



**HAL**  
open science

# High-resolution data assimilation through stochastic subgrid tensor and parameter estimation from 4DEnVar

Yin Yang, Etienne Mémin

► **To cite this version:**

Yin Yang, Etienne Mémin. High-resolution data assimilation through stochastic subgrid tensor and parameter estimation from 4DEnVar. Tellus A, 2017, 19 p. hal-01500140

**HAL Id: hal-01500140**

**<https://inria.hal.science/hal-01500140>**

Submitted on 2 Apr 2017

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

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

# High-resolution data assimilation through stochastic subgrid tensor and parameter estimation from 4D<sub>En</sub>Var

By Yin Yang<sup>1\*</sup> and Etienne Mémin<sup>2</sup>

<sup>1</sup>ENSEEIH, 2 Rue Charles Camichel, 31000 Toulouse, France,

<sup>2</sup>INRIA, Campus de Beaulieu, 35042 Rennes cedex, France

(Manuscript received xx xxxx xx; in final form xx xxxx xx)

## ABSTRACT

In this paper we explore a dynamical formulation allowing to assimilate high resolution data in a large-scale fluid flow model. This large-scale formulation relies on a random modeling of the small-scale velocity component and allows to take into account the scale discrepancy between the dynamics and the observations. It introduces a subgrid stress tensor that naturally emerges from a modified Reynolds transport theorem adapted to this stochastic representation of the flow. This principle is used within a stochastic shallow water model coupled with an 4D<sub>En</sub>Var assimilation technique to estimate both the flow initial conditions and the inhomogeneous time-varying subgrid parameters. The performance of this modeling has been assessed numerically with both synthetic and real world data. Our strategy has shown to be very effective in providing a more relevant prior/posterior ensemble in terms of the dispersion compared to other tests using the standard shallow water equations with no subgrid parameterization or with simple eddy viscosity models. We also compared two localization techniques. The results indicate the localized covariance approach is more suitable to deal with the scale discrepancy related errors.

*Keywords: in this paper*

## 1. Introduction

The numerical simulation of all the scales of geophysical fluid flows, up to the smallest dissipation scales, remains nowadays completely intractable due to the dimension of the problem to handle (which scales as the cube of the Reynolds number). For atmospheric flows for instance, the spatial scales range from several millimeters to thousands of kilometers. The temporal scale is also huge and goes from decades to seconds. Many of the physical processes related to the submeso-scale and micro-scale of the geophysical systems are left unresolved. However, due to nonlinear interactions between scales, it is important to understand the actions of those unresolved components to take them into account in a large-scale representation of the dynamics.

For several years there has been a growing interest in modeling the unresolved processes through random terms. In that respect, the simplest method consists in introducing random forcing terms into the flow dynamics equations (Buizza et al., 1999; Leith, 1990; Mason and Thomson, 1992; Shutts, 2005; Slingo

and Palmer, 2011). This kind of methods is often coupled to the usual subgrid-scale (SGS) parameterization used in deterministic large-scale approaches where the statistically averaged flow on a coarse grid are solved directly and the effects of the small unresolved scale on the large resolved scales are modeled thereafter. Those models of the Reynolds stress reflect primarily the dissipation due to turbulence while stochastic subgrid closure terms have been proposed to model also the energy backscattering from small to large scales (Frederiksen et al., 2013).

Another way of considering the stochastic nature of the neglected scales is to study directly the governing differential equations driven by stochastic processes. To that end, a stochastic dynamics modeling is proposed in Mémin (2014). In this approach, which constitutes the fundamental cornerstone to this article, the author proposed to introduce the stochasticity from the very beginning in the dynamics physical derivation.

The basic idea is built on the assumption that the Lagrangian fluid particle displacement results from a smooth velocity component and a fast oscillating stochastic velocity component uncorrelated in time,

$$\mathbf{X}_t = \mathbf{X}_{t_0} + \int_{t_0}^t \mathbf{w}(\mathbf{X}_s, s) ds + \int_{t_0}^t \boldsymbol{\sigma}(\mathbf{X}_s, s) d\mathbf{B}_s, \quad (1)$$

\* Corresponding author.  
E-mail: iamyangyin@gmail.com

with  $\mathbf{X}_t$  denoting the flow map at time  $t$ ,  $\mathbf{w}$  the smooth fluid particles velocity corresponding to the sought “large-scale” solution, and  $d\mathbf{B}_t$  a  $d$ -dimensional Wiener increment ( $d=2,3$ ). The term  $\boldsymbol{\sigma}(x,t)d\mathbf{B}_t$  represents the fast-changing fluctuation associated to turbulence; it has a much lower time-scale than  $\mathbf{w}$ , the smooth component. Both  $\boldsymbol{\sigma}$  and  $\mathbf{w}$  require to be solved or modeled. Stochastic calculus principles provides then the rules to derive the expressions describing the evolution of functions of such flows.

Large-scale models are usually assessed and calibrated by comparing simulated results and data. Opposite to the sparse heterogeneous network of *in situ* measurements, satellite image data exhibit a spatial resolution that is finer than the resolved scales at which the state variables are solved through a filtered dynamics. This still increasing scale discrepancy due to the rapid technological progress of satellite sensors advocates the setup of methods enabling to calibrate the numerical flow model from high-resolution data. The data assimilation (DA) approach can indeed serves this purpose. Note that image data assimilation is a research topic drawing more and more attention since the last decades and several strategies have been proposed to deal with these difficulties, (Beyou et al., 2013; Bereziat and Herlin, 2015; Chabot et al., 2015; Corpetti et al., 2009; Cuzol and Mémin, 2009; Huot et al., 2013; Papadakis and Mémin, 2008; Souopgui, 2010; Titaud et al., 2010).

Traditionally, data assimilation techniques are widely used to recover as exactly as possible the initial condition given a dynamics model, a certain span of time called assimilation window and some observations across this window (Talagrand, 1997; Blum et al., 2009). The initial condition is crucial for Numerical Weather Prediction (NWP) systems but the inference of the models’ parameters is also tractable through the data assimilation process. In a broader sense, the assimilation technique discussed in this article can be viewed as a learning process of a dynamic model on the basis of the observations available by adjusting the model’s parameters.

Data assimilation techniques are classically categorized into two groups. The first group, referred to as variational assimilation techniques, is formulated from the optimal control theory and variational calculus (Lions, 1971; Le Dimet and Talagrand, 1986). The second group of “ensemble methods” originates from Monte Carlo Bayesian filtering principles (Evensen, 1994; Gordon et al., 1993). The parameter estimation problem can also be divided in the same spirit into those two groups.

In the former approach, the estimation consists in considering the unknown parameters as a control parameter as proposed by Navon (1998). Consequently, the objective of the variational system is to find the optimal initial condition and the associated optimal parameters for which a system trajectory fits best the available observations. Several works in this vein can be found in (Cacuci et al., 2013; Navon, 1998; Zhu and Navon, 1999; Pulido and Thurn, 2005; Kazantsev, 2012). The latter approach is a direct extension of standard ensemble Kalman filter methods. Within this framework, one can choose freely the

specific ensemble filtering schemes in terms of specific applications (Zupanski and Zupanski, 2006; Tong and Xue, 2008; Yang and Delsole, 2009; Bocquet and Sakov, 2013b; Ruiz et al., 2013; Tandeo et al., 2014; Sawada et al., 2015).

In this article, we propose to use an ensemble-based variational scheme which gathers the advantages of both the variational assimilation schemes and the ensemble-based sequential methods (Yang et al., 2015). In addition to the obvious opportunity of not requiring any adjoint dynamical models, ensemble-based schemes provide other benefits especially when considering stochastic dynamics. As a matter of fact ensemble prediction systems are becoming more and more crucial in meteorology and oceanography. In that respect, under-dispersion of the ensemble greatly jeopardize forecasting quality (Slingo and Palmer, 2011). Such underestimation of the ensemble spread is usually related to the unawareness of the uncertainties associated with the dynamic system. Recent studies show that the quality of the ensemble spread can be significantly improved by not only perturbing the initial condition but also some parameters related to the dynamics (Wei et al., 2013; Yang et al., 2015). Those facts advocate the setup of physically relevant stochastic dynamical models coupling the random production terms and the dissipation term.

Within this paper, we focus on the specification through high-resolution image data assimilation of the dynamics model uncertainties. A fully justified argument on the derivation of the stochastic fluid flow dynamics is not the main purpose here. This framework is briefly reviewed in section 2, then we introduce on this basis the stochastic shallow water system. In section 3, we investigate methods for estimating a central object called the quadratic variation tensor from the data. Model and experimental settings, results and discussions with some concluding remarks are provided in the following sections.

## 2. Stochastic modeling

Analogously to the standard deterministic case, the derivation procedure from the physical conservation laws of the Navier-Stokes equations is based primarily on the Reynolds transport theorem (RTT). In our context the RTT consists of a stochastic expression of the rate of change of a given quantity (mass, momentum and energy) in a material volume. A comprehensive derivation process of this theorem (Mémin, 2014) is presented in appendix A for reader convenience.

Applied to scalar function,  $q$ , the expression of the RTT within a volume transported by a stochastic flow (1) with an incompressible small-scale velocity component ( $\nabla \cdot \boldsymbol{\sigma} = 0$ ), is given by:

$$d \int_{\mathcal{V}(t)} q dx = \int_{\mathcal{V}(t)} [d_t q + \underbrace{\nabla \cdot (q(\mathbf{w} - \frac{1}{2} \nabla \cdot \mathbf{a}))}_{\bar{w}}] dt - \frac{1}{2} \sum_{ij} \frac{\partial}{\partial x_i} (a_{ij} \frac{\partial}{\partial x_j} q) dt + \nabla q \cdot \boldsymbol{\sigma} d\mathbf{B}_t dx. \quad (2)$$

This modified RTT involves the time increment of function  $q$  (the differential of  $q$  at a fixed point) instead of the time derivative. A diffusion operator emerges also naturally. For clarity's sake, this term is designated as "subgrid stress tensor" following the protocols of large eddies simulation (LES). However, its construction is quite different. It is not based on Boussinesq's eddy viscosity assumption nor on any structural turbulence models (Sagaut, 2006) but arises from stochastic calculus rules. It expresses the mixing process exerted by the fast oscillating velocity component. This diffusion term is directly related to the small-scale component through the variance tensor,  $\mathbf{a}$ , defined from the diagonal of the uncertainty covariance:

$$\mathbf{a}(\mathbf{x})\delta(t-t')dt = \mathbb{E} \left( (\boldsymbol{\sigma}(\mathbf{x}, t)d\mathbf{B}_t) (\boldsymbol{\sigma}(\mathbf{x}, t')d\mathbf{B}_{t'})^T \right).$$

The variance tensor corresponds to an eddy viscosity term (with units in  $m^2 s^{-1}$ ).

A corrective advection term,  $\tilde{\mathbf{w}} = \mathbf{w} - 1/2 \nabla \cdot \mathbf{a}$ , appears also in the stochastic RTT formulation. This correction expresses the influence of the small scales inhomogeneity on the large ones. A drift is engendered from the regions associated with maximal variance (maximal turbulent kinetic energy - TKE) toward the area of minimum variance (e.g. minimal TKE). For homogeneous noise (i.e. within an homogeneous turbulence assumption), the variance tensor is constant and this corrective advection does not come into play.

Through this modified RTT, stochastic versions of the mass and momentum conservation equations can be (almost) directly derived. Incompressibility conditions can for instance be immediately deduced from the RTT applied to  $q = 1$  and the flow jacobian ( $J$ ):

$$d \int_{\mathcal{V}(t)} d\mathbf{x} = \int_{\mathcal{V}(t)} \nabla \cdot \tilde{\mathbf{w}}(\mathbf{x}, t) dt d\mathbf{x} = \int_{\mathcal{V}(t_0)} d(J(\mathbf{X}_t(\mathbf{x}), t)) d\mathbf{x} = \int_{\mathcal{V}(t_0)} (J \nabla \cdot \tilde{\mathbf{w}})(\mathbf{X}_t(\mathbf{x}), t) dt d\mathbf{x}. \quad (3)$$

Together with the incompressibility of the random term, the incompressibility condition reads thus:

$$\nabla \cdot \boldsymbol{\sigma} = 0 \text{ and } \nabla \cdot \tilde{\mathbf{w}} = 0. \quad (4)$$

Imposing, *a priori*, an incompressibility condition to the large-scale component,  $\mathbf{w}$ , simplifies the previous constraint as:

$$\nabla \cdot \boldsymbol{\sigma} = 0 \text{ and } \nabla \cdot \mathbf{w} = \nabla \cdot (\nabla \cdot \mathbf{a})^T = 0. \quad (5)$$

Note that for an homogeneous unresolved velocity, the last condition is naturally satisfied, since this unresolved velocity component is associated with a constant variance tensor,  $\mathbf{a}$ . It should be stressed that condition (4) corresponds to a weaker form of the incompressibility constraint in the sense that it does not impose an incompressibility condition to the large-scale component but only to the effective drift. It has the advantage not to include a cumbersome condition on the variance tensor.

In the following section we will briefly present how this principle can be applied to derive a stochastic expression of the shallow water equations. Those expressions are derived from the ap-

plication of the stochastic RTT to the linear momentum conservation principle. Interested readers may consult (Mémmin, 2014) for a thorough development or (Resseguier et al., 2016a;b;c) for the derivation and the analysis of general geophysical flow models and their quasigeostrophic approximations.

### 2.1. The stochastic 2D nonlinear shallow water model

Considering the usual assumptions underlying the constitution of the Shallow water model: the depth is shallow compared to the horizontal dimension, the friction forces and the vertical shear on the velocity horizontal components can be neglected, the external force is only due to gravity, we get the following 2D momentum equations for the horizontal velocity components (where superscript "h" stands for the horizontal coordinates)

$$\frac{\partial \mathbf{u}^h}{\partial t} + \nabla^h \cdot (\mathbf{u} \mathbf{u}^T) - \frac{1}{2} \sum_{(ij)^h} \partial_{x_i} \partial_{x_j} (a_{ij}^h \mathbf{u}) = -\frac{1}{\rho} \nabla^h p, \quad (6)$$

$$\tilde{\mathbf{u}}^h = \mathbf{u}^h - \frac{1}{2} \nabla \cdot \mathbf{a}^h \quad (7)$$

$$\nabla^h d\hat{p} = -\rho (\mathbf{w}^h \nabla^h) (\boldsymbol{\sigma} d\mathbf{B}_t)^h. \quad (8)$$

The last equation arises from a balance of the advected Brownian term and the random pressure,  $d\hat{p}$ , associated to the random small-scale component. This equation stems from a separation principle between the smooth terms and the Brownian terms. It can be noted that in those equations the noise components are assumed to depend only on the horizontal components. Thus the diffusion tensor as well as the variance tensor do not depend on depth. These equations are complemented by the three incompressibility constraints (5). The noise component (and its related variance tensor) as well as the horizontal velocity component can be interpreted as depth-averaged components of the 3D velocity and uncertainty components respectively. Integrating the incompressibility constraints ( $\nabla \cdot \tilde{\mathbf{w}} = 0$ ) and ( $\nabla \cdot \boldsymbol{\sigma} = 0$ ) along the depth from the bottom  $h_b$  to the surface  $h_s$  we get from the effective drift expression (7):

$$\nabla^h \cdot \tilde{\mathbf{u}}(h_s - h_b) = w|_{h_s} - w|_{h_b},$$

and

$$(\boldsymbol{\sigma} d\mathbf{B}_t)^z|_{h_s} = (\boldsymbol{\sigma} d\mathbf{B}_t)^z|_{h_u}.$$

We may noticed that for ( $\nabla \cdot \nabla \cdot \mathbf{a} = 0$ ) we recover the traditional balance between the divergence of the horizontal averaged velocity components and the vertical motions at the interfaces. In the same way as in the traditional setting, considering the difference of the kinematic boundary conditions at the surface and the still bottom, and introducing the previous relation for the depth variable,  $h = h_s - h_b$ , the shallow-water continuity equation reads:

$$d_t h + \left( \nabla^h \cdot (h \tilde{\mathbf{u}}^h) - \frac{1}{2} \sum_{(i,j)^h} \partial_{x_i} (a_{ij} \partial_{x_j} h) \right) dt + \nabla^h \cdot (h \boldsymbol{\sigma} d\mathbf{B}_t) = 0. \quad (9)$$

It should be remarked that this continuity equation involves the horizontal components of the (divergence free) effective advection term:  $\tilde{\mathbf{u}}^h = \mathbf{u}^h - \frac{1}{2}\nabla^h \cdot \mathbf{a}^h$ . The term  $(h\nabla^h \cdot \nabla^h \cdot \mathbf{a}^h)$  plays the role of an additional forcing condition that ensues from the small-scale inhomogeneity. For smooth enough fluctuation variances this forcing is negligible.

Introducing the hydrostatic balance (which is well respected in the shallow water model due to a negligible vertical acceleration) the pressure (after depth integration and with a boundary condition provided by the atmospheric pressure at the water-air interface) and its gradient reads respectively

$$p|_z = \rho g(h_s - z) + p|_{atm}, \text{ and } \nabla p|_z = \rho g \nabla h_s \quad (10)$$

We get hence the following horizontal linear momentum conservation equation:

$$\frac{\partial \mathbf{u}}{\partial t} + (\tilde{\mathbf{u}}^h \cdot \nabla^h) \mathbf{u}^h - \frac{1}{2} \sum_{(ij)^h} \partial_{x_i} (a_{ij} \partial_{x_j} \mathbf{u}^h) = -g \nabla^h h_s. \quad (11)$$

It is important to outline that the free surface acts here as a random forcing term on the momentum equations. It is driven by the stochastic evolution of the height variable (9). For a deterministic height evolution such as the one associated to the mean depth evolution conditionally to a deterministic velocity field, the momentum equation would provide the velocity fields attached precisely to this particular (mean) instance of the evolving free-surface.

The uncertainty formalism allows us to obtain the 2D shallow water equation in the form of (9, 11). It can be observed that the dissipative terms emerged both in the continuity equation and in the momentum conservation equation. It is similar in spirit to the eddy viscosity assumption used in LES and Reynold average numerical simulation (RANS). Nevertheless, our subgrid-scale model, takes its origin in the stochastic representation of the uncertainty associated to the model itself. No extra assumption is needed. However, their similarities indeed bring us another perspective for the interpretation of the subgrid-scale effect. For instance, it can be shown (Mémin, 2014) that setting the variance tensor as:

$$\mathbf{a}^h = 2\nu_h \mathbb{I}_h, \quad (12)$$

with the eddy viscosity coefficient

$$\nu_h = C \Delta_x \Delta_y \sqrt{(\partial_x u)^2 + (\partial_y v)^2 + \frac{1}{2}(\partial_y u + \partial_x v)^2}, \quad (13)$$

corresponds, for smooth enough rate of strain tensor norm, to the famous Smagorinsky (1963) subgrid model.

## 2.2. Conservative form

The conservative form of shallow water equations is usually more useful when implementing the finite-volume method. It can be inferred directly by combining equation (11) and equation (9). Alternatively, one can deduce such a form by integrating equation (9) along the depth from the bottom  $h_b$  to the surface  $h_s$ . Finally, we have, (note we dropped in the following the

subscript  $h$  for sakes of clarity; all the vectorial variables and spatial derivative operators represent planar vectors and derivations w.r.t. horizontal components respectively)

$$\begin{aligned} d_t(h\mathbf{u}) + \nabla \cdot \left( (h\mathbf{u}\tilde{\mathbf{u}}^T) + \frac{1}{2}g\eta^2\mathbb{I} \right) dt \\ - \sum_{(i,j)} \left( \frac{1}{2}\partial_{x_i} (a_{ij}\partial_{x_j}(h\mathbf{u})) - a_{ij}\partial_{x_j} h \partial_{x_i} \mathbf{u} \right) dt \\ + \mathbf{u} \nabla \cdot (h\sigma d\mathbf{B}_t) = 0, \quad (14) \end{aligned}$$

As previously indicated, if the focus is only on the ensemble average depth evolution (denoted by  $\bar{h}$ ) or if the noise is assumed to live on iso-height surfaces (i.e  $\sigma d\mathbf{B}_t \cdot \nabla h = 0$ ), the above equation greatly simplifies as it boils down to a deterministic system:

$$\frac{\partial \bar{h}}{\partial t} + \nabla \cdot (\bar{h}\tilde{\mathbf{u}}) - \frac{1}{2} \sum_{(i,j)} \partial_{x_i} (a_{ij}\partial_{x_j} \bar{h}) = 0, \quad (15a)$$

$$\begin{aligned} \frac{\partial (\bar{h}\mathbf{u})}{\partial t} + \nabla \cdot \left( (\bar{h}\mathbf{u}\tilde{\mathbf{u}}^T) + \frac{1}{2}g\eta^2\mathbb{I} \right) \\ - \sum_{(i,j)} \left( \frac{1}{2}\partial_{x_i} (a_{ij}\partial_{x_j}(h\mathbf{u})) - a_{ij}\partial_{x_j} h \partial_{x_i} \mathbf{u} \right) = 0. \quad (15b) \end{aligned}$$

This system, as opposed to the usual deterministic Shallow water model, still includes additional subgrid terms that depend on the variance of the small-scale random component.

## 2.3. 1D stochastic shallow water equation

The tests shown in the next sections are carried out with a 1D stochastic shallow water equation since it is less time-consuming but remains representative of the considered dynamics. The 1D stochastic shallow water system simply reads,

$$\begin{aligned} d_t h + \left( \partial_x \left( h \left( u - \underbrace{\frac{1}{2}\partial_x a_{xx}}_{\tilde{u}} \right) - \frac{1}{2}\partial_x (a_{xx}\partial_x h) \right) dt \right. \\ \left. + \partial_x h (\sigma d\mathbf{B}_t)_x = 0, \quad (16a) \end{aligned}$$

$$\begin{aligned} d_t(hu) + \left( \partial_x(hu\tilde{u} + \frac{1}{2}g\eta^2) - \frac{1}{2}\partial_x(a_{xx}\partial_x(hu)) \right. \\ \left. - a_{xx}\partial_x h \partial_x u \right) dt + u \partial_x h (\sigma d\mathbf{B}_t)_x = 0, \quad (16b) \end{aligned}$$

and the mean depth and its associated horizontal fields satisfy,

$$\partial_t \bar{h} + \partial_x (\bar{h}\tilde{u}) - \frac{1}{2}\frac{\partial^2}{\partial x^2} (a_{xx}\bar{h}) = 0, \quad (17a)$$

$$\begin{aligned} \partial_t (\bar{h}u) + \partial_x (\bar{h}u\tilde{u} + \frac{1}{2}g\eta^2) \\ - \left( \frac{1}{2}\partial_x (a_{xx}\partial_x (\bar{h}u)) - a_{xx}\partial_x \bar{h} \partial_x u \right) = 0. \quad (17b) \end{aligned}$$

As stated before, one of the main motivations of our approach is to propose methods for the assimilation of fine resolution image data. An interesting idea consists thus in estimating the diagonal quadratic variation tensor,  $\mathbf{a}$ , or to model the (plain) diffusion tensor,  $\sigma$ , from the statistical variations of small scale

velocity measurements extracted from image data (Heitz et al., 2010).

In the following experiments, we will restrain our interests to the horizontal mean flow fields described in Eqs. (17) and (15). However the use of the full stochastic model with the noise forcing terms is also possible alternative (Miller et al., 1999) that will be explored in a future study for data assimilation methods expressed in a stochastic filtering context.

### 3. Estimation of the quadratic variation tensor from data

#### 3.1. Estimation by explicit formulation

Several immediate and simple strategies can be used to define the variance tensor. As detailed previously, an explicit form of this tensor can be given so as to recover classical subgrid models. However the specific statistical nature of this tensor allows us to specify a variety of empirical forms for this tensor (Kadri Harouna and Mémin, 2016). For instance, it can be defined from local velocity covariances matrices estimated on local spatial or temporal neighborhood. The previous estimation methods are easy to implement, but constitute only rough approximations as they rely either on the resolved components or on the flow velocity measurements, which are in practice large-scale observations (Heitz et al., 2010). A more accurate way of estimating the parameters is through a data assimilation process.

#### 3.2. Estimation through data assimilation process

Data assimilation techniques consist in estimating the evolution of a given state variables, denoted  $\mathbf{x} \in \mathbb{R}^n$ , by associating measurements,  $\mathcal{Y} \in \mathbb{R}^m$ , of this system with a dynamics model. The evolution model will be denoted in a general way as

$$\partial_t \mathbf{x}(t, x) + \mathbb{M}(\mathbf{x}(t, x), \mathbf{a}(t, x)) = 0, \quad (18)$$

$$\mathbf{x}(t_0, x) = \mathbf{x}_f^0(x) + \boldsymbol{\omega}(x). \quad (19)$$

It involves a non autonomous differential operator  $\mathbb{M}$  – usually nonlinear – that depends in our case on the variance tensor,  $\mathbf{a}$ . The second equation defines the initial state,  $\mathbf{x}(t_0, x)$ , which is assumed to be known up to a zero Gaussian noise,  $\boldsymbol{\omega}$ , of covariance  $\mathbf{B}$ . The integration of the initial forecast guess,  $\mathbf{x}_f^0$ , is called the forecast trajectory,  $\mathbf{x}_f(t, x)$ . It is usually fixed from a previous assimilation result on an anterior temporal window. Note that depending on the context,  $\mathbf{x}_f$  is also referred as the background. The observations,  $\mathcal{Y}$ , and the state variable  $\mathbf{x}$  are linked together through an observation operator  $\mathbb{H}$

$$\mathcal{Y}(t, x) = \mathbb{H}(\mathbf{x}(t, x)) + \boldsymbol{\epsilon}(t, x). \quad (20)$$

The error,  $\boldsymbol{\epsilon}$ , between the measurements and the state is assumed to be a zero mean Gaussian random field with a covariance tensor,  $\mathbf{R}$ . This noise is assumed to be uncorrelated from the state variable.

Optimal parameter can be found through data assimilation

techniques along with the system state. The strategies are formulated as the minimization with respect to control variables (here the initial condition and the variance tensor) of an energy functional encoding a distance between the data and a state variable trajectory with low values of the parameters norm. In our approach, we express the cost function in an incremental form (Courtier et al., 1994), with a Mahalanobis norm  $\|f\|_A^2 = (f, A^{-1}f)$  where  $A^{-1}$  is an inverse covariance matrix and  $(f, g)$  is the  $L^2$  inner product :

$$J(\delta \mathbf{x}_0, \delta \mathbf{a}) = \frac{1}{2} \|\delta \mathbf{x}_0(x)\|_{\mathbf{B}_x}^2 + \frac{1}{2} \|\delta \mathbf{a}\|_{\mathbf{B}_a}^2 + \frac{1}{2} \int_{t_0}^{t_f} \|\partial_x \mathbb{H} \delta \mathbf{x}(t, x, \mathbf{a}) - \mathbf{D}(t, x, \mathbf{a})\|_{\mathbf{R}}^2 dt. \quad (21)$$

The initial analysis filed (denoted by subscript  $a$ ) are related to the increments to be sought through

$$\mathbf{x}_a^0 = \mathbf{x}_f^0 + \delta \mathbf{x}_0, \quad \mathbf{a}_a = \mathbf{a}_f + \delta \mathbf{a}. \quad (22)$$

The innovation vector  $\mathbf{D}(t, x, \mathbf{a})$  is defined as:

$$\mathbf{D}(t, x, \mathbf{a}) = \mathcal{Y}(t, x) - \mathbb{H}(\varphi_t(\mathbf{x}_f, \mathbf{a}_f)), \quad (23)$$

where  $\varphi_t(\mathbf{x}_0, \mathbf{a}_0)$  is the flow map (viewed here as a function of the initial condition and of the initial parameters). In addition the increment  $\delta \mathbf{x}(t, x, \mathbf{a})$  is driven by the tangent-linear model linearized around the current trajectory  $\mathbf{x}_f$ ,

$$\delta \mathbf{x}(t, x, \mathbf{a}) = \partial_x \varphi_t(\mathbf{x}_f, \mathbf{a}_f) \delta \mathbf{x}_0.$$

We are interested in applying a particular DA method, namely a group of ensemble variational scheme (4DEnVar), to our state/parameter estimation problem. 4DEnVar (Liu et al. (2008); Buehner et al. (2010)) features a 4D ensemble-based covariance fields in a variational system. The strategy described here is a direct extension of the scheme discussed in Yang et al. (2015). The extension of 4DEnVar for parameter estimation is done by the so-called augmented state vector technique. If we assume that the parameter variables are random variables, then the extension of state-space based methods to parameter-space is straightforward by encapsulating the parameters. The state augmentation technique has shown to be effective in diverse applications (Tong and Xue (2008); Kang et al. (2011); Simon et al. (2015); Sawada et al. (2015)). The parameter estimation is also related to the estimation of model error from ensemble-based approach. The model error can take simple forms of model bias (Dee, 2005; Baek et al., 2006) or functional forms through parameterization.

#### 3.2.1. Augmented 4DEnVar algorithm

The constitution of a corresponding joint estimation algorithm is illustrated below within an augmented state context. First we need to define  $\mathbf{s}_0$  as the initial augmented state vector  $[\mathbf{x}_0, \mathbf{a}_0] \in \mathbb{R}^{(n+p)}$  where  $p$  is the degree of freedom of  $\mathbf{a}$ . In an incremental framework, we are looking for  $\delta \mathbf{s}_0$  defined

by  $\mathbf{s}_0 - \mathbf{s}_f$ . The cost function in terms of the augmented state increment vector  $\delta \mathbf{s}_0$  reads:

$$J(\delta \mathbf{s}_0) = \frac{1}{2} \|\delta \mathbf{s}_0\|_{\mathbf{B}_s^f}^2 + \frac{1}{2} \int_{t_0}^{t_f} \|\widetilde{\partial_{\mathbf{x}} \mathbb{H}} \widetilde{\partial_{\mathbf{x}} \varphi}_t(\mathbf{s}_f) \delta \mathbf{s}_0 - \mathbf{D}(t, x, \mathbf{a})\|_{\mathbf{R}}^2 dt, \quad (24)$$

where the symbol  $\widetilde{\bullet}$  denotes the augmented version of the tangent linear model or the tangent observation model respectively. Firstly, the extended tangent linear model operator  $\widetilde{\partial_{\mathbf{x}} \varphi}_t$  is composed of the parameter evolution model as well as the tangent linear dynamic model operator  $\partial_{\mathbf{x}} \varphi_t$ . Secondly, in a fully and directly observable case, the observation operator corresponds to the identity matrix. In our case, the validity of  $\mathbf{a}$  hinges on the quality of the analysis trajectory driven by the model, therefore, rather than to simply fit to the noisy empirically observed  $\mathbf{a}$  values with few samples, we will assume that  $\mathbf{a}$  is non-observable. Let us note that a prior information on the variance tensor  $\mathbf{a}_0$  is included in the background regulation term. The observation operator or its linear tangent expression takes the form of  $(\mathbb{I}_{m \times n}, \mathbf{0}_{m \times p})^T$ .

The augmented state ensemble, denoted by  $\mathbf{S}_f \in \mathbb{R}^{(n+p) \times N}$ , is obtained by perturbing the background  $\mathbf{s}_f$ ;  $N$  is the number of ensemble members. We defined the ensemble anomaly matrix as,

$$\mathbf{A}'_s := \frac{1}{\sqrt{N-1}} (\mathbf{S}_f^1 - \overline{\mathbf{S}}_f, \dots, \mathbf{S}_f^N - \overline{\mathbf{S}}_f), \quad \mathbf{A}'_s \in \mathbb{R}^{(n+p) \times N} \quad (25)$$

where the operator  $\overline{\bullet}$  takes the ensemble mean and the  $j$ th perturbed ensemble member is noted  $\mathbf{S}_f^j = [\mathbf{x}^j, \mathbf{a}^j]$ . The background covariance  $\mathbf{B}_s^f$  is approximated by the ensemble anomaly matrix:

$$\mathbf{B}_s^f = \mathbf{A}'_s \mathbf{A}'_s{}^T = \begin{pmatrix} \mathbf{B}_{xx} & \mathbf{B}_{ax}^T \\ \mathbf{B}_{ax} & \mathbf{B}_{aa} \end{pmatrix}, \quad (26)$$

where the sub-matrix  $\mathbf{B}_{xx}$  and  $\mathbf{B}_{aa}$  correspond to the auto-covariance matrix and  $\mathbf{B}_{ax}$  contains the cross covariance between the state and the parameter. As suggested by several studies, a localization procedure enables to robustify the construction of ensemble-based covariance to sampling errors. Due to the parameters' spatial structure, it is necessary to introduce similar localization to the parameter-related covariances. In this study we have tested two different localization techniques referred to as ‘localized covariance’ and ‘local analysis’ respectively. The former relies on a Schur product between the global covariance matrix and an isotropic correlation function,

$$\mathbf{P}_s^f = \mathcal{C}_s \odot \mathbf{A}'_s \mathbf{A}'_s{}^T = \begin{pmatrix} \mathcal{C}_{xx} & \mathcal{C}_{ax}^T \\ \mathcal{C}_{ax} & \mathcal{C}_{aa} \end{pmatrix} \odot \begin{pmatrix} \mathbf{B}_{xx} & \mathbf{B}_{ax}^T \\ \mathbf{B}_{ax} & \mathbf{B}_{aa} \end{pmatrix}, \quad (27)$$

to filter out the long range spatial correlations. The spatial correlation matrix  $\mathcal{C}_{\bullet\bullet}$  is defined from a polynomial approximation of a Gaussian function with compact support and a hard cut-off (Gaspari and Cohn, 1999). Same values of the cutoff distance

are adopted for all  $\mathcal{C}_{\bullet\bullet}$ . The latter consist in bounding – or localizing – directly the region of influence of a given observation on the grid to a restricted area.

The 4DEnVar consists in applying a change of variable with the square-root of the error covariance matrix, by setting  $\delta \mathbf{s}_0 = \mathbf{U} \delta \mathbf{z}_0$ , where  $\delta \mathbf{z}_0$  is the control weight vector and  $\mathbf{U}$  is the square root of  $\mathbf{P}_s^f$  calculated through a spectral decomposition (Buehner, 2005) or simply  $\mathbf{A}'_s$  in the case of no localization. We can define the following scheme:  $\mathbf{U} = \mathbf{A}'_s \mathbf{D}$ . The control variable transformation is also relevant as a preconditioning technique. The cost function in terms of control variable reads,

$$J(\delta \mathbf{z}_0) = \frac{1}{2} \|\delta \mathbf{z}_0\|^2 + \frac{1}{2} \int_{t_0}^{t_f} \|\widetilde{\partial_{\mathbf{x}} \mathbb{H}} \widetilde{\mathbf{B}}_t^{\frac{1}{2}} \delta \mathbf{z}_0 - \mathbf{D}(t, x, \mathbf{a})\|_{\mathbf{R}}^2 dt.$$

where  $\widetilde{\mathbf{B}}_t^{\frac{1}{2}} = \widetilde{\partial_{\mathbf{x}} \varphi}_t(\mathbf{s}_f) \mathbf{U}$ . The  $\widetilde{\mathbf{B}}_t^{\frac{1}{2}}$  contains a set of samples based on an expression of the ensemble integrated in time. Note that for the Shur-product based localization of this empirical matrix, we implemented  $(\widetilde{\partial_{\mathbf{x}} \varphi}_t(\mathbf{s}_f) \mathbf{A}'_s) \mathbf{D}$  rather than  $\widetilde{\partial_{\mathbf{x}} \varphi}_t(\mathbf{s}_f) (\mathbf{A}'_s \mathbf{D})$  (which comes to consider a time invariant localization).

In order to handle high nonlinearity either related to the model or the observations, many authors advocated the use of an iterative approach (Gu and Oliver, 2007; Kalnay and Yang, 2010; Sakov et al., 2012; Bocquet and Sakov, 2013a). The algorithm in Yang et al. (2015) is characterized by an iterative resampling strategy of the prior ensemble and an update of the ensemble sensitivity matrix within the increment minimization steps. It adopts the form of a Gauss-Newton scheme with nested loops that shows the relevance of this approach in particular in the case of a badly known background information and incomplete observations. In our current problem, with the background information bounded by the coarse resolution and the absence of observed parameter fields, it is indeed appropriate to introduce such iterative approach. Here, the iterative update concerns two parts: on the one hand, the innovation vector of the current outer loop is updated by the analysis trajectory based on the analysis state of the previous outer loop driven by the nonlinear model; on the other hand, the ensemble anomaly matrix is updated through the posterior ensemble fields. The ensemble update can either be derived from the posterior ensemble built on perturbed observations or by a direct transformation of the prior perturbations.

**3.2.1.1. Perturbed observation with global analysis** In the first approach, at the  $k$ th outer loop iteration, a parallel realization of minimization with regard to each member of the initial ensemble is conducted,

$$\delta \mathbf{S}_0^{(j),k} = \mathbf{U} \delta \mathbf{Z}_0^{(j),k}, \quad j = 1, \dots, N. \quad (28)$$

$$J(\delta \mathbf{Z}_0^{(j),k}) = \frac{1}{2} \|\delta \mathbf{Z}_0^{(j),k}\|^2 +$$

$$\frac{1}{2} \int_{t_0}^{t_f} \|\widetilde{\partial_{\mathbf{x}} \mathbb{H}} \widetilde{\partial_{\mathbf{x}} \varphi}_t(\mathbf{s}_0^k) \mathbf{U} \delta \mathbf{Z}_0^{(j),k} - \mathbf{D}^{(j),k}(t, x)\|_{\mathbf{R}}^2 dt,$$

where  $\mathbf{D}^{(j)}$  denotes the  $j$ th innovation vector calculated from perturbed observation field and  $\mathbf{Z}_0$  denotes the control weight matrix. Finally, the updated initial ensemble field and its perturbation matrix read:

$$\begin{aligned} \mathbf{S}_0^{(j),k+1} &= \mathbf{S}_0^{(j),k} + \mathbf{U} \delta \widehat{\mathbf{Z}}_0^{(j),k}, \quad j = 1, \dots, N. \\ \mathbf{A}_s^{k+1} &= \frac{1}{\sqrt{N-1}} (\mathbf{S}_0^{(1),k+1} - \overline{\mathbf{S}_0^{k+1}}, \dots, \mathbf{S}_0^{(N),k+1} - \overline{\mathbf{S}_0^{k+1}}). \end{aligned} \quad (29)$$

**3.2.1.2. Ensemble transformation with local analysis** The square root transformation approach corresponds to a mean preserving transformation as used in Ensemble Transform Kalman filter (ETKF). The updated background ensemble anomaly matrix reads,

$$\mathbf{A}_s^{k+1} = \mathbf{A}_s^{k} \left\{ \mathbb{I} + \int_{t_0}^{t_f} \widetilde{\mathbf{B}}_t^{\frac{1}{2}T} (\partial_{\mathbf{x}} \mathbb{H})^* \mathbf{R}^{-1} \partial_{\mathbf{x}} \mathbb{H} \widetilde{\mathbf{B}}_t^{\frac{1}{2}} dt \right\}^{-\frac{1}{2},k} \mathbf{V}. \quad (30)$$

The transformation matrix is the square root of inverse Hessian<sup>1</sup>. The arbitrary orthogonal matrix  $\mathbf{V}$  is used to center the posterior ensemble on the updated initial condition/analysis. In this approach a single minimization process with respect to the background state is sufficient since the ensemble does not need to be updated explicitly. Finally the updated initial ensemble fields are,

$$\mathbf{S}_0^{(j),k+1} = \widehat{\mathbf{S}}_0^{k+1} + \sqrt{N-1} \mathbf{A}_s^{(j),k+1}, \quad j = 1, \dots, N \quad (31)$$

where  $\widehat{\mathbf{S}}_0^{k+1}$  corresponds to the updated augmented vector,

$$\widehat{\mathbf{S}}_0^{k+1} = \widehat{\mathbf{S}}_0^k + \mathbf{A}_s^{k} \delta \widehat{\mathbf{Z}}_0^k. \quad (32)$$

Since the minimization is done in the ensemble space, we use here the local analysis approach. This computationally efficient approach proposed in Ott et al. (2004) and Hunt et al. (2007) consists to lead the analysis step through an EnKF procedure in a local spatial region. The minimization and the error covariance update steps are done around each grid point considering the flow representation and the observations within a region of small size. Within the local domain, observation error is gradually increased when further away from the analysis grid point (Greybush et al., 2011). Note that the ensemble forecast step must be done globally with the full non-linear dynamic model.

**3.2.1.3. Deterministic global analysis** We also tested another strategy capable of generating global deterministic ensemble anomaly matrix without perturbing the observations. This formulation is close in spirit with the DEnKF method proposed in Sakov and Oke (2008). In this approach, the posterior ensemble

<sup>1</sup> Note that the relationship between the Hessian and the covariance matrix holds rigorously in a linear sense; in a nonlinear scenario, this relationship is only an approximation. As the minimization algorithm LBFGS relies on an approximation of the inverse Hessian matrix  $\mathcal{H}^{-1}$ , we can use this byproduct to evaluate equation (30).

is obtained by using half of the Kalman gain therefore preventing the level of posterior error from underestimation in the context of unperturbed observations. Here we use the same format for the analysis ensemble which can be found in Fairbairn et al. (2013):

$$\mathbf{S}_0^{k+1} \leftarrow \overline{\mathbf{S}_0^{k+1}} + \frac{1}{2} (\mathbf{S}_0^k - \overline{\mathbf{S}_0^k}) + \frac{1}{2} (\mathbf{S}_0^{k+1} - \overline{\mathbf{S}_0^{k+1}}). \quad (33)$$

Let us note, that the iterative schemes described above are also used in the ensemble cycling procedure to move from one assimilation window to the next.

### 3.2.2. Parameter evolution and uncertainties

It is essential to realize that the parameters involved in geophysical models can be global or local in time and/or space. The parameters are generally time-dependent and their values may change when a new observation occurs. However, in this work we will make the assumption that the parameters exhibit only slow temporal variation compared to the model state variation. In DA approaches, it is even custom to assume a constant value of parameters during the assimilation window. This consists in forecasting the analysis state, at the end of the  $i$ th assimilation window, to the beginning of the  $i+1$ th cycle with the optimum parameter  $\mathbf{a}_a^i$ :

$$\mathbf{x}_f^{i+1} = \varphi_{i+1}(\mathbf{x}_a^i, \mathbf{a}_a^i), \quad (34)$$

and to propagate the parameters using a persistent model,

$$\mathbf{a}_f^{i+1} = \mathbf{a}_a^i. \quad (35)$$

However, this persistent model can be a source of instability to the model integration with limited ensemble numbers. Gottwald and Majda (2013) explained this instability by the contribution of the unrealistic large covariance terms. In order to damp the rapid temporal change of model parameter, we set,

$$\mathbf{a}_a^i = \mathbf{a}_f^i + \beta_i \delta \widehat{\mathbf{a}}^i, \quad (36)$$

where  $\delta \widehat{\mathbf{a}}^i$  is the optimizer obtained at the  $i$ th cycle and  $\beta_i$  is a damping coefficient that needs be tuned in order to maximize the ensemble spread while not causing filter divergence based on trial-and-error principle.

In the framework of the augmented state technique, the parameters are treated as random variables. So their distribution must be set implicitly or explicitly. In our case we can use the estimators proposed in section 3.1. as prior values. The errors associated with those prior values of  $\mathbf{a}$  are assumed to be centered Gaussian fields, with variance given by an uncertainty representation of the Smagorinsky model:

$$\text{var}(\mathbf{a}) = C \frac{1}{2} \sum_{ij} \|\partial_{x_i} \delta w^j + \partial_{x_j} \delta w^i\| \mathbb{I}, \quad (37)$$

note  $\delta \mathbf{w}$  is defined as variation of high resolution velocity  $\mathbf{w} - \bar{\mathbf{w}}_k$  on local spatial region,  $\mathcal{C}(x_k)$ , centered on the coarse grid points indexed by  $k$ .



### 3.2.3. Ensemble Inflation

Ensemble methods typically require an explicit variance inflation scheme in order to maintain a relevant ensemble spread. Although such scheme is widely used in practice by either additive or multiplicative inflation formulations, it reflects a lack of knowledge on how to address properly the model errors responsible for the maintenance of the ensemble spread.

Here we adopt the multiplicative inflation (Anderson and Anderson, 1999) by multiplying the ensemble anomaly matrix with a tunable factor  $\gamma$  ranging from (1.0, 1.3) for each cycle. This strategy is applied to the standard model as a reference run but not to the stochastic model. Because one advantage of our stochastic model is to provide an inflation-free DA scheme. More precisely, our combined approach discussed in section 2. and 3.2. can be seen as an implicit inflation scheme where the corresponding scale discrepancy errors are dealt by the stochastic dynamics.

However, as shown in next section, the ensemble transformation approach with local analysis is particularly unable to produce sufficient ensemble spread leading to the filter divergence. To tackle such bad behavior, we employed a relaxation scheme in which part of the posterior ensemble perturbation is relaxed back to the prior perturbations (Zhang et al., 2004).

$$\mathbf{S}_{0,\text{relaxed}}^{k+1} = \overline{\mathbf{S}_0^{k+1}} + (1 - \alpha)(\mathbf{S}_0^k - \overline{\mathbf{S}_0^k}) + \alpha(\mathbf{S}_0^{k+1} - \overline{\mathbf{S}_0^{k+1}}). \quad (38)$$

with  $0 \leq \alpha \leq 1$ . This approach is similar to the conditional covariance inflation method used in Aksoy et al. (2006).

### 3.2.4. Parameter estimation effect

It can be emphasized that although the estimation of the initial condition and the parameter proceed simultaneously, the estimation process of the initial condition is nevertheless not directly related to the auxiliary covariance terms ( $\mathbf{B}_{aa}, \mathbf{B}_{ax}$ ) partly associated to the parameter  $\mathbf{a}$ . The  $\mathbf{a}$ , however, interferes in the quality of the analysis trajectory in two ways: firstly, it explicitly manifests itself on the model integration in the form of the estimated parameter  $\hat{\mathbf{a}}$ ; secondly, it implicitly affects the analysis state through the ensemble spread from which the propagation of the ensemble perturbation matrix is calculated. Depending on the context, either of the two factors can exhibit large impact to the result (Ruiz and Pulido, 2015; Koyama and Watanabe, 2010). Note that in our ensemble-based 4DVar approach, the parameter is not only related to the forecast-analysis cycling process, but also can affect the ensemble fields generated between two consecutive outer loops.

### 3.2.5. Parameter identifiability

Before we carry out the experiments, it is important to assess the parameter identifiability. This is especially pertinent in our case as the quadratic variation tensor  $\mathbf{a}$  is non-observable. This concept, following the definition in Navon (1998), can be

viewed as a criteria to decide whether or not the parameter of interest can be inferred from the data. With an ensemble-based method, an efficient way of examining the parameter identifiability consists in calculating the correlation coefficient (Tong and Xue, 2008).

$$\text{cor}(\mathbf{x}, \mathbf{a}) = \frac{\text{cov}(\mathbf{x}, \mathbf{a})}{\sqrt{\text{var}(\mathbf{x})\text{var}(\mathbf{a})}}. \quad (39)$$

A strong correlation suggests that any changes in the parameter space will heavily affect the state space through the cross-covariance part of the error covariance matrix and vice versa.

## 4. Model and experimental settings

In order to evaluate the methods proposed, we have first carried out experiments on a one-dimensional stochastic shallow water model (Eqs.(17)) with synthetic data. We extended then this procedure on a 2D stochastic shallow water model (Eqs.(15)) with both synthetic and real image data.

### 4.1. Model numerical scheme

The systems studied here concentrates on the large scale evolution of an average depth. We note that the subgrid-scale model associated with the quadratic tensor  $\mathbf{a}$  emerges both in the continuity equation and the momentum conservation equation. The finite volume method is directly applicable to these equations (Bradford and Sanders, 2002). Indeed, the numerical flux formulation of the new subgrid-scale terms is analog to the viscous terms.

### 4.2. Experimental settings

Several experiments have been conducted to evaluate the performance of the stochastic model. We first tested the 1D stochastic shallow water equation. The 1D domain is of length  $L_x = 6000$  km with the initial surface height  $h(x, 0)$ .

$$h(x, 0) = H_0 - \frac{fU_0x}{g} + A\xi, \quad (40)$$

where  $H_0 = 5000\text{m}$ ,  $f = 1.03 \times 10^{-4}\text{s}^{-1}$  is the Coriolis parameter,  $U_0 = 40\text{m/s}$  and  $g$  is the gravity acceleration. Let us remark that though the Coriolis force is not described in the original formulation for sake of clarity, it can be added without any difficulty to the model (see Resseguier et al. (2016a;b;c)). An additional noise is considered with  $A = 10$  as its amplitude. And  $\xi$  is a random Gaussian covariance field computed as the spectral methods discussed in Evensen (2003) (with decorrelation length equals to 20%L). The initial velocity field is inferred from the geostrophic relation. The true state used to construct the finer observation is obtained by integrating the standard shallow water equation with the initial condition (40) on 401 grid points. The coarse-resolution background state with the same initial condition is simulated with the stochastic model on only 101 grid points. The time step  $\Delta t$  is set to 37.5s for the

true state model and to 150s for the coarse-resolution model in order to satisfy the different CFL conditions. The synthetic observation is extracted from the true state every 600s by adding i.i.d Gaussian noise to the true state. Each assimilation window is set as 1800s therefore containing 3 observations. In order to maintain the balance of state variables for ensemble fields, the analysis started at 6000s.

For the 2D case, a very similar setup is implemented. The length of the domain is the same as  $L_x$  in 1D case. The width is  $L_y = 2400$  km. The initial height surface is a 2D field with the extra error term as a 2D random Gaussian field. The background state resolution is  $21 \times 51$  and the observation resolution is  $81 \times 201$ . The time step  $\Delta t$  is set to 10s for the true state model and to 50s for the coarse-resolution model. The synthetic observations are sampled every 150s from the true state. Each assimilation window is set as 450s therefore also containing 3 observations.

For both cases, we tested the cycling windows strategies in order to adapt the temporal variation of parameters. Here the parameters values change at the beginning of every assimilation window. In terms of the iterative scheme, we found it necessary to introduce similar parameter evolution model between two outer loops. For the 1D case, we firstly carried out a run of 30 assimilation windows in order to show the stability of the scheme. Then we adopted 6 windows along with a forecast stage whose duration equals 3 assimilation windows to show the details of error growth in the first few cycles. Similarly, 5 windows are chosen for the 2D case. Note that in these synthetic cases, the coarse grid nodes of the background state coincide with those of fine synthetic observations.

Finally, we applied this stochastic model to those Kinect-captured images. Kinect is a depth sensor developed by Microsoft as the motion sensing input device for game consoles. Those images related to the work of Combes et al. (2015) are captured by a Kinect sensor to reflect the evolution of a unidirectional wave generated by an initial free surface height difference  $\Delta h = 1$  cm on a grid of  $222 \times 88$  pixels. The flow is bounded in a rectangle flat tank of size  $L_x \times L_y = 250$  cm  $\times$  100 cm. The flow surface was located between 680 and 780 mm from the device. When the attenuation coefficient of the liquid is larger than  $113 \text{ m}^{-1}$ , the Kinect sensor displays a mean measurement error of 0.5 mm with standard deviations of about 0.5 mm for both flat and sinus-like surfaces. The sensor captures successfully sinus-like varying elevations with spatial periods smaller than 20 mm and amplitudes smaller than 2 mm. However, these raw data suffer large areas of missing information. Therefore, the same image pre-processing as in Yang et al. (2015) has been done in order to fix the issues related to the observation errors' spatial distribution and the background state, etc. For the observation errors: in terms of a point located in the unobserved region, the observation error is set to be dependent on the distance. The longer the distance, the larger the error. The observation error is bounded by a maximal value of 60% of the height difference  $\Delta h$ . Within the observed region, we set the observation error homogeneously to the instrument error

$\sigma_o = 0.5\% \Delta h$ . The simulation is executed on a  $124 \times 49$  rectangle grids mapped on the same domain. In this case, the initial background was completely unknown, hence it was set to a filtered observation with interpolated values on the missing data regions at the initial time. Since the initial background quality is bad, the system is designed to assimilate repeatedly the first few observations in order to obtain a faster 'spin-up'. So we chose five overlapping assimilation windows over nine observations with each window containing 5 observations.

In all cases, the observation error covariance matrix  $\mathbf{R}$  is assumed to be diagonal and all the observations are local in the sense that all the observations are taken at a certain spatial location. We employed different strategies to construct the ensemble in the synthetic case and in the real case. For the synthetic cases, as there is no explicit initial error associated with the background, the ensemble was initialized as a function of the parameters. And the prior parameters are determined from an empirical estimation of the variance tensor on local spatio-temporal regions,  $\mathcal{C}(x_k, t)$ , centered on the coarse grid points indexed by  $k$

$$\mathbf{a}(t_i, x_k) = C \frac{\delta t}{n^2 - 1} \sum_{i, l \in \mathcal{C}} [(\mathbf{w}(t_i, x_l) - \langle \mathbf{w}(t, x_c) \rangle) (\mathbf{w}(t_i, x_l) - \langle \mathbf{w}(t, x_c) \rangle)^T], \quad (41)$$

where  $\delta t$  is the characteristic time step used to regularize the unit of  $\mathbf{a}$  and  $\langle \bullet \rangle$  denotes the spatial-temporal mean. The error uncertainty is computed by equation (37). For the real cases, parameters are perturbed in addition to the initial state Gaussian perturbations to initialize the ensemble. Since no velocity observation is available at the initial time, the parameter initial observation is given by the initial ensemble variance:

$$\mathbf{a} = \frac{C \delta t}{N - 1} \sum_{j=1}^N [\mathbf{w}^j(t) - \bar{\mathbf{w}}(t)] [\mathbf{w}^j(t) - \bar{\mathbf{w}}(t)]^T. \quad (42)$$

### 4.3. Comparison criterion

With a reference state available in both synthetic cases, we used the Root Mean Square Error (RMSE) to measure the differences between the values predicted by a model or an estimator and the reference value. Such RMSE is defined through  $L_2$  norm:

$$\text{RMSE} = \|\mathbf{x}^{\text{target}} - \mathbf{x}^{\text{ref}}\|_2, \quad (43)$$

In the 2D synthetic case, we check the lengthwise velocity besides the surface height. The lengthwise velocity is the primary velocity component along  $L_x$ . In the real image case, since there is no ground truth, we compare the background/analysis at discrete time levels against the observations as well as the surface height profile obtained by different methods.

## 5. Results and discussions

### 5.1. 1D case results, synthetic observations

As for the 1D experiment, the spatial mean of the correlation coefficient calculated from equation (39) are  $\overline{\text{COR}}(h, a_{xx}) = 0.1649$  and  $\overline{\text{COR}}(u, a_{xx}) = 0.1804$ . Such an order of magnitude of the correlation coefficient ensures the identifiability of  $\mathbf{a}$  through the state variables.

A 30 pure assimilation cycles' test has been performed for the 4DEnVar schemes with perturbed observations and localized covariance. In order to evaluate the robustness of the proposed scheme and the effectiveness of the stochastic model, we carried out one test with the model under uncertainty (with suffix 'stochastic-model') and two tests on the "standard-model" in which no subgrid parameterization is introduced and the ensemble is generated by perturbing the initial condition state with the uncertainties computed a few time steps after the starting of the experiment. The two tests on the "standard-model" differs on whether or not the ensemble multiplicative inflation technique discussed in section 3.2.3. is used or not. The figure (1) shows the RMSE evolution in the course of the 30 cycles (here we use PO prefix denoting perturbed observations and LC prefix denoting localized covariance). The results are clearly in favor of our proposed inflation-free uncertainty model. The ensemble inflation technique improved to a certain extent the result but further improvement can hardly be obtained by simply enlarging the inflation coefficient  $\gamma$  since this causes numerical instability for the standard model.

The RMSE curves plotted in figure (2) represent the results obtained by the different assimilation configurations mentioned in section 3.2. After a 6 windows cycling of assimilation procedure, we run a forecast stage, whose duration equals the length of 3 assimilation windows. We have discussed in section 3.2.4. about the two different sources of the parameter estimation effect. The purpose of a prevision stage is to study only the effect upon the numerical integration using the prescribed stochastic model. In terms of the performance with the model under uncertainty, we have found that with no damping in parameter evolution model, insufficient number of ensemble members ( $< 50$ ) lead to model blow-up characterized by unrealistic model variables' values. This is not shown in figure 2 since it makes the figure scale unreadable. By introducing adequate damping factor  $\beta \in [0.2, 0.5]$  which attenuates the change rate of variance tensor  $\mathbf{a}$ , the black curve displayed in figure 2 only experiences a strong divergence in the first cycle, while this behavior is dampened at later cycles. The analysis trajectories across the assimilation stage show that increasing the ensemble size allows achieving a considerable improvement in the analysis quality in terms of the initial condition and indirectly of the parameters. We also find that the localization technique with optimally tuned localization scales enables improving the result for ensemble of small dimension. Note that compared to a localization procedure, increasing the ensemble size do not improve significantly

the analysis performance. In fact the RMSE curve using 512 sampling numbers and no localization scheme was very close to the one with only 32 members. Therefore it has not been included in the figure for sake of readability.

In terms of the forecast stage (starting from 15000 in figure 2), the forecast trajectories associated with better analysis states from the last cycle yield less error and distinct error growth schematic compared to the background. Conversely, the error evolutions with the standard model (with or without inflation) coincide closely with the one of the background. This reflected the power of the subgrid modeling in producing more relevant analysis state. Nevertheless, a similar divergence can be found in all the forecast trajectories. This behavior indicates that the effect of the parameters here is only slightly significant with respect to forecast. This seems to indicate that the strategy consisting to keep the variance tensor constant along the assimilation window (and therefore for the whole forecast) does not provide a slower slope of the error growth at long terms. Its assimilation performance enables nevertheless to get forecast results of much better quality. Choosing a time-dependent variance tensor may lead to better result for long-terms forecasting.

Eventually we conjectured that the effectiveness of the corresponding methods is actually related to the ensemble spread. The ensemble spread curves shown in figure (3) correspond to three identical experiment setup (No localization scheme with same sampling numbers) but with the state driven by the stochastic model and the standard model with or without inflation scheme respectively. Note that the first cycle is not shown here since during this cycle the ensemble spread is dominated by arbitrary noise. We observed that the ensemble spread for stochastic model is significantly larger than the spreads associated with the standard model (with or without inflation). The spread curve of the standard model without inflation not only decreases faster but also collapses of nearly 100% after the fourth cycle. The inflation technique do increase the ensemble spread but its power is intrinsically limited by the model stability. On the other hand, we observed that the ensemble spread of the stochastic model also decreases but is still maintained at a rather high level across the sliding windows. A relatively high-level ensemble spread explains why the forecast state is corrected effectively at the beginning of each cycle in terms of the RMSE curve (e.g. the black dotted line in figure 2); similarly, a collapsed ensemble related to the standard model loses gradually the ability of improving the forecast state.

We have also conducted a study on the different terms composing the subgrid model in Eqs.(17). The configuration of different cases are shown in table (1). The RMSE evolution along 3 cycles are illustrated in figure (4). We can conclude that the stochastic shallow water model works equally well with the case  $\mathbf{a}_I$  (complete terms presented) and the case  $\mathbf{a}_{II}$  (no term associated with the gradient of  $\mathbf{a}$  in the momentum equation). Case  $\mathbf{a}_{III}$  does not take account of the terms associated with the gradient of  $\mathbf{a}$  in both mass and momentum conservation equations. The initial analysis resulting from this case is relatively bad

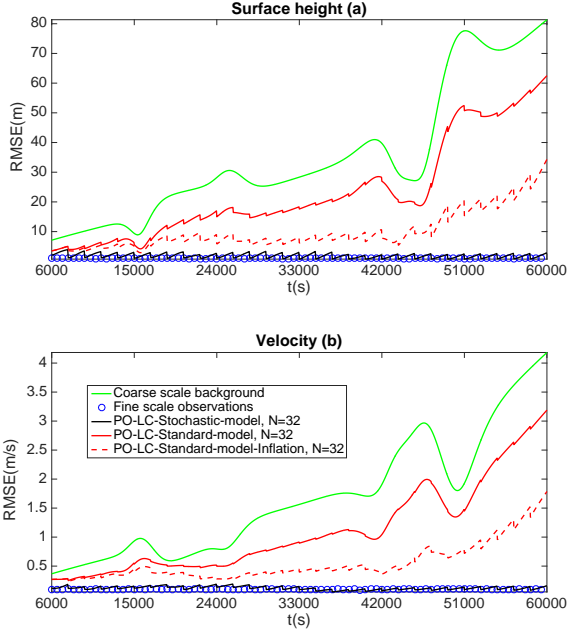


Fig. 1. RMSE comparison in terms of free surface height (a) and velocity (b) between various configurations. PO: Perturbed observation. LC: Localized covariance approach. Standard model: no subgrid tensor parameterization. N: number of ensemble member.

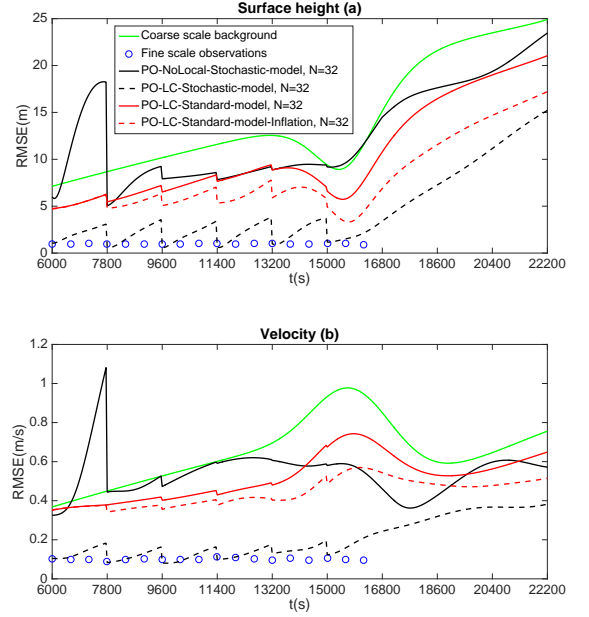


Fig. 2. RMSE comparison in terms of free surface height (a) and velocity (b) between various configurations of 4DVar. PO: Perturbed observation. LC: Localized covariance approach. Standard model: no subgrid tensor parameterization. N: number of ensemble member. 6 DA cycle phase:  $t \in [6000, 16800]$ s. Forecast phase:  $t \in [15000, 22200]$ s

Table 1. Different combination of the terms composing the subgrid model.

	Momentum equation	Continuity Equation
$\mathbf{a}_I$	$-\frac{1}{2}a_{xx}\partial_x(hu_x) - \frac{1}{2}a_{xx}\partial_x(uh_x) - \partial_x a_{xx}\partial_x(hu)$	$-\frac{1}{2}a_{xx}\partial_{xx}h - \partial_x a_{xx}\partial_x h$
$\mathbf{a}_{II}$	$-\frac{1}{2}a_{xx}\partial_x(hu_x) - \frac{1}{2}a_{xx}\partial_x(uh_x)$	$-\frac{1}{2}a_{xx}\partial_{xx}h - \partial_x a_{xx}\partial_x h$
$\mathbf{a}_{III}$	$-\frac{1}{2}a_{xx}\partial_x(hu_x) - \frac{1}{2}a_{xx}\partial_x(uh_x)$	$-\frac{1}{2}a_{xx}\partial_{xx}h$
$\mathbf{a}_{IV}$	$-\frac{1}{2}a_{xx}\partial_x(hu_x)$	none

compared to the previous two cases, which indicates the relevance of the missing modeling part of the continuity equation in generating larger ensemble spread. Case  $\mathbf{a}_{IV}$  assumes a similar form as the Smagorinsky-like eddy viscosity model described in the previous section without any subgrid term in the continuity equation. Besides the significantly lower quality of initial analysis state, the evolution of RMSE diverges faster and to a greater extent compared to other 3 counterparts. This test suggests that the subgrid modeling term presented in the continuity equation is crucial to keep an accurate representation of the small-scale uncertainty constituting a more balanced ensemble spread between the height and the velocity fields. Note that case  $\mathbf{a}_I$  is the default case for all the other tests conducted in this article.

## 5.2. 2D case results, synthetic observations and real data

Under the 2D experimental setup, due to the sizable degrees of freedom of the system, the localization-free paradigm basically does not work anymore. So here we focus on the comparison of two localization techniques: localized covariance (LC) and local analysis (LA). Sakov and Bertino (2011) have shown the similarities of the two types of localization scheme within the framework of EnKF under some restricted conditions. In our practice, other factors interfere so that the similarity of the two different localization scheme may break down. One of factors is the effectiveness of the minimization scheme since our localization techniques are constructed on a variational basis. The background error statistics as well as the relevance of the initial ensemble also contribute to the inequality. Another important factor is the ensemble update schemes used in combination with LC or LA technique. We use perturbed observation scheme (PO) with LC and ensemble transform scheme (ET) in conjunction with LA. In order to separate the two factors (the ensemble update scheme and the localization scheme), we employed another scheme (section 3.2.1.3.) using global covariance localization and a deterministic ensemble update approach (denoted as DE-LC) without perturbing the observations. All those have

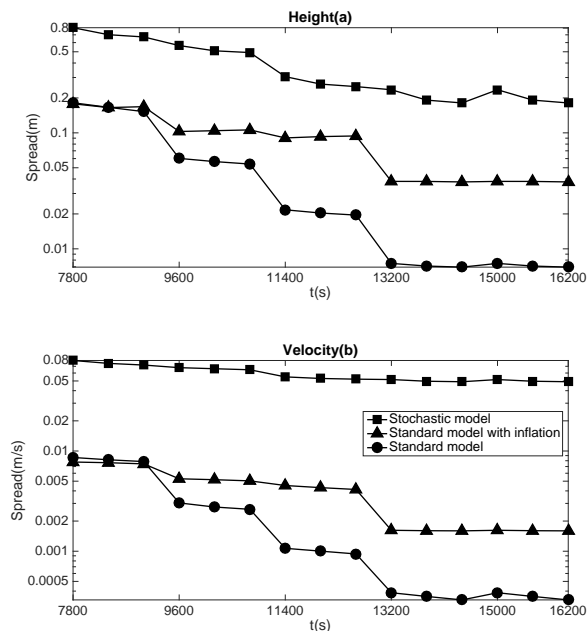


Fig. 3. Ensemble spread in terms of free surface height (a) and velocity (b): standard model (circle) with ensemble in function of initial states; standard model (triangle) with ensemble in function of initial states but with multiplicative inflation technique; subgrid model (square) with ensemble in function of initial parameters.

been run for the 2D stochastic large-scale dynamics described eqs. (15a,15b).

To be able to diagnose the performances, the first step is to study the sensitivities of localization scales. The localization scale is termed as cut-off distance in terms of LC and local domain size in terms of LA respectively. Figure 5 shows the mean of RMSE of height per cycle against different localization scales. We observed that the LC approach tolerates a larger scale while the LA approach clearly is more in favor of a smaller scale. Although the average RMSE monotonically increased in function of the local domain size when using LA, the local domain size smaller than  $10\%L_x$  contains only several grid points which results in numerical instability. Eventually, we chose  $20\%L_x$  as the cutoff distance for LC and  $10\%L_x$  as the local domain size for LA.

The RMSE results regarding the 2D Synthetic Results are shown in figure (6). In terms of the group of three LC scheme, the error evolution of the localized covariance (LC) cases with stochastic model indicated consistent sensible results. The result of a PO-LC test using the standard shallow water model and multiplicative inflation technique shows relatively higher error and a diverged velocity RMSE from the 2nd cycle compared to those associated to the stochastic model using LC scheme. Between two ensemble update schemes using LC and the stochas-

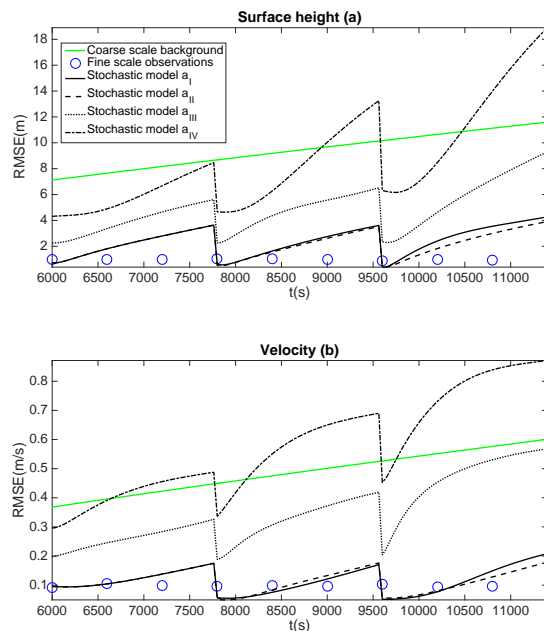


Fig. 4. RMSE comparison in terms of free surface height (a) and velocity (b) between various subgrid model configurations in table 1 of stochastic shallow water model (17). The meaning of different  $\alpha$  is shown in table 1.

tic subgrid modeling, the DE scheme yield comparable result in height field but slightly worse result in velocity field. In terms of ET-LA scheme, as shown by the black solid line in figure (6), a similar catastrophic filter divergence to the 1D case can be found in the first cycle with underdamped parameter model. We can conclude that the damping parameter model is essential to the first cycle for the parameters estimated in the first cycle are badly constrained. After optimally tuning the value of  $\beta$ , we can see that the analysis state for LA approach (the black dashed line in figure (6)) is corrected at the 1st cycle and its trajectory does not diverge anymore. However, the analysis across later cycles is characterized by a monotonically increasing error caused by the collapsing of ensemble spread. Because when the ensemble spread is too small, the state variables will conform to the background and the observation will barely exert any effect. Parameter ensemble collapsing is a common problem encountered by many studies of parameter estimation in an augmented state vector framework (Aksoy et al., 2006; Simon et al., 2015). Regarding our test, firstly, after the 1st cycle, the estimated parameters have to be severely tampered before they can be used to evolve the analysis ensemble to the beginning of the 2nd cycle (therefore formulating the forecast ensemble). This damping process is a source of loss of signal in the ensemble parameter fields. Secondly, during the assimilation process, the variances of both the state variables and the parameters are naturally decreased. We concluded that the reason why the ensemble fields

deteriorates in LA is linked to the form of calculating the ensemble anomaly fields in local domains. Since the noise that we are trying to characterize in this paper is mainly due to the non-consistent spatial and temporal sampling numerical steps. Consequently, using a transformation matrix of small dimension can lead to excessive small-scale variations and eliminate the longer correlations since it restricts the directions of the variations of increments in the ensemble space. To verify this, we plot the ensemble anomalies' PSD (power spectral density) in terms of height with respect to the forecast ensemble, the analysis ensemble of LC and LA respectively for the 1st cycle in figure 7. We observed that after integrating the stochastic model for some time, the forecast ensemble is clearly dominated by few leading modes. The analysis ensemble yielded by LC preserves considerably the spectral form while the magnitude associated with each mode decreases. The result clearly reveals why the LC approach is more robust even with a relative small ensemble size. However, the computational cost associated with LC approach of 64 members ensemble is already considerably high. The analysis ensemble produced by LA, on the contrary, spreads the energy over much more modes. These small-scales energy variations are then very likely smoothed out by the diffusive subgrid model.

In order to make LA work better, we adopt the relaxation methods discussed in section 3.2.3. in which  $\alpha$  is tuned as 0.9 for state ensemble. As shown by the black dotted line in figure 6, the LA approach with relatively large ensemble generates sensible analysis comparable with the LC experiments.

One possible fix besides this relaxation method is to use left-multiplication ensemble transformation schemes which is in the state vector space. Note that this scheme is both compatible with LC and LA. However, the left-multiplied form will require the minimization process to be done in state vector space which is more computational demanding. Another possible approach would be to introduce a static full-rank B and to apply a 3DVar-FGAT in order to add more directions in the incremental field which is not solely covered by the ensemble B as suggested in Gustafsson and Bojarova (2014). This will in fact add a hybrid taste in our EnVar scheme, which constitutes an interesting perspective for advanced study. From the perspective of image data, it is also beneficial to introduce correlation into the observation error covariance matrix (Chabot et al., 2015). This remains another interesting outlook for future work.

In terms of the real image data depicting the free surface height of a laboratory tank, figure (8a) shows the evolution of the mean surface height of the wave crest region using different methods. We particularly chose to focus on this region as the flow is dominated by a single wave. Figure (8b) shows the RMSE between background/analysis state and the observation at discrete observation times. Note the analysis trajectories denoted by 4DVar, 4DEnVar-PO-LC and 4DEnVar-ET-LA are the same as in figure 8 in Yang et al. (2015), which are produced by the standard model without any subgrid parameterizations. No ensemble inflation technique is used in this test. Basically,

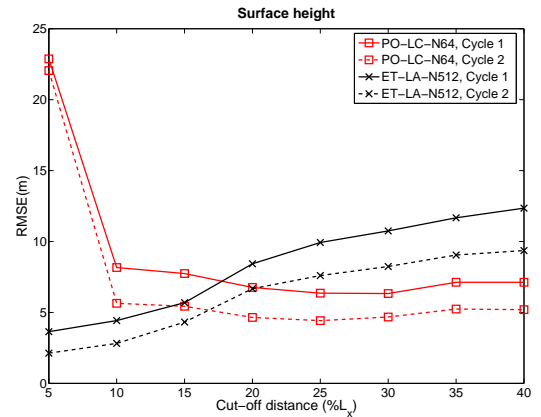


Fig. 5. The RMSE of height comparison in the 1st and 2nd cycle as a function of Cut-off distance (LC, localized covariance) or local domain size (LA, local analysis).

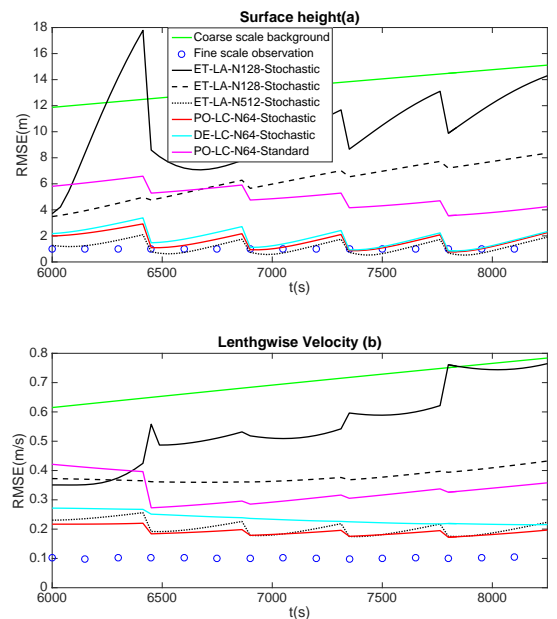


Fig. 6. RMSE comparison in terms of free surface height (a) and lengthwise velocity (b) between various configurations of 4DEnVar. ET, ensemble transform. PO, perturbed observation. DE, deterministic ensemble. LC, localized covariance. LA, local analysis. ET-LA-N128-Stochastic (solid line) case with  $\beta = 1$  shows ensemble 'blow up'; ET-LA-N128-Stochastic (dashed line) with  $\beta = 0.5$  shows ensemble collapse; ET-LA-N512-Stochastic (dotted line) used ensemble relaxation inflation  $\alpha = 0.9$  technique for state ensemble; PO-LC-N64-Standard used standard model and multiplicative inflation technique for state ensemble.

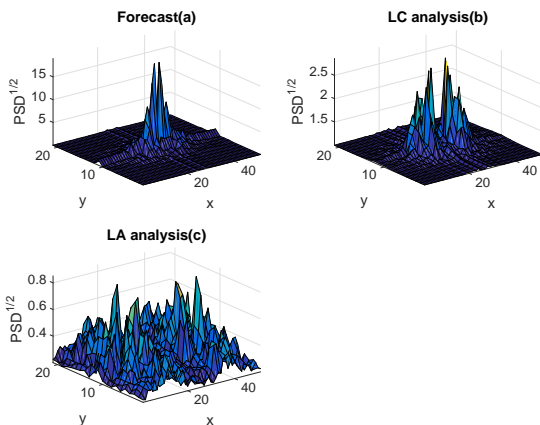


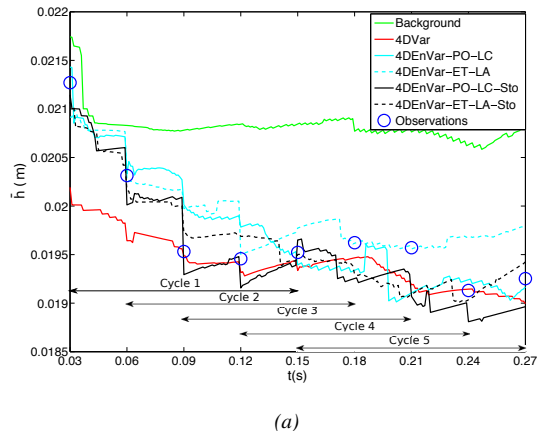
Fig. 7. Power spectral density of ensemble anomalies for the forecast ensemble (a), the analysis ensemble by LC (b) and LA (c).

the group of 4DEnVar outperforms the 4DVar in terms of the surface height at the beginning of the assimilation window. We observed an underestimation of the surface height for 4DVar in figure (8a), while the group of 4DEnVar follows quite well the first 2 observations. The analysis trajectories generated by EnVar schemes with stochastic models (with Sto suffix) show a slight improvement between the 3rd and the 6th observations. Their trajectories started to attenuate in the last cycle compare to their counterpart using standard model. This local phenomenon is again validated in the global RMSE results of figure (8b). We can observe that the RMSE values regarding the stochastic models between the 3rd and the 6th observations are consistent with the 4DVar case. Lower RMSE associated with stochastic models are also found in the last cycle (black symbols versus cyan symbols). We recalled in this experiment the initial background state is unknown, which makes the numerical performance of the model highly unstable. Hence, the parameterization procedure can only tolerate a relatively low magnitude of the variance tensor, which cannot induce significant ensemble perturbations. All in all, the stochastic version has been found more stable than the other schemes.

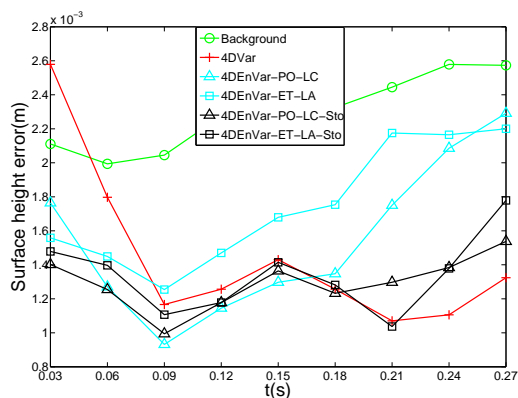
We have also plotted the ensemble spread versus time in figure (9). As in the synthetic case, the ensemble spread associated with subgrid model is larger than those of the standard model for the whole assimilation windows. However, the difference of ensemble spread between the two models is less evident compared to the synthetic case, which indicate the perturbed initial states are mainly responsible for generating relevant background error estimations especially for the later cycles. This behavior also sheds light on the divergence of the analysis trajectory.

## 6. Concluding remarks

In this paper, we started by presenting shortly the principle of the stochastic fluid dynamics proposed in Mémin (2014), which



(a)



(b)

Fig. 8. Comparison of various configurations of 4DEnVar w.r.t. (a) Mean surface height of the wave crest region as a function of time and (b) The RMSE against observation at observation time. 4DVar, 4DEnVar-PO-LC and 4DEnVar-ET-LA are tested with the standard model; 4DEnVar-PO-LC-Sto and 4DEnVar-ET-LA-Sto are tested with the stochastic model. LC, localized covariance. PO, perturbed observation. LA, local analysis. ET, ensemble transformation.

offers us a way to reinterpret the so called subgrid stress models used in CFD for many years. In this framework, the subgrid models are theoretically sound as they are directly related to the small-scale uncertainty components and inferred from physical conservation laws. They are therefore not prone to the assumptions underlying the eddy viscosity models (to list few of them: filter permutation with differentiation, specific shape of the dissipation, isotropic turbulence, diffusion alignment with the rate of strain proper directions).

Then we have focused on the numerical evaluation of the stochastic model with a coherent scheme 4DEnVar proposed in Yang et al. (2015). The idea is this: since we argue that the subgrid-scale stress term attached to the uncertainties plays an essential role in formulating the effects of the uncertainties on the resolved scales, we can indeed try to estimate these effects



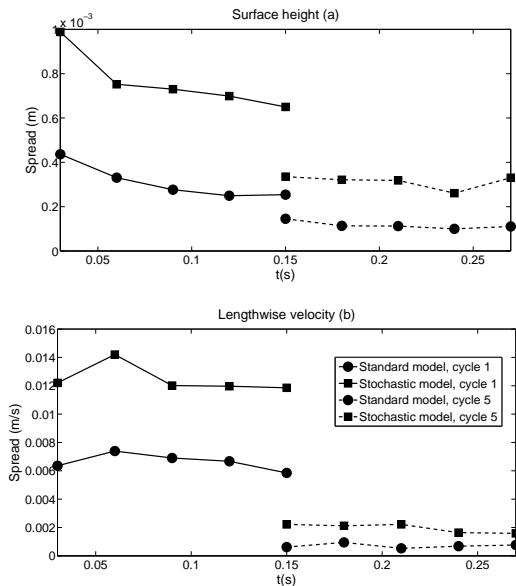


Fig. 9. Ensemble Spread in terms of free surface height (a) and lengthwise velocity (b) for: standard model cycle 1 and cycle 5 (circle with solid/dashed line); stochastic model cycle 1 and cycle 5 (square with solid/dashed line).

through high-resolution data by a parametrization of the associated terms. In our 4DEnVar scheme, the unknown augmented initial vector is adjusted to fit the 4D observations simultaneously. Besides, due to its cycling and iterative ensemble update strategy, we are able to fully utilize the 4D Ensemble in order to meet the needs of dealing with bad background and incomplete observations.

The numerical results have been encouraging. Using the stochastic shallow water model with synthetic data sets, the 4DEnVar scheme is tested to compensate the gap between the coarse resolution background state and fine resolution observations. The results clearly indicated the pertinence of the quantification of the unknown small scale physical processes or numerical artifacts. We have a much better analysis state in global quality in terms of spatial errors; and the error growth tendency of the analysis trajectory with the estimated subgrid-scale stress parameters is distinct from other cases driven by the standard model (with/without inflation), which suggests that more relevant information from the observation can be extracted and contribute to the formulation of a more correct and coherent analysis state. Such an improvement is largely due to two related subjects. The stochastic model introduced more relevant ensemble spread growth and avoided the common catastrophe of ensemble collapse due to the insufficient ensemble spread. The ensemble-based DA scheme, therefore benefited from these 4D ensemble in formulating the necessary components (innovation vector, ensemble anomaly matrix, etc.), and consequently gen-

erated more accurate analysis states and parameter fields for the stochastic model. It is important to note that the subgrid-scale stress term appeared in the stochastic model, although only labeled as resolution-related, can actually embrace broader types of errors as long as these errors can be inferred from the discrepancies between the dynamics and the high-resolution data.

The numerical results are also discussed in detail with two localization techniques. In ensemble-based methods, localization is useful not only in removing the unrealistic correlation, but also in enlarging the control space to a space with greater degree of freedom. With 2D synthetic or real case, using localization is obligatory. The localized covariance approach is proved to be more robust in the sense that the posterior ensemble produced by such scheme not only kept its magnitude, but also maintained to a large extent its energy spectral form. The spectral form is important in terms of our experimental setups since it prevented the small scale energy variations from diffusing by the stochastic model, which is exactly the case with the local analysis scheme. The results revealed by the real data case is less convincing due to the numerical instabilities associated with the variance tensor. The analysis trajectory only showed a slight improvement with a more stable behavior in the analysis. More work needs to be done in order to provide a better initial background, which eventually allows generating a more pertinent prior ensemble. Employing the full stochastic model with the noise term constitutes an interesting perspective for this path.

Our results suggested that the two localization techniques both suffered some limitations in their current forms. In the localized covariance approach, the posterior ensemble can only be sought through an ensemble analysis, which is computational demanding. The local analysis, although much cheaper and easy to implement, has shown some difficulties in bridging the gap between the local and the global representations. Following works will investigate new localization scheme that do not have the above restrictions. One way to improve the posterior ensemble efficiency and accuracy is to construct the estimation scheme on adjoint models. By introducing the adjoint models, the optimization is done in the state space rather than the ensemble space. Although maintaining and computing an adjoint model is cumbersome and time-consuming, using a state space formulation allows to express the left transformation of the prior ensemble to the posterior ensemble (Bocquet, 2015). In this form, the ensemble transformation takes the form of EAKF (Anderson, 2001; Sakov and Bertino, 2011). In the future we are planning to adopt similar strategy with complex data sets. Since the scheme is formulated on variational base, the transition is straightforward. Adding a static full-rank B and applying a 3DVar-FGAT (Gustafsson and Bojarova, 2014) can also be beneficial to the posterior ensemble quality as more directions are permitted to grow in the incremental field. This path constitutes another promising direction to follow.



## References

- Aksoy, A., Zhang, F., and Nielsen-Gammon, J. W. (2006). Ensemble-based simultaneous state and parameter estimation in a two-dimensional sea-breeze model. *Monthly Weather Review*, 134(10):2951–2970.
- Anderson, J. L. (2001). An ensemble adjustment Kalman filter for data assimilation. *Mon. Wea. Rev.*, 129(12):2884–2903.
- Anderson, J. L. and Anderson, S. L. (1999). A monte carlo implementation of the nonlinear filtering problem to produce ensemble assimilations and forecasts. *Monthly Weather Review*, 127(12):2741–2758.
- Baek, S.-J., Hunt, B. R., Kalnay, E., Ott, E., and Szunyogh, I. (2006). Local ensemble Kalman filtering in the presence of model bias. *TellusA*, 58(3):293–306.
- Berezziat, D. and Herlin, I. (2015). Coupling dynamic equations and satellite images for modelling ocean surface circulation. *Com. in Comp. and Inf. Science*, 550:191–205.
- Beyou, S., Cuzol, A., Gorthi, S., and Mémin, E. (2013). Weighted ensemble transform kalman filter for image assimilation. *TellusA*, 65:18803.
- Blum, J., Le Dimet, F.-X., and Navon, I. M. (2009). Data Assimilation for Geophysical Fluids. In *Handbook of Numerical Analysis, Vol. XIV*, pages 385–441. Elsevier.
- Bocquet, M. (2015). Localization and the iterative ensemble kalman smoother. *Quarterly Journal of the Royal Meteorological Society*.
- Bocquet, M. and Sakov, P. (2013a). An iterative ensemble Kalman smoother. *Quarterly Journal of the Royal Meteorological Society*, 140(682):1521–1535.
- Bocquet, M. and Sakov, P. (2013b). Joint state and parameter estimation with an iterative ensemble Kalman smoother. *Nonlinear Processes in Geophysics*, 20(5):803–818.
- Bradford, S. F. and Sanders, B. F. (2002). Finite-volume model for shallow-water flooding of arbitrary topography. *Journal of Hydraulic Engineering*, 128:289–298.
- Buehner, M. (2005). Ensemble-derived stationary and flow-dependent background-error covariances: Evaluation in a quasi-operational NWP setting. *Quart. J. Roy. Meteor. Soc.*, 131(607):1013–1043.
- Buehner, M., Houtekamer, P. L., Charette, C., Mitchell, H. L., and He, B. (2010). Intercomparison of Variational Data Assimilation and the Ensemble Kalman Filter for Global Deterministic NWP. Part I: Description and Single-Observation Experiments. *Mon. Wea. Rev.*, 138(5):1550–1566.
- Buizza, R., Miller, M., and Palmer, T. (1999). Stochastic representation of model uncertainties in the ecmwf ensemble prediction system. *Quarterly Journal Royal Meteorological Society*, 125:2887–2908.
- Cacuci, D., Navon, I., and Ionescu-Bujor, M. (2013). *Computational Methods for Data Evaluation and Assimilation*. CRC press.
- Chabot, V., Nodet, M., Papadakis, N., and Vidard, A. (2015). Accounting for observation errors in image data assimilation. *Tellus A*, 67:23629.
- Combes, B., Heitz, D., Guibert, A., and Mémin, E. (2015). A particle filter to reconstruct a free-surface flow from a depth camera. *Fluid Dynamics Research*. minor revision.
- Corpetti, T., Héas, P., Mémin, E., and Papadakis, N. (2009). Pressure image assimilation for atmospheric motion estimation. *TellusA*, 61(1):160–178.
- Courtier, P., Thépaut, J.-N., and Hollingsworth, A. (1994). A strategy for operational implementation of 4d-var, using an incremental approach. *Quart. J. Roy. Meteor. Soc.*, 120(519):1367–1387.
- Cuzol, A. and Mémin, E. (2009). A stochastic filtering technique for fluid flow velocity fields tracking. *IEEE Tran. on Pattern Analysis and Machine Intelligence*, 31(7):1278–1293.
- Da Prato, G. and Zabczyk, J. (1992). *Stochastic equations in infinite dimensions*. Cambridge University Press.
- Dee, D. P. (2005). Bias and data assimilation. *Quarterly Journal of the Royal Meteorological Society*, 131(613):3323–3343.
- Evensen, G. (1994). Sequential data assimilation with a non linear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *J. Geophys. Res.*, 99 (C5)(10):143–162.
- Evensen, G. (2003). The ensemble Kalman filter: Theoretical formulation and practical implementation. *Ocean Dynamics*, 53:343–367.
- Fairbairn, D., Pring, S. R., Lorenc, A. C., and Roulstone, I. (2013). A comparison of 4DVar with ensemble data assimilation methods. *Quart. J. Roy. Meteor. Soc.*, 140(678):281–294.
- Frederiksen, J. S., O’Kane, T. J., and Zidikheri, M. J. (2013). Subgrid modelling for geophysical flows. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 371(1982):20120166.
- Gaspari, G. and Cohn, S. E. (1999). Construction of correlation functions in two and three dimensions. *Quart. J. Roy. Meteor. Soc.*, 125(554):723–757.
- Gordon, N., Salmond, D., and Smith, A. (1993). Novel approach to nonlinear/non-gaussian bayesian state estimation. *IEEE Processing-F*, 140(2).
- Gottwald, G. A. and Majda, A. (2013). A mechanism for catastrophic filter divergence in data assimilation for sparse observation networks. *Nonlinear Processes in Geophysics*, 20(5):705–712.
- Greybush, S. J., Kalnay, E., Miyoshi, T., Ide, K., and Hunt, B. R. (2011). Balance and Ensemble Kalman Filter Localization Techniques. *Monthly Weather Review*, 139(2):511–522.
- Gu, Y. and Oliver, D. S. (2007). An Iterative Ensemble Kalman Filter for Multiphase Fluid Flow Data Assimilation. *Spe Journal*, 12(04):438–446.
- Gustafsson, N. and Bojarova, J. (2014). Four-dimensional ensemble variational (4D-En-Var) data assimilation for the High Resolution Limited Area Model (HIRLAM). *Nonlinear Processes in Geophysics*, 21(4):745–762.
- Heitz, D., Memin, E., and Schnoerr, C. (2010). Variational fluid flow measurements from image sequences: synopsis and perspectives. *Exp. in Fluids*, 48(3):369–393.
- Hunt, B. R., Kostelich, E. J., and Szunyogh, I. (2007). Efficient data assimilation for spatiotemporal chaos: A local ensemble transform Kalman filter. *Physica D*, 230:112–126.
- Huot, E., Herlin, I., and Papari, G. (2013). Optimal orthogonal basis and image assimilation: Motion modeling. In *IEEE Int. Conf. on Comp. Vision*, Sydney, Australia.
- Kadri Harouna, S. and Mémin, E. (2016). Stochastic representation of the Reynolds transport theorem: revisiting large-scale modeling. working paper or preprint.
- Kalnay, E. and Yang, S.-c. (2010). Accelerating the spin-up of Ensemble Kalman Filtering. *Quarterly Journal of the Royal Meteorological Society*, 136(651):1644–1651.
- Kang, J.-S., Kalnay, E., Liu, J., Fung, I., Miyoshi, T., and Ide, K. (2011). variable localization in an ensemble kalman filter: Application to the carbon cycle data assimilation. *Journal of Geophysical Research: Atmospheres (1984–2012)*, 116(D9).
- Kazantsev, E. (2012). Sensitivity of a shallow-water model to parame-

- ters. *Nonlinear Analysis: Real World Applications*, 13(3):1416–1428.
- Koyama, H. and Watanabe, M. (2010). Reducing forecast errors due to model imperfections using ensemble kalman filtering. *Monthly Weather Review*, 138(8):3316–3332.
- Kunita, H. (1990). *Stochastic flows and stochastic differential equations*. Cambridge University Press.
- Le Dimet, F.-X. and Talagrand, O. (1986). Variational algorithms for analysis and assimilation of meteorological observations: theoretical aspects. *Tellus*, 38A:97–110.
- Leith, C. (1990). Stochastic backscatter in a subgrid-scale model: plane shear mixing layer. *Phys. of Fluids*, 2(3):1521–1530.
- Lions, J. L. (1971). *Optimal control of systems governed by PDEs*. Springer-Verlag, New-York.
- Liu, C., Xiao, Q., and Wang, B. (2008). An Ensemble-Based Four-Dimensional Variational Data Assimilation Scheme. Part I: Technical Formulation and Preliminary Test. *Mon. Wea. Rev.*, 136(9):3363–3373.
- Mason, P. and Thomson, D. (1992). Stochastic backscatter in large-eddy simulations of boundary layers. *J. of Fluid Mech.*, 242:51–78.
- Mémin, E. (2014). Fluid flow dynamics under location uncertainty. *Geophysical and Astrophysical Fluid Dynamics*, 108(2):119–146.
- Miller, R. N., Carter, E. F., and Blue, S. T. (1999). Data assimilation into nonlinear stochastic models. *Tellus A*, 51(2):167–194.
- Navon, I. (1998). Practical and theoretical aspects of adjoint parameter estimation and identifiability in meteorology and oceanography. *Dynamics of Atmospheres and Oceans*, 27(1):55–79.
- Ott, E., Hunt, B. R., Szunyogh, I., Zimin, A. V., E. J. Kostelich, M. C., Kalnay, E., Patil, D. J., and Yorke, J. A. (2004). A local ensemble Kalman filter for atmospheric data assimilation. *Tellus*, 56A:415–428.
- Papadakis, N. and Mémin, E. (2008). Variational Assimilation of Fluid Motion from Image Sequence. *SIAM Journal on Imaging Sciences*, 1(4):343–363.
- Pulido, M. and Thuburn, J. (2005). Gravity wave drag estimation from global analyses using variational data assimilation principles. i: Theory and implementation. *Quarterly Journal of the Royal Meteorological Society*, 131(609):1821–1840.
- Resseguier, V., Mémin, E., and Chapron, B. (2016a). Geophysical flows under location uncertainty, part I random transport and general models. Technical report, <https://hal.inria.fr/hal-01391420v2/document>.
- Resseguier, V., Mémin, E., and Chapron, B. (2016b). Geophysical flows under location uncertainty, part II Quasigeostrophic models and efficient ensemble spreading. Technical report, <https://hal.inria.fr/hal-01391420v2/document>.
- Resseguier, V., Mémin, E., and Chapron, B. (2016c). Geophysical flows under location uncertainty, part III SQG and frontal dynamics under strong turbulence. Technical report, <https://hal.inria.fr/hal-01391420v2/document>.
- Ruiz, J. and Pulido, M. (2015). Parameter estimation using ensemble-based data assimilation in the presence of model error. *Monthly Weather Review*, 143(5):1568–1582.
- Ruiz, J. J., Pulido, M., and Miyoshi, T. (2013). Estimating model parameters with ensemble-based data assimilation: Parameter covariance treatment. *Journal of the Meteorological society of Japan*, 91(4):453–469.
- Sagaut, P. (2006). *Large eddy simulation for incompressible flows: an introduction*. Springer Science & Business Media.
- Sakov, P. and Bertino, L. (2011). Relation between two common localisation methods for the enfk. *Computational Geosciences*, 15(2):225–237.
- Sakov, P. and Oke, P. R. (2008). A deterministic formulation of the ensemble Kalman filter: an alternative to ensemble square root filters. *Tellus A*, 60(2):361–371.
- Sakov, P., Oliver, D. S., and Bertino, L. (2012). An Iterative EnKF for Strongly Nonlinear Systems. *Mon. Wea. Rev.*, 140(6):1988–2004.
- Sawada, M., Sakai, T., Iwasaki, T., Seko, H., Saito, K., and Miyoshi, T. (2015). Assimilating high-resolution winds from a Doppler lidar using an ensemble Kalman filter with lateral boundary adjustment. *TellusA*, 67(0):1–13.
- Shutts, G. (2005). A kinetic energy backscatter algorithm for use in ensemble prediction systems. *Quarterly Journal of the Royal Meteorological Society*, 612:3079–3012.
- Simon, E., Samuelsen, A., Bertino, L., and Mouysset, S. (2015). Journal of Marine Systems. *Journal of Marine Systems*, 152(C):1–17.
- Slingo, J. and Palmer, T. (2011). Uncertainty in weather and climate prediction. *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, 369(1956):4751–4767.
- Smagorinsky, J. (1963). General circulation experiments with the primitive equations: I. the basic experiment\*. *Monthly weather review*, 91(3):99–164.
- Souopgui, I. (2010). *Assimilation d’images pour les fluides géophysiques*. PhD thesis, Université Joseph-Fourier-Grenoble I.
- Talagrand, O. (1997). Assimilation of observations, an introduction. *J. Meteor. Soc. Japan*.
- Tandeo, P., Pulido, M., and Lott, F. (2014). Offline parameter estimation using EnKF and maximum likelihood error covariance estimates: Application to a subgrid-scale orography parametrization. *Quarterly Journal of the Royal Meteorological Society*, 141(687):383–395.
- Titaut, O., Vidard, A., Souopgui, I., and Le Dimet, F.-X. (2010). Assimilation of image sequences in numerical models. *Tellus A*, 62(1):30–47.
- Tong, M. and Xue, M. (2008). Simultaneous estimation of microphysical parameters and atmospheric state with simulated radar data and ensemble square root kalman filter. part i: Sensitivity analysis and parameter identifiability. *Monthly weather review*, 136(5):1630–1648.
- Wei, M., Jacobs, G., Rowley, C., Barron, C., Hogan, P., Spence, P., Smedstad, O., Martin, P., Muscarella, P., and Coelho, E. (2013). The impact of initial spread calibration on the refo ensemble and its application to lagrangian dynamics. *Nonlinear Processes in Geophysics*, 20(5):621–641.
- Yang, X. and Delsole, T. (2009). Using the ensemble Kalman filter to estimate multiplicative model parameters. *TellusA*, 61(5):601–609.
- Yang, Y., Robinson, C., Heitz, D., and Mémin, E. (2015). Enhanced ensemble-based 4dvar scheme for data assimilation. *Computers and Fluids*, 115(C):201–210.
- Zhang, F., Snyder, C., and Sun, J. (2004). Impacts of initial estimate and observation availability on convective-scale data assimilation with an ensemble kalman filter. *Monthly Weather Review*, 132(5):1238–1253.
- Zhu, Y. and Navon, I. (1999). Impact of parameter estimation on the performance of the fsu global spectral model using its full-physics adjoint. *Monthly Weather Review*, 127(7):1497–1517.
- Zupanski, D. and Zupanski, M. (2006). Model error estimation employing an ensemble data assimilation approach. *Monthly Weather Review*, 134(5):1337–1354.

## APPENDIX A: Stochastic Reynolds transport theorem

We derive in this appendix the expression of the Reynolds transport theorem for the drift plus noise decomposition: In a Lagrangian stochastic picture, the infinitesimal displacement associated to a trajectory  $X_t$  of a particle is noted:

$$d\mathbf{X}_t = \mathbf{w}(\mathbf{X}_t, t)dt + \boldsymbol{\sigma}(\mathbf{X}_t, t)d\mathbf{B}_t. \quad (\text{A1})$$

The random field involved in this equation is defined for all points of the fluid domain,  $\Omega$  through the kernel  $\check{\sigma}(\cdot, \cdot, t)$  associated to the diffusion operator  $\boldsymbol{\sigma}(\cdot, t)$

$$(\boldsymbol{\sigma}(\mathbf{x}, t)\mathbf{f})^i \triangleq \sum_j \int_{\Omega} \check{\sigma}^{ij}(\mathbf{x}, \mathbf{y}, t) f^j(\mathbf{y}, t) d\mathbf{y}. \quad (\text{A2})$$

This operator is assumed to be of finite norm for any orthonormal basis of the associated Hilbert space  $\mathcal{H}^d$  and to have a null boundary condition on the domain frontier:

$$\boldsymbol{\sigma}(\mathbf{x}, t)\mathbf{f} = 0 \quad \text{on } \partial\Omega, \quad \forall \mathbf{f} \in \mathcal{H}^d. \quad (\text{A3})$$

Function  $\mathbf{B}_t$  denotes a  $d$ -dimensional Brownian function<sup>2</sup> and the resulting  $d$ -dimensional random field,  $\boldsymbol{\sigma}(\mathbf{x}, t)d\mathbf{B}_t \in \mathcal{H}^d$ , is a centered vectorial Gaussian function correlated in space and uncorrelated in time with covariance tensor:

$$Q_{ij}(\mathbf{x}, \mathbf{y}, t, t') = \int_{\Omega} \check{\sigma}^{ik}(\mathbf{x}, \mathbf{y}', t) \check{\sigma}^{jk}(\mathbf{y}', \mathbf{y}, t) d\mathbf{y}' \delta(t-t') dt.$$

We observe that those random fields have a (mean) bounded norm:  $\mathbb{E}\|\boldsymbol{\sigma}d\mathbf{B}_t\|^2 = \text{tr } \mathbf{Q} < \infty$ , where the trace of the covariance tensor is given for any basis  $\{\mathbf{e}_k \mid k \in \mathbb{N}\}$  of  $\mathcal{H}^d$  as  $\text{tr } \mathbf{Q} = \sum_k \langle \mathbf{e}_k, \mathbf{Q}\mathbf{e}_k \rangle_{\mathcal{H}}$ . In the following we will note the diagonal of the covariance tensor as:  $\mathbf{a}(\mathbf{x})dt = \mathbf{Q}(\mathbf{x}, \mathbf{x})$ . Tensor,  $\mathbf{a}$ , that will be referred to as the variance tensor

$$\mathbf{a}(\mathbf{x})\delta(t-t')dt = \mathbb{E} \left( (\boldsymbol{\sigma}(\mathbf{x}, t)d\mathbf{B}_t) (\boldsymbol{\sigma}(\mathbf{x}, t')d\mathbf{B}_{t'})^T \right),$$

is by definition a symmetric positive definite matrix at all spatial points,  $\mathbf{x}$ . This quantity corresponds to the time derivative of the so-called quadratic variation process:

$$\begin{aligned} \mathbf{a}(\mathbf{x})dt &= \int_{\Omega} \check{\sigma}(\mathbf{x}, \mathbf{z})\check{\sigma}^T(\mathbf{x}, \mathbf{z})d\mathbf{z}dt \triangleq \boldsymbol{\sigma}(\mathbf{x})\boldsymbol{\sigma}(\mathbf{y})^T dt \\ &= d \left\langle \int_0^t \boldsymbol{\sigma}(\mathbf{x}, s)d\mathbf{B}_s, \int_0^t \boldsymbol{\sigma}(\mathbf{x}, r)d\mathbf{B}_r \right\rangle. \end{aligned}$$

The notation  $\langle f, g \rangle$  stands for the quadratic cross-variation process of  $f$  and  $g$ . This central object of Stochastic Calculus, can be interpreted as the covariance along time of the increments of  $f$  and  $g$ . The quadratic variation process is briefly presented in Appendix B.

The drift term,  $\mathbf{w}$ , of Lagrangian expression (A1), represents

<sup>2</sup> Formally it is a cylindrical  $I_d$ -Wiener process (see Da Prato and Zabczyk (1992) for more information on infinite dimensional Wiener process and cylindrical  $I_d$ -Wiener process).

the "smooth" – differentiable – part of the flow whereas the random part,

$$\dot{\mathbf{W}} = \boldsymbol{\sigma} \frac{d\mathbf{B}_t}{dt}, \quad (\text{A4})$$

figures the small-scale velocity component associated to a much thinner time-scale. This component is modeled as a delta-correlated random field in time as it represents a highly non regular process at the resolved time scale. In this work, we assume that this small-scale random component is incompressible and therefore associated to a divergence-free diffusion tensor:  $\nabla \cdot \boldsymbol{\sigma} = 0$ . This assumption, which obviously does not prevent the drift component (and therefore the whole field) to be compressible, leads to much simpler modeling and remains realistic for the models considered in this study.

Let us consider now a spatially regular enough scalar function  $\phi$  of compact support, transported by the stochastic flow (A1) and that vanishes outside volume  $\mathcal{V}(t)$  and on its boundary  $\partial\mathcal{V}(t)$ . As this function is assumed to be transported by the stochastic flow, it constitutes a stochastic process defined from its initial time value  $g$ :

$$\phi(\mathbf{X}_t, t) = g(\mathbf{x}_0).$$

We will assume that both functions have bounded spatial gradients. Besides, the initial function  $g : \Omega \rightarrow \mathbb{R}$  vanishes outside the initial volume  $\mathcal{V}(t_0)$  and on its boundary. Let us point out that in this construction, function  $\phi$  cannot be a deterministic function. As a matter of fact, if it was the case, its differential would be given by a standard Ito formula

$$\begin{aligned} \left( \partial_t \phi + \nabla \phi \cdot \mathbf{w} + \frac{1}{2} \sum_{i,j} d\langle X_t^i, X_t^j \rangle \frac{\partial^2 \phi}{\partial x_i \partial x_j} \right) dt + \\ \nabla \phi \cdot \boldsymbol{\sigma} d\hat{\mathbf{B}}_t = 0, \quad (\text{A5}) \end{aligned}$$

which here cancels as  $\phi$  is transported by the flow. A separation between the slow deterministic terms and the fast Brownian term would yield a null uncertainty or a specific uncertainty orthogonal to  $\nabla \phi$ . This would lead us back to the deterministic case, which is not the general goal followed here.

As  $\phi$  is a random function, the differential of  $\phi(\mathbf{X}_t, t)$  involves the composition of two stochastic processes. Its evaluation requires the use of a generalized Ito formula usually referred in the literature to as the Ito-Wentzell formula (see theorem 3.3.1, Kunita, 1990). This extended Ito formula incorporates in the same way as the classical Ito formula a quadratic variation terms related to process  $\mathbf{X}_t$  but also co-variation terms between  $\mathbf{X}_t$  and the gradient of the random function  $\phi_t$ . Its expression is in our case given by

$$\begin{aligned} d\phi(t, \mathbf{X}_t) &= d_t \phi + \nabla \phi \cdot d\mathbf{X}_t + \sum_i d \left\langle \frac{\partial \phi}{\partial x_i}, X_t^i \right\rangle dt + \\ &= \frac{1}{2} \sum_{i,j} d\langle X_t^i, X_t^j \rangle \frac{\partial^2 \phi}{\partial x_i \partial x_j} dt \\ &= 0. \quad (\text{A6}) \end{aligned}$$

It can be immediately checked that for a deterministic function, the standard Ito formula is recovered since the co-variations terms between  $\mathbf{X}_t$  and  $\nabla\phi_t$  cancel.

It follows from A6 that for a fixed grid point, function  $\phi(\mathbf{x}, t)$  is solution of a stochastic differential equation of the form

$$d_t\phi(\mathbf{x}, t) = v(\mathbf{x}, t)dt + \mathbf{f}(\mathbf{x}, t) \cdot d\mathbf{B}_t, \quad (\text{A7})$$

where the Brownian term must compensate the Brownian term of (A6). The quadratic variation term involved in (A6) is given through (appendix B) as

$$d\langle X_t^i, X_t^j \rangle = a^{ij}(\mathbf{x}, t) \quad (\text{A8})$$

$$\begin{aligned} &= \sum_k \sigma^{ik}(\mathbf{x}, t) \sigma^{kj}(\mathbf{x}, t) \\ &= \sum_k \int_{\Omega} \check{\sigma}^{ik}(\mathbf{x}, \mathbf{y}, t) \check{\sigma}^{kj}(\mathbf{y}, \mathbf{x}, t), \quad (\text{A9}) \end{aligned}$$

and the covariation term reads

$$d\left\langle \frac{\partial\phi_t}{\partial x_i}, X_t^i \right\rangle = \sum_j \int_{\Omega} \frac{\partial \check{f}^{jj}}{\partial x_i}(\mathbf{x}, \mathbf{y}, t) \check{\sigma}^{ij}(\mathbf{y}, \mathbf{x}, t). \quad (\text{A10})$$

In these expressions  $\check{f}$  (resp.  $\check{\sigma}$ ) designates the kernel associated to operator  $\mathbf{f}$  (resp.  $\sigma$ ). Now, identifying first the Brownian term and next the deterministic terms of equations (A6) and (A7), we infer

$$\begin{aligned} \check{f}(\mathbf{x}, \mathbf{y}, t)^T &= -\nabla\phi(\mathbf{x}, t)^T \check{\sigma}(\mathbf{x}, \mathbf{y}, t), \\ v(\mathbf{x}, t) &= -\nabla\phi \cdot \mathbf{w} + \sum_{i,j} \frac{1}{2} a_{ij} \frac{\partial^2 \phi}{\partial x_i \partial x_j} + \nabla\phi \cdot \frac{\partial \sigma^{\bullet j}}{\partial x_i} \sigma^{ij}, \\ &= -\nabla\phi \cdot \mathbf{w} + \nabla\phi \cdot (\nabla \cdot \mathbf{a})^T + \frac{1}{2} \sum_{i,j} a_{ij} \frac{\partial^2 \phi}{\partial x_i \partial x_j}, \end{aligned}$$

and finally get

$$d_t\phi = \mathcal{L}\phi dt - \nabla\phi \cdot \sigma d\mathbf{B}_t, \quad (\text{A11})$$

$$\mathcal{L}\phi = -\nabla\phi \cdot (\mathbf{w} - \frac{1}{2} (\nabla \cdot \mathbf{a})^T) + \frac{1}{2} \nabla \cdot (\mathbf{a} \nabla \phi). \quad (\text{A12})$$

This differential at a fixed point,  $\mathbf{x}$ , defines the equivalent in the deterministic case of the material derivative of a function transported by the flow. The differential of the integral over a material volume of the product  $q\phi$  is given by

$$\begin{aligned} d \int_{\mathcal{V}(t)} q\phi(\mathbf{X}_t, t) d\mathbf{x} &= d \int_{\Omega} q\phi d\mathbf{x}, \\ &= \int_{\Omega} (d_t q\phi + q d_t\phi + d_t(q, \phi)) d\mathbf{x}, \end{aligned}$$

where the first line comes from  $\phi(t, \mathbf{x}) = 0$  if  $\mathbf{x} \in \Omega \setminus \mathcal{V}(t)$  and the second one from the integration by part formula of two Ito processes. Hence, from (A11) this differential is

$$\int_{\Omega} (d_t q\phi + q\mathcal{L}\phi + \nabla\phi \cdot \mathbf{a} \nabla q) dt d\mathbf{x} - \int_{\Omega} q \nabla\phi \cdot \sigma d\mathbf{B}_t.$$

Introducing  $\mathcal{L}^*$  the (formal) adjoint of the operator  $\mathcal{L}$  in the space  $L^2(\Omega)$  with Dirichlet boundary conditions, this expression can be written as

$$\int_{\Omega} \left( (d_t q + \mathcal{L}^* q - \nabla \cdot (\mathbf{a} \nabla q)) dt + \nabla \cdot (q \sigma d\mathbf{B}_t) \right) \phi d\mathbf{x}.$$

With the complete expression of  $\mathcal{L}^*$  (remarking that the second right-hand term of A12 is self-adjoint) and if  $\phi(\mathbf{x}, t) \rightarrow \mathbb{1}_{\mathcal{V}(t)/\partial\mathcal{V}(t)}$ , where  $\mathbb{1}$  stands for the characteristic function, we get the sought form of this differential:

$$\int_{\mathcal{V}(t)} \left[ d_t q + \left( \nabla \cdot (q\mathbf{w} - \frac{1}{2} \nabla \cdot (q\mathbf{a})^T) \right) dt + \nabla q \sigma d\mathbf{B}_t \right] d\mathbf{x},$$

## APPENDIX B: Quadratic variation and covariation

We recall here the notion of quadratic variation and co-variation, which play a central role in stochastic calculus. We will here restrict ourselves to standard Ito processes. Quadratic variation and co-variation correspond respectively to the variance and covariance of the process increments along time. The quadratic co-variation process denoted as  $\langle \mathbf{X}, \mathbf{Y} \rangle_t$ , (respectively the quadratic variation for  $\mathbf{Y} = \mathbf{X}$ ) is defined as the limit in probability over a partition  $\{t_1, \dots, t_n\}$  of  $[0, t]$  with  $t_1 < t_2 < \dots < t_n$ , and a partition spacing  $\delta t_i = t_i - t_{i-1}$ , noted as  $|\delta t|_n = \max_i \delta t_i$  and such that  $|\delta t|_n \rightarrow 0$  when  $n \rightarrow \infty$ :

$$\langle \mathbf{X}, \mathbf{Y} \rangle_t = \lim_{|\delta t|_n \rightarrow 0} \sum_{i=0}^{n-1} (\mathbf{X}(t_{i+1}) - \mathbf{X}(t_i)) (\mathbf{Y}(t_{i+1}) - \mathbf{Y}(t_i))^T.$$

For Brownian motion these covariations can be easily computed and are given by the following rules  $\langle B, B \rangle_t = t$ ,  $\langle B, h \rangle_t = \langle h, B \rangle_t = \langle h, h \rangle_t = 0$ , where  $h$  is a deterministic function and  $B$  a scalar Brownian motion.