Manuscript prepared for Hydrol. Earth Syst. Sci. Discuss.
with version 3.5 of the LaTeX class copernicus_discussions.cls.
Date: 4 November 2013

Emulation of an Ensemble Kalman filter algorithm on a flood wave propagation model

S. Barthélémy[1], **S. Ricci**[1], **O. Pannekoucke**[1,2], **O. Thual**[1,3], **and P.O. Malaterre**[4]

[1]URA 1875/CERFACS, Toulouse, France
[2]CNRM-GAME, UMR 3589, Toulouse, France
[3]INPT, CNRS, IMFT, Toulouse, France
[4]UMR G-EAU, Irstea, Montpellier, France

Correspondence to: S. Barthélémy (barthelemy@cerfacs.fr)

Abstract

This study describes the emulation of an Ensemble Kalman Filter (EnKF) algorithm on a 1D flood wave propagation model. This advection-diffusion model is forced at the upstream boundary through a random variable with stationnary gaussian statistics and a correlation function in time with gaussian shape. When the advection time is large compared to the diffusion time, an analytic study, confirmed by numerical experiments, shows that the covariance functions of the propagated signal anomaly keeps its gaussian shape in the absence of observation. In the case of a one observation point assimilation where synthetical observations are generated by adding an error to a true state, the dynamic of the background error covariance functions is no longer simple and a numerical approach using an EnKF algorithm is prefered. It is shown that the application of a Best Linear Unbiased Estimator (BLUE) algorithm using the background error covariance matrix converged from an EnKF algorithm, provides the same results as the EnKF but with a cheaper computational cost, thus allowing for the use of data assimilation in the context of real time flood forecasting. Numerical experiments lead to the construction of a parametrized matrix that emulates the background error covariance matrix converged from the EnKF. This method is then applied to a case with two different observation point with different error statistics providing results close to those of the EnKF but with a significantly reduced computational cost.

1 Introduction

With flood frequency likely to increase as a result of altered precipitation patterns triggered by climate change (Drogue et al., 2004) there is a growing need for improved flood modeling. While significant advances have been made in recent years in hydraulic data assimilation (DA) for water level and discharge prediction (Schumann et al., 2009; Biancamaria et al., 2011), as well as for parameters correction (Pappenberger et al., 2005; Durand et al., 2010), using insitu as well as remote sensing data (Andreadis et al., 2007; Neal et al., 2009) they are yet to be fully taken advantage of in the operational forecast of flood and inondation areas. Recent studies

have shown the benefit hydrology and hydraulics can draw from the progress of DA approaches using either variational inverse problem (Valstar et al., 2004), particle filtering (Matgen et .al, 2010; Giustarini et al., 2011), Extended Kalman Filter (Thirel et al., 2010), Ensemble Kalman Filter (EnKF) for state updating (Moradkhani et al., 2005b; Weerts et al., 2006), or for dual state-parameter estimation (Moradkhani et al., 2005a; Hendricks and Kinzelbach, 2008). In the field of hydraulic, amongst the numerous research studies in DA that aims at overcoming the limitations of the hydraulics models, only a few are formulated in an operational setting and demonstrate the peformance gained from DA (Madsen and Skotner, 2005; Malaterre et al., 2010; Weerts et al., 2010; Jean-Baptiste et al., 2011; Ricci et al., 2011).

For most DA algorithms, the description of the background error covariance matrix is essential but fastidious. Indeed the background error covariance matrix plays a keyrole in DA as this matrix spreads the information brought by the observations over the domain and between the state variables. These covariances must be modeled as approximations of the true covariances of backgound errors. In sequential DA algorithms such as the Kalman Filter (KF) algorithm this matrix is propagated by the dynamic of the model and is updated each time an observation is available (Bouttier, 1993, 1994), namely for each cycle of analysis. The KF algorithm decomposes in two steps. A first step where the state vector and the analysis error covariance matrix are updated using information from the observations. In the second step, the analyzed state vector and the analysis error covariance matrix are propagated from the current assimilation cycle to the next to respectively describe the new background state vector and the new background error covariance matrix. This last step requires the computation of the tangent linear of the dynamical model and computation of the matrix products. As stochastic methods (Evensen, 2009; Moradkhani et al., 2004, 2005b) offer an alternative method to these issues, they require a large number of model integrations and should often be complemented with cost reduction methods such as localization (Tippett et al., 2003; Szunyogh et al., 2008). For large dimension problem and also when the computationnel time is an issue (for operational purpose for instance), the explicit formulation and propagation of the covariances is not possible and the covariance matrix should be described using a model. This is done for instance by parametrizing the covariances using smoothing functions and balance relationships (Daley, 1991; Weaver et al., 2005)

that are formulated with a limited number of parameters such as correlation length scales and standard deviations (Weaver and Courtier, 2001; Pannekoucke et al., 2008a). Reference studies from (Fukumori et al., 1993; Gelb, 1974) describe how the asymptotic behaviour of the Kalman Filter algorithm allows to identify an invariant background error covariance matrix under conditions of observability and controlability. Assuming the statistics of the errors in the system remain stationnary, this asymptotic matrix can thus be used in a simplified KF algorithm, at a much reduced computational cost. In the framework of data assimilation applied to hydraulics, numerous study present the results of the implementation of a KF or EnKF. Only a few studies address the modelling of the covariance matrix itself and the evolution of the background error statistics. (Shiiba et al., 2000) approximates the background error covariance matrix in a KF algorithm with a lower rank matrix using an eigen vector decomposition. (Madsen and Skotner, 2005) proposed an invariant formulation of the KF using a panel of simple covariance functions at the observation point whereas (Ricci et al., 2011) uses at the observation points a steady covariance functions. While these studies offer a practical solution for the implementation a reduced-cost data assimilation algorithm, the mathematical framework that justifies the choice of the correlation functions at the observation point is quite limited.

In this paper, a parametrization for the background error covariance matrix in a sequential filter is proposed as a reduced cost alternative of the EnKF in order to allow for the use of DA in the context of operational flood forecasting and the form of the parametrization is fully justified. The first step of the study stands in the implementation of an EnKF from which the background error covariance model parametrization is derived and validated. It is shown that the parametrized filter successfully emulates the EnKF at a much reduced cost and that the parametrization can be extended for a different observation network. In the context of operational flood-forecasting, running an ensemble of 10 or 20 integrations of a 1D or 2D shallow-water models is not compatible with real-time constraints, computational cost but also with current operational chain implementation. For that reason, the study is carried out on a simplified 1D diffusive flood wave propagation model that approximates the Saint-Venant equations usually derived in 1D and 2D hydraulic models used for flood forecasting. With this simplified model, a large number of members for the EnKF are used, so that we obtain less

than 1% accuracy compared to the asymptotic limit of an infinite number of members. The uncertainty in the model are supposed to be due to uncertainties in the upstream forcing that are not easy to take into account in the formulation of the KF algorithm which is why an ensemble method is favored.

It is first shown that without assimilation, the evolution of the water level anomaly (WLA) covariances, initially prescribed as gaussian, can be described analytically and validated with an ensemble approach via the computation of the covariance matrix $\mathbf{B}_e$. When the diffusion is small, the covariances shape remain gaussian with an increasing correlation length scale and decreasing variance as the signal propagates. Then DA experiments are carried out in the framework of Observing System Experiment (OSE) with a steady observation network. It is shown that an initial correlation function of gaussian shape turns into an anisotropic function at the observation point, with a shorter correlation length scale downstream of the observation point than upstream (which is one of the functions proposed in Madsen et al. 2005), and that the error variance of the state is significantly reduced downstream of the observation point. The resulting converged matrix $\mathbf{B}_{\mathrm{EnKF}}$ can be used as an invariant background error covariance matrix with the deterministic DA algorithm BLUE to emulate the EnKF (EEnKF) with a significantly reduced computational cost. A parametrized model for background error correlation length scale $L_p(x)$ and variance $\sigma^2(x)$ in $\mathbf{B}_{\mathrm{EnKF}}$ is finally established over the whole domain. Along with a diffusion operator (Pannekoucke and Massart, 2008b; Mirouze and Weaver, 2010; Weaver and Mirouze, 2012), this model allows to emulate the EnKF for observing networks where the number of observations and the observation error statistics vary.

The outline of the paper is as follows : Section 2 describes the diffusive flood wave propagation model. It also provides theoretical proof for the generation of a signal with a Gaussian spatial covariance function and for the evolution of the correlation function and length scale without assimilation. The numerical validation, with an ensemble approach, for these theoretical results is presented as well. A brief description of the ensemble based DA algorithms used in the paper and of the diffusion operator are given in Section 3. In Section 4 the results of the EnKF and EEnKF algorithms regarding the evolution of water level and its error statistics are outlined. The parametrization of the reduction of the background error correlation length scale

and variance at the observation point as a function of the observation and background error statistics is also presented and used when the observing network is modified. Some conclusive remarks and perspectives are given in Section 5.

2 Dynamic of the flood wave propagation model

2.1 The flood wave propagation model

2.1.1 Equation

The shallow-water equations can be approximated by the diffusive flood wave propagation equation when the river slope is important:

$$\frac{\partial h}{\partial t} + c\frac{\partial h}{\partial x} = \kappa \frac{\partial^2 h}{\partial x^2}, \tag{1}$$

where h is the Water Level Anomaly (WLA), namely a perturbation to the equilibrium state (h_m, U_m) (where U_m denotes the velocity) such that $U_m = K_{\rm s}(\sin\gamma)^{1/2} h_m^{2/3}$, with $K_{\rm s}$ the Strickler coefficient. Equation (1) is a classical advection-diffusion equation where $c = \frac{5U_m}{3}$ is the advection speed and $\kappa = \frac{U_m h_m}{2\tan\gamma}$ is the diffusion coefficient, with a constant slope γ on a 1D domain defined for $x \in [0, L]$, with $L = 200$ km, discretised in $N = 200$ points. An open boundary condition is imposed downstream with $\frac{\partial h}{\partial t}(L,t) + c\frac{\partial h}{\partial x}(L,t) = 0$ and the upstream boundary condition h_{up} is described in Sect. 2.1.2.

A fourth order Runge-Kutta (RK4) scheme is used in place of an Euler first order temporal scheme allowing for a proper diffusion in the numerical resolution of Eq. (1). Still, it should be noted that for high frequency signals the RK4 scheme can also lead to spurious dispersion thus implying a lower limit for the choice of the initial correlation length (this limit is estimated numerically). This effect is small compared to an explicit Euler scheme (see Annexe 1).

2.1.2 The upstream forcing

The upstream boundary condition is imposed by $h(0,t) = h_{up}(t)$, where h_{up} is characteristic of a flow up to a multiplicative constant. Here h_{up} is modeled as a stationnary Gaussian random process characterised by a temporal auto-covariance function in time $\rho_t(\delta t) = \left\langle h_{up}(t) h^*_{up}(t+\delta t) \right\rangle$ that has a gaussian shape of correlation time scale τ, $\rho_t(\delta t) = q_m^2 e^{-\frac{\delta t^2}{2\tau^2}}$. The Gaussian hypothesis on the random process h_{up} implies that it can be fully described by the first and second order moments of its distribution and the Kalman filter equations can be used. This forcing translates into a WLA signal with a spatial covariance function that has a gaussian shape as shown in Sect. 2.2.1. This choice is made in order to prescribe a known covariance function for the WLA signal and study how it is evolved by the flood wave propagation model.

For the purpose of the theoretical derivation we now switch to a spectral representation of h_{up}. The construction of the upstream forcing requires the formulation of $h_{up}(t)$ using Fourier transform, $h_{up}(t)$ can be written as a sum of harmonic signals

$$h_{up}(t) = \int_{\mathbb{R}} h_{up,\omega} e^{-i\omega t} d\omega. \tag{2}$$

The identification of the $h_{up,\omega}$ coefficients relies on the knowledge of the covariance function $\rho(\delta t)$. Due to the stationnarity of the random process, the complex amplitudes $h_{up,\omega}$ are uncorrelated so that $\left\langle h_{up,\omega} h^*_{up,\omega'} \right\rangle = \rho_\omega \delta(\omega - \omega')$ where δ is the Dirac distribution and where the energy spectrum $\rho_\omega = \frac{\tau}{\sqrt{2\pi}} q_m^2 e^{-\frac{\omega^2\tau^2}{2}}$ is the Fourier transform of

$$\rho_t(\delta t) = \int_{\mathbb{R}} \rho_\omega e^{-i\omega\delta t} d\omega \tag{3}$$

from the *Wiener-Khintchine* theorem.

Since the random process is Gaussian , $h_{up,\omega}$ in Eq. (2) can be written as

$$h_{up,\omega} = \zeta_\omega l_\omega \tag{4}$$

where ζ_ω is a complex gaussian random variable whose module has zero mean and standard deviation 1 and $\zeta_\omega = \zeta^*_{-\omega}$ and the l_ω are complex number of arbitrary phases. Using the definition of the correlation function in time of h_{up} one can show that

$$\rho_t(\delta t) = \int_{\mathbb{R}} |l_\omega|^2 e^{-i\omega\delta t} d\omega. \tag{5}$$

By identification of the two expressions of ρ in Eq. (3) and Eq. (5) it comes that $\mid l_\omega \mid = \sqrt{\rho_\omega}$. Therefore, from a numerical point of view the upstream forcing h_{up} can be built as the inverse Fourier transform of $h_{up,\omega} = \zeta_\omega l_\omega$.

2.2 Study of the covariances dynamic

Given an upstream forcing with a known temporal covariance function for the propagation model, the description of the WLA covariance function is first described analytically in Sect. 2.2.1 and corroborated with an numerical ensemble approach in Sect. 2.2.2 when no data are assimilated. The study of the covariance dynamics when data assimilation is applied is only studied with a numerical approach in Sect. 3 since no analytical solution is available.

2.2.1 Analytical study of the covariances dynamic

Knowing the characteristics of the temporal covariance of the boundary condition flow $h_{up}(t)$, the spatial covariance of the WLA state can be derived. Given the linearity of the problem, the solution $h(x,t)$ can be formulated as the superposition of modal solutions. Assuming that the forcing is a sinusoidal function $h_{up}(t) = q_\omega e^{-i\omega t}$, a modal solution for Eq. (1) is of the form

$$h(x,t) = q_\omega e^{-i\omega t} h_\omega(x) \tag{6}$$

where q_ω is the magnitude of the mode that in the particular case of the upstream forcing can be identified to $\zeta_\omega \mid l_\omega \mid$ in Eq. (4). For any forcing $h_{up}(t)$, the general solution reads

$$h(x,t) = \int_{\mathbb{R}} q_\omega e^{-i\omega t} h_\omega(x) d\omega. \tag{7}$$

In the case of advection only ($\kappa = 0$), the solution of the form given in Eq. (6) is $h_\omega(x) = e^{i\omega\frac{x}{c}}$. Thus the general solution reads

$$h(x,t) = \int_{\mathbb{R}} q_\omega e^{-i\omega t} e^{i\omega\frac{x}{c}}\, d\omega. \tag{8}$$

of which the spatial covariance function ρ is a gaussian, defined as:

$$\begin{aligned}\rho(x, x+\delta x) &= \langle h(x,t) h^*(x+\delta x, t)\rangle \\ &= \int_{\mathbb{R}} |\, q_\omega\,|^2\, e^{i\omega\frac{\delta x}{c}}\, d\omega = \rho\left(\frac{\delta x}{c}\right)\end{aligned} \tag{9}$$

In summary, in the case of advection only, a forcing signal with a gaussian temporal covariance function translates into a WLA signal with a gaussian spatial covariance function of constant length scale $L_0 = c\tau$ and constant variance $\sigma_0^2 = q_\omega^2$.

In the case of advection and small diffusion, which reads $\kappa \ll cx$, a straightforward expansion leads to,

$$h_\omega(x) = e^{(\frac{c}{2\kappa} - \frac{\sqrt{c^2 - 4i\omega\kappa}}{2\kappa})x} \approx e^{i\omega\frac{x}{c} - \frac{\omega^2\kappa}{c^3}x} \tag{10}$$

and a more elaborated asymptotic analysis shows that ρ can locally be approximated by

$$\rho(x, x+\delta x) \approx \int_{\mathbb{R}} \frac{q_m^2}{\sqrt{2\pi}} L_p(0) e^{-\frac{\omega^2 L_p^2(x)}{2}} e^{-i\omega\delta x}\, d\omega = q_m^2 \frac{L_p(0)}{L_p(x)} e^{-\frac{\delta x^2}{2L_p(x)^2}}, \tag{11}$$

that is a locally Gaussian covariance function of correlation length scale :

$$L_p(x) = \sqrt{L_0^2 + 4\kappa\frac{x}{c}} \tag{12}$$

and variance :

$$\sigma^2(x) = \sigma^2(0)\frac{L_p(0)}{L_p(x)}. \tag{13}$$

In summary, in the case of advection and small diffusion a forcing signal with a gaussian temporal covariance function translates into a WLA signal with a spatial covariance function that can be approximated by a gaussian of length scale $L_p(x)$.

2.2.2 Validation with an ensemble approach

These theoretical results are validated computing the covariance matrix $\mathbf{B}_e$ of an ensemble of N_e WLA states $\mathbf{x}^k = (h_{1,k}, \cdots, h_{N,k})$ on the 1D domain $[0, L]$ where N is the number of grid point, generated with different forcings $h_{up,k}(t)$ with k $\in [1, N_e]$ that follows the statistics described in Sect. 2.1.2. The correlation length scale L_p is computed by the gaussian based approximation (Pannekoucke et al., 2008a):

$$L_p(x) = \frac{\delta x}{\sqrt{2(1-\rho(x, x+\delta x))}} \tag{14}$$

where $\rho(x, x+\delta x)$ is the correlation estimated from data or analytical formula such as Eq. (11).

Fig. 1-a displays the covariance function at 3 different points of the 1D domain for the advection only case (dashed lines) and for the advection-diffusion case (solid lines), for $N_e = 10000$ (so that the sampling noise is less than $10^{-2}\ m^2$), $\sigma^2(0) = 1\ m^2$ and $\tau = 5 \times 10^3 s$. In the first case, the initially gaussian function is advected. The characteristics of the covariance function remain unchanged as illustrated in Fig. 1-b where L_p, estimated from Eq. (14) is constant (dashed thick line), and in agreement with the theoretical value L_0 (dashed thin line). When diffusion occurs, the covariance function of the WLA state is diffused as shown in Fig.1-a and L_p increases with x (solid thick line in Fig. 1-b), still in agreement with the theoretical value from Eq. (12) (solid thin line in Fig. 1-b). The WLA variance represents the maximum amplitude of the covariance functions in Fig. 1-a. The initial value prescribed at $1\ \mathrm{m}^2$ remains constant for the advective case (dashed lines) and decreases with x for the advective-diffusive (solid lines) case in agreement with the theoretical results in Eq. (13).

3 Data assimilation algorithms

The classical equations for EnKF are presented in Sect. 3.1 while the EEnKF is presented in Sect. 3.2. For these algorithms, the background error covariance matrix is characterized by the correlation length scale $L_p(x)$ and the variance $\sigma^2(x)$ for which a parametrized model is presented is Sect. (4.3). Given these information only, the diffusion operator described in Sect. (3.3) allows to fully describe a covariance matrix to be used in the EEnKF for various observing networks.

3.1 The Ensemble Kalman Filter algorithm

The EnKF algorithm (Evensen, 2009) is implemented on Eq. (1), using an OSE (Observing System Experiment) framework. A reference run is integrated using a given forcing $h_{up}^{\rm true}(t)$, to simulate the *true* WLA $h^{\rm true}(x,t)$. The observation

$$h^{\rm obs}(x_{obs},t) = h^{\rm true}(x_{obs},t) + \epsilon^o(t) \tag{15}$$

is then calculated in the middle of the 1D domain $x_{\rm obs} = \frac{L}{2}$ where $\epsilon^o(t)$ is a Gaussian noise defined by its standard deviation σ^o ($\sigma^o = 0.2354\ m$ in the following), thus defining the observation vector $\mathbf{y}^o$. The background trajectories $h_k^b(x,t)$ for the ensemble approach are integrated using a perturbed set of forcing $h_{up,k}(t)$ with k $\in [1,N_e]$, defining the background vectors $\mathbf{x}^{b,k}(t)$ for the DA analysis at time t. The observation frequency is set to 3 model time steps. In the following DA is applied over a cycle between two observation times $t=i$ and $t=i+1$ (assimilation cycle $i+1$).

As illustrated on Fig. 2 for the assimilation cycle $i+1$, the ensemble of previously analyzed states $\mathbf{x}_i^{a,k}$ are propagated by the diffusive flood wave model $M_{i,i+1}$ from the observation time i to $i+1$ to provide the background states $\mathbf{x}_{i+1}^{b,k} = M_{i,i+1}(\mathbf{x}_i^{a,k})$ over which the background error covariance matrix $\mathbf{B}_{\rm EnKF,i+1}$ is computed.

$$\mathbf{B}_{\rm EnKF,i+1} = \frac{1}{N_e-1}\sum_{k=1}^{N_e}\left(\mathbf{x}_{i+1}^{b,k} - \bar{\mathbf{x}}_{i+1}\right)\left(\mathbf{x}_{i+1}^{b,k} - \bar{\mathbf{x}}_{i+1}\right)^T. \tag{16}$$

where $\bar{\mathbf{x}}_{i+1} = \frac{1}{N_e}\sum_{k=1}^{N_e}\mathbf{x}_{i+1}^{b,k}$ and T stands for the transposition operator.

The assimilation step at $i+1$ consists in assimilating a perturbed observation vector $\mathbf{y}_{i+1}^{o} + \epsilon_{i+1}^{o,k}$ (Burgers el al., 1998) to correct the background vector $\mathbf{x}_{i+1}^{b,k}$, using the Kalman Filter gain matrix $\mathbf{K}_{\mathrm{EnKF,i+1}}$:

$$\mathbf{x}_{i+1}^{a,k} = \mathbf{x}_{i+1}^{b,k} + \mathbf{K}_{\mathrm{EnKF,i+1}}(\mathbf{y}_{i+1}^{o} + \epsilon_{i+1}^{o,k} - \mathbf{H}(\mathbf{x}_{i+1}^{b,k})) \quad \text{with} \tag{17}$$

$$\mathbf{K}_{\mathrm{EnKF,i+1}} = \mathbf{B}_{\mathrm{EnKF,i+1}}\mathbf{H}^T(\mathbf{H}\mathbf{B}_{\mathrm{EnKF,i+1}}\mathbf{H}^T + \mathbf{R})^{-1}. \tag{18}$$

Assuming that the observation network remains the same, after 1000 assimilation cycles $\mathbf{B}_{\mathrm{EnKF,i}}$ converges to a steady matrix denoted $\mathbf{B}_{\mathrm{EnKF}}$ (its associated correlation matrix is $\mathbf{C}_{\mathrm{EnKF}}$, $\mathbf{C}_{\mathrm{EnKF}}$ is computed with $\Sigma\mathbf{B}_{\mathrm{EnKF}}\Sigma^T$ where Σ is the diagonal matrix with the inverse of the standard deviation of $\mathbf{B}_{\mathrm{EnKF}}$ on its diagonal). (Li and Xiu, 2008) showed that ensemble errors due to the Monte Carlo sampling in EnKF can be dominant compared to other errors (numerical or model errors) and that in order to estimate converged statistics in $\mathbf{B}_{\mathrm{EnKF,i+1}}$ a large number of members N_e is required. Indeed it can be shown that $N_e = 10000$ members are necessary to reduce the sampling noise on the background error covariance matrix to less than $10^{-2}m^2$.

3.2 Emulation of the EnKF algorithm

The BLUE (Best Linear Unbiased Estimator, Bouttier and Courtier 1999) algorithm can be viewed as a simplification of the Kalman Filter in which the background error covariance matrix is not propagated over the cycles. The analysis equation Eq. (17) is applied sequentially with a constant matrix $\mathbf{B}_{\mathrm{BLUE}}$:

$$\mathbf{K}_{\mathrm{BLUE}} = \mathbf{B}_{\mathrm{BLUE}}\mathbf{H}^T(\mathbf{H}\mathbf{B}_{\mathrm{BLUE}}\mathbf{H}^T + \mathbf{R})^{-1}. \tag{19}$$

If $\mathbf{B}_{\mathrm{BLUE}} = \mathbf{B}_e$ (where $\mathbf{B}_e$ is the covariance matrix computed without assimilation in Sect. 2.2.2), one misses the fact that the background error covariance matrix should be impacted by the previous assimilation steps. When the BLUE algorithm is applied with $\mathbf{B}_{\mathrm{BLUE}} = \mathbf{B}_{\mathrm{EnKF}}$, the

algorithm is called the Emulated EnKF and is denoted by EEnKF. Once $\mathbf{B}_{\mathrm{EnKF}}$ is computed, this algorithm only requires the integration of a single member (with a single upstream forcing h_{up}) and its computational cost is thus significantly lower than the EnKF.

Still, the BLUE algorithm can of course be applied to an ensemble of forcings $h_{up,k}$ so that a comparison of the correlation length scale and variance of the EnKF and this ensemble of BLUE algorithms can be made, as pictured in Fig. 2. In the following, the ensemble of BLUE analysis using $\mathbf{B}_{\mathrm{BLUE}} = \mathbf{B}_e$ or $\mathbf{B}_{\mathrm{BLUE}} = \mathbf{B}_{\mathrm{EnKF}}$ are called respectively $\mathrm{EnBLUE}_{\mathrm{B_e}}$ and $\mathrm{EnBLUE}_{\mathrm{B_{EnKF}}}$. The correlation length scale and variance of the covariance matrix computed over the members for both algorithm are compared to those of $\mathbf{B}_{\mathrm{EnKF}}$ in Sect. 4.

3.3 The diffusion operator

The columns and lines of the matrix $\mathbf{B}_{\mathrm{EnKF}}$ contain the discretization of the background error covariance functions for each grid point of the domain. In this section we present how the covariance functions in $\mathbf{B}_{\mathrm{EnKF}}$ can be fully described simply using the diagnosed correlation length scale and variance when the EnKF is converged. This can be done using the diffusion operator (Weaver and Courtier, 2001) that comes down to formulating the matrix, vector product $\mathbf{Bx}$ instead of formulating the matrix $\mathbf{B}$. The diffusion operator is a mathematical tool that allows to model covariances and correlations for DA algorithm. It is widely used in meteorology and oceanography, where the dimension of the control vector is large and the covariance matrix should be formulated as an operator (applied to a vector) rather than as a matrix. The solution of a 1D pseudo-diffusion equation on an infinite domain for a pseudo-time T reads

$$\eta(x,T) = \frac{1}{\sqrt{4\pi\tilde{\kappa}T}} \int_{-\infty}^{+\infty} e^{-\frac{(x-x')^2}{4\tilde{\kappa}T}} \eta_0(x')\,dx' \tag{20}$$

where $\tilde{\kappa}$ is the constant pseudo-diffusion coefficient, η_0 is the initial condition and η vanishes as $x \to \pm\infty$. Equation (20) is the convolution product of $g(x) = \frac{1}{\sqrt{4\pi\tilde{\kappa}T}} e^{-\frac{x^2}{4\tilde{\kappa}T}}$ and $\eta_0(x)$. Since g is positive definite, Eq. (20) describes a covariance operator correlation length scale $L_p = \sqrt{2\tilde{\kappa}T}$ applied to η_0. The corresponding correlation operator is obtained with a normalization

factor $\lambda = \sqrt{4\pi\tilde{\kappa}T}$. The multiplication of the desired variance σ^2 finally describes the expected covariance operator. This result can be extended to heterogeneous diffusion tensor $\tilde{\kappa}(x) = \frac{L_p^2(x)}{2T}$ (Pannekoucke and Massart, 2008b). In the following, $L_p(x)$ and $\sigma^2(x)$ are specified with the analytical equations Eq. (12) and Eq. (13) away from the observation points and with the abacus from EnKF experiments (presented in Sect. 4) at the observation point. From a numerical point of view, the pseudo-diffusion model is applied at each grid point to a Dirac function. The result is then normalized and multiplied by the background error variance to provide the background error covariance function at the grid point and thus the complete background error covariance matrix with correlation length scale $L_p(x)$ and variance $\sigma^2(x)$.

4 Results

4.1 Comparison of EnKF and EEnKF results

The EnKF is applied for the observing network described in Sect. 3.1 with the initial background error covariance matrix $\mathbf{B}_e$ computed over the integrated members without DA (Sect. 2.2.2) - the associated correlation function is noted $\mathbf{C}_e$. Figure 3-a illustrates how the initially isotropic correlation function in $\mathbf{C}_e$ (dashed line) at the observation point ($x_{obs} = 100$) is modified by the analysis and propagation steps of the EnKF algorithm, at the end of the assimilation procedure. Considering a steady observation network, the shape of the correlation function in $\mathbf{C}_{\mathrm{EnKF}}$ (solid line) converges towards an anisotropic function with a shorter correlation length scale downstream of the observation point than upstream. The correlation between the observation point and its neighbors is reduced since information at the observation point is introduced at this location by the analysis procedure through the observation vector at the previous analysis cycles.

The correlation length scale L_p and variance σ^2 are computed for the EnKF as well as for the ensemble of BLUE analysis using either the initial covariance matrix $\mathbf{B}_e$ ($\mathrm{EnBLUE}_{\mathrm{B_e}}$ algorithm) or $\mathbf{B}_{\mathrm{EnKF}}$ ($\mathrm{EnBLUE}_{\mathrm{B_{EnKF}}}$ algorithm). Equation (14) is used to estimate the correlation length scale L_p at each point away from the observation point. Fig. 4-a shows that the evolu-

tion of the background error correlation length scale for the EnKF (thick solid line) follows the theory (thin solid line) upstream of the observation point. At the observation point, where a discontinuity occurs, the upstream and downstream correlation length scales (respectively L_p^- and L_p^+) differ. For the EnKF algorithm (thick solid line), the reduction of the correlation length scale spreads over the entire domain downstream of the observation point. This result is well reproduced by the $\mathrm{EnBLUE}_{\mathbf{B}_{\mathrm{EnKF}}}$ (thick dashed line that overlaps the thick solid one) algorithm, whereas for $\mathrm{EnBLUE}_{\mathbf{B}_{\mathrm{e}}}$ (thin dashed line), the reduction is local. The upstream correlation length scale at the observation point L_p^- is approximated by Eq. (12) with $x = 100km$, meaning that the assimilation has no impact on the correlation function upstream of the observation point, which is consistent with the flood wave approximation in the propagation model. The downstream correlation length scale at the observation point L_p^+ should be extrapolated from the ensemble based estimate function for $x > 100km$ that is of the form given in Eq. (12). L_p^+ can thus be used in place of L_0 in Eq. (12) to derive the analytical function in Eq. (23) for the rest of the domain.

The variances of the background error covariance matrix are presented in Fig. 4-b. The standard deviations at the observation point before assimilation and after assimilation are respectively $\sigma^- = \sqrt{\mathbf{B}_e(x_{obs},x_{obs})}$ and $\sigma^+ = \sqrt{\mathbf{B}_{\mathrm{EnKF}}(x_{obs},x_{obs})}$. The error variance is significantly reduced at the observation point and beyond with the EnKF (thick solid line), compared to the initially prescribed variances (thin solid line). However, when $\mathbf{B}$ is kept invariant and isotropic ($\mathrm{EnBLUE}_{\mathbf{B}_{\mathrm{e}}}$), the reduction of the variance is only located in the close neighboring of the observation point (thin dashed line); the invariant matrix is not optimal. In this case, the merits of using a DA algorithm that evolves the background error statistics with the dynamics are demonstrated; the shorten length scale of $\mathbf{B}_{\mathrm{EnKF}}$ prevents from overcorrecting downstream of the observation once information from the observation is taken into account. The $\mathrm{EnBLUE}_{\mathbf{B}_{\mathrm{EnKF}}}$ algorithm shows the same results (thick dashed line that overlaps the thick solid line) as the EnKF.

Since the analysis for the different members in the $\mathrm{EnBLUE}_{\mathbf{B}_{\mathrm{EnKF}}}$ are independent, these results demonstrate that the computation of the EnKF converged background covariance matrix can be achieved at first and then used with the EEnKF with a single analysis algorithm such

as BLUE with a much reduced computational cost, thus emulating the EnKF. This result is of particular interest in the framework of real-time forecasting where a single analysis is usually carried out instead of an ensemble of analysis (what we are looking for is the forecasted WLA state and not its covariance matrix). Assuming that an EnKF analysis has previously been carried out, the real-time data assimilation procedure can be achieved with a non expensive EEnKF algorithm, using $\mathbf{B}_{\mathrm{EnKF}}$ as the invariant background error covariance matrix for a single EEnKF analysis. Indeed the computational cost of an EnKF with members that are run sequentially depends linearly on the total number of members. In this study 10000 members are used for the EnKF to ensure that the sampling noise of the background error covariance matrix is less than $10^{-2}m^2$, see Sect. (2.2.2), this implies that an EEnKF with a single analysis is 10000 times cheaper than a full EnKF.

This approach is illustrated in Fig. 5 where the improvement of the WLA for a single BLUE analysis (with $\mathbf{B} = \mathbf{B}_e$) and for a single EEnKF is shown: the analyzed state h^a for the EEnKF (thick dashed line) is significantly closer to the true state h^{true} (thin solid line) than the analyzed state for the BLUE using $\mathbf{B}_e$(thin dashed line).

4.2 Influence of the observation error standard deviation on the correlation length scale and the variance

Both the reduction of the background error variance and the correlation length scale depend on the ratio $r = \frac{\sigma^-}{\sigma_o}$. In section (4.1) σ_o has been chosen so that $r = 3$. In the following σ^- is assumed to be fixed and σ_o varies to represent different observation error statistics so that r ranges from 0.5 to 4. Fig. 6 shows the diagnosed correlation length scale (a), correlation functions at x_{obs} (b) and variance over the domain (c) with the EnKF for different values of the ratio r. When $r = 0.5$, the observation error standard deviation is large (the observations are not reliable), the correlation function at the observation point is close to the initial isotropic one $\left(\frac{L_p^+}{L_p^-} \simeq \frac{1}{2}\right)$, the reduction of variance is small, hence the analyzed WLA remains close to the background. On the contrary when $r = 4$ the observation error standard deviation is small (the observations are reliable), the correlation function at the observation point evolves into an anisotropic function

$\left(\frac{L_p^+}{L_p^-} \ll 1\right)$, the reduction of variance at the observation point and downstream of that point is significant, hence the assimilation provides good results and the analyzed WLA is brought closer to the true state. Figures 6-(a) and (b) show that when the observation error decreases (r increases), the ratio $\frac{L_p^+}{L_p^-}$ decreases (the anisotropy of the correlation function at the observation point increases). Similarly, Fig. 6-(c) shows that when the observation error decreases (r increases), the ratio $\frac{\sigma^+}{\sigma^-}$ decreases. It should be noted that the data assimilation algorithm leads to a reduction of the variance inside the interval $\mathbf{I}_\varepsilon = [x_{obs} - \varepsilon^-; x_{obs} + \varepsilon^+]$ $\left(\text{where } \varepsilon^- = 2.L_p^- \text{ and } \varepsilon^+ = 2 \times \max_{r \in [0.5;4]} \left\{L_p^+\right\}\right)$. In order to describe the variance outside this interval downstream of the observation point with Eq. (13), the variance in $x_{obs} + \varepsilon^+$ should be estimated and used in place of $\sigma^2(0)$. In the following, for $x \in \mathbf{I}_\varepsilon$, the ratio $\frac{\sigma^+(x)}{\sigma^-(x)}$ is defined with $\sigma^-(x) = \sqrt{\mathbf{B}_e(x,x)}$ and $\sigma^+(x) = \sqrt{\mathbf{B}_{\text{EnKF}}(x,x)}$.

A set of EnKF experiments is achieved for $r \in [0.5;4]$ by 0.5 increments of r. It is shown that the relation between r and $\frac{L_p^+}{L_p^-}$ at the observation point can be described with an abacus built from a linear regression in logarithmic scales, represented in dashed lines on Fig. 7-(a), while the results of the EnKF are represented with thick solid lines:

$$\ln\left(\frac{L_p^+}{L_p^-}\right) = \alpha \ln(r) + \beta. \tag{21}$$

Similarly, for any $x \in \mathbf{I}_\varepsilon$, the relation between r and $\frac{\sigma^+(x)}{\sigma^-(x)}$ can be described with an abacus built from a linear regression in logarithmic scales using the same set of EnKF experiments. The linear regresion is shown in Fig. 7-b with a thick dashed line, for $x = x_{obs} + \varepsilon^+$. In the following, $\sigma^-(x_{obs} + \varepsilon^+)$ and $\sigma^+(x_{obs} + \varepsilon^+)$ are respectively noted σ_ε^+ and σ_ε^- and are represented in Fig. 6-(c) for $r = 0.5$:

$$\ln\left(\frac{\sigma_\varepsilon^+}{\sigma_\varepsilon^-}\right) = \gamma \ln(r) + \delta. \tag{22}$$

Abacus in Eqs. (21-22) have been established for $x_{obs} = 100$ but numerical experiments show that the coefficients α, β, γ and δ do not depend on the position of the observation point

x_{obs}. These equations allow for the quantification of the impact of the assimilation and the dynamics on the background correlation length scale and variance. More importantly, they lead to the parametrization of the correlation length over the whole domain (see Sect. (4.3)) allowing for the modeling of the converged background error covariance matrix from the EnKF. Using this parametrization, the EnKF can be emulated for any observation network described by the number of observations, their locations and the variance of their respective error (see Sect. (4.4)).

4.3 Parametrized model for correlation length scale and variance reduction at the observation point

This section presents how Eqs. (21-22-23-24) provide a parametrized model for the background error correlation length scale $L_p(x)$ and variance $\sigma^2(x)$, for any observation error variance σ_o^2. These information are used to model the converged background error covariance matrix from the EnKF using the diffusion operator presented in Sect. (3.3) in order to integrate the EEnKF. While the methodology is applicable for any $r \in [0.5;4]$, illustrations are given here for $r = 0.75$.

Away from the observation point the expression of the correlation length scale $L_p(x)$ is derived from the analytical expression in Eq. (12):

$$L_p(x) = \sqrt{L_p^2(0) + 4\kappa\frac{x}{c}},\ x \leq x_{obs} \quad \text{and} \quad L_p(x) = \sqrt{(L_p^+)^2 + 4\kappa\frac{x-x_{obs}}{c}},\ x > x_{obs} \tag{23}$$

where L_p^+ is the downstream correlation length scale at the observation point computed using the abacus from Eq. (21) with $L_p^- = \sqrt{L_p^2(0) + 4\kappa\frac{x_{obs}}{c}}$. Figure 8-(a) illustrates how the correlation length scale obtained from the EnKF (thick solid line) compares with the parametrized one (thick dashed line). It should be noted that here, the EnKF is run for validation purpose only.

The variance $\sigma^2(x)$ away from the observation point is derived from the analytical expres-

sion in Eq. (13):

$$\begin{aligned}
\sigma^2(x) &= \sigma(0)^2 \frac{L_p(0)}{\sqrt{L_p^2(0)+4\kappa\frac{x}{c}}}, && x \leq x_{obs}-\varepsilon^- \\
\sigma^2(x) &= \left(\sigma_\varepsilon^+\right)^2 \frac{L_p(x_{obs}+\varepsilon^+)}{\sqrt{L_p^2(x_{obs}+\varepsilon^+)+4\kappa\frac{x-(x_{obs}+\varepsilon^+)}{c}}}, && x > x_{obs}+\varepsilon^+
\end{aligned} \tag{24}$$

where $L_p(x_{obs}+\varepsilon^+)$ is computed with Eq. (23) and $\left(\sigma_\varepsilon^+\right)^2$ is the background error variance at $x_{obs}+\varepsilon^+$ computed using the abacus from Eq. (22). In the neighboring of the observation, the reduction of variance is computed from the application of the Eq. (22) for $x \in \mathbf{I}_\varepsilon$. Depending on how many points in $\mathbf{I}_\varepsilon$ are used to derive linear regressions as shown in Eq. (22), the reduction of variance is more or less finely described. Figure 8-(b) illustrates how the variance obtained from the EnKF compares with the parametrized variance using only 2 points in $\mathbf{I}_\varepsilon$ (thin dashed line) or 25 points (thick dashed line that almost overlaps the thick solid line). It is shown that when a crude approximation for the variance in the neighboring of the observation point is used, the results of the EEnKF are degradated.

4.4 Application to any observation network

The parametrized model for background error correlation length scale and variance is validated for DA experiments with a single observation point and various observation error standard deviations. Here the results are extended to an observing network with two observation points $x_{obs,1}=50$ and $x_{obs,2}=150$ with respective observation error variance $\sigma_{o,1}^2$ and $\sigma_{o,2}^2$. The distance between $x_{obs,1}$ and $x_{obs,2}$ is bigger than the background error correlation length scale diagnosed previously. In order to estimate the correlation length scale and variance in the neighboring of $x_{obs,2}$ and downstream, the impact of the assimilation at $x_{obs,1}$ should be taken into account.

The parametrized correlation length scale $L_p(x)$ for this observation network is given by:

$$
\begin{aligned}
L_p(x) &= \sqrt{L_p^2(0) + 4\kappa\tfrac{x}{c}}, && x \le x_{obs,1} \\
L_p(x) &= \sqrt{(L_{p_1}^+)^2 + 4\kappa\tfrac{x - x_{obs,1}}{c}}, && x_{obs,1} < x \le x_{obs,2} \\
L_p(x) &= \sqrt{(L_{p_2}^+)^2 + 4\kappa\tfrac{x - x_{obs,2}}{c}}, && x > x_{obs,2}
\end{aligned}
\tag{25}
$$

where $L_{p_1}^+$ and $L_{p_2}^+$ are the downstream correlation length scales at point $x_{obs,1}$ and $x_{obs,2}$ respectively. $L_{p_1}^+$ and $L_{p_2}^+$ are computed using the abacus from Eq. (21) with the uptream correlation length scales $L_{p_1}^- = \sqrt{L_p^2(0) + 4\kappa\frac{x_{obs,1}}{c}}$ and $L_{p_2}^- = \sqrt{\left(L_{p_1}^+\right)^2 + 4\kappa\frac{x_{obs,2} - x_{obs,1}}{c}}$.

The parametrized variance $\sigma^2(x)$ is given by:

$$
\begin{aligned}
\sigma^2(x) &= \sigma(0)^2 \frac{L_p(0)}{\sqrt{L_p^2(0) + 4\kappa\frac{x}{c}}}, && x \le x_{obs,1} - \varepsilon^- \\
\sigma^2(x) &= (\sigma_{\varepsilon,1}^+)^2 \frac{L_p(x_{obs,1} + \varepsilon^+)}{\sqrt{L_p^2(x_{obs,1} + \varepsilon^+) + 4\kappa\frac{x - (x_{obs,1} + \varepsilon^+)}{c}}}, && x_{obs,1} + \varepsilon^+ < x \le x_{obs,2} - \varepsilon^- \\
\sigma^2(x) &= (\sigma_{\varepsilon,2}^+)^2 \frac{L_p(x_{obs,2} + \varepsilon^+)}{\sqrt{L_p^2(x_{obs,2} + \varepsilon^+) + 4\kappa\frac{x - (x_{obs,2} + \varepsilon^+)}{c}}}, && x > x_{obs,2} + \varepsilon^+
\end{aligned}
\tag{26}
$$

where $L_p(x_{obs,1} + \varepsilon^+)$ and $L_p(x_{obs,2} + \varepsilon^+)$ are computed using the abacus Eq. (25) and where $\sigma_{\varepsilon,1}^+$ and $\sigma_{\varepsilon,2}^+$ are computed using Eq. (22) with respectively $\sigma_{\varepsilon,1}^-$ and $\sigma_{\varepsilon,2}^-$ described from Eq. (13).

Thus the background error correlation length scale and variance are fully parameterized as shown in Fig. 9 and are very close to the correlation length scale and variance diagnosed with the EnKF (here again, run is validation purpose only). The parametrized $L_p(x)$ and $\sigma^2(x)$ are used with the diffusion operator to model the converged background error covariance matrix from the EnKF with two observation points, $\widetilde{\mathbf{B}}_{\mathrm{EnKF}}$. The analyzed WLA for the EEnKF algorithm using the constant $\widetilde{\mathbf{B}}_{\mathrm{EnKF}}$ matrix (only one member to integrate) compares very well with that

of the EnKF algorithm as shown in Fig. 10, for a much reduced computational cost, which is compatible with real-time flood-forecasting constraint.

This method can be extended to any observation network building on the characterization of the correlation length scale and variance reduction that established in Eqs. (21-22) in the neighboring of an observation point and the analytical solutions in Eqs. (12-13) away from the observation point.

5 Summary and conclusions

This study describes the evolution of the background error covariance matrix with an EnKF algorithm in the framework of OSE, for a steady observation network, meaning that the observation frequency and locations remain the same through the assimilation cycles. It was shown that the filter converges to an optimal and invariant matrix, characterized, at the observation point by an anisotropic correlation function with a shorter correlation length scale downstream of the observation point and a reduction of the error variance. As the model is forced at its upstream boundary with a random variable characterized by a gaussian correlation function over time, the background error correlation length scale and the background error variance away from the observation points are described by analytical equations. The correlation length scale and variance reduction at the observation point are described by abacus as functions of the observation error variance. The construction of such abacus requires the integration of a very small set of EnKF experiments and can be used for any observation point. Thus a parametrization of the background error correlation length scale and variance is proposed over the entire simulation domain for any observation network given the number of observations, their locations and their respective error variance.

The parametrized model was then used to build the invariant matrix using a diffusion operator, which is a convenient tool especially for large dimension problems. This methods allows to emulate the EnKF at a much reduced computational cost with a deterministic BLUE algorithm where the background error covariance matrix does not evolve in time. It was shown that the resulting algorithm, denoted by EEnKF (for Emulated EnKF), leads to similar results to the EnKF allowing for the use of DA for real-time flood forecasting.

A perspective for this work is to study how the background error statistics evolve with the full shallow-water equations instead of the flood wave propagation model. This would give a closer idea of what to expect with an operational hydraulic model such as MASCARET, MIKE or LISFLOOD. In this context, it is expected that the impact of the assimilation would also spread upstream of the observation points thus leading to the reduction of the background error correlation length scale and variance on both side of the observing stations. Whether the

resulting correlation would be isotropic or not, and to what extend still need to be investigated.

Appendix A

Let us consider the advection-diffusion equation :

$$\begin{cases} \partial_t h + c\partial_x h = \kappa \partial_x^2 h & (x,t) \in [0,L] \times \mathbb{R}^+ \\ h(x,0) = h_0(x) & x > 0 \\ h(0,t) = h_{up}(t) & t > 0 \\ \partial_t h + c\partial_x h = 0 & t > 0 \end{cases} \tag{27}$$

Let us denote $\Delta x = \frac{L}{N}$ the space step and Δt the time step. Let us denote also h_j^i the value of the discrete solution of (27) at point $(i\Delta t, j\Delta x)$. Using the finite difference method one can write the following numerical scheme for equation (27) :

$$\begin{cases} \frac{h_j^{i+1} - h_j^i}{\Delta t} + c\frac{h_{j+1}^i - h_{j-1}^i}{2\Delta x} = \kappa \frac{h_{j+1}^i - 2h_j^i + h_{j-1}^i}{\Delta x^2} & j = 2, \dots, N-1 \\ h_j^0 = h_0(j\Delta x) & j = 1, \dots, N \\ h_j^i = q(i\Delta t) & j = 1 \\ \frac{h_j^{i+1} - h_j^i}{\Delta t} + c\frac{h_j^i - h_{j-1}^i}{\Delta x} = 0 & j = N \end{cases} \tag{28}$$

The scheme (28) allows for the numerical resolution of the advection-diffusion equation with an error. Now let us demonstrate why the numerical solution of (28) is a solution of the equation (31). Using Taylor expansion one can write :

$$\begin{aligned} \frac{u_j^{i+1} - u_j^i}{\Delta t} &= (\partial_t u)_j^i + \frac{\Delta t}{2}(\partial_t^2 u)_j^i + \mathcal{O}(\Delta t^2) \\ \frac{u_{j+1}^i - u_{j-1}^i}{2\Delta x} &= (\partial_x u)_j^i + \frac{\Delta x^2}{6}(\partial_x^3 u)_j^i + \mathcal{O}(\Delta x^3) \\ \frac{u_{j+1}^i - 2u_j^i + u_{j-1}^i}{\Delta x^2} &= (\partial_x^2 u)_j^i + \frac{\Delta x^2}{12}(\partial_x^4 u)_j^i + \mathcal{O}(\Delta x^3) \end{aligned} \tag{29}$$

Thereafter to lighten the notations we note u instead of u_j^i the value of u at point $(i\Delta t, j\Delta x)$. It comes from (28) and (29) that u solves the following equation :

$$\partial_t u + c\partial_x u - \kappa\partial_x^2 u + \underbrace{\frac{\Delta t}{2}\partial_t^2 u + \Delta x^2\left(\frac{c}{6}\partial_x^3 u - \frac{\kappa}{12}\partial_x^4 u\right) + \mathcal{O}(\Delta t^2, \Delta x^3)}_{\varepsilon} = 0 \tag{30}$$

Using the derivation of the advection-diffusion equation with respect to time one can write the expession of the temporal derivatives $\partial_t^2 u$ with respect to the spatial derivatives : $\partial_t^2 u = c^2\partial_x^2 u - 2c\kappa\partial_x^3 u + \kappa^2\partial_x^4 u$ that allows for, using (30), writing :

$$\frac{\partial \nu}{\partial t} + c\frac{\partial \nu}{\partial x} + \underbrace{\left(\frac{c\Delta x^2}{6} - c\kappa\Delta t\right)}_{\mu}\frac{\partial^3 \nu}{\partial x^3} = \underbrace{\left(\kappa - \frac{c^2\Delta t}{2}\right)}_{\kappa'}\frac{\partial^2 \nu}{\partial x^2} - \left(\frac{\kappa^2\Delta t}{2} - \frac{\kappa\Delta x^2}{12}\right)\frac{\partial^4 \nu}{\partial x^4} + \mathcal{O}(\Delta t^2, \Delta x^3) \tag{31}$$

Practically, the numerical model solves Eq. (31) instead of Eq. (28) inducing spurious dispersion due to the term μ (particularly for high frequencies) and spurious diffusion due to the term κ' (as neither c nor Δt is equal to zero, the numerical diffusion κ' is not equal to the physical diffusion κ). Numerical experiments highlighted that the numerical model can under or over estimate the diffusion and for $\kappa' < 0$ the scheme is unstable because the CFL (Courant-Friedrichs-Lewy) condition is not verified (Quarteroni et al. [12]).

References

Andreadis, K.M., Clark, E.A., Lettenmaier, D.P. and Alsdorf, D.E.: Prospects for river discharge and depth estimation through assimilation of swath-altimetry into a raster-based hydrodynamics model, Geophys. Res. Lett., 34, L10403, doi:10.1029/2007GL029721, 2007.

Biancamaria, S., Durand, M., Andreadis, K.M. , Bates, P.D., Boone, A., Mognard, N.M., Rodríguez, E., Alsdorf, D.E., Lettenmaier, D.P. and Clark, E.A.: Assimilation of virtual wide swath altitmetry to improve Arctic river modeling, Remote Sens. Environ., vol. 115, no. 2, pp. 373-381, 2011.

Bouttier, F.: The dynamics of error covariances in an assimilation system. Tellus A, 45, 408–423, 1993.

Bouttier, F.: A dynamical estimation of error covariances in an assimilation system. Mon. Wea. Rev., 122, 2376–2390, 1994.

Daley, R.: Atmospheric data analysis. Cambridge atmospheric and space science series. Cambridge University Press, 1991.

Drogue, G., Pfister, L., Leviandier, T., El Idrissi, A., Iffly, J.-L., Matgen, P., Humbert, J., and Hoffmann, L.: Simulating the spatio-temporal variability of streamflow response to climate change scenarios in a mesoscale basin, J. Hydrol., 293, 255–269, 2004.

Durand, M., Lee-Lang, F., Lettenmaier, D.P., Alsdorf, D.E., Rodriguez, E., and Esteban-Fernandez, D.: The Surface Water and Ocean Topography Mission: Observing Terrestrial Surface Water and Oceanic Submesoscale Eddies, IEEE, 98, 766–779, 2010.

Evensen, G. : Data Assimilation - The Ensemble Kalman Filter, Springer, 2009.

Fukumori, I., Benveniste, J., Wunsch, C., and Haidvogel, D. B.: Assimilation of sea surface topography into an ocean circulation model using a steady-state smoother. J. Phys. Oceanogr, Vol. 23, Issue 8, 1831-1855, 1993.

Gelb, A. : Applied optimal estimation, The MIT press, 1974.

Jean-Baptiste, N., Malaterre, P.-O., Dorée, C. and Sau, J.: Data assimilation for real-time estimation of hydraulic states and unmeasured perturbations in a 1D hydrodynamic model, Journal of Mathematics and Computers in Simulation, Vol. 81, Issue 10, 2201-2214, 2011.

Giustarini, L., Matgen, P., Hostache, R., Montanari, M., Plaza, D., Pauwels,V. R. N., De Lannoy, G. J. M., De Keyser, R., Pfister, L., Hoffmann, L., and Savenije, H. H.G.: Assimilation SAR-derived water level data into a hydraulic model: a case study, Hydrol. Earth Syst. Sci., 8, 2103–2144, 2011.

Hendricks, F.H.J., and Kinzelbach, W.: Real-time groundwater flow modeling with the Ensemble Kalman Filter: Joint estimation of states and parameters and the filter inbreeding problem, Water Resour. Res., 44, W09408, doi:10.1029/2007WR006505, 2008.

Madsen, H., and Skotner, C.: Adaptive state updating in real-time river flow forecasting - a combined filtering and error forercasting procedure, J. Hydrol., 308, 302–312, 2005.

Li, J., and Xiu, D.: A generalized polynomial chaos based ensemble Kalman filter with high accuracy, J. Comput. Phys., 228, 5454–5469, 2009.

Malaterre, P.-O., Baume, J.-P., Jean-Baptiste, N., and Sau, J.: Calibration of open channel flow models: a system analysis and control engineering approach, SimHydro 2010: Hydraulic modeling and uncertainty, 2-4 June 2010, Sophia Antipolis.

Matgen, P., Montanari, M., Hostache, R., Pfister, L., Hoffmann, L., Guingla, D.P., Pauwels, V., De Lannoy, G., De Keyser, R., and Savenije, H.H.G.: Towards the sequential assimilation of SAR-derived water stages into hydraulic models using the Particle Filter: proof of concept, Hydrol. Earth Syst. Sci, 14, 1773–1785, 2010.

Mirouze, I., and Weaver, A.T.: Representation of correlation functions in variational assimilation using an implicit diffusion operator, Q. J. R. Meteorol. Soc., 136, 1421–1443, DOI:10.1002/qj.643, 2010.

Moradkhani, H., Hsu, K.-L. Gupta, H.V., and Sorooshian, S.: Improved streamflow forecasting using self-organizing radial basis function artificial neural networks, J. Hydro., 295, 246–262, 2004.

Moradkhani, H., Sorooshian, S, Gupta, H.V., and Houser, P.-R.: Dual state-parameter estimation of hydrological models using ensemble Kalman filter, Advances in Water Resources, 28, 135–147, 2005a.

Moradkhani, H., Hsu, K.-L., Gupta, H.V., and Sorooshian, S.: Uncertainty assessment of hydrologic model states and parameters: Sequential data assimilation using the particle filter, Water Resour. Res., 41, W05012, doi:10.1029/2004WR003604, 2005b.

Neal, J. C., Atkinson, P. M., and Hutton, C. W.: Flood inundation model updating using an ensemble Kalman filter and spatially distributed measurements. J. Hydro., Volume 336, Issue 3, 401-415, 2007.

Neal, J., Schumann, G., Bates, P., Buytaert, W., Matgen, P., and Pappenberger, F.: A data assimilation approach to discharge estimation from space, Hydrol. Process., 23, 3641—3649. doi: 10.1002/hyp.7518, 2009.

Pannekoucke O., Berre, L., and Desroziers, G.: Background-error correlation length scale estimates and their sampling statistics, Q. J. R. Meteorol. Soc., 134, 497–508, 2008.

Pannekoucke, O., and Massart, S.: Estimation of the local diffusion tensor and normalization for heterogeneous correlation modelling using a diffusion equation, Q. J. R. Meteorol. Soc., 134, 1425–1438, 2008.

Pappenberger, F., Bevena, K., Horrittb, M., and Blazkovac, S.: Uncertainty in the calibration of effective roughness parameters in HEC-RAS using inundation and downstream level observations, J. Hydrol., 302, 46–49, 2005.

Ricci, S., Piacentini, A., Thual, O., Le Pape, E., and Jonville, G.: Correction of upstream flow and hydraulics state with data assimilation in the context of flood forecasting, Hydrol. Earth Syst. Sci, Vol. 15, 1–21, 2011.

Schumann, G., Bates, P.D., Horritt, M.S., Matgen, P., and Pappenberger, F.: Progress in integration of remote sensing-derived flood extent and stage data and hydraulic models, Rev. Geophys., 47, RG4001, doi:10.1029/2008RG000274, 2009.

Shiiba, M., Laurenson, X., and Tachikawa, Y.: Real-time stage and discharge estimation by a stochastic-dynamic flood routing model. Hydrol. Process., Vol. 14, Issue 3, 481-495, 2000.

Szunyogh, I., Kostelich, E., Gyarmati, G., Kalnay, E., Hunt, B., Ott, E., Satterfield, E., and Yorke, J.: A local ensemble transform Kalman filter data assimilation system for the NCEP global model, Tellus A, 60, 113–130, 2008.

Tippett, M.K., Anderson, J.L., Bishop, C. H., Hamill, T.M., and Whitaker, J. S.: Ensemble Square Root Filters, Mon. Wea. Rev., 131, 1485–1490, 2003.

Thirel, G., Martin, E., Mahfouf, J.-F., Massart, S., Ricci, S., and Habets, F.: A past discharges assimilation system for ensemble streamflow forecasts over France, Hydrol. Earth Syst. Sci., 14, 1623–1637, 2010.

Valstar, J.R., McLaughlin, D.B., te Stroet, C.B.M., and van Geer, F.C.: A representer-based inverse method for groundwater flow and transport applications, Water Resour. Res., 40, W05116, doi:10.1029/2003WR002922, 2004.

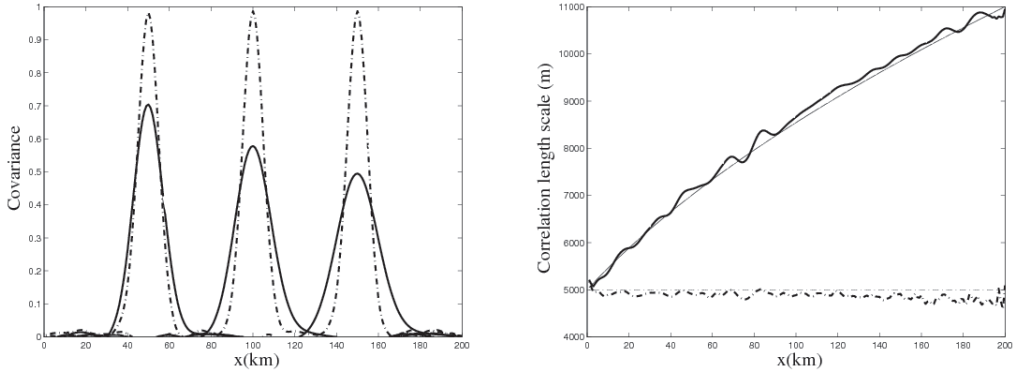
Weaver, A., and Courtier, P.: Correlation modelling on the sphere using a generalized diffusion equation, Q. J. R. Meteorom. Soc.,127, 1815–1846, 2001.

Weaver, A.T., Deltel, C., Machu, E. , Ricci, S., and Daget, N.: A multivariate balance operator for variational ocean data assimilation, Q. J. R. Meteorol. Soc., 131, 3605–3625, 2005.

Weaver, A.T., and Mirouze, I.: On the diffusion equation and its application to isotropic and anisotropic correlation modelling in variational assimilation, Q. J. R. Meteorol. Soc., DOI:10.1002/qj.1955, 2012.

Weerts, A.H., and El Serafy, G.Y.H.: Particle filtering and ensemble Kalman filtering for state updating with hydrological conceptual rainfall-runoff models, Water Resour. Res., 42, W09403, doi:10.1029/2005WR004093., 2006.

Weerts, A.H., El Serafy, G.Y. , Hummel, S., Dhondia, J., and Gerritsen, H.: Application of generic data assimilation tools (DATools) for flood forecasting purposes, Computer & Geosciences, Vol. 36, Issue 4, 453–463, 2010.



(a) (b)

Fig. 1. (a) Covariance function from $\mathbf{B}_e$ for $x = 50$, 100, 150, (b) Correlation length scale over the domain. The case of advection only is represented with dashed lines and the case of advection-diffusion is represented with solid lines. In (b), the results from the theoretical analysis are represented with thin lines and the results from the numerical analysis are represented with thick lines.

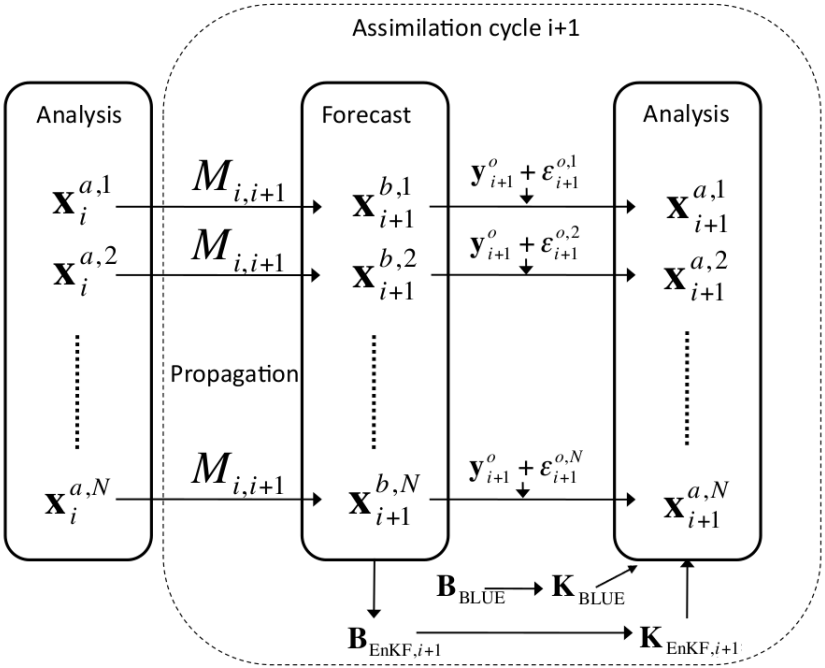


Fig. 2. Ensemble data assimilation algorithms, assimilation cycle $i+1$.

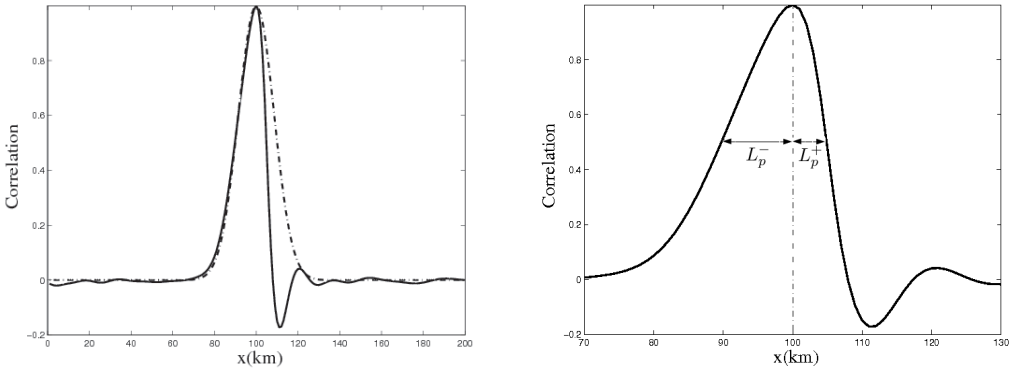
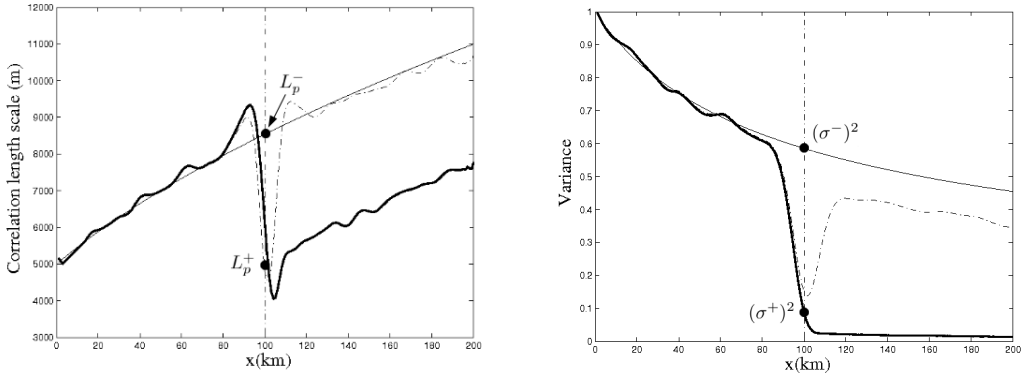


Fig. 3. Background error correlation function at the observation point, for the initial correlation matrix $\mathbf{C}_e$ (dashed line) and for $\mathbf{C}_{\mathrm{EnKF}}$ (solid line) a- on the whole domain, $x \in [0;200]$ and b- on the domain $x \in [70;130]$ with the upstream and downstream correlation length scale, L_p^- and L_p^+ respectively.



(a) (b)

Fig. 4. a- Background error correlation length scale $L_p(x)$ and b- Variance σ^2, in theory without assimilation (thin solid line), for $\mathbf{B}_{\mathrm{EnKF}}$ (thick solid line), for $\mathrm{EnBLUE}_{\mathrm{B_e}}$ (thin dashed line) and for $\mathrm{EnBLUE}_{\mathbf{B}_{\mathrm{EnKF}}}$ (thick dashed line). As $\mathbf{B}_{\mathrm{EnKF}}$ and $\mathrm{EnBLUE}_{\mathbf{B}_{\mathrm{EnKF}}}$ provide the same results for the variance and the correlation length scales the corresponding curves overlap.

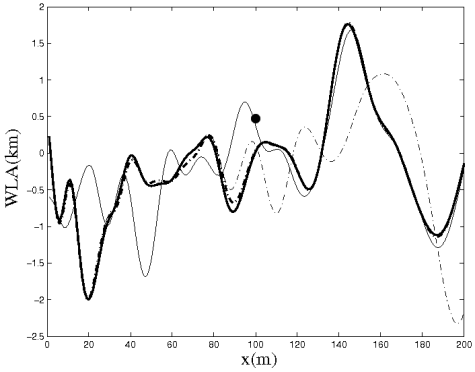


Fig. 5. WLA for the EnKF analysis (thick solid line for h^a_{EnKF}), the single BLUE analysis with $\mathbf{B}_e$ (thin dashed line for $h^a_{\mathrm{BLUE,B_e}}$) and the single EEnKF analysis with $\mathbf{B}_{\mathrm{EnKF}}$ (thick solid line for $h^a_{\mathrm{BLUE,B_{EnKF}}}$). The observation is denoted by a black dot, the true state h^{true} by a thin solid line.

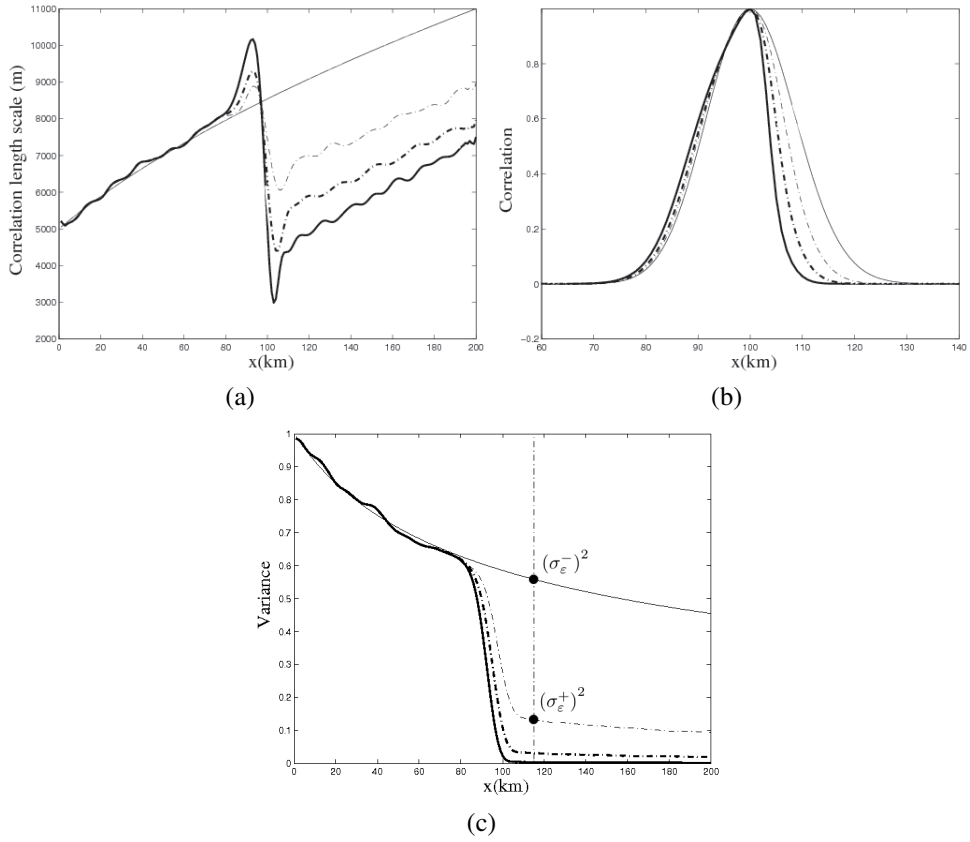
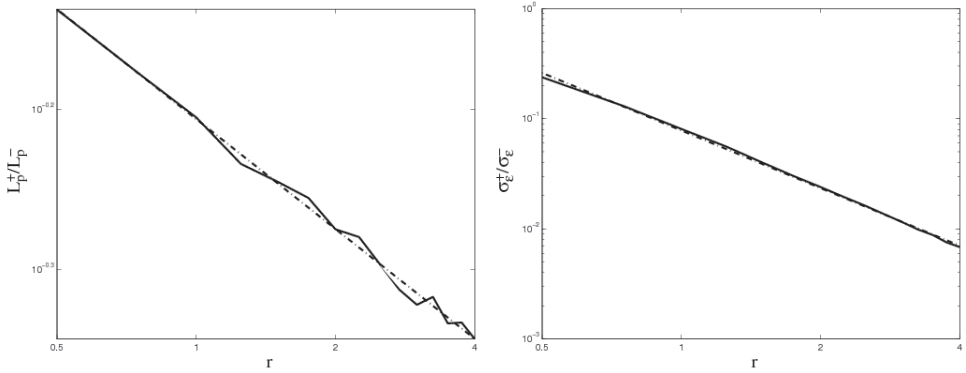
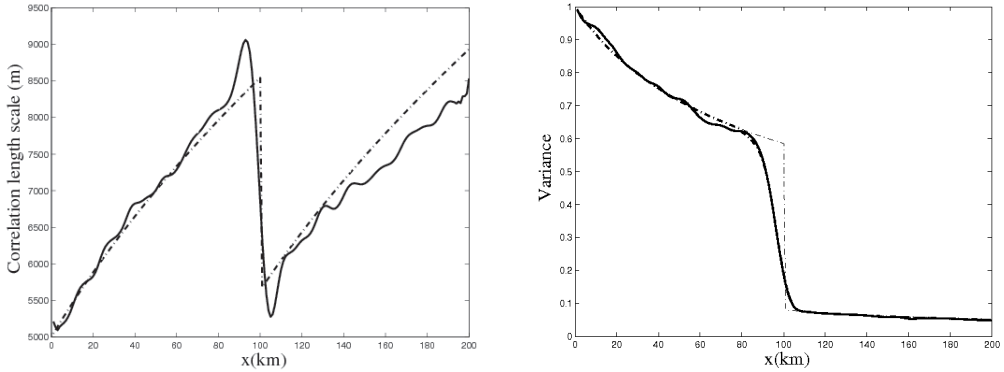


Fig. 6. a- Correlation length scale $L_p(x)$ b- Correlation function at x_{obs} c-Variance σ^2 in theory without assimilation (thin solid line), for $r = 0.5$ (thin dashed line), for $r = 1.75$ (thick dashed line) and for $r = 4$ (thick solid line).



(a) (b)

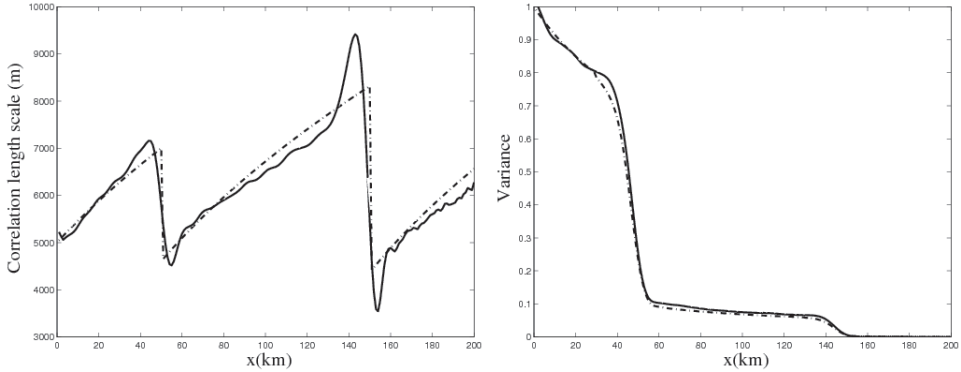
Fig. 7. Abacus for (a) $\frac{L_p^+}{L_p^-}$ and (b) $\frac{\sigma_\varepsilon^+}{\sigma_\varepsilon^-}$ as function of the ratio r in solid lines and linear regression in logarithmic scales in dashed lines.



(a) (b)

Fig. 8. (a) Background error correlation length scale and (b) variance computed with the EnKF (thick solid lines) and with the parametrized model for $r = 0.75$, using 25 points in $\mathbf{I}_\varepsilon$ (thick dashed line) or 2 points (thin dashed line) for the description of the variance reduction

.



(a) (b)

Fig. 9. a- Background error correlation length scale $L_p(x)$ and b- variance $\sigma^2(x)$ for the EnKF (solid line) and the parametrized model (dashed line).

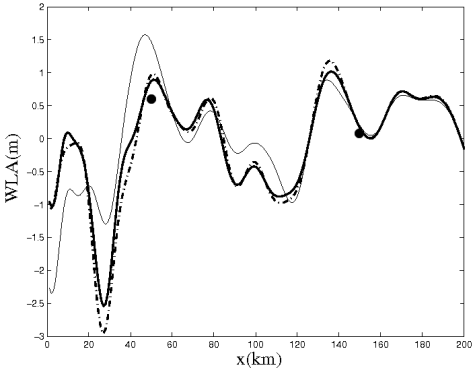


Fig. 10. Analyzed WLA for the EnKF (thick solid line) and the EEnKF (thick dashed line). The true state is represented by the thin solid line and the observations by the black dots.