



**2-D geophysical flows with a VEnKF**

Z. Mussa et al.

This discussion paper is/has been under review for the journal Nonlinear Processes in Geophysics (NPG). Please refer to the corresponding final paper in NPG if available.

# Data assimilation of two-dimensional geophysical flows with a Variational Ensemble Kalman Filter

**Z. Mussa, I. Amour, A. Bibov, and T. Kauranne**

Lappeenranta University of Technology, Lappeenranta, Finland

Received: 30 January 2014 – Accepted: 10 March 2014 – Published: 31 March 2014

Correspondence to: Z. Mussa (zubeda.mussa@lut.fi)

Published by Copernicus Publications on behalf of the European Geosciences Union & American Geophysical Union.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures



Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



## Abstract

The Variational Ensemble Kalman Filter (VEnKF), a recent data assimilation method that combines a variational assimilation of the Bayesian estimation problem with an ensemble of forecasts, is demonstrated in two-dimensional geophysical flows using a Quasi-Geostrophic (QG) model and a shallow water model. Using a synthetic experiment, a two layer QG model with model bias is solved on a cylindrical  $40 \times 20$  domain. The performance of VEnKF on the QG model with increasing ensemble size is compared with the classical Extended Kalman Filter (EKF). It is shown that although convergence can be achieved with just 20 ensemble members, increasing the number of members results in a better estimate that approaches the one produced by EKF.

In the second test case, a 2-D shallow water model is described using a real dam-break experiment. The VEnKF algorithm was used to assimilate observations obtained from a modified laboratory dam-break experiment with a two-dimensional setup of sensors at the downstream end. The wave meters are placed parallel to the direction of the flow alongside the flume walls to capture both cross flow and stream flow.

In both test cases, VEnKF was able to predict genuinely two-dimensional flow patterns when the sensors had a two-dimensional geometry and was stable against model bias in the first test case.

In the second test case, the experiments are complemented with an empirical study of the impact of observation interpolation on the stability of the VEnKF filter. In this study, a novel Courant–Friedrichs–Lewy type filter stability condition is observed that relates ensemble variance to the time interpolation distance between observations.

The results of the two experiments shows that VEnKF is a good candidate for data assimilation problems and can be implemented in higher dimensional nonlinear models.

**NPGD**

1, 403–446, 2014

## 2-D geophysical flows with a VEnKF

Z. Mussa et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

⏪

⏩

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



# 1 Introduction

The increasing power of computers and various automatic measurement devices has created a new domain for algorithmic research that develops methods for producing optimal state estimates of a system to be studied that combine a computer model of corresponding system dynamics with real – often indeed real-time – measurements. The methods suggested for this kind of studies go by a variety of names, such as methods for parameter estimation, data assimilation methods, Markov chain Monte Carlo methods, particle filters and so on. Most of these methods are statistical by their construction and often fit the paradigm of Bayesian estimation theory, where the computer model produces a prior estimate that is then complemented to the posterior estimate by some algorithm based on the Bayes formula.

In the current article we shall focus on the problem of state estimation of two-dimensional systems in geophysics, more precisely atmospheric and hydrological models, with a recent ensemble data assimilation method that combines a variational formulation of the Bayesian estimation problem with an ensemble of forecasts, namely the Variational Ensemble Kalman Filter (VENKF) introduced in Solonen et al. (2012). The current section reviews related previous research. Section two presents the variational ensemble Kalman filter with related earlier Kalman filter algorithms. The third section discusses both example cases, namely a synthetic two-dimensional atmospheric flow and a real two-dimensional hydraulic river flow, and presents the corresponding assimilation results. Section four concludes the paper with a brief discussion of the results.

## 1.1 Data assimilation in geophysical and atmospheric sciences

In recent years, data assimilation has become common in the field of geophysics and atmospheric sciences (van Leeuwen, 2011; Blum et al., 2009; Wu et al., 2008). Some of the challenges in estimating geophysical systems are the nonlinearity of the corresponding system dynamics (Miller et al., 1994; Blum et al., 2009; van Leeuwen,

## 2-D geophysical flows with a VENKF

Z. Mussa et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures



Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



## 2-D geophysical flows with a VEnKF

Z. Mussa et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

◀

▶

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



2011); sensitivity to initial conditions; and estimating the error covariance matrix for high-dimensional state vectors (Kuznetsov et al., 2003; Hamill et al., 2001). Various data assimilation methods have been applied in some fields of geophysics, such as meteorology and oceanography. These methods belong to two main groups, namely sequential assimilation methods (e.g. nudging, Particle Filter methods and Kalman filter methods) and lumped and variational assimilation methods (e.g. Optimal Interpolation, three-dimensional and four-dimensional variational data assimilation).

Optimal interpolation has long been used in numerical weather prediction (NWP) (Daley, 1993; Bertino et al., 2003). However, the use of optimal interpolation in hydrological models with complicated flow patterns introduces instabilities (Heemink and Metzelaar, 1995). Kalman filtering is an optimal sequential data assimilation method. However, it can only be applied to linear models with a Gaussian assumption of model and observation error covariances (see e.g., Dee, 1991; Heemink and Metzelaar, 1995; Auvinen et al., 2010; Bertino et al., 2003; Bardsley et al., 2013; McMillan et al., 2013). Incorporating the Kalman filter with repeated linearizations of a nonlinear dynamical systems leads to the Extended Kalman Filter (EKF) that can be used for nonlinear models. EKF has the advantage of being robust and it gives accurate estimates of the state. However, it is computationally expensive for large scale problems.

Particle Filters have also been used in geophysical flows e.g. in NWP (van Leeuwen, 2011; van Leeuwen and Ades, 2013). Particle filtering uses Sequential Monte Carlo techniques whereby the state is represented in the form of a posterior probability density function conditioned on the measurements available at that time. Particle filters do not require the Gaussian assumption on the error terms. van Leeuwen (2011) has suggested that the probability density function of highly nonlinear models can be represented by a number of points or particles so that each particle represent a full model state.

The Ensemble Kalman Filter (EnKF) (Evensen, 1994) has also been used in NWP and in hydrological models. EnKF uses a Monte Carlo approach such that the error

## 2-D geophysical flows with a VEnKF

Z. Mussa et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

◀

▶

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



covariance matrices are replaced by the corresponding sample covariance matrices calculated from an ensemble and the ensemble of states is propagated in time using the fully nonlinear model (Evensen, 1994; Reichle et al., 2002; Bertino et al., 2003; Hoteit et al., 2007; McMillan et al., 2013). Variational methods such as 3DVAR and 4DVAR are also commonly used. However, their use is limited by the need of a tangent linear and an adjoint model for the evaluation of the gradient of the cost function which leads to a high computational cost. These methods solve the underlying maximum a posterior optimization problem that measures the model to data misfit (Bertino et al., 2003). Navon (2009) gives a review of these methods in numerical weather prediction applications. Other studies in NWP include that of Fisher et al. (2009).

### 1.2 Data assimilation in hydrological and coastal models

The use of data assimilation in hydrological studies has been a great challenge and has not been applied very often. In many numerical simulations based on hydrological flows, several researchers have developed numerical models based on different numerical schemes, such as finite difference, finite volume and finite element methods. In different numerical simulations, comparisons have been made between simulation results and experimental results (Liang and Marche, 2009; Chang et al., 2011; Tseng and Chu, 2011; Zhou et al., 2004). Liu et al. (2012) review the challenges and opportunities on the use of effective data assimilation in operational hydrological forecasting. They point out that before the adoption of data assimilation techniques in hydrological modeling, the challenges which need to be taken care of include the high non-linearity of hydrological processes, high dimensionality of the state vector and the need to use large samples when using ensemble methods. As it was pointed out by Heemink and Metzelaar (1995) there is a need of numerical hydrological models to incorporate water level measurements using data assimilation technique so as to improve forecast.

Reichle et al. (2002) has shown in a study of hydrologic data assimilation that although EnKF underestimates the forecast error covariance when 100 ensemble

**2-D geophysical flows with a VEnKF**

Z. Mussa et al.

Title Page	
Abstract	Introduction
Conclusions	References
Tables	Figures
◀	▶
◀	▶
Back	Close
Full Screen / Esc	
Printer-friendly Version	
Interactive Discussion	



members were used, it can well estimate the error covariance with 500 ensemble members so that the soil moisture can be estimated as a key variable in hydrologic and atmospheric models. Xu and Daley (2002) used data assimilation to test a representer algorithm on an unstable 2-D shallow water system. The representer algorithm minimizes a cost function of a 4DVAR assimilation scheme that measures the fit of the analysis to the observations, to the forecast model and to the initial state estimate (Xu and Daley, 2002). The accelerated representer algorithm was found to be computationally more efficient than the cyclic representer algorithm. However, the cyclic representer algorithm produced a more accurate analysis than the accelerated representer algorithm.

Tinka et al. (2009) presented a Quadratic Programming (QP) method for assimilating Lagrangian sensor measurement data into a shallow water equation model. Strub et al. (2009) compared two data assimilation algorithms applied to river flow, the QP as described by Tinka et al. (2009) and the EnKF. The performance of the two methods was compared using a twin experiment and their results concluded that the QP based algorithms is of low computational cost and presents a better balance between cost and accuracy than EnKF. Bélanger and Vincent (2004) used a 4-Dimensional Variational data Assimilation (4DVAR) to forecast floods. The 4DVAR method was applied to shallow water equations and used a steepest descent minimization technique to minimize the cost function which for their case is the mean deviation between the model solution and the available measurements. Though 4DVAR is computationally expensive in high dimensional problems and it needs an adjoint model to find the gradient required in 4DVAR, the method was suggested as a good one for improving hydrological models. In dealing with boundary conditions, Kazantsev (2012) applied a 4-D variational data assimilation technique to a non-linear shallow water model so as to control the discretization of the derivative and interpolation error near the boundary, with the advantage of making the solution closer to observations after assimilation.

Another approach in data assimilation which has been used in hydrological flow was introduced by Lei et al. (2012). The method is known as Hybrid Nudging Ensemble

**2-D geophysical flows with a VEnKF**

Z. Mussa et al.

Title Page

Abstract Introduction

Conclusions References

Tables Figures

⏪ ⏩

◀ ▶

Back Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



Kalman Filter (HNEKF) and it combines EnKF and Observation Nudging with the aim of achieving more gradual and continuous data assimilation. When applied to a shallow water model, the method was found to have a smaller RMS error than nudging and EnKF alone. However, it costs quite a lot of CPU time, as does EnKF, since they both require an ensemble of forecasts whereas nudging needs much less CPU time.

The current paper extends an earlier study on a dam-break experiment (Amour et al., 2013), in which we introduced data assimilation of wave meter data into a river model that has originally been presented by Martin and Gorelick (2005). In that paper we demonstrated the capability of a recently developed data assimilation method, the variational ensemble Kalman filter by Solonen et al. (2012), to produce better results than pure simulation when applied to a hydrological model. It was shown earlier by Solonen et al. (2012) that the problems encountered in using the Kalman filter are hereby solved by the use of low memory estimations of state and observation covariance matrices. These are obtained using the LBFGS optimization method by using the low-rank Hessian matrix estimates it produces to replace the full error covariance matrices in EKF.

In the previous work, only a one-dimensional set of observations was available. In such a situation, the analysis of the flume is limited to one-dimensional corrections to the simulated flow. In this case, data assimilation produces a series of crests and troughs in the flow direction only, and no cross-flow is introduced to the analysis. In the current paper, we apply the VEnKF to modified hydrological flume data of a dam-break experiment conducted in a laboratory by Bellos et al. (1991), that simulates the presence of a two-dimensional set of observations with a known cross-flow pattern. We also apply VEnKF to simulated two-dimensional observations of a barotropic vorticity equation.

## 2 Assimilation Methods

In this section we discuss the data assimilation methods that are used in our experiments. The implementation of the Saint-Venant shallow water model, which is exploited as the second example case in the present study, does not provide linearization (e.g. tangent-linear and adjoint) codes. Therefore, we are limited to use of ensemble methods, which can be executed directly on top of a nonlinear model. In this paper we only consider sequential data-assimilation performed using VEnKF that provides a good low-memory approximation of the EKF and does not require linearization codes for transition operator.

### 2.1 Ensemble Kalman Filter

We begin by reviewing the well known Ensemble Kalman filter (EnKF) that is derived via straightforward sampling approximation of the EKF formulas. First, we consider the following coupled system of stochastic equations:

$$\mathbf{x}_{k+1} = \mathcal{M}_k(\mathbf{x}_k) + \boldsymbol{\varepsilon}_k, \quad (1)$$

$$\mathbf{y}_{k+1} = \mathcal{H}_{k+1}(\mathbf{x}_{k+1}) + \boldsymbol{\eta}_{k+1}. \quad (2)$$

Here  $\mathbf{x}_k$  denotes  $N$ -dimensional model state vector, which fully describes the phenomena being studied at a given time instant  $k$ ,  $\mathcal{M}_k$  is transition operator that propagates model state vector from time instant  $k$  to time instant  $k + 1$ ,  $\mathbf{y}_{k+1}$  is  $M$ -dimensional vector of observations collected at time instant  $k + 1$ ,  $\mathcal{H}_{k+1}$  is observation operator which maps the model state vector space to the observation vector space. Finally, the terms  $\boldsymbol{\varepsilon}_k$  and  $\boldsymbol{\eta}_{k+1}$  are random vectors with known covariances  $\mathbf{C}_{\boldsymbol{\varepsilon}_k}$  and  $\mathbf{C}_{\boldsymbol{\eta}_{k+1}}$  respectively. The role of stochastic addenda is to merely represent prediction error of transition operator  $\mathcal{M}_k$  in Eq. (1) and to account for the measurement noise in observation model (2). Procedure of data assimilation provides a statistically consistent way of estimating the model state vector at time instant  $k + 1$  given the corresponding observation  $\mathbf{y}_{k+1}$  and the state estimate  $\mathbf{x}_k^{\text{est}}$  with covariance  $\mathbf{C}_k^{\text{est}}$  at preceding time

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

◀

▶

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



instant  $k$ . Transition operator  $\mathcal{M}_k$  and observation operator  $\mathcal{H}_k$  as well as covariances  $\mathbf{C}_{\varepsilon_k}$  and  $\mathbf{C}_{\eta_{k+1}}$  are also assumed to be known. Often, the stochastic terms  $\varepsilon_k$  and  $\eta_{k+1}$  are supposed to be normally distributed, although this assumption may be relaxed.

Prior to the discussion of the ensemble approximation, we present the basic EKF, which under certain conditions derives in some sense statistically optimal estimate for state vector  $\mathbf{x}_{k+1}$  (see Simon, 2006). The algorithmic formulation of the EKF reads as follows:

i. Prediction step.

a. Compute prediction:  $\mathbf{x}_{k+1}^p = \mathcal{M}_k \left( \mathbf{x}_k^{\text{est}} \right)$ .

b. Propagate estimate covariance:  $\mathbf{C}_{k+1}^p = \mathbf{M}_k^{\text{TL}} \mathbf{C}_k^{\text{est}} \mathbf{M}_k^{\text{AD}} + \mathbf{C}_{\varepsilon_k}$ .

ii. Correction step.

a. Compute Kalman gain:  $\mathbf{G}_{k+1} = \mathbf{C}_{k+1}^p \mathbf{H}_{k+1}^{\text{AD}} \left( \mathbf{H}_{k+1}^{\text{TL}} \mathbf{C}_{k+1}^p \mathbf{H}_{k+1}^{\text{AD}} + \mathbf{C}_{\eta_{k+1}}^p \right)^{-1}$ .

b. Compute the state estimate:  $\mathbf{x}_{k+1}^{\text{est}} = \mathbf{x}_{k+1}^p + \mathbf{G}_{k+1} \left( \mathbf{y}_{k+1} - \mathbf{H}_{k+1}^{\text{TL}} \left( \mathbf{x}_{k+1}^p \right) \right)$ .

c. Compute the covariance estimate:  $\mathbf{C}_{k+1}^{\text{est}} = \mathbf{C}_{k+1}^p - \mathbf{G}_{k+1} \mathbf{H}_{k+1}^{\text{TL}} \mathbf{C}_{k+1}^p$ .

Here  $\mathbf{H}_{k+1}^{\text{TL}}$ ,  $\mathbf{M}_k^{\text{TL}}$  denote tangent-linear codes of the operators  $\mathcal{H}_{k+1}$  and  $\mathcal{M}_k$  evaluated at  $\mathbf{x}_k^{\text{est}}$  and  $\mathbf{x}_{k+1}^p$ , respectively,  $\mathbf{H}_{k+1}^{\text{AD}}$  and  $\mathbf{M}_k^{\text{AD}}$  are the corresponding adjoint codes. As was previously pointed out, the presented algorithm derives an optimal estimate for  $\mathbf{x}_{k+1}$ . However, it requires explicit matrix storage to operate covariances  $\mathbf{C}_k^{\text{est}}$  and  $\mathbf{C}_{k+1}^p$  (covariances  $\mathbf{C}_{\varepsilon_k}$  and  $\mathbf{C}_{\eta_{k+1}}$  are often assumed being diagonal or implemented by low-memory sub-routines). Therefore, in large-scale cases such matrix storage becomes infeasible and the task of data assimilation requires a special treatment.

One of the obvious ways to overcome the emitted problem is to replace the problematic parts of the EKF with corresponding ensemble approximations. This is

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

◀

▶

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



the main idea behind the EnKF (see Evensen, 2003). Let us consider a bunch of  $N$ -dimensional random vectors  $\mathbf{s}_{k,i} \sim \mathcal{N}(\mathbf{x}_k^{\text{est}}, \mathbf{C}_k^{\text{est}})$ , where  $k \in \mathbb{N}$ ,  $i = 1, \dots, S$ , and  $S$  is the ensemble cardinality. Consider an  $N$ -by- $S$  matrix  $\mathbf{X}_k$  depending on  $\mathbf{s}_{k,i}$ , which is defined by the following:

$$\mathbf{X}_k = ((\mathbf{s}_{k,1} - \bar{\mathbf{s}}_k), \dots, (\mathbf{s}_{k,S} - \bar{\mathbf{s}}_k)) / \sqrt{S-1}. \quad (3)$$

Here  $\bar{\mathbf{s}}_k$  denotes the mean of ensemble  $\mathbf{s}_{k,i}$ . A single EnKF data assimilation step defines procedure of propagating  $\mathbf{s}_{k,i}$  to  $\mathbf{s}_{(k+1),i}$ .

i. Prediction step.

- a. Move the state estimate and ensemble forward over the time:  $\mathbf{s}_{(k+1),i}^p = \mathcal{M}_k(\mathbf{s}_{k,i}) + \mathbf{e}_{k,i}^p, i = 1, \dots, S$ .
- b. Calculate prediction covariance matrix using ensemble obtained on the previous step:  $\mathbf{C}_{k+1}^p = \mathbf{X}_k \mathbf{X}_k^T$ .

ii. Correction step.

- a. Compute Kalman gain:  $\mathbf{G}_{k+1} = \mathbf{C}_{k+1}^p \mathbf{H}_{k+1}^{\text{AD}} \left( \mathbf{H}_{k+1}^{\text{TL}} \mathbf{C}_{k+1}^p \mathbf{H}_{k+1}^{\text{AD}} + \mathbf{C}_{\eta_{k+1}}^p \right)^{-1}$ .
- b. Update ensemble members:

$$\mathbf{s}_{(k+1),i} = \mathbf{s}_{(k+1),i}^p + \mathbf{G}_{k+1} \left( \mathbf{y}_{k+1} - \mathbf{H}_{k+1}^{\text{TL}} \mathbf{s}_{(k+1),i}^p + \mathbf{n}_{(k+1),i} \right).$$

- c. Calculate the next state estimate as the sample mean of ensemble  $\mathbf{s}_{(k+1),i}$ :  $\mathbf{x}_{k+1}^{\text{est}} = \bar{\mathbf{s}}_{(k+1),i}$ .

Here vectors  $\mathbf{e}_{k,i}^p$  and  $\mathbf{n}_{(k+1),i}$  are realizations of random terms  $\varepsilon_k$  and  $\eta_{k+1}$  respectively. Obviously, the EnKF algorithm closely mimics the steps of the EKF, while the

**2-D geophysical flows with a VEnKF**

Z. Mussa et al.

Title Page	
Abstract	Introduction
Conclusions	References
Tables	Figures
◀	▶
◀	▶
Back	Close
Full Screen / Esc	
Printer-friendly Version	
Interactive Discussion	



impracticable matrix storage is avoided since the cardinality of the ensemble  $S$  is usually significantly smaller compared to the problem dimension.

The ensemble Kalman filter can be implemented directly on top of a nonlinear model as it does not require either tangent-linear or adjoint codes. However, the algorithm has numerous disadvantages such as propagation ensemble degradation as pointed out by Houtekamer and Mitchell (1998) and Zupanski (2005). In the next section we discuss an approach that attempts to overcome the drawbacks of the EnKF by introducing the ensemble approximations into the variational formulation of the EKF suggested by Auvinen et al. (2010).

## 2.2 Variational Ensemble Kalman Filter

In this section we describe another approximation of the EKF called the Variational Ensemble Kalman filter. We begin by considering the following quadratic function:

$$J(\mathbf{x}|\mathbf{y}_{k+1}) = \frac{1}{2} \left( \mathbf{x} - \mathbf{x}_{k+1}^p \right)^T \left( \mathbf{C}_{k+1}^p \right)^{-1} \left( \mathbf{x} - \mathbf{x}_{k+1}^p \right) + \frac{1}{2} \left( \mathbf{y}_{k+1} - \mathbf{H}_{k+1}^{\text{TL}} \mathbf{x} \right)^T \left( \mathbf{C}_{k+1} \right)^{-1} \left( \mathbf{y}_{k+1} - \mathbf{H}_{k+1}^{\text{TL}} \mathbf{x} \right). \quad (4)$$

It can be proved (see Simon, 2006) that the minimizer of this function equals to the model state vector estimate  $\mathbf{x}_{k+1}^{\text{est}}$  computed by EKF. Furthermore, the inverse Hessian of  $J(\mathbf{x}|\mathbf{y}_{k+1})$  equals to covariance  $\mathbf{C}_{k+1}^{\text{est}}$  as defined by the EKF. Hence, EKF is equivalent to generalized least squares problem (4). The idea to replace EKF algebraic formulas by equivalent optimization task is referred to as the variational formulation of the extended Kalman filter. However, it inherits the memory-related pitfalls of the EKF since minimization of Eq. (4) requires inversion of  $\mathbf{C}_{k+1}^p$  as well as storage of  $\left( \mathbf{C}_{k+1}^p \right)^{-1}$  and the inverse Hessian of Eq. (4). These issues can be circumvented by leveraging sampled approximations as demonstrated in the previous section and by

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

◀

▶

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



plugging them into the variational formulation (4). This is the basic idea behind the variational ensemble Kalman filter.

In order to formulate the VENKF we again begin by considering a bundle of  $N$ -dimensional random vectors  $\mathbf{s}_{k,i} \sim \mathcal{N}(\mathbf{x}_k^{\text{est}}, \mathbf{C}_k^{\text{est}})$  (here we assume that model state vector as well as its covariance estimated at time instance  $k$  are known). Therefore, the prediction step now can be formulated as follows:

$$\mathbf{x}_{k+1}^p = \mathcal{M}_k(\mathbf{x}_k^{\text{est}}), \quad (5)$$

$$\mathbf{s}_{(k+1),i}^p = \mathcal{M}_k(\mathbf{s}_{k,i}). \quad (6)$$

Next, we redefine matrix  $\mathbf{X}_k$  from the previous section as now the mean of ensemble  $\mathbf{s}_{(k+1),i}$  is considered known and equal to predicted state  $\mathbf{x}_{k+1}^p$ :

$$\mathbf{X}_k = \left( (\mathbf{s}_{k,1} - \mathbf{x}_k^p), \dots, (\mathbf{s}_{k,S} - \mathbf{x}_k^p) \right) / \sqrt{S}, \quad (7)$$

where  $S$  as previously denotes the cardinality of ensemble  $\mathbf{s}_{k,i}$ . Hence, the sampled approximation for the prior covariance can be defined by leveraging prior ensemble  $\mathbf{s}_{(k+1),i}^p$  computed on prediction step leading to the following:

$$\mathbf{C}_{k+1}^p = \mathbf{X}_k \mathbf{X}_k^T + \mathbf{C}_{\varepsilon_k}. \quad (8)$$

This sampled approximation allows to programmatically implement the prior covariance  $\mathbf{C}_{k+1}^p$  as a low-memory subroutine since following Eq. (8), the computation of a matrix-vector product would only require storage of  $\mathbf{X}_k$  (as before, it is assumed that  $\mathbf{C}_{\varepsilon_k}$  is diagonal or implemented as a low-memory subroutine). Nevertheless, minimization of Eq. (4) makes use of  $[\mathbf{C}_{k+1}^p]^{-1}$ , which can be obtained by applying the Sherman-Morrison-Woodbury (SMW) matrix identity:

$$[\mathbf{C}_{k+1}^p]^{-1} = \mathbf{C}_{\varepsilon_k}^{-1} - \mathbf{C}_{\varepsilon_k}^{-1} \mathbf{X}_k \left( \mathbf{I} + \mathbf{X}_k^T \mathbf{C}_{\varepsilon_k}^{-1} \mathbf{X}_k \right)^{-1} \mathbf{X}_k^T \mathbf{C}_{\varepsilon_k}^{-1}. \quad (9)$$

**2-D geophysical flows with a VENKF**

Z. Mussa et al.

Title Page	
Abstract	Introduction
Conclusions	References
Tables	Figures
◀	▶
◀	▶
Back	Close
Full Screen / Esc	
Printer-friendly Version	
Interactive Discussion	



Here, it is assumed that covariance  $\mathbf{C}_{\varepsilon_k}$  can be easily inverted due to its simple structure. Moreover, since  $\mathbf{I} + \mathbf{X}_k^T \mathbf{C}_{\varepsilon_k}^{-1} \mathbf{X}_k$  is an  $S$ -by- $S$  matrix and the ensemble size  $S$  is usually much smaller compared to the problem dimension  $N$ , the inversions in EQ. (9) are considered feasible.

Minimization of Eq. (4) is done by the L-BFGS unconstrained optimizer described in Nocedal and Wright (1999). The L-BFGS is a Quasi-Newton method, which uses the history of its iterations in order to approximate the inverse Hessian of the target cost function. Furthermore, the L-BFGS usually converges to the optimal point having a qualified inverse Hessian approximation in much smaller amount of iterations than the dimension of the problem. These characteristics of the method can be leveraged to minimize Eq. (4) as well as to compute its inverse Hessian, wherein both tasks are completed in single pass. The same idea may be used instead of SMW matrix identity to obtain  $[\mathbf{C}_{k+1}^p]^{-1}$  (see Solonen et al., 2012). However, the L-BFGS only provides an approximation for the inverse Hessian of the target cost function, so formula (9) is suggested as the one preferable to use.

Finally, putting together Eqs. (5), (7), (8), (9) and the argumentation concerning the L-BFGS we can formulate the VEnKF algorithm:

i. Prediction step.

- a. Compute prior model state and propagate the ensemble members as defined in Eq. (5).
- b. Define the approximative prior covariance operator  $\mathbf{C}_{k+1}^p$  in accordance with Eq. (8).
- c. Apply SMW matrix identity or L-BFGS in order to define low-memory operator representation of the inverse prior covariance  $[\mathbf{C}_{k+1}^p]^{-1}$ .

2-D geophysical flows with a VEnKF

Z. Mussa et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

⏪

⏩

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



ii. Correction step.

a. Apply L-BFGS to minimize Eq. (4). Assign  $\mathbf{x}_{k+1}^{\text{est}}$  to the minimizing point and  $\mathbf{C}_{k+1}^{\text{est}}$  to the L-BFGS approximation of inverse Hessian of the cost function, (Eq. 4)

b. Generate new ensemble  $\mathbf{s}_{(k+1),i} \sim \mathcal{N}(\mathbf{x}_{k+1}^{\text{est}}, \mathbf{C}_{k+1}^{\text{est}})$ .

The attractive feature in the presented algorithm is that the operating ensemble is regenerated at every assimilation round, which allows to avoid the ensemble inbreeding inherent to EnKF. It should be mentioned, that the given formulation of the VEnKF algorithm does not reveal efficient implementation, but only explains the main steps of the procedure. Further details can be found in the paper by Solonen et al. (2012).

### 3 Test cases with 2-D flows

#### 3.1 A synthetic atmospheric flow with the barotropic vorticity equation

In this section we present a synthetic case of the two-layer Quasi-Geostrophic model (QG-model) introduced in Pedlosky (1987). This model provides an example of chaotic dynamics which can be run at a large-scale setting with reasonable computational cost. In the present study the QG-model is coupled with VEnKF in an artificial data assimilation experiment, which is devoted to empirically demonstrate the “correctness of the concept” and illustrate the advantages of the VEnKF over EnKF in case of large-scale data assimilation.

The QG-model simulates flat double-layered geostrophic (slow) wind motion. Geometrically, the model resides on a cylindrical surface divided into two interacting “atmospheric” layers located one above the other. The lower layer is affected by an orography component, which models surface irregularities of the bottom. Due to the

## 2-D geophysical flows with a VEnKF

Z. Mussa et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

◀

▶

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



## 2-D geophysical flows with a VEnKF

Z. Mussa et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

◀

▶

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



cylindrical domain, periodical boundary conditions are employed along the longitudinal direction. Figure 1 provides an illustration of the geometrical layout. In the figure variables  $\bar{U}_1$  and  $\bar{U}_2$  denote mean zonal wind speeds in the upper and lower layers, respectively. These values are used to define constant boundary conditions at the top and at the bottom of the cylinder. Hereafter in this section the subindex  $i = 1, 2$  is used as reference to the top and the bottom model layers.

Operating components of the QG-model are potential vorticity  $q_i$  and stream function (analog of pressure)  $\psi_i$ . The relation between these components is described by the following equations:

$$q_1 = \nabla^2 \psi_1 - F_1 (\psi_1 - \psi_2) + \beta y, \quad (10)$$

$$q_2 = \nabla^2 \psi_2 - F_2 (\psi_2 - \psi_1) + \beta y + R_s, \quad (11)$$

where  $F_i$  are layer interaction parameters,  $\beta$  is northward gradient of the Coriolis parameter, and  $R_s$  denotes two-dimensional orography surface. It can be shown (see Fandry and Leslie, 1984) that the coupled system of Eqs. (10)–(11) is invertible. Hence, there is a one-to-one correspondence between potential vorticity and stream function, wherein the latter relates to the zonal wind  $u_i$  and meridional wind  $v_i$  by the following dependency:

$$u_i = -\frac{\partial \psi_i}{\partial y}, \quad v_i = \frac{\partial \psi_i}{\partial x}. \quad (12)$$

Therefore, stream function  $\psi_i$  can be thought of as potential of the two-dimensional field  $(v_i, -u_i)$ .

Finally, it is assumed that the QG-model obeys the following conservation law:

$$\frac{D_i q_i}{Dt} = 0. \quad (13)$$

Here operator  $\frac{D_i}{Dt} = \frac{\partial}{\partial t} + u_i \frac{\partial}{\partial x} + v_i \frac{\partial}{\partial y}$  denotes substantial derivative. Equations (10), (11), (12), and (13) define the governing PDE system of the QG-model. The given

system is described in non-physical terms specified via the standard procedure of non-dimensionalization given below:

$$F_i = \frac{f_0^2 L^2}{\dot{g} D_i},$$

$$\dot{g} = g \frac{\Delta\theta}{\bar{\theta}},$$

$$R_s = \frac{S(x, y)}{\eta D_2},$$

$$\beta = \beta_0 \frac{L}{U},$$

where  $D_i$  is the depth of the corresponding model layer,  $g$  is acceleration of gravity,  $\Delta\theta$  is the potential temperature change across the layer interface,  $\bar{\theta}$  is the mean potential temperature,  $S(x, y)$  is dimensional orography,  $f_0$  is the Coriolis parameter,  $\beta_0$  is the dimensional northward derivative of the Coriolis parameter,  $L$  and  $U$  are the main length and velocity scales respectively.

In our experiments Eqs. (10)–(13) are integrated using a semi-Lagrangian approach (refer for example to Staniforth and Côté, 1991) coupled with finite-difference schemes. This numerical method is based on the core ideas of solving the QG-model equations explained by Fandry and Leslie (1984).

The test runs employ the VEnKF algorithm applied on top of the QG-model. More precisely, the model is instantiated twice in a twin experiment, where the first instance (hereafter, the truth run) simulates the “nature” and is used to generate observations and the second instance (hereafter, the biased run) runs with different layer depths and is leveraged as prediction model. Both model instances were run at dimension of 40-by-20 grid nodes in each layer thus having 1600 degrees of freedom. The layer depths used in the truth run were 6000 m for the top layer and 4000 m for the bottom layer. In the biased run they were set to 5500 and 4500 m, respectively. The rest of parameters were the same in both runs. The observations extracted from the truth

**2-D geophysical flows with a VEnKF**

Z. Mussa et al.

Title Page	
Abstract	Introduction
Conclusions	References
Tables	Figures
◀	▶
◀	▶
Back	Close
Full Screen / Esc	
Printer-friendly Version	
Interactive Discussion	



## 2-D geophysical flows with a VEnKF

Z. Mussa et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

◀

▶

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



run were perturbed by normally distributed zero-mean noise with standard deviation equal to 0.1. In addition, prior to start of the actual data assimilation the truth and the biased runs for simulated for two weeks of the model time. This was done to establish divergence between the initial estimate of the VEnKF and the first bundle of observations.

The dimension of the problem in the described numerical experiments was still small enough to allow the use of the EKF. Therefore, we compared performance of the VEnKF estimates against those of the classical EKF and against the “tourist’s brochure” (climatology), which is a fixed estimate equal to the mean field of the target model. “Tourist’s brochure” represents the mean model behavior within the observation history and thus corresponds to the “average climate”. As the measure for the estimates quality we chose the root mean square error (RMSE) defined as follows:

$$E_k = \frac{\|\mathbf{x}_k^{\text{truth}} - \mathbf{x}_k^{\text{est}}\|}{\sqrt{N}},$$

where  $\mathbf{x}_k^{\text{truth}}$  is the real model state at time instant  $k$  (which remains unknown in real cases, but is available in the described artificial setting),  $\mathbf{x}_k^{\text{est}}$  is filter estimate at time instant  $k$ , and  $N$  is the dimension of the model. The results of the numerical experiments are shown in Figs. 2 and 3.

In the data assimilation runs coupled with the QG-model we assumed that the measurement noise was known. Hence, the observation error covariance  $\mathbf{C}_{\varepsilon_{k+1}}$  was set to  $0.1\mathbf{I}$  (here  $\mathbf{I}$  is identity matrix of corresponding dimension). Prediction error was specified by two parameters: error variance in a single node of the discretization grid and correlation between two nodes lying in different layers one above the other. The former parameter was set to 0.2 and the latter one was assigned to 0.5. In-layer correlations of the model error were not taken into account. Hence, prediction error covariance was defined as follows:

$$\mathbf{C}_{\varepsilon_k} = \begin{pmatrix} 0.2\mathbf{I} & 0.5\mathbf{I} \\ 0.5\mathbf{I} & 0.2\mathbf{I} \end{pmatrix},$$

where  $\mathbf{I}$  was 800-by-800 identity matrix. The number of the L-BFGS iterations as well as the length of stored iteration history in the VEnKF was set to 50 (as was pointed out earlier, the iteration history is used to generate approximative inverse Hessian of Eq. (4)).

5 The RMSE values of the VEnKF estimates computed with different ensemble sizes are illustrated in Fig. 2. It can be seen that the filter converges already for ensemble cardinality of 20, but stable behavior is observed only with at least 50 ensemble members. Predictably, increasing the ensemble size results in better estimation quality, although even for 400 ensemble members the RMSE values are inferior to that of the  
10 EKF, which is natural due to the approximation error.

Figure 3 contains the forecast skill curves for the VEnKF executed at different ensemble cardinalities as well as for the EKF. The plot demonstrates that when the VEnKF stabilizes (i.e. starting from 50 ensemble members) the effective forecast range stays at the same 7 days mark regardless of the ensemble size growth. Expectedly,  
15 the EKF overcomes its approximation providing about 1 day longer range of effective forecast.

---

**2-D geophysical flows with a VEnKF**

Z. Mussa et al.

---

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

⏪

⏩

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



## 3.2 A real hydraulic two-dimensional flow

### 3.2.1 2-D shallow water equations

The governing equations are the depth-averaged shallow water equations as given in Martin and Gorelick (2005):

$$5 \quad \frac{\partial U}{\partial t} + U \frac{\partial U}{\partial x} + V \frac{\partial U}{\partial y} = -g \frac{\partial \eta}{\partial x} + \varepsilon \left( \frac{\partial^2 U}{\partial x^2} + \frac{\partial^2 U}{\partial y^2} \right) + \gamma_T \frac{(U_a - U)}{H} - S_{fx} - fV, \quad (14)$$

$$\frac{\partial V}{\partial t} + U \frac{\partial V}{\partial x} + V \frac{\partial V}{\partial y} = -g \frac{\partial \eta}{\partial y} + \varepsilon \left( \frac{\partial^2 V}{\partial x^2} + \frac{\partial^2 V}{\partial y^2} \right) + \gamma_T \frac{(V_a - V)}{H} - S_{fy} - fU, \quad (15)$$

$$\frac{\partial \eta}{\partial t} + \frac{\partial(HU)}{\partial x} + \frac{\partial(HV)}{\partial y} = 0. \quad (16)$$

10 where  $U$  is the depth averaged  $x$ -direction velocity component,  $V$  is the depth averaged  $y$ -direction velocity component,  $\eta$  is the free surface elevation,  $g$  is the gravitational constant,  $t$  is time,  $\varepsilon$  is the horizontal eddy viscosity,  $f$  is the Coriolis parameter and  $H = h + \eta$  is the total water depth, where  $h$  is the water depth measured from the undisturbed water surface,  $\gamma_T$  is the wind stress coefficient,  $U_a$  and  $V_a$  are wind velocity components in the  $x$  and  $y$  direction respectively,  $S_{fx}$  and  $S_{fy}$  are the bottom friction terms in  $x$  and  $y$  direction, respectively. The relationship of  $H$ ,  $h$ , and  $\eta$  are as shown in Fig. 4.

15 Top friction and bottom friction boundaries are given by Eqs. 17 and 18, respectively

$$\nu \frac{\partial U}{\partial z} = \gamma_T (U_a - U), \quad \nu \frac{\partial V}{\partial z} = \gamma_T (V_a - V). \quad (17)$$

$$S_{fx} = gU \frac{\sqrt{U^2 + V^2}}{Cz^2}, \quad S_{fy} = gV \frac{\sqrt{U^2 + V^2}}{Cz^2} \quad (18)$$

20 where  $Cz$  is the Chezy coefficient,  $\nu$  is the kinetic velocity coefficient, and  $z$  indicates vertical direction (Martin and Gorelick, 2005).

### 3.2.2 Numerical scheme

A combination of a semi-implicit, semi-Lagrangian time stepping and a finite volume discretization is employed to solve the equations on a rectangular grid (Martin and Gorelick, 2005). This scheme provides a stable solution, even for a time step larger than the Courant–Friedrichs–Levy (CFL) restriction, defined by

$$\text{CFL} = w \frac{\Delta t}{\Delta x_j} \quad (19)$$

where  $w$  is the velocity component in  $x_j$  direction,  $\Delta t$  is the time step size and  $\Delta x_j$  is the cell dimension in the  $x_j$  direction of flow (Martin and Gorelick, 2005). The CFL relates fluid velocity and time step size with computational cell size, and requires that the CFL should be smaller than 1. Please refer Martin and Gorelick (2005) for more on numerical approximations.

### 3.2.3 Boundary conditions

The model identifies itself the location of water/land boundaries using Eqs. (20) and (21) (Martin and Gorelick, 2005):

$$H_{i+1/2,j}^{N+1} = \max(0, h_{i+1/2,j} + \eta_{i,j}^{N+1}, h_{i+1/2,j} + \eta_{i+1,j}^{N+1}), \quad (20)$$

$$H_{i,j+1/2}^{N+1} = \max(0, h_{i,j+1/2} + \eta_{i,j}^{N+1}, h_{i,j+1/2} + \eta_{i,j+1}^{N+1}). \quad (21)$$

Two types of radiation boundaries have been set, which are:

- i. The projection of velocity normal to the domain boundary:

$$\frac{\partial U}{\partial t} + U_{\text{upw}} \frac{\partial U}{\partial n} = 0 \quad (22)$$

where  $U_{\text{upw}}$  is the upwinded normal direction velocity component, and  $n$  is the direction normal to the domain boundary (Martin and Gorelick, 2005).

**2-D geophysical flows with a VEnKF**

Z. Mussa et al.

Title Page	
Abstract	Introduction
Conclusions	References
Tables	Figures
◀	▶
◀	▶
Back	Close
Full Screen / Esc	
Printer-friendly Version	
Interactive Discussion	



ii. To limit wave reflections at open boundaries, the following condition is imposed

$$\frac{\partial \eta}{\partial t} + C_n \frac{\partial \eta}{\partial n} = 0 \tag{23}$$

where  $C_n$  is the propagation velocity from grid points around the boundary (Martin and Gorelick, 2005).

**3.3 Two-dimensional ensemble generation**

For defining the shore boundary, VEnKF does not account for additional prior knowledge beyond the observations. This fact means that in case of problems on bounded domains, there is no way to include information about the boundaries into the Kalman filter analysis. If the prediction model automatically maintains the boundaries in accordance with defined constraints, one can simply reduce the data assimilation analysis to the inner part of the model domain. However, this approach is complicated when the boundaries change over the time.

In our experiments we use a strategy, which allows us to account for evolving boundaries more flexibly, albeit it does not guarantee that the boundaries will be preserved exactly as required by the model constraints. Information about the boundaries is included into the model uncertainty description, i.e. into the model error covariance  $\mathbf{C}_{\varepsilon_k}$ . This changes the analytical representation of the boundaries into a probabilistic description, which means that there is no absolute certainty about where the boundaries are located, but there is more confidence about evolution of boundaries than that of the model.

In the dam-break case studied here, we have prior knowledge about the shore line and we have high certainty that there can be no water in the middle of the river bank. Therefore, we define the model covariance  $\mathbf{C}_{\varepsilon_k}$  so that the state elements that are confined to the river bank have variances much smaller than the variances assigned to the rest of the state. This strategy shifts the responsibility of maintaining the boundaries to the data assimilation analysis.

### 3.3.1 Setup of the meters with two rows of observations

A dam-break experiment of Bellos et al. (1991) is tested with new modification. At the downstream end the meters are placed as in the original experiment and as shown in Martin and Gorelick (2005). Parallel wave meters are placed along the flow in the downstream end. The goal of this setup is to examine if the VEnKF can learn and predict cross-flow along the flume, which is the true case for river flows. The meters are placed in the same original position in the  $y$  direction but pushed up and down from the flume midline for downstream meters by  $4\Delta x$ , i.e.  $y' = y$  and  $x' = x \pm 4\Delta x$ , where  $x'$  and  $y'$  are new wave meter positions downstream in  $x$  and  $y$  direction, respectively. This makes total of 8 wave meters downstream as can be seen in Fig. 5. Other parameters are as in Martin and Gorelick (2005).

With these new positions of the meters, data are collected by assuming there are cross flows along the flume. To accomplish this assumption we simulate a new set of observations at the downstream end by superimposing a sinusoidal wave across the water flow on the true experimental observations. The sine wave is chosen so that the wave crests are not in a straight angle to the channel walls, so that the observations correspond to genuine cross-flow patterns, and that the water level at the sensors does not drop to zero at any time of the simulation. To make it more natural, we add normally distributed noise with mean 0 and standard deviation of 0.001 to the resulting data set.

### 3.3.2 Interpolation in time

As mentioned above, the data set of observations has a major disadvantage of sparsity, both in time and in space. More precisely, that means that at a time instance only a small number of sensors among those installed along the flume were producing actual measurements, whereas the time instances had no alignment with the model integration time step set to 0.1 s. This sparsity is a challenge to data assimilation methods, since the amount of data obtained from the measurements is usually not enough to expose bias in the prediction model. Therefore simple interpolation

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

◀

▶

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



techniques were deployed to reduce the negative impact caused by the sparsity of the observations.

As we expect the flume to be continuous both over time and space, the time interpolation can be performed if the measurements obtained from each wave-meter are frequent enough. In the case discussed, the smallest data rate of incoming measurements was demonstrated by the second wave-meter and was equal to  $7.07 \times 10^{-1}$  measurements per second, which means that the longest average time period between two successive observations was about 0.14 s. As this period is about the same length as the model integration time step, we expect satisfactory results from the time interpolation.

The interpolation procedure was organized as follows. We began by discretizing the time axis with a discretization step of 0.1 s. Thereafter, every time instance related to a measurement obtained from a wave-meter installed in the flume was aligned with the time discretization grid by rounding the time instances to the closest grid point. Since the time grid resolution is smaller than the rate of incoming measurements, some of the time grid points were left with no related observation. These gaps were filled by piecewise cubic interpolation defined by Hermite Interpolating Polynomials (Fritsch and Carlson, 1980). The curves given in Fig. 6 demonstrate the resulting interpolation and the original measurement data obtained from the 2nd sensor.

### 3.3.3 Gaussian interpolation in space

As noted previously, the published observation data we use in our experiments was defined for seven spatial locations only, which is much less than the model state dimension. Therefore, for each sensor, the corresponding observations were extrapolated to a small neighbourhoods of their spatial location.

We specified a square patch that defines the extrapolation neighbourhood for each actual observation. These neighbourhoods were specified symmetrically with their centres aligned to the spatial locations of the sensors. Thereafter, the extrapolated values were defined by a product of an actual measurement and a Gaussian weight,

## 2-D geophysical flows with a VEnKF

Z. Mussa et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

⏪

⏩

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



where the weights were computed from a discretized Gaussian hill with its top aligned to the sensor location and the discretization defined by the spatial grid and the extrapolation neighbourhood.

We should note that the Gaussian kernels were allowed to be anisotropic, i.e. to have different standard deviations along the spatial axes. An example of the discussed spatial interpolation computed for the 2nd sensor measurements is shown in Fig. 7. We used a smaller Gaussian patch of 5-by-5, so that we do not spread the cross wave effect on a wide area, since the difference might grow bigger than the one estimated by the gaussian interpolation if we use a bigger square patch.

### 3.3.4 VEnKF parameters

The state vector for the assimilation is defined as the vector of heights at the center of a grid point. The complete state vector comprises free surface elevation  $\eta$  and horizontal velocities  $u$  in the  $x$  direction and  $v$  in the  $y$  direction for the entire domain; that is  $\mathbf{x} = [\eta \ u \ v]^T$ . The model has therefore altogether 16000 spatial degrees of freedom. With the interpolations, the ensembles are sampled in every time step of the assimilation. The observation error covariance and the model error covariance are both assumed to be diagonal matrices. The observation operator  $\mathbf{H}^{\text{TL}}$  from Eq. 4 is a linear operator that maps the state vector to the observation space corresponding to all grid points covered by the interpolated data, but restricted to the water height values only.

### 3.3.5 Results with two rows of observations

The result shown by VEnKF proves the capability of VEnKF to capture the cross flows along the flume. As it can be seen in Figs. 9 and 10. In both cases we can see a reasonable balance between the measurements and the model prior estimate. For the first three upstream sensors, the flow captured by the VEnKF is realistic and no cross flows are detected, as can be seen in Fig. 8, whereas for the downstream sensors

## 2-D geophysical flows with a VEnKF

Z. Mussa et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

◀

▶

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



sinusoidal oscillations indicate the presence of cross-flows that can also be seen in a video of the simulation.

### 3.4 Observation interpolation and Kalman filter stability

When studying the impact of observation interpolation with two rows of observations and the superimposed sine wave solution, a curious phenomenon was observed. There appears to be a relationship between the interpolation distance of observations in time, and the spread of the ensemble used in re-sampling. The latter quantity is directly proportional to the assumed variance of the state vector and is therefore an analogue of covariance inflation that is often used with Ensemble Kalman Filters.

The nature of this relationship is such that the shorter the distance in time between observations, the smaller the spread of the ensemble can be, and the analysis still captures the wave-like pattern of the solution. But if the interpolation distance of observations is too long, the VEnKF algorithm misses the troughs of the true solution, unless the spread of the ensemble is also increased correspondingly. In this case, VEnKF diverges, even if there is no numerical instability. This behaviour can be seen in Figs. 11, 12, and 13, where the analysis begins to diverge from the true solution if ensemble variance is too small.

As a further analysis, the relationship between interpolation distance and spread was studied over a range of values of both. The metric used in the comparison is the difference in Euclidean norm between the analysis and the true solution. Figure 14 plots this relationship in logarithmic coordinates and shows the level curves of this distance as a function of observation interpolation distance in time  $\Delta t$  and of ensemble variance  $\sigma^2$ . The level curves are seen to be almost linear in logarithmic scale.

Based on a visual study of the slopes of the level curves with respect to the  $\log(\sigma)$  and  $\log(\Delta t)$  axes in Fig. 14, we can see that there appears to be a power law that relates the standard deviation of the ensemble and the time interval between observations that has to be met for the VEnKF algorithm not to diverge. This empirical power law that

## 2-D geophysical flows with a VEnKF

Z. Mussa et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

◀

▶

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



guarantees filter convergence looks roughly like

$$\Delta t \sim \sigma^6 \quad \text{or} \quad \Delta t \sim \text{var}^3.$$

The exact nature and reasons behind these particular exponents remains a subject for further investigation.

## 5 4 Conclusions

In an earlier study by the current authors Amour et al. (2013), one of the current flow examples, the dam break example, was assimilated with VEnKF, but in that case the observations of the flow were located along a one-dimensional flow contour only, even if the flow model was based on the two-dimensional shallow water equations. The analysis increments resulting from Kalman filtering in that case tended to be almost one-dimensional structures that are constant in the cross-flow direction.

In the current article, we have seen that the Variational Ensemble Kalman Filter VEnKF is able to capture genuinely two-dimensional realistic flow patterns in geophysical flows, both in the case of a synthetic atmospheric flow with a biased model and a modified real hydraulic flow. In the latter case, real observations were interpolated in time to prevent filter divergence. By the two-dimensionality of the flow pattern we mean that both the real flow and the real or synthetic observations of the flow have a two-dimensional geometry.

When the stability of the VEnKF filter was studied as a function of ensemble variance and interpolation distance in time between observations, an empirical relationship between these two quantities was detected that establishes a power law between the two quantities above as a sufficient condition for filter stability.

*Acknowledgements.* We would like to acknowledge the financial support from Lappeenranta University of Technology and the Science and Technology Higher Education Project through the University of Dar es Salaam and Dar es Salaam University College of Education.

## 2-D geophysical flows with a VEnKF

Z. Mussa et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

◀

▶

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



## References

- Amour, I., Mussa, Z., Bibov, A., and Kauranne, T.: Using ensemble data assimilation to forecast hydrological flumes, *Nonlin. Processes Geophys.*, 20, 955–964, doi:10.5194/npg-20-955-2013, 2013. 409, 428
- 5 Auvinen, H., Bardsley, J., Haario, H., and Kauranne, T.: The variational Kalman filter and an efficient implementation using limited memory BFGS, *Int. J. Numer. Meth. Fl.*, 64, 314–335, doi:10.1002/flid.2153, 2010. 406, 413
- Bardsley, J. M., Parker, A., Solonen, A., and Howard, M.: Krylov space approximate Kalman filtering, *Numer. Linear Algebr.*, 20, 171–184, doi:10.1002/nla.805, 2013. 406
- 10 Bélanger, E. and Vincent, A.: Data assimilation (4D-VAR) to forecast flood in shallow-waters with sediment erosion, *J. Hydrol.*, 300, 114–125, doi:10.1016/j.jhydrol.2004.06.009, 2004. 408
- Bellos, C., Soulis, J., and Sakkas, J.: Computation of two-dimensional dam-break induced flows, *Adv. Water Resour.*, 14, 31–41, 1991. 409, 424
- 15 Bertino, L., Evensen, G., and Wackernagel, H.: Sequential data assimilation techniques in oceanography, *Int. Stat. Rev.*, 71, 223–241, 2003. 406, 407
- Blum, J., Le Dimet, F.-X., Navon, I. M., et al.: Data assimilation for geophysical fluids, *Handbook of Numerical Analysis: Computational Methods for the Atmosphere and the Oceans*, 14, 385–441, 2009. 405
- 20 Chang, T.-J., Kao, H.-M., Chang, K.-H., and Hsu, M.-H.: Numerical Simulation of shallow water dam break flows in open channels using smoothed particle hydrodynamics, *Hydrology*, 408, 78–90, 2011. 407
- Daley, R.: *Atmospheric Data Analysis*, Cambridge Atmospheric and Space Science Series, Cambridge University Press, 1993. 406
- 25 Dee, D. P.: Simplification of the Kalman filter for meteorological data assimilation, *Q. J. Roy. Meteor. Soc.*, 117, 365–384, doi:10.1002/qj.49711749806, 1991. 406
- Evensen, G.: Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte-Carlo methods to forecast error statistics, *J. Geophys. Res.*, 99, 10143–10162, doi:10.1029/94JC00572, 1994. 406, 407
- 30 Evensen, G.: The Ensemble Kalman Filter: theoretical formulation and practice implementation, *Ocean Dynam.*, 53, 343–367, doi:10.1007/s10236-003-0036-9, 2003. 412

**NPGD**

1, 403–446, 2014

## 2-D geophysical flows with a VEnKF

Z. Mussa et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

◀

▶

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



## 2-D geophysical flows with a VENKF

Z. Mussa et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

◀

▶

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



- Fandry, C. and Leslie, L.: A Two-Layer Quasi-Geostrophic Model of Summer Trough Formation in the Australian Subtropical Easterlies, *J. Atmos. Sci.*, 41, 807–818, 1984. 417, 418
- Fisher, M., Nocedal, J., Trémolet, Y., and Wright, S. J.: Data assimilation in weather forecasting: a case study in PDE-constrained optimization, *Optim. Eng.*, 10, 409–426, 2009. 407
- 5 Fritsch, F. and Carlson, R.: Monotone Piecewise Cubic Interpolation, *Society of Industrial and Applied Mathematics Journal on Numerical Analysis*, 17, 238–246, doi:10.1137/0717021, 1980. 425
- Hamill, T. M., Whitaker, J. S., and Snyder, C.: Distance-dependent filtering of background error covariance estimates in an ensemble Kalman filter, *Mon. Weather Rev.*, 129, 2776–2790, 2001. 406
- 10 Heemink, A. and Metzelaar, I.: Data Assimilation into a numerical shallow water flow model: a stochastic optimal control approach, *J. Marine Syst.*, 6, 145–158, 1995. 406, 407
- Hoteit, I., Triantafyllou, G., and Korres, G.: Using low-rank ensemble Kalman filters for data assimilation with high dimensional imperfect models, *JNAIAM*, 2, 67–78, 2007. 407
- 15 Houtekamer, P. and Mitchell, H.: Data assimilation using an ensemble Kalman filter technique, *Mon. Weather Rev.*, 126, 796–811, doi:10.1175/1520-0493(1998)126<0796:DAUAEK>2.0.CO;2, 1998. 413
- Kazantsev, E.: Boundary conditions control for a shallow-water model, *Int. J. Numer. Meth. Fl.*, 68, 625–641, 2012. 408
- 20 Kuznetsov, L., Ide, K., and Jones, C.: A method for assimilation of Lagrangian data, *Mon. Weather Rev.*, 131, 2247–2260, 2003. 406
- Lei, L., Stauffer, D. R., and Deng, A.: A hybrid nudging-ensemble Kalman filter approach to data assimilation. Part II: application in a shallow-water model, *Tellus A*, 64, 18485, doi:10.3402/tellusa.v64i0.18485, 2012. 408
- 25 Liang, Q. and Marche, F.: Numerical resolution of well-balanced shallow water equations with complex source terms, *Adv. Water Resour.*, 32, 873–884, 2009. 407
- Liu, Y., Weerts, A. H., Clark, M., Hendricks Franssen, H.-J., Kumar, S., Moradkhani, H., Seo, D.-J., Schwanenberg, D., Smith, P., van Dijk, A. I. J. M., van Velzen, N., He, M., Lee, H., Noh, S. J., Rakovec, O., and Restrepo, P.: Advancing data assimilation in operational hydrologic forecasting: progresses, challenges, and emerging opportunities, *Hydrol. Earth Syst. Sci.*, 16, 3863–3887, doi:10.5194/hess-16-3863-2012, 2012. 407
- 30 Martin, N. and Gorelick, S. M.: MODFreeSurf2D: A MATLAB surface fluid flow model for rivers and streams, *Comput. Geosci.*, 31, 926–946, 2005. 409, 421, 422, 423, 424

## 2-D geophysical flows with a VEnKF

Z. Mussa et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

◀

▶

◀

▶

Back

Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion



- McMillan, H. K., Hreinsson, E. Ö., Clark, M. P., Singh, S. K., Zammit, C., and Uddstrom, M. J.: Operational hydrological data assimilation with the recursive ensemble Kalman filter, *Hydrol. Earth Syst. Sci.*, 17, 21–38, doi:10.5194/hess-17-21-2013, 2013. 406, 407
- Miller, R. N., Ghil, M., and Gauthiez, F.: Advanced Data Assimilation in Strongly Nonlinear Dynamical Systems, *J. Atmos. Sci.*, 51, 1037–1056, 1994. 405
- Navon, I. M.: Data assimilation for numerical weather prediction: a review, in: *Data Assimilation for Atmospheric, Oceanic and Hydrologic Applications*, Springer, 21–65, 2009. 407
- Nocedal, J. and Wright, S.: *Numerical Optimization*, chap. Limited-Memory BFGS, Springer-Verlag, New York, 224–227, 1999. 415
- Pedlosky, J.: *Geophysical Fluid Dynamics*, chap. Geostrophic Motion, Springer-Verlag, New York, 22–57, 1987. 416
- Reichle, R. H., McLaughlin, D. B., and Entekhabi, D.: Hydrologic Data Assimilation with the Ensemble Kalman Filter, *Mon. Weather Rev.*, 130, 103–114, 2002. 407
- Simon, D.: Optimal state estimation, Kalman,  $H_\infty$ , and nonlinear approaches, chap. The Extended Kalman filter, Wiley-Interscience, Hoboken, 400–403, 2006. 411, 413
- Solonen, A., Hakkarainen, J., Auvinen, H., Amour, I., Haario, H., and Kauranne, T.: Variational Ensemble Kalman Filtering Using Limited Memory BFGS, *Electronic Transaction on Numerical Analysis*, 39, 271–285, 2012. 405, 409, 415, 416
- Staniforth, A. and Côté, J.: Semi-lagrangian integration schemes for atmospheric models review, *Mon. Weather Rev.*, 119, 2206–2223, 1991. 418
- Strub, I. S., Percelay, J., Tossavainen, O.-P., and Bayen, A. M.: Comparison of two data assimilation algorithms for shallow water flows, *Netw. Heterog. Media*, 4, 409–430, doi:10.3934/nhm.2009.4.409, 2009. 408
- Tinka, A., Strub, I. S., Wu, Q., and Bayen, A. M.: Quadratic Programming based data assimilation with passive drifting sensors for shallow water flows, in: *48th Conference on Decision and Control*, Shanghai, China, 16–18 December 2009, 7614–7620, 2009. 408
- Tseng, M. H. and Chu, C. R.: The simulation of dam break flows by an improved predictor-corrector TVD schemes, *Adv. Water Resour.*, 23, 637–643, 2011. 407
- van Leeuwen, P. J.: Efficient nonlinear data assimilation in geophysical fluid dynamics, *Comput. Fluids*, 46, 52–58, doi:10.1016/j.compfluid.2010.11.011, 2011. 405, 406
- van Leeuwen, P. J. and Ades, M.: Efficient fully nonlinear data assimilation in geophysical fluid dynamics, *Comput. Geosci.*, 55, 16–27, 2013. 406

Wu, L., Mallet, V., Bocquet, M., and Sportisse, B.: A comparison study of data assimilation algorithms for ozone forecasts, *J. Geophys. Res.*, 113, D20310, doi:10.1029/2008JD009991, 2008. 405

Xu, L. and Daley, R.: Data assimilation with a barotropically unstable shallow water system using representer algorithms, *Tellus A*, 54, 125–137, 2002. 408

Zhou, J., Causon, D. M., Mingham, C. G., and Ingram, D. M.: Numerical prediction of dambreak flows in general geometries with complex bed topography, *J. Hydraul. Eng.-ASCE*, 130, 332–340, 2004. 407

Zupanski, M.: Maximum likelihood ensemble filter: theoretical aspects, *Mon. Weather Rev.*, 133, 1710–1726, doi:10.1175/MWR2946.1, 2005. 413

**2-D geophysical flows with a VEnKF**

Z. Mussa et al.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

⏪

⏩

◀

▶

Back

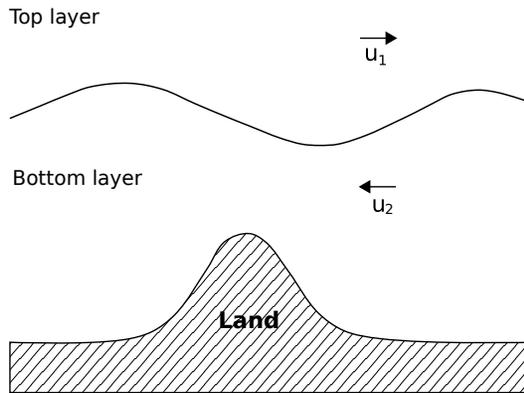
Close

Full Screen / Esc

Printer-friendly Version

Interactive Discussion





**Fig. 1.** Geometric layout of the QG-model.

**2-D geophysical flows with a VEnKF**

Z. Mussa et al.

Title Page

Abstract Introduction

Conclusions References

Tables Figures

◀ ▶

◀ ▶

Back Close

Full Screen / Esc

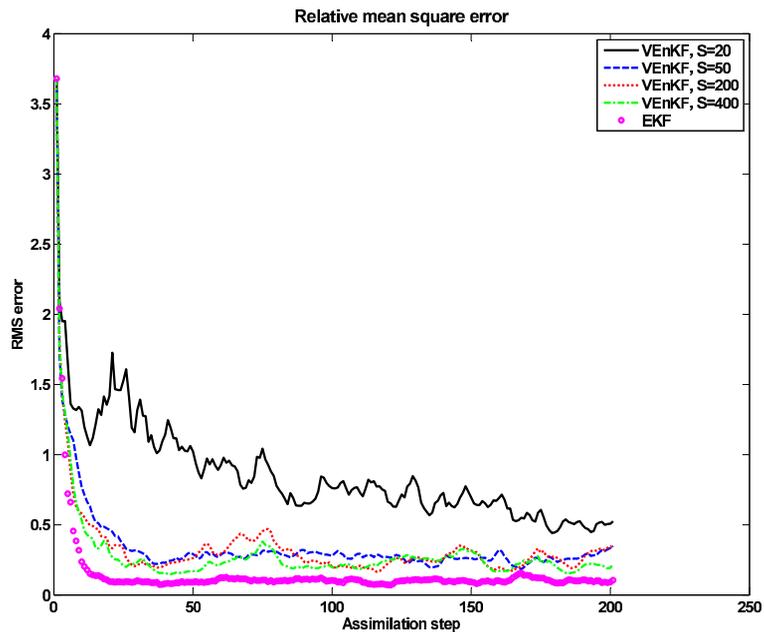
Printer-friendly Version

Interactive Discussion



## 2-D geophysical flows with a VEnKF

Z. Mussa et al.

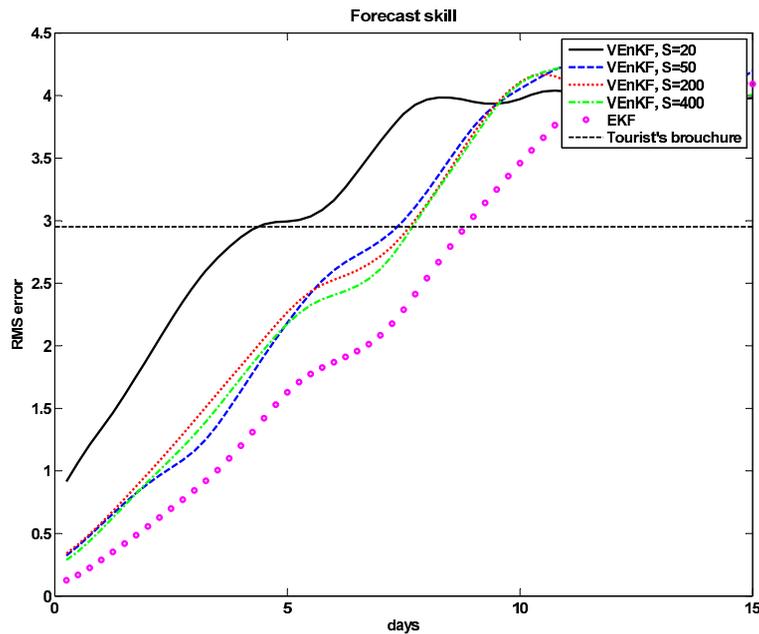


**Fig. 2.** Root mean square error of the estimates obtained from the data assimilation of the QG-model.

[Title Page](#)[Abstract](#)[Introduction](#)[Conclusions](#)[References](#)[Tables](#)[Figures](#)[I ◀](#)[▶ I](#)[◀](#)[▶](#)[Back](#)[Close](#)[Full Screen / Esc](#)[Printer-friendly Version](#)[Interactive Discussion](#)

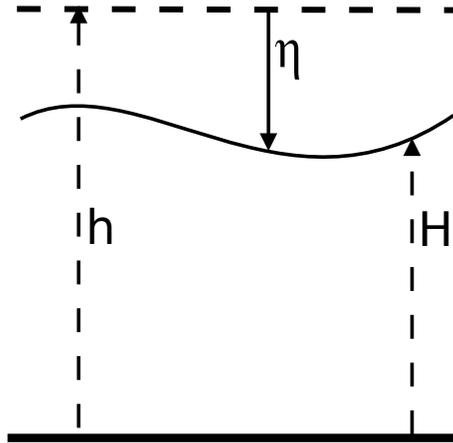
**2-D geophysical flows with a VEnKF**

Z. Mussa et al.



**Fig. 3.** Forecast skill of VEnKF at different ensemble sizes and that of EKF on the QG-model.

[Title Page](#)[Abstract](#)[Introduction](#)[Conclusions](#)[References](#)[Tables](#)[Figures](#)[◀](#)[▶](#)[◀](#)[▶](#)[Back](#)[Close](#)[Full Screen / Esc](#)[Printer-friendly Version](#)[Interactive Discussion](#)



**Fig. 4.** Relationship between  $H$ ,  $h$ , and  $\eta$ .

**2-D geophysical flows with a VEnKF**

Z. Mussa et al.

Title Page

Abstract Introduction

Conclusions References

Tables Figures

◀ ▶

◀ ▶

Back Close

Full Screen / Esc

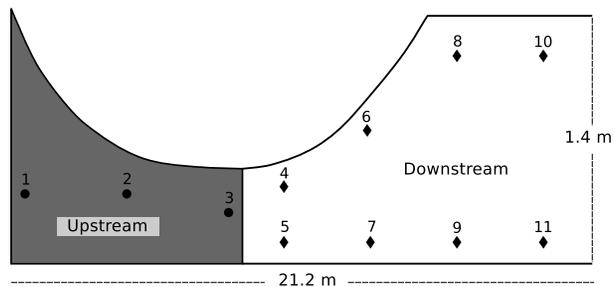
Printer-friendly Version

Interactive Discussion



**2-D geophysical flows with a VEnKF**

Z. Mussa et al.



**Fig. 5.** Parallel setup of wave meters at downstream end (plan view).

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

◀

▶

◀

▶

Back

Close

Full Screen / Esc

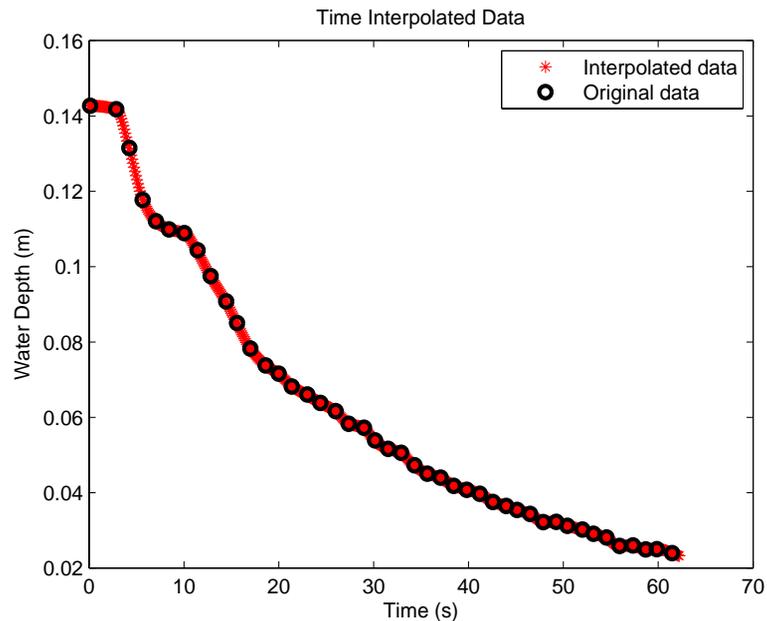
Printer-friendly Version

Interactive Discussion



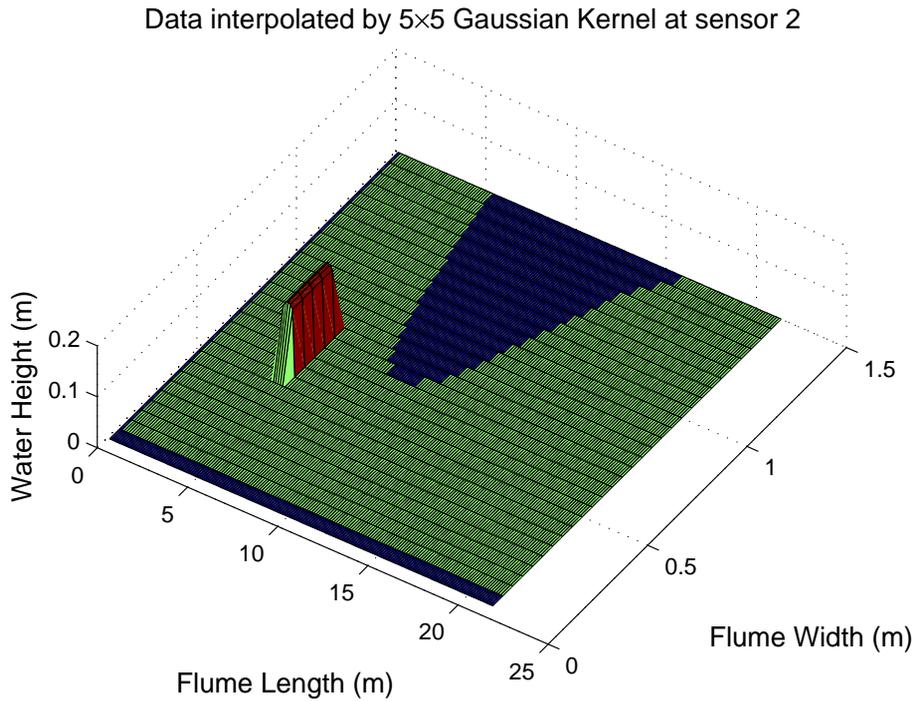
**2-D geophysical flows with a VEnKF**

Z. Mussa et al.

**Fig. 6.** Time interpolated water depth at Sensor Number 2.[Title Page](#)[Abstract](#)[Introduction](#)[Conclusions](#)[References](#)[Tables](#)[Figures](#)[◀](#)[▶](#)[◀](#)[▶](#)[Back](#)[Close](#)[Full Screen / Esc](#)[Printer-friendly Version](#)[Interactive Discussion](#)

**2-D geophysical flows with a VEnKF**

Z. Mussa et al.



**Fig. 7.** Space interpolated water depth at Sensor Number 2.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

◀

▶

◀

▶

Back

Close

Full Screen / Esc

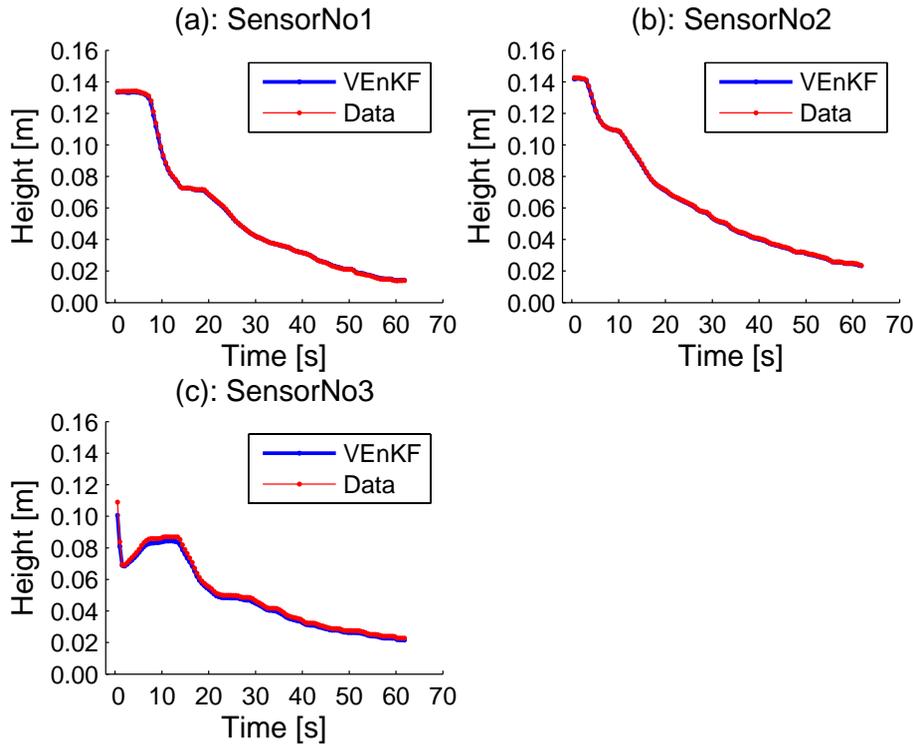
Printer-friendly Version

Interactive Discussion



## 2-D geophysical flows with a VEnKF

Z. Mussa et al.



**Fig. 8.** Upstream meters: no cross flows recorded by the VEnKF as was expected.

Title Page

Abstract Introduction

Conclusions References

Tables Figures

◀ ▶

◀ ▶

Back Close

Full Screen / Esc

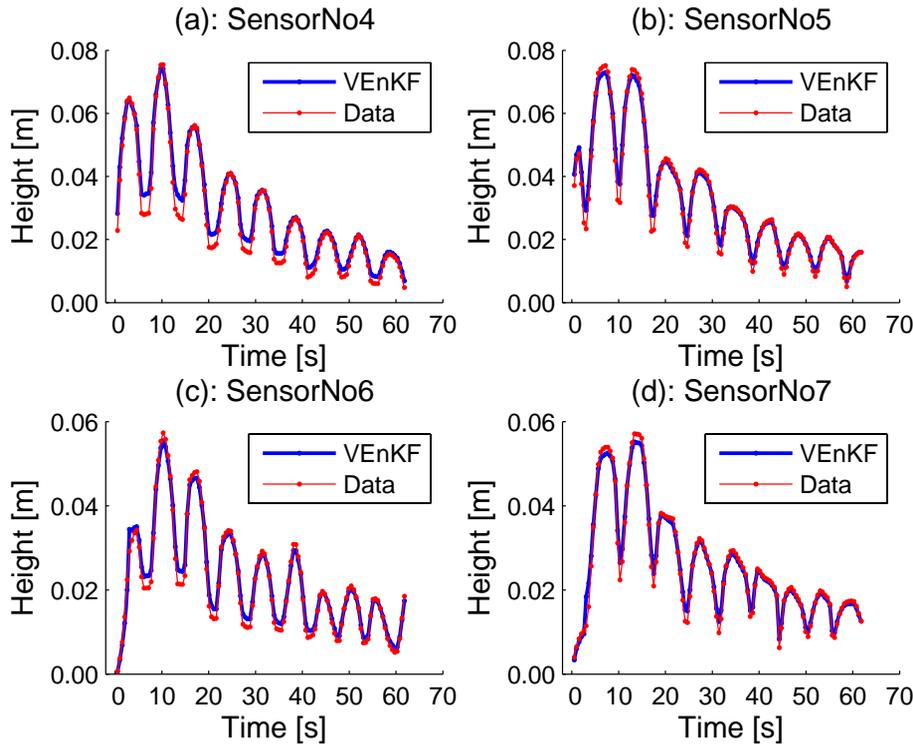
Printer-friendly Version

Interactive Discussion



**2-D geophysical flows with a VEnKF**

Z. Mussa et al.



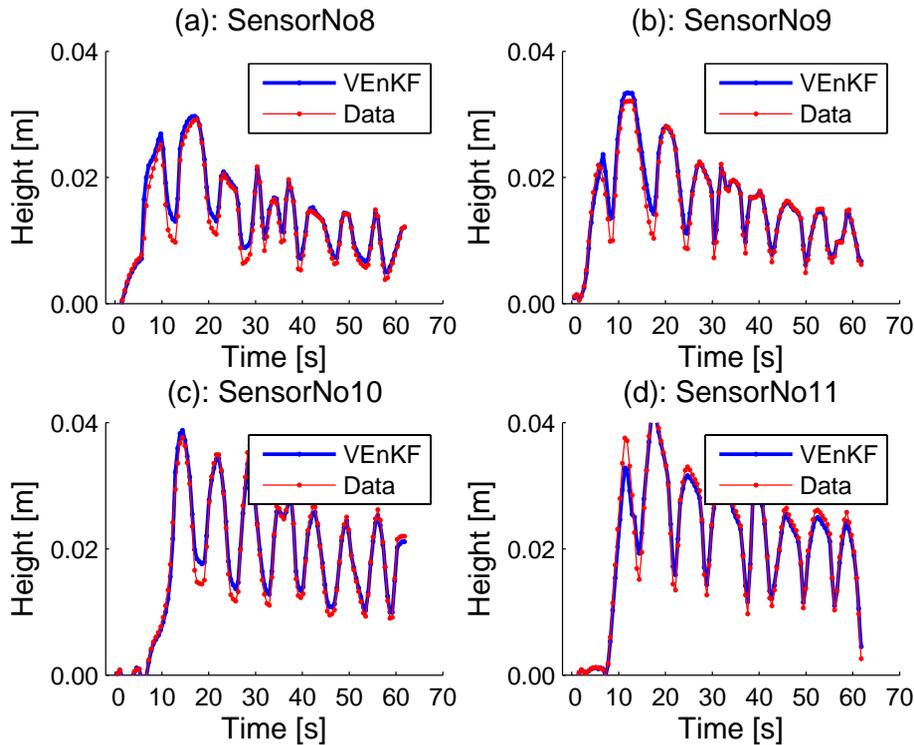
**Fig. 9.** The VEnKF captures well the cross flows for the downstream locations.

Title Page	
Abstract	Introduction
Conclusions	References
Tables	Figures
◀	▶
◀	▶
Back	Close
Full Screen / Esc	
Printer-friendly Version	
Interactive Discussion	



**2-D geophysical flows with a VEnKF**

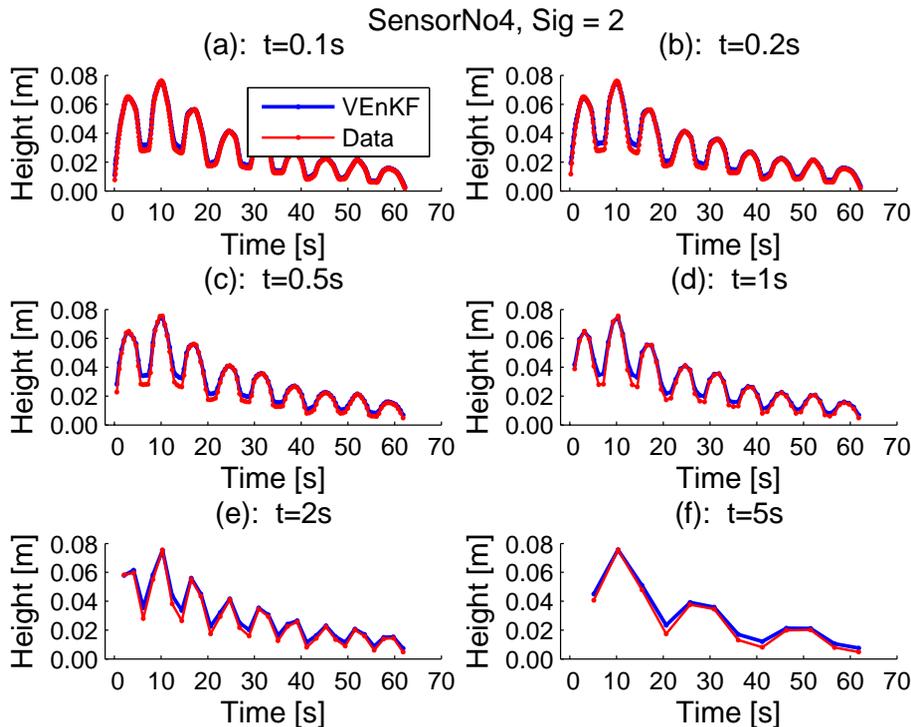
Z. Mussa et al.



**Fig. 10.** The VEnKF captures well the cross flows for the downstream locations.

## 2-D geophysical flows with a VEnKF

Z. Mussa et al.



**Fig. 11.** Results showing VEnKF converges to the true measurements with all observation intervals if ensemble variance is sufficient ( $\sigma^2 = 4$ ). Note the aliasing of the sine wave to a lower frequency wave when the observation interval exceeds the wave frequency at  $5 \Delta t = 5\text{s}$  and the estimation problem violates the Nyquist limit. The filter then converges to the aliased solution.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

◀

▶

◀

▶

Back

Close

Full Screen / Esc

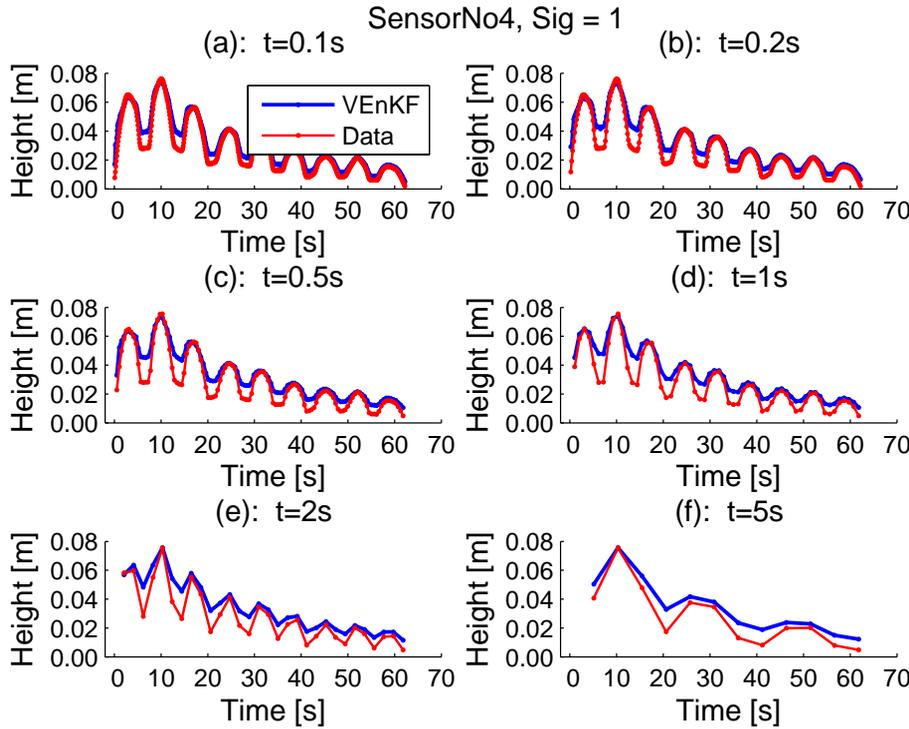
Printer-friendly Version

Interactive Discussion



**2-D geophysical flows with a VEnKF**

Z. Mussa et al.



**Fig. 12.** Border-line filter divergence with different observation intervals and border-line ensemble variance  $\sigma^2 = 1$ .

Title Page

Abstract      Introduction

Conclusions      References

Tables      Figures

◀      ▶

◀      ▶

Back      Close

Full Screen / Esc

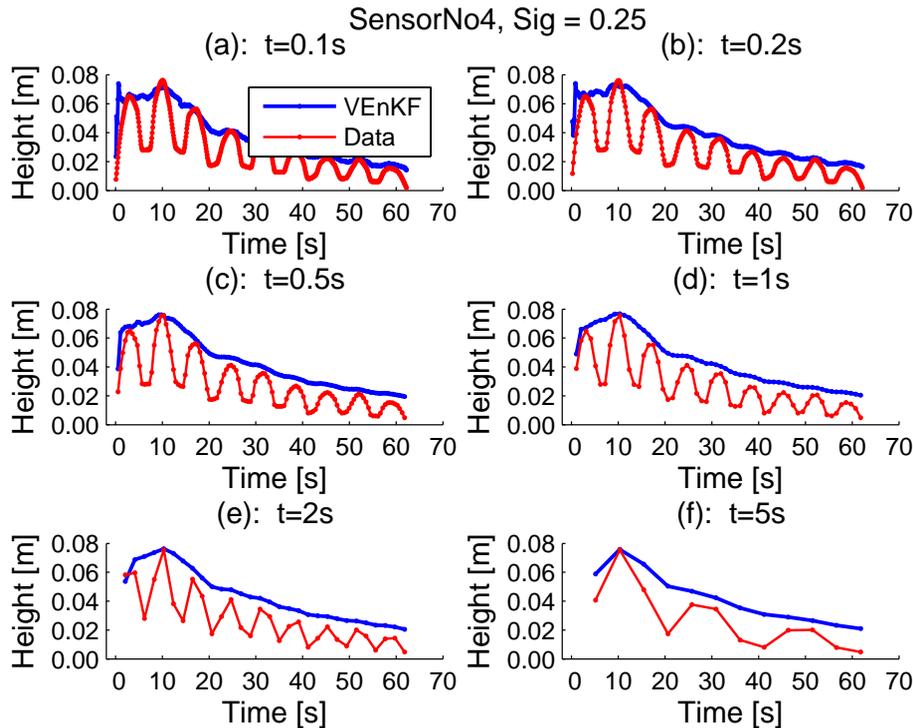
Printer-friendly Version

Interactive Discussion



## 2-D geophysical flows with a VEnKF

Z. Mussa et al.



**Fig. 13.** Results showing VEnKF divergence at all observation intervals with excessively small ensemble variance ( $\sigma^2 = 0.625$ ). The solutions remain numerically stable in all cases.

Title Page

Abstract

Introduction

Conclusions

References

Tables

Figures

◀

▶

◀

▶

Back

Close

Full Screen / Esc

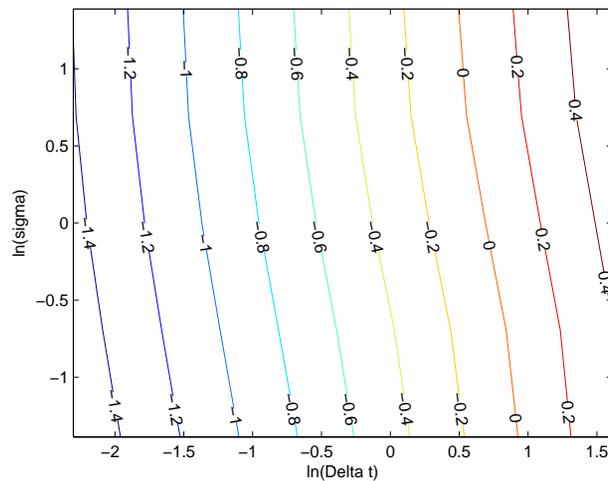
Printer-friendly Version

Interactive Discussion



**2-D geophysical flows with a VEnKF**

Z. Mussa et al.



**Fig. 14.** Empirical level curves for the difference in Euclidean norm between the analysis and true solution as a function of the logarithm of the observation interval  $\Delta t$  and the logarithm of ensemble standard deviation  $\sigma$ .

[Title Page](#)[Abstract](#)[Introduction](#)[Conclusions](#)[References](#)[Tables](#)[Figures](#)[Back](#)[Close](#)[Full Screen / Esc](#)[Printer-friendly Version](#)[Interactive Discussion](#)