NUMERICAL SCHEMES FOR STOCHASTIC BACKSCATTER IN
THE INVERSE CASCADE OF QUASIGEOSTROPHIC
TURBULENCE

IAN GROOMS†‡, YOONSANG LEE†§, AND ANDREW J. MAJDA†§

Abstract. Backscatter is the process of energy transfer from small to large scales in turbulence; it is crucially important in the inverse energy cascades of two-dimensional and quasigeostrophic turbulence, where the net transfer of energy is from small to large scales. A numerical scheme for stochastic backscatter in the two-dimensional and quasigeostrophic inverse kinetic energy cascades is developed and analyzed. Its essential properties include a local formulation amenable to implementation in finite difference codes and non-periodic domains, smooth behavior at the coarse grid scale, and realistic temporal correlations. The scheme is related to a simplified version of stochastic superparameterization [Grooms and Majda, J. Comp. Phys. 271, (2014)]; the simplified version allows detailed numerical analysis, focusing on the spatial and temporal correlation structure of the modeled backscatter. The method is demonstrated in an idealized setting of quasigeostrophic turbulence using a low-order finite difference code, where it produces a good approximation to the results of a spectral code with more than 5 times higher nominal resolution. When the essential properties of the new scheme are integrated into the full stochastic superparameterization, which includes downscale potential energy transfer in addition to kinetic energy backscatter, the results are in excellent agreement with the high resolution reference simulation. The method is of particular importance to eddy-permitting ocean modeling applications in the inverse cascade regime, but it should also serve to improve stochastic backscatter algorithms in regimes of turbulence with a net downscale cascade of energy.

Key words. Backscatter; Quasigeostrophic Turbulence; Stochastic Modeling

1. Introduction. Backscatter is the process of energy transfer from small to large scales in turbulence. Quasigeostrophic (QG) and two-dimensional (2D) turbulence, which are approximate models for large-scale geophysical turbulence, both include regimes with a net transfer of kinetic energy from small to large scales [5,7]. While the dynamical regime of the atmosphere precludes significant net transfer to large scales, there is evidence of an inverse transfer of kinetic energy in the oceans [31,32]. Ocean models used in coupled climate simulations are currently unable to resolve the full range of scales that correspond to quasigeostrophic dynamics, and must model the effects of the unresolved scales. When the dynamics transfer net energy upscale across the model grid scale, a backscatter model is crucial to maintain an accurate energy level for the resolved scales.

Representations of backscatter are included in many engineering and atmospheric turbulence models, where the backscatter is much weaker. For example, backscatter is incorporated in large-eddy simulations (LES) of three-dimensional isotropic turbulence either by including a stochastic term in the equations for the resolved dynamics [19,24,25,30], or through nonlinear deterministic models [2,10,26]. In larger-scale geophysical models backscatter has also been represented by stochastic forcing terms [3,9,15,16,28] and by nonlinear deterministic models [15,27,28]. These models have been applied in situations where the net transfer of kinetic energy is either downscale or negligible, whereas, as noted above, application to non-eddy-resolving ocean models requires the ability to represent a net upscale transfer. Backscatter

† Center for Atmosphere Ocean Science, Courant Institute of Mathematical Sciences, New York University, 251 Mercer St., New York, NY 10012
‡ Corresponding author, email address: grooms@cims.nyu.edu. I. Grooms is supported by the NSF CMG program grant DMS-1025468, and by ONR MURI grant N00014-12-1-0912.
§ Y. Lee and A. J. Majda are supported by ONR MURI grant N00014-12-1-0912.
schemes are also often formulated in terms of spherical harmonics [3,9,16], which is appropriate for applications to fluids in a spherical shell, like the atmosphere. But application to domains with boundaries, like the ocean, precludes the use of Fourier modes or spherical harmonics in specifying the backscatter model; a local approach is required, similar to those in [15,19,24,25,28,30].

The approach taken here is based on a simplified form of stochastic superparameterization [12–14,22] that has previously been applied to quasigeostrophic turbulence in the inverse cascade regime. The novelty of the current investigation is the application of the simplified method in an under-resolved model with low-order finite-difference numerics, in analogy with the configuration of non-eddy-resolving ocean models. Direct application of the original method in this low-order finite-difference code gives poor results. The method was originally developed in a setting where the under-resolved model used a spectral discretization [12,13]; the huge impact of simply changing the numerics underscores the sensitivity of models to the details of backscatter parameterizations in the inverse cascade regime.

The failure of the original method in the context of a low-order code results primarily from the fact that the low-order model is unable to efficiently transfer energy injected near the coarse grid scale to larger scales. This is remedied by modifying the spatial correlation structure of the backscatter scheme so that it injects zero energy at the coarse grid scale, and smoothly transitions to the original scheme at larger, more well-resolved scales. The scheme is further improved by adding temporal correlation to the stochastic backscatter forcing, since forcing of the large scales by the unresolved small scales has a finite decorrelation time. The updated scheme gives good results in a quasigeostrophic model experiment, but fails to give a perfect result because it only models the inverse transfer of kinetic energy and not the direct transfer of potential energy from large to small scales. The full stochastic superparameterization scheme of [13] models transfers of both potential and kinetic energy, and incorporation of the essential features of the new scheme into the full superparameterization gives excellent results.

The failure of the original method in the context of a low-order code results primarily from the fact that the low-order model is unable to efficiently transfer energy injected near the coarse grid scale to larger scales. This is remedied by modifying the spatial correlation structure of the backscatter scheme so that it injects zero energy at the coarse grid scale, and smoothly transitions to the original scheme at larger, more well-resolved scales. The scheme is further improved by adding temporal correlation to the stochastic backscatter forcing, since forcing of the large scales by the unresolved small scales has a finite decorrelation time. The updated scheme gives good results in a quasigeostrophic model experiment, but fails to give a perfect result because it only models the inverse transfer of kinetic energy and not the direct transfer of potential energy from large to small scales. The full stochastic superparameterization scheme of [13] models transfers of both potential and kinetic energy, and incorporation of the essential features of the new scheme into the full superparameterization gives excellent results.

The paper is organized as follows. Section 2 presents and analyses the stochastic kinetic energy backscatter scheme, including the incorporation of spatial and temporal correlation. Section 3 presents a high resolution simulation for comparison with low resolution models in following sections. The results of low resolution models without backscatter is presented in §4, to emphasize the need for backscatter. Sections 5 and 6 show the results of backscatter simulations with and without spatial and temporal correlations; §7 presents the results of the full stochastic superparameterization scheme of [13] with the new spatial and temporal correlation properties. The results are reviewed and conclusions are offered in §8.

2. The Stochastic Backscatter Model. Quasigeostrophic (QG) and incompressible two-dimensional (2D) fluid dynamics are governed by conservation of a scalar $q$, which is vorticity in 2D and potential vorticity in QG:

\[
\partial_t q + \nabla_h \cdot (u q) = \dot{q}
\]  

(2.1)

where $\dot{q}$ denotes forcing and dissipation terms and $\nabla_h = (\partial_x, \partial_y)$. The scalar $q$ is related to a streamfunction $\psi$ for the incompressible horizontal velocity $u = (-\partial_y \psi, \partial_x \psi)$ by an elliptic inversion

\[
\nabla_h^2 \psi + \mathbb{L} \psi = q
\]  

(2.2)
where $L$ is a linear operator. In 2D dynamics $L = 0$ and in QG it is a second-order negative semi-definite differential operator in $z$. Sections 3-7 will focus on the simplest model of geophysical turbulence, the two-equal-layer QG model [35], where the vertical direction is discretized into two levels $j = 1, 2$ and where the elliptic inversion is

$$\nabla_h^2 \psi_j + \frac{k_j^2}{2} (\psi_i - \psi_j) = q_j, \ i \neq j.$$  \hspace{1cm} (2.3)

The deformation radius $k_j^{-1}$ is related to the nonzero eigenvalue of $L$. The discussion generalizes to multiple vertical levels in a straightforward manner.

In the two-layer case there is a scalar conservation equation of the form (2.4) for each layer:

$$\partial_t q_j + \nabla \cdot (u_j q_j) = \hat{q}_j.$$  \hspace{1cm} (2.4)

The unforced, inviscid dynamics with $\hat{q}_j = 0$ conserve energy in the form

$$\frac{1}{2} \iint \sum_j |u_j|^2 + \frac{k_j^2}{2} (\psi_1 - \psi_2)^2 \, dx \, dy$$

where the integral of $k_j^2 (\psi_1 - \psi_2)^2/4$ is the potential energy.

An equation for the large-scale dynamics can be derived by applying a low-pass spatial filter, denoted by $\langle \cdot \rangle$, to (2.4)

$$\partial_t \overline{q}_j + \nabla \cdot (\overline{u}_j \overline{q}_j) = \nabla \cdot (\overline{u}_j q_j - \overline{u}_j \overline{q}_j) + \overline{\hat{q}}_j.$$ \hspace{1cm} (2.5)

Examples of low-pass filters include truncation to a set of large-scale Fourier modes, or a local Gaussian convolution, among others [21] (see also §3 and §6).

The subgridscale flux of potential vorticity can be further decomposed as follows

$$\overline{u}_j \overline{q}_j - \overline{u}_j \overline{q}_j = (\overline{u}_j \overline{q}_j - \overline{u}_j \overline{q}_j) + \overline{u}_j \overline{q}_j + \overline{u}_j \overline{q}_j$$ \hspace{1cm} (2.7)

where the prime denotes the small-scale part $(\cdot)' = (\cdot) - \langle \cdot \rangle$. These terms can all be modeled separately [2, 8, 10, 20]; we instead model their combined effect while formally assuming the filter is a Reynolds average, i.e. that it has the same properties as a statistical average. This approach follows the development of stochastic superparameterization [12, 13, 22], and has been used e.g. by [30] in the development of a stochastic backscatter scheme in incompressible three-dimensional turbulence. Under this assumption the divergence of the subgridscale potential vorticity flux becomes

$$\nabla \cdot (\overline{u}_j \overline{q}_j) = \nabla \cdot (\overline{u}_j \nabla^2 \psi_j) + \frac{k_j^2}{2} \nabla \cdot (\overline{u}_j (\psi_j - \psi_j)). $$ \hspace{1cm} (2.8)

The horizontal Laplacian of $\psi_j'$ is the subgridscale vorticity $\nabla_h \psi_j' = \zeta_j' = \partial_x v_j' - \partial_y u_j'$. The subgridscale potential vorticity flux $\overline{u}_j \overline{q}_j'$ includes the vorticity flux $\overline{u}_j' \zeta_j'$, whose divergence can be further expanded as

$$\nabla_h \cdot (\overline{u}_j \overline{\zeta}_j) = \left( \partial_x^2 - \partial_y^2 \right) \overline{u}_j \overline{v}_j' + \partial_{xy} \left( \overline{v}_j'^2 - \overline{u}_j'^2 \right). $$ \hspace{1cm} (2.9)

This is the curl of the divergence of the Reynolds stress tensor $\overline{u}_j' \otimes \overline{u}_j'$. It has been shown by [31] that the subgridscale vorticity flux divergence is responsible for the inverse kinetic energy cascade in two-layer QG turbulence. The stochastic kinetic energy
The backscatter scheme developed here models the subgridscale vorticity flux divergence, ignoring the remaining part of the subgridscale potential vorticity flux.

The effect of subgridscale Reynolds stresses can be modeled in general as the sum of a deterministic drain on the resolved scales plus a stochastic forcing [16, 17, 21]. Careful analysis shows that the deterministic drain can be modeled to a good approximation as a hyperviscous drain of momentum, which in this context implies that the drain component of the vorticity flux divergence takes the form $\nu_n (\nabla^2) \zeta_j$ where $n \geq 0$ depends on the grid scale and $\nu_n > 0$ is the turbulent hyperviscosity coefficient [16]. Eddy-permitting ocean models typically use a biharmonic viscosity ($n = 2$), which we adopt here, leaving the coefficient $\nu_4$ as a tunable parameter.

The stochastic backscatter scheme developed here is based on the ‘uncorrelated’ closure from [13, section 3.3]. The subgridscale velocity in this closure is based on modeling the subgridscale dynamics by randomly-oriented plane waves with wavenumbers in the direction $\theta$. In this closure the subgridscale velocity is assumed to be completely independent of the large-scale flow. Note that the assumption of independence is relaxed in the full stochastic superparameterization developed in [13]; the spatial and temporal correlation properties developed in this section for the backscatter scheme are incorporated into stochastic superparameterization in §7. Under the assumption of independence from the large-scale flow, the multiscale description of the subgridscale velocity field in [13] results in subgrid stresses of the form

$$\overline{u_j'v_j'} = \frac{E_0}{2} \sin(2\theta), \quad (\overline{u_j'}^2 - (\overline{u_j'})^2) = E_0 \cos(2\theta)$$

(2.10)

where $E_0$ is a constant subgridscale kinetic energy and $\theta$ is a random field, white in space and time and with a uniform distribution $\theta \in [0, 2\pi)$. These stresses would also be generated by subgridscale velocities with amplitude $\sqrt{E_0}$ in the direction orthogonal to $\theta$, i.e.

$$u_j' = \sqrt{E_0} \sin(\theta), \quad v_j' = \sqrt{E_0} \cos(\theta).$$

(2.11)

The direction $\theta$ is depth-independent, although a more general vertical correlation structure is clearly possible. In this scheme the Reynolds stresses are thus identical in each layer; in the full stochastic superparameterization scheme the Reynolds stresses are not completely set by the angle $\theta$, and can be different in each layer even when $\theta$ is the same in each layer. The backscatter scheme of [13], combined with a biharmonic drain, thus takes the form

$$\nabla \cdot (\overline{u_j'\zeta_j'}) = \frac{E_0}{2} (\partial_x^2 - \partial_y^2) \sin(2\theta) + E_0 \partial_{xy} \cos(2\theta) + \nu_4 \nabla^4 \zeta_j$$

(2.12)

The backscatter amplitude $E_0$ and hyperviscosity coefficient $\nu_4$ are treated here as tunable parameters, although if a prediction were available for the subgridscale kinetic energy, e.g. through the subgridscale energy model of [13, section 3.4], it could be used to provide a spatio-temporally variable prediction of $E_0$.

2.1. Spatial Correlation. We now demonstrate that the model (2.10) leads to a random forcing of the large-scale dynamics with a $k^3$ spectrum. The large-scale potential vorticity equation (2.6) has the form

$$\partial_t \overline{\eta_j} = -\nabla \cdot (\overline{u_j \eta_j}) + \frac{E_0}{2} (\partial_y^2 - \partial_x^2) \sin(2\theta) - E_0 \partial_{xy} \cos(2\theta) - \nu_4 \nabla^4 \zeta_j + \overline{\eta}.$$  

(2.13)
Recall that the random field $\theta$ has uniform distribution on the periodic interval $[0, 2\pi)$ and that it is spatially white, i.e. the values at different horizontal locations are independent. These properties of $\theta$ imply that the terms $\sin(2\theta)$ and $\cos(2\theta)$ are delta-correlated (white) in space, have zero mean, and are uncorrelated with each other since

$$E[\sin(2\theta)\cos(2\theta)] = \frac{1}{2\pi} \int \sin(2\theta)\cos(2\theta) d\theta = 0$$

(2.14)

where $E$ denotes the expectation.

Recalling that the Fourier transform of a delta-correlated (white) random field is uniform, the Fourier transform of the backscatter forcing terms is proportional to

$$\frac{1}{2}(k_x^2 - k_y^2)A_k + k_xk_yB_k$$

(2.15)

where $A_k$ and $B_k$ are uncorrelated random variables with zero mean and equal variance. The expected value of the square amplitude of the Fourier transform is thus proportional to $k^4$, and the azimuthally-averaged (1D) spectrum is proportional to $k^5$. This implies that the stochastic backscatter scheme forces most strongly at the coarse grid scale.

For comparison, a $k^4$ spectrum is expected for momentum backscatter in homogeneous isotropic three-dimensional turbulence [17, 21, 30]. Schumann [30] points out that although this can be justified by appeal to the Eddy-Damped Quasi-Normal Markovian (EDQNM) model of turbulence, it is also expected based solely on the assumption that the components of the Reynolds stress are independent, and spatially uncorrelated. These are also the only assumptions that go into the proof of the $k^5$ spectrum for the current model, which suggests that the $k^5$ behavior is a robust quality of the backscatter spectrum in quasigeostrophic turbulence.

The backscatter scheme developed above was shown in [13] to give good results when used with a low resolution model with spectral numerics, but in §5 we demonstrate that it performs poorly in a model with a second order finite-difference discretization. The poor performance is a unique aspect of modeling in a regime where the energy backscattered from unresolved scales to scales near the coarse grid scale is then transferred to even larger scales in an inverse cascade; the nonlinear dynamics of the inverse cascade are not as well represented in the second-order finite difference model as in the spectral model, especially near the coarse grid scale.

In the theory of large-eddy simulation (LES) of homogeneous isotropic three-dimensional turbulence it is common to identify the solution of a low resolution model with a low-pass filtered version of the true dynamics, as in our derivation of the large-scale dynamics (2.6). If the low resolution model uses a spectral numerical method then it is natural to use a Fourier truncation as the filter, projecting the true dynamics onto the spectrum of Fourier modes represented by the low resolution model. In contrast, it is more appropriate to identify low-order finite difference models with a locally smoothed version of the true dynamics, e.g. convolution with a Gaussian. Leslie & Quarini [21] showed that the sharp $k^4$ backscatter spectrum is associated with a Fourier truncation filter, whereas the backscatter spectrum associated with a smooth Gaussian filter is attenuated near the coarse grid scale, implying that the spectrum of the backscatter forcing in a low-order finite difference code should be attenuated near the coarse grid scale. This can be heuristically associated with the
Fig. 2.1. Angle-integrated (1D) backscatter forcing spectra that result from the three local-average smoothers: the three-point average $S_3$, the five-point average $S_5$, and the three-point average applied twice $S_3^2$. The plots assume a $96 \times 96$ coarse grid, with Nyquist wavenumber 48.

fact that low order finite-difference methods represent the dynamics of the smallest resolved scales less accurately than spectral methods.

To adapt the backscatter scheme for use in the finite-difference code we apply spatial smoothing to the Reynolds stresses. This spatial smoothing should be defined by a local average in order to be easily implementable in a finite-difference code. Also, since the Nyquist wavenumber is represented on the coarse grid but its dynamics are not resolved we require the smoothing operator to damp the spectrum of the Reynolds stresses to zero at the Nyquist wavenumber. There is only one symmetric three-point moving average that completely removes the Nyquist wavenumber, and its coefficients are $1/4$, $1/2$, $1/4$. Application of this average in the $x$ and $y$ directions modifies the Fourier coefficients of the backscatter forcing terms to

$$\cos^2\left(\frac{k_x\Delta x}{2}\right)\cos^2\left(\frac{k_y\Delta y}{2}\right)\left[\frac{1}{2}(k_x^2 - k_y^2)A_k + k_xk_yB_k\right]$$

(2.16)

where $\Delta x$ and $\Delta y$ are the grid spacing and $A_k$ and $B_k$ are as in equation (2.15). The 2D forcing spectrum is proportional to $\cos^4(k_x\Delta x/2)\cos^4(k_y\Delta y/2)k^4$, which is no longer isotropic, but reduces the forcing to zero for modes at the Nyquist wavenumbers $k_x = \pi/\Delta x$ and $k_y = \pi/\Delta y$. This particular choice of smoothing is ad hoc, but serves to reduce the forcing smoothly to zero at the coarse grid scale without affecting the forcing at large scales, and is easily implemented. In §5 we compare to two alternate smoothers: double application of the $1/4$, $1/2$, $1/4$ average, and application of a five-point moving average with coefficients $-1/16$, $1/4$, $5/8$, $1/4$, and $-1/16$. Whereas the three-point smoother damps the coefficient of a Fourier mode by a factor of $\cos^2(k_x\Delta x/2)\cos^2(k_y\Delta y/2)$, the five-point smoother damps the coefficient of a
Fourier mode by

\[ \frac{1}{4} \cos^2(k_x \Delta x/2)(3 - \cos(k_x \Delta x)) \cos^2(k_y \Delta y/2)(3 - \cos(k_y \Delta y)). \]  

(2.17)

This smoothing operator sets the Nyquist wavenumbers to zero, but damps the remaining modes less than the three-point average.

The angle-averaged (1D) forcing spectra that result from these three smoothing operators are plotted in figure 2.1, assuming a 96×96 coarse grid. There are significant differences between these forcing spectra, both in the amplitude of the peak forcing and in the wavenumber of the peak forcing. Nevertheless, in §5 we find that there is little qualitative difference in the results for these three different smoothers. The main difference is that the optimal forcing amplitude \( E_0 \) is different for the three filters; it is higher for the double smoother, and lower for the five-point smoother.

It is worth noting that the finite difference approximation to the derivatives of the Reynolds stress terms in equation (2.12) does not reproduce exactly a \( k^5 \) forcing spectrum due to truncation errors. For example, the standard centered second-order difference approximation of \( \partial_x^2 \) multiplies a single Fourier mode by \(-2 \sin^2(k_x \Delta x/2)/\Delta x^2\) rather than \(-k_x^2\). However, as shown in §5 the smoothing effect of the second-order finite-difference approximation is not sufficient to give good results in the finite difference code without additional smoothing of the Reynolds stresses.

Leith [19], Mason & Thomson [25], and Schumann [30] developed stochastic backscatter schemes for low-order codes. Only Mason & Thomson included spatial smoothing of the Reynolds stresses (via the same three-point average used here), but Schumann noted that it was not necessary since good results could be obtained without it. We expect that the primary difference is that in their applications the net energy transfer was downscale. Energy backscattered near the grid scale of the coarse model simply dissipates away under the linear action of coarse-model viscosity, and this process is accurately represented even by low-order numerics, whereas in the inverse cascade it must be nonlinearly transferred to larger scales, and low-order numerics do not represent these nonlinear dynamics accurately near the grid scale.

2.2. Temporal Correlation. Leith [19], Mason & Thomson [25], and two of the current authors [12,13] have developed backscatter schemes where the subgrid-scale stresses are white (delta-correlated) in time; this is implemented simply by updating the random variables at each time step [12,13,19], or every other time step [25]. Of course, real subgrid stresses have a nonzero decorrelation time because they are generated by deterministic subgrid-scale dynamics, and Schumann [30], following Bertoglio [4], incorporated a nonzero decorrelation time in his model of stochastic backscatter.

To give the subgrid stresses a realistic nonzero decorrelation time we model the random field \( \theta \) at each coarse grid point by a Weiner process on the circle

\[ d\theta = \sigma dW. \]  

(2.18)

We prove in the appendix that this leads to an exponential lagged autocorrelation with a decorrelation time of \((2\sigma^2)^{-1}\) for the modeled Reynolds stress terms \( u'u' \) and \( (v')^2 - (u')^2 \). The proof makes use of the Fokker-Planck equation for (2.18) on a circle, which has the uniform distribution as a stationary solution. The evolution of the Dirac delta under the action of the Fokker-Planck equation is given in terms of its Fourier series, which together with some trigonometric identities allows us to
prove that the lagged autocorrelation function of $\sin(2\theta)$ or $\cos(2\theta)$ is exponential with decay scale $(2\sigma^2)^{-1}$.

Although the spatially-discretized fluid equations with stochastic backscatter are SDEs, the time integration methods used with these backscatter schemes are ad hoc, because the SDEs are generally not in the standard form for which numerical methods are well studied, i.e. $dX = f(X,t)dt + g(X,t)dW$ where $W$ is a Weiner process. Leith [19] notes that in order for his method to be robust to changing time steps the amplitude of the forcing should vary as $(\Delta t)^{-1/2}$ where $\Delta t$ is the time step size; since the forcing is multiplied by $\Delta t$ in the time integration scheme the overall method is analogous to the Euler method for Itô SDEs, which multiplies the stochastic forcing term by $(\Delta t)^{1/2}$ instead of $\Delta t$. Although the backscatter scheme developed here does not result in a standard-form SDE, we find similar behavior in §6: when the backscatter is white in time the scheme is robust to changing time steps when the amplitude $E_0$ is scaled by the square root of the time step.

We implement the backscatter forcing as follows. When the forcing is white in time the integration scheme multiplies the forcing terms by $(\Delta t)^{1/2}$ instead of $\Delta t$ and the forcing terms are held constant through all stages of the Runge–Kutta time step. When the forcing has an exponential decorrelation as described above, the integration scheme multiplies the forcing terms by $\Delta t$ as usual, and the angle $\theta$ evolves under equation (2.18) during the internal stages of the Runge–Kutta time step. We find in §6 that both of these schemes are robust to changes in the time step.

3. An Eddy-Resolving Baseline Simulation. Development of the backscatter model is provided here in the context of the two-layer quasigeostrophic (QG) model

$$
\partial_t q_1 = -\nabla \cdot (u_1 q_1) + \partial_x q_1 - k_d^2 v_1 - \nu_8 \nabla^8 q_1,
$$

$$
\partial_t q_2 = -\nabla \cdot (u_2 q_2) + \partial_x q_2 + k_d^2 v_2 - \nu_8 \nabla^8 q_2 - r \nabla^2 \psi_2 \tag{3.1}
$$

$$
q_1 = \nabla^2 \psi_1 + \frac{k_d^2}{2} (\psi_2 - \psi_1), \quad q_2 = \nabla^2 \psi_2 - \frac{k_d^2}{2} (\psi_2 - \psi_1),
$$

where $q_j$ is the potential vorticity in the upper ($j = 1$) and lower ($j = 2$) layers, $\nabla^2 \psi_j = \zeta_j$ is the relative vorticity, 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 $\nu_8$ is the hyperviscous Reynolds number. A large-scale zonal (east-west, $x$-direction) vertical shear is applied with equal and opposite unit velocities in the upper and lower layers; this imposed velocity is associated with the terms $-1/k_d^2(\partial_x q_j + k_d^2 v_j)$. The imposed vertical shear leads to a baroclinic instability that drives the system dynamics [35]. The doubly-periodic, two-layer QG model driven by imposed shear is a classical idealized model of large-scale turbulence in the atmosphere and oceans.

A high resolution reference simulation of equations (3.1) in a $2\pi$-periodic domain with a $512 \times 512$ computational grid was run to provide a baseline for comparison with the low resolution simulations of the next sections. The simulations use a Fourier spectral method with semi-implicit fourth-order Runge–Kutta time integration; details of the numerical method are exactly the same as [13]. As in [12, 13], the friction parameter is $r = 16$ and the deformation wavenumber is $k_d = 50$, but we here use a smaller viscous coefficient $\nu = 10^{-17}$, which is the same as used by [34]. The different viscous coefficient has minimal impact on the results compared to [13], affecting only the smallest scales. The simulation starts from a low-energy initial condition, and
Figure 3.1. Properties of the high-resolution reference simulation: (a) Growth rate of linear baroclinic instability, (b) Snapshot of upper-layer potential vorticity \( q_1 \), (c) Time-averaged kinetic (solid) and potential (dashed) energy spectra. The thin vertical line in (c) indicates the deformation wavenumber \( k_d = 50 \).

statistics are computed after the total kinetic energy is saturated, which occurs at approximately \( t = 4 \).

Figure 3.1(a) shows the growth rate of the linear instability that drives the turbulence. Specifically, it shows the growth rate of Fourier modes of the linearized system (3.1); the most unstable modes at fixed \( k_x \) have \( k_y = 0 \) so only these modes are shown. The most unstable mode has \( k_x \approx 31.3 \), and the instability drops to negligible rates for \( k > k_d = 50 \). Figure 3.1(b) is a snapshot of the upper-layer potential vorticity \( q_1 \), which is dominated by small-scale coherent vortices. Figure 3.1(c) shows the time-averaged kinetic (solid) and potential (dashed) energy spectra; the thin vertical line indicates the deformation wavenumber \( k_d = 50 \). The imposed shear is equivalent to the effect of an imposed streamfunction in the upper and lower layers of the form \((-1)^j y\), which is associated with a meridional potential vorticity gradient of the form \((-1)^j k_d^2\). The dynamics generate a net potential vorticity flux \( v_j q_j \) that acts counter to this imposed gradient. The net potential vorticity flux is equal and opposite in each layer, and is proportional to the area integral of \((v_1 + v_2)(\psi_1 - \psi_2)/4\), which we refer to as the ‘heat flux’. The reference simulation generates a net, time-averaged heat flux of 226.

As described in §2, the backscatter amplitude \( E_0 \) can be related to the subgridscale kinetic energy via equation (2.11); a similar relation holds for the full stochastic superparameterization scheme of [13]. Although we treat \( E_0 \) as a tunable parameter in the following sections, it is of interest to diagnose the subgridscale kinetic energy for comparison with the tuned values of \( E_0 \). To diagnose the subgrid kinetic energy we define a low-pass filter \( \langle \cdot \rangle \) by multiplying the Fourier coefficients of the high resolution reference simulation by \( 1 - (k/48)^{12} \) for \( k \leq 48 \), and by 0 for \( k > 48 \), where \( k = (k_x^2 + k_y^2)^{1/2} \). This has the effect of smoothly transitioning from large scales with \( k < 48 \) to small scales with \( k \geq 48 \), where 48 is the Nyquist wavenumber of the coarse model grid. A corresponding high-pass filter is defined by \( (\cdot)' = (\cdot) - \langle \cdot \rangle \). For comparison with \( E_0 \) we diagnose

\[
(u_j')^2 + (v_j')^2
\]

from the high resolution reference simulation. The distribution is skewed to large values with a median of 57 and a mean of 268. We tested low-pass filters with exponents 8 and 10 instead of 12, and found that the diagnosed subgridscale energy is not very sensitive to the exponent in the filter,
4. Low-Order and Spectral Model Results at Eddy-Permitting Resolutions without Backscatter. In this section we describe a low-order code for the two-layer QG equations that is analogous to large-scale ocean models; we call this code the ‘ocean’ code to distinguish it from the code that uses fully spectral numerics. We then compare simulations of the spectral and ocean codes at various low resolutions and without backscatter schemes, emphasizing the differences between the codes and the need for backscatter.

Most large-scale ocean models use low-order numerical methods based on finite differences, finite elements, or finite volumes. Also, when the spatial grid is sufficient to allow some eddy variability (‘eddy-permitting’), they typically use biharmonic diffusion of momentum. The ‘ocean’ code solves the large-scale QG equations (2.13) in the following form

\[
\begin{align*}
\partial_t \mathbf{\eta}_1 &= -\nabla \cdot (\mathbf{u}_1 \mathbf{\eta}_1) - \partial_x \mathbf{\eta}_1 - k_2^2 \mathbf{v}_1 - \nu_4 \nabla^4 \mathbf{\eta}_1, \\
\partial_t \mathbf{\eta}_2 &= -\nabla \cdot (\mathbf{u}_2 \mathbf{\eta}_2) + \partial_x \mathbf{\eta}_2 + k_2^2 \mathbf{v}_2 - \nu_4 \nabla^4 \mathbf{\eta}_2 - r \nabla^2 \mathbf{\psi}_2
\end{align*}
\]

using the energy- and enstrophy-conserving second-order finite-difference discretization of the Jacobian developed by Arakawa [1]. All other spatial derivatives are approximated using spectral Fourier methods, and in this section there is no backscatter, i.e. \( E_0=0 \). Time integration is achieved using the third-order Runge–Kutta scheme of [6], paired with a PI.3.4 adaptive time step algorithm [33].

We run simulations using both the spectral and ocean codes on \( N \times N \) grids with at \( N = 96, 128, 192, 256, \) and \( 384 \); the viscous coefficients \( \nu_8 \) and \( \nu_4 \) are tuned at each resolution to give results with reasonable energy spectra. The goal is to find the minimal resolution necessary to achieve good results in each code. Figure 4.1 shows the time-averaged kinetic and potential energy spectra from simulations with the spectral code at \( N = 512, 192, \) and \( 128 \). The simulation with \( N = 192 \) used \( \nu_8 = 10^{-14} \), and the simulation with \( N = 128 \) used \( \nu_8 = 5 \times 10^{-13} \). Although the simulation with \( N = 192 \) is not perfect, the most energetic modes are approximately correct and the mean heat flux is 238, quite close to the true value of 226. At \( N = 128 \)
the total energy drops significantly, and the flux drops to 109, approximately half the correct value; the poor results at $N = 128$ cannot be ameliorated by tuning the viscosity.

Figure 4.2 shows the time-averaged kinetic and potential energy spectra from simulations with the spectral code at $N = 512$, and from the ocean code at $N = 384$, 256, and 96. There is similar behavior to the spectral model results, but at higher resolution: the results at $N = 384$ are acceptable (the heat flux is 245), but the results at $N = 256$ have too little energy and too little heat flux (172), which can’t be improved by tuning the viscosity. The results at $N = 96$ are terrible, but they are included for comparison with results in the following sections that use $N = 96$ and have a backscatter parameterization.

The significant differences between the performance of the under-resolved spectral and ocean codes are striking. The spectral code performs much better at low resolution than the ocean code, which underscores the importance of developing parameterizations in the more difficult and more realistic context of the ocean code. The spectral and ocean codes have different viscous operators and Runge–Kutta integration schemes, but we have verified that the discretization of the advection terms is primarily responsible for the differences in performance by running simulations using the same biharmonic viscosity and Runge–Kutta scheme as the ocean code but using the spectral Jacobian (not shown).

5. Stochastic Backscatter: Spatial Smoothing. In this section we examine the effect of spatial smoothing on the backscatter model. The ocean code is used to solve the equations

$$
\partial_t \bar{q}_1 = -\nabla \cdot (\mathbf{u}_1 \bar{q}_1) + \frac{E_0}{2} \left( \partial_y^2 - \partial_x^2 \right) [S \sin(2\theta)] - E_0 \partial_y [S \cos(2\theta)] - \partial_x \bar{q}_1 - k^2 \bar{q}_1 - \nu_4 \nabla^4 \bar{\eta}_1, \quad (5.1)
$$

FIG. 4.2. Time-averaged kinetic (left) and potential (right) energy spectra from simulations with the spectral code at $N = 512$, and from the ocean code at $N = 384$, 256, and 96.
\[
\partial_t \overline{\eta_2} = -\nabla \cdot (\overline{\eta_2} \overline{\eta_2}) + \frac{E_0}{2} \left( \partial_y^2 - \partial_x^2 \right) [S \sin(2\theta)] - E_0 \partial_{xy} [S \cos(2\theta)] \\
+ \partial_x \overline{\eta_2} + k_2^2 \overline{\eta_2} - r \nabla^2 \overline{\psi_2} - \nu_4 \nabla^4 \overline{\zeta}_2 
\]  

(5.2)

where \( S \) denotes a smoothing operator. We denote no smoothing by \( S_0 \), the three-point moving average with coefficients \( 1/4, 1/2, 1/4 \) in each direction by \( S_3 \), and the five-point moving average with coefficients \( -1/16, 1/4, 5/8, 1/4, -1/16 \) in each direction by \( S_5 \). Double application of the three-point moving average is denoted \( S^2_3 \).

In all simulations in this section \( \theta \) is a random field that is white in time. Following [12,13] the random field \( \theta \) is updated at the beginning of each Runge–Kutta time step, and held constant through all stages. The ocean code uses a time step \( \Delta t = 5 \times 10^{-5} \); for comparison, the spectral simulation with \( N = 512 \) adaptively chooses a time step on the order of \( 2.5 \times 10^{-5} \), so \( \Delta t = 5 \times 10^{-5} \) should be sufficiently small in the coarse code at \( N = 96 \). Furthermore, as shown in the next section the results are robust to varying the step size by a factor of 2 in either direction.

In all simulations both \( E_0 \) and \( \nu_4 \) are tuned to optimize the kinetic and energy spectra in comparison with the spectral reference simulation. Overall the results are less sensitive to viscosity than to the backscatter amplitude, so for the sake of brevity we choose to present results only for \( \nu_4 = 2 \times 10^{-5} \). This value gives reasonable results with all types of smoothing operators.

Figure 5.1 shows the time-averaged kinetic and potential energy spectra from simulations with no smoothing; in one simulation the derivatives of the backscatter terms are evaluated using a spectral numerical method, and in the other they are evaluated using centered second-order finite differences. Both simulations use the same backscatter amplitude \( E_0 = 76 \). Neither simulation gives good results (and the results can’t be qualitatively improved by using different \( E_0 \) or \( \nu_4 \)), but the simulation with spectral differencing gives worse results. This simply illustrates the beneficial smoothing effect of finite differences, but shows that this smoothing by itself is insufficient. All remaining simulations in this section, and in \( \S 6 \) and \( \S 7 \), use second order
finite differences to approximate the derivatives acting on the Reynolds stress terms.

Figure 5.2 shows the time-averaged kinetic and potential energy spectra from simulations where three different spatial smoothing operators are applied to the Reynolds stress terms: the five-point local average $S_5$, the three-point local average $S_3$, and the three-point average applied twice $S_2^2$. The $S_5$ simulation uses $E_0 = 69$, the $S_3$ simulation uses $E_0 = 92$, and the $S_2^2$ simulation uses $E_0 = 123$; these values are between the median value of 57 and the mean value of 268 diagnosed from the reference simulation. The primary difference in the results is in the kinetic energy spectra, at scales between the peak and the grid scale. As might be expected, the simulation with the least smoothing ($S_5$) has more kinetic energy at these scales than the simulation with the most smoothing ($S_2^2$).

All the coarse simulations have relatively accurate kinetic energy spectra, but the potential energy spectra all have a pronounced bump at small scales. This pileup of potential energy at small scales may be related to the dual energy cascade of quasigeostrophic turbulence [29, 31]: potential energy is transferred to scales near the deformation radius, where it is converted to kinetic energy before being cascaded back to large scales. In the low resolution model simulations the coarse grid Nyquist wavenumber is $k = 48$, and the dynamics of potential energy conversion near the deformation wavenumber $k_d = 50$ are poorly resolved. Thus, potential energy cascades to small scales, but piles up instead of being converted to kinetic energy. We show in §7 that the full superparameterization closure, which includes terms responsible for absorption of the downscale cascade of potential energy, removes this bump in the potential energy spectrum.

The simulations all have similar heat flux, which is too large. In comparison with the reference simulation’s heat flux of 226, $S_5$ generates 292, $S_3$ generates 310, and $S_2^2$ generates 312. The fact that the heat flux is too large for all of these simulations may be related to the bump in the potential energy spectrum, and like that bump it is improved by the full superparameterization algorithm in §7.

The main results of this section are that it is crucial to smooth the $k^5$ backscatter
spectrum in a low-order code, and that good results can be obtained with several easily-implemented local averages. The degree of smoothing applied to the Reynolds stresses is inversely proportional to the value of \( E_0 \) required to achieve good results: more smoothing requires larger \( E_0 \).

6. Stochastic Backscatter: Temporal Correlation. In this section we incorporate temporal correlation into the backscatter by making the random angle \( \theta \) a Weiner process on the circle, as described in §2.2. We set the decorrelation time by comparison with the high resolution reference simulation by measuring the decorrelation time of the subgridscale Reynolds stress terms.

Specifically, we define a low-pass filter \( \langle \cdot \rangle \) (as in §3) by multiplying the Fourier coefficients of the high resolution reference simulation by \( 1 - (k/48)^{12} \) for \( k \leq 48 \), and by 0 for \( k > 48 \), where \( k = (k_x^2 + k_y^2)^{1/2} \). This has the effect of smoothly transitioning from large scales with \( k < 48 \) to small scales with \( k \geq 48 \), where 48 is the Nyquist wavenumber of the coarse model grid. The low-pass filtered fields correspond loosely to the solution of a finite-difference model on a 96 \( \times \) 96 grid, and they obey equation (2.6). The backscatter model is intended to approximate the subgridscale vorticity flux divergence, which has the form

\[
\nabla \cdot (\overline{u_j \zeta_j - u_j \overline{\zeta_j}}) = \\
(\partial_x^2 - \partial_y^2) (\overline{v_j} \overline{v_j} - \overline{\overline{v_j} v_j}) + \partial_{xy} \left[ \left( (v_j)^2 - (\overline{u_j})^2 \right) - (\overline{v_j}^2 - \overline{\overline{v_j}}^2) \right].
\]

(6.1)

Time series of the subgridscale Reynolds stress terms \( \overline{u_j v_j - u_j \overline{v_j}} \) and \( \overline{(v_j)^2 - (u_j)^2} - \overline{(u_j)^2} - \overline{(v_j)^2} \) were computed from the reference simulation along a line of 16 equispaced points at constant \( x \) in each layer. The lagged autocorrelation function was then computed at for each Reynolds stress term at each location. Recall that the lagged autocorrelation function for a zero-mean stationary stochastic process \( A(t) \) is defined as

\[
C(\tau) = \frac{\mathbb{E} [A(t)A(t + \tau)]}{\text{Var}[A(t)]}
\]

(6.2)

where \( \text{Var}[\cdot] \) denotes the variance. The expectation \( \mathbb{E} \) and the variance are computed using a time average. An approximate decorrelation time is computed from the lagged autocorrelation function.
autocorrelation function as

$$\tau_d = \int_0^{\infty} C(\tau) d\tau,$$

which is exact for exponential lagged autocorrelation functions $C(\tau) = \exp\{-\tau/\tau_d\}$.

Figure 6.1 shows the average of all the lagged autocorrelation functions for each Reynolds stress term and at each location compared to an approximate exponential lagged autocorrelation function with $\tau_d = 1/600$, which fits the data reasonably well. This decorrelation time implies a value of $\sigma^2 = 300$ for the temporally correlated model of $\theta$ (equation (2.18), and this value is used in simulations.

To test the effect of the addition of temporal correlation we run a simulation with the three-point average $S_3$, time step $\Delta t = 5 \times 10^{-5}$, and viscosity $\nu_4 = 2 \times 10^{-5}$, and with empirical decorrelation time corresponding to $\sigma^2 = 300$. Figure 6.2 compares this simulation to the analogous result with no temporal correlation. The kinetic and potential energy spectra are essentially indistinguishable, but there is a small difference in heat flux: the simulation with temporal correlation has a heat flux of 293, as compared to a heat flux of 310 without temporal correlation (the reference simulation has a heat flux of 226). The main difference is in the optimal backscatter amplitude: the simulation without temporal correlation uses $E_0 = 92$, whereas the simulation with temporal correlation uses $E_0 = 1625$. Part of the discrepancy is related to the fact that $E_0$ is scaled by the square root of the time step in the white noise scheme, but not in the scheme with temporal correlation.

The robustness of the schemes with and without temporal correlation has been tested by running simulations using time steps half as long and twice as long as those given above. In both cases the results are so similar that the energy spectra are indistinguishable in plots (not shown). Similarly, the heat flux varies by less than 1% as the time step is changed.

Overall the addition of temporal correlation seems to have minimal effect on the simulation results. The optimal backscatter amplitude $E_0 = 1625$ for the scheme with
temporal correlation is much too large in comparison with the values diagnosed from the reference simulation (median 57, mean 268). On the other hand the Reynolds stresses diagnosed from the reference simulation clearly have a nonzero decorrelation time. The addition of temporal correlation is clearly not as crucial to success as the addition of spatial correlation, although it remains unclear whether it may have some hidden benefit to the large-scale dynamics.

7. Stochastic Superparameterization. In this section we incorporate spatial and temporal smoothing into the full stochastic superparameterization scheme of [12, 13]. The primary goal is to demonstrate that the full stochastic superparameterization improves the potential energy spectrum and net heat flux as compared to the simplified backscatter scheme of the previous sections.

Stochastic superparameterization makes use of a formal multiscale ansatz to model the dynamics of the unresolved scales in turbulence; the subgrid dynamics are approximated by local linear stochastic models that respond to the local large-scale velocity and vorticity gradient [12,13,22]. It also makes use of a field of random angles, and the subgrid-scale Reynolds stress terms have the form

\begin{align}
    \overline{u'_j v'_j} &= \frac{A}{2} F_{u,j}(\overline{\mu_1}, \overline{\mu_2}, \nabla_h q_1, \nabla_h q_2, \theta; \epsilon) \sin(2\theta), \\
    \overline{(v'_j)^2} - \overline{(u'_j)^2} &= AF_{u,j}(\overline{\mu_1}, \overline{\mu_2}, \nabla_h q_1, \nabla_h q_2, \theta; \epsilon) \cos(2\theta). 
\end{align}

In the simplified backscatter model the functions $F_{u,j}$ are identical constants (compare to equation (2.10)). Note that the Reynolds stresses in stochastic superparameterization are not identical in the upper and lower layers, like they are in the simplified model. The simplified backscatter model ignores part of the potential vorticity flux divergence $k_{\theta}^2 \psi_i'(\psi_i' - \psi_j')$ where $i \neq j$. In stochastic superparameterization these...
components of the potential vorticity flux take the form

\[
\begin{align*}
\overline{u}_1'(\psi'_1 - \psi'_2) &= AF_b(\pi_1, \pi_2, \nabla_h \pi_1, \nabla_h \pi_2, \theta; \epsilon) \cos(\theta), \\
\overline{v}_1'(\psi'_1 - \psi'_2) &= AF_b(\pi_1, \pi_2, \nabla_h \pi_1, \nabla_h \pi_2, \theta; \epsilon) \sin(\theta), \\
\overline{u}_2'(\psi'_2 - \psi'_1) &= -\overline{u}_1'(\psi'_1 - \psi'_2), \quad \overline{v}_2'(\psi'_2 - \psi'_1) = -\overline{v}_1'(\psi'_1 - \psi'_2).
\end{align*}
\]  

(7.2)

In stochastic superparameterization the functions \(F_{u,j}\) and \(F_b\) are computed from solutions of the local stochastic models of the subgrid-scale dynamics; the stochastic subgrid dynamics evolve for a time length \(\epsilon^{-1}\) while the large scales are held fixed, and \(\epsilon\) is considered a tunable parameter. The stochastic subgrid scale dynamics are assumed to evolve on length scales corresponding to wavenumbers with magnitude \(k = \sqrt{k_x^2 + k_y^2} \geq 48\), which is the Nyquist wavenumber of the coarse grid. The coefficient \(A\) is a tunable parameter analogous to \(E_0\).

We apply the stochastic superparameterization algorithm exactly as described in [13] with the following modifications: (i) the Reynolds stress terms and the potential vorticity flux are smoothed using the three-point average \(S_3\), and their derivatives are approximated with centered second-order finite differences, and (ii) the random angles \(\theta\) evolve according to equation (2.18) with the empirical value \(\sigma^2 = 300\). Figure 7.1 compares the results of the full stochastic superparameterization to the simplified backscatter scheme; both simulations use the empirical decorrelation time corresponding to \(\sigma^2 = 300\), and the three-point average \(S_3\). The stochastic superparameterization scheme uses \(A = 17500\), and \(\epsilon = 800\). The kinetic energy spectra of the simplified scheme and the superparameterization are very similar, differing only near the coarse grid scale. The potential energy spectrum of the superparameterization scheme does not have the bump at small scales, and is a good fit to the potential energy spectrum of the reference simulation. The superparameterization has a net heat flux of 280, which is smaller than the heat flux of 293 from the simplified scheme, but still bigger than the value of 226 from the reference simulation. Overall, the full stochastic superparameterization scheme with spatial and temporal correlations improves over the simplified backscatter scheme.

8. Conclusions. Kinetic energy backscatter is the process of kinetic energy transfer from small to large scales in turbulence. We have developed stochastic schemes to model backscatter in the inverse kinetic energy cascade of quasigeostrophic turbulence, where backscatter drives a net transfer from small to large scales. The main proposed application of these schemes is in eddy-permitting ocean modeling, where low-resolution models do not resolve the energy transfer from subgrid scales. In addition to improving the accuracy of low-resolution models, stochastic backscatter is expected to improve ensemble filtering and prediction schemes [3], and our development in the context of superparameterization connects naturally with multiscale filtering methods [11,18,23].

In §2.1 we showed that a simplified stochastic backscatter scheme from [13], based on stochastic superparameterization, corresponds to an isotropic random forcing with spectrum proportional to \(k^3\) where \(k\) is the wavenumber magnitude, i.e. it forces very strongly near the coarse grid scale. This backscatter spectrum is analogous to the backscatter spectrum for smaller-scale three-dimensional isotropic turbulence [17,21,30]. In §5 we showed that the scheme of [13] performs poorly in a low resolution model using second-order finite differences to approximate the nonlinear advection terms, whereas it was shown in [13] that the scheme performed well in a low resolution
model with spectral numerics. We argue that the poor performance is due to the fact that the energy backscattered near the coarse grid scale is supposed to be transferred to larger scales by the nonlinear dynamics, but the low-order model is not able to represent this process accurately.

The use of a low order numerical method requires the numerical solution to be relatively smooth, even in a low resolution model, and we showed in §5 that very good results can be obtained by smoothing the forcing with a local average. This has the effect of damping the backscatter forcing spectrum to zero at the coarse grid scale, so that the backscatter injects energy at a scale where the nonlinear dynamics are better resolved by the low-order numerics.

The spatially-smoothed backscatter scheme of §5 corresponds to a forcing that is delta-correlated (white) in time. Since the backscatter is supposed to model the nonlinear effects of unresolved deterministic dynamics, it is more realistic for the backscatter model to have a nonzero decorrelation time. We showed in §2.2 and in the Appendix how to add temporal correlation to the backscatter scheme in such a way that the backscatter has an exponential lagged autocorrelation function. The appropriate decorrelation timescale was diagnosed from a high resolution reference simulation in §6, and this was incorporated into the backscatter scheme. Incorporation of the nonzero decorrelation time had a minimal impact on the results, although it remains possible that there is some hidden benefit. For example, the short-term prediction and filtering performance of the model might be improved through the use of nonzero temporal decorrelation time.

The stochastic backscatter scheme with spatial and temporal correlation gives good results for the time-averaged kinetic energy spectrum, but the results invariably include a bump in the potential energy spectrum at small scales. This bump is related to the fact that the stochastic kinetic energy backscatter scheme does not include a mechanism for absorbing the downscale cascade of potential energy. In §7 we showed that incorporating spatial smoothing and temporal correlation into the full stochastic superparameterization algorithm of [12,13] gives excellent results overall, matching both the kinetic and potential energy spectra quite accurately.

Appendix. In this section we prove the following theorem

**Theorem A.1.** The Reynolds stress terms modeled as \( \sin(2\theta) \) and \( \cos(2\theta) \) using

\[
d\theta = \sigma \, dW
\]  

(A.1)

have stationary lagged autocorrelation functions of the form \( \exp\{-2\sigma^2 \tau\} \).

**Proof.** The Fokker-Planck equation for (A.1) is simply

\[
\partial_t p = \frac{\sigma^2}{2} \partial^2_\theta p.
\]  

(A.2)

On a periodic domain of size \( 2\pi \) the invariant measure is uniform \( \mathbf{p} = (2\pi)^{-1} \). The stationary lagged autocorrelation function for the sine term is defined as

\[
C(\tau) = 2E[\sin(2\theta_t) \sin(2\theta_{t+\tau})] = 2E[\sin(2\theta_t) \sin(2(\theta_t + \Delta_\tau))]
\]  

(A.3)

where \( \tau \) is the lag, \( \theta_t \) denotes \( \theta \) at time \( t \), and \( \Delta_\tau = \theta_{t+\tau} - \theta_t \). Dependence on \( t \) drops out in the stationary state, i.e. the long time limit.

The proof relies on an explicit expression for the probability density of \( \Delta_\tau \), which is a Dirac delta distribution at zero lag \( \tau = 0 \). This can be obtained from the Fourier
expansion of the Dirac delta distribution and its explicit evolution under the Fokker-Planck equation (A.1). Specifically, the probability distribution for $\Delta \tau$ is

$$p(\Delta \tau) = \frac{1}{2\pi} + \frac{1}{\pi} \sum_{n=1}^{\infty} e^{-n^2 \sigma^2 \tau / 2} \cos(n \Delta \tau). \quad (A.4)$$

The proof that the lagged autocorrelation function (A.3) is exponential proceeds as follows. First, simple trigonometric identities imply

$$C(\tau) = 2 \mathbb{E} \left[ \sin(2\theta_t) \left( \sin(2\theta_t) \cos(2\Delta \tau) + \cos(2\theta_t) \sin(2\Delta \tau) \right) \right]. \quad (A.5)$$

Second, the non-anticipating property of solutions to (A.1) implies that $\theta_t$ is independent of $\Delta \tau$ for $\tau > 0$; the above expression therefore simplifies to

$$C(\tau) = 2 \left( \mathbb{E} \left[ \sin^2(2\theta_t) \right] \mathbb{E} \left[ \cos(2\Delta \tau) \right] + \mathbb{E} \left[ \sin(2\theta_t) \cos(2\theta_t) \right] \mathbb{E} \left[ \sin(2\Delta \tau) \right] \right). \quad (A.6)$$

Since we are interested in the stationary lagged autocorrelation function the distribution of $\theta_t$ is considered uniform, which implies that the second term above drops out and the expected value of $\sin^2(2\theta_t)$ is $1/2$, leaving

$$C(\tau) = \mathbb{E} \left[ \cos(2\Delta \tau) \right]. \quad (A.7)$$

This expectation can be written as an integral against the known probability distribution of $\Delta \tau$, which gives

$$C(\tau) = \int_{-\pi}^{\pi} \left( \frac{1}{2\pi} + \frac{1}{\pi} \sum_{n=1}^{\infty} e^{-n^2 \sigma^2 \tau / 2} \cos(n \Delta \tau) \right) \cos(2\Delta \tau) d\Delta \tau = e^{-2\sigma^2 \tau}. \quad (A.8)$$

This proves that the stationary lagged autocorrelation function for the sine term is exponential with decorrelation time scale $(2\sigma^2)^{-1}$; the proof for the cosine term is the same. ☐

REFERENCES

20
Backscatter in Quasigeostrophic Turbulence


