Preventing catastrophic filter divergence using adaptive additive inflation for baroclinic turbulence

Yoonsang Lee*, Andrew J. Majda and Di Qi

Center of Atmosphere and Ocean Science, Courant Institute of Mathematical Sciences, New York University, New York, New York

**Corresponding author address:* Yoonsang Lee, Courant Institute of Mathematical Sciences, 251 Mercer St., New York, NY 10012

E-mail: ylee@cims.nyu.edu

ABSTRACT

Ensemble based filtering or data assimilation methods have proved to be indispensable tools in atmosphere and ocean science as they allow computationally cheap, low dimensional ensemble state approximation for extremely high dimensional turbulent dynamical systems. For sparse, accurate and infrequent observations, which are typical in data assimilation of geophysical systems, ensemble filtering methods can suffer from catastrophic filter divergence which frequently drives the filter predictions to machine infinity. A two-layer quasi-geostrophic equation which is a classical idealized model for geophysical turbulence is used to demonstrate catastrophic filter divergence. The mathematical theory of adaptive covariance inflation by *Tong et al.* and covariance localization are investigated to stabilize the ensemble methods and prevent catastrophic filter divergence. Two forecast models, a coarse-grained ocean code, which ignores the small-scale parameterization, and stochastic superparameterization (SP), which is a seamless multi-scale method developed for large-scale models without scale-gap between the resolved and unresolved scales, are applied to generate large-scale forecasts with a coarse spatial resolution 48×48 compared to the full resolution 256×256. The methods are tested in various dynamical regimes in ocean with jets and vorticities, and catastrophic filter divergence is documented for the standard filter without inflation. Using the two forecast models, various kinds of covariance inflation with or without localization are compared. It shows that proper adaptive additive inflation can effectively stabilize the ensemble methods without catastrophic filter divergence in all regimes. Furthermore, stochastic SP achieves accurate filtering skill with localization while the ocean code performs poorly even with localization.

1. Introduction

Ensemble based filters or data assimilation methods, including the ensemble Kalman filter (EnKF; Evensen 2003) and ensemble square root filters, the ensemble transform Kalman filter (ETKF; Bishop et al. 2001) and the ensemble adjustment Kalman filter (EAKF; Anderson 2001), provide accurate statistical estimation of a geophysical system combining a forecast model and observations. These methods quantify the uncertainty of the system using an ensemble which samples the information of the system. For geophysical systems which are complex high-dimensional and thus require enormously huge computational costs for long time integration, the ensemble based methods are indispensable tools for data assimilation as the methods allow computationally cheap and low dimensional state approximation. Due to the simplicity and efficiency of the ensemble based filters, these methods are widely applied to various fields of geophysical science such as numerical weather prediction (Kalnay 2003).

Despite their successful applications in geophysical applications, ensemble based filters suffer from small ensemble size due to the high dimensionality and expensive computational costs (frequently referred as "curse of ensemble numbers" (Majda and Harlim 2012)) which can lead to filter divergence. Sampling errors due to insufficient ensemble size and imperfect model errors often yield underestimation of the uncertainty in the forecast and thus filters trust the forecast with larger confidence than the information given by observations. Inaccurate uncertainty quantification in the forecast fails to track the true signal and thus filter performance degrades, which is called filter divergence (Majda and Harlim 2012). Also insufficient ensemble size can lead to spurious overestimation of cross correlations between otherwise uncorrelated variables (Hamill et al. 2001; Whitaker et al. 2009; Sakov and Oke 2008) which also affect filter performance. Covariance inflation, which inflates the prior covariance and pulls the filter back toward observations, is one among various methods to remedy the filter divergence (Anderson 2001). For the overestimation of cross correlations between uncorrelated variables, localization which multiplies the covariances between prior state variables and observation variables by a correlation function with local support is a powerful method to correct the overestimated cross correlations (Houtekamer and Mitchell 2001).

Catastrophic filter divergence (Harlim and Majda 2010; Gottwald and Majda 2013) is another important issue hindering the applications of the ensemble based methods to high dimensional systems especially in the case with sparse and infrequent observations and small observation errors. Catastrophic filter divergence drives the filter predictions to machine infinity although the underlying system remains in a bounded set. In data assimilation of geophysical systems in the ocean, observations are often sparse and infrequent. In observations of ocean dynamics such as sea surface temperature, observations become accurate using various techniques such as tropical moored buoys, ocean reference status, and surface drifting buoys. But observations are still inadequate and sparse to sample over the vast surface and the interior of the ocean.

It is shown rigorously in Kelly et al. 2015 that catastrophic filter divergence is not caused by numerical instability, instead the analysis step of filters generates catastrophic filter divergence. Although covariance inflation and localization stabilize filters and improve accuracy, they cannot avoid catastrophic filter divergence. In Harlim and Majda 2010, it is demonstrated that ensemble based methods with constant covariance inflation still suffer from catastrophic filter divergence. In this study we also see that covariance localization decreases the occurrence of catastrophic filter divergence but does not prevent catastrophic divergence.

To avoid catastrophic filter divergence, a judicious model error using linear stochastic models was studied in Harlim and Majda 2010 with skillful results in some parameter regimes. Recently a simple remedy of catastrophic filter divergence without using linear stochastic models has been proposed through rigorous mathematical arguments and tested for the Lorenz-96 model in Tong et al. 2016. The approach in Tong et al. 2016 adaptively inflates covariance with minimal additional costs according to the distribution of the ensemble. The strength of inflation is determined by two statistics of the ensemble, 1) ensemble innovation which measures how far predicted observations are from actual observations and 2) cross covariance between observed and unobserved variables (see (12) and (13) in Section 3 respectively). If the filter is malfunctioning based on these two statistics, inflation is triggered and becomes larger when filters stray further into malfunction.

In this study we demonstrate catastrophic filter divergence of ensemble based filters in the two-layer quasi-geostrophic equations, which are classical idealized models for geophysical turbulence (Salmon 1998). The adaptive inflation method is then proposed for this two-layer system to avoid catastrophic filter divergence. Both a coarse-grained ocean code, which ignores the subgrid scale parameterization, and stochastic superparameterization (Grooms et al. 2015b), which is a seamless multi-scale method developed for large-scale models without scale-gap between the resolved and unresolved scales, are applied to generate forecasts with a coarse spatial resolution 48×48 for each layer compared to the full resolution 256×256 which generates true signals. We test ensemble methods for various dynamical regimes in the ocean corresponding to idealized low, mid and high latitude states and document that catastrophic filter divergence occurs for ensemble based methods even with localization unless adaptive inflation is applied. Ensemble filtering for the two-layer quasi-geostrophic equations using these forecast models, the ocean code and stochastic superparameterization, has already been studied in Grooms et al. 2015a to investigate the effect of constant inflation on accounting for model errors without catastrophic filter divergence. In this study we test a very sparse observation network which observes only 4×4 points of the upper layer stream function with a small observation error variance corresponding to 1% of the total variance of the stream function to represent the typical realistic scenario with sparse high quality data, which leads to catastrophic filter divergence.

Using both the ocean code and stochastic superparameterization, various kinds of covariance inflation with or without localization are compared. We verify that proper adaptive covariance inflation can effectively stabilize the ensemble based filters uniformly without catastrophic filter divergence in all test regimes. Furthermore, stochastic superparameterization achieves accurate filtering skill with localization while the ocean code performs poorly even with localization.

The structure of this paper is as follows. In Section 2 we briefly review an ensemble method, the Ensemble Adjustment Kalman Filter (Anderson 2001) with covariance inflation and localization. The adaptive inflation method to prevent catastrophic filter divergence is described in Section 3 including how to choose parameters of the adaptive method. In Section 4 the two-layer quasi-geostrophic equation with baroclinic instability is described and two coarse-grained forecast models, the ocean code and stochastic superparameterization, are explained. Numerical experiments with various inflation strategies with or without localization are reported in Section 5 along with stabilized and improved filtering results using the adaptive inflation method. In Section 6 we conclude this paper with discussion.

2. Ensemble Filtering

In this section we briefly describe the Ensemble Adjustment Kalman Filter (EAKF; Anderson 2001) which in our experience is a more stable and accurate scheme than other popular ensemble based methods (Majda and Harlim 2012). We assume that the true signal is generated by a nonlinear mapping $\psi_n : \mathbb{R}^d \to \mathbb{R}^d$

$$u_n = \psi_n(u_{n-1}) \tag{1}$$

where $u_n \in \mathbb{R}^d$ is a state vector at the n-th observation time. We consider a linear observation of u_n by an observation operator $H : \mathbb{R}^d \to \mathbb{R}^q$ with a rank q

$$z_n = Hu_n + \xi_n \tag{2}$$

where ξ_n is a mean zero Gaussian noise with a variance σ independent in different times and space grid points. For an easy exposition of the adaptive inflation in Section 3 we use a decomposition of the state variable u_n into observed and unobserved variables $x_n \in \mathbb{R}^q$ and $y_n \in \mathbb{R}^{d-q}$ respectively so that $x_n = Hu_n$ and $y_n \in Ker(H)$.

As other ensemble based filters, EAKF uses ensemble members $\{v_n^{(k)}\}_{k=1}^K$ to represent statistical properties of the state but uses only the first and second order moments (that is, mean and covariance) to update each ensemble member. First, EAKF generates prior predictions by solving a forecast model for each ensemble member

$$\tilde{v}_n^{(k)} = \widetilde{\psi}_n(v_{n-1}^{(k)}), \quad k = 1,2,...,K \tag{3}$$

where $\widetilde{\psi}_n$ is an approximate forecast model to the true dynamics ψ_n. From the forecast ensemble $\{\tilde{v}_n^{(k)}\}_{k=1}^K$, the prior mean $\overline{v}_n^f$ and covariance C_n^f are given by

$$\overline{v}_n^f = \frac{1}{K}\sum_{k=1}^K \tilde{v}_n^{(k)} \tag{4}$$

and

$$C_n^f = \frac{1}{K-1}\sum_{k=1}^K \left(\tilde{v}_n^{(k)} - \overline{v}_n^f\right) \otimes \left(\tilde{v}_n^{(k)} - \overline{v}_n^f\right) \tag{5}$$

respectively. With these prior mean and covariance, the standard Kalman formula using observation $z_n \in \mathbb{R}^q$ gives the following posterior mean and covariance

$$\overline{v}_n^a = \overline{v}_n^f - C_n H^T (I + H^T C_n H)^{-1} (H\overline{v}_n^f - z_n). \tag{6}$$

and

$$C_n^a = C_n^f - C_n^f H^T (I + H^T C_n^f H)^{-1} H C_n^f \tag{7}$$

respectively. For a ensemble perturbation matrix $V \in \mathbb{R}^{d\times K}$ whose k-th column is given by the ensemble perturbation $\delta v_n^{(k)} = \tilde{v}_n^{(k)} - \overline{v}_n^f$, EAKF finds an adjustment matrix $A_n \in \mathbb{R}^{K\times K}$ so that the adjusted ensemble satisfies the posterior covariance (7)

$$\frac{1}{K-1} V_n A_n \otimes V_n A_n = C_n - C_n H^T (I + H^T C_n H)^{-1} H C_n. \tag{8}$$

Once the adjustment matrix A_n is calculated, the posterior ensemble is obtained by adding the adjusted perturbation to the posterior mean. That is $v_n^{(k)} = \overline{v}_n^a + s_n^{(k)}$ where $s_n^{(k)}$ is the k-th column of the ensemble perturbation matrix $V_n A_n$.

Covariance inflation overcomes some problems caused by sampling errors due to insufficient ensemble numbers or an imperfect model and requires only a minimal additional cost to the original EAKF. The covariance inflation introduces more uncertainty in the prior covariance so that the filter has more weight on the information given by observations. That is, for a constant λ_n which determines the strength of inflation, covariance inflation inflates the prior covariance

$$C_n^f \leftarrow (I+\lambda_n)C_n^f. \tag{9}$$

for multiplicative inflation or

$$C_n^f \leftarrow C_n^f + \lambda_n I \tag{10}$$

for additive inflation. Then the ensemble is modified to satisfy the inflated prior covariance by spreading the ensemble for the multiplicative inflation and by adding additional noise for the additive inflation. Although covariance inflation improves filter skill in many applications it is reported that constant inflation does not prevent catastrophic filter divergence with sparse and accurate observation networks (Harlim and Majda 2010).

3. Adaptive Additive Inflation

A simple remedy in Tong et al. 2016 to stabilize ensemble based filters by preventing catastrophic filter divergence is to adaptively trigger the inflation and change the strength λ_n. Although the adaptive inflation method of Tong et al. 2016 works both for the multiplicative inflation (9) and additive inflation (10), we focus on the simpler additive inflation in this study. The inflation strength λ_n of (10) is determined by two statistics of the ensemble

$$\lambda_n = c_a \Theta_n (1+\Xi_n) \mathbb{1}_{\{\Theta_n > M_1 \text{ or } \Xi_n > M_2\}} \tag{11}$$

where c_a is a tunable positive constant, Θ_n is a measure related to the innovation process $H\tilde{v}_n^{(k)} - z_n$ in a standard Kalman filter

$$\Theta_n := \frac{1}{K}\sum_{k=1}^{K} \|H\tilde{v}_n^{(k)} - z_n\|^2, \tag{12}$$

Ξ_n is the l_2 norm of the cross covariance between the observed and unobserved variables

$$\Xi_n = \left\| \frac{1}{K-1}\sum_{k=1}^{K} \left(\tilde{x}_n^{(k)} - \bar{x}_n\right) \otimes \left(\tilde{y}_n^{(k)} - \bar{y}_n\right) \right\|, \quad \tilde{v}_n^{(k)} = (\tilde{x}_n^{(k)}, \tilde{y}_n^{(k)}), \tilde{x}_n^{(k)} = H\tilde{v}_n^{(k)} \tag{13}$$

and M_1 and M_2 are fixed positive thresholds to decide whether the filter is performing well or not. The first statistical information Θ_n measures the accuracy of the prediction, that is, how far the predicted observations are from actual observations. The second statistical information Ξ_n is an important factor because large cross covariance can magnify a small error in the observed component and impose it on the unobserved variables. Hence the adaptive inflation can be regarded as a control of these two statistics to prevent catastrophic filter divergence. Note that these two factors are in fact derived from a rigorous mathematical argument for nonlinear stability of finite ensemble filters which can be found in Tong et al. 2016.

In contrast to the conventional covariance inflation which modifies the prior ensemble to satisfy the inflated covariance, the EAKF with adaptive additive inflation does not modify the prior ensemble to inflate covariance; the additive inflation can make the rank of the posterior covariance larger than or equal to d while its rank cannot exceed $K-1$ where K is the ensemble size. Thus, in adaptive additive inflation, we use the inflated prior covariance (10) to calculate the posterior mean while the posterior covariance does not change. That is, instead of (6), the posterior mean is defined as

$$\bar{v}_n^a = \bar{v}_n^f - \tilde{C}_n^f H^T (I + H^T \tilde{C}_n^f H)^{-1} (H\bar{v}_n^f - z_n). \tag{14}$$

using

$$\tilde{C}_n^f = C_n^f + \lambda_n I \tag{15}$$

where the posterior covariance is the same as (7), that is, no inflated prior covariance.

The two thresholds M_1 and M_2 of (11) are important factors as they differentiate poor forecasts from properly working forecasts. Using an elementary benchmark of accuracy which should be surpassed by filters, we use the following aggressive thresholding (Tong et al. 2016). The thresholds are given by

$$M_1 = \|H\|^2 \text{Error}_{bench} + 2q\sigma \tag{16}$$

and

$$M_2 = \frac{K}{2K-2} \text{Error}_{bench} \tag{17}$$

where the benchmark for accuracy Error_{bench} is the mean-square error of an estimator using an invariant probability measure of the model

$$\text{Error}_{bench} := \mathbb{E}(u_n - \mathbb{E}(u_n|z_n)|^2. \tag{18}$$

As the invariant measure of the model is not available, aggressive thresholding uses a Gaussian approximation to the invariant measure using climatological properties, mean and covariance. Then the conditional distribution given observation z_n is a Gaussian measure and can be computed exactly which gives the following formula

$$\text{Error}_{bench} = \text{tr}\left(cov(u_n) - cov(u_n)H^T(I + Hcov(u_n)H^T)^{-1}Hcov(u_n)\right). \tag{19}$$

4. Model equations and forecast models

In atmosphere and ocean science, quasi-geostrophic equations are widely used as classical idealized models of geophysical turbulence (Salmon 1998). In this study we use a two-layer quasi-geostrophic equation as the model equation to observe catastrophic filter divergence in high dimensional data assimilation and test the adaptive additive inflation to prevent catastrophic filter divergence. The system is maintained by baroclinic instability imposed by vertical shear flows and shows interesting features in geophysical turbulence such as inverse cascade of energy and zonal jets. After describing the model equation in Section a, two coarse-grained forecast models, an ocean code which ignores the subgrid scales and another forecast method with stochastic parameterization of the subgrid scales, are explained in Section b.

a. Two-layer quasi-geostrophic equations

Our model equation to generate high dimensional geophysical turbulence is the following two-layer quasi-geostrophic equation in a doubly periodic domain used in Grooms and Majda 2014; Majda and Grooms 2014; Grooms et al. 2015a; Lee et al. 2016 to generate baroclinic turbulence

$$\begin{aligned}
\partial_t q_1 &= -\mathbf{v}_1\cdot\nabla q_1 - \partial_x q_1 - (k_\beta^2 + k_d^2)v_1 - \nu\Delta^4 q_1,\\
\partial_t q_2 &= -\mathbf{v}_2\cdot\nabla q_2 + \partial_x q_2 - (k_\beta^2 - k_d^2)v_2 - r\Delta\psi_2 - \nu\Delta^4 q_2,\\
q_1 &= \Delta\psi_1 + \frac{k_d^2}{2}(\psi_2 - \psi_1),\\
q_2 &= \Delta\psi_2 - \frac{k_d^2}{2}(\psi_2 - \psi_1).
\end{aligned} \tag{20}$$

Here q_j is the potential vorticity in the upper ($j = 1$) and lower ($j = 2$) layers, k_d is the deformation wavenumber, r is a linear Ekman drag coefficient at the bottom layer of the flows, k_β is an nondimensional constant resulting from the variation of the vertical projection of Coriolis frequency with latitude and the velocity field $\mathbf{v}_j = (u_j, v_j) = (-\partial_y\psi_j, \partial_x\psi_j)$ for the stream function ψ_j. To stabilize the equation by absorbing a downscale cascade of enstrophy at the smallest scales while leaving other scales nearly inviscid for interesting dynamics at large-scales, we use a hyper-dissipation $\Delta^4 q_j$ with a hyperviscosity ν, which is commonly used in turbulence simulations. To maintain nontrivial dynamics of (20) by baroclinic instability, a large-scale zonal vertical shear is applied with equal and opposite unit velocities which are related to the terms $(-1)^j(\partial_x q_j + k_d^2 v_j)$ in (20).

Following the experiments in Grooms et al. 2015a and Lee et al. 2016, we test three different regimes corresponding to low, mid and high latitude ocean models by changing the β-plane effect k_β and the bottom drag r (see Table 1 for the parameter values of the three test regimes). While the deformation wavenumber k_d is fixed at 25, we use a fine resolution of 256×256 grid points for each layer to generate true signals in our data assimilation experiments. The hyperviscosity ν is set to 1.28×10^{-15} and we use a pseudo-spectral space discretization while the time integration uses a fourth order semi-implicit Runge-Kutta method by incorporating an exponential integration for the linear stiff dissipation term. Time step is fixed at 2×10^{-5} for all test regimes.

In the high latitude case (or the f-plane case), the quasi-geostrophic equation is dominated by spatially homogeneous and isotropic flows (see Figure 1 for snapshots of the upper and lower layer stream function). In the mid and low latitude cases which have the β-plane effect, the flows organize into inhomogeneous and anisotropic structure such as zonal jets.

b. Forecast models with and without stochastic parameterization

As a forecast model in data assimilation of the true signal given by (20), we consider two forecast models on a low resolution 48×48 grid points, 1) an ocean code which uses only a coarse grid without parameterizing the small scales and 2) stochastic superparameterization which parameterizes the effect of the small scales by modeling the small scales as randomly oriented plane waves (Majda and Grooms 2014; Grooms et al. 2015b). Note that these two forecast models are imperfect models as they approximate the true signal on a low resolution grid. Thus in data assimilation using ensemble based methods, there is an error from the imperfect model in addition to the sampling error due to a small ensemble size.

The first forecast model, which we call the ocean code, solves the following approximation to (20) which replaces the hyper dissipation by a biharmonic dissipation of relative vorticity $\omega_j =$

$\Delta\psi_j$

$$\begin{aligned}
\partial_t q_1 &= -\mathbf{v}_1 \cdot \nabla q_1 - \partial_x q_1 - (k_\beta^2 + k_d^2) v_1 - \nu_2 \Delta^2 \omega_1, \\
\partial_t q_2 &= -\mathbf{v}_2 \cdot \nabla q_2 + \partial_x q_2 - (k_\beta^2 - k_d^2) v_2 - r\Delta\psi_2 - \nu_2 \Delta^2 \omega_2.
\end{aligned} \tag{21}$$

This replacement is to mimic the biharmonic dissipation commonly used in eddy-permitting ocean models (Griffies and Hallberg 2000). By analogy with ocean models and some atmospheric models, the ocean code also uses the second order energy- and enstrophy-conserving Arakawa finite differencing (Arakawa 1966) for the nonlinear advection terms $\mathbf{v}_j \cdot \nabla q_j, j = 1,2$. For time integration, we use a second order Runge-Kutta integration with the same exponential integrator for the linear stiff term and a time step fixed at 5×10^{-4}.

We consider another forecast model called stochastic superparameterization which uses stochastic parameterization of the subgrid scales using randomly oriented plane waves for the subgrid scales. The subgrid scales are generally not zero and influence the evolution of the resolved scales. Especially in quasi-geostrophic turbulence which includes regimes with a net transfer of kinetic energy from small to large scales (Charney 1971), it is important to accurately model the effects of the under-resolved eddies to obtain accurate properties of the system such as energy spectrum. Stochastic superparameterization is developed as a multiscale model for turbulence without scale-gap between the resolved and unresolved scales (Grooms and Majda 2014; Majda and Grooms 2014). Among various versions of stochastic superparameterization, we use the most recent version developed in Grooms et al. 2015b to deal with arbitrary boundary conditions using finite difference numerics for the large scales.

The stochastic superparameterization forecast model solves (21) using the same second order finite differencing for the nonlinear term but with additional terms $SGS_j, j = 1,2$ obtained from stochastic subgrid scale parameterization

$$\begin{aligned}
\partial_t q_1 &= -\mathbf{v}_1 \cdot \nabla q_1 - \partial_x q_1 - (k_\beta^2 + k_d^2) v_1 - \nu_2 \Delta^2 \omega_1 + SGS_1, \\
\partial_t q_2 &= -\mathbf{v}_2 \cdot \nabla q_2 + \partial_x q_2 - (k_\beta^2 - k_d^2) v_2 - r\Delta\psi_2 - \nu_2 \Delta^2 \omega_2 + SGS_2.
\end{aligned} \tag{22}$$

The parameterization terms $SGS_j, j = 1,2$ are computed by modeling the subgrid scale as randomly oriented plane waves. Under this modeling of the subgrid scales, stochastic superparameterization replaces the nonlinear terms of the subgrid scale equation using additional damping and white noise forcing which yields quasilinear equation conditional to the resolved scale variable. We also use the method in Grooms et al. 2015b to impose temporal correlations in the parameterization by using a Wiener process model for the orientation of the plane waves. Because the subgrid scales are solved in formally infinite domains, this approach has no scale-gap between the resolved and subgrid scales. Also, the stochastic modeling of the subgrid scales generate the missing instability of the subgrid scales using deterministic parameterization of the subgrid scales. Note that we use the same time integration as in the ocean code, thus the difference between the ocean code and stochastic superparameterization comes from the parameterization terms $SGS_j, j = 1,2$.

Figure 2 shows the time averaged kinetic energy (KE) spectra

$$\mathrm{KE} = \frac{1}{2} \int |\nabla \psi_1|^2 + |\nabla \psi_2|^2$$

by the direct numerical method (black), stochastic superparameterization (blue) and the ocean code (red) using biharmonic viscosity $\nu_4 = 1.0 \times 10^{-7}$ and $\nu_4 = 1.6 \times 10^{-4}$ obtained by tuning to match the energy spectra. Although the ocean code has much weaker dissipation than stochastic superparameterization, the ocean code has smaller energies than stochastic superparameterization while stochastic superparameterization captures the correct large-scale kinetic energy spectra; the small energy of the ocean code cannot be improved further by tuning the biharmonic viscosity coefficient. This result implies that the ocean code could have filter divergence by inappropriately capturing the uncertainty in the forecast due to small energy of the resolved scales. On the other hand, it is shown that stochastic parameterization can act to reduce model error (Shutts 2005; Frenkel et al. 2012) and it increases ensemble spread which yields an effect similar to covariance inflation. In the next section, we will see that stochastic superparameterization requires smaller covariance inflation than the ocean code as the ocean code has large model errors which cannot be improved by covariance inflation.

5. Catastrophic filter divergence and numerical experiments

In this section we demonstrate catastrophic filter divergence for all three test regimes regardless of the two forecast models, the ocean code and stochastic superparameterization, with sparse high quality observations which are infrequent in time. Catastrophic filter divergence is effectively prevented using the adaptive additive inflation for both forecast methods. Stochastic superparameterization achieves accurate filtering skill with localization while the ocean code fails to achieve accurate skill even with localization.

a. Filtering setup

For EAKF, we use a sequential update of observations used in Anderson 2001 which avoids explicit computation of the SVD in (8) by processing observations individually. The true signal is given by a fine resolution solution of (20) using a resolution 256×256 for each layer and a fourth order semi-implicit Runge-Kutta integration with a time step 2×10^{-5}. The two forecast models, the ocean code and stochastic superparameterization, use the same coarse resolution 48×48 for each layer and a second order semi-implicit Runge-Kutta with a time step 5×10^{-4}.

We observe only the upper layer stream function, analogous to observation of sea surface height, on a sparse 4×4 uniform grid while the stream function in the lower layer is completely unobserved. Observation error variances correspond to 1% of the stream function variance for each test regime. Following the idea of Keating et al. 2012, the eddy turnover time $T_{eddy} = 2\pi Z^{-1/2}$ (where Z is the time-averaged total enstrophy $q_1^2 + q_2^2$) is comparable to 0.006 for all test regimes and we use infrequent observations with an observation interval 0.008. Note that using the time step of the forecast models, 5×10^{-4}, this observation interval requires 16 time integrations for each forecast step. The ensemble size is 17 and thus the prior covariance is not necessarily rank deficient. This number is small compared to the dimension of the forecast model which is general in real data assimilation.

For each filtering test, we run 1000 assimilation cycles and take the last 600 cycles to measure filter performance using the time averaged RMS error (RMSE)

$$\text{time averaged RMSE} := \frac{1}{600} \sum_{n=401}^{1000} \|\bar{v}_n - u_n\| \tag{23}$$

and pattern correlation (PC)

$$\text{time averaged PC} := \frac{1}{600} \sum_{n=401}^{1000} \frac{\langle \bar{v}_n, u_n \rangle}{\|\bar{v}_n\| \|u_n\|} \tag{24}$$

respectively where $\langle , \rangle$ is the l_2-inner product.

For covariance inflation, we test several combinations of inflation methods - for the inflation strength λ_n in (10), no inflation (noI) $\lambda_n = 0$, constant inflation (CI) $\lambda_n = c_c$ for a constant c_c, adaptive inflation (AI) λ_n by (11) and constant+adaptive inflation (CAI)

$$\lambda_n = c_c + c_a \Theta_n (1 + \Xi_n) \mathbb{1}_{\{\Theta_n > M_1 \text{ or } \Xi_n > M_2\}} \tag{25}$$

(see Table 2 for the tuned c_c and c_a used in this study). The thresholds for adaptive inflation are given by the aggressive thresholding (16) and (17) where the benchmark for accuracy, Error_{bench} is given by 10, 166 and 155 for the low, mid and high latitude cases respectively from the reference simulations. Along with these inflation methods, we also use covariance localization with the compactly supported fifth-order piecewise rational function from Gaspari and Cohn 1999. The localization radius (where influence of observation is zero) is set to 8 forecast gird points. The distance between two adjacent observation points are 12 and thus the square region centered at each observation point is marginally updated from the other observation points.

b. Filter experiments - catastrophic filter divergence and stabilization

If no inflation is applied, EAKF has catastrophic filter divergence for both forecast models. Figure 3 shows a sequence of snapshots of the low latitude case upper layer stream function by the ocean code without inflation and localization (observation points are marked with black circles). At the 570th cycle, the filter still works capturing the meridional structure of the low latitude case but as more cycles go on, instability develops at unobserved grid points which eventually diverges to machine infinity after the 600th cycle. The first row of Figure 4 shows time series of the RMS errors by each forecast method when they suffer from catastrophic filter divergence. The RMS errors increase gradually but they eventually diverge to machine infinity. The two forecast models run slightly longer with localization but localization fails to prevent catastrophic filter divergence. The second row of Figure 4 shows time series of RMS errors with the constant+adaptive inflation where the cycles at which adaptive inflation is triggered is marked with dots. In the ocean code case with no localization, the adaptive inflation is triggered at the beginning and stops although the filter still degrades. Inflation is triggered again when the filter fails to capture the true signal. The ocean code with localization triggers the adaptive inflation most of the time and obtains a stable result but also fails to achieve accurate filtering skill. In the stochastic superparameterization case with adaptive inflation and localization, adaptive inflation is triggered only 99 times out of 1000 cycles where most of the adaptive inflation is triggered at the beginning and infrequently triggered later as the filter is performing well.

The occurrence percentage of catastrophic filter divergence out of 100 different runs is in Table 3. With no localization and inflation, the filter suffers from catastrophic filter divergence more than 75% for both the ocean code and stochastic superparameterization. The constant inflation stabilizes the filter slightly but it does not prevent catastrophic filter divergence perfectly. The constant inflation (CI) with no localization has a higher percentage of divergence than the no inflation case for the stochastic superparameterization forecast model. Through stochastic parameterization of subgrid scales, stochastic superparameterization has more variability than the ocean code and thus additional constant inflation is not necessary.

Adaptive inflation (AI) with and without localization significantly decreases the number of occurrence of catastrophic filter divergence but the ocean code fails to prevent catastrophic filter divergence entirely. For the constant+adaptive inflation (CAI), all methods are stable even without localization. Note that for stochastic superparameterization, both AI and CAI work well preventing catastrophic filter divergence while the ocean code fails to prevent the divergence in the AI case. As we discussed before, stochastic superparameterization has enough ensemble spread through stochastic parameterization of the subgrid scales and thus when adaptive inflation is already applied, constant inflation plays a marginal role in improving filter skill.

For the stabilized filters with the constant+adaptive inflation, we compare the filter performance using the time averaged posterior RMS errors and pattern correlations (the performance difference between the adaptive inflation (AI) and constant+adaptive inflation (CAI) is marginal when there is no catastrophic filter divergence). In the low latitude case (shown in Table 4), both the ocean code and the superparameterization methods fail to achieve accurate filtering skill without localization. The RMS errors are larger than the standard deviation of the stream function and both forecast methods do not capture the correlation with the true signal. When localization is combined with adaptive inflation, it helps to increase filtering skill for both methods. The superparameterization has significantly improved results; RMS error is smaller than 50% of the standard deviation of the stream function and pattern correlation is larger than 90% for both layers. Although the lower layer stream function is completely unobserved, the adaptive filter achieves accurate filter skill. The ocean code result is improved using localization but it still suffers from standard filter divergence with RMS error larger than the standard deviation of the stream function.

In the mid latitude case, the superparameterization still has skillful filtering skill and is superior to the ocean code although the perofrmance is slightly degraded compared to the low latitude case as the mid latitude is more turbulent than the low latitude case. The RMS error by superparameterization with adaptive inflation and localization is about 30% smaller than the standard deviation and pattern correlations are larger than 75% (see Table 5 for the mid latitude case RMS errors and pattern correlations). On the other hand, the ocean code does not show any significant skill even with adaptive inflation and localization. In the mid latitude case, the ocean code using adaptive inflation displays comparable results with and without localization, and both fail to achieve meaningful filtering results. For the superparameterization, on the other hand, significantly improvement in filter skill can be achieved using localization (see the second row of Figure 5 the time series of RMS errors with adaptive inflation). As the RMS errors are more fluctuating than the low latitude case, the adaptive inflation is triggered most of the time for all combination of inflation and localization.

The last test regime, high latitude case, is the most difficult test case as it is strongly turbulent and dominated by homogeneous and isotropic vortical flows with no spatial structure. In this test regime, stochastic superparameterization with constant+adaptive inflation and localization still achieves a smaller RMS error and a larger pattern correlation than the ocean code though the improvement by superparameterization is more marginal. The observed upper layer RMS error is 10% smaller than the standard deviation while the unobserved lower layer RMS error is only 5% smaller than the standard deviation (Table 6).

6. Conclusions

Ensemble based filtering methods are indispensable tools in atmosphere and ocean science as they provide computationally cheap and low dimensional ensemble state estimation for extremely high dimensional turbulent systems. But these methods can suffer from catastrophic filter divergence which drives the forecast predictions to machine infinity especially when the observation is sparse, accurate and infrequent although the underlying true signal remains bounded. Using an idealized model for the geophysical turbulence of the ocean, the two-layer quasi-geostrophic equation with baroclinic instability, and a sparse observation network which is general in real applications, we were able to see catastrophic filter divergence of the ensemble adjustment Kalman filter, which is one of the most stable and accurate ensemble methods.

The constant covariance inflation and localization, which are widely used methods to account for the sampling errors due to insufficient ensemble size and model errors from imperfect forecast models, stabilize the filter but fail to prevent the catastrophic filter divergence. Increasing the observation size or ensemble number can help to prevent catastrophic filter divergence but this approach is practically prohibitive and sometimes impossible as it requires enormous amount of financial and computer resources to cover the vast surface of the ocean. Instead we followed the adaptive inflation approach of Tong et al. 2016 to prevent catastrophic filter divergence. The adaptive approach requires a minimal additional computational cost compared to the standard ensemble based methods and uses only two low order statistics of the ensemble, the ensemble innovation and cross covariance between observed and unobserved variables.

We tested the adaptive inflation using two forecast models, the ocean code without parameterization of the subgrid scales and stochastic superparamterization which parameterizes the subgrid scales by modeling them as randomly oriented plane waves. Although both forecast models are stabilized with the adaptive inflation, stochastic superparameterization displays filtering skill superior to the ocean code. When the ensemble method is combined with localization and adaptive inflation, stochastic superparameterization achieves RMS errors smaller than the climatological error while the ocean code still suffers from the standard filter divergence with RMS errors comparable to the climatological error.

As we have shown in this study, covariance inflation is an important and useful technique in ensemble based methods to improve filtering skill. There are another class of adaptive inflation techniques such as Anderson 2007 and Ying and Zhang 2015. Although the adaptive inflation in Tong et al. 2016 is based on rigorous mathematical arguments, it would be interesting to test other adaptive inflation methods to avoid catastrophic filter divergence like the blended filter (Majda et al. 2014; Qi and Majda 2015) that combines a particle filter in a low-dimensional subspace and efficient Kalman filter in the orthogonal part. As it is investigated in Harlim and Majda 2010 through a linear stochastic model for the forecast, a judicious model error could be alternative to prevent catastrophic filter divergence.

Acknowledgments. The research of A.J. Majda is partially supported by Office of Naval Research grant ONR MURI N00014-12-1-0912 and DARPA 25-74200-F4414. Y. Lee is supported as a postdoctoral fellow by these grants. D. Qi is supported as a graduate research assistant by the ONR grant.

References

Anderson, J. L., 2001: An ensemble adjustment kalman filter for data assimilation. *Mon. Weather Rev.*, **129**, 2884–2903.

Anderson, J. L., 2007: An adaptive covariance inflation error correction algorithms for ensemble filters. *Tellus*, **59**, 210–224.

Arakawa, A., 1966: Computational design for long-term numerical integration of the equations of fluid motion: Two-dimensional incompressible flow part i. . *Comput. Phys.*, **1**, 119–143.

Bishop, C. H., B. J. Etherton, and S. J. Majumdar, 2001: Adaptive sampling with the ensemble transform kalman filter part i : theoretical aspects. *Mon. Weather Rev.*, **129**, 420–436.

Charney, J. G., 1971: Geostrophic turbulence. *J. Atmospheric Sci.*, **28**, 1087–1095.

Evensen, G., 2003: The ensemble kalman filter : theoretical formulation and practical implementation. *Ocean dynamics*, **53**, 343–367.

Frenkel, Y., A. J. Majda, and B. Khouider, 2012: Using the stochastic multicloud model to improve convective parameterization: A paradigm example. *J. Atmos. Sci.*, **69**, 1080–1105.

Gaspari, G., and S. Cohn, 1999: Construction of correlation functions in two and three dimensions. *Quart. J. Roy. Meteor. Soc.*, **125**, 723–757, doi:10.10002/qj.49712555417.

Gottwald, G. A., and A. J. Majda, 2013: A mechanism for catastrophic filter divergence in data assimilation for sparse observation networks. *Nonlinear Processes in Geophysics*, **20**, 705–712.

Griffies, S. M., and R. W. Hallberg, 2000: Biharmonic friction with a smagorinsky-like viscosity for use in large-scale eddy-permitting ocean modes. *Mon. Weather Rev.*, **128**, 2935–2946.

Grooms, I., Y. Lee, and A. J. Majda, 2015a: Ensemble filtering and low resolution model error: Covariance inflation, stochastic parameterization, and model numerics. *Mon. Wea. Rev.*, **143**, 3912–3924.

Grooms, I., Y. Lee, and A. J. Majda, 2015b: Numerical schemes for stochastic backscatter in the inverse cascade of quasigeostrophic turbulence. *Multiscale Modeling and Simulation*, **13**, 1001–1021.

Grooms, I., and A. J. Majda, 2014: Stochastic superparameterization in quasigeostrophic turbulence. *Journal of Computational Physics*, **271**, 78–98.

Hamill, T. M., J. S. Whitaker, and C. Snyder, 2001: Distance-dependent filtering of background covariance estimates in an ensemble kalman filter. *Mon. Weather Rev.*, **129**, 2776–2790.

Harlim, J., and A. J. Majda, 2010: Catastrophic filter divergence in filtering nonlinear dissipative systems. *Comm. Math. Sci.*, **8**, 27–43.

Houtekamer, P. L., and H. L. Mitchell, 2001: A sequential ensemble kalman filter for atmospheric data assimilation. *Mon. Weather Rev.*, **129**, 123–137.

Kalnay, E., 2003: *Atmospheric modeling, data assimilation, and predictability*. Cambridge university press.

Keating, S., A. J. Majda, and K. Smith, 2012: New methods for estimating ocean eddy heat transport using satellite altimetry. *Mon. Weather Rev.*, **140**, 1703–1722, doi:10.1175/MWR-D-11-00145.1.

Kelly, D., A. Majda, and X. T. Tong, 2015: Concrete ensemble kalman filters with rigorous catastrophic filter divergence. *PNAS*, **112**, 10 589–10 594, doi:10.1073/pnas.1511063112.

Lee, Y., A. J. Majda, and D. Qi, 2016: Stochastic superparameterization and multiscale filtering of turbulent tracers, preprint, submitted to Journal of Computational Physics.

Majda, A. J., and I. Grooms, 2014: New perspectives on superparameterization for geophysical turbulence. *Journal of Computational Physics*, **271**, 60–77.

Majda, A. J., and J. Harlim, 2012: *Filtering complex turbulent systems*. Cambridge university press, Cambridge UK.

Majda, A. J., D. Qi, and T. P. Sapsis, 2014: Blended particle filters for large-dimensional chaotic dynamical systems. *PNAS*, **111**, 7511–7516, doi:10.1073/pnas.1405675111.

Qi, D., and A. J. Majda, 2015: Blended particle methods with adaptive subspaces for filtering turbulent dynamical systems. *Physica D*, **298–299**, 21–41, doi:10.1016/j.physd.2015.02.002.

Sakov, P., and P. Oke, 2008: A deterministic formulation of the ensemble kalman filter : An alternative to ensemble square root filters. *Tellus*, **60**, 415–428.

Salmon, R., 1998: *Lectures on Geophysical Fluid Dynamics*. Oxford university press, Oxford, UK.

Shutts, G., 2005: A kinetic energy backscatter algorithm for use in ensemble prediction systems. *Quart. J. Roy. Meteor. Soc.*, **131**, 3079–3102.

Tong, X. T., A. J. Majda, and D. Kelly, 2016: Nonlinear stability of the ensemble kalman filter with adaptive covariance inflation. *Comm. Math. Sci*, in press, arXiv:1507.08319.

Whitaker, J. S., G. P. Compo, and J. N. Thepaut, 2009: A comparison of variational and ensemble-based data assimilation systems for reanalysis of sparse observation. *Mon. Weather Rev.*, **137**, 1991–1999.

Ying, Y., and F. Zhang, 2015: An adaptive covariance relaxation method for ensemble data assimilation. *Quarterly Journal of the Royal Meteorological Society*, **141**, 2898–2906.

LIST OF TABLES

TABLE 1: Parameters of (20) for three test regimes. Other parameters are fixed at $\nu = 1.28 \times 10^{-15}$ and $k_d = 25$

	k_β	r
Low	$k_d^2/2$	0.5
Mid	$k_d^2/4$	2
High	0	8

TABLE 2: Constant and adaptive inflation parameters c_c and c_a for each test regime

	Ocean		SP	
	c_c	c_a	c_c	c_a
Low	3×10^{-3}	5×10^{-4}	1×10^{-4}	5×10^{-4}
Mid	2×10^{-3}	4×10^{-4}	2×10^{-4}	1×10^{-5}
High	3×10^{-3}	1×10^{-4}	1×10^{-3}	1×10^{-5}

	Low		Mid		High	
no localization	Ocean	SP	Ocean	SP	Ocean	SP
noI	78%	84%	98%	97%	90%	85%
CI	63%	87%	80%	76%	45%	57%
AI	3%	0%	2%	0%	5%	0%
CAI	0%	0%	0%	0%	0%	0%

	Low		Mid		High	
with localization	Ocean	SP	Ocean	SP	Ocean	SP
noI	40%	24%	19%	38%	44%	64%
CI	15%	11%	9%	12%	22%	8%
AI	1%	0%	0%	0%	0%	0%
CAI	0%	0%	0%	0%	0%	0%

TABLE 3: Occurrence percentage of catastrophic filter divergence out of 100 different runs with and without localization. No inflation (noI), constant (CI), adaptive (AI) and constant+adaptive (CAI) inflation methods.

TABLE 4: Low latitude case. Stream function estimation for both layers. Posterior RMS errors and pattern correlations in parenthesis

no localization	Ocean, CAI	SP, CAI	Std of stream ftn
Upper layer	4.35 (0.05)	3.18 (0.02)	3.21
Lower layer	4.40 (0.01)	3.24 (0.04)	3.07

with localization	Ocean, CAI	SP, CAI	Std of stream ftn
Upper layer	3.76 (0.62)	1.37 (0.93)	3.21
Lower layer	3.90 (0.50)	1.41 (0.91)	3.07

TABLE 5: Mid latitude case. Stream function estimation for both layers. Posterior RMS errors and pattern correlations in parenthesis

no localization	Ocean, CAI	SP, CAI	Std of stream ftn
Upper layer	11.99 (0.26)	11.62 (0.30)	12.59
Lower layer	12.58 (0.23)	12.13 (0.29)	12.03
with localization	Ocean, CAI	SP, CAI	Std of stream ftn
Upper layer	11.24 (0.35)	7.73 (0.78)	12.59
Lower layer	12.26 (0.23)	8.14 (0.77)	12.03

TABLE 6: High latitude case. Stream function estimation for both layers. Posterior RMS errors and pattern correlations in parenthesis

no localization	Ocean, CAI	SP, CAI	Std of stream ftn
Upper layer	14.12 (0.25)	11.95 (0.22)	12.71
Lower layer	14.67 (0.24)	12.98 (0.18)	12.21
with localization	Ocean, CAI	SP, CAI	Std of stream ftn
Upper layer	13.01 (0.31)	11.38 (0.43)	12.71
Lower layer	13.21 (0.26)	11.53 (0.42)	12.21

LIST OF FIGURES

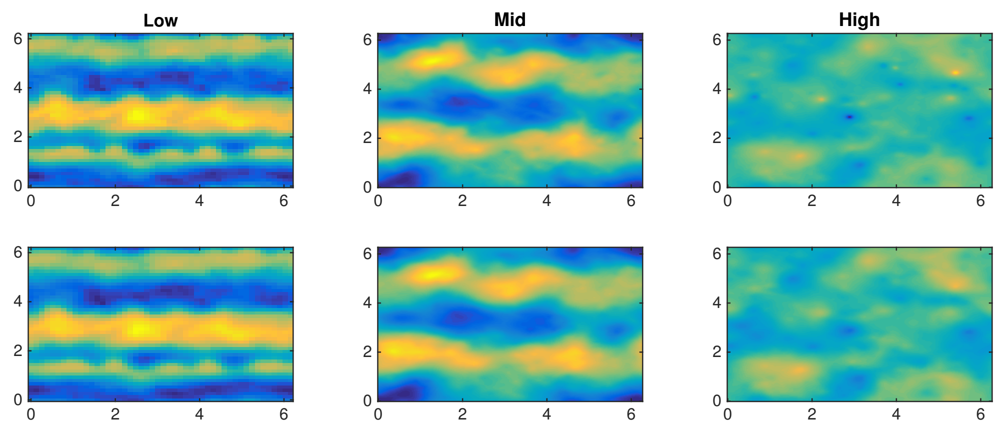


FIG. 1: Snapshots of stream functions ψ_j, $j = 1,2$. Upper layer (top row) and lower layer (bottom row). Low (first column), mid (second column) and high (third column) latitude cases.

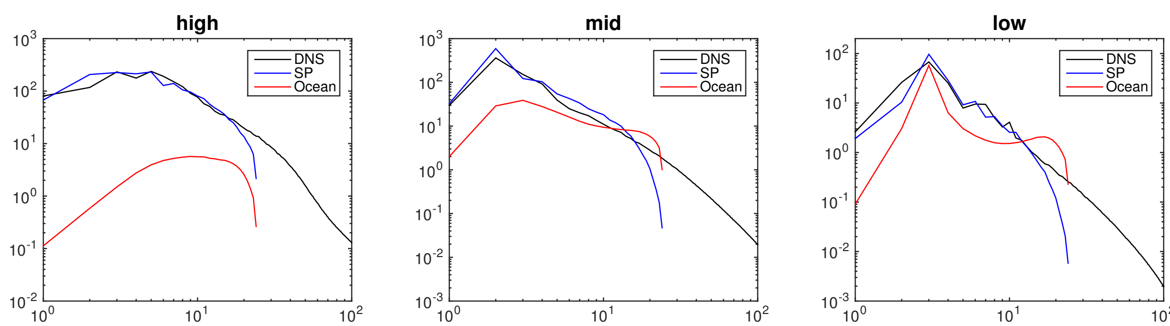
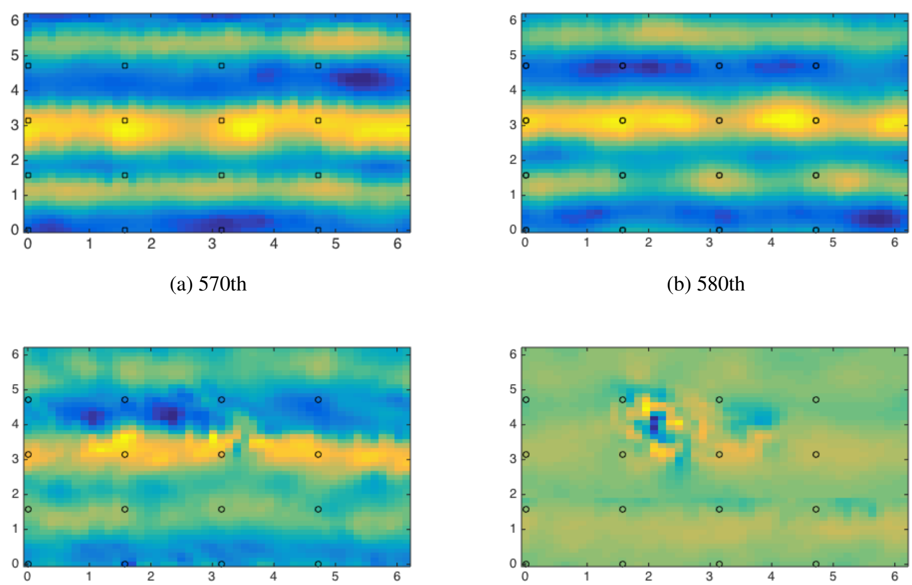


FIG. 2: Time averaged total kinetic energy (KE) spectra by direct numerical reference (black), stochastic superparameterization (blue) and ocean code (red)



(a) 570th (b) 580th

(c) 590th (d) 600th

FIG. 3: Low latitude case. Snapshots of posterior upper layer stream functions by Ocean code at 570th, 580th, 590th, and 600th cycles. Observation points are marked with circles. Catastrophic filter divergence is invoked after the 600th cycle.

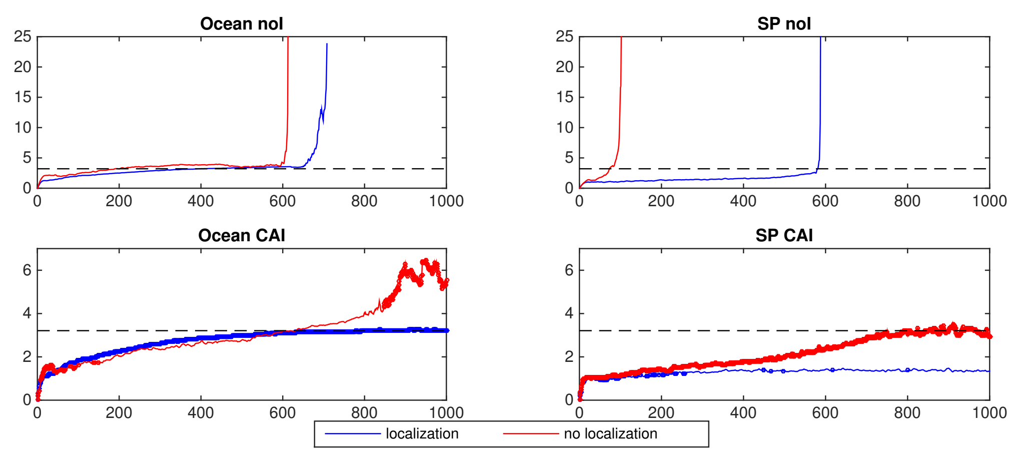


FIG. 4: Low latitude case. Time series of upper layer RMS error. The cycles at which inflation is triggered are marked with filled circles. Standard deviation of the stream function in dash line.

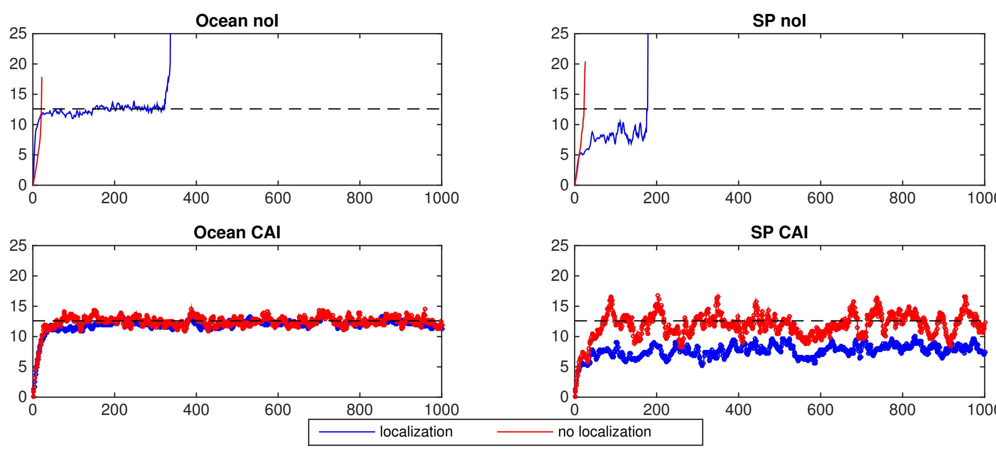


FIG. 5: Mid latitude case. Time series of upper layer RMS error. The cycles at which inflation is triggered are marked with filled circles. Standard deviation of the stream function in dash line.

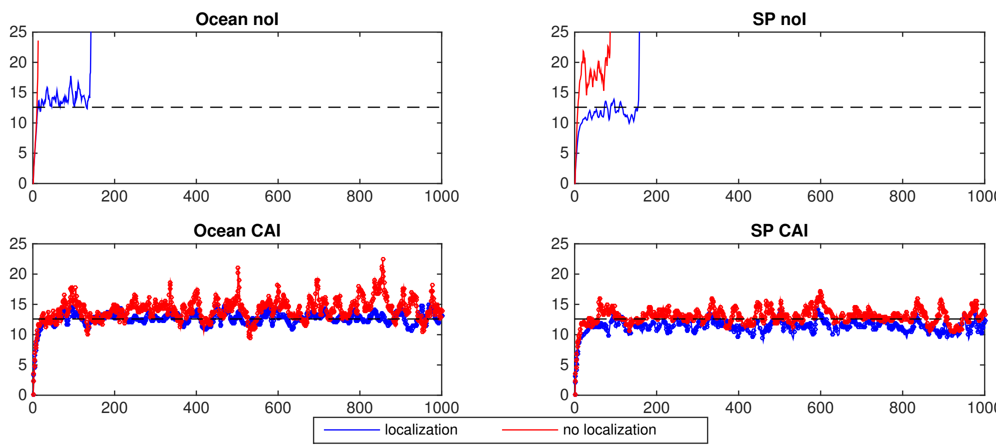


FIG. 6: High latitude case. Time series of upper layer RMS error. The cycles at which inflation is triggered are marked with filled circles. Standard deviation of the stream function in dash line.