Data Assimilation on a Flood Wave Propagation Model : Emulation of an Ensemble Kalman Filter Algorithm

Sébastien Barthélémy[1] - Sophie Ricci[1] - Olivier Pannekoucke[2]
Olivier Thual[1,3] - Pierre-Olivier Malaterre[4]

[1]CERFACS/URA1875 - [2]CNRM/GAME - [3]INPT, CNRS, IMFT - [4]IRSTEA

November 13th 2012

Objective and outline

Objective

Emulation of an Ensemble Kalman Filter (EnKF) algorithm to allow for the use of data assimilation in the context of operationnal flood forecasting with a low computationnal cost.

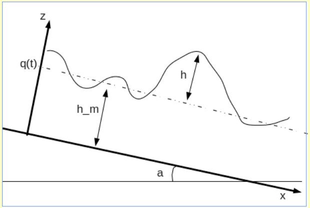
1. Advection-diffusion model and random forcing
2. Water level covariance functions without observations : theory versus ensemble computation
3. Impact of the assimilation on the water level covariance functions : towards the emulation of the EnKF

Outline

1. Advection-diffusion model and random forcing
2. Water level covariance functions without observations : theory versus ensemble computation
3. Impact of the assimilation on the water level covariance functions : towards the emulation of the EnKF
   - Data assimilation chain and algorithms
   - Data assimilation results
   - Emulation of the EnKF algorithm

Model equations and boundary conditions

The 1D advection-diffusion equation



Inclined channel with free surface

$$\begin{cases} \frac{\partial h}{\partial t} + c\frac{\partial h}{\partial x} = \kappa \frac{\partial^2 h}{\partial x^2} \ , \\ \qquad \forall (x,t) \in [0,L] \times \mathbb{R}^+ \\ h(x,0) = h_0(x) \ , \ x > 0 \\ h(0,t) = h_{up}(t) \\ \frac{\partial h}{\partial t}(L,t) + c\frac{\partial h}{\partial x}(L,t) = 0 \end{cases}$$

- h denotes the Water Level Anomaly (WLA) : perturbation to the equilibrium state h_m ;
- h_{up} is a random water level ;
- Models the shallow-water equations when the slope is important ;

The upstream forcing h_{up}

Construction of the upstream forcing

$h_{up}(t)$ is a random variable with gaussian statistics and gaussian temporal covariance function $\rho(\delta t) = q_m^2 e^{-\frac{\delta t^2}{2\tau^2}}$ where :

- τ is the correlation time scale ;
- q_m^2 is the variance ;

Under mathematical considerations one can write :

$$h_{up}(t) = \int_{\mathbb{R}} \zeta_\omega \sqrt{\rho_\omega} e^{-i\omega t} d\omega \tag{1}$$

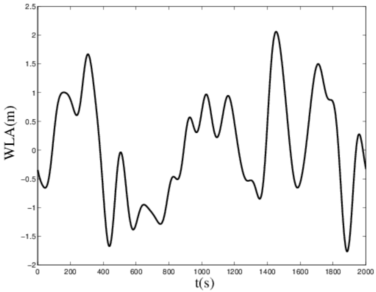
where $\rho_\omega = \frac{\tau}{\sqrt{2\pi}} q_m^2 e^{-\frac{\omega^2 \tau^2}{2}}$ and the ζ_ω follow the normal distribution $\mathcal{N}(0,1)$, are uncorrelated and $\zeta_\omega = \zeta_{-\omega}^*$;

Illustration of the upstream forcing h_{up}

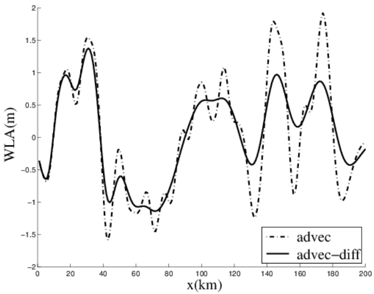
The model equation and the upstream boundary condition

$$\begin{cases} \frac{\partial h}{\partial t} + c\frac{\partial h}{\partial x} = \kappa \frac{\partial^2 h}{\partial x^2} \ , & \forall (x,t) \in [0,L] \times \mathbb{R}^+ \\ h(0,t) = h_{up}(t) & \forall t > 0 \end{cases}$$

An upstream forcing with its corresponding propagated WLA



Random upstream forcing



Corresponding WLAs

Outline

1. Advection-diffusion model and random forcing
2. Water level covariance functions without observations : theory versus ensemble computation
3. Impact of the assimilation on the water level covariance functions : towards the emulation of the EnKF
   - Data assimilation chain and algorithms
   - Data assimilation results
   - Emulation of the EnKF algorithm

Study of the covariance function of h without assimilation

Covariance function of h

$$\Sigma(x, x+\delta x) = \langle h(x) h^*(x+\delta x) \rangle$$

The case of advection only ($\kappa = 0$)

The covariance function reads : $\Sigma(x, x+\delta x) = \Sigma_m e^{-\frac{\delta x^2}{2L_p^2(0)}}$

- Constant correlation length scale : $L_p(0)$
- Constant variance : $\Sigma_m = q_m^2$

The case of advection and small diffusion ($\kappa \ll cx$)

An asymptotic expansion leads to : $\Sigma(x, x+\delta x) \approx \Sigma_m(x) e^{-\frac{\delta x^2}{2L_p^2(x)}}$

- Correlation length scale : $L_p(x) = \sqrt{L_p^2(0) + 4\kappa \frac{x}{c}}$
- Variance : $\Sigma_m(x) = q_m^2 \frac{L_p(0)}{L_p(x)}$

Covariance functions for the upstream forcing and the WLA

Approximation of the covariance matrix with an ensemble method

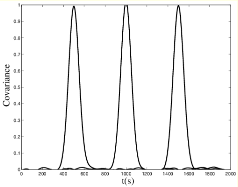
- Discretisation of the WLA in $n = 200$ grid points,

$$\mathbf{X}_k = (h_1^k, \dots, h_n^k)$$

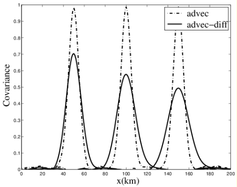
- The covariance matrix of the WLA is approximated by :

$$\mathbf{B}_e = \frac{1}{N}\sum_{k=1}^{N}(\mathbf{X}_k - \overline{\mathbf{X}})(\mathbf{X}_k - \overline{\mathbf{X}})^T \qquad N = 10000$$

Covariance functions for the upstream forcing and the WLA



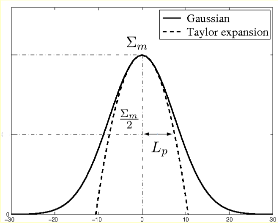
Temporal covariance functions for the upstream forcing for $t = 500s, 1000s, 1500s$



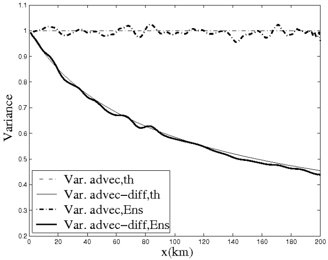
Spatial covariance functions for the WLA for $x = 50km$, $x = 100km, 150km$

Validation of the analytical results with an ensemble method

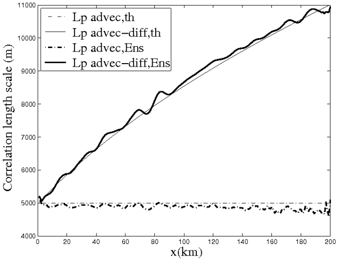
Diagnosis of the variance and the correlation length



- The variance Σ_m is the maximum of the gaussian ;
- The correlation length scale is the distance for which the second order Taylor expansion of the correlation function is equal to 1/2 ;



Variance



Correlation length scale

Outline

Outline

1. Advection-diffusion model and random forcing
2. Water level covariance functions without observations : theory versus ensemble computation
3. Impact of the assimilation on the water level covariance functions : towards the emulation of the EnKF
   - Data assimilation chain and algorithms
   - Data assimilation results
   - Emulation of the EnKF algorithm

Data assimilation chain for Kalman Filter algorithm (KF)

Discretisation of the WLA in $n = 200$ grid points,

$$\mathbf{X}_i = (h_1^i, \ldots, h_n^i)$$

Analysis step at assimilation cycle i

Update of the state vector : $\mathbf{X}_i^a = \mathbf{X}_i^b + \mathbf{K}_i(\mathbf{Y}_i^o - \mathbf{H}\mathbf{X}_i^b)$ where :

- $\mathbf{X}_i^a$ and $\mathbf{X}_i^b$ are the analysis and the background ;
- $\mathbf{K}_i = \mathbf{B}_i\mathbf{H}^T(\mathbf{H}\mathbf{B}_i\mathbf{H}^T + \mathbf{R})^{-1}$ is the gain matrix ;
- $\mathbf{B}_i$ is the background error covariance matrix ;
- $\mathbf{H}$ is the observation operator ;
- $\mathbf{R}$ is the observation error covariance matrix ;
- $\mathbf{Y}_i^o$ is a synthetic observation generated from the true state by adding an error following $\mathcal{N}(0, \sigma_o^2)$

Update of the analysis error covariance matrix : $\mathbf{A}_i = (\mathbf{I} - \mathbf{K}_i\mathbf{H})\mathbf{B}_i$ where $\mathbf{I}$ is the identity matrix

Data assimilation chain for Kalman Filter algorithm

Propagation step from cycle i to cycle $i+1$

- Propagation of the state vector : $\mathbf{X}^b_{i+1} = \mathcal{M}_{i,i+1}\mathbf{X}^a_i$
- Propagation of the background error covariance matrix :

$$\mathbf{B}_{i+1} = \mathbf{M}_{i,i+1}\mathbf{A}_i\mathbf{M}^T_{i,i+1}$$

  where $\mathbf{M}_{i,i+1}$ is the tangent linear of the model.
- When $\mathbf{B}_i$ is not propagated we talk of a sequence of Best Linear Unbiased Estimator (BLUE)

From the Kalman Filter to the Ensemble Kalman Filter

- Propagation of $\mathbf{B}_i$ with the KF algorithm is costly and requires the computation of the tangent linear of the model.
- Problem to define a relevant error on the upstream forcing with the KF.
- An ensemble approach is then prefered to compute $\mathbf{B}_i$ and define easily an error on the upstream forcing.

Derived algorithms from the assimilation chain

Derived algorithms

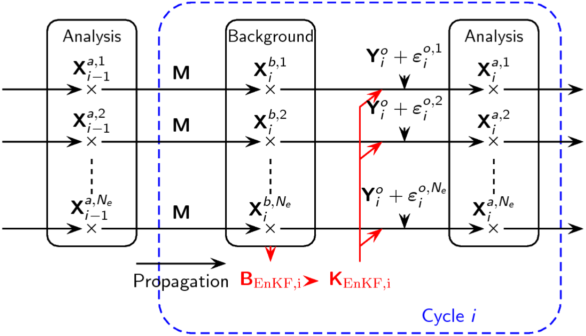
Algorithms derived from the data assimilation chain :

- EnKF where $\mathbf{B}_i = \mathbf{B}_{EnKF,i}$ is updated at each observation ;
- EnBLUE where $\mathbf{B}_i = \mathbf{B}_e$ remains constant and is computed without assimilation ;
- EEnKF where $\mathbf{B}_i = \mathbf{B}_{EnKF}$ remains constant and is the converged matrix from the EnKF ;

The question is ...

Can we use an algorithm where $\mathbf{B}_i$ is constant (EnBLUE or EEnKF) instead of an algorithm that requires the propagation of all the members to update $\mathbf{B}_i$ at each assimilation cycle (EnKF) ?

Assimilation scheme for the EnKF

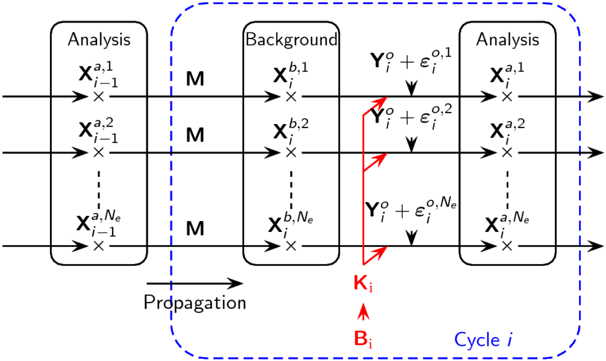


$\mathbf{B}_{EnKF,i}$ is updated at each observation with an ensemble method :

$$\mathbf{B}_{EnKF,i} = \frac{1}{N}\sum_{k=1}^{N}(\mathbf{X}_i^{b,k} - \overline{\mathbf{X}_i})(\mathbf{X}_i^{b,k} - \overline{\mathbf{X}_i})^T \qquad N = 10000$$

The analysis error covariance matrix computed after the last assimilation cycle is called $\mathbf{B}_{EnKF}$.

Assimilation scheme for the EnBLUE and the EEnKF

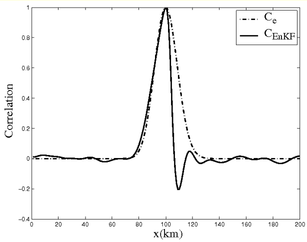


$\mathbf{B}_i$ is constant, it is either $\mathbf{B}_e$ (EnBLUE) or $\mathbf{B}_{EnKF}$ (EEnKF). The analysis error covariance matrix computed after the last assimilation cycle is called $\mathbf{B}_{EnBLUE}$ (EnBLUE) or $\mathbf{B}_{EEnKF}$ (EEnKF)

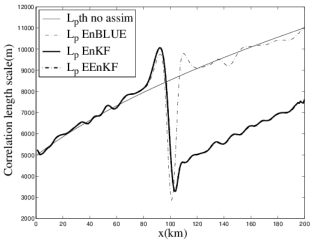
Outline

Data Assimilation Results - Part I

Impact of the assimilation on the correlation length



Correlation functions
at the observation point

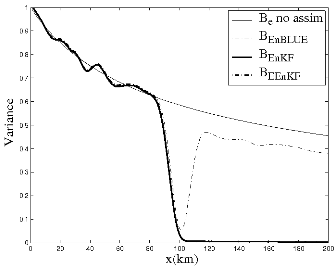


Correlation length
over the domain

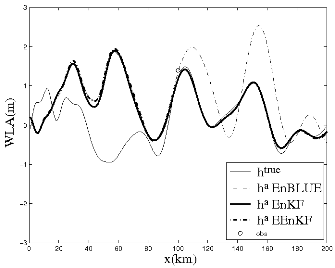
- The correlation function at the observation point turns into an anisotropic function with a shorten correlation length downstream ;
- In the case of the EnKF and EEnKF the reduction of correlation length is propagated downstream by the dynamics of the model whereas it is just local with the EnBLUE ;

Data Assimilation Results - Part II

Impact of the assimilation on the variance and the WLA



Variance

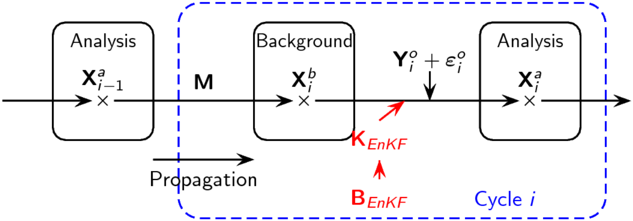


WLA over the domain

- $\mathbf{B}_{EnKF}$ and $\mathbf{B}_{EEnKF}$ provide the same results in terms of variance and WLA with the same upstream forcing whereas in the case of $\mathbf{B}_{EnBLUE}$ the reduction of variance and the correction of the WLA are just local round the observation point ;

Outline

Emulation of the EnKF algorithm



- EnKF and EEnKF provide the same result for the correlation length scale, the variance and the analysed WLA ;
- Design of an algorithm derived from the EEnKF : sequence of BLUE using $\mathbf{B}_{EnKF}$ with only one member ;
- Reduction of the computationnal cost allowing for the use of data assimilation in the context of real-time flood forecasting ;

Conclusions and further work

Conclusions

- Theoretical and numerical study of the covariance of the propagated water level signal without assimilation ;
- The assimilation of water level observations has an impact on the correlation length scale and improves the results in terms of variance and WLA in the case of the EnKF and the EEnKF algorithms ;
- The use of a matrix computed with assimilation in a sequence of BLUE algorithm allows for the emulation of the EnKF algorithm and the use of data assimilation in the context of flood forecasting ;

Further work

Modelling of the matrix $\mathbf{B}_{EnKF}$ without computing the EnKF algorithm using a diffusion operator with a different observation network.