Ensemble filtering and low resolution model error: Covariance inflation, stochastic parameterization, and model numerics

I. GROOMS,* Y. LEE, AND A. J. MAJDA

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

ABSTRACT

The use of under-resolved models in ensemble data assimilation schemes leads to two kinds of model errors: truncation errors associated with discretization of the large-scale dynamics, and errors associated with interactions with subgrid scales. Multiplicative and additive covariance inflation can be used to account for model errors in ensemble Kalman filters, but they do not reduce the model error. Truncation errors can be reduced by increasing the accuracy of the numerical discretization of the large-scale dynamics, and subgrid scale parameterizations can reduce errors associated with subgrid scale interactions. Stochastic subgrid scale parameterizations both reduce the model error and inflate the ensemble spread, so their effectiveness in ensemble assimilation schemes can be gaged by comparing with covariance inflation techniques. We compare the effects of covariance inflation, stochastic parameterizations, and model numerics in two-layer periodic quasigeostrophic turbulence on an f-plane and on a β-plane. Covariance inflation improves the performance of a benchmark model with no parameterizations and second-order numerics. Fourth-order spatial discretization and the stochastic parameterizations, alone and in combination, are superior to covariance inflation. In our experiments fourth-order numerics and stochastic parameterizations lead to similar levels of improvement in filter performance even though the climatology of models without stochastic parameterizations is poor.

1. Introduction

Data assimilation is the process of combining dynamical models and observations to estimate the state of a dynamical system (Kalnay 2002; Evensen 2009; Majda and Harlim 2012). There are many data assimilation algorithms, and ensemble Kalman filters (EnKFs; Evensen 1994) are a class of algorithms that are particularly amenable to applications in atmospheric and oceanic data assimilation (e.g. Houtekamer et al. 2005; Whitaker et al. 2008; Szunyogh et al. 2008; Houtekamer et al. 2009). Computational costs prevent ensemble data assimilation systems from using models that resolve all the active scales of the atmosphere or ocean. The use of low resolution models leads to two kinds of error: model error and representation error, the latter being associated with the contribution of subgrid scales to the observations. The present investigation is focused on model errors associated with low resolution models.

Model error can come from many sources, e.g. incorrect parameters; the focus here is on two kinds of model error associated with low resolution: numerical/truncation error in modeling large-scale self interaction, and error associated with subgrid scale interactions. The true state of the dynamical system can be partitioned into a large-scale part that is represented on the low resolution model grid, and a subgrid scale part that is not represented. The low resolution model attempts to predict the evolution of the large-scale part. Even when the subgrid scale part is zero, there are model errors associated with truncation errors in the numerical discretization of the large-scale dynamics; this type of model error can be reduced by increasing the accuracy of the numerical discretization. Of course, the subgrid scales are generally not zero and they influence the evolution of the large scales. Subgrid scale parameterizations attempt to model the effect of the subgrid scales on the large scales, and improving subgrid scale parameterizations can reduce large-scale model error. Nevertheless, because the state of the subgrid scales is not precisely known there is in principle a nonzero minimum of possible model error due to the uncertainty concerning the state of the subgrid scales; information barriers of this type are discussed e.g. by Branicki and Majda (2012).

Model errors need to be accounted for in ensemble data assimilation algorithms, and one common approach in the context of EnKFs is covariance inflation, either multiplicative (Anderson and Anderson 1999) or additive (Mitchell and Houtekamer 2000). Multiplicative inflation multiplies the prior ensemble perturbations by an inflation factor that can vary in space and time. The inflation factor can be

*Corresponding author address: Ian Grooms, Courant Institute of Mathematical Sciences, 251 Mercer St., New York, NY 10012
E-mail: grooms@cims.nyu.edu

hand-tuned, which involves considerable computational expense, or it can be adaptively estimated as part of the filtering algorithm (e.g. Anderson 2007, 2009; Li et al. 2009; Miyoshi 2011). Additive inflation is implemented by drawing random samples from a specified model error distribution, and adding these to the prior ensemble before the analysis update. It assumes that model error is independent of the model state, which is convenient but typically erroneous. Methods of estimating the model error distribution are discussed by, e.g. Zupanski and Zupanski (2006). Additive inflation is better suited to accounting for model error than multiplicative inflation (Whitaker and Hamill 2012), though the latter can mitigate the effects of many different kinds of errors in ensemble filtering, like sampling errors. Covariance inflation techniques attempt to account for model error in the analysis, but they do not reduce the model error.

Additive inflation is analogous to stochastic parameterization, the difference being that random perturbations are added to the forecast before every assimilation cycle, versus being added at every time step of the numerical time integration scheme. In contrast to additive inflation, stochastic parameterizations can act to *reduce* model error (e.g. Berner et al. 2012; Frenkel et al. 2012). Since stochastic parameterizations typically increase ensemble spread they have an effect similar to covariance inflation, and can thus be viewed as accounting for as well as reducing model error. Unlike additive inflation, stochastic parameterization accounts for model error without assuming that it is independent of forecast error: even when the stochastic noise terms added to the model are independent of the model state, their integrated effect depends on the model.

A well-designed stochastic parameterization should outperform covariance inflation, so the latter can be used as a benchmark for evaluating stochastic parameterizations in the context of ensemble data assimilation (Whitaker and Hamill 2012). Houtekamer et al. (2009) compared additive inflation and stochastic parameterization in a low resolution ensemble filtering context and found that additive inflation had a significant positive impact on performance, while a stochastic backscatter parameterization (SKEBS; Berner et al. 2009) had no positive impact. Whitaker and Hamill (2012) found that additive inflation and SKEBS had similarly positive impacts on the performance of an ensemble assimilation scheme.

We perform ensemble data assimilation experiments in the idealized setting of two-layer, doubly-periodic quasigeostrophic (QG) turbulence with observations of the top layer streamfunction, analogous to observations of sea surface height. Stochastic subgrid scale parameterizations have been developed for this setting by Grooms and Majda (2013, 2014, hereafter GM13 and GM14) and Grooms et al. (2015, hereafter GLM15). GM13 and GM14 developed 'stochastic superparameterization' (SP) for this setting (the connection of stochastic SP with the the approach of Randall et al. (2013) is discussed by Majda and Grooms (2014)); stochastic SP generates a stochastic forcing of quasigeostrophic potential vorticity conditional on the local large-scale variables. GLM15 extended the stochastic SP algorithm to include temporal correlation, and developed a simplified backscatter scheme that is independent of the state of the large-scale variables.

In this setting we find that additive inflation has no positive impact, uniform (non-adaptive) multiplicative inflation has a positive impact, and stochastic parameterization has a greater positive impact on assimilation performance. We also find that moving from second-order to fourth-order discretization leads to significant improvement in the performance of the assimilation algorithm, comparable to the effect of the stochastic parameterizations. The use of fourth-order numerics in combination with the stochastic parameterization yields our best results, slightly better than the use of the stochastic parameterization alone.

The configuration of the high-resolution truth model and of several imperfect low resolution models is described in section 2, and the configuration of the ensemble assimilation system is described in section 3. The results of the assimilation experiments are presented in section 4, and conclusions are offered in section 5.

2. Two-Layer Quasigeostrophic Model Configuration

The high resolution model used in these experiments is essentially the same as the model used in GM13, GM14, and GLM15, the primary difference being that the domain used here is one quarter the size, to reduce computational cost. The model evolves the solution of the nondimensional two-equal-layer quasigeostrophic (QG) equations in a doubly-periodic domain forced by an imposed zonal baroclinic shear. The governing equations are

$$\begin{aligned}
\partial_t q_1 &= -\nabla\cdot(\boldsymbol{u}_1 q_1) - \partial_x q_1 - (k_\beta^2 + k_d^2)v_1 - \nu_8\nabla^8 q_1,\\
\partial_t q_2 &= -\nabla\cdot(\boldsymbol{u}_2 q_2) + \partial_x q_2 - (k_\beta^2 - k_d^2)v_2 - r\nabla^2\psi_2 - \nu_8\nabla^8 q_2\\
q_1 &= \nabla^2\psi_1 + \frac{k_d^2}{2}(\psi_2 - \psi_1),\\
q_2 &= \nabla^2\psi_2 - \frac{k_d^2}{2}(\psi_2 - \psi_1),
\end{aligned} \tag{1}$$

where q_j is the potential vorticity in the upper ($j = 1$) and lower ($j = 2$) layers, the velocity-streamfunction relation is $u_j = -\partial_y\psi_j$, $v_j = \partial_x\psi_j$, k_d is the deformation wavenumber (k_d^{-1} is the deformation radius), the coefficient r specifies the strength of linear bottom friction (Ekman drag) and ν_8 is the hyperviscous Reynolds number. The dynamics can also be described in terms of barotropic and baroclinic modes, the former being given by the vertical average $q_t = (q_1 + q_2)/2 = \nabla^2\psi_t$ and the latter by the vertical difference $q_c = (q_1 - q_2)/2 = (\nabla^2 - k_d^2)\psi_c$. Subscripts t

and c are used throughout to denote barotropic and baroclinic components, respectively.

The simulations are carried out in a square periodic domain of half the width of the domains used in GM13, GM14, and GLM15, but with the same grid resolution; the computational grid has 256×256 points. The deformation wavenumber is $k_d = 25$, and the hyperviscous Reynolds number is $\nu_8 = 1.28 \times 10^{-15}$. We consider two of the three parameter regimes from GM14: f-plane with $k_\beta = 0$ and $r = 8$, and β-plane with $k_\beta^2 = k_d^2/2$ and $r = 0.5$. The experiments were also performed for the third parameter regime in GM14, with $k_\beta^2 = k_d^2/4$, but the results were qualitatively similar to the β-plane case considered here and are omitted for brevity.

The f-plane scenario is dominated by small-scale vortices with spatially homogeneous statistics. In the β-plane scenario the flow organizes into three zonal jets (see figure 1c) that act as a barrier to meridional transport. The meridional heat flux is defined as

$$\text{Heat Flux} \equiv \iint v_t \psi_c \mathrm{d}x\mathrm{d}y \tag{2}$$

where v_t is the barotropic meridional velocity and ψ_c is the baroclinic streamfunction. The dynamics generate a heat flux in response to the potential vorticity gradient associated with the imposed zonal baroclinic shear. In the f-plane case the (nondimensional) climatological heat flux is 442, and in the β-plane case the climatological heat flux is 2.3. The massive reduction in heat flux in the β-plane case is partly due to the strong zonal jets, and is partly due to the fact that the β-plane case is less energetic. GM14 showed that in both scenarios the heat flux is generated by scales resolved on the coarse model grid.

Imperfect Models

We consider several different imperfect models on a low resolution grid of 48×48 points. One model, denoted 'Spectral' in the results, simply uses the same governing equations and spectral discretization as the perfect model (with no subgrid scale parameterization), but uses a different value of the hyperviscous Reynolds number ν_8, tuned to produce optimal climatology. The f-plane scenario uses $\nu_8 = 1.28 \times 10^{-10}$ and the β-plane scenario uses $\nu_8 = 6.4 \times 10^{-11}$.

The remaining imperfect models all use biharmonic vorticity diffusion instead of the hyperviscous potential vorticity diffusion of the perfect model, in order to mimic the biharmonic viscosity commonly used in eddy-permitting ocean models (following GLM15). The remaining imperfect models thus solve the following alternative equations

$$\begin{aligned}\partial_t q_1 = &-\nabla\cdot(\boldsymbol{u}_1 q_1) \\ &- \partial_x q_1 - (k_\beta^2 + k_d^2)v_1 - \nu_4\nabla^4\omega_1 + \mathrm{SGS}_1, \end{aligned}\tag{3}$$

$$\begin{aligned}\partial_t q_2 = &-\nabla\cdot(\boldsymbol{u}_2 q_2) \\ &+ \partial_x q_2 - (k_\beta^2 - k_d^2)v_2 - \nu_4\nabla^4\omega_2 - r\nabla^2\psi_2 + \mathrm{SGS}_2 \end{aligned}\tag{4}$$

where ν_4 is the biharmonic Reynolds number, $\omega_j = \nabla^2\psi_j$ is the relative vorticity, and SGS_j are stochastic subgrid scale parameterizations.

By analogy with ocean models and some atmospheric models, the remaining imperfect models do not use a spectral discretization of the nonlinear terms. Instead, they use either the second-order or fourth-order energy- and enstrophy-conserving finite-difference discretizations of Arakawa (1966). Models using the second-order discretization are denoted FD2, and models using the fourth-order discretization are denoted FD4.

To separate the effects of the numerical discretization we run models with and without stochastic subgrid scale (SGS) parameterizations. We consider two stochastic SGS parameterizations: a backscatter scheme from GLM15 that is white in time and independent of the model state, and stochastic superparameterization (SP) from GM14 that is temporally correlated (following GLM15) and dependent on the model state. (Experiments were also run with the temporally-correlated backscatter scheme from GLM15, but the results were almost identical to the white-in-time scheme and are omitted. Also note that the stochastic SP scheme includes a backscatter component.) Both of the stochastic parameterization schemes depend on one primary tunable parameter A related to the amplitude of the backscatter. In the f-plane case $A = 600$ for the uncorrelated backscatter and $A = 6750$ for the stochastic SP scheme (for both the FD2 and FD4 models); the difference in amplitude is due to the difference in temporal correlation. In the β-plane case $A = 30$ for the uncorrelated backscatter in the FD2 model, $A = 350$ for stochastic SP in the FD2 model, and $A = 390$ for stochastic SP in the FD4 model. In the f-plane case the stochastic SP scheme is configured as in GLM15; for the β-plane case it uses $\varepsilon = 25$, $\gamma_0 = 15$, and $\sigma^2 = 67.5$ (for the meaning of these internal parameters see GM13, GM14, or GLM15).

In the f-plane case the FD2 and FD4 models all use the same viscosity coefficient $\nu_4 = 1.6 \times 10^{-4}$. The models are more sensitive to this parameter in the β-plane case, so the values are tuned to improve the model climatology. The FD2 model without stochastic SGS terms uses $\nu_4 = 8 \times 10^{-5}$, and the FD4 model without stochastic SGS terms uses $\nu = 4.8 \times 10^{-5}$. All models with stochastic SGS terms use $\nu_4 = 1.6 \times 10^{-4}$. The values of ν_4 and A were tuned so that the climatological energy spectra and heat flux of the imperfect models would be as accurate as possible; these parameters were not tuned to optimize the performance of the ensemble assimilation schemes.

The time-mean streamfunction spectra $|\hat{\psi}_1|^2 + |\hat{\psi}_2|^2$ (where $\hat{\psi}_i$ is the Fourier coefficient of ψ_i) of the perfect and imperfect models are compared in figure 1, along with the time- and zonal-mean structure of the zonal barotropic

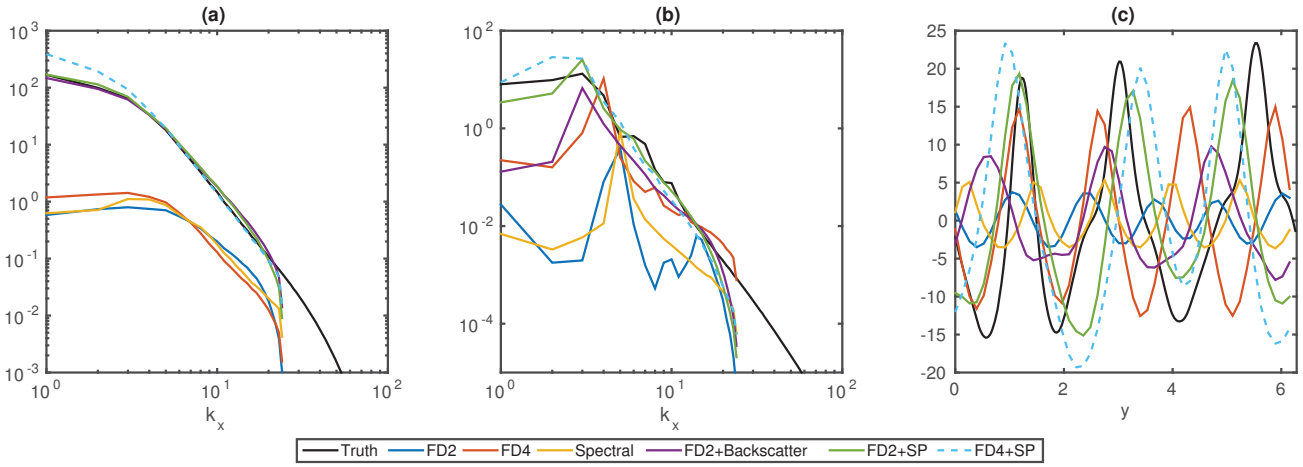

FIG. 1. Perfect and imperfect model climatology: (a) Time-mean streamfunction spectrum $|\hat{\psi}_1|^2 + |\hat{\psi}_2|^2$ for the f-plane scenario. (b) Time-mean streamfunction spectrum for the β-plane scenario. (c) Time- and zonal-mean zonal barotropic jet profiles $(u_1 + u_2)/2$.

TABLE 1. Time-mean $\pm$ standard deviation of the nondimensional heat flux for the perfect and imperfect models in the f-plane and β-plane scenarios.

Model	f-plane	β-plane
Truth	442±94	2.2±0.4
FD2	91±9	1.7±0.2
FD4	72±9	7.0±0.6
Spectral	63±7	0.6±0.1
FD2, Backscatter	580±111	4.3±0.7
FD2, Stoch. SP	569±99	6.7±1.2
FD4, Stoch. SP	671±170	4.8±1.0

jets that develop in the β-plane case. In the f-plane case (figure 1a) the imperfect models without stochastic SGS terms have far too little variability, by nearly two orders of magnitude, whereas all methods with stochastic SGS terms have accurate spectra. There is no time-mean structure in the f-plane case.

In the β-plane case the perfect model develops three barotropic zonal jets, whose time- and zonal mean profiles are shown in black in figure 1c. The imperfect models with stochastic SP (FD2 and FD4) both have three reasonably accurate jets. The FD4 model without SGS terms has four jets, and the FD2 model with backscatter has three weak jets. The remaining models have too many jets that are too weak. The streamfunction spectra are shown in figure 1b. The models with SP are the most accurate, followed by the FD4 model without SGS terms and the FD2 model with backscatter, and finally by the FD2 and Spectral models without SGS terms.

The heat flux generated by the models is presented in table 1. The accuracy of the heat flux generated by the imperfect models is not consistent across the two parameter regimes. In the f-plane case the models without stochastic SGS parameterizations all have far too little energy, and as a result generate far too little heat flux (less than 100, compared to the true value of 442), while the models with stochastic parameterizations all produce too much heat flux. In the β-plane case the models exhibit a wide range of heat fluxes, with the least-accurate model (FD2) having the best heat flux. The Spectral model has too little heat flux, and the remaining models all have too much.

3. Ensemble Assimilation System Configuration

The ensemble assimilation experiments use the EAKF (Anderson 2001) with 100 ensemble members and with observations of the upper layer streamfunction ψ_1, analogous to observations of sea surface height. The observations are taken on a regular 16×16 grid, and the observational errors are independent with zero mean and with variance 15 (f-plane) or 1 (β-plane); the observations are assimilated serially. Observation errors in both cases are about 10% of the climatological variability of ψ_1. Our observational grid has a similar resolution (in comparison with the deformation radius) as the recent experiments of Keating et al. (2012); the observational grid does not resolve the deformation radius, but it does resolve most of the large-scale structure of ψ_1. The subgrid scales contribute an extremely small amount to the observations, which can be seen in the extremely low amplitude of the small-scale part ($k > 25$) of the spectra shown in figure 1; we therefore ignore the representation error, even though the stochastic SP could provide a spatio-temporally varying estimate of the representation error (Grooms et al. 2014).

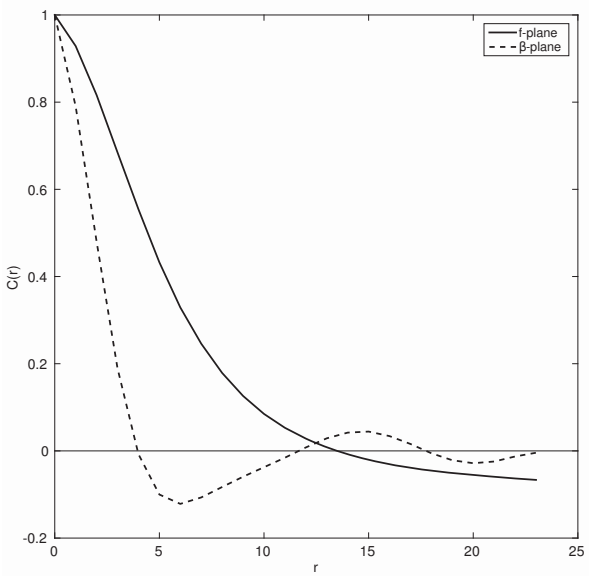

FIG. 2. Angle-averaged spatial correlation functions for ψ_1 in the f-plane scenario (solid) and β-plane scenario (dashed).

Following Keating et al. (2012) we compute the eddy turnover time $T_{\text{eddy}} = 2\pi Z^{-1/2}$ (where Z is the time-averaged total enstrophy $q_1^2 + q_2^2$); in both scenarios the turnover time is approximately 0.006. Taking this as a point of comparison, we performed experiments using observations at intervals of 0.004 and 0.008. The differences between the models were smaller using the shorter observation time, so results are presented here using observations taken every 0.008 time units. The imperfect model ensembles were initialized by adding random samples from a homogeneous, spatially uncorrelated Gaussian random field with variance equal to the observational error variance to the exact state of the large-scale part of the perfect model. The first 200 assimilation cycles are discarded and the last 300 are used to compute performance statistics.

All assimilation experiments use covariance localization with the compactly-supported fifth-order piecewise rational function from Gaspari and Cohn (1999). The localization radius (the distance at which the influence of observations goes to zero) is set to 16 coarse grid points in all experiments. Although better results might have been obtained by tuning the localization radius, good results were obtained using a radius of 16 points, which allows a fair comparison of all the imperfect models. The approximate spatial correlation function for the upper layer streamfunction ψ_1 was computed for both scenarios as

$$C(\delta x, \delta y) = \frac{\mathbb{E}\left[\tilde{\psi}_1(x,y)\tilde{\psi}_1(x+\delta x, y+\delta y)\right]}{\text{Var}\left[\tilde{\psi}_1\right]} \qquad (5)$$

where $\tilde{\psi}_1$ is ψ_1 minus its time mean, and the expectation and variance are computed over all grid points and over the 500 time instances used in the assimilation experiments. The spatial correlation function is approximately isotropic in both cases (not shown). The angle-averaged correlation functions $C(r)$ are shown in figure 2, for comparison with the 16-point localization radius.

In the experiments with the unparameterized FD2 model we used multiplicative and additive inflation, alone and in combination. We simply tested constant multiplicative inflation factors between 1 and 10%; larger inflation factors led to large errors in the heat flux estimation. To implement an additive inflation we diagnosed 500 samples of model error for the unparameterized FD2 model, and used the results to develop an algorithm to generate random samples from an approximate error distribution, as described in the appendix. The diagnosed model error in the β-plane scenario was very inhomogeneous, with large model error variance located near the peaks of the zonal jets. As a result, we only developed an additive error approximation for the f-plane scenario, where the model errors are approximately homogeneous. The additive inflation method was implemented following Mitchell and Houtekamer (2000) by simply adding zero-mean samples from the approximate model error distribution to the prior ensemble before the assimilation cycle.

4. Results

The results of the assimilation experiments for the f-plane and β-plane scenarios are presented in tables 2 and 3, respectively. The results for the FD2 model with neither inflation nor subgrid scale parameterization are presented in boldface, and serve as a baseline for evaluating improvements. For both scenarios the time-mean RMS errors in the forecast and analysis of ψ_1 and ψ_2 are presented, along with the forecast and analysis heat flux, and the RMS errors in the heat flux. Note that the errors are calculated by comparing the low resolution model solution to the large-scale part of the true solution.

In the f-plane case the RMS errors in the components of velocity are all relatively high, and are fairly similar for all the models: between 21 and 26 for the top layer and between 19 and 15 for the bottom layer, compared to climatological variability of 33 and 27 for the top and bottom layers. Nevertheless, despite these relatively large errors, the pattern correlation (PC) in the velocity is relatively high and better differentiates between the models; table 2 therefore shows the pattern correlations for the velocity, rather than the RMS errors. The instantaneous pattern correlation (PC) between the analysis upper layer meridional velocity v_1^a and the true upper layer meridional velocity v_1^t is defined as

$$\text{PC} = \frac{\iint v_1^a v_1^t \,\mathrm{d}x\mathrm{d}y}{\|v_1^a\|_2 \|v_1^t\|_2} \qquad (6)$$

where $\|\cdot\|_2$ denotes the L^2 norm over the domain.

TABLE 2. Filter performance statistics for the f-plane scenario. The format is Forecast → Analysis, where RMS errors (RMSE) are shown for ψ_i and pattern correlations (PC) for $\boldsymbol{u}_i$. The heat flux format is Forecast (time-)Mean → Analysis Mean (RMS Forecast Error → RMS Analysis Error). The standard deviation of the observation error is $\sqrt{15} \approx 3.87$. The standard deviation of the climatological variance of the perfect model is 12 for ψ_1 and 11 for ψ_2; the time-mean heat flux of the perfect model is 430, with standard deviation 125. All values are nondimensional. FD2+ is the FD2 model with additive inflation, and FD2× uses 10% multiplicative inflation.

Model	RMSE ψ_1	RMSE ψ_2	PC u_1	PC v_1	PC u_2	PC v_2	Heat Flux Mean, RMSE
Spectral	3.9 → 3.2	3.5 → 2.7	0.71 → 0.76	0.72 → 0.76	0.79 → 0.83	0.79 → 0.84	365 → 385 (78 → 59)
FD4	4.2 → 3.4	3.7 → 3.0	0.69 → 0.73	0.69 → 0.74	0.77 → 0.81	0.77 → 0.81	380 → 400 (67 → 51)
FD2	**4.4 → 3.6**	**4.0 → 3.3**	**0.64 → 0.69**	**0.65 → 0.70**	**0.73 → 0.77**	**0.73 → 0.77**	**400 → 425 (56 → 45)**
FD2+	4.4 → 3.6	3.9 → 3.1	0.62 → 0.67	0.63 → 0.67	0.73 → 0.77	0.74 → 0.78	330 → 375 (114 → 74)
FD2×	4.1 → 3.3	3.6 → 2.9	0.66 → 0.71	0.67 → 0.72	0.75 → 0.79	0.76 → 0.80	380 → 415 (68 → 44)
FD2, Backscatter	4.0 → 3.3	3.6 → 2.9	0.68 → 0.73	0.69 → 0.74	0.76 → 0.81	0.77 → 0.81	400 → 430 (49 → 37)
FD2, Stoch. SP	4.0 → 3.2	3.5 → 2.8	0.69 → 0.74	0.70 → 0.75	0.77 → 0.81	0.77 → 0.82	400 → 425 (50 → 40)
FD4, Stoch. SP	3.9 → 3.1	3.4 → 2.7	0.71 → 0.76	0.72 → 0.76	0.79 → 0.83	0.79 → 0.84	380 → 400 (63 → 48)

a. f-plane

In the f-plane scenario the baseline FD2 model exhibits modest accuracy despite model error and a relatively sparse observation network: RMS errors for the analysis estimate of the streamfunctions (3.6 and 3.3, for ψ_1 and ψ_2) are less than the observation error variance (3.9), and the velocity components have relatively high pattern correlations of 0.69 to 0.77. The heat flux is also quite accurate, with an analysis mean of 425 in comparison with the true value of 430. The accuracy of the model in this short-range filtering setting is in striking contrast to its poor climatology, as shown for the overall energy of the model in figure 1a and for the heat flux in table 1.

The use of additive inflation, denoted FD2+, has no significant effect on estimates of streamfunctions or velocities, but degrades the performance of the heat flux assimilation. In contrast, 10% multiplicative inflation, denoted FD2×, leads to a noticeable improvement in estimates of both streamfunctions and velocities; the effect on heat flux estimation is mildly detrimental, but the analysis heat flux is still reasonably accurate at 415 versus the true value of 430. As the inflation factor increases from 0 to 10% the streamfunction estimation improves slowly, while heat flux estimation steady degrades. The combination of additive and multiplicative inflation was similar to the use of additive inflation alone (not shown).

The use of higher-order numerics reduces model error, and leads to significant improvement in the assimilation performance: the FD4 model without inflation is comparable to the FD2 model with 10% multiplicative inflation. The Spectral model reduces the model error further, resulting in even more accurate estimates for the streamfunctions and velocities. The assimilation performance for heat flux degrades when moving from the FD2 to the FD4 and Spectral models; this mirrors the decrease in accuracy of the heat flux climatology shown in table 1.

Addition of a white-noise backscatter independent of the model variables in the FD2 model leads to significant improvements in filtering the streamfunctions and velocities, and a small improvement in heat flux. The backscatter reduces model error, as seen in figure 1, and inflates the forecast covariance. The results are better than the FD4 model and better than the FD2 model with 10% inflation, and are nearly comparable to the Spectral model, but with better heat flux. The temporally-correlated, state-dependent stochastic SP scheme with the FD2 model results in only slightly better performance than the backscatter method. Pairing the stochastic SP scheme with the FD4 model leads to further incremental improvements in estimating the streamfunctions and velocities, but has a mildly detrimental impact on heat flux estimation.

For the f-plane scenario covariance inflation, stochastic parameterization, and high-order numerics all lead to improvements in assimilation quality, and the combination of high-order numerics with stochastic SP is more accurate than covariance inflation with the FD2 model.

b. β-plane

In the β-plane scenario the baseline FD2 model exhibits modest accuracy. RMS errors for the analysis estimate of the streamfunctions, 0.97 and 0.86 for ψ_1 and ψ_2, are just less than the observation error variance (unity). The zonal velocities u_1 and u_2, which are dominated by three large-scale jets, are estimated with modest accuracy, having RMS errors in the analysis of 5.1 and 3.5 in the top and bottom layers, compared to climatological variability of 11 and 10, respectively. In contrast, the meridional velocities v_1 and v_2 are estimated very poorly: the RMS analysis errors are 6.0 and 4.1 for the top and bottom layers, compared to climatological variability of 5 and 4. As a result of the poor estimation of the meridional velocities, the heat flux estimate is off by a factor of 2, with a

TABLE 3. Filter performance statistics for the β-plane scenario. The format is Forecast $\rightarrow$ Analysis, where RMS errors (RMSE) are shown for both ψ_i and $\boldsymbol{u}_i$. The heat flux format is Forecast (time-)Mean $\rightarrow$ Analysis Mean (RMS Forecast Error $\rightarrow$ RMS Analysis Error). The standard deviation of the observation error is 1. The standard deviation of the climatological variance of the perfect model is 3.3 for ψ_1, 3.1 for ψ_2, 11 for u_1, 10 for u_2, 5 for v_1, and 4 for v_2. The time-mean heat flux of the perfect model is 2.3, with standard deviation 0.3. All values are nondimensional. FD2$\times$ is the FD2 model with 4% multiplicative inflation.

Model	RMSE ψ_1	RMSE ψ_2	RMSE u_1	RMSE v_1	RMSE u_2	RMSE v_2	Heat Flux Mean, RMSE
Spectral	0.52 $\rightarrow$ 0.49	0.41 $\rightarrow$ 0.39	3.2 $\rightarrow$ 3.1	3.5 $\rightarrow$ 3.4	2.1 $\rightarrow$ 2.0	2.3 $\rightarrow$ 2.2	1.6 $\rightarrow$ 1.7 (0.8 $\rightarrow$ 0.7)
FD4	0.70 $\rightarrow$ 0.66	0.55 $\rightarrow$ 0.52	4.0 $\rightarrow$ 3.9	4.6 $\rightarrow$ 4.4	2.6 $\rightarrow$ 2.5	3.2 $\rightarrow$ 3.0	2.3 $\rightarrow$ 2.3 (0.4 $\rightarrow$ 0.4)
FD2	**0.99 $\rightarrow$ 0.97**	**0.87 $\rightarrow$ 0.86**	**5.1 $\rightarrow$ 5.1**	**6.1 $\rightarrow$ 6.0**	**3.6 $\rightarrow$ 3.5**	**4.1 $\rightarrow$ 4.1**	**4.6 $\rightarrow$ 4.6 (2.4 $\rightarrow$ 2.4)**
FD2$\times$	0.77 $\rightarrow$ 0.72	0.61 $\rightarrow$ 0.58	4.8 $\rightarrow$ 4.7	5.5 $\rightarrow$ 5.3	3.0 $\rightarrow$ 3.0	3.6 $\rightarrow$ 3.5	5.7 $\rightarrow$ 5.6 (3.5 $\rightarrow$ 3.4)
FD2, Backscatter	0.75 $\rightarrow$ 0.70	0.61 $\rightarrow$ 0.58	4.1 $\rightarrow$ 4.0	4.8 $\rightarrow$ 4.6	2.7 $\rightarrow$ 2.7	3.4 $\rightarrow$ 3.2	1.0 $\rightarrow$ 1.0 (1.3 $\rightarrow$ 1.3)
FD2, Stoch. SP	0.70 $\rightarrow$ 0.66	0.55 $\rightarrow$ 0.51	3.9 $\rightarrow$ 3.8	4.5 $\rightarrow$ 4.3	2.6 $\rightarrow$ 2.5	3.1 $\rightarrow$ 2.9	1.9 $\rightarrow$ 2.0 (0.5 $\rightarrow$ 0.5)
FD4, Stoch. SP	0.62 $\rightarrow$ 0.58	0.49 $\rightarrow$ 0.45	3.6 $\rightarrow$ 3.5	4.0 $\rightarrow$ 3.8	2.4 $\rightarrow$ 2.3	2.7 $\rightarrow$ 2.6	2.0 $\rightarrow$ 2.1 (0.4 $\rightarrow$ 0.4)

value of 4.6 compared to the true value of 2.3. The climatological heat flux in the FD2 model is 1.7, as shown in table 1, but its climatological energy is too small (figure 1a). The data assimilation procedure increases the energy level of the FD2 model above its climatology, leading to an erroneous increase in the model's heat flux.

Similar to the f-plane case, multiplicative inflation degrades heat flux estimation. Streamfunction estimation improves as the multiplication factor increases up to an optimum value of 4%, past which it degrades. As shown in table 3, 4% inflation improves the filter performance significantly, decreasing RMS errors in the analysis estimates of ψ_1 and ψ_2 from 0.97 to 0.72 and from 0.86 to 0.58, respectively, while also improving estimates of the velocity. Nevertheless, it does degrade the heat flux estimate.

The use of higher-order numerics in the FD4 model reduces model error, and leads to significant improvement in the assimilation performance for all variables – streamfunctions, velocities, and heat flux. This is somewhat surprising since the FD4 model has too-large climatological heat flux of 7. The FD4 model without inflation performs better than the FD2 model with optimized (albeit non-adaptive) multiplicative inflation. The Spectral model reduces the model error further, resulting in even more accurate estimates for the streamfunctions and velocities, but a slightly worse estimate for heat flux, in line with the too-low climatological heat flux of the Spectral model (table 1).

The FD2 model with white-noise backscatter is slightly better than the FD2 model with inflation when it comes to filtering the streamfunctions and velocities and a small improvement in heat flux, though the latter is still too small. This is somewhat surprising since the climatological heat flux for the white-noise backscatter model is too large, at 4.3 (table 1). The FD2 with backscatter is not quite as accurate as the FD4 model. The FD2 model with stochastic SP improves on the white-noise backscatter, and is very similar to the FD4 model, and the FD4 model with stochastic SP leads to further improvements in streamfunction and velocity estimates. The Spectral model without a stochastic parameterization has better estimates of streamfunctions and velocities than the FD models with stochastic SP, and comparable (though slightly worse) heat flux.

5. Conclusions

Ensemble data assimilation systems for atmosphere and ocean science use computational models that are unable to resolve all the active dynamical scales. Two types of model error result from the use of low resolution models: truncation error in the numerical discretization of the large-scale dynamics, and errors associated with subgrid scale interactions. Additive and multiplicative covariance inflation are methods for accounting for model errors in ensemble data assimilation algorithms (Anderson and Anderson 1999; Mitchell and Houtekamer 2000), but inflation only accounts for model errors and does not reduce them. The use of high-order numerical schemes and subgrid scale parameterizations reduce low resolution model errors; stochastic subgrid scale parameterizations also inflate ensemble spread, similar to additive and multiplicative inflation. Covariance inflation can act as a baseline for evaluating the effect of stochastic parameterizations in ensemble data assimilation systems (Whitaker and Hamill 2012), but previous studies using a particular stochastic subgrid scale parameterization (SKEBS; Berner et al. 2009) have not found significant improvements compared to covariance inflation (Houtekamer et al. 2009; Whitaker and Hamill 2012). We compare inflation and stochastic parameterization in the context of idealized quasigeostrophic turbulence, and find that a simple model-independent white-noise backscatter scheme (from GLM15) is comparable to tuned, non-adaptive multiplicative inflation, while stochastic superparameterization (GM13, GM14) gives better results. We attribute this

success to the ability of a well-designed stochastic parameterization to both reduce and account for model error.

The performance of ensemble assimilation systems based on low resolution models can be improved through the use of stochastic parameterizations, by carefully estimating and accounting for model error distributions (Zupanski and Zupanski 2006), and through sophisticated adaptive covariance inflation techniques (Anderson 2007, 2009; Li et al. 2009; Miyoshi 2011). In our setting the most straightforward way to improve the assimilation performance was to move from a second-order discretization to a fourth-order discretization. The effect of higher-order numerics may seem counterintuitive because the model does not resolve the true solution; the explanation is that the large-scale part of the true solution that is *represented* on the coarse model grid is not equally well *resolved* by different numerical methods. Our stochastic subgrid scale parameterizations had a positive impact on the assimilation performance, with an effect similar to the effect of using higher-order numerics. The stochastic parameterizations had the added benefit of improving the low resolution model climatology, much more so than the use of higher-order numerics.

Covariance inflation will remain an important and useful technique in ensemble data assimilation. But the ability of stochastic parameterizations to reduce model error, and not just to account for it, underscores the need for further development and improvement of stochastic parameterizations in oceanic and atmospheric models.

Acknowledgments. The authors gratefully acknowledge funding from ONR MURI grant N00014-12-1-0912.

APPENDIX

Model Error Parameterization

The model error for the FD2 model is estimated by initializing the FD2 model from the large-scale state of the perfect model, running both models forwards for an interval of $dt = 0.008$, and evaluating the difference in the predictions. This process is repeated using the same time series with 500 samples as is used in the filtering tests, producing 500 samples of model error for ψ_1 and ψ_2.

In the β-plane scenario the model error is inhomogeneous, as shown in figure A1. The center panel shows that the time-mean error in the barotropic streamfunction $\psi_t = (\psi_1 + \psi_2)/2$ is negligible in comparison with the time-mean barotropic streamfunction (left panel) and the standard deviation of the error (right panel). However, the standard deviation of the error is very inhomogeneous, with strong peaks at the latitude of the zonal jets. The model error for the baroclinic component $\psi_c = (\psi_1 - \psi_2)/2$ exhibits similar behavior (not shown).

The model error in the f-plane scenario is approximately homogeneous and isotropic, as shown in figure A2. The left panel shows the time-average of the square amplitude of the barotropic model error ($|\hat{\psi}_t|^2$, where $\hat{\psi}_t(k_x, k_y)$ is the Fourier coefficient), and the center panel shows the time-average of the square amplitude of the baroclinic model error. The right panel shows the angle-averaged values from the left (barotropic, solid) and center (baroclinic, dashed) panels, along with the approximate version generated by the following *ad hoc* sampling algorithm (circles).

f-Plane Model Error Sampling Algorithm:

1.a Generate 48^2 independent samples of a standard normal random variable, and arrange on the 48×48 coarse grid.

1.b Take the discrete Fourier transform. For wavenumbers with $k = (k_x^2 + k_y^2)^{1/2} < 24$, multiply the Fourier coefficients by $\exp\{-((k-6)/10)^2/2\}$; set the remaining coefficients to zero.

1.c Take the inverse discrete Fourier transform, and rescale so that the sample has unit variance. This is the unscaled barotropic error sample.

2.a Repeat step 1.a to generate a new random field.

2.b Take the discrete Fourier transform. For wavenumbers with $k < 30$ multiply the coefficients by $\sin^2(\pi k/30)$; set the remaining coefficients to zero.

2.c Take the inverse discrete Fourier transform, and rescale so that the sample has unit variance.

2.d Add the unscaled barotropic error sample from step 1.c to the sample from step 2.c and divide by 2. The result is the unscaled baroclinic error sample.

3 Multiply the unscaled barotropic sample by 1.47 and the unscaled baroclinic sample by 0.61 to get the barotropic and baroclinic error samples. The top layer sample is the sum of the barotropic and baroclinic samples, and the bottom layer sample is the barotropic minus the baroclinic sample.

This error sampling algorithm is clearly *ad hoc*. However, it produces a homogeneous and isotropic random field with properties very similar to the diagnosed model error. For example, the diagnosed standard deviation for the barotropic model error is 1.49, while the algorithm has standard deviation 1.47. The diagnosed standard deviation for the baroclinic model error is 0.63, while the algorithm has standard deviation 0.61. The diagnosed local correlation between the barotropic and baroclinic error is 0.48, while the algorithm generates a correlation of 0.5. The diagnosed correlation between the model error in the top and bottom layers is 0.75, which is matched exactly by the

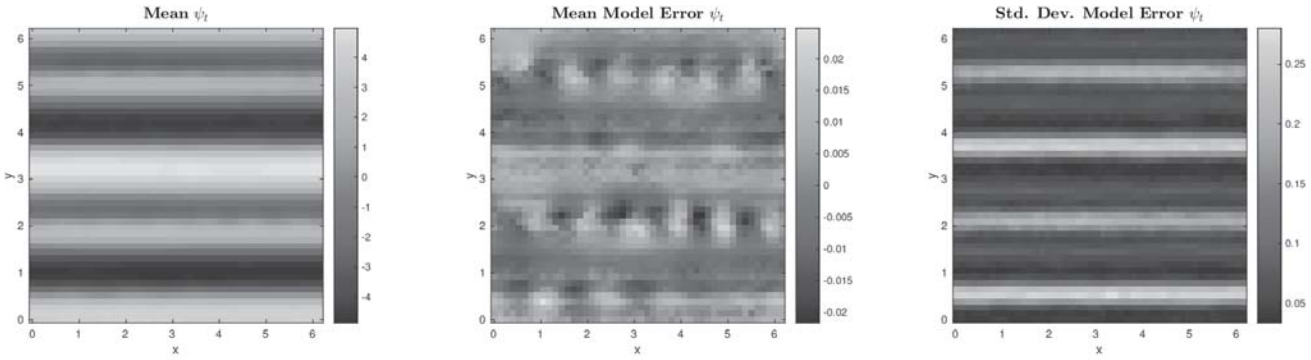

FIG. A1. Time-mean barotropic streamfunction (left), time-mean model error in the barotropic streamfunction for the FD2 model (center), and standard deviation of model error in the barotropic streamfunction for the FD2 model (right).

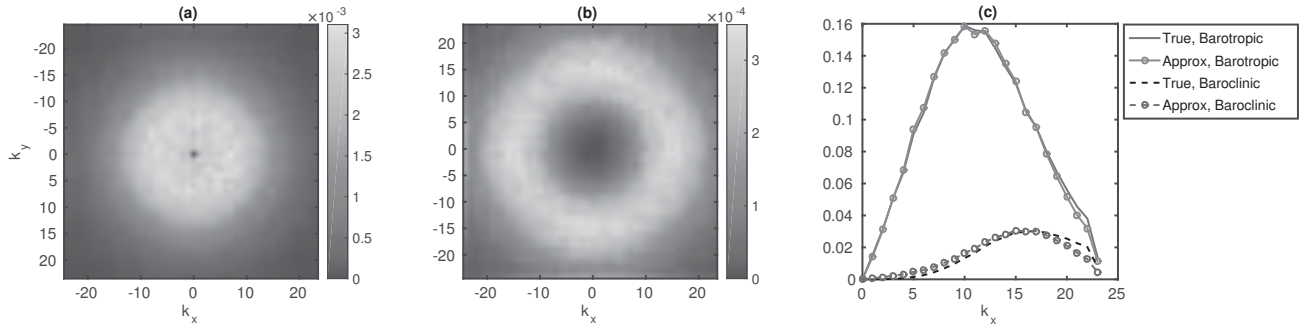

FIG. A2. f-plane FD2 streamfunction model error statistics: (a) Barotropic model error spectrum, (b) Baroclinic model error spectrum, (c) angle-integrated barotropic error spectrum from the diagnostics (solid) and from the approximate sampling algorithm (solid, circles), and angle-integrated baroclinic error spectrum from the diagnostics (dashed) and from the approximate sampling algorithm (dashed, circles).

algorithm. Finally, the barotropic and baroclinic model error spectra are accurately reproduced by the algorithm, as shown in the right panel of figure A2.

References

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

Anderson, J. L., 2007: An adaptive covariance inflation error correction algorithm for ensemble filters. *Tellus*, **59 (2)**, 210–224.

Anderson, J. L., 2009: Spatially and temporally varying adaptive covariance inflation for ensemble filters. *Tellus*, **61 (1)**, 72–83.

Anderson, J. L., and S. L. Anderson, 1999: A Monte Carlo implementation of the nonlinear filtering problem to produce ensemble assimilations and forecasts. *Mon. Wea. Rev.*, **127 (12)**, 2741–2758.

Arakawa, A., 1966: Computational design for long-term numerical integration of the equations of fluid motion: Two-dimensional incompressible flow. Part I. *J. Comp. Phys.*, **1 (1)**, 119–143.

Berner, J., T. Jung, and T. Palmer, 2012: Systematic model error: the impact of increased horizontal resolution versus improved stochastic and deterministic parameterizations. *J. Climate*, **25 (14)**, 4946–4962.

Berner, J., G. J. Shutts, M. Leutbecher, and T. N. Palmer, 2009: A spectral stochastic kinetic energy backscatter scheme and its impact on flow-dependent predictability in the ECMWF ensemble prediction system. *J. Atmos. Sci.*, **66 (3)**, 603–626.

Branicki, M., and A. J. Majda, 2012: Quantifying uncertainty for predictions with model error in non-Gaussian systems with intermittency. *Nonlinearity*, **25 (9)**, 2543.

Evensen, G., 1994: Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. *J. Geophys. Res.*, **99 (C5)**, 10 143–10 162.

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

Frenkel, Y., A. J. Majda, and B. Khouider, 2012: Using the stochastic multicloud model to improve tropical convective parameterization: A paradigm example. *J. Atmos. Sci.*, **69 (3)**, 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.

Grooms, I., Y. Lee, and A. J. Majda, 2014: Ensemble Kalman filters for dynamical systems with unresolved turbulence. *J. Comp. Phys.*, **273**, 435–452, doi:10.1016/j.jcp.2014.05.037.

Grooms, I., Y. Lee, and A. J. Majda, 2015: Numerical schemes for stochastic backscatter in the inverse cascade regime of quasi-geostrophic turbulence. *Multiscale Modeling & Simulation*, submitted; manuscript available at http://www.cims.nyu.edu/~grooms/.

Grooms, I., and A. J. Majda, 2013: Efficient stochastic superparameterization for geophysical turbulence. *Proc. Natl. Acad. Sci. (USA)*, **110**, 4464–4469, doi:10.1073/pnas.1302548110.

Grooms, I., and A. J. Majda, 2014: Stochastic superparameterization in quasigeostrophic turbulence. *J. Comp. Phys.*, **271**, 78–98, doi:10.1016/j.jcp.2013.09.020.

Houtekamer, P., H. L. Mitchell, and X. Deng, 2009: Model error representation in an operational ensemble Kalman filter. *Mon. Wea. Rev.*, **137 (7)**, 2126–2143.

Houtekamer, P. L., H. L. Mitchell, G. Pellerin, M. Buehner, M. Charron, L. Spacek, and B. Hansen, 2005: Atmospheric data assimilation with an ensemble kalman filter: Results with real observations. *Mon. Wea. Rev.*, **133 (3)**, 604–620.

Kalnay, E., 2002: *Atmospheric modeling, data assimilation, and predictability*. Cambridge University Press.

Keating, S., A. Majda, and K. Smith, 2012: New methods for estimating ocean eddy heat transport using satellite altimetry. *Mon. Wea. Rev.*, **140**, 1703–1722.

Li, H., E. Kalnay, and T. Miyoshi, 2009: Simultaneous estimation of covariance inflation and observation errors within an ensemble Kalman filter. *Quart. J. Roy. Meteor. Soc.*, **135 (639)**, 523–533.

Majda, A., and J. Harlim, 2012: *Filtering Complex Turbulent Systems*. Cambridge University Press.

Majda, A. J., and I. Grooms, 2014: New Perspectives on Superparameterization for Geophysical Turbulence. *J. Comp. Phys.*, **271**, 60–77, doi:10.1016/j.jcp.2013.09.014.

Mitchell, H. L., and P. Houtekamer, 2000: An adaptive ensemble Kalman filter. *Mon. Wea. Rev.*, **128 (2)**, 416–433.

Miyoshi, T., 2011: The Gaussian approach to adaptive covariance inflation and its implementation with the local ensemble transform Kalman filter. *Mon. Wea. Rev.*, **139 (5)**, 1519–1535.

Randall, D., M. Branson, M. Wang, S. Ghan, C. Craig, A. Gettelman, and J. Edwards, 2013: A community atmosphere model with superparameterized clouds. *EOS*, **94 (25)**, 221–228.

Szunyogh, I., E. J. Kostelich, G. Gyarmati, E. Kalnay, B. R. Hunt, E. Ott, E. Satterfield, and J. A. Yorke, 2008: A local ensemble transform Kalman filter data assimilation system for the NCEP global model. *Tellus*, **60 (1)**, 113–130.

Whitaker, J. S., and T. M. Hamill, 2012: Evaluating methods to account for system errors in ensemble data assimilation. *Mon. Wea. Rev.*, **140 (9)**, 3078–3089.

Whitaker, J. S., T. M. Hamill, X. Wei, Y. Song, and Z. Toth, 2008: Ensemble data assimilation with the NCEP global forecast system. *Mon. Wea. Rev.*, **136 (2)**, 463–482.

Zupanski, D., and M. Zupanski, 2006: Model error estimation employing an ensemble data assimilation approach. *Mon. Wea. Rev.*, **134 (5)**, 1337–1354.