

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27

**Validation and application of an ensemble Kalman filter in the  
Selat Pauh of Singapore**

Jun Wei and Paola Malanotte-Rizzoli

Department of Earth, Atmospheric and planetary Sciences  
Massachusetts Institute of Technology  
Cambridge, MA 02139

junwei@mit.edu  
rizzoli@mit.edu

*December 2008*

1 **Abstract**

2

3

4

5

6

7

8

9

10

11

12

13

14

15

16

17

The effectiveness of an ensemble Kalman filter (EnKF) is assessed in the Selat Pauh of Singapore using observing system simulation experiment (OSSE). Perfect and imperfect model experiments are considered. The perfect model experiments examine the EnKF in reducing the initial perturbations with no model errors. Current velocity at 15 observational sites from the true ocean is assimilated every hour into the false ocean. While EnKF reduces the initial velocity error during the first few hours, it fails after one tidal cycle ( $\sim 12$  hours) due to the rapid convergence of the ensemble members. The imperfect model experiments consider both initial errors and model errors. The model errors are mimicked through wind perturbations. A random perturbation  $\varepsilon$  is applied independently to each ensemble member to maintain the ensemble spread. The assimilation results showed that the success of EnKF depends essentially on the presence of  $\varepsilon$ , yet it is not sensitive to the magnitude of  $\varepsilon$ . Although all experiments were made with EnKF only, the results could be applicable in general to all other ensemble-based data assimilation methods.

# 1 1. Introduction

2 Numerical ocean models have become capable of simulating the four dimensional  
3 evolution of the ocean with the flexible resolution that is essential to reproduce and  
4 understand the processes not captured by the limited observations. In spite of this  
5 progress, ocean models still contain errors because of the incomplete representation of  
6 ocean physics. Thus it is necessary to correct for the model errors and this can be done  
7 through data assimilation. Data assimilation methods synthesize the model solution with  
8 the available limited observations to obtain an optimal solution (analysis) which can be  
9 used as the new initial condition for model forecasting. In general, the analysis procedure  
10 minimizes the difference (error) between the model states and the observations using  
11 least-square methods. In this study, we focus on the Ensemble Kalman Filter (EnKF), one  
12 of the most advanced sequential assimilation methods, introduced in oceanography by  
13 Evensen [1994].

14 The original Kalman filter is developed for a linear system and the extended  
15 Kalman filter (EKF) extends its basic algorithm to non-linear problems by linearizing the  
16 non-linear function around the current estimate. For a realistic ocean model, however, it  
17 is difficult to directly apply EKF because of the computational burden associated with  
18 large model dimensions, which produces a computational requirement far beyond today  
19 computational capability. Thus a number of suboptimal approximations have been  
20 proposed. Among them are the filters based on ensemble estimation [*Evensen, 1994;*  
21 *Houtekamer and Mitchell, 1998; Tippett et al., 2003; Zang and Malanotte-Rizzoli, 2003;*  
22 *Chen et al., 2008*].

1           The EnKF predicts a flow-dependent error covariance through a set of ensemble  
2 forecasts computed directly from non-linear models. The success of EnKFs for large  
3 scales is in part due to the global and continuous coverage of satellite data which provides  
4 stringent constraints on the ocean models. In contrary, the studies of Kalman filters in  
5 coastal areas are very limited. *Chen et al.* [2008] present an extensive application of the  
6 ensemble filters with limited, point-wise measurements in three coastal problems. They  
7 used the twin experiment approach considering the model perfect and 20 ensemble  
8 members. They showed that the ensemble-based filters (EnKF and EnSKF) are very  
9 successful in an idealized tide- and buoyancy-driven problem in an estuary. In this study  
10 we examine the EnKF scheme in a realistic coastal region, the Selat Pauh of Singapore.  
11 Perfect and imperfect model experiments are carried out to examine the reliability and  
12 limitation of the EnKF. The paper is organized as follows. Section 2 describes the study  
13 area and the ocean model. Section 3 briefly discusses the experimental design. Section 4  
14 presents the OSSE results and section 5 gives a conclusive discussion.

15

## 16   2. Study area and ocean model

17           In Malay, ‘selat’ means strait or channel. Selat Pauh is located to the west of the  
18 Singapore Strait and is surrounded by four major islands, Pulau Bukom, Pulau Busing,  
19 Pulau Sudong and Pulau Semakau. Selat Pauh is about 8 km long in west-east and 2 km  
20 wide in south-north (Fig. 1a). The maximum depth is about 30 meters. The circulation in  
21 the Selat Pauh is dominated by barotropic tidal flow with a weak vertical stratification. Its  
22 subtidal circulation is characterized with vortices induced by tide-island interactions and  
23 the wind driven circulation.

1           The model used in this study is an unstructured grid, three-dimensional, free  
2 surface, primitive-equation Finite Volume Coastal Ocean Model (FVCOM), developed  
3 by *Chen et al.* [2003; 2006a; 2006b] to which the reader is referred for details. The model  
4 grid covers the large Selat Pauh area and is closed by a circular open boundary (Fig.1a).  
5 Horizontal resolution is 20 m along the coast of the islands and increases to 500 m at the  
6 open boundary. The model is configured with 5 sigma levels in vertical. The model  
7 simulation is driven by realistic tide along the open boundary and wind stress at the  
8 surface. The tidal forcing is obtained from a regional Singapore Strait model and the wind  
9 stress is from 6-hour NCEP reanalysis which is uniform over the domain. The model is  
10 spun up under homogeneous condition and is integrated from January 3 to 12, 2007. No  
11 temperature and salinity are considered in this study.

12           We chose east Selat Pauh (hereafter referred as ESP domain) as assimilation  
13 domain to focus on small scales. The ESP domain is a 2.5 km long and 2 km wide west-  
14 east channel (Fig. 1b). The control run (true ocean) and the simulation with wrong initial  
15 conditions and/or wrong forcing (false ocean) are integrated with the large Selat Pauh  
16 domain (including the ESP domain), while the analysis procedure is conducted only in  
17 the ESP domain. The observations are 15 moorings shown in Fig. 1b.

18

### 19   3. Experimental design

20           The algorithm of EnKF is described in *Kalnay* [2003] to which the reader is  
21 referred for details. The EnKF theory is based on the statistical estimation of sufficient  
22 ensemble members. However, large ensemble size requires large computational resource  
23 and degrades its feasibility. In this study we use 20 ensemble members following *Chen et*

1 *al.* [2008], which has shown its efficiency in nonlinear tidal and buoyancy driven  
2 problems.

3 A localization scheme is necessary for coastal ocean data assimilations. First,  
4 from the statistical point of view the forecast error covariance contains large sampling  
5 errors amongst those distant grid points [*Anderson, 2001; Hamill et al. 2001*], and,  
6 theoretically, the localization scale should be consistent with the coherent scale of the  
7 flow structure. We use a localization scheme introduced by *Gaspari and Cohn* [1999]. A  
8 smooth function with 500 m cut-off radius is applied to each observation point when  
9 calculating the error covariance. The cut-off radius defines the utmost distance that an  
10 observation might affect the neighboring grid points.

11 OSSE with twin experiment approach provides a complete knowledge of the true  
12 ocean state and is widely used to assess the data assimilation method. All OSSE  
13 experiments in this study are listed in Table 1. The perfect model experiments examine  
14 the EnKF in reducing the initial perturbations with no model errors. There are three  
15 model runs: the true ocean, the false ocean and the assimilation. The true ocean is the  
16 unconstrained model simulation. The false ocean run is integrated in the same period but  
17 with a set of wrong initial conditions. In the assimilation run, velocity from the true ocean  
18 is assimilated every hour into the false ocean. The imperfect model experiments consider  
19 both initial and model errors. The true ocean is the same as the perfect model. In the false  
20 ocean the wind stress is perturbed to mimic model errors, i.e. the errors induced by  
21 inaccurate wind forcing. The perturbed wind is constructed from the true wind plus a  
22 weighted difference between the true wind and the wrong wind

23 
$$\mathbf{W}_i = \mathbf{W}_{true} + (\mathbf{W}_{wrong} - \mathbf{W}_{true}) \times (\alpha + \varepsilon_i) \quad i = 1, 2 \dots ens \quad (1)$$

1 where  $\alpha$  is a weighting factor to control the magnitude of the wind perturbation and  $\varepsilon$  is a  
2 small random perturbation applied to each ensemble member.

3

#### 4 4. OSSE results

5           The true ocean state is the model simulation of January 8 – 12, 2007. The  
6 circulation in the ESP domain is barotropic as no temperature and salinity are calculated  
7 in the model. The velocity field is mainly controlled by the boundary tidal forcing and  
8 slightly modified by the wind forcing. The false ocean run is initialized with a set of  
9 wrong initial conditions (ensemble members). The initial ensemble members are  
10 constructed from a random sampling of the model states of January 3 – 7. Each ensemble  
11 member is integrated forward without data assimilation and the mean of ensemble  
12 forecasts is regarded as the false ocean state and is used to calculate the RMS errors with  
13 respect to the true ocean. Fig 2a,b show the RMS errors of the false ocean (only first 48-  
14 hour results are showed). The initial velocity error decreases dramatically in the first 12  
15 hours ( $\sim 1$  tidal cycle) and remains at the level of  $10^{-3}$  cm/s. Since the model used for  
16 false ocean run is identical to the model of the true ocean run and thus the model is  
17 subject only to the initial errors, the rapid dissipation of initial errors is due to the  
18 enforcement of the ‘true’ boundary tidal forcing.

19           In the assimilation run, the model is constrained by the EnKF. Fig 2a,b also  
20 show the analysis RMS errors after data assimilation. The analysis RMS errors are  
21 smaller than the false ocean errors in the first 12 hours due to the data assimilation, while  
22 afterward they quickly converge to the false ocean errors implying the incapableness of  
23 the EnKF. As both the false ocean errors and the analysis errors decrease rapidly, to

1 reveal the error reduction due to the assimilation it is useful to calculate the normalized  
2 error reduction as  $[1 - \text{RMS}(\text{analysis}) / \text{RMS}(\text{false\_ocean})]$  (Fig 2c). The normalized error  
3 reduction is above 0.8 in the first few hours and gradually decreases to around zero,  
4 which indicates that EnKF fails to reduce the errors with respect to the false ocean after  
5 12 hours. The incapableness of EnKF is due to the rapid convergence of the ensemble  
6 members. Fig 2d shows the ensemble spread over the ESP domain. The spread decreases  
7 from  $10^{-2}$  cm/s to  $10^{-4}$  cm/s at hour 12 and remains afterward. This result is not surprising.  
8 Since we consider model perfect, all ensemble members are driven by the same boundary  
9 tidal forcing and the wind forcing. Therefore, the ensemble members eventually ‘forced’  
10 to the true ocean state. Even though after each analysis procedure the analysis is  
11 randomly perturbed to a set of new ensemble members used as initial conditions for the  
12 next assimilation step, it is showed that the initial perturbations are mostly swamped out  
13 by the dominant tidal forcing during the 1-hour model forecasting.

14 In the imperfect model experiments the model errors are mimicked through  
15 perturbations of the wind stress. The wrong wind is simply taken from a different time  
16 period. Fig 3a shows the difference (square root of  $[(\mathbf{U}_{\text{wrong}} - \mathbf{U}_{\text{true}})^2 + (\mathbf{V}_{\text{wrong}} - \mathbf{V}_{\text{true}})^2]$ )  
17 between the true wind and the wrong wind. The maximum difference is about 6 m/s at  
18 hour 36. In eq. (1)  $\varepsilon$  is a random number with zero mean and standard deviation of  $\langle \varepsilon \rangle$ .  
19 Fig 3b shows the RMS errors of the false ocean and the analysis with perturbation factors  
20  $\alpha = 0.5$  and  $\langle \varepsilon \rangle = 0.5$  (case 1). In the false ocean, the RMS error decreases in the first 12  
21 hours, and then gradually increases due to the inaccurate wind forcing. The analysis RMS  
22 error is significantly reduced throughout the 48 hours. The normalized error reduction  
23 indicates that the data assimilation reduces up to 90% errors with respect to the false

1 ocean after a few hours (Fig. 3c). A remarkable degradation of the error reduction at hour  
2 40 is likely due to the large wind stress perturbation. Fig 3d compares the ensemble  
3 spreads for the imperfect model and the perfect model. The ensemble spread of the  
4 imperfect model is well maintained between  $10^{-2}$  and  $10^{-3}$  cm/s and is about 10 times  
5 bigger than the perfect model.

6         The results of the imperfect model experiment (case 1) indicate that the success  
7 of the EnKF is highly sensitive to the ensemble spread and consequently the perturbation  
8 factors  $\alpha$  and  $\varepsilon$ . Fig 4 show the results of a series of sensitivity experiments with different  
9 values of  $\alpha$  and  $\varepsilon$  (cases 1 – 6 in Table 1). In the experiments of  $\varepsilon$ ,  $\alpha$  is set to 0.5 and  $\varepsilon = 0$ ,  
10 0.2 and 0.5. The normalized error reductions for  $\varepsilon = 0.2$  and 0.5 are similar, while the  
11 error reduction for  $\varepsilon = 0$  rapidly decreases to zero if no random perturbation is applied to  
12 the ensemble members. The ensemble spreads of cases 1 - 3 also confirm that the failure  
13 of the EnKF in the case 3 is due to the convergence of the ensemble members. In the  
14 experiments of  $\alpha$ ,  $\varepsilon$  is set to 0.2 and  $\alpha = 0, 0.2$  and 0.5. EnKF succeeds in all three cases  
15 and all ensemble spreads remain the order of  $10^{-3}$  cm/s. It is noted that the difference  
16 between the case 6 and the perfect model experiment is the presence of  $\varepsilon$  (0.2), which  
17 indicates that the incapableness of the EnKF in the perfect model can be essentially  
18 improved by applying a small random perturbation to each ensemble member. .

19

20

## 21 5. Summary and discussions

22         OSSE experiments are designed to assess the effectiveness of the EnKF data  
23 assimilation method in the Selat Pauh of Singapore. Perfect and imperfect model

1 experiments are carried out. 20 ensemble members are used, based on the successful  
2 experience in the idealized experiments [*Chen et al.*, 2008]. In the perfect model  
3 experiments, while the EnKF reduces the initial velocity error in the first few hours, it  
4 fails after 1 tidal cycle due to the rapid convergence of the ensemble members. In the  
5 imperfect model experiments, the model errors are mimicked through perturbations of the  
6 wind stress. The EnKF successfully reduces the velocity errors throughout the 48 hours  
7 assimilation run with the presence of a random perturbation ( $\varepsilon$ ) applied independently to  
8 each ensemble member. The assimilation-induced error reduction is up to 90% with  
9 respect to the false ocean.

10 Our model results are consistent with the EnKF theory that the success of the  
11 EnKF depends on the diversity of the ensemble members. Since our model only considers  
12 barotropic circulation, the flow evolution is badly controlled by the boundary tidal  
13 forcing. In the perfect model experiments, the tidal forcing is identical to the true ocean  
14 and thus the ensemble members converge rapidly in spite of initial perturbations. In the  
15 imperfect model, the tidal forcing is the same while the wind forcing is perturbed. Two  
16 perturbation factors ( $\alpha$  and  $\varepsilon$ ) are used to control the wind perturbations.  $\alpha$  controls the  
17 major difference between the true wind and the perturbed wind, and  $\varepsilon$  controls the  
18 magnitude of random perturbations to each ensemble member. The presence of  $\varepsilon$  is  
19 essential to the effectiveness of the EnKF as the EnKF succeeds in all the cases with  $\varepsilon$  but  
20 fails in the case without  $\varepsilon$ . The results also suggest that the success of EnKF is not  
21 seriously sensitive to the magnitude of  $\varepsilon$ , but the presence of  $\varepsilon$ .

22 *Houtekamer and Mitchell* [1998] found that the ensemble members tend to  
23 converge if the same observation is used to update the ensemble analysis. To prevent the

1 convergence of the ensemble members, they proposed to update the ensemble analysis  
2 with a perturbed set of observations. *Bergers et al.* [1998] and *Whitaker and Hamill*  
3 [2002] also discussed the need of perturbed observations. Our model results showed that  
4 for a barotropic regime, even with a set of perturbed observations, the ensemble members  
5 still converge rapidly in a few hours, which directly causes the failure of the EnKF in the  
6 perfect and imperfect (case 3) model experiments. On the other hand, a small random  
7 perturbation to each ensemble member effectively prevents the convergence of the  
8 ensemble members and essentially improves the effectiveness of the EnKF.

## Acknowledgment

1

2

3

4

5

This study was funded by the Office of Naval Research (ONR) and by the Singapore National Research Foundation (NRF) through the Singapore-MIT Alliance for Research and Technology (SMART) and Center for Environmental Sensing and Monitoring (CENSAM).

## 1 **References**

- 2 Anderson, J. L (2001), An ensemble adjustment filter for data assimilation. *Mon. Wea.*  
3 *Rev.*, 129, 2884-2903.
- 4 Burgers, G., P. J. van Leeuwen and G. Evensen (1998), Analysis scheme in the ensemble  
5 Kalman filter. *Mon. Wea. Rev.*, 126, 1719-1724.
- 6 Chen, C., H. Liu and R. C. Beardsley (2003), An unstructured, finite-volume, three-  
7 dimensional, primitive equation ocean model: application to coastal ocean and  
8 estuaries. *J. Atmos. Oceanic Tech.*, 20, 159-186.
- 9 Chen, C., R.C. Beardsley, and G. Cowles (2006a), An unstructured grid, finite-volume  
10 coastal ocean model-FVCOM user manual, School for Marine Science and  
11 Technology, University of Massachusetts Dartmouth, New Bedford, Second  
12 Edition. *Technical Report SMAST/UMASSD-06-0602*, 318pp.
- 13 Chen, C, R. C. Beardsley and G. Cowles (2006b), An unstructured grid, finite-volume  
14 coastal ocean model (FVCOM) system. Special Issue entitled “Advance in  
15 Computational Oceanography”, *Oceanography*, 19(1), 78-89.
- 16 Chen, C., P. Malanotte-Rizzoli, J. Wei, R. C. Beardsley, Z. Lai, P. Xue, S. Lyu, Q. Xu, J.  
17 Qi, and G. Cowles (2008), Application and comparison of Kalman filters for  
18 coastal ocean problems: An experiment with FVCOM. *J. Geophys. Res.*, in  
19 revision.
- 20 Evensen, G. (1994), Sequential data assimilation with a nonlinear quasi-geostrophic  
21 model using Monte Carlo methods to forecast error statistics. *J. Geophys. Res.*, 99,  
22 10,143-10,162.
- 23 Hamill, T. M., J. S. Whitaker and C. Snyder (2001), Distance-dependent filtering of

1 background error covariance estimates in an ensemble Kalman filter. *Mon. Wea.*  
2 *Rev.*, 129, 2776-2790.

3 Gaspri, G. and S. E. Cohn (1999), Construction of correlation functions in two and three  
4 dimensions. *Quart. J. Roy. Meteor. Soc.*, 125, 723-757.

5 Houtekamer, P. and H. L. Mitchell (1998), Data assimilation using an ensemble Kalman  
6 filter technique. *Mon. Wea. Rev.*, 126, 796-811.

7 Kalnay, E. (2003), Atmospheric modeling, data assimilation, and predictability.  
8 Cambridge University Press, 341pp.

9 Tippett, M. K., Anderson, J.L., Bishp, C.H., Hamill, T. and Whitaker, J.S. (2003),  
10 Ensemble square root filters. *Mon. Wea. Rev.*, 131, 1485-1490.

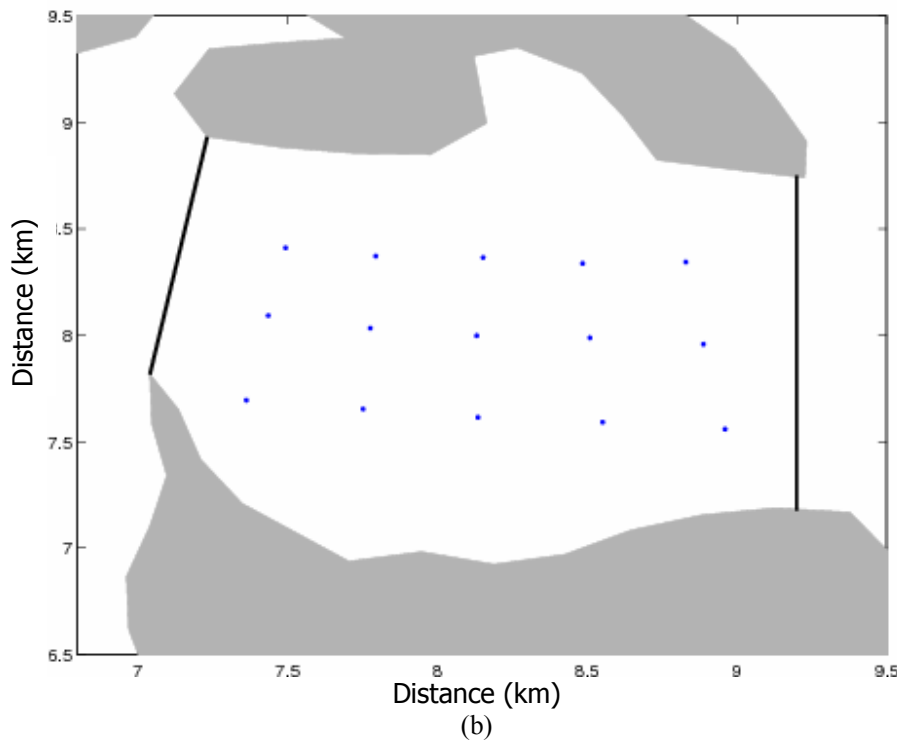
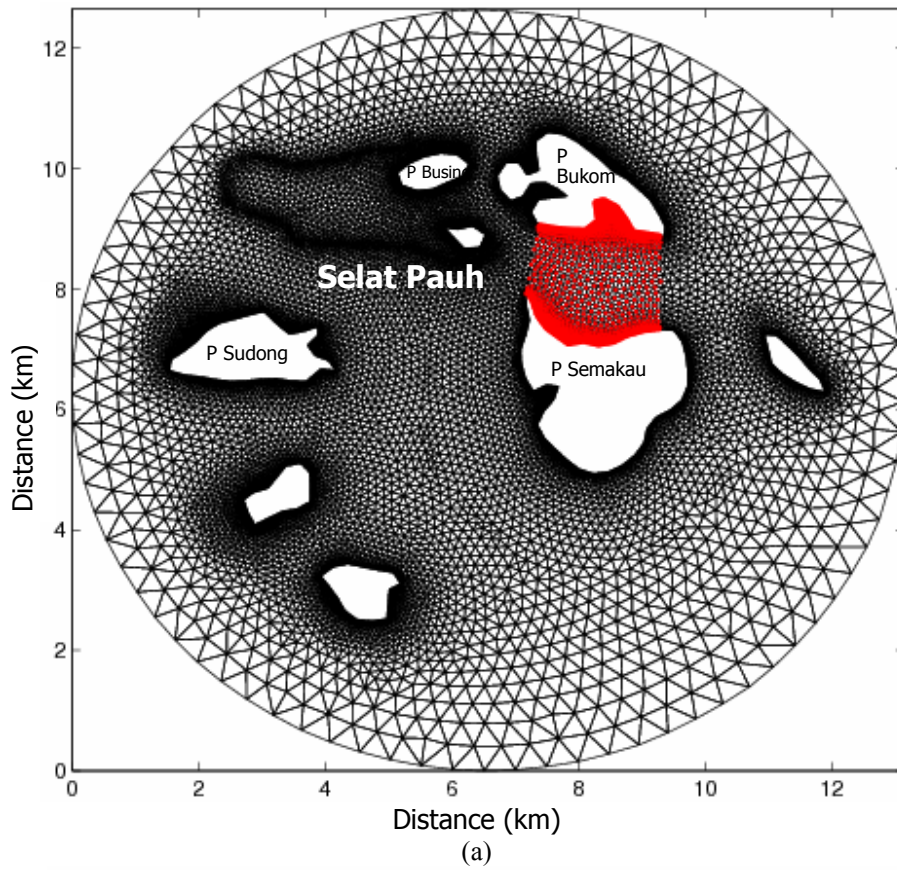
11 Whitaker, J. S. and T. M. Hamill (2002), Ensemble data assimilation without perturbed  
12 observations. *Mon. Wea. Rev.*, 130, 1913-1924.

13 Zang, X. and P. Malanotte-Rizzoli (2003), A comparison of assimilation results from the  
14 Ensemble Kalman filter and the Reduced-Rank Extended Kalman filter.  
15 *Nonlinear Processes in Geophysics*, 10, 6,477-6,491.

1 Table 1: Experiment list

<b>Perfect model</b>	True ocean	Unconstrained model simulation
	False ocean	Forecast with wrong initial conditions
	Assimilation	Assimilation with velocity from true ocean
<b>Imperfect model</b>	False ocean	Forecast with perturbed wind stress
	Experiments of $\varepsilon$	Case 1: $\alpha = 0.5, \langle \varepsilon \rangle = 0.5$
		Case 2: $\alpha = 0.5, \langle \varepsilon \rangle = 0.2$
		Case 3: $\alpha = 0.5, \langle \varepsilon \rangle = 0$
	Experiments of $\alpha$	Case 4: $\alpha = 0.5, \langle \varepsilon \rangle = 0.2$
		Case 5: $\alpha = 0.2, \langle \varepsilon \rangle = 0.2$
		Case 6: $\alpha = 0, \langle \varepsilon \rangle = 0.2$

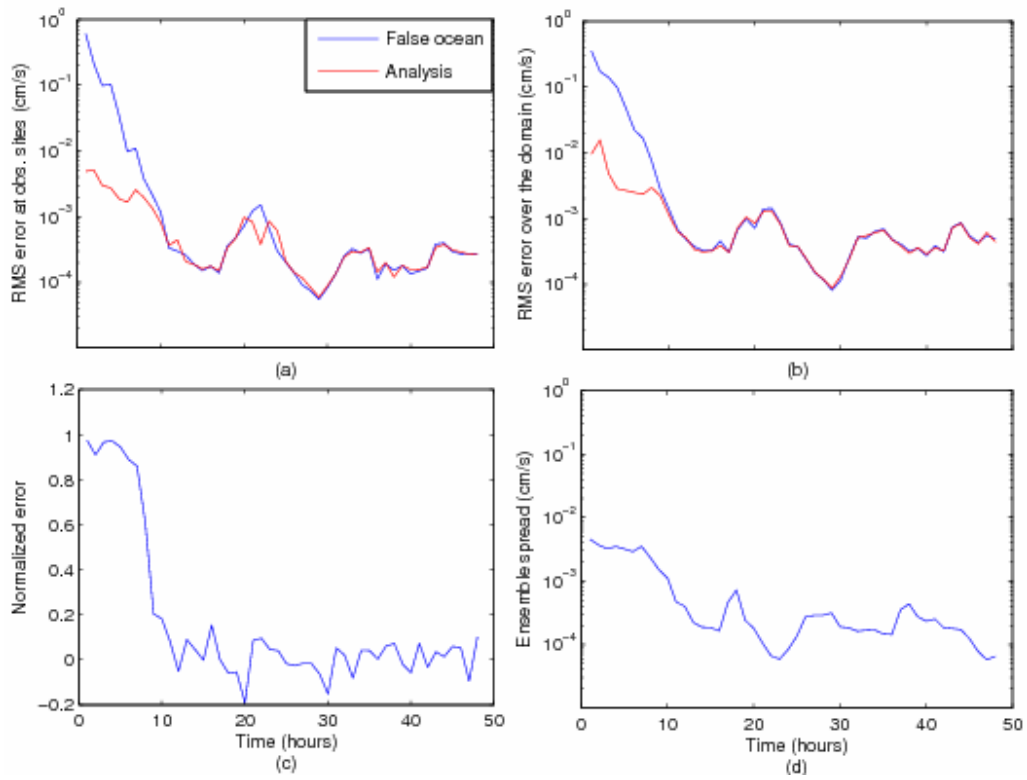
2



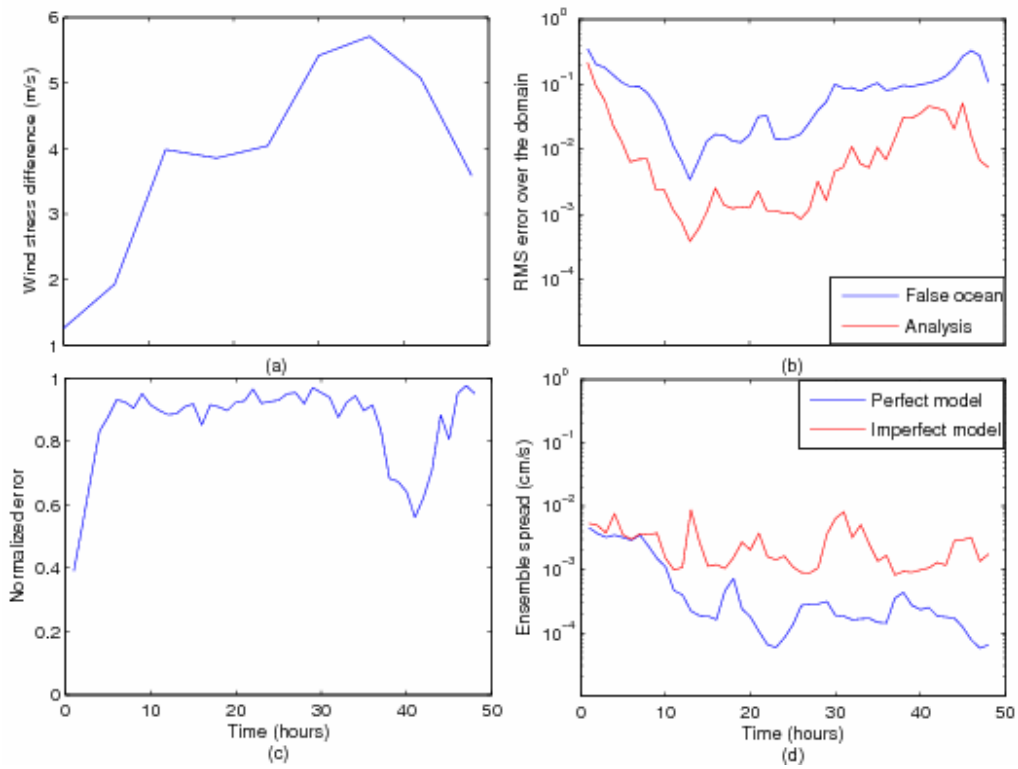
1  
2  
3

4  
5  
6  
7  
8

Fig. 1 (a) Large Selat Pauh area and the model grid. The east Selat Pauh (ESP) domain is marked as red region. (b) The ESP domain with 15 observational moorings.

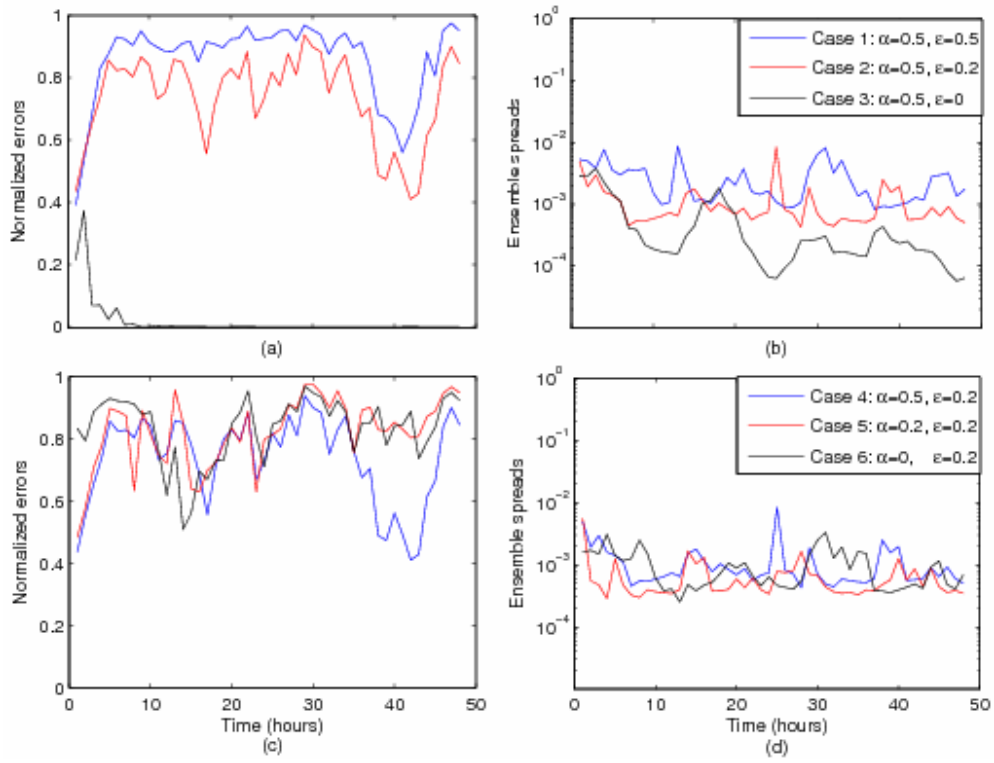


1  
 2  
 3 Fig. 2 Perfect model experiments. (a) RMS errors at observational sites; (b) RMS errors  
 4 over the domain; (c) Normalized error reduction  $[1 - \text{RMS}(\text{analysis}) / \text{RMS}(\text{false ocean})]$ ; (d)  
 5 Ensemble spread.



1  
2  
3  
4  
5  
6

Fig. 3 Imperfect model experiments. (a) Difference between the wrong wind and the true wind; (b) RMS errors over the domain; (c) Normalized error reduction; (d) Ensemble spreads.



1  
2  
3  
4  
5  
6

Fig. 4 Sensitivity cases 1 – 3: (a) Normalized error reductions and (b) Ensemble spreads. Sensitivity cases 4 – 6: (c) Normalized error reductions and (d) Ensemble spreads.