Front type implementation for flood extent and flood modelling image assimilation

Q. Bonassies[1](quentin.bonassies@cerfacs.fr), S. Ricci[1], T. H. Nguyen[1], A. Piacentini[1], M. Sadki[1], C. Fatras[2], A. Andral[2], S. Pena Luque[3], and R. Rodriguez Suquet[3]

[1] CECI, CERFACS/CNRS UMR 5318, Toulouse, 31057, France

[2] Collecte Localisation Satellites (CLS), 31520, Ramonville Saint-Agne, France

[3] Centre National d'Etudes Spatiales (CNES), Toulouse, 31401, France

***Abstract* – Flood forecasting with 2D hydrodynamic model is a powerful tool for risk assessment and decision makers as they provide water elevation maps over a computational domain. However, those models have uncertainties inherent to them due to the data used to construct them. Data assimilation (DA) is an efficient way to reduce these uncertainties. The assimilation of flood extent maps, derived from synthetic aperture radar (SAR) observations or, in the near future, 2D water surface elevation (WSE) observed from the Surface Water and Ocean Topography (SWOT) satellite is a hot topic as these data provide information in the floodplain. In this work, we propose an innovative framework to directly assimilate the location and the shape of the flood extent, by treating these data as interfaces (front-type) information. This approach allows to overcome the limitations of classical amplitude error correction, therefore allowing for the correction of flood edge position and/or deformation errors. To deal with complex front topology, an object-oriented approach based on the Chan-Vese (CV) contour fitting functional typically used in image processing was proposed. This study demonstrates the capability of the CV measure in a context of 2D flood forecasting to formulate the discrepancy between observed and simulated flood extents, used for DA analysis based on an ETKF algorithm for sequential parameter estimation. An OSSE experiment on a small toy model is presented to show the merit of the CV distance implemented in an ETKF algorithm.**

Keywords: Chan-Vese, Front Data, Data Assimilation, Remote Sensing, TELEMAC-2D, OSSE

I. INTRODUCTION

Between 1995 and 2015, floods accounted for 43% of the global natural hazards making it one of the most devastating and frequent weather-related disasters worldwide. Estimating and reducing the risk of flooding has become essential and attracted significant attention in the hydrometeorological and remote sensing (RS) communities, as well as the public and private sectors [1]. Flood studies aim at mitigating the impacts of flooding on population, as they allow flood management services to better identify flood-prone areas and reduce socio-economic losses with timely and reliable alerts. Moreover, the world population reached 8 billion in 2022 and is still in constant growth. Associated urbanisation growing is often located in floodplains and inundation-prone areas. Thus, increasing vulnerability to flood events and putting pressure on operational flood forecasting services [2]. Several international initiatives have joined efforts in forecasting and monitoring river hydrodynamics, in order to provide Decision Support Systems with accurate flood forecasting capability. Risk assessment relies on Earth Observations (EO) and river hydrodynamic software. The latter solves Shallow Water (SW) equations that are used to predict water surface elevations (WSE) and discharge in the riverbed or floodplains, further used to assess flood risks.

Hydrodynamic models are powerful tools for flood forecasting systems as they provided a fine resolution description of the flow in time and space. However, various sources of uncertainties in the model and in the inputs, for instance, hydrological forcing, initial conditions, model parameters and structures, translate into uncertainties in the model outputs. Data assimilation (DA) reduces these uncertainties by combining the model forecasts with various types of observations, such as in-situ gauge and satellite EO data. In the present work, we use the TELEMAC-2D (T2D) software (www.opentelemac.org), implemented for a simplified test channel case characterized by a narrow parabolic-shaped bathymetry, a slightly inclined topography, and an artificial hill located at mid-length of the channel. An ensemble-based DA filtering algorithm, namely the Ensemble Transformed Kalman Filter (ETKF) is used to sequentially combine observed flood extents, as wet/dry interface position, with T2D WSE simulated maps to improve flood forecasts. The ETKF is here favoured with respect to the Ensemble Kalman Filter (EnKF). Indeed, the computation of the covariance matrices is formulated in the ensemble subspace instead of in the state or observation subspace. A localisation feature is added to the ETKF using a geolocalized polygon that includes the ensemble of front lines. This DA algorithm is carried out within the context of an Observing System Simulation Experiment (OSSE). Synthetical flood extents, which mimic the flood extent maps derived from synthetic aperture radar (SAR) images, are assimilated to account for the error in the upstream inflow. The *a priori* water inflows differ from the true inflow, and the ETKF analysis aims at bringing the *a priori* value closer to the true and consequently bring the simulated fronts closer to the synthetical fronts.

Over the past decades, the literature on DA for hydrodynamic models mainly focused on the assimilation of in-situ or RS-derived WSE observation (*e.g.* [3], [4]). This stems from the fact that WSE is a state variable in any hydraulic model, thereby rendering the DA more straightforward. WSE maps can be derived from SAR images, by combining RS-derived flood extent maps and topography data, yet such

methods still require further research (*e.g.* [5]). Alternatively, fully automated methods for mapping flood extents from satellite images are described in the literature (*e.g.* [6], [7]). Machine learning algorithms, for instance integrated in FloodML, allow to retrieve flood extents from SAR images ([12]). Recent research works have developed methods to directly assimilate RS-derived flood extents, water masks or probabilistic flood maps, in hydraulic models and flood forecasting chains (*e.g.* [8], [9] & [10]), thereby paving the way for operational applications. Cooper *et al.* [11] proposed an observation operator that directly considers the SAR backscatter values as observations in order to bypass the flood edge identification or flood probability estimation processes. Another approach stands in the computation and the assimilation of wet surface ratio (WSR) in floodplain areas from SAR-derived flood extents [13]. Satellite SAR imagery data provides flood extent information with a large coverage from the observed backscatter (BS) values. Satellite SAR systems are capable of monitoring flood events with day-and-night imaging ability, regardless of weather conditions. In addition, the recently launched wide-swath radar interferometry SWOT satellite is able to compute water surface elevation (WSE) maps from its interferometric observations. In this paper, we propose a framework for the direct assimilation of flood extent location and shape (derived from Sentinel-1 SAR images), considering the wet-dry interface as an innovative information. Such front-type information aims to overcome the limitations of classical amplitude error correction and allows for the correction of flood edge position and/or deformation errors.

In several research fields, such as wildfire modelling ([15], [16] & [17]), tumour growth monitoring ([18], [19]) and oil spill detection and surveillance ([20]), DA algorithms have been developed to treat images as front-type information. Such approach was developed by Rochoux *et al.* ([16]) for data-driven surface wildfire propagation at a regional scale, based on a fire front mapping and shape recognition. In this work, the position of markers, *i.e.* the finite set of points positioned along the observed and the simulated fire front, are matched and compared. The resulting discrepancies were used in an ETKF algorithm to reduce the errors in the propagation model parameters and state. This prototype was evaluated with a synthetic and controlled grassland fire experiment. However, it was shown that this marker-based method was suboptimal for complex front geometry due to the computation of Euclidean distances between markers. Indeed, to deal with complex front topology, an object-oriented approach derived from the Chan-Vese (CV) contour fitting function used in image processing ([14]) was proposed in [15]. This front shape similarity measure was implemented and evaluated for a field-scale experiment in [17]. A similar idea is here applied for the assimilation of SAR-derived flood extent in the context of flood modelling. However, it should be noted that the comparison of non-hydrometric observations, *e.g.* flood edge locations, or flood probability measures, with the model outputs is not straightforward and requires the development of appropriate observation operators. In the present study, the ETKF algorithm is implemented for an OSSE experiment using a 2D hydrodynamic model as forward model. RS-derived front-type data are assessed with respect to simulated flood extent using the CV metrics.

II. Materials and Methods

A. TELEMAC-2D solver

Free-surface hydraulic modelling is principally governed by the Shallow Water equations (SWE, also known as Saint-Venant equations derived from Navier-Stokes Equations), which express mass and momentum conservation averaged in the vertical dimension. In this work, the hydrodynamic numerical model TELEMAC-2D is used to simulate and predict the water level (denoted by H [m]) and velocity (with horizontal components denoted by u and v [m/s]) from which flood risks can be assessed. It solves the SWE with an explicit first-order time integration scheme, a finite-element scheme and an iterative conjugate gradient method. A complete description of the underlying theoretical approach is provided in [21]. At each point within the mesh representing the model topography and bathymetry (for mesh nodes in the river channel), the results of the simulation are water height and velocity averaged over the vertical axis.

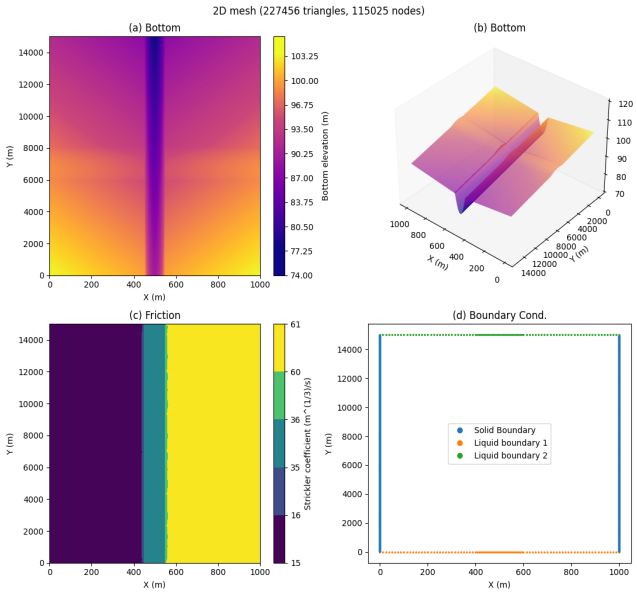


Figure 1. Topography and bathymetry of the mesh (a and b), Friction coefficients (c) and boundary condition (d). For illustration purposes, the x-axis and y-axis are not proportionally scaled.

B. TELEMAC-2D test case

A 15km-long parabolic-shape channel with a uniform slope and simplified floodplain test case was set for our work (Figure 1). The riverbed is characterized by a parabolic geometry (10m maximum depth). The catchment has a constant slope of 0.8% from upstream to downstream (x-axis). The floodplain has a uniform slope of 8% in the perpendicular direction to the river centerline (y-axis). Two topographic extrusions at around y = 7000 m are added in the floodplains. The extrusion is generated with the following equations:

$$T = 10 - [(x - 500)^2 + (y - 7000)^2] * 1.10^5$$

with x ∈ [0; 450] ∪ [550; 1000] and y ∈ [6000; 8000]

$$Z = T + \min(|T|)$$

The topography and bathymetry of the model are displayed in Figure 1 (a) and (b).

The upstream and downstream boundary condition, respectively represented in orange and green in Figure 1 (d) are prescribed with a constant inflow and a rating curve. The Strickler coefficient is prescribed to 15 $m^{1/3}/s$ in the left floodplain, to 60 $m^{1/3}/s$ in the right floodplain and to 35 $m^{1/3}/s$ in the channel (Figure 1 - (c)). The triangular mesh contains around 115,000 nodes, and the global mean mesh edge length is 12.5 m. Within the riverbed, the triangles are oriented with a mean edge horizontal length of 6.25 m and a mean edge vertical length of 12.5m. Floodplains are described on an unstructured mesh with a mean edge length of 12.5 m.

Preliminary analyses on the impact of the mesh density, the computational time steps and the friction coefficients were carried out. It was shown that an efficient computation with reasonable CPU resource cost (around 2 min per run on 6 nodes of Intel Xeon Gold 6140 core) was achieved with a time step of 10 s for this 115,000-node mesh. In addition, such analyses showed that the dynamics of the flow is mostly driven by the inflow, and that the impact of a variation in friction coefficients is of lesser importance. This motivates the choice of the DA control vector as a corrective term to the inflow in this preliminary work.

A spin-up simulation with a constant inflow of 5,000 m^3/s is integrated over 48 h to provide the initial condition for all ensemble members of the DA strategy. These members are characterized by an inflow value prescribed between 4000 m^3/s and 6000 m^3/s.

OSSE Framework

For the DA ensemble test case, two OSSE experiments are set. The OSSE framework is based on a deterministic reference simulation (denoted as *the truth*) with a chosen set of parameters, in our case, a prescribed corrective term to the inflow named β. β is supposed to be a random variable with a gaussian distribution characterised by a zero-mean and a standard deviation (std) here chosen to 700 m^3/s. This factor is added to a constant *a priori* water flow rate prescribed as the inflow boundary condition for each ensemble member. Two different ensembles of 60 members with *a priori* settings are generated for our following tests: (i) 4,500 and (ii) 5,500 m^3/s, respectively experiment A and B (noted exp.A and exp.B).

Synthetical observations are generated from the truth simulation water level (WL) maps, thresholded to express a water mask and to mimic SAR derived flood extent maps.

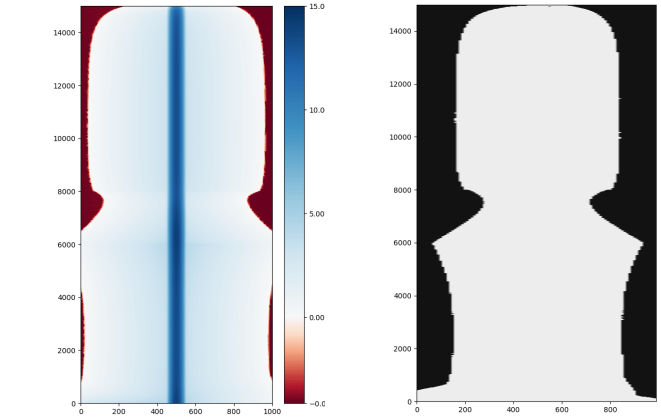
C. *Wet-dry pixels interface and level set variable*

Level-set (LS) map and LS functions are closely related to shape optimization and topology analysis. It is used in several research domains to compare and analyse the evolution of an object in images as it allows a straightforward treatment of topology changes (see [22]). In hydrodynamic study, it can be used to take into account the changes and deformations of the wet/dry interfaces of a flood map. The wet/dry interfaces are extracted from the WL map, thus defining the LS function noted Φ. It is here defined as follows:

$$\Phi(x, y) = WL(x, y) - WL_{fr}$$

where (x, y) denotes the (lon,lat) coordinates of a grid cell and WL_{fr} is the threshold between wet and dry areas (WL_{fr} = 5 cm in our case). The LS value is null at the interface between wet and dry areas. An example of a level-set map for exp.B is shown in Figure 2 (a).

In the framework of OSSE, this simulated interface position is compared to synthetical water mask derived from the truth simulation, also thresholded to WL_{fr} and represented as a water mask as shown in Figure 2 (b).



(a) (b)

Figure 2.
(a) Simulated level-set function computed from T2D water level (from exp.B)
(b) Observed binary flood extent map from a synthetical SAR image (from the truth WL map)

D. *The Chan-Vese contour fitting functional*

CV distances derived from the CV functional are based on two scalars, noted as C0 and C1, which rely on the formulation of the contingency map (Figure 3) between the observed and the simulated images. In the contingency map, each pixel is identified as one of the four following outcomes:

- False Negative (FN - in yellow in Figure 3) if the pixel is flooded in the observation but dry in the simulation,
- True Negative (TN - in blue in Figure 3) if the pixel is non flooded in the observation and dry in the simulation,
- False Positive (FP - in green in Figure 3) if the pixel is non flooded in the observation and wet in the simulation,
- True Positive (TP - in orange in Figure 3) if the pixel is flooded in the observation and wet in the simulation.

C0 quantifies the mismatch between the observed and the simulated flooded area. It is the ratio between the number of flooded pixels that the simulation misses (FN) and the total of the non-flooded pixels in the simulation (TN + FN).

C1 quantifies the match between the observed and the simulated flooded area. It is the ratio between the number of

flooded pixels that the simulation hits (TP) and the total of flooded pixels in the simulation (TP + FP).

C1 and C0 are then defined as follow:

$$C_0 \approx \frac{FN}{TN + FN} \; and \; C_1 \approx \frac{TP}{TP + FP}$$

In the following, a buffer around the fronts is determined to reduce the computational cost of the CV metric and the subsequent algebra in the DA algorithm. The contingency map (Figure 3) is only computed in this buffer zone near the fronts (*i.e.* white pixels outside the buffer are not taken into account). The buffer localisation is described in subsection II-F. The observed image is considered as the reference binary map and the LS simulation as the experiment to be assessed with respect to this reference.

Figure 3 displays the contingency map for two typical scenarios. In Figure 3 (a), the simulated flooded area is totally included within the observed flooded area, hence FP = 0 and TP+FP = TP, so that 0 < C0 < 1 (TN ≠ 0, FN ≠ 0) and C1=1. In Figure 3 (b) the simulated flooded area fully contains the observed flooded area, thus FN = 0, therefore C0 = 0 and 0 < C1 < 1 (TP ≠ 0, FP ≠ 0).

Two other scenarios are also possible in general: (i) the perfect agreement scenario, when the simulated flooded surface perfectly matches the observed flooded surface, *i.e.* FN = 0 and FP = 0, so that C0=0 and C1=1 (TN ≠ 0, TP ≠ 0); and (ii) the total disagreement scenario, where the simulated flooded surface and the observed flooded area do not overlap nor intersect, thus TP = 0 and FN ≠ TN+FN, so that 0 < C0 < 1 and C1 = 0 (TN ≠ 0, FP ≠ 0). This last scenario is not possible in our case since the riverbed is wet in both simulation and observation. An intermediate state, more common in practice, is the partial agreement, when the simulated flooded surface partially overlaps the observed flooded area, leading to TP ≠ 0, FN ≠ 0, so that 0 < C0 < 1 and 0 < C1 < 1.

The CV functional can be seen as a discrepancy functional between two LS. We decompose the CV functional in two functionals of the following form:

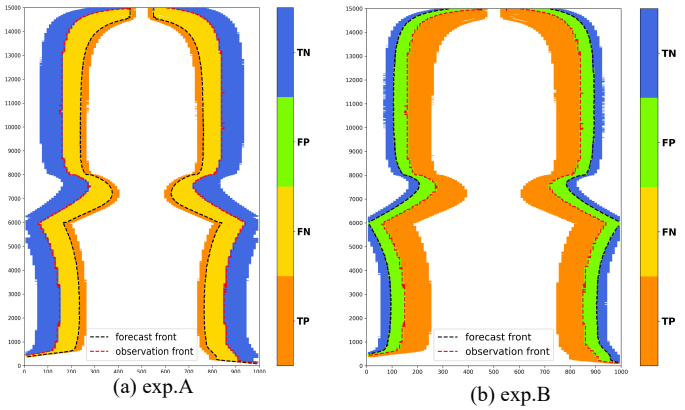
$$\mathcal{J} = \mathcal{J}^+ + \mathcal{J}^- \qquad (1)$$

where:

$$\mathcal{J}^+ = \int_\Omega Hv(\Phi)\left(\Phi^{obs} - Cmax\left(\Phi^{obs}, \Phi\right)\right)^2 dxdy$$

$$\mathcal{J}^- = \int_\Omega \left(1 - Hv(\Phi)\right)\left(\Phi^{obs} - Cmin\left(\Phi^{obs}, \Phi\right)\right)^2 dxdy$$

where Ω is the buffer area. ϕ^{obs} is the binary observed map and ϕ the simulated LS function. Hv the Heaviside function, it is equal to 1 if ϕ is non-negative and to 0 otherwise. Cmax and Cmin are respectively the max and min between C0 and C1. $\mathcal{J}$ will be used as an analytical metric for results assessment in Section III.



(a) exp.A (b) exp.B

Figure 3. Examples of contingency maps
blue = TN; green = FP; yellow = FN; orange = TP
dashed black line=forecast front; dashed red line=observed front

From $\mathcal{J}^+$ and $\mathcal{J}^-$, which are squared functionals, we define Chan-Vese distances that are compatible with the ETKF algorithm:

$$D\left(\Phi^{obs}, \Phi\right) = D^+\left(\Phi^{obs}, \Phi\right) + D^-\left(\Phi^{obs}, \Phi\right) \qquad (2)$$

D^+ and D^- are defined as follow:

$$D^+\left(\Phi^{obs}, \Phi\right) = Hv(\Phi)\left(\Phi^{obs} - Cmax\left(\Phi^{obs}, \Phi\right)\right) \; and$$

$$D^-\left(\Phi^{obs}, \Phi\right) = \left(1 - Hv(\Phi)\right)\left(\Phi^{obs} - Cmin\left(\Phi^{obs}, \Phi\right)\right).$$

E. Buffer localisation

The observation operator in the ETKF is formulated in the observation space, here described as the number of pixels. This dimension is reduced when a buffer zone is prescribed.

The buffer selects only the points that are the most informative near the ensemble of fronts (from observation image and the simulated runs). It consists in three steps (illustrated by Figure 4): (i) extract all the fronts in the ensemble LS maps and the binary flood extend; (ii) compute for each front line a polygon around it with a buffer chosen as twice the size of the pixel; (iii) merge all the buffer polygons in order to build a unique buffer polygon that is their union altogether.

F. Ensemble-based data assimilation algorithm (ETKF)

1) Description of the control vector

The control vector, denoted by **x** composed of a single parameter noted β, that is an additive factor to the water flow rate Qup at the upstream boundary condition. **x** is of size (1x Ne) where Ne is the ensemble size. In the following, **x** is noted $\mathbf{x}^{\mathbf{f}}$ or $\mathbf{x}^{\mathbf{a}}$ for forecast and analysis respectively. It is indexed with i that represents the i-th member between [1, Ne].

The implemented DA algorithm consists of a cycled deterministic ETKF where each cycle involves one or several binary flood extent observations. Each assimilation cycle c covers a time window $T = [t_{start}, t_{end}]$ of 3h-duration where $n_{obs,c}$ are assimilated. In our case, $n_{obs,c}$ is the number of pixels to be assimilated from all the observed images over the cycle c.

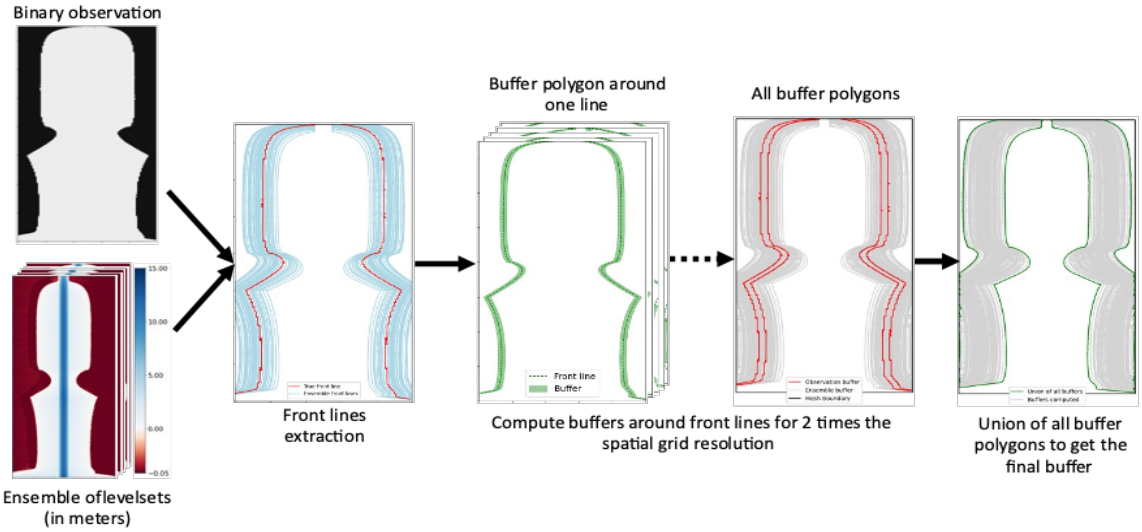


Figure 4. Framework for the buffer selection

Over a DA cycle and for each member of the ensemble, the *a priori* Qup is used to run the *a priori* simulation with the hydrodynamic deterministic model with Qup = 4,500 (exp.A) or 5,500 m3/s (exp. B). This is the forecast step; it defines the background or *a priori* hydraulic state. The simulated WL map is expressed as a water mask that is the model equivalent to the observed flood extent at the observation time. The CV discrepancy between observed and simulated flood extent for each member is used to determine the misfits and the covariance matrices, leading to the estimation of the analysis of β. This is the analysis step, it describes the a posteriori, or analysed hydraulic state. These steps are further detailed in the following. For each member, the updated control vector is used to perform the analysed trajectory with the hydrodynamic model. For each DA cycle, the analysed trajectory starts with a spin-up of 2h at t_{start} - 2h in order to reduce the impact of the initial condition on the analysis. There is 1h of overlap between the 3h-assimilation window and the 2h-spin-up window. It provides a final analysed state at t_{start} + 2h which correspond to the end of the assimilation window for cycle c.

2) ETKF forecast step

An ETKF is used in our DA experiment. The ETKF performs the linear algebra of the analysis mostly in the ensemble subspace, which is of a smaller dimension than that of the observation space. For cycle c, the forecast step consists of the propagation in time, over T, of the background control and model state vectors.

The control vector x^i is then of size one, the observation space is of space n_{obs} and the hydraulic state subspace of size m. We consider a time t included in the assimilation cycle c. $\mathcal{M}_c$: $R \rightarrow R^n$ denote the hydraulic model at the cycle c, of size n.

The background hydraulic state, $s_c^{f,i}$, associated with the ensemble i-th member of the control background vector is equal to:

$$s_c^{f,i} = \mathcal{M}_c\left(s_{c-1}^{a,i}, x_c^{f,i}\right) \quad (3)$$

where $x_c^{f,i}$ is the forecasted control vector resulting from the previous analysis cycle. $s_{c-1}^{a,i}$ is the analysed state from the previous cycle that is used as the initial condition for the current cycle. For the first cycle, it is provided by a restart file. Along DA cycles, x_c^f is generated using $x_{c-1}^{a,\,i}$ and a random perturbation in order to avoid the ensemble collapse.

The equivalent of control vector in the observation space for each member, denoted by $y_t^{f,i}$, stems from:

$$y_c^{f,i} = \mathcal{H}_c\left(s_c^{f,i}\right) \quad (4)$$

where $\mathcal{H}_c$: $R^m \rightarrow R^{nobs}$ is the observation operator in cycle c, from the model state space to the observation space, that selects, extracts and eventually interpolates model outputs at times and locations of the observation vector $\mathbf{y^o}$.

3) ETKF analysis step

Details on the ETKF can be found in [23], [24] and [25].

The EnKF methods, including the ETKF variant, are based on the Kalman filter equations:

$$\boldsymbol{x}^a = \boldsymbol{x}^f + \mathbf{K}\,\mathrm{D}(\boldsymbol{y}^o; \boldsymbol{y}^f) \quad (5)$$

$$\mathbf{P}^a = (\mathbf{I} - \mathbf{KH})\mathbf{P}^f \quad (6)$$

$$\mathbf{K} = \mathbf{P}^f\mathbf{H}^{\mathrm{T}}(\mathbf{H}\mathbf{P}^f\mathbf{H}^{\mathrm{T}} + \mathbf{R})^{-1} \quad (7)$$

with $\mathbf{x^a}$ the analysed control vector of the ensemble (of size (1 x Ne)), $\mathbf{K}$ the Kalman gain (of size (Ne x n_{obs})), $\mathbf{P^a}$ the analysis error covariance matrix (of size (n x n)), $\mathbf{P^f}$ the forecast error covariance matrix (of size (n x n)) and $\mathbf{R}$ the observation error covariance matrix (of size (n_{obs} x n_{obs})) which is taken diagonal in this study. D($\mathbf{y^o}$; $\mathbf{y^f}$) is the innovation vector between the observation $\mathbf{y^o}$ and the model equivalent related to the forecast ensemble states. In this work, D is the CV distance defined in subsection II-E.

The principle of the EnKF methods is to stochastically the covariances matrices $\mathbf{P}$ within the ensemble.

$$\bar{\mathrm{x}} = \frac{1}{\mathrm{Ne}} \sum_{\mathrm{i}=0}^{\mathrm{Ne}} \mathrm{x_i} \tag{8}$$

is the control ensemble mean and **P** reads:

$$\mathbf{P} = \frac{1}{\mathrm{Ne}-1} \sum_{i=0}^{Ne} (x_i - \bar{x})(x_i - \bar{x})^T = \frac{1}{\mathrm{Ne}-1} \mathbf{X}\mathbf{X}^{\mathrm{T}} \tag{9}$$

with $[X]_i = (x_i - \bar{x})$ the matrix of the ensemble anomalies. We note $\mathbf{X_a}$ the analysed anomaly vector and $\mathbf{X_f}$ the forecasted anomaly vector.

For each member of the ensemble, the *a priori* is updated during the analysis step:

$$x^{a,i} = x^{f,i} + \mathbf{K}\, \mathrm{D}(y^o; y^{f,i}). \tag{10}$$

In the EnKF, in order to avoid ensemble collapse, an ensemble of perturbed observations is used in place of the deterministic observation vector. This way, (6) is validated in a statistical way.

In contrast to the EnKF, the ETKF updates in an explicit way the ensemble means and anomalies inside the ensemble subspace instead of working the algebra in the observation space (as in (3)). The ensemble mean is updated using the analysis (1) and the anomalies are updated with an explicit transformation represented by the ensemble transform matrix **T**:

$$\mathbf{X}_a = \mathbf{X}_f \mathbf{T} \tag{11}$$

The analysed anomalies are defined such that the analysed covariance matrix respect (6), given (9) and (11):

$$\mathbf{P}^a = (\mathrm{I} - \mathbf{K}\mathbf{H})\mathbf{P}^f = (\mathrm{I} - \mathbf{K}\mathbf{H})\mathbf{X}_f\mathbf{X}_f^{\mathrm{T}} \tag{12}$$

$$\mathbf{P}^a = \mathbf{X}_a\mathbf{X}_a^{\mathrm{T}} = \mathbf{X}_f\mathbf{T}\mathbf{T}^{\mathrm{T}}\mathbf{X}_f^{\mathrm{T}} \tag{13}$$

If we develop (12) according to (7) and (9) we can choose $\mathbf{T}\mathbf{T}^{\mathrm{T}}$ as:

$$\mathbf{T}\mathbf{T}^{\mathrm{T}} = \left(\mathrm{I} - \frac{1}{Ne-1}\mathbf{Y}_f^T \left(\frac{1}{Ne-1}\mathbf{Y}_f\mathbf{Y}_f^T + \mathbf{R}\right)^{-1}\right) \tag{14}$$

$$and\ \mathbf{Y}_f = \mathbf{H}\mathbf{X}_f \tag{15}$$

(14) is equivalent, according to the Sherman-Morrison-Woodbury formula [23], to:

$$\mathbf{T}\mathbf{T}^{\mathrm{T}} = \left(I * (Ne-1) + \mathbf{Y}_f^T\mathbf{R}^{-1}\mathbf{Y}_f\right)^{-1}. \tag{16}$$

In the present work, $\mathbf{Y}_{\mathrm{f,i}}$ is computed as the CV distance between the mean ensemble background and each member of the ensemble:

$$\left[\mathbf{Y}_f\right]_i = \mathbf{H}\left[\mathbf{X}_f\right]_i = \frac{D(Hv(\bar{y}^f); y^{f,i})}{\sqrt{Ne-1}} \text{ and } \bar{y} = \frac{1}{Ne}\sum_{i=1}^{Ne} y^{f,i}$$

[26] propose to decompose **T** in eigenvectors such that the ensemble mean is preserved. **T** is a positive symmetric matrix. The orthogonal matrix **U** (generally chosen as the identity matrix) is introduced so that **T** decomposes as:

$$\mathbf{T} = \sqrt{Ne-1}C\Gamma^{-\frac{1}{2}}C^T\mathbf{U} \tag{17}$$

where *C* are the eigenvectors (because **T** is symmetric) and **Γ** the eigenvalues matrix of **T**. The analysed thus reads:

$$\mathbf{X}_a = \mathbf{X}_f\mathbf{T} = \sqrt{Ne-1}\mathbf{X}_f C\Gamma^{-\frac{1}{2}}C^T\mathbf{U} \tag{18}$$

and the ensemble mean can is updated:

$$\bar{x}^a = \bar{x}^f + \mathbf{X}_f\mathbf{w_a} \tag{19}$$

$$with\ \mathbf{w}_a = (C\Gamma^{-1}C^T)\mathbf{Y}_f^T\mathbf{R}^{-1}\mathrm{D}(y^o; \bar{y}^f) \tag{20}$$

with $\mathrm{D}(y^o; \bar{y}^f)$ the CV distance between the observation and the ensemble mean.

For each member, the analysed control vector is updated as:

$$x^{a,i} = \bar{x}^a + \mathbf{X}_a^i \tag{21}$$

where $\mathbf{X}_a^{\mathrm{i}}$ is the i-th column of the matrix $\mathbf{X_a}$.

We can now compute the new analysed state for the next cycle with the hydrodynamic model and the analysed control vector with:

$$s_c^{a,i} = \mathcal{M}_c\left(s_{c-1}^{a,i}, x_c^{a,i}\right) \tag{22}$$

III. EXPERIMENTAL SETTING AND RESULTS

The result of the DA for the corrective term on the upstream forcing using the CV metric to compute the discrepancy between observed and simulated front is assessed in this section. In the following, these results are displayed with two types of figures with similar layout for exp.A ($Q_{up,\ a\ priori}$ = 4,500 m3/s) and exp.B ($Q_{up,\ a\ priori}$ = 5,500 m3/s). Four DA cycles are carried out.

Figure 5 and Figure 7, represent the ensemble fronts for the forecast (subplot a) and the analysis (subplot b) for the last time of the last DA cycle, at physical time t = 124,000 s with a start at t0 = 86,400 s. Members of the ensemble (forecast or analysed) are plotted in thin grey lines, the ensemble mean (forecast or analysis) is plotted in green, and the observed front is plotted in red.

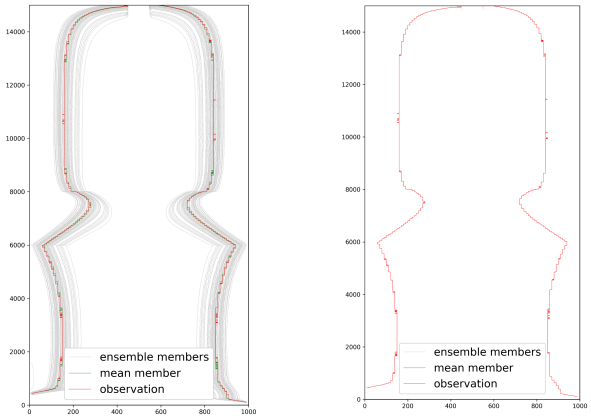
In Figure 6 and Figure 8, subplots (a) represents the evolution of the ensemble means (upper panels) and the standard deviations (bottom panels) over the DA cycles for the forecast control vector (in blue) and the analysis control vector (in red). Subplots (b) show the evolution of the mean CV functional (eq. (1)) computed between the ensemble members' levelset and the observed binary images used for results analysis. The smaller the CV functional the better; it reaches zero if the two images used for the comparison are perfectly equal.

A. Results for experiment A

In exp.A, the *a priori* is lower than the truth. Since the true value for Qup is 5000m^3/s, β is expected to be positive and to be close to 500 m3/s (because of the observation error) allowing for an inflow forecasted and analysed close to the truth.

For a given DA cycle, the buffer limits the observation space so that the number of pixels is about 50,000 for one image instead of 150,000 for the whole domain. In the present case where three images are assimilated over T for cycle c, n_{obs} = 150,000.

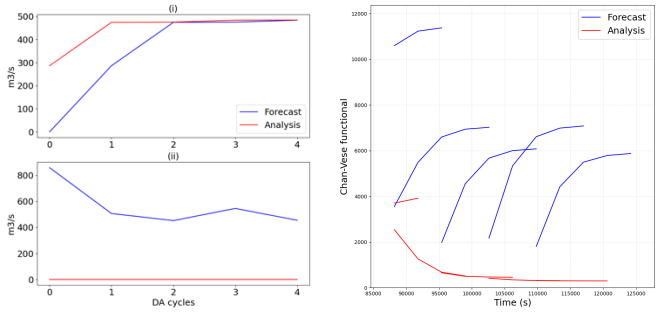
The forecasted and analysed front lines for the last DA cycle are shown in Figure 5 (a) and (b) respectively. The analysis is brought closer to the truth so that the members, mean and observations seem to overlap.



(a) forecast front lines (b) analysis front lines

Figure 5. Exp.A - Front lines of the ensemble members (in light grey), the observation (in red) and the ensemble mean (in green) for the last time of the last cycle of DA.

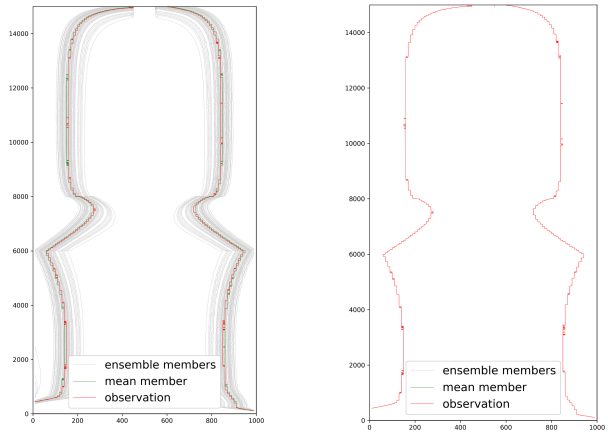
This illustrates how the DA algorithm is capable of correctly increasing the inflow to almost reach the truth inflow value.



(a) Summary of the analysis phases. (i) ensemble mean; (ii) ensemble std (b) Chan-Vese functional evolution in time.

Figure 6. Analysis metric for Exp.A.
In blue the forecast, in red the analysis

The upper panel in subplot a in Figure 6 shows the evolution of the ensemble mean for β along the DA cycles. It should be noted that, as expected, the analysis as well as the forecast is reaches almost 500 m^3/s which give the truth value when added to the *a priori* value of 4,500 m^3/s. Since the truth is prescribed as a constant value, the analysis converges to this value along the DA cycles and the std of the ensemble decreases to zero. This shows that the ETKF DA efficiently brings the *a priori* close to the truth, with a high certainty in spite of the limited number of members in the ensemble. Indeed, before the analysis, the forecast std values are found between 400 and 800 m^3/s. The values are greatly reduced with the DA algorithm. Figure 6 (b) displays the Chan-Vese functional over time from eq. (1). This clearly shows the improvement brought by the DA analysis with the cost function metrics. As expected, the forecast and analysis cost functions decrease over the DA cycles, with a smaller $\mathcal{J}$ value for the analysis. The CV functional for the analysis runs decreases almost exponentially. Even though it does not reach zero (perfect state) for the ensemble, it converges quite well towards it. Indeed, we put a strong trust on observations so that we expect that the analysis phase brings the control vector near the truth.

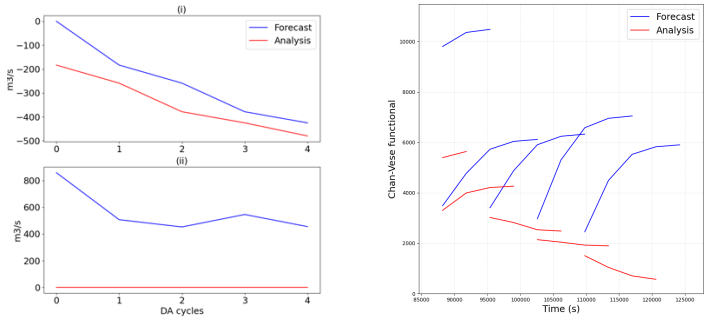


(a) forecast front lines (b) analysis front lines

Figure 7. Exp.B - Front lines of the ensemble members (in light grey), the observation (in red) and the ensemble mean (in green) for the last last time of the last cycle of DA.

B. *Results for experiment B*

In exp.B, a second *a priori* setting was tested when the inflow ensemble mean is greater than the truth, *i.e.* 5,500 m^3/s compared to the true value 5,000 m^3/s. As opposed to exp.A, here the beta factor is expected to be negative and approach - 500 m^3/s, which allows for decreased inflow closer to the truth.



(a) Summary of the analysis phases. (i) ensemble mean; (ii) ensemble std (b) Chan-Vese functional evolution in time.

Figure 8. Analysis metric for Exp.B.
In blue the forecast, in red the analysis

Similarly to Section III A, Figure 7 (a) and (b) show the forecasted and analysed front lines for the last DA cycle. For the analysis all members, mean and observations seem to overlap.

Figure 8 (a) (upper panel) shows the evolution of the mean ensemble beta factor evolution. The analysis as well as the forecast reaches - 500 m^3/s which give the truth value when added to the *a priori* value of 5,500 m^3/s.

It seems that the convergence over the DA cycles is slower than in expA. This suggests that the non-linearity between the inflow and the flood extent is stronger when removing water from the system, which may relate to the incapacity of the model to wash out water from the floodplain as evapotranspiration processes are not accounted for. This question is currently being investigated. The ensemble inflow std value of the analysis is almost zero (Figure 8 (a) bottom panel) meaning that the ETKF DA is most certain. Similarly, to expA, the CV cost function significantly decreases over the DA cycles with reduced values for the analysis with respect to the forecast as shown in Figure 8 (b), with a lesser efficiency than for exp.A.

IV. Discussion and Conclusion

In conclusion, the implementation of the ETKF with the CV metric for front distance estimation shows promising results. The OSSE framework allows for a sanity validation of the algorithm here applied to a simplified test case. The construction of a buffer zone to reduce the size of the observations space allows to reduce the computational cost of the ETKF DA algorithms. For now, the buffer selection phase is done for each cycle in the algorithm. This phase could be adapted to be more efficient.

Further work will focus on more advanced test cases where the simulated and observed fronts intersect. This could result from uncertainties in topography and Strickler coefficient.

The perspective for this work is to apply the CV-ETKF strategy to real test cases, in unsteady conditions, for instance for the Garonne Marmandaise model, and extend the observations to other satellite-derived flood extents, for instance from SWOT.

Acknowledgement

The authors acknowledge the primary works of E. Combes during his internship on the Chan-Vese functional for hydrodynamics T2D models.

References

[1] United Nations Office for Disaster Risk Reduction, 2018, The human Cost of Weather-related Disasters, Available from: https://www.unisdr.org/files/46796_cop21weatherdisastersreport2015.pdf

[2] F. Dottori, L. Mentaschi, A. Bianchi, L. Alfieri, and L. Feyen. Adapting to rising river flood risk in the EU under climate change. Joint Research Centre (JRC), 2020.

[3] R. Hostache, X. Lai, J. Monnier, and C. Puech. Assimilation of spatially distributed water levels into a shallow-water flood model. Part II: Use of a remote sensing image of Mosel River. J Hydrol, 390(3-4):257–268, 2010.

[4] H. Oubanas, I. Gejadze, P.-O. Malaterre, et al. Discharge estimation in ungauged basins through variational data assimilation: The potential of the swot mission. Water Resour. Res., 54(3):2405–2423, 2018.

[5] F. Cian, M. Marconcini, P. Ceccato, and C. Giupponi. Flood depth estimation by means of high-resolution SAR images and lidar data. Nat. Hazards Earth Syst. Sci., 18(11):3063–3084, 2018.

[6] M. Chini, R. Hostache, L. Giustarini, and P. Matgen. A hierarchical split-based approach for parametric thresholding of SAR images: Flood inundation as a test case. IEEE Trans. Geosci. Remote Sens., 55(12):6975–6988, 2017.

[7] Sandro Martinis, Jens Kersten, and André Twele. A fully automated TerraSAR-X based flood service. ISPRS J. Photogramm. Remote Sens., 104:203–212, 2015.

[8] B. Revilla-Romero, N. Wanders, P. Burek, P. Salamon, and A. de Roo. Integrating remotely sensed surface water extent into continental scale hydrology. J. Hydrol., 543:659–670, 2016.

[9] A. Dasgupta, R. Hostache, R. Ramsankaran, et al. A mutual information-based likelihood function for particle filter flood extent assimilation. Water Resour. Res., 57(2):e2020WR027859, 2021.

[10] C. Di Mauro, R. Hostache, P. Matgen, et al. Assimilation of probabilistic flood maps from SAR data into a coupled hydrologic/hydraulic forecasting model: a proof of concept. Hydrol. Earth Syst. Sci., 25:4081–4097, 2021.

[11] E.S. Cooper, S.L. Dance, J. García-Pintado, N.K. Nichols, and P.J. Smith. Observation operators for assimilation of satellite observations in fluvial inundation forecasting. Hydrol. Earth Syst. Sci., 23(6):2541–2559, 2019.

[12] S. Peña-Luque, S. Ferrant, M.C.R. Cordeiro, et al. Sentinel-1 & 2 multitemporal water surface detection accuracies, evaluated at regional and reservoirs level. Remote Sens., 13(16), 2021.

[13] Nguyen, T. H., Ricci, S., Piacentini, A., Fatras, C., Kettig, P., Blanchet, G., et al. (2022). Dual state-parameter assimilation of SAR-derived wet surface ratio for improving fluvial flood reanalysis. Water Resources Research, 58, e2022WR033155. https://doi.org/10.1029/2022WR033155

[14] T. F. Chan and L. A. Vese, "Active contours without edges," in IEEE Transactions on Image Processing, vol. 10, no. 2, pp. 266-277, Feb. 2001, doi: 10.1109/83.902291.

[15] Rochoux, M. C., Collin, A., Zhang, C., Trouvé, A., Lucor, D., and Moireau, P. (2018). Front shape similarity measure for shape-oriented sensitivity analysis and data assimilation for eikonal equation. ESAIM: Proceedings and Surveys, 63:258-279

[16] Rochoux, M. C., Emery, C., Ricci, S., Cuenot, B., Trouvé, A. (2014). Towards predictive data-driven simulations of wildfire spread – Part II: Ensemble Kalman Filter for the state estimation of a front-tracking simulator of wildfire spread, Natural Hazards and Earth System Sciences, doi:0.5194/nhess-15-1721-2015

[17] Zhang, C., Collin, A., Moireau, P., Trouvé, A., Rochoux, M. C. (2018). Front shape similarity measure for data-driven simulations of wildland fire spread based on state estimation: Application to the RxCADRE field-scale experiment. Proceedings of the Combustion Institute

[18] Martins, M., Ferreira Jr, S., and Vilela, M. (2007). Multiscale models for the growth of avascular tumors. Physics of Life Reviews, 4(2):128-156.

[19] Collin, A., Chapelle, D., and Moireau, P. (2015). A Luenberger observer for reaction-diffusion models with front position data. J Comput Phys, 300(C):288-307.

[20] Li, L., Le Dimet, F.-X., Ma, J., Vidard, A. (2017). A level-set based image assimilation method: Potential applications for predicting the movement of oil spills. IEEE Transactions on Geoscience and Remote Sensing, Institute of Electrical and Electronics Engineers. doi:10.1109/TGRS.2017.2726013

[21] M. Hervouet, Hydrodynamics of free surface flows: modelling with the finite element method. Wiley Online Library, 2007, vol. 360.

[22] van Dijk, N.P., Maute, K., Langelaar, M. et al. Level-set methods for structural topology optimization: a review. Struct Multidisc Optim 48, 437–472 (2013). https://doi.org/10.1007/s00158-013-0912-y

[23] M. Asch, M. Bocquet, and M. Nodet, Data assimilation: methods, algorithms, and applications. SIAM, 2016.

[24] P. Sakov et P. R. Oke (2008a). A deterministic formulation of the Ensemble Kalman Filter : an alternative to ensemble square root filters. Tellus A, 60(2):361–371.

[25] B. R. Hunt, E. J. Kostelich et I. Szunyogh (2007). Efficient data assimilation for spatiotemporal chaos : A local ensemble transform Kalman filter. Physica D, 230:112 – 126.

[26] E. Ott, B. R. Hunt, I. Szunyogh, A. V. Zimin, E. J. Kostelich, M. Corazza, E. Kalnay, D. Patil et J. A. Yorke (2004). A local Ensemble Kalman filter for atmospheric data assimilation. Tellus A, 56:415–428.