TEST MODELS FOR FILTERING WITH SUPERPARAMETERIZATION*

JOHN HARLIM† AND ANDREW J. MAJDA‡

Abstract. Superparameterization is a fast numerical algorithm to mitigate implicit scale separation of dynamical systems with large-scale, slowly varying 'mean', and smaller scale, rapidly fluctuating 'eddy' term. The main idea of superparameterization is to embed parallel highly resolved simulations of small scale eddies on each grid cell of coarsely resolved large scale dynamics. In this paper, we study the effect of model errors in using superparameterization for filtering multiscale turbulent dynamical systems. In particular, we use a simple test model, designed to mimic typical multiscale turbulent dynamics with small-scale intermittencies without local statistical equilibriation conditional to the large scale mean dynamics, and, simultaneously, force the large scale dynamics through eddy flux terms. In this paper, we consider the Fourier domain Kalman filter for filtering regularly spaced sparse observations of the large scale mean variables. We find high filtering and statistical prediction skill with superparameterization (identical to the skill with perfect model), beyond conventional approaches such as the "bare-truncation model" that ignores completely the eddy-fluxes and the "equilibrium closure" model that crudely approximates the eddy fluxes with classical averaging theory. We show that this high filtering skill is robust even for very sparse observation networks and turbulent signals with a very steep, -6, spectrum. This is a counter-example to naive thinking that the small-scale processes are not so important in multiscale turbulent dynamics with steep energy spectrum. We find that the high filtering skill with superparameterization is robust for small enough scale gap, provided that the filter prior model satisfies the classical linear controllability condition. We will demonstrate a spectacular failure of filtering deterministically forced true signals with the exactly perfect model that does not satisfy controllability and a dramatic improvement when the controllability condition is restored with additional stochastic forcings. This result reconfirms and justifies the counter-intuitive viewpoint that judicious model errors (or noises) can help filtering turbulent signals.

Key words. superparameterization, Kalman filter, no scale separation, data assimilation

AMS subject classifications. 93E11, 62M20, 62L12,65C20, 76F55, 34E13

1. Introduction.

An important scientific problem in applied sciences and engineering is to model signals from nature with multiscale features. In many applications ranging from modeling polymeric fluids to weather and climate dynamics, we are interested in predicting the large scale 'mean' dynamics which is typically observed on time scales longer than that of the unobserved small scale 'eddy' component. The practical difficulty in multiscale modeling is that we typically can't resolve the eddies and thus model errors through various parameterizations are unavoidable on the mean macroscopic level. For example, in the climate modeling, the planetary waves of interest have spatial structure on the order of 1000 to 10,000 km with weekly or monthly time scale and in the long run the behavior of these waves depends on the energy inverse cascade from the small scale processes such as the dry and moist convection, clouds that occur on a spatial structure between 100 m to 100 km and time scales between hours to days. Practically, it is typically computationally expensive to resolve these small scale processes even if we understood their dynamics and at the

*The research of J.H. was partially supported by the ONR grant N00014-11-1-0310 and A.J.M. was partially supported by the ONR DRI grant N00014-10-1-0554 and N00014-11-1-0306, the NSF grant DMS-0456713 and the NSF CMG program grant DMS-1025468. Both authors are also partially supported by the ONR MURI grant N00014-12-1-0912.

†Department of Mathematics, North Carolina State University, NC, 27695, United States (jharlim@ncsu.edu).

‡Courant Institute of Mathematical Sciences, New York University, NY, 10012, United States (jonjon@cims.nyu.edu)

same time we can't completely ignore them.

Various approaches were introduced to mitigate such issues, including superparameterization that has been extensively used in simulating combustion problems [16] and cloud resolving convection parameterizations [6, 5, 30] with some success. The main idea of superparameterization is to embed parallel highly resolved simulations of small scale eddies on each grid cell of coarsely resolved large scale dynamics. Typically, the eddies are resolved on a horizontal periodic domain and they only interact with each other through the large scale mean dynamics. The goal of this paper is to assess the effect of model errors in using superparameterization for data assimilation in a transparent fashion. As in the practical setting, we focus on filtering (or assimilating) sparse observations of the large scale mean variables and no observations of the small scale eddies.

To achieve this goal, we use a test model for superparameterization that was first introduced in [23]; this model is a version of the Gaussian closure models for turbulence [21, 7]. The test model is designed to mimic typical multiscale turbulent dynamics with small-scale intermittencies without local statistical equilibriation conditional to the mean dynamics, and, simultaneously, forces the mean equations through nonlinear eddy flux-like terms. Here, we called it flux-like term because the average is defined over the coordinate of the embedded domain in addition to the standard statistical average. In the remainder of the paper, we will just refer it as the eddy flux term. For analytical tractability, we set the mean dynamics, conditional to the nonlinear eddy flux terms, to solve a linear PDE; in this fashion, the mean variables are nonlinear processes. Practically, the traditional superparameterization introduces various model errors depending on the implementation details [6, 5, 20, 30]; in this paper, we are interested to assess the effect of model errors due to spatial discretization for data assimilation. This is realized by defining the test model to embed small scale dynamics for the eddies on a parallel infinite domain and by defining the superparameterization as a numerical scheme that replaces this infinite domain with an L-periodic domain with a finite scale gap, L^{-1}, [23]. Essentially, the eddies in the our test model do not have direct interaction between each other, and therefore, we do not assess model errors due to elimination of nonlocal eddy interactions. A detailed study of statistical prediction skill for this type of model errors is fully reported in [7] for complex dispersive turbulent waves.

Here, we implement the recently developed Fourier domain Kalman filtering strategy for regularly spaced sparse observations [9, 24]. We choose this strategy rather than the standard physical space filter to avoid spurious nonlocal correlations when the filter covariance matrix is suboptimally estimated [14, 11], in addition to the fact that the test model mean dynamics, conditional to the eddy flux forcing, is linear with explicit solutions in discrete Fourier coordinate. In this fashion, we assimilate nonlinear turbulent signals, technically, with a linear filter driven by nonlinear eddy flux additive forcings. This strategy is not only numerically beneficial but it also produces accurate solutions beyond the standard physical space approach as we reported in various contexts (see [25, 24] and the references therein). We find that filtering with model errors due to superparameterization with small enough scale gap (large L) is superior to filtering with model errors due to other approximations such as the bare truncation model, which completely ignores estimating the small scale eddy flux terms, and the statistical equilibrium closure model, which applies the classical averaging principle [17, 19, 28] to crudely estimate the eddy flux terms. We will show that this high filtering skill is robust even for very sparse observation networks and

turbulent signals with a very steep, -6, spectrum. We will also demonstrate that the classical linear filtering controllability condition [1, 12] is necessary for accurate filtering with superparameterization and not sufficient for accurate filtering with the bare truncation model and equilibrium closure models. We will also show a spectacular failure of filtering the deterministically forced true signal with the exactly perfect model that does not satisfy the controllability and a dramatic improvement when controllability condition is restored with additional stochastic forcings. This result reconfirms and justifies a counter-intuitive thinking that judicious model errors (or noises) can sometimes help filtering, as we encountered in many different data assimilation contexts [24]. A detailed outline of the remainder of this paper is presented next.

The test model for superparameterization is discussed in detailed in Section 2. In there, we also discuss the related other approximations: the bare-truncation model and the equilibrium closure model. In Section 3, we discuss superparameterization. In Section 4, we define our filtering problem and review the basic Fourier domain Kalman filtering algorithm. In Section 5, we show the numerical results. We end the paper with a short summary and a brief discussion of future work in Section 6.

2. Test models for superparameterization. Superparameterization is a fast numerical algorithm to mitigate implicit scale separation of dynamical systems with large-scale, slowly varying 'mean', $\bar{u}$, and smaller scale, rapidly fluctuating 'eddy' term, u' [6, 5, 20]. Here, we describe a relevant test model with two spatial scales, $X, x \in \mathbb{R}^n$, where $X = \epsilon x$, and two temporal scales, t, τ, where $\tau = t\epsilon^{-1}$, with $\epsilon < 1$, a scale separation parameter [23, 20, 26]. In particular, we decompose a turbulent field

$$U = \bar{u}(X,t) + u'(X,x,t,\tau) \in \mathbb{R}^s. \tag{2.1}$$

into its slowly varying mean, $\bar{u}(X,t)$, and fluctuation, $u'(X,x,t,\tau)$, where $\bar{u}' \equiv 0$, similar to a Reynold averaging formulation in turbulence [10, 27, 29]. We model u' as an unbiased Gaussian random field that is stationary in x for fixed (X,t,τ) with an eddy flux term represented by an $s \times s$ covariance matrix, $cov(u')(X,t,\tau) \equiv \overline{u'u'}$. Given any integrable function $f(t,\tau)$ and a fixed ϵ, we define

$$\langle f \rangle(t) = \epsilon \int_0^{\epsilon^{-1}} f(t,\tau) d\tau \tag{2.2}$$

as the empirical time average of f over the fast time scale.

An in [23], we consider the following scalar ($s = 1$) multiscale test model:

$$\frac{\partial \bar{u}}{\partial t} + P(\partial_X)\bar{u} = \langle cov(u') \rangle (X,t) + F_{ext}(X,t), \tag{2.3}$$

$$\frac{\partial u'}{\partial \tau} + P'(\bar{u}, \partial_x)u' = -(-\Gamma(\partial_x)u' + \sigma(x)\dot{W}(\tau)). \tag{2.4}$$

In (2.3), we will choose P to be a linear constant coefficient differential operator that represents relevant mean dynamics. In this fashion, the test model exhibits nontrivial large-scale turbulent dynamics only through the nonlinear covariance eddy flux term, $\langle cov(u') \rangle$, that reflects the turbulent fluctuations from the small-scales, and a large scale external forcing term, F_{ext}, only when this forcing has a stochastic component. We consider small-scale dynamics in (2.4) with a constant coefficient differential operator P' that depends explicitly on the mean variable $\bar{u}$ with varying stability and

instability features without statistical steady state, mimicking real applications (see [6, 5, 30] for cloud-resolving convection parameterization). In general, the dynamical equations for the eddies are strongly nonlinear; here, for analytical tractability, we use Ornstein-Uhlenbeck process with damping operator Γ and white noise forcing $\sigma(x)\dot{W}(x,\tau)$ to model nonlinear interaction between eddies [23]. This test model is a special example of the Gaussian closure models for turbulent dynamical systems introduced by the second author in [21]. In particular, we can write (2.4) in Fourier space as a linear stochastic differential equation

$$\frac{d\hat{u}'_k}{d\tau} + \tilde{P}'(\bar{u}, \mathrm{i}k)\hat{u}'_k = -\gamma_k \hat{u}'_k + \sigma_k \dot{W}_k, \tag{2.5}$$

where $\tilde{P}'$ is the eigenvalue of the differential operator P' associated with eigenfunction e^{ikx} for constant $\bar{u}$, $P'e^{ikx} = \tilde{P}'e^{ikx}$. Similarly, γ_k satisfies $\Gamma e^{ikx} = \gamma_k e^{ikx}$. Here, the Fourier coefficient is defined by the following spectral integral [31],

$$u'(X,x,t,\tau) = \int_{\mathbb{R}^n} \hat{u}'_k(X,t,\tau)e^{\mathrm{i}kx}dW_k, \tag{2.6}$$

In (2.5), the stochastic noise term in (2.4) is defined as $\sigma(x)\dot{W}(\tau) \equiv \int_{\mathbb{R}^n} e^{\mathrm{i}kx}\sigma_k dW_k$, and by the definition in (2.6), we have

$$cov(u')(X,t,\tau) = \int_{\mathbb{R}^n} C_k(X,t,\tau)dk, \tag{2.7}$$

where $C_k \equiv \overline{\hat{u}'_k(\hat{u}'_k)^*}$, and subscript '*' denotes complex conjugate. From (2.5), we deduce a linear deterministic ODE with coefficients depending on $\bar{u}$ for the covariance C_k,

$$\begin{aligned} \frac{dC_k}{d\tau} &= -(\tilde{P}'_k + (\tilde{P}'_k)^* + \gamma_k + \gamma_k^*)C_k + \sigma_k\sigma_k^*, \\ C_k(\tau = 0) &= C_{k,0}, \end{aligned} \tag{2.8}$$

where $\tilde{P}'_k \equiv \tilde{P}'(\bar{u}, \mathrm{i}k)$. For fixed $\bar{u}$, the initial value problems in (2.8) have explicit solutions given by,

$$C_k(\tau) = e^{-2\lambda_k\tau}C_{k,0} + \frac{\sigma_k^2}{2\lambda_k}(1 - e^{-2\lambda_k\tau}), \tag{2.9}$$

where

$$\lambda_k = \frac{(\tilde{P}'_k + (\tilde{P}'_k)^*) + (\gamma_k + \gamma_k^*)}{2}. \tag{2.10}$$

Given (2.3)-(2.10), now we discuss the numerical implementation details to finalize the description of the test model. In particular, given initial conditions $\bar{u}$ and $C_{k,0}$ at time t, we numerically integrate (2.3)-(2.4) in the following sequence:

1. Update $\tilde{P}'_k = \tilde{P}'(\bar{u}, \mathrm{i}k)$ in (2.10) and compute the turbulent fluctuation $\langle cov(u')\rangle$ using the spectral integral (2.7) and empirical time average (2.2) with constant ϵ,

$$\begin{aligned} \langle cov(u')\rangle &= \epsilon \int_0^{\epsilon^{-1}} \int_{\mathbb{R}^n} C_k(\tau)\, dk\, d\tau \\ &= \int_{\mathbb{R}^n} \Big[\frac{\sigma_k^2}{2\lambda_k} + \frac{\epsilon}{2\lambda_k}(1 - e^{-2\lambda_k\epsilon^{-1}})\Big(C_{k,0} - \frac{\sigma_k^2}{2\lambda_k}\Big)\Big]dk \end{aligned} \tag{2.11}$$

   Here, we effectively freeze the large scale time t.

2. Integrate the large-scale PDE in (2.3) with large time step Δt on a coarsely resolved period domain. Numerically, we assume that the turbulent fluctuation $\langle cov(u')\rangle(X,t)$ is constant over the time interval $(t, t+\Delta t)$. With such an approximation, notice that the integration step here involves two discrete Fourier forward and inverse transforms for $\bar{u}$ and $\langle cov(u')\rangle$. Then we repeat these two steps.

From the implementation steps described above, notice that $\bar{u}$ and u' are resolved on different domains. Furthermore, we proceed in step 1 above, locally, conditional to $\bar{u}$ at grid point X_j. In this sense, pairwise small scale solutions $\langle cov(u')\rangle(X,t)$ obtained by freezing $\bar{u}$ at two distinct locations $X_j \neq X_\ell$ do not interact directly. With this implementation, we implicitly define the notation '$-'$ as an average over the coordinate of the embedded domain in additional to the standard statistical average. Essentially, the test model consists of parallel solutions of highly resolved small scale eddies on an infinite domain that are locally embedded to coarsely resolved large scale dynamics on a periodic domain and interact to each other indirectly through the large scale dynamics. With this setup, we have an unambiguous test model [23, 21] to assess the potential of recovering the mean dynamics in the presence of model errors due solely to spatial discretization when superparameterization is used [6, 5, 30].

2.1. Other closure approximations. Notice that in the two-step above, we never compute the evolution of the fluctuations, $u'(\tau)$, since we have explicit solutions for its covariances. In real applications, the smaller-scale dynamics are highly nonlinear and obtaining $\langle cov(u')\rangle$ is a nontrivial task. In that case, various closure approximations can be introduced to estimate $\langle cov(u')\rangle$, avoiding resolving $u'(\tau)$ in detail.

A poor man's approach is to completely ignore the small-scale nonlinear eddy flux term by setting $\langle cov(u')\rangle = 0$ in (2.3). We call this approach the **bare-truncation model**. In our setup, if the large-scale linear operator P is associated with a dissipative dynamics and the external forcing F_{ext} is deterministic, the bare-truncation model will produce steady state solutions without any internal feature to develop turbulence or chaotic dynamics.

The classical averaging theory [17, 19, 28] states that for a clear scale separation, $\epsilon \to 0$, if the conditional invariant measure $p_\infty(u'|\bar{u})$ exists, we can approximate the mean dynamics in (2.3) with an effective dynamics by replacing $\langle cov(u')\rangle$ with an approximate eddy flux (if it exists), obtained by averaging with respect to the conditional invariant measure $p_\infty(u'|\bar{u})$. These assumptions are fundamental in the Heterogeneous Multiscale Methods (HMM) [4], which is a numerical method designed to estimate $p_\infty(u'|\bar{u})$ when its explicit expression is not available.

In our setup, these assumptions correspond to restricting $\lambda_k > 0$. In this case, C_k converges to a statistical equilibrium $\sigma_k^2(2\lambda_k)^{-1}$ as $\epsilon \to 0$ (or as $\tau = \epsilon^{-1} \to \infty$). If C_k converges quickly (or much faster than ϵ^{-1}), then the empirical average in (2.2) is not too sensitive to initial condition $C_{k,0}$ for small $\epsilon \ll 1$, and we have a **statistical equilibrium closure model** given by

$$\frac{\partial \bar{u}}{\partial t} + P(\partial_X)\bar{u} = \langle cov(u')\rangle_\infty(X,t) + F_{ext}(X,t), \tag{2.12}$$

where

$$\langle cov(u')\rangle_\infty(X,t) \equiv \lim_{\epsilon\to 0} \epsilon \int_0^{\epsilon^{-1}} \int_{\mathbb{R}^n} C_k(X,t,\tau)\, dk\, d\tau = \int_{\mathbb{R}^n} \frac{\sigma_k^2}{2\lambda_k} dk. \tag{2.13}$$

In nature, the separation of scales may not be satisfied with $\epsilon \approx 10^{-1}$. Additionally, for some $\bar{u}$ the eddy dynamics have no statistical equilibrium when $\lambda_k \leq 0$ for some k. In this situation, the energy spectrum $C_k(\tau)$ grows without bound as $\tau \to \infty$ and such dynamics won't be captured by the statistical equilibrium closure model above.

In our test model, we allow λ_k to fluctuate between positive and negative depending on $\bar{u}$ to generate small-scale intermittencies without local statistical equilibriation to nontrivially affect the mean dynamics [23]. This is what we will discuss next.

2.2. Small-scale intermittency. To model intermittency, as in [23], we choose

$$\frac{\tilde{P}'_k + (\tilde{P}'_k)^*}{2} = -f(\bar{u})A_k,$$

where

$$A_k = \bar{A}e^{-\delta|k|}|k|^2 \tag{2.14}$$

is a decaying factor for high wavenumber and $f(\bar{u})$ is a function that depends on mean $\bar{u}$. In (2.14), $\delta = 0.1$ and $\bar{A}$ are chosen such that $\max_k A_k = 1$. In [23], they considered a cubic polynomial f; in that case, f has no upper bound and the covariance $\langle cov(u')\rangle$ is sensitive to small changes in $\bar{u}$ that can lead to numerical instability in the quadrature; to avoid this problem, the covariance was always truncated below a maximal value set to 10 and the initial spectrum was chosen to equal the equilibrium spectrum without large-scale interaction, $C_{k,0} = \sigma_k^2(2\gamma_k)^{-1}$.

Here, we choose quadratic f such that it has an upper bound and hence the numerical instability in the quadrature is avoided when we proceed to step 1 above with initial spectrum $C_{k,0} = \sigma_k^2(2\gamma_k)^{-1}$. In this fashion, we avoid the adhoc covariance truncation adopted in [23]. The choice of equilibrium initial spectrum at each time t in our implementation is merely for numerical simplicity; in real applications, we may need to estimate $C_{k,0}$ at time t, possibly with a small-scale reinitialization approach introduced in [8, 13], and this is beyond the scope of this paper. To motivate the choice of quadratic function for f, consider

$$\lambda_k = \frac{\gamma_k + \gamma_k^*}{2} - f(\bar{u})A_k \geq -\alpha,$$

where $\alpha > 0$. Suppose that the damping term $\gamma_k > 0$ and we consider $\bar{u} \in \mathbb{R}$, then $f : \mathbb{R} \to (-\infty, (\gamma_k + \alpha)A_k^{-1})$. When $\bar{u}$ is such that

$$-\infty < f(\bar{u}) < \frac{\gamma_k}{A_k}, \tag{2.15}$$

the small-scale dynamics (2.8) has statistical equilibrium

$$\lim_{\tau\to\infty} C_k(\tau) = \frac{\sigma_k^2}{2(\gamma_k - f(\bar{u})A_k)}.$$

But, when $\bar{u}$ is such that

$$\frac{\gamma_k}{A_k} < f(\bar{u}) \leq \frac{\gamma_k + \alpha}{A_k}, \tag{2.16}$$

there is no statistical equilibrium.

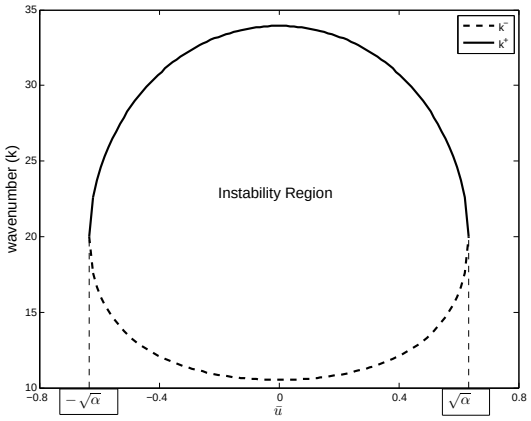

FIG. 2.1. *Boundary of unstable modes as a function of u.*

In our numerical example in this paper, we choose quadratic

$$f(\bar{u}) = \gamma_k + \alpha - \bar{u}^2, \tag{2.17}$$

as an alternative to the cubic f in [23]; note that this choice doesn't change the underlying physics. This means that for a given $\bar{u}$, the instability occurs on modes $k \in [k^-, k^+]$, where $k^\pm$ with $0 \le k^- < k^+$ solve

$$\gamma_k - f(\bar{u})A_k = 0. \tag{2.18}$$

For quadratic f in (2.17), the condition (2.16) implies that there is no statistical equilibrium when

$$-\sqrt{\alpha} < \bar{u} < \sqrt{\alpha}. \tag{2.19}$$

The instability region, corresponding to no statistical equilibriation, is shown in Fig. 2.1, for $\alpha = 0.4$.

2.3. Numerical Simulations. We numerically simulate the test model in (2.3)-(2.8) with a linear differential operator for large scale dynamics defined by

$$P(\partial_X) = A\partial_X^3 - \nu\partial_X^2 + c\partial_X + d, \tag{2.20}$$

where $A = 1$ is the dispersion coefficient, $c = 1$ is the advection coefficient, $\nu = 10^{-8}$ is the diffusion coefficient, and $d = 10^{-2}$ is the linear damping coefficient. For the small-scale dynamics, we choose uniform damping $\gamma_k = 1$ with a -5/3 turbulent spectrum,

$$\frac{\sigma_k^2}{2\gamma_k} = \frac{E_0}{(1+|k|)^{5/3}}, \tag{2.21}$$

with pre-constant $E_0 = 0.1$. In this sense, the statistical equilibrium for the small-scale equation when $\bar{u}$ is ignored is an energetic turbulent field without scale separation. We resolve the large-scale dynamics on $N = 128$ equally spaced grid points on a one-dimensional ($n = 1$) 2π-periodic domain while the smaller scale dynamics is resolved up to wavenumber $|k| \le K_{max} = 1000$. We use an adaptive quadrature rule to

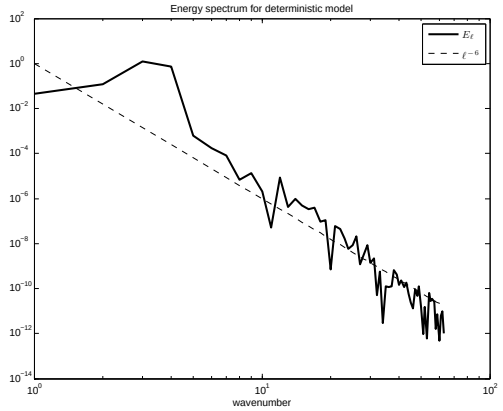

FIG. 2.2. *Empirically estimated large-scale energy spectrum of the deterministically forced system (thick solid) compared to the ℓ^{-6} spectrum. The mean dynamics has $|\ell| \leq N/2 = 64$ modes since we only resolve it on $N = 128$ equally spaced grid points.*

integrate (2.11) with a fixed initial condition $C_{k,0} = E_0(1 + |k|)^{-5/3}$ for numerical simplicity as discussed above in Section 2.2. The large-scale dynamics are integrated following step 2 above with time step $\Delta t = 0.5$. We consider modest scale separation with $\epsilon = 0.1$, and large-scale external forcings through

$$F_{ext} = \bar{F} + \Lambda(X) \cdot \dot{W}(t), \tag{2.22}$$

where we set $\bar{F} = -0.5$.

In our first simulation, we consider a large-scale deterministic forcing by setting $\Lambda(X) \cdot \dot{W}(t) = 0$. In Fig. 2.2, we show the empirically estimated large-scale energy spectrum from numerical solutions of the test model, which is nothing but the temporal variance of the Fourier coefficients. Notice that the energy spectrum is close to ℓ^{-6} spectrum for $\ell > 5$. In Fig. 2.3, we plot the corresponding solutions of the deterministic forcing case. Panels (a) and (b) show the snapshots of the solution and the covariance of the small-solution at time $t = 400.1$, respectively; panel (c) depicts $\bar{u}$ and $\langle cov(u') \rangle$ at grid point $X = 3.0925$ as functions of time. Here, the choices of snapshot time $t = 400.1$ and grid point $X = 3.0925$ are arbitrary. Notice that whenever $\bar{u}$ is within $(-\sqrt{\alpha}, \sqrt{\alpha})$, which is denoted by the two horizontal dashes on panels (a) and (c), the nonlinear eddy flux term, $\langle cov(u') \rangle$, is large as expected. These large excursions in $\langle cov(u') \rangle$ are the intermittent local instabilities described in Section 2.2.

We also consider stochastically forced large scale dynamics through,

$$\Lambda(X_j) \cdot \dot{W}(t) \equiv \sum_{|\ell| \leq N} e^{\mathrm{i}\ell X_j} \hat{\sigma}_\ell \dot{W}_\ell(t), \quad X_j = j2\pi/N, \tag{2.23}$$

a spatially correlated noise that is white in time. Here, we choose $\hat{\sigma}_\ell$ such that the large scale dynamics in (2.3) without the small scale covariance term, $\langle cov(u') \rangle$, has an energy spectrum of ℓ^{-6}. For this case, the empirically estimated large-scale energy spectrum is close to ℓ^{-3} spectrum for $\ell > 5$ (see Fig. 2.4). The corresponding solutions for this case physically resemble the deterministic forcing case; the only difference is that the small scale intermittent local instabilities occur at different locations and times (results not shown).

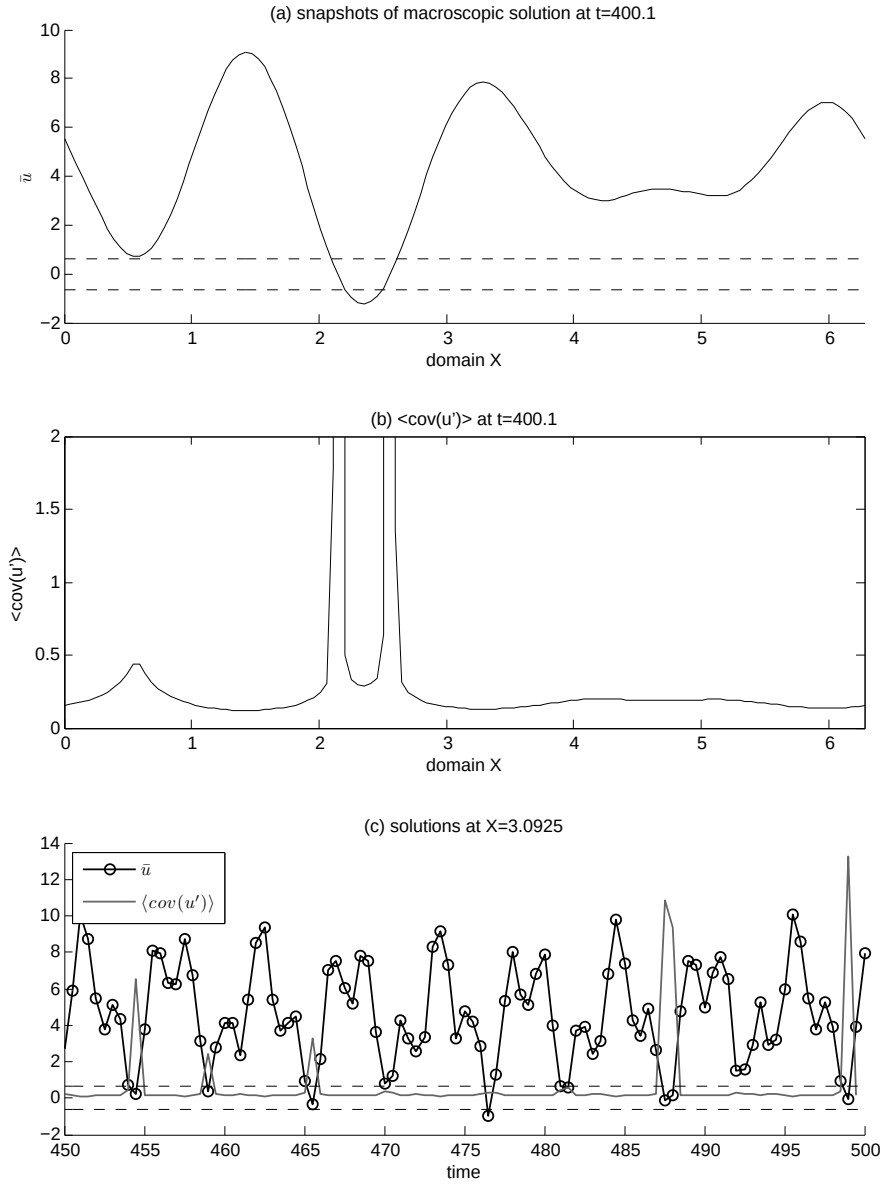

FIG. 2.3. *Deterministically forced solutions: (a) Snapshots of $\bar{u}$ at time $t = 400.1$; (b) Small-scale eddy fluxes correspond to $\bar{u}$ at time $t = 400.1$; (c) Timeseries of $\bar{u}$ and $\langle cov(u')\rangle$ at $X = 3.0925$.*

In Fig. 2.5, we show the space-time plot of the mean state, $\bar{u}$, and the associated small-scale intermittent unstable eddy fluxes, $\langle cov(u')\rangle$, that occur when $\bar{u} \in (-\sqrt{\alpha}, \sqrt{\alpha})$. In the remainder of this paper, we refer to these numerical solutions as the truth.

3. Superparameterization. The main idea of superparameterization is to retain the large-scale dynamics in (2.3), but make various space-time discrete approximations in solving the small-scale dynamics in (2.4) to reduce the computational

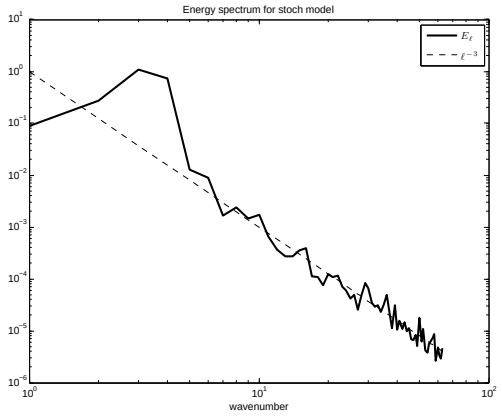

FIG. 2.4. *Empirically estimated large-scale energy spectrum of the stochastically forced system (thick solid) compared to the ℓ^{-3} spectrum. The mean dynamics has $|\ell| \leq N/2 = 64$ modes since we only resolve it on $N = 128$ equally spaced grid points.*

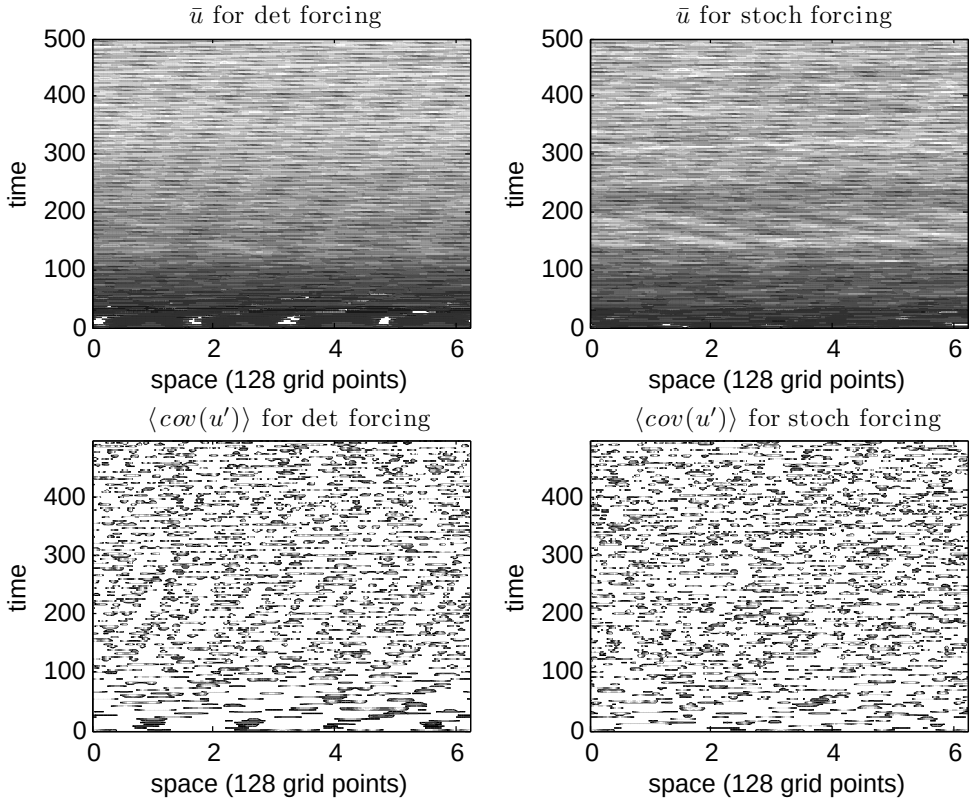

FIG. 2.5. *Space-time plot for the deterministically forced (left panels) and stochastically forced (right panels) solutions. In the bottom panels, the darker curves correspond to large small-scale intermittent local instabilities.*

cost. The traditional superparameterization introduces an artificial scale gap, L^{-1}, and solves the small-scale dynamical equations in (2.4) locally on a periodic domain [6, 5, 30]. Such an approximation introduces two types of model errors: first, model errors due to finite spatial and temporal discrete approximations. Second, model errors due to truncation of the direct interaction between nonlocal fluxes.

The test model in (2.3)-(2.8) is designed in such a way that superparameterization only introduces model errors of the first type (see Section 2). Indeed, this test model is readily a "superparameterization" with small-scale dynamics resolved on parallel embedded infinite domains (see [7]). With this setup, superparameterization simply replaces the stochastic integral in (2.6) with a discrete sum of random variables over the lattice with wavenumbers $k_j = j/L$. Practically, we approximate the covariance integral in (2.11) with,

$$\langle cov(u')\rangle_L = \frac{1}{L^n}\sum_j \Big[\frac{\sigma_{k_j}^2}{2\lambda_{k_j}} + \frac{\epsilon}{2\lambda_{k_j}}(1-e^{-2\lambda_{k_j}\epsilon^{-1}})\Big(C_{k_j,0} - \frac{\sigma_{k_j}^2}{2\lambda_{k_j}}\Big)\Big], \qquad (3.1)$$

which is nothing but a trapezoidal rule. In realistic applications, superparameterization approximates (3.1) even further by a suitable reduced model on a one-dimensional L-periodic domain, replacing the full n-dimensional periodic domain [16, 6, 5]. There is also a sparse-time variant of superparameterization [30] that judiciously approximates the eddy flux term, $\langle cov(u')\rangle$, with that from a reduced time integration of the small-scale dynamics. In this paper, we will not explore these two approximations; we only consider resolving the test model on a one-dimensional domain ($n = 1$) and focusing on the effect of the trapezoidal approximation in (3.1) for data assimilation purposes.

Now, let's discuss some numerical results of simulating the test model in (2.3)-(2.8) with superparameterization with various values of scale gap, measured by L^{-1}. In Fig. 3.1, we show the simulations of the deterministically forced case with superparameterization with $L = 2$. Notice that this approximation reproduces qualitative solutions with local intermittent instabilities, similar to those of the truth (compare with Fig. 2.3); note that the small-scale intermittencies do not occur at the same instances and locations since the test model is a chaotic dynamical system. By chaos, we mean sensitivity to the choices of unstable modes when different integration rule is employed in approximating (2.11) in addition to sensitivity to initial conditions.

On the other hand, the superparameterization with $L = 0.1$ produces stable equilibrium dynamics (see Fig. 3.2). Here, the failure in superparameterization is attributed to incorrect subsampling on the unstable wavenumbers of the small scale dynamics. For this specific example, when $\bar{u} \in (-\sqrt{\alpha}, \sqrt{\alpha})$, we resolve at most two unstable wavenumbers, $k_j = j/L = 10j$, where $j = 2, 3$ (see Fig. 2.1). With such a sparse sample, the eddy flux converges to $\langle cov(u')\rangle_L = 1.008$, uniformly in space. Consequently, the large-scale dynamics converges to

$$\begin{aligned}\hat{u}_{\ell,\infty} \equiv \lim_{t\to\infty}\hat{u}_\ell(t) &= \lim_{t\to\infty}\frac{\langle cov(u')\rangle_L + F_{ext}}{p(\mathrm{i}\ell)}\delta_{\ell,0}\Big(1-e^{p(\mathrm{i}\ell)t}\Big)\\ &= \frac{\langle cov(u')\rangle_L + F_{ext}}{d+\nu\ell^2}\delta_{\ell,0} = \begin{cases}50.8, & \ell = 0,\\ 0, & \ell \neq 0,\end{cases}\end{aligned}$$

which implies $\bar{u} \to \hat{u}_{0,\infty} = 50.8$, a trivial constant mean dynamics, as $t \to \infty$ (this is consistent with Fig. 3.2). In the above derivation, we define $\delta_{\ell,0}$ to equal zero when $\ell \neq 0$ and one only when $\ell = 0$.

In Table 3.1, we report the statistical prediction skill of the large scale variable, $\bar{u}$, obtained from superparameterization with various values of L compared to the truth (see Section 2.3) for the deterministically and stochastically forced cases. In particular, we compare the mean, covariance, spatial and temporal correlations, computed from statistical steady state solutions for the time interval $[200, 500]$. Notice that for both type of forcings, superparameterization with $L = 2$ produces the best statistical estimates. For $L = 0.5, 1$, we still reproduce the same physics with small-scale intermittent instabilities (results are not shown) but the statistical estimates deteriorate (see Table 3.1). When $L = 0.1$, the statistical estimates are not skillful with very large mean, as we discussed earlier. Intuitively, the improved statistical estimates with larger L is not so surprising since the trapezoidal approximation in (3.1) produces smaller errors in estimating the covariance integral (2.11). Notice also that the stochastically forced case has slightly smaller statistical estimates compared to the deterministically forced case. The smaller values of mean and covariances suggest that the large-scale stochastic forcing decreases $\bar{u}$ and simultaneously enhances the occurence of small-scale local intermittent instabilities.

TABLE 3.1

Statistical prediction skills: Mean, variance, spatial, and temporal correlations of $\bar{u}$ computed from statistical steady state solutions at time interval $[200, 500]$ from superparameterization with various values of scale gap, L, and the truth for large-scale deterministic and stochastic forcings.

	Mean	Cov	SpCorr	TemCorr
Det True	4.30	6.13	2.36	112.80
SP $L = 2$	4.32	6.00	2.38	113.62
SP $L = 1$	4.18	5.36	2.41	114.94
SP $L = 0.5$	3.83	3.68	2.51	120.07
SP $L = 0.1$	48.79	2.79	3.14	149.92
Stoch True	4.03	5.76	2.29	110.88
SP $L = 2$	4.16	5.67	2.38	113.17
SP $L = 1$	3.85	4.89	2.36	112.94
SP $L = 0.5$	3.81	3.92	2.48	118.14
SP $L = 0.1$	48.35	5.52	3.13	149.82

4. Filtering with superparameterization. We consider solutions of the test model in (2.3)-(2.8) as the underlying truth signal. We define our filtering problem with partial observations of the large-scale variable, $\bar{u}$, at discrete time and space. In particular, we consider observations at every p model grid points through the following observation model,

$$v_{j,m} = \bar{u}(X_j, t_m) + \sigma^o_{j,m}, \quad \sigma^o_{j,m} \sim \mathcal{N}(0, r^o), \tag{4.1}$$

where $X_j = j2\pi/M$ and $M = N/p$ denotes the total number of observations at time t_m (recall that $N = 128$ are the total grid points of the large scale mean dynamics). The observations $v_{j,m}$ are corrupted by spatially and temporally uncorrelated Gaussian noises with mean zero and variance r^o and available at every discrete time $t_m = mt_{obs}$, with observation time interval t_{obs}.

We will compare various filtering strategies, including:

1. The true filter as a diagnostic. This scheme essentially utilizes the test model described in Section 2 as the filter prior model.

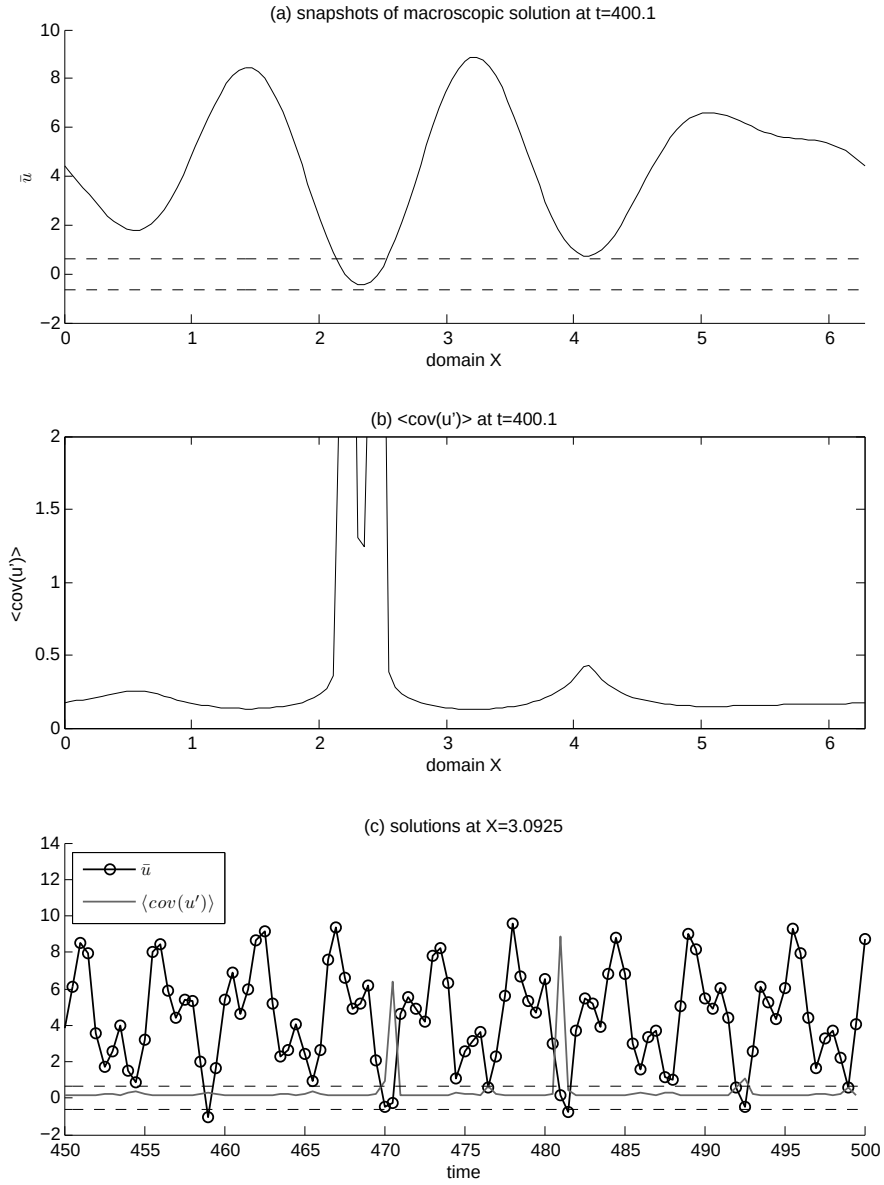

FIG. 3.1. *Deterministically forced superparameterized approximate solutions with $L = 2$: (a) Snapshots of $\bar{u}$ at time $t = 400.1$; (b) Small-scale eddy fluxes correspond to $\bar{u}$ at time $t = 400.1$; (c) Timeseries of $\bar{u}$ and $\langle cov(u')\rangle$ at $X = 3.0925$.*

2. The bare-truncation filter. This filtering strategy utilizes the bare-truncation model described in Section 2.1 as the filter prior model.
3. The equilibrium closure filter. This filtering strategy utilizes the statistical equilibrium closure model described in Section 2.1 as the filter prior model.
4. The superparameterization filter. This filtering scheme utilizes the superparameterizated model described in Section 3 as the filter prior model.

Below, we describe these four filtering strategies and the basic filtering algorithm for

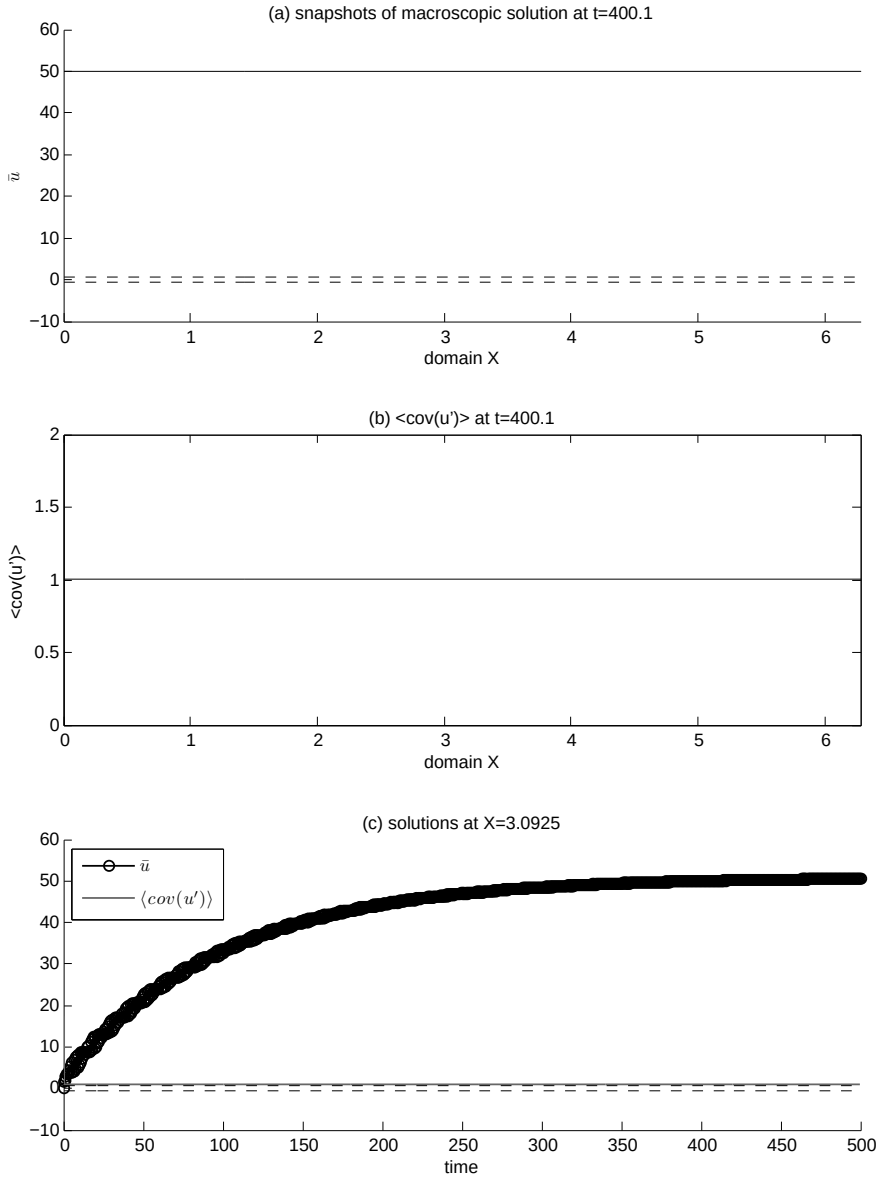

FIG. 3.2. *Deterministically forced superparameterized approximate solutions with* $L = 0.1$*: (a) Snapshots of* $\bar{u}$ *at time* $t = 400.1$*; (b) Small-scale eddy fluxes correspond to* $\bar{u}$ *at time* $t = 400.1$*; (c) Timeseries of* $\bar{u}$ *and* $\langle cov(u')\rangle$ *at* $X = 3.0925$*.*

regularly spaced sparse observations in (4.1) in detail.

4.1. Fourier Domain Kalman Filters for regularly spaced sparse observations. We consider implementing the four filtering strategies above in Fourier space rather than in physical space; this strategy is not only numerically beneficial but it also produces accurate solutions beyond the standard physical space approach as we reported in various contexts (see [25, 24] and the references therein). Essentially, this approach applies a spatial localization in Fourier domain to avoid spurious non-

local correlations that often occur when the filter covariance matrix is sub-optimally estimated [14, 11]. Below, we briefly describe the Fourier domain Kalman filter in detail [24, 25, 22, 9].

In Fourier space, the observation model in (4.1) is expressed as follows:

$$\hat{v}_{n,m} = \sum_{n_j \in \mathcal{A}(n)} \hat{u}_{n_j,m} + \hat{\sigma}^o_{n,m}, \quad |n| \le M/2 \tag{4.2}$$

where $\hat{v}_{n,m}$ is defined through $v_{j,m} = \sum_{|\ell| \le M/2} \hat{v}_{n,m} e^{\mathrm{i}nX_j}$, with $X_j = 2j\pi/M$ and $M = N/p = 128/p$; the coefficient $\hat{u}_{\ell,m} \equiv \hat{u}_\ell(t_m)$ is defined through discrete Fourier transform $\bar{u}(X_j, t) = \sum_{|\ell| \le N/2} \hat{u}_\ell(t) e^{\mathrm{i}\ell X_j}$, with $X_j = 2j\pi/N$ and $N = 128$. In (4.2), the noise term is Gaussian with mean zero and variance r^o/M, $\hat{\sigma}^o_{n,m} \sim \mathcal{N}(0, r^o/M)$, whereas the term $\mathcal{A}(n)$ denotes the aliasing set of wavenumber n defined as [9, 24]:

$$\mathcal{A}(n) = \Big\{ n_j = n + Mj \,|\, j \in \mathbb{Z},\, |n_j| \le \frac{N}{2} \Big\}. \tag{4.3}$$

For example, if observations are available at every $p = 4$ grid points of the total model grid points $N = 128$, we have a total of $M = N/p = 32$ observations. The corresponding discrete Fourier coefficients are available for wavenumbers $n = -15, -14, \ldots, -1, 0, 1, \ldots, 16$. The aliasing sets of wavenumbers $0, 1, 2, \ldots$ are given by

$$\begin{aligned} \mathcal{A}(0) &= \{-32, 0, 32, 64\}, \\ \mathcal{A}(1) &= \{-63, -31, 1, 33\}, \\ \mathcal{A}(2) &= \{-62, -30, 2, 34\}, \\ &\vdots \\ \mathcal{A}(16) &= \{-48, -16, 16, 48\}. \end{aligned}$$

In this case, the observation model in (4.2) couples $p = 4$ modes that are aliased to each other; any pair of Fourier coefficients belonging to different aliasing sets are uncorrelated. Hence, the Fourier domain Kalman filter consists of block diagonal filters, where each block solves a $p \times p$ filtering problem [9, 22, 24].

Let's define $\vec{u}_{n,m} = (\hat{u}_{n_1,m}, \ldots, \hat{u}_{n_p,m})^T$ and $G = (1, \ldots, 1) \in \mathbb{R}^{1\times p}$, where subscript '$T$' denotes a transpose. Then we can rewrite the observation model in (4.2) in vector form,

$$\hat{v}_{n,m} = G\vec{u}_{n,m} + \hat{\sigma}^o_{n,m}.$$

Given a prior mean state $\vec{u}^-_{n,m} = (\hat{u}^-_{n_1,m}, \ldots, \hat{u}^-_{n_p,m})^T$ and a $p\times p$ positive definite prior error covariance matrix, $R^-_{n,m}$, the standard Kalman filter formula [1] provides the posterior mean estimate, $\vec{u}^+_{n,m} = (\hat{u}^+_{n_1,m}, \ldots, \hat{u}^+_{n_p,m})^T$, and error covariance matrix, $R^+_{n,m}$, in explicit form,

$$\vec{u}^+_{n,m} = \vec{u}^-_{n,m} + K_{n,m}(\hat{v}_{n,m} - G\vec{u}_{n,m}), \tag{4.4}$$

$$R^+_{n,m} = (\mathcal{I}_p - K_{n,m}G)R^-_{n,m}, \tag{4.5}$$

where $\mathcal{I}_p$ denotes a $p \times p$ identity matrix and $K_{n,m}$ denotes a $p \times 1$ Kalman gain matrix,

$$K_{n,m} = R^-_{n,m}G^T(GR^-_{n,m}G^T + r^o/M)^{-1}, \tag{4.6}$$

where the computational cost involves only a scalar inversion and various matrix-vector multiplications.

For the test model in (2.3)-(2.8), the Fourier domain filtering is natural since the large scale solutions are obtained by solving the following linear SDE for each Fourier coefficient, conditional to $\hat{C}_\ell$,

$$\frac{d\hat{u}_\ell}{dt} = -p(\mathrm{i}\ell)\hat{u}_\ell + \hat{C}_\ell + \bar{F}\delta_{\ell,0} + \hat{\sigma}_\ell \dot{W}_\ell, \quad |\ell| \le N/2, \tag{4.7}$$

where $p(\mathrm{i}\ell)$ is the eigenvalue of linear differential operator $P(\partial_X)$ in (2.3). In our special numerical example in (2.20), $p(\mathrm{i}\ell) = (d + \nu k^2) + \mathrm{i}(ck - Ak^3)$ is a selective decay damping when $\nu \neq 0$. In (4.7), $\hat{C}_\ell$ denotes the discrete Fourier coefficient of the covariance,

$$\langle cov(u')\rangle(X_j, t) = \sum_{|\ell|\le N/2} \hat{C}_\ell(t) e^{\mathrm{i}\ell X_j}, \tag{4.8}$$

and we assume $\hat{C}_\ell$ is constant within a small integration step $(t, t + \Delta t)$ as described in Section 2.3. Therefore, in terms of $\hat{u}_\ell$ alone, the SDE in (4.7) is linear.

For the true filter, we use $\langle cov(u')\rangle(X, t)$ obtained by integrating (2.11) with adaptive quadrature as discussed in Section 2.3. For the bare-truncation model, we set $\langle cov(u')\rangle = 0$. For the equilibrium closure filter, we set $\langle cov(u')\rangle = \langle cov(u')\rangle_\infty$ from (2.13). For the superparameterization, we use $\langle cov(u')\rangle_L$ obtained from the trapezoidal approximation in (3.1). The last two terms in (4.7) correspond to the Fourier coefficients of the large-scale external forcing defined in (2.22) and (2.23). For the deterministic forcing case, we set $\hat{\sigma}_\ell = 0$.

Given a posterior mean state, $\hat{u}^+_{\ell,m}$, and a $p \times p$ error covariance matrix, $R^+_{n,m}$, at time t_m, both obtained from (4.4)-(4.6), we integrate the dynamical equation in (4.7) to obtain the prior mean, $\hat{u}^-_{\ell,m+1}$, and error covariance, $R^-_{\ell,m+1}$, at the next observation time, $t_{m+1} = t_m + t_{obs}$. In discrete-time formulation, conditional to $\hat{C}_{\ell,m}$, the prior statistics are explicitly given by:

$$\hat{u}^-_{\ell,m+1} = F_\ell \hat{u}^+_{\ell,m} + \frac{1}{p(\mathrm{i}\ell)}(1 - F_\ell)(\hat{C}_{\ell,m} + \bar{F}\delta_{\ell,0}), \quad |\ell| \le N/2 \tag{4.9}$$

$$R^-_{n,m+1} = \mathcal{F}_n R^+_{n,m} \mathcal{F}^*_n + \mathcal{Q}_n, \quad |n| \le M/2, \tag{4.10}$$

where $F_\ell = \exp(-p(\mathrm{i}\ell)t_{obs})$. In (4.10), we define $p \times p$ diagonal matrices for the dynamical operator and system noise covariance for Fourier coefficients in aliasing set $\mathcal{A}(n)$,

$$\mathcal{F}_n = \mathrm{diag}(F_{n_1}, \ldots, F_{n_p}), \quad \mathcal{Q}_n = \mathrm{diag}(Q_{n_1}, \ldots, Q_{n_p}),$$

where $Q_\ell = \frac{\hat{\sigma}^2_\ell}{2p(\mathrm{i}\ell)}(1 - |F_\ell|^2)$ is the variance of the solutions of (4.7); we denote $|u|^2 = u \cdot u^*$. The recursive formula in (4.4), (4.5), (4.6), (4.9), and (4.10) define the Fourier domain Kalman filter [9, 24]. The four filtering strategies above differ through their forcing terms, $\hat{C}_{\ell,m}$. Note that for a special case where we observe at all grid points, $p = 1$, the Fourier domain Kalman filter is completely diagonal with $M = N$ (see [22, 2, 24]).

The famous Kalman and Bucy [12] filtering stability criterion suggests that the filtered solutions are stable when the observability condition is satisfied. Furthermore, if the dynamical model is also controllable, then it is possible to obtain accurate Kalman

filtered solutions. In our case above, the necessary and sufficient conditions for observability are $F_{n_i} \neq F_{n_j}$ for any pair of $i, j \in \mathcal{A}(n)$ where $i \neq j$ (see Chapter 3 of [24]). These conditions are clearly satisfied in our test model since $P(\partial_X)$ in (2.20) defines a selective damping $p(\mathrm{i}\ell)$ when $\nu \neq 0$, and therefore, the Fourier domain Kalman filter defined through recursive steps in (4.4)-(4.6), (4.9) , (4.10) with prior model in (4.7) always satisfies the observability condition. The controllability condition is satisfied if $\mathcal{Q}_n$ has a full rank [24]; therefore, filtering with the deterministically forced prior model with $\hat{\sigma}_\ell = 0$ does not satisfy the controllability condition. On the other hand, filtering with the stochastically forced prior model always satisfies the controllability condition. With this theoretical guideline, we expect less accurate filtered solutions with deterministically forced prior models compared to the stochastically forced prior models even when the true signals are solutions of the same deterministically forced models; in some sense, the system noise becomes our "friend".

5. Filter performance on test models. In our numerical experiments below, we set the observation noise variance $r^o = 1.41$ (which is roughly about $23 - 25\%$ of the covariance of $\bar{u}$, depending on the forcing case, as recorded in Table 3.1) and the observation time interval $t_{obs} = 0.5$ (much shorter than the temporal correlation, see Table 3.1). We will use the average RMS error and spatial correlation (SC) between the posterior mean state $\vec{u}_m^+ = (\bar{u}^+(X_1, t_m), \ldots, \bar{u}^+(X_N, t_m))$ and the truth $\vec{u}_m = (\bar{u}(X_1, t_m), \ldots, \bar{u}(X_N, t_m))$ to measure the filtering skill,

$$\mathrm{RMS} = \frac{1}{T - T_0} \sum_{m=T_0+1}^{T} \sqrt{\left\langle (\vec{u}_m^+ - \vec{u}_m)^2 \right\rangle_N}, \tag{5.1}$$

$$\mathrm{SC} = \frac{1}{T - T_0} \sum_{m=T_0+1}^{T} \frac{\left\langle (\vec{u}_m^+ - \langle \vec{u}_m^+ \rangle_N)(\vec{u}_m - \langle \vec{u}_m \rangle_N) \right\rangle_N}{\sqrt{\langle (\vec{u}_m^+ - \langle \vec{u}_m^+ \rangle_N)^2 \rangle_N \langle (\vec{u}_m - \langle \vec{u}_m \rangle_N)^2 \rangle_N}}. \tag{5.2}$$

In (5.1)-(5.2), the temporal average ignores the first $T_0 = 100$ of the total $T = 1,000$ assimilation steps; we also define spatial average with $\langle \vec{u}_m \rangle_N = N^{-1} \sum_{j=1}^{N} \bar{u}(X_j, t_m)$.

5.1. Stochastically forced prior models. First, let's discuss numerical results from filtering stochastically forced truth signals with stochastically forced prior models. We'll discuss the more subtle case of filtering deterministically forced truth signals in Section 5.2.

In Fig. 5.1, we show the average RMS errors and SC of all four strategies described in Section 4 for regularly sparse observations in (4.1) with $p = 4, 8, 16, 32$. Notice the filtering skill with model errors from the best to the worst are the superparameterization filter with $L = 2$, the equilibrium closure filter, and the bare-truncation filter, subsequently. Second, the filtering skill of superparameterization with $L = 2$ (squares) is indistinguishable from that of the true filter (dashes); both schemes produce the smallest RMS and highest SC. As observations become sparser (or p increases), the filtering skill degrades as expected; the SP filter with $L = 2$ RMS is roughly 0.76 (still much below the observation error, $\sqrt{r^o}$ =1.18) and SC is roughly 0.9.

In Figs. 5.2-5.5, we show snapshots of the large scale posterior and prior mean estimates, $\bar{u}^+$ and $\bar{u}^-$ compared to the hidden true state, $\bar{u}$, at $t_m = 500$ (or after $T = 1000$ assimilation steps), for regularly sparse observations with $p = 4, 8, 16, 32$, respectively. Comparing these figures, we can see in detail the degradation of the accuracy of the filtered solutions for each method when observations are sparser. For $p = 4$, estimates from all four filtering strategies are nearly identical except near

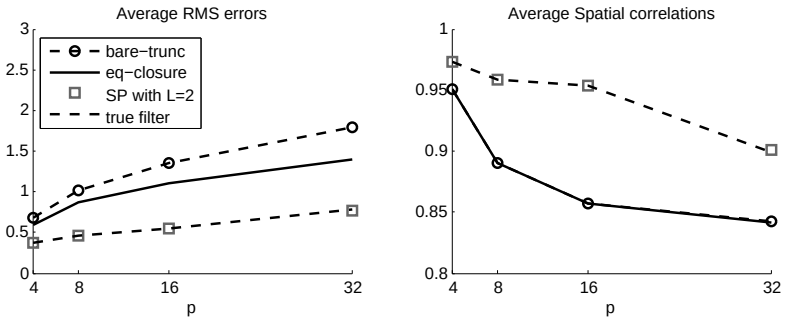

FIG. 5.1. *Average RMS errors and SC (spatial correlations) as functions of p.*

$X = 5$ where the bare truncation and equilibrium closure filters' estimates are slightly inaccurate. Notice the similarity of SP filter with $L = 2$ and the true filter even in a very sparse observation network case with $p = 32$ (see Fig. 5.5); here the bare truncation and equilibrium estimates are much worse compared to the SP filter.

In Fig. 5.6, we plot the average RMS and SC from SP filter for sparse observations with $p = 8$ as functions of scale gap, L. Here, we only simulate with $L = 0.1, 0.5, 1, 2$, similar to those in Table 3.1. Notice the sudden degradation of filtering skill with $L = 0.1$ (see also Fig. 5.7) as we expected since this scale gap produces stable dynamics without local intermittent instabilities as discussed in Section 3 (recall the results in Fig. 3.2 and Table 3.1). On the other hand, the filtering skill with $L = 0.5, 1, 2$ are comparable (again, see Fig. 5.7); essentially, the filtering skill of the very cheap SP with $L = 0.5$ is comparable to that of $L = 2$ which we know is almost identical to the skill of the true filter (see Fig. 5.1).

5.2. Controllability. In Fig. 5.8, we show numerical results of filtering the deterministically forced true signal with the perfect deterministically forced prior filter model for sparse observations with $p = 4$. Both the true and SP filtered solutions reproduce the wave pattern inaccurately. The bare truncation and equilibrium closure estimates converge to $\bar{u}^+ \approx -49.29$ and -31.3, respectively, without any turbulent fluctuations. These inaccurate filtered solutions are due to the lack of controllability as we discussed earlier in Section 4 when the prior model noise covariance is zero, $Q_\ell = 0$. Numerically, the prior covariance update in (4.10) is so small when $Q_\ell = 0$ that the Kalman gain matrix in (4.6) has $K_{n,m} \approx 0$, and subsequently, the uncontrollable filter trusts the prior mean estimates completely. For bare truncation and equilibrium closure, the filtered solutions are nothing but the stable equilibrium solutions of these models. See also Table 5.1 for their corresponding average RMS errors and SC. We also find similar numerical outcomes for smaller observation noise variance, r^o (results are not shown).

In Fig. 5.9, we consider filtering exactly the same observations as in Fig 5.8 with a stochastically forced prior model, where $Q_\ell \neq 0$ is chosen such that the large-scale dynamics in (2.3) without nonlinear covariance term, $\langle cov(u') \rangle$, has ℓ^{-6} spectrum, similar to the spectrum of the deterministically forced true signals (see Fig 2.2). Notice the dramatic improvement in the filtering skill (see Fig. 5.9 and Table 5.1) when controllability condition is restored. The choice of ℓ^{-6} spectrum is completely adhoc, one can investigate further for the better choices and we are not interested in that aspect here. Our point is to show the importance of the controllability condition in

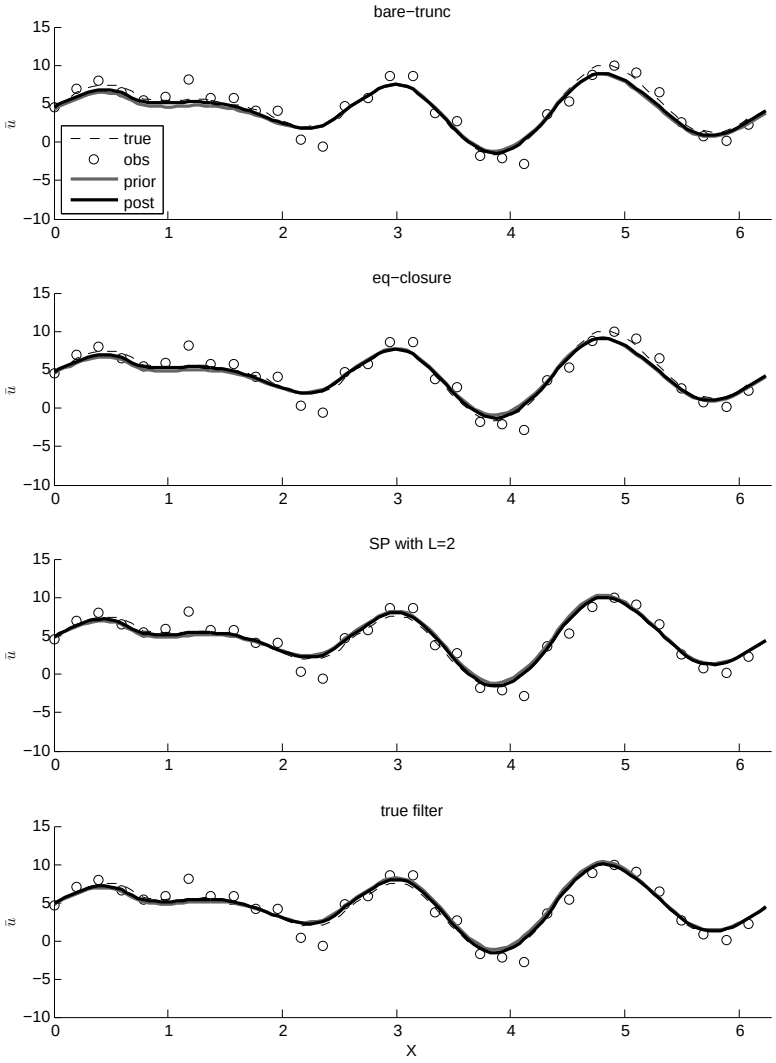

FIG. 5.2. *Case $p = 4$: (a) Snapshots of at $t_m = 500$ (or after $T = 1000$ assimilation steps). The bare truncation filter (first row), the equilibrium closure filter (second row), the superparameterization filter with $L = 2$ (third row), and the true filter (fourth row).*

TABLE 5.1
Average RMS errors and SC for filtering deterministic truth with and without controllability.

	Not controllable		Controllable	
Scheme	RMS	SC	RMS	SC
bare truncation	43.1004	0.2314	0.6331	0.9677
eq-closure	28.7007	0.2784	0.5362	0.9677
SP with $L = 2$	1.1180	0.8364	0.2775	0.9859
true filter	1.1320	0.8325	0.2777	0.9859

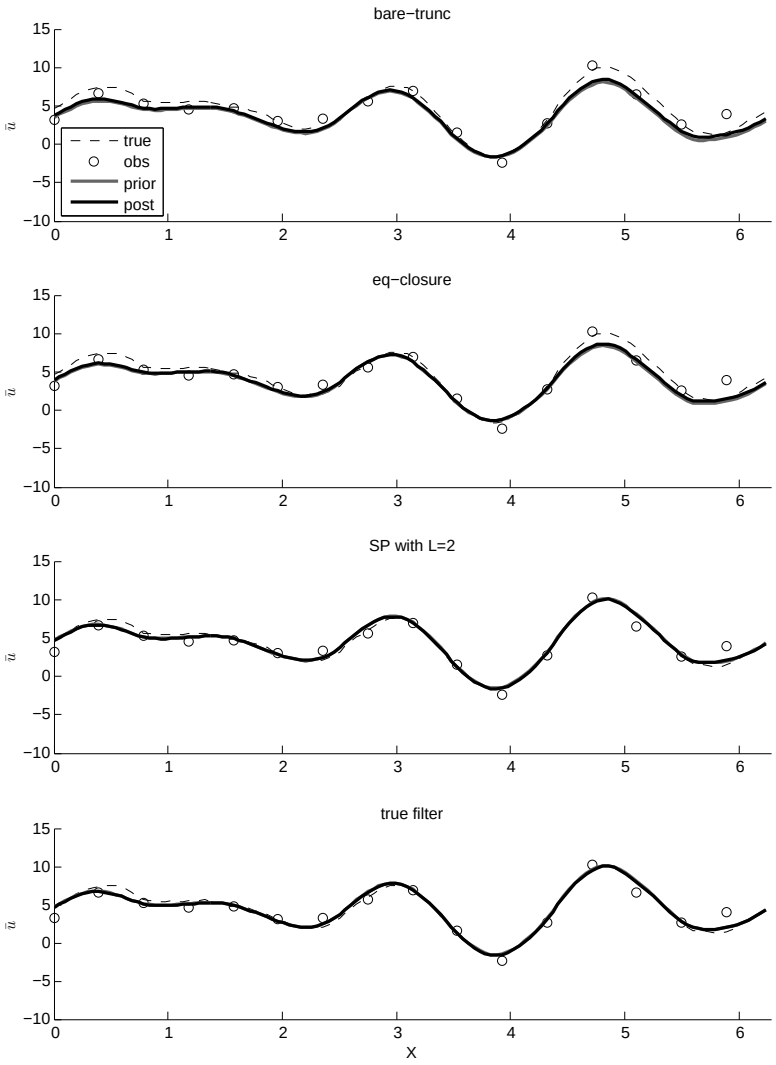

FIG. 5.3. *Case $p = 8$: (a) Snapshots of at $t_m = 500$ (or after $T = 1000$ assimilation steps). The bare truncation filter (first row), the equilibrium closure filter (second row), the superparameterization filter with $L = 2$ (third row), and the true filter (fourth row).*

obtaining accurate filtered solutions. A similar result holds for filtering stochastically forced true signals; in that case, we obtain inaccurate filtered solutions when we use deterministically forced prior filter models (results are not shown). We should note that all the results in Section 5.1 hold for deterministically forced true signals provided that the filter prior model satisfies controllability condition, that is, $Q_\ell \neq 0$.

If we are able to observe the large scale dynamics at a very high spatial resolution with small errors (which may not be the case in real applications), then the filtered solutions will discard the prior model and completely rely on these observations. To solidify this argument, we show the RMS errors and SC for simulations with no observations error $r^o = 0$ but satisfying the controllability condition in Figure 5.10. Notice that there are no difference in skills when $p = 4$ (observations are spatially

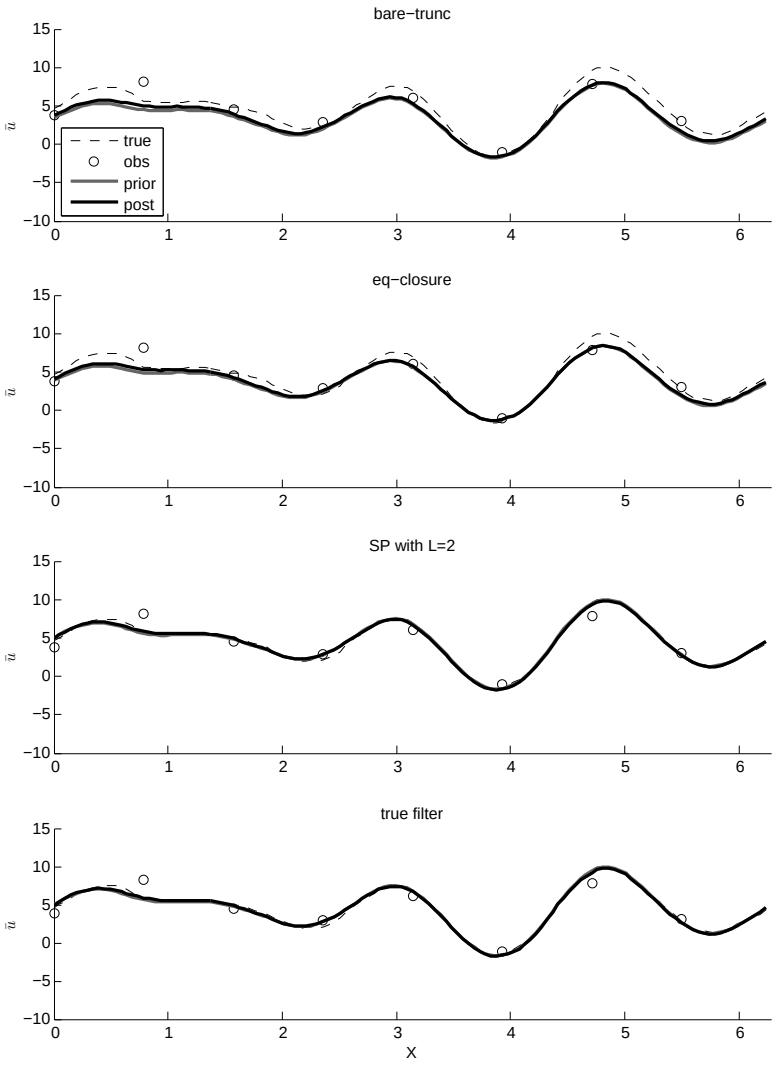

FIG. 5.4. *Case $p = 16$: (a) Snapshots of at $t_m = 500$ (or after $T = 1000$ assimilation steps). The bare truncation filter (first row), the equilibrium closure filter (second row), the superparameterization filter with $L = 2$ (third row), and the true filter (fourth row).*

dense). For sparse observations, the study in this section suggests that the choice of prior model is very important. Notice the inaccurate filtered solutions from both the bare-truncation and equilibrium closure filters compared to that of the SP with $L = 2$, even when controllability is satisfied (see Figs. 5.9, 5.10, and Table 5.1). An additional remarkable fact is that the true signal here has a very steep spectrum, with ℓ^{-6}; one may be led to believe that for such a steep spectrum, the small scale dynamics may not be that important. Our numerical results demonstrate that we can't ignore the small scale covariance $\langle cov(u') \rangle$ nor replace it with the equilibrium closure estimate $\langle cov(u') \rangle_\infty$. Indeed, we have to carefully estimate this small scale covariance. Even superparameterization fails when the scale gap is too large with $L = 0.1$.

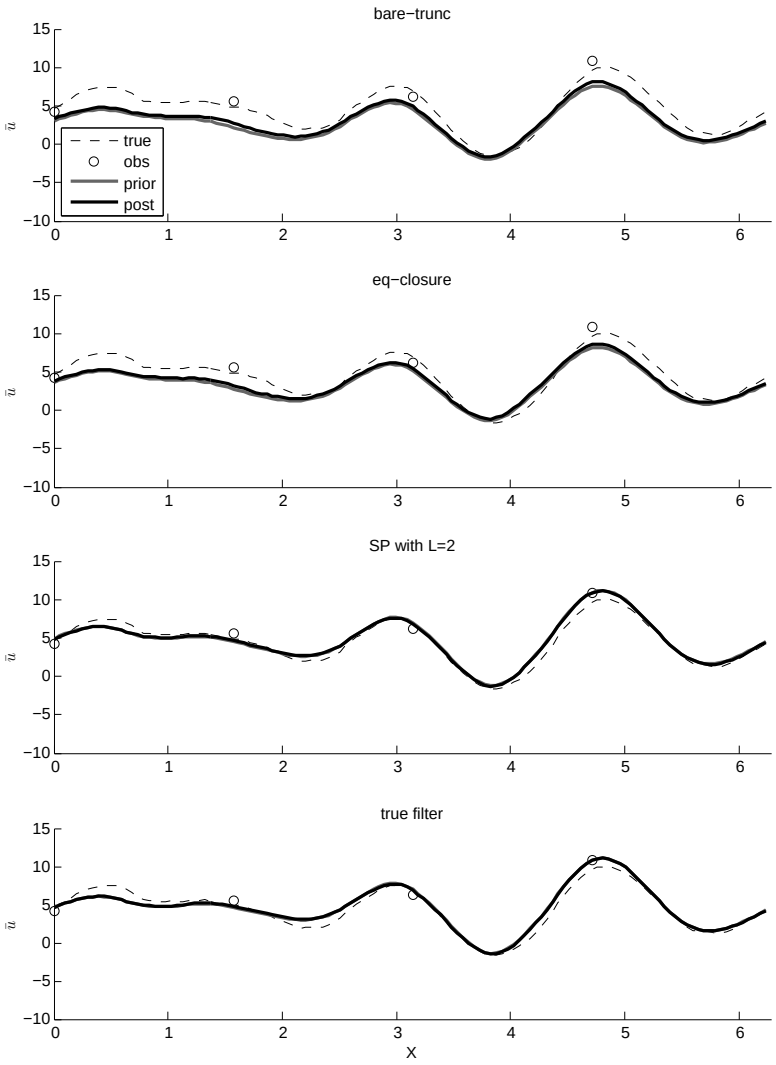

FIG. 5.5. *Case $p = 32$: (a) Snapshots of at $t_m = 500$ (or after $T = 1000$ assimilation steps). The bare truncation filter (first row), the equilibrium closure filter (second row), the superparameterization filter with $L = 2$ (third row), and the true filter (fourth row).*

6. Concluding Discussions. In this paper, we study the effect of model errors in using superparameterization for data assimilation with a simple test model with semi-analytic solutions. The test model is designed to mimic typical multiscale turbulent dynamics [15, 16, 6, 18] with small-scale intermittencies without local statistical equilibriation conditional to the large scale dynamics, which simultaneously force the large scale dynamics.

With such an idealized test model, we unambiguously compare the effect of model errors, introduced through various approximations of the small-scale eddy fluxes, in filtering regularly spaced sparse observations and predicting equilibrium statistics. In particular, we compare three different approaches with model errors: the bare truncation model, the equilibrium closure model, and the superparameterized model. We

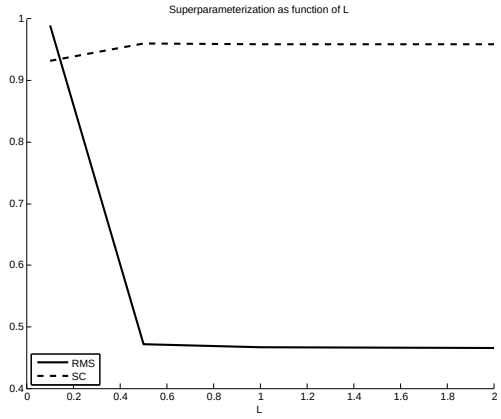

FIG. 5.6. *Case $p = 8$: Average RMS errors and spatial correlations as functions of L.*

find robust high filtering and statistical prediction skill with superparametrized model (with skill almost identical to that of the true model), when the superparameterization scale gap is not too large (L is not too small). The high filtering skill with superparameterization is robust even when the observation network is very sparse, as long as the filter models satisfy the controllability condition with sufficient nonzero system noise covariance, $Q_\ell \neq 0$. A spectacular failure of filtering the deterministically forced true signal with the exactly perfect model that does not satisfy controllability and a dramatic improvement when controllability condition is restored with additional stochastic forcings reconfirm that judicious model errors (or noises) can sometimes help filtering, as we encountered in different data assimilation contexts [24]. On the other hand, the other two approaches, the bare truncation and equilibrium closure models, have much lower skill in filtering as well as statistical prediction even when the true signal has a very steep spectrum with ℓ^{-6} and controllability condition is satisfied! In this case, these two models produce steady state solutions (without turbulence) since the small-scale eddy flux terms that generate turbulence are ignored or not properly estimated. Furthermore, a violation of controllability condition sets the filtered solutions to be the steady state solutions, ignoring any given observations. This is a counter-example to naive thinking that the small-scale processes are not so important in multiscale turbulent dynamics with steep energy spectrum.

In the near future, we will investigate the effect of other kinds of model errors in filtering with superparameterization: (1) When the small-scale dynamics are resolved on n-dimensional periodic domain, additional model errors are introduced in the classical superparameterization with suitable reduced models on a one-dimensional L-periodic domain, replacing the n-dimensional periodic domain [16, 6, 5]; (2) Model errors through a sparse-time variant of superparameterization [30] that judiciously approximates the eddy fluxes, $\langle cov(u') \rangle$, with reduced time integration of the small-scale dynamics; (3) model errors from elimination of direct nonlocal interaction between smaller scale processes for complex nonlinear dispersive wave turbulence [7]. A much more difficult problem is to filter observations of mixed typed, involving functions of the large and small scale variables, which often arise in some applications; for example, in geophysical fluid applications, the observed temperature, moisture, and velocity necessarily mix both the slow and fast modes [3]. In a more realistic setting, we may need to appropriately account for the initial conditions of the unobserved

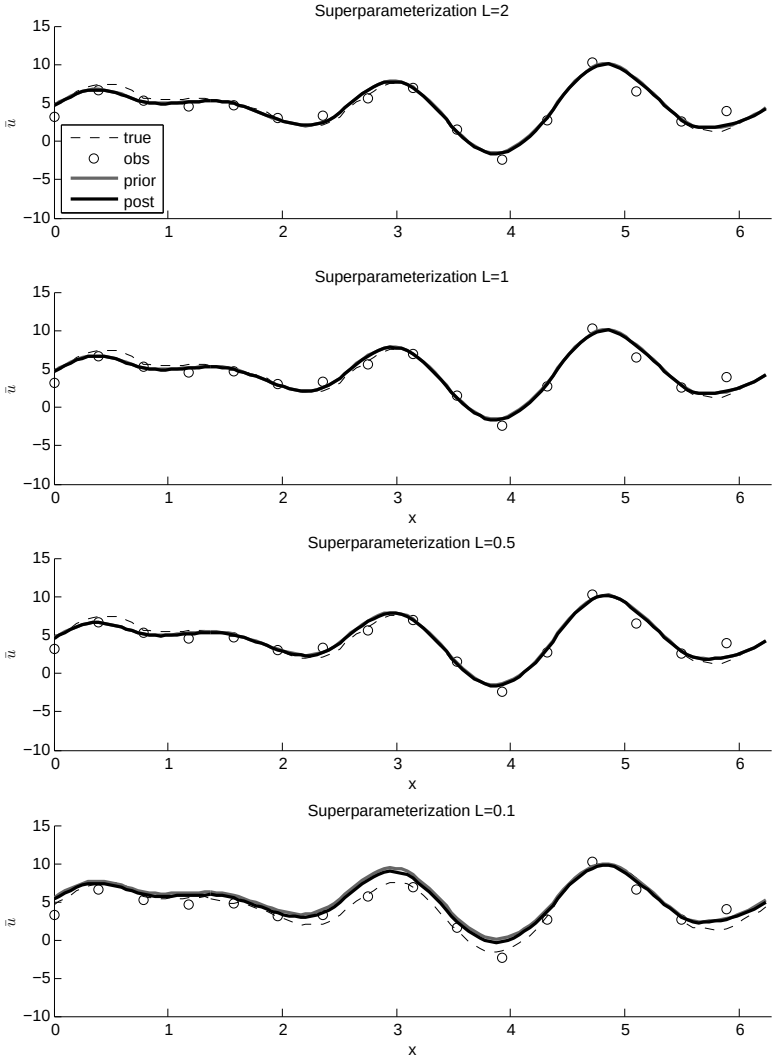

FIG. 5.7. *Snapshots of at $t_m = 500$ (or after $T = 1000$ assimilation steps) for $p = 8$ for superparameterization with various L.*

eddies such that they are statistically consistent with the large scale variables, $\bar{u}$. We will explore the possibility of utilizing the small-scale reinitialization approach introduced in [8, 13] as well as other options.

REFERENCES

[1] B.D. ANDERSON AND J.B. MOORE, *Optimal filtering*, Prentice-Hall Englewood Cliffs, NJ, 1979.
[2] E. CASTRONOVO, J. HARLIM, AND A.J. MAJDA, *Mathematical test criteria for filtering complex systems: plentiful observations*, Journal of Computational Physics, 227 (2008), pp. 3678–3714.
[3] DALEY, *Atmospheric data analysis*, Cambridge University Press, New York, 1991.
[4] W. E, B. ENGQUIST, X. LI, W. REN, AND E. VANDEN-EIJNDEN, *The Heterogeneous Multi-Scale Method: A review*, Commun. Comput. Phys, 2 (2007), pp. 367–450.

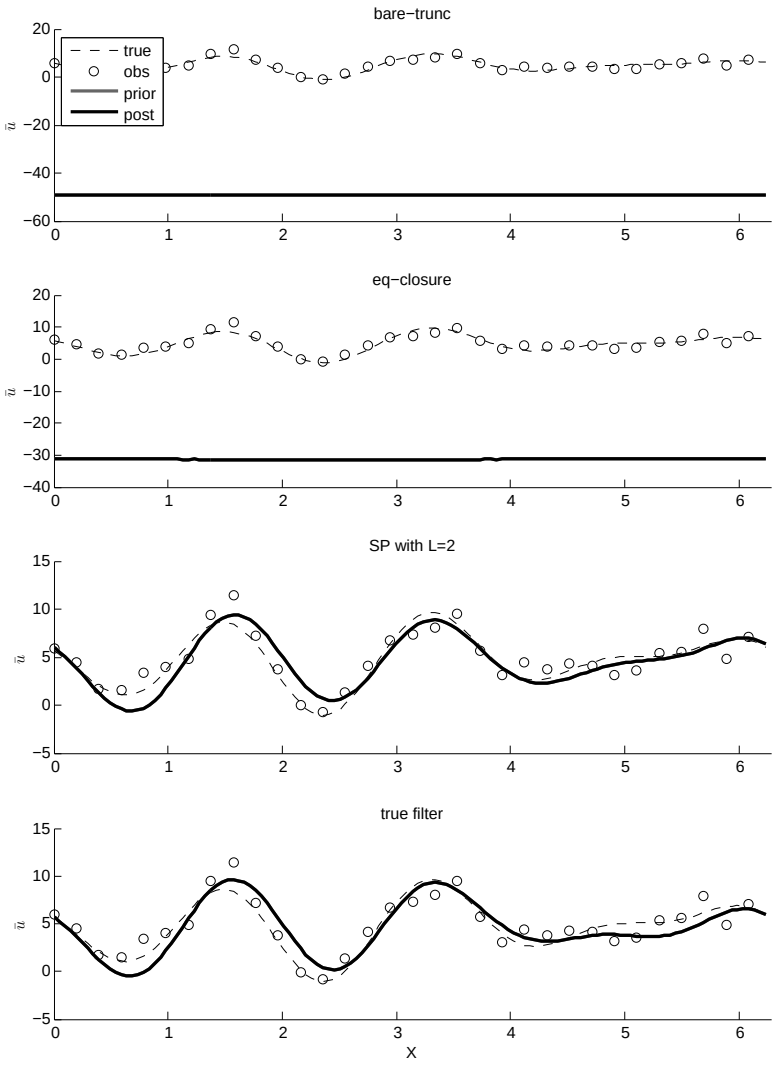

FIG. 5.8. *Case $p = 4$ filtered with $Q_\ell = 0$: (a) Snapshots of at $t_m = 500$ (or after 1000 assimilation steps). The bare truncation filter (first row), the equilibrium closure filter (second row), the superparameterization filter with $L = 2$ (third row), and the true filter (fourth row).*

[5] W.W. GRABOWSKI, *Coupling cloud processes with the large-scale dynamics using the Cloud-Resolving Convection Parameterization (CRCP)*, J. Atmos. Sci., 58 (2001), pp. 978–996.
[6] W.W. GRABOWSKI AND P.K. SMOLARKIEWICZ, *CRCP: a Cloud Resolving Convection Parameterization for modeling the tropical convecting atmosphere*, Physica D, 133 (1999), pp. 171–178.
[7] I. GROOMS AND A.J. MAJDA, *Stochastic superparameterization in a one-dimansional model for wave turbulence*, submitted to Multiscale Model. Simul., (2012).
[8] J. HARLIM, *Numerical strategies for filtering partially observed stiff stochastic differential equations*, J. Comput. Phys., 230 (2011), pp. 744–762.
[9] J. HARLIM AND A.J. MAJDA, *Mathematical strategies for filtering complex systems: Regularly spaced sparse observations*, Journal of Computational Physics, 227 (2008), pp. 5304–5341.
[10] J.O. HINZE, *Turbulence*, McGraw-Hill classic textbook reissue series, McGraw-Hill, 1975.
[11] B.R. HUNT, E.J. KOSTELICH, AND I. SZUNYOGH, *Efficient data assimilation for spatiotemporal*

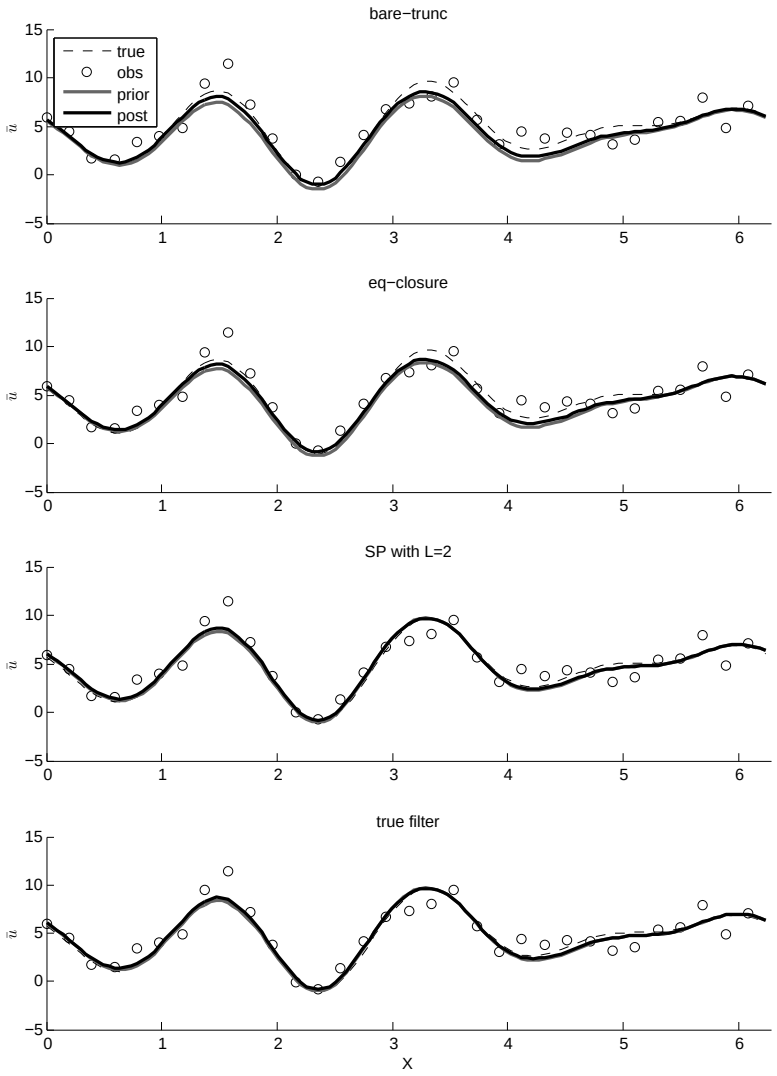

FIG. 5.9. *Case $p = 4$ filtered with Q_ℓ corresponds to ℓ^{-6}: (a) Snapshots of at $t_m = 500$ (or after 1000 assimilation steps). The bare truncation filter (first row), the equilibrium closure filter (second row), the superparameterization filter with $L = 2$ (third row), and the true filter (fourth row).*

chaos: a local ensemble transform Kalman filter, Physica D, 230 (2007), pp. 112–126.

[12] R.E. KALMAN AND R. BUCY, *New results in linear filtering and prediction theory*, Trans. AMSE J. Basic Eng., 83D (1961), pp. 95–108.

[13] E.L. KANG AND J. HARLIM, *Filtering partially observed multiscale systems with heterogeneous multiscale methods-based reduced climate models*, Monthly Weather Review, 140 (2012), pp. 860–873.

[14] C.L. KEPPENNE, *Data assimilation into a primitive-equation model with a parallel ensemble Kalman filter*, Monthly Weather Review, 128 (2000), pp. 1971–1981.

[15] A.R. KERSTEIN, *A linear- eddy model of turbulent scalar transport and mixing*, Combustion Science and Technology, 60 (1988), pp. 391–421.

[16] ———, *One-dimensional turbulence: model formulation and application to homogeneous turbulence, shear flows, and buoyant stratified flows*, Journal of Fluid Mechanics, 392 (1999),

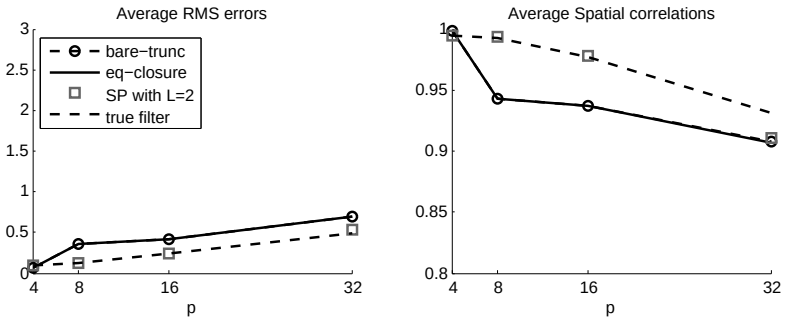

FIG. 5.10. *Controllable and no observation errors ($r^o = 0$): Average RMS errors and SC (spatial correlations) as functions of p.*

pp. 277–334.

[17] R. KHASMINSKII, *On averaging principle for it stochastic differential equations*, Kybernetika, Chekhoslovakia (in Russian), 4 (1968), pp. 260–279.

[18] R. KLEIN AND A.J. MAJDA, *Systematic multiscale models for deep convection on mesoscales*, Theo. Comp. Fluid Dyn., 20 (2006), pp. 525–551.

[19] T.G. KURTZ, *Semigroups of conditional shifts and approximations of Markov processes*, Annals of Probability, 3 (1975), pp. 618–642.

[20] A.J. MAJDA, *Multiscale models with moisture and systematic strategies for superparameterization*, Journal of the Atmospheric Sciences, 64 (2007), pp. 2726–2734.

[21] ———, *Challenges in climate science and contemporary applied mathematics*, Communications on Pure and Applied Mathematics, 65 (2012), pp. 920–948.

[22] A.J. MAJDA AND M.J. GROTE, *Explicit off-line criteria for stable accurate time filtering of strongly unstable spatially extended systems*, Proceedings of the National Academy of Sciences, 104 (2007), pp. 1124–1129.

[23] ———, *Mathematical test models for superparametrization in anisotropic turbulence*, Proc. Nat. Acad. Sci., 2106 (2009), pp. 5470–5474.

[24] A.J. MAJDA AND J. HARLIM, *Filtering Complex Turbulent Systems*, Cambridge University Press, UK, 2012.

[25] A.J. MAJDA, J. HARLIM, AND B. GERSHGORIN, *Mathematical strategies for filtering turbulent dynamical systems*, Discrete and Continuous Dynamical Systems A, 27 (2010), pp. 441–486.

[26] A.J. MAJDA AND Y. XING, *New multi-scale models on mesoscales and squall lines.*, Commun. Math. Sci., 8 (2010), pp. 113–134.

[27] A.S. MONIN, A.M. IAGLOM, AND J.L. LUMLEY, *Statistical fluid mechanics: mechanics of turbulence*, no. v. 2 in Statistical Fluid Mechanics: Mechanics of Turbulence, MIT Press, 1975.

[28] G. PAPANICOLAOU, *Some probabilistic problems and methods in singular perturbations*, Rocky Mountain J. Math, 6 (1976), pp. 653–673.

[29] A.A.R. TOWNSEND, *The Structure of Turbulent Shear Flow*, Cambridge monographs on mechanics and applied mathematics, Cambridge University Press, 1980.

[30] Y. XING, A.J. MAJDA, AND W.W. GRABOWSKI, *New efficient sparse space-time algorithms for superparameterization on mesoscales*, Monthly Weather Review, 137 (2009), pp. 4307–4323.

[31] A.M. YAGLOM, *An Introduction to the Theory of Stationary Random Functions*, Dover Phoenix Editions, Dover Publications, 2004.