the DA algorithm at Peyrehorade is translated into a correction over the entire network by the univariate covariance function in **B**, i.e. at Urt. Thanks to the multivariate covariance function between water level and discharges in **B**, the discharge is also significantly improved at the observing stations of Peyrehorade and Urt.

Previous studies by [1] implemented an invariant formulation of the background error covariance matrix for the control of the hydraulic state. In this formulation, the shape of the background error correlation function is prescribed with two half-Gaussian functions, with shorter correlation length scale upstream of the observation point than downstream. The length scales were estimated in the framework of an idealized study and the assumption was made that they were not varying neither in time nor due to the geometry of the river. With the IEnKF, the covariance functions are stochastically estimated amongst the members so that they are coherent with the dynamic of the flow as well as with the hydraulic network characteristics. In the following, the covariance functions are illustrated for the 2011 flood event before the flood peak in Figs. 9-10 respectively for univariate (Z,Z) and multivariate (Z,Q) statistics both associated with the observing station in Peyrehorade. These functions are plotted along the reaches 6-5-2-1. The Peyrehorade station is represented with a vertical red dashed line. The vertical black dashed lines represent the separation between two reaches (that correspond to confluences) and the vertical blue dashed line

represents the position of the dam on reach 6. The IEnKF water level univariate covariance function shows important discontinuity where the geometry of the river varies, for instance at the location of the dam on reach 6 as well as upstream of the network where the bathymetry features discontinuities. Downstream of the dam, the Adour river is under tidal influence, the covariance function is smooth and decreases towards the imposed maritime boundary condition defined by the observed water level. The IEnKF water level/discharge multivariate covariance function presents a discontinuity at each confluence between reaches since discharge is an additive variable. Thanks to the flow dependent description of the **B** matrix, the IEnKF algorithm provides an optimal and spatially distributed correction in water level and discharge. As the error correlation length scales are large, the impact of a local observation is spread onto the entire hydraulic network both for water level and discharge. The corrected hydraulic state is thus used as initial conditions for further forecast.

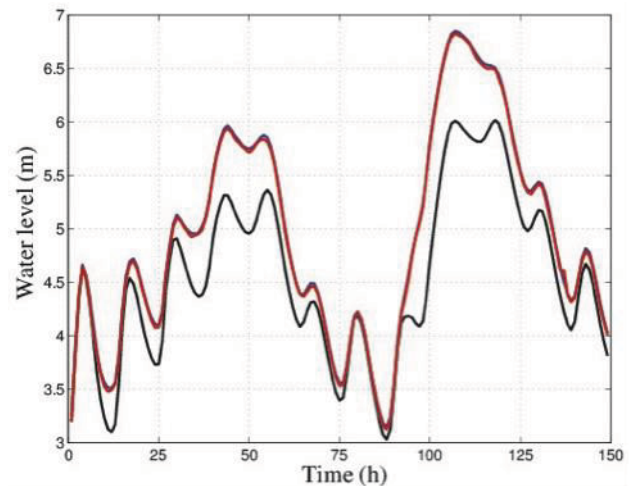


Figure 5. Water level at Peyrehorade at the time of analysis for SE, for 2011 flood event. The observations are represented with the blue line, the free run is represented with the black line and the analysis is the red line (on top of the blue line).

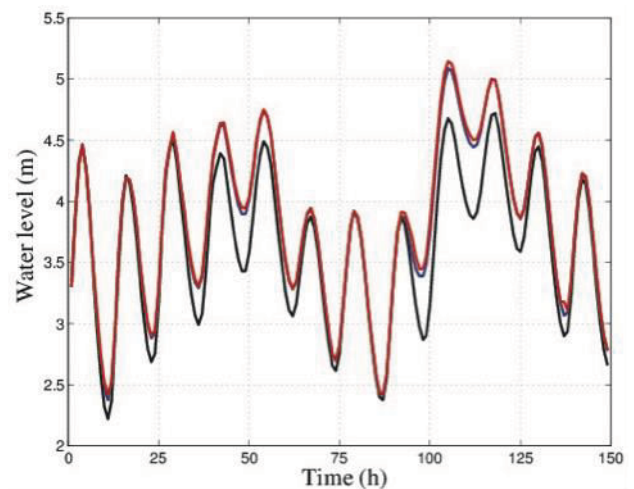


Figure 6. Same caption as Fig. 5 but for the observing station of Urt (SE).

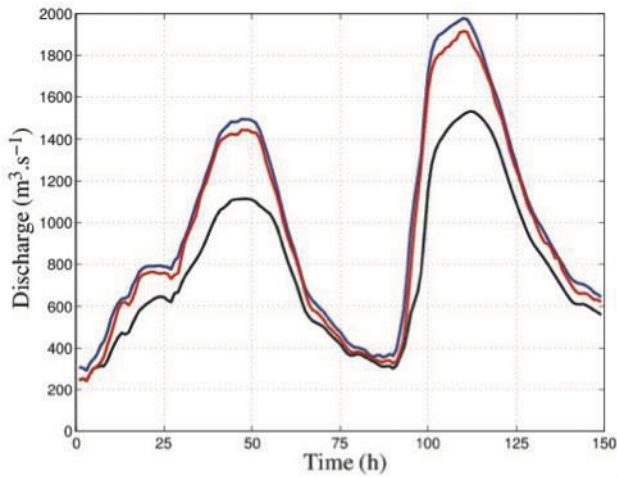


Figure 7. Same caption as Fig. 5 but for the discharge at Peyrehorade (SE).

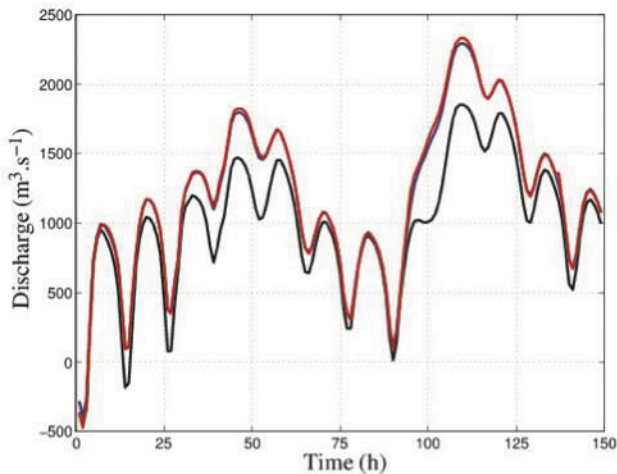


Figure 8. Same caption as Fig. 5 but for the discharge at Urt (SE).

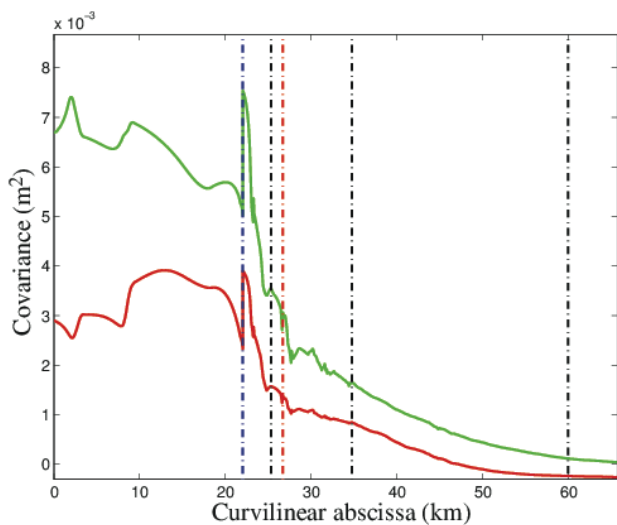


Figure 9. Representation of the water level univariate background error covariance functions associated to Peyrehorade computed with the IEnKF along reaches 6-5-2-1 for two different simulation times. The vertical red dashed line represents the Peyrehorade observing station. Vertical black dashed lines represent the separation between two reaches and the vertical blue

dashed line represents the position of the dam on reach 6.

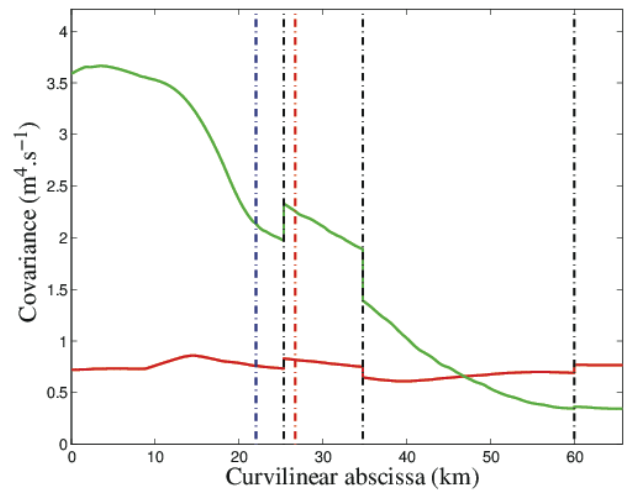


Figure 10. Same caption as Fig. 9 for the multivariate water level/discharge background error covariance functions.

Figures 11-12 respectively present the 2-*h* and the 4-*h* forecast water level at Peyrehorade and Urt respectively. While there is a significant improvement for short-range forecast (1-*h* to 3-*h* forecast), the improvement resulting from DA decreases as the forecast lead-time increases. Indeed, the impact of the correction of the hydraulic state is limited in time and the analysed water level drifts back towards the free run as other sources of uncertainties remain uncorrected by the EnKF. This highlights the need for the extension of the control vector to model parameters and hydrological forcing to improve medium to long forecasts (3-*h* to 6-*h* forecast and 6-*h* to 24-*h* forecast respectively).

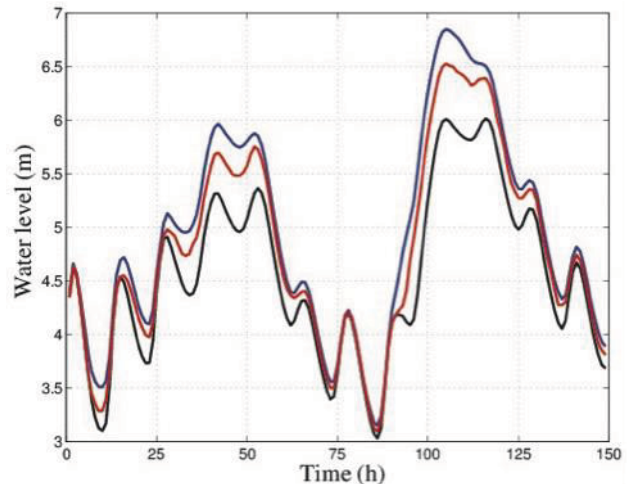


Figure 11. Forecasted water level at Peyrehorade for SE for a 2-*h* lead-time. The observations are represented with the blue line, the free run is the black line and the analysis is the red line.

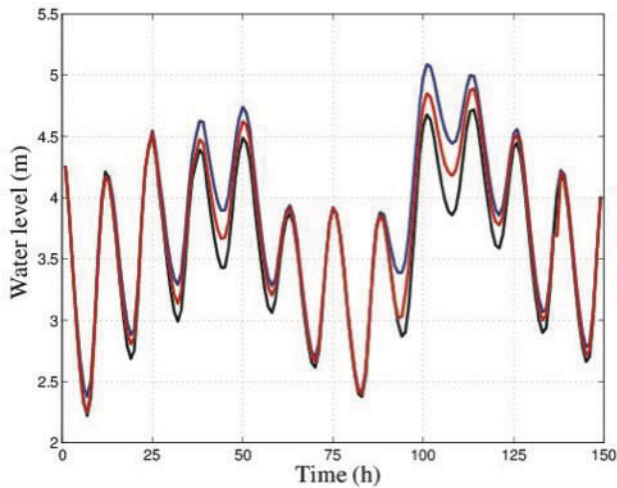


Figure 12. Forecasted water level at Urt for SE for a 4-*h* lead-time. The observations are represented with the blue line, the free run is the black line and the analysis is the red line.

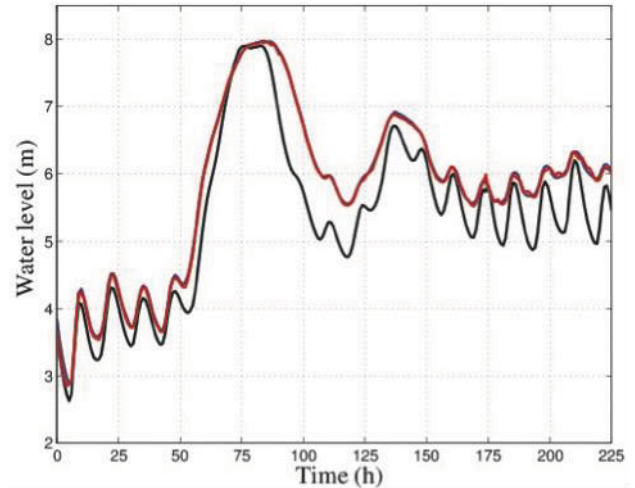


Figure 13. Water level at Peyrehorade at the time of analysis for RE for the 2014 flood event. The observations are represented with the blue line, the free run is the black line and the analysis is the red line.

4.2 Real data experiment results

In the context of RE, the EnKF is applied to a set of eight flood events on the Adour catchment from 2009 to 2014. Water level observations are assimilated hourly at Peyrehorade and Pont-Blanc. In the following illustrations are given for the 2014 flood event.

The results of the sequential application of the EnKF for the 2014 flood event are displayed in Fig. 13 at Peyrehorade and Fig. 14 at Urt for water level. At the time of analysis in Fig. 13, the assimilation leads to excellent results as previously observed for SE. At Urt, the water level is improved except for high flow including at the flood peak that is over-estimated. At this location, the improvement is less obvious than it was for SE as model error now limits the impact of the DA algorithm. During the beginning of the flood event, the lack of flood plain modelling in the vicinity of Urt is penalizing and the infinitely high banks assumption in the 1D model results in an erroneous dynamic of the water line that causes errors in the water level covariance functions and leads in turn to an inappropriate correction of the water line. Thus, in spite of the assimilation of observed water level at Peyrehorade, the simulated water level at Urt remains over-estimated for high flow on reaches 6 and 7. After the flood peak, for medium flow, the 1D model is realistic and the simulated water level is improved at Urt. It should be noted that this problem does not appear in SE as the model error is present both in the reference run and in the ensemble runs.

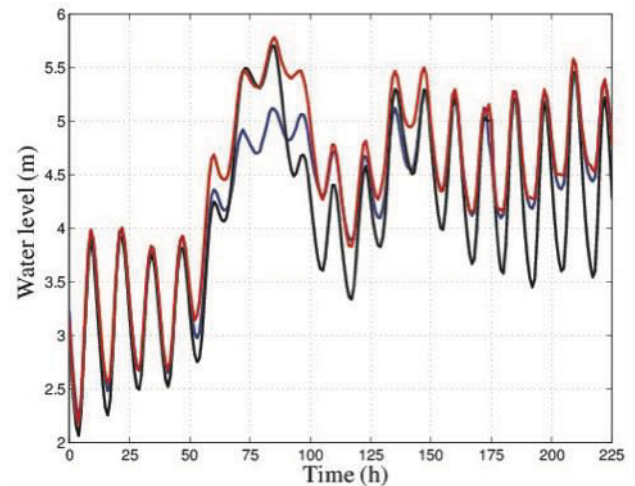


Figure 14. Same caption as Fig. 13 but for Urt.

Table 1 summarizes the improvement in the mean RMSE with DA computed over eight flood events for forecast lead times ranging from +0 *h* to +12 *h* depending on the observing station on the hydraulic network. DA of observations at Peyrehorade and Pont-Blanc significantly improves the RMSE for short to medium lead-time for water level at these locations. The water level is also improved at Lesseps where no observation is assimilated. However, the impact at Urt is close to 0 on average due to model errors in this area.

Observing point	Forecast lead times (<i>h</i>)	RMSE
Peyrehorade	0	87.91
	1	74.18
	3	40.59
	6	12.08
Urt	0	-3.4
	2	5.9
	5	5.61
	10	-0.99

Lesseps	0	30.33
	2	11.01
	6	0.77
	12	-1.36
Pont-Blanc	0	70.95
	1	45.44
	3	10.2
	6	3.8

Table 1. Mean improvement of the RMSE at the stations of Peyrehorade, Urt, Lesseps and Pont-Blanc for increasing forecast lead times for RE; statistics computed for the eight flood events between 2009 and 2014.

5 Conclusions

This study describes the application of the IEnKF algorithm on the "Adour maritime" hydraulic network using the 1D hydrodynamic code MASCARET. The EnKF is an ensemble-based DA algorithm in which the background error covariance statistics are stochastically and accurately estimated using several model trajectories based on perturbed upstream discharges. An inflation factor is applied to preserve the spread of the ensemble over the assimilation cycles. The model state functions are found to be closely related to the network geometry and characterized by a large spatial extent. They are also time dependant mostly due to the time varying description of the upstream and downstream forcing along the flood event but also eventually due to the variability of the hydraulic parameters and geometry of the river. It is demonstrated on a synthetic experiment as well as on a real experiment that the IEnKF allows for the correction of both water level and discharge at the observing stations and all over the entire network and provides better results, in terms of analysed and forecast water level and discharge for short forecast lead-times (1-3 h), compared to a free run without assimilation. In conclusion, the use of various techniques of data assimilation considered in this article could improve the understanding of hydraulic processes, the mapping of flooded areas, and flood forecasting on www.vigicrues.gouv.fr.

6 References

- Ricci S., Piacentini A., Thual O., Le Pape E. and Jonville G., (2011). Correction of upstream flow and hydraulics state with a data assimilation in the context of flood forecasting. *Hydrol. Earth Syst. Sci* **15**, 1-21.
- Habert J., Ricci S., Thual O., Le Pape E., Piacentini A., Goutal N., Jonville G., Rochoux M., (2016). Reduction of the uncertainties in the water level-discharge relation of a 1D hydraulic model in the context of operational flood forecasting. *Journal of Hydrology*, **532**, 52-64.
- Barthélémy S., PhD (2015) Assimilation de données ensembliste et couplage de modèles hydrauliques 1D-2D pour la prévision des crues en temps réel. Application au réseau hydraulique Adour maritime, Université de Toulouse, France.
- Goutal, N. and Maurel, F., (2002). A finite volume solver for 1D shallow water equations applied to an actual river. *Int. J. Numer. Meth. Fluids*, **38 (2)**, 1-19.
- Beven, K., Freer, J., 2001. A dynamic topmodel. *Hydrol. Process.*, **15**, 1993-2011.
- Boyaval, S. (2012). A fast Monte-Carlo method with a Reduced Basis of Control Variates applied to Uncertainty Propagation and Bayesian Estimation. *CMAME*, 241-244.
- Wiener P. (1938). The Homogeneous Chaos. *American Journal of Mathematics*. Vol. **60(4)**, 897-936.
- Ghanem R. and P. Spanos (1991). *Stochastic finite elements, A spectral approach*, Dover.
- Evensen G. (1994). Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *Journal of Geophysical Research*, **99(C5)**, 10143-10162.
- Burgers G., Jan Van Leeuwen P. and Evensen G. (1998). Analysis Scheme in the Ensemble Kalman Filter. *Monthly Weather Review*, **126(6)**, 1719-1724.
- Anderson J. L. and Anderson S. L. (1999). A Monte Carlo Implementation of the Nonlinear Filtering Problem to Produce Ensemble Assimilations and Forecasts, *Monthly Weather Review*, **127**, 2741-2758.
- Anderson J. L. (2007). An adaptive covariance inflation error correction algorithm for ensemble filters, *Tellus, Series A, Dynamic Meteorology and Oceanography*, **59**, 210-224.
- Desroziers G., Berre L., Chaonik B. and Poli P. (2005). Diagnosis of observation, background and analysis-error statistics in observation space. *Quarterly Journal of the Royal Meteorological Society*, **131(2005)**, 3385-3396.
- Buis S., Piacentini A. and Déclat D. (2006). PALM: a computational framework for assembling high-performance computing applications. *Concurrency and Computation: Practice and experience*, **18(2)**, 231-245.
- Li H., Kalnay E. and Miyoshi T. (2009). Simultaneous estimation of covariance inflation and observation errors within an Ensemble Kalman Filter. *Quarterly Journal of the Royal Meteorological Society*, **135(February)**, 523-533.