Research in the Mathematical Sciences manuscript No.
(will be inserted by the editor)

Parameter estimation in the stochastic superparameterization of two-layer quasigeostrophic flows

Estimation of subgrid-scale modeling parameters in the stochastic superparameterization of two-layer quasigeostrophic turbulence

Yoonsang Lee

Received: date / Accepted: date

Abstract Geophysical turbulence has a wide range of spatiotemporal scales that requires a multiscale prediction model for efficient and fast simulations. Stochastic parameterization is a class of multiscale methods that approximates the large-scale behaviors of the turbulent system without relying on scale separation. In the stochastic parameterization of unresolved subgrid-scale dynamics, there are several modeling parameters to be determined by tuning or fitting to data. We propose a strategy to estimate the modeling parameters in the stochastic parameterization of geostrophic turbulent systems. The main idea of the proposed approach is to generate data in a spatiotemporally local domain and use physical/statistical information to estimate the modeling parameters. In particular, we focus on the estimation of modeling parameters in the stochastic superparameterization, a variant of the stochastic parameterization framework, for an idealized model of synoptic scale turbulence in the atmosphere and oceans. The test regimes considered in this study include strong and moderate turbulence with complicated patterns of waves, jets, and vortices.

Keywords stochastic parameterization · multiscale · parameter estimation · quasigeostrophic turbulence

Mathematics Subject Classification (2010) 65C20 · 76F55 · 86-08

The author is supported by NSF DMS-1912999 and the Burke award at Dartmouth College.

Mathematics Department
Dartmouth College
27 N Main St, Hanover NH, 03755, USA
Tel.: +1 (603) 646-3178
E-mail: yoonsang.lee@dartmouth.edu

1 Introduction

Geophysical fluid systems involve a tremendously wide range of spatiotemporal scales. In the atmosphere and oceans, the spatial scale varies from millimeters to tens of thousands of kilometers, while the time scale varies from seconds to hundreds of years [34,37]. In the numerical simulation of such systems, it is essential to develop a multiscale prediction model that is tractable by the current (or the near future) generation computing powers. A challenge in multiscale modeling is that the unique and relevant properties of such systems are characterized by a complex interplay of different scale dynamics. Quasigeostrophic turbulence includes regimes where there is a kinetic energy transfer from small to large scales, an inverse cascade of energy [5,7], and the small (or unresolved subgrid) scale affects and is affected by the large (or resolved coarse) scale. Therefore, it is crucial to represent the effect of unresolved small scale, either analytically or numerically, to close the large-scale dynamics, which is called subgrid-scale parameterization in geophysical fluid systems.

There are several classes of subgrid-scale parameterization strategies for geostrophic turbulence. A deterministic approach uses a physical model for the effect of the unresolved small scale, for example, the Smagorinsky model in LES. In [3,30], nonlinear deterministic models have been used in a LES modeling of three-dimensional isotropic turbulence. The test problems considered in these models correspond to regimes where the net transfer of kinetic energy is downscale with a negligible inverse cascade of energy. The cloud-resolving method for tropical atmospheric convection (or Superparameterization) by Grabowski [12] and its extensions [19,6] belong to another class of subgrid-scale parameterization method. In a numerical computation point of view, this strategy is closely related to the Heterogeneous Multiscale Method (HMM) framework [1,10]. The main idea of HMM is to run subgrid-scale simulations in local domains to directly estimate the effect of the unresolved small-scale dynamics on the large-scale dynamics. A typical local domain size of these methods is comparable to the coarse grid mesh, and thus these methods depend strongly on scale separation.

The stochastic parameterization is a multiscale framework that can handle a continuous range of scales without scale separation. It uses a stochastic process for the subgrid-scale effect to mimic random and chaotic motions of the unresolved small-scale turbulence. The stochastic approach using random forcing terms has been successfully applied to atmospheric and oceanic turbulent flows [4,18,31]. It has also improved the deterministic LES by incorporating stochastic backscatters that mimic the inverse cascade of energy that is missed by the deterministic model [26,33]. In the subgrid-scale parameterization of the tropical atmosphere, a stochastic modeling that uses additional damping and a random forcing showed successful prediction skills in capturing the synoptic-scale variability of the Madden-Julian oscillation (MJO) [36].

In the stochastic parameterization of the unresolved scale, it is indispensable to introduce several modeling parameters, such as damping coefficients, random forcing terms, etc. These parameters are estimated by tuning or fit-

ting to observation data or statistical information. In many applications, observation or measurement data are limited or only partially available, which makes it difficult to estimate the correct modeling parameters. Even if the necessary information is available, tuning parameters, or optimization over high-dimensional parameter space, is a challenging computational task.

The goal of this paper is to propose an estimation method for the modeling parameters using a minimal requirement of data. Particularly, we focus on the stochastic superparameterization (SSP) [16,17,27,14], a variant of the stochastic subgrid-scale parameterization method for quasigeostrophic turbulence. In the stochastic superparameterization of quasigeostrophic turbulence, there are six parameters to be estimated. These parameters are determined by the information of the true system, such as eddy turnover time, linear instability rate, and are tuned to match energy spectra. The proposed method estimates five parameters out of them by generating subgrid-scale data in a local domain. By assuming that energy spectra are available, the remaining parameter, an enhanced viscosity, is tuned to match the energy spectra. We note that the tuning process of the proposed method is more tractable than the standard SSP as the number of the tuning parameters is low.

This paper has the following structures. We first describe a turbulence model in Section 2, two-layer quasigeostrophic turbulence, as an idealized model of synoptic-scale turbulence in the atmosphere and oceans. As stringent test cases for the proposed method, we consider two test regimes, strong and moderate turbulence. The first regime is isotropic and homogeneous with a strong inverse cascade of energy, breaking wave patterns, etc. The latter regime is anisotropic and inhomogeneous with zonal direction strong jets. In Section 3, we review the stochastic superparameterization for the model problem and discuss modeling parameters to estimate. Section 4 explains the proposed method to determine the modeling parameters in the stochastic superparameterization. We also investigate how the local domain size affects the estimation process and accuracy. Numerical results of the stochastic superparameterization using the modeling parameters from the proposed method can be found in Section 5. The paper concludes with discussions in Section 6.

2 Model problem: two-layer quasigeostrophic turbulence

As a turbulence model for subgrid-scale parameterization, we consider a two-layer quasigeostrophic equation in a two-dimensional space,

$$\begin{aligned}
\partial_t q_1 &= -\mathbf{u}_1 \cdot \nabla(q_1) - \partial_x q_1 - (k_\beta^2 + k_d^2) v_1 - \nu \nabla^4 q_1, \\
\partial_t q_2 &= -\mathbf{u}_2 \cdot \nabla(q_2) + \partial_x q_2 - (k_\beta^2 - k_d^2) v_2 - r \nabla^2 \psi_2 - \nu \nabla^4 q_2, \\
q_1 &= \nabla^2 \psi_1 + \frac{k_d^2}{2}(\psi_2 - \psi_1), \\
q_2 &= \nabla^2 \psi_2 - \frac{k_d^2}{2}(\psi_2 - \psi_1).
\end{aligned} \tag{1}$$

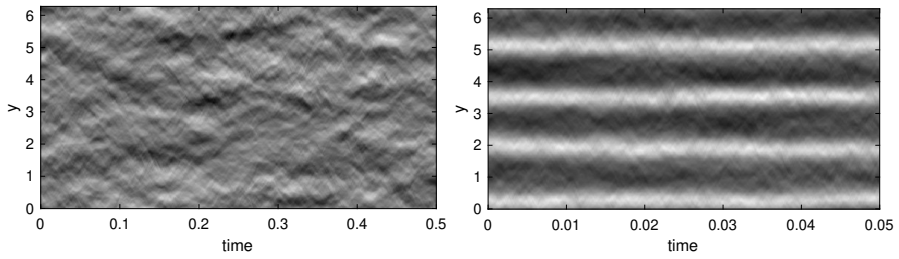
Regime	k_d	k_β	r	ν
Strong turbulence	50	0	16	1×10^{-7}
Moderate turbulence	50	25	4	1×10^{-7}

Table 1: Parameters of the model equation (1) for each test regime. k_d: deformation wavenumber, k_β nondimensional coefficient related to the Coriolis force, r: Ekman drag coefficient, ν: biharmonic viscosity.

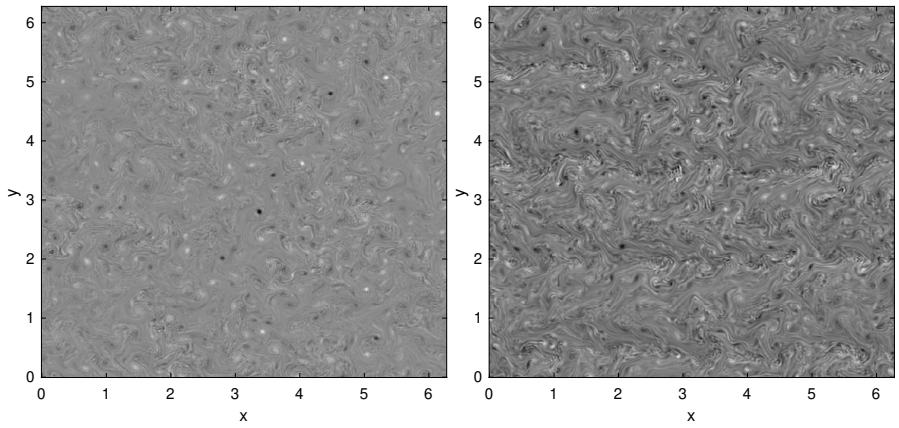
Here ψ_j is the stream function and q_j is the potential vorticity in the upper ($j = 1$) and lower ($j = 2$) layers. The velocity field $\mathbf{u}_j$ follows the velocity-stream function relation $\mathbf{u}_j = (u_j, v_j) = (-\partial_y\psi, \partial_x\psi)$. k_d is a deformation wavenumber, r is the linear Ekman drag coefficient at the bottom layer, and ν is the biharmonic viscosity that is commonly used in eddy permitting ocean models [13]. We use the biharmonic viscosity to dissipate small scale while maintaining the large-scale dynamics of the model. The term related to a nondimensional k_β results from the variation of the vertical projection of Coriolis frequency with latitude. Also the model has a large-scale zonal vertical shear that results in the term$(-1)^j(\partial_x q_j + k_d^2 v_j)$. In [37], it is shown that when $k_\beta < k_d$, the imposed vertical shear leads to a baroclinic instability that maintains the model dynamics without an additional explicit force term. The quasigeostrophic equation is often expressed equivalently in terms of barotropic $q_{bt} = \frac{q_1+q_2}{2}$ and baroclinic $q_{bc} = \frac{q_1-q_2}{2}$ variables, which correspond to the vertical average and the vertical difference.

This model is a classical idealized model of synoptic-scale turbulent flows in the atmosphere and oceans [37] with features such as complicated patterns of waves, jets, and vortices. To generate the features, we consider two test regimes, 1) strong and 2) moderate turbulent flows. These two regimes are different by varying the Coriolis effect k_β and the Ekman drag r while the deformation wavenumber k_d and the viscosity ν are fixed at 50 and 1×10^{-7} respectively for all regimes. For the strong turbulent regime, we set $k_\beta = 0$ and $r = 16$. For the moderate turbulent regime, there is a Coriolis effect with $k_\beta = 25$ but with a weaker drag $r = 4$ (see Table 1 for each test regime parameters). A similar model and test regimes have been used for testing stochastic parameterization in [17,14,23] but with a hyperviscosity $-\nabla^8 q_j$ instead of the biharmonic viscosity used in the current study.

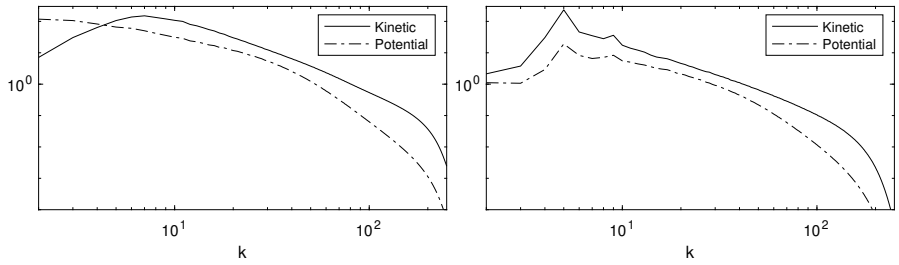
Our reference simulation solves the model equation in a domain $\Omega = (0, 2\pi)^2$ with a doubly-periodic boundary condition. For a spatial discretization, we use the Fourier spectral method for the linear terms while the nonlinear term is discretized by the finite difference method by Arakawa [2] that preserves energy and enstrophy. Regarding time integration, we use a fourth-order Runge-Kutta time integrator. The spatial resolution is 512×512 for each layer, which is comparable to the finest resolution of the baroclinic turbulence simulation in [35]. We fix the time step at 2×10^{-5} for all simulations. Each simulation is initialized with a small energy randomly generated potential vorticity field and runs until it reaches a quasi-stationary state, i.e., the total energy is fully developed.



(a) Time series of zonal average of barotropic velocity component u_{bt}



(b) Snapshots of barotropic potential vorticity



(c) Energy spectra

Fig. 1: Direct numerical reference simulations of the model (1). Strong (left) and moderate (right) turbulent regimes.

Figure 1(a) and (b) show time series of the zonal (x-direction) average of barotropic velocity component $u_{bt} = \frac{u_1+u_2}{2}$ and snapshots of barotropic potential vorticity q_{bt} (left column: strong regime, right column: moderate regime). The strong regime shows homogeneous isotropic turbulence dominated by small-scale coherent vortices. The moderate regime, on the other hand, shows anisotropic and inhomogeneous structures with four jets. The jets are asymmetric with stronger zonal velocity. The inverse energy cascades

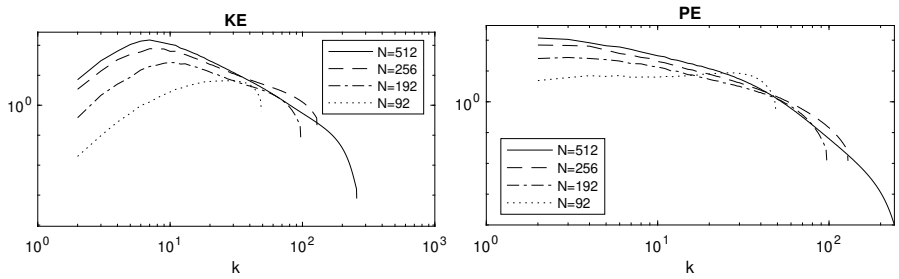
from small scale to large scale and maintains the large-scale jets (see [9] for a study of large-scale jet structure of quasigeostrophic turbulence). As the formation of the jets are related to the small scale, it is a challenging task to capture jets using multiscale modeling that does not resolve all active subgrid-scale components. Figure 1(c) shows energy spectra of the kinetic energy (KE) $\sum_j \frac{1}{2}\iint_\Omega |\mathbf{u}_j|^2 dxdy$ and the potential energy (PE) $\frac{k_d^2}{2}\iint_\Omega (\psi_1-\psi_2)^2 dxdy$ as functions of the wavenumber $k=|\mathbf{k}|$ (here $\mathbf{k}$ is the wave vector), whose sum (i.e., the total energy) is conserved by the model for the unforced and inviscid case. The model has both downscale and upscale energy cascade. The interaction between the y-direction velocity and the imposed baroclinic potential vorticity gradient generates potential vorticity and energy at the large scale determined by the bottom friction and k_β. The potential energy cascades downscale to the small scale determined by the deformation radius, and is converted to barotropic kinetic energy [32]. The kinetic energy, on the other hand, cascades upscale and becomes barotropic at the large scale, which is further dissipated by bottom friction. In addition to energy spectra, the model has heat flux that is related to the background temperature gradient with the imposed mean shear. The heat flux $\iint (v_1+v_2)(\psi_1-\psi_2)/4$ is weaker in the moderate regime than in the strong regime as the zonal jets obstruct the heat transport in the y-(meridional) direction. From the reference simulation, the time averaged heat fluxes are 223 and 20 for the strong and the moderate regimes respectively.

A numerical test shows that the model requires a spatial resolution higher than 256×256 for each layer. Various spatial resolutions N^2 for $N=256, 192$ and 96, are tested with a tuned biharmonic viscosity to match the energy spectra. The test result (Figure 2) shows that none of the resolutions can achieve the correct spectra even with the tuned viscosity. In Section 3 and 4, we show that the stochastic superparameterization with a coarse resolution 96×96 approximates the energy spectra and heat flux well, including the jet structure of the moderate turbulent regime.

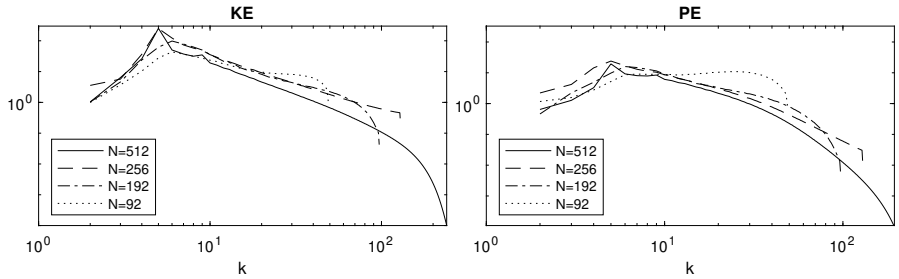
3 Stochastic parameterization for subgrid-scale variables

In this section, we review the stochastic parameterization method, particularly the stochastic superparameterization, and discuss its characteristics and limitations. In the subgrid-scale parameterization, it is typically assumed that a full resolution variable has a natural decomposition into large-scale variables $\overline{(\cdot)}$ and subgrid-scale variables $(\cdot)'$. An application of the large-scale and subgrid-scale projections to the model equation yields the equations for the large-scale and subgrid-scale variables

$$\begin{aligned}
\partial_t \overline{q}_1 &= -\overline{\mathbf{u}}_1\cdot\nabla(\overline{q}_1) - \overline{\mathbf{u}_1'\cdot\nabla q_1'} - \partial_x \overline{q}_1 - (k_\beta^2+k_d^2)\overline{v}_1 - \nu\nabla^4\overline{q}_1,\\
\partial_t \overline{q}_2 &= -\overline{\mathbf{u}}_2\cdot\nabla(\overline{q}_2) - \overline{\mathbf{u}_2'\cdot\nabla q_2'} + \partial_x \overline{q}_2 - (k_\beta^2-k_d^2)\overline{v}_2 - \nu\nabla^4\overline{q}_2 - r\nabla^2\overline{\psi}_2,
\end{aligned} \tag{2}$$



(a) Strong turbulent regime



(b) Moderate turbulent regime

Fig. 2: Kinetic energy (KE; left) and potential energy (PE; right) spectra from various spatial resolutions. Viscosity is tuned to match the spectra.

and

$$\begin{aligned}\partial_t q_1' =& -\mathbf{u}_1'\cdot\nabla q_1' - \overline{\mathbf{u}}_1\cdot\nabla q_1' - \mathbf{u}_1'\cdot\nabla\overline{q}_1 - \partial_x q_1' - (k_\beta^2+k_d^2)v_1' - \nu\nabla^4 q_1',\\ \partial_t q_2' =& -\mathbf{u}_2'\cdot\nabla q_2' - \overline{\mathbf{u}}_2\cdot\nabla q_2' - \mathbf{u}_2'\cdot\nabla\overline{q}_2 + \partial_x q_2' - (k_\beta^2-k_d^2)v_2' - \nu\nabla^4 q_2'\\ & - r\nabla^2\psi_2',\end{aligned} \tag{3}$$

respectively. The main issue of the large-scale scale equation, which is called 'closure problem in multiscale modeling [29], is that the subgrid-scale effect on the large-scale variables, $\overline{\mathbf{u}_j'\cdot\nabla q_j'}$, requires the subgrid-scale variables $\mathbf{u}_j'$ and $\nabla q_j'$ that are not available unless the subgrid-scale equation is solved. In many applications, this term must be accurately approximated as it has a nontrivial effect on the coarse-scale dynamics. The inverse energy cascade of geostrophic turbulent flows is an example that the small-scale dynamics affect the large-scale dynamics. Among different methods to account for the subgrid-scale effect, we focus on the stochastic parameterization method.

The standard stochastic parameterization [18,36] models the subgrid-scale effect as a stochastic system typically using the following form

$$\overline{\mathbf{u}_j'\cdot\nabla q_j'} \approx -d_j\overline{q}_j + F_j \tag{4}$$

where $d > 0$ is a damping coefficient, $\overline{q}$ is a large-scale variable, and F is a random force. The coefficients d and F are tuning parameters determined

by fitting the system to observation/climatological data. This approach has been successfully applied to many interesting problems in geophysical turbulent problems including the synoptic-scale variability of the Madden-Julian oscillation [36].

3.1 Stochastic superparameterization

The stochastic superparameterization (SSP) [16,17] is a variant of the stochastic parameterization approach for subgrid-scale parameterization. Instead of the stochastic modeling directly for the effect of the subgrid-scale on the large-scale dynamics, the stochastic superparameterization uses stochastic modeling for the subgrid-scale dynamics (3). By replacing the nonlinear term in (3) (not in (2)) with additional damping and random forcing, the subgrid-scale equation becomes a linear system conditioned to the large-scale variables. The modified subgrid-scale equation is then solved to estimate the effect of the subgrid-scale variables on the large-scale equation. That is, the stochastic superparameterization calculates the subgrid-scale effect by calculating the subgrid-scale dynamics using a cheap stochastic model. The standard stochastic parameterization, on the other hand, mimics the subgrid-scale effect directly using a stochastic model. The idea of the stochastic superparameterization is similar to the heterogeneous multiscale method [1,10] in that both methods generate subgrid-scale information to estimate the subgrid-scale effect on the large-scale dynamics. The main difference is that the stochastic superparameterization uses stochastic modeling for the subgrid-scale dynamics while the heterogeneous multiscale method uses the original subgrid-scale equation under scale separation. The model in the stochastic superparameterization is cheap to solve and also enables to cover a broader range of scales without scale separation.

For completeness, we review some technical details of the stochastic superparameterization [17,14]. SSP replaces the nonlinear term in the subgrid-scale equation (3) by an additional damping $-\Gamma q_j'$ and a random force F_j'

$$\begin{aligned}\partial_t q_1' \approx &-\Gamma q_1' + F_1' - \overline{\mathbf{u}}_1 \cdot \nabla q_1' - \partial_x q_1' - \mathbf{u}_1' \cdot \nabla \bar{q}_1 - (k_\beta^2 + k_d^2) v_1' - \nu \nabla^4 q_1', \\ \partial_t q_2' \approx &-\Gamma q_2' + F_2' - \overline{\mathbf{u}}_2 \cdot \nabla q_2' + \partial_x q_2' - \mathbf{u}_2' \cdot \nabla \bar{q}_2 - (k_\beta^2 - k_d^2) v_2' - \nu \nabla^4 q_1' \\ &- r \nabla^2 \psi_2'.\end{aligned} \tag{5}$$

Γ is a positive definite pseudo-differential operator, and the random force F_j' are correlated in space and white in time. One key idea used in SSP is a point approximation for multiscale modeling. This approximation imposes an artificial dynamical scale separation where the subgrid-scale equations are understood in an embedded domain. For each set of the large-scale variables $\overline{\mathbf{u}_j}$ and $\bar{q}_j$ defined on the coordinates (x, y, t), SSP introduces new coordinates $(\tilde{x}, \tilde{y}, \tau; x, y, t)$ for the subgrid-scale variables. All variations of the subgrid-scale variables are interpreted on the new coordinates $(\tilde{x}, \tilde{y}, \tau; x, y, t)$, which

yields

$$\begin{aligned}\partial_\tau q_1' \approx &- \Gamma q_1' + F_1' - \overline{\mathbf{u}}_1 \cdot \nabla_{(\tilde{x},\tilde{y})} q_1' - \partial_{\tilde{x}} q_1' - \mathbf{u}_1' \cdot \nabla_{(\tilde{x},\tilde{y})} \overline{q}_1 - (k_\beta^2 + k_d^2) v_1 \\ &- \nu \nabla^4_{(\tilde{x},\tilde{y})} q_1', \\ \partial_\tau q_2' \approx &- \Gamma q_2' + F_2' - \overline{\mathbf{u}}_2 \cdot \nabla_{(\tilde{x},\tilde{y})} q_2' + \partial_{\tilde{x}} q_2' - \mathbf{u}_2' \cdot \nabla_{(\tilde{x},\tilde{y})} \overline{q}_2 - (k_\beta^2 - k_d^2) v_2 \\ &- \nu \nabla^4_{(\tilde{x},\tilde{y})} q_1' - r \nabla^2_{(\tilde{x},\tilde{y})} \psi_2'\end{aligned} \tag{6}$$

where $\nabla_{\tilde{x},\tilde{y}} = (\partial_{\tilde{x}}, \partial_{\tilde{y}})$. In the point approximation, the large-scale projection is understood as an average over the new coordinates. Also note that in equation (6), the large-scale variables are understood as constants while in equation (5), the large-scale variables vary over the domain.

For fixed large-scale variables, the Fourier transform of equation (6) (by suppressing the notation $(\cdot)'$) is

$$\begin{aligned}\frac{d\hat{q}_1}{d\tau} &= -i(\overline{\mathbf{U}}_1 \cdot \mathbf{k})\hat{q}_1 - (i\mathbf{k} \times \nabla \overline{Q}_1)\hat{\psi}_1 + A_{1,\mathbf{k}} \dot{W}_{1,\mathbf{k}} - (\gamma_{\mathbf{k}} + \nu k^4)\hat{q}_1 \\ \frac{d\hat{q}_2}{d\tau} &= -i(\overline{\mathbf{U}}_2 \cdot \mathbf{k})\hat{q}_2 - (i\mathbf{k} \times \nabla \overline{Q}_2)\hat{\psi}_2 + A_{1,\mathbf{k}} \dot{W}_{1,\mathbf{k}} - (\gamma_{\mathbf{k}} + \nu k^4)\hat{q}_2 + r k^2 \hat{\psi}_2,\end{aligned} \tag{7}$$

where $\mathbf{k} = (k_x, k_y)$ is the wave vector, $k = |\mathbf{k}|$, $W_{j,\mathbf{k}}$ are independent complex Wiener processes, and $A_{j,\mathbf{k}}$ are complex constants. Here we introduced new large-scale variables $\overline{\mathbf{U}}_j = \overline{\mathbf{u}}_j - (-1)^j \hat{\mathbf{x}}$ and $\overline{Q}_j = \overline{q}_j + (k_\beta^2 - (-1)^j k_d^2) y$. The damping operator Γ is defined to satisfy

$$\widehat{\Gamma q_j'} = \gamma_{\mathbf{k},j} \hat{q}_j \tag{8}$$

where $\gamma_{\mathbf{k},j}$ is a positive real number for each $\mathbf{k}$. Through the relation between q_j' and ψ_j', we can represent the above equation as an equation for $\hat{\psi}_j$

$$d \begin{pmatrix} \hat{\psi}_1 \\ \hat{\psi}_2 \end{pmatrix} = \mathbf{L}_{\mathbf{k}} \begin{pmatrix} \hat{\psi}_1 \\ \hat{\psi}_2 \end{pmatrix} d\tau + \sigma_{\mathbf{k}} d\mathbf{W}_{\mathbf{k}} \tag{9}$$

where $\sigma_{\mathbf{k}} = \begin{pmatrix} A_{1,\mathbf{k}} & 0 \\ 0 & A_{2,\mathbf{k}} \end{pmatrix}$, $\mathbf{W}_{\mathbf{k}} = \begin{pmatrix} W_{1,\mathbf{k}} \\ W_{2,\mathbf{k}} \end{pmatrix}$, and the linear operator $\mathbf{L}_{\mathbf{k}}$ (that depends on the large-scale variables)

$$\begin{aligned}\mathbf{L}_{\mathbf{k}} = &- (\gamma_k + \nu k^4)\mathbf{I} + -i\mathbf{Q}_k^{-1} \begin{pmatrix} \overline{\mathbf{U}}_1 \cdot \mathbf{k} & 0 \\ 0 & \overline{\mathbf{U}}_2 \cdot \mathbf{k} \end{pmatrix} \mathbf{Q}_k \\ &+ \mathbf{Q}_k^{-1} \begin{pmatrix} -i\mathbf{k} \times \nabla \overline{Q}_1 & 0 \\ 0 & r k^2 - i\mathbf{k} \times \nabla \overline{Q}_2 \end{pmatrix},\end{aligned} \tag{10}$$

for the 2×2 identity matrix $\mathbf{I}$ and $\mathbf{Q}_k = \begin{pmatrix} -(k_d^2/2 + k^2) & k_d^2/2 \\ k_d^2/2 & -(k_d^2/2 + k^2) \end{pmatrix}$. In [17], it is shown that the subgrid-scale effect $\overline{\mathbf{u}_1' \cdot \nabla q_1'}$ has the following representation

$$\overline{\mathbf{u}_j' \cdot \nabla q_j'} = \frac{(-1)^j k_d^2}{2} \nabla \cdot \overline{\mathbf{u}_j'(\psi_1' - \psi_2')} + (\partial_x^2 - \partial_y^2)\overline{u_j' v_j'} + \partial_{xy}\left(\overline{(v_j')^2} - \overline{(u_j')^2}\right). \tag{11}$$

As we interpret the large-scale projection as the average over the small-scale coordinates, the subgrid-scale effect is derived from the average of $|\hat{\psi}_i\hat{\psi}_j|$ (see equation (18) - (23) for the representations of each term in (11) in terms of $\hat{\psi}_i\hat{\psi}_j^*$). Thus, it is sufficient to solve for the covariance matrix $C_{\mathbf{k}} = \begin{pmatrix} |\hat{\psi}_1|^2 & \hat{\psi}_1\hat{\psi}_2^* \\ \hat{\psi}_1^*\hat{\psi}_2 & |\hat{\psi}_2|^2 \end{pmatrix}$, whose evolution is governed by

$$\frac{d}{d\tau}C_{\mathbf{k}} = \mathbf{L}_{\mathbf{k}}C_{\mathbf{k}} + C_{\mathbf{k}}\mathbf{L}_{\mathbf{k}}^* + \sigma_{\mathbf{k}}\sigma_{\mathbf{k}}^* \tag{12}$$

using the Ito formula. We note that this system has an interaction with the large-scale variables through the dependence of $\mathbf{L}_{\mathbf{k}}$ on $\overline{\mathbf{U}}_j$ and $\overline{Q}_j$.

In solving for $C_{\mathbf{k}}$, there are several parameters left to be determined. First, the damping coefficient $\gamma_{\mathbf{k},j}$ is assumed to be equal for both layers and is modeled as isotropic and proportional to the nonlinear subgrid timescale for each $k = |\mathbf{k}|$

$$\gamma_{\mathbf{k},j} = \gamma_{\mathbf{k}} = \gamma_k = \begin{cases} \gamma_0(k/k_d)^{2/3} & \text{for } k \le k_d, \\ \gamma_0 & \text{for } k \ge k_d \end{cases}. \tag{13}$$

In SSP, the base damping coefficient γ_0 is chosen so that it is slightly more than sufficient to damp the linear instability of the imposed shear. To find the noise level $\sigma_{\mathbf{k}}$, SSP specifies the equilibrium state of the covariance matrix. Using the ideas of fully developed quasigeostrophic turbulence, in particular, the subgrid-scale dynamics are isotropic (see [17] for details), SSP models the equilibrium covariance as

$$C_{\mathbf{k},eq} = An_k \begin{pmatrix} \frac{2(2k^2+k_d^2)}{1+\alpha} & k_d^2 \\ k_d^2 & \frac{2\alpha(2k^2+k_d^2)}{1+\alpha} \end{pmatrix} \tag{14}$$

where A is a tunable eddy amplitude parameter, α is the energy ratio between layers, and

$$n_k = \begin{cases} 0 & \text{for } k < k_0 \text{ and } k > k_{max} \\ (4k^{14/3}(k^2+k_d^2))^{-1} & \text{for } k_0 \le k \le k_d \\ k_d^{4/3}(4k^6(k^2+k_d^2))^{-1} & \text{for } k_d \le k \le k_{max}. \end{cases} \tag{15}$$

Here k_0 is the Nyquist wavenumber of the coarse resolution and k_{max} is 256, which is the Nyquist wavenumber of the reference simulation using 512×512 grid points for each layer. As the equilibrium covariance is specified, the noise level $\sigma_{\mathbf{k}}$ is chosen to satisfy

$$\sigma_{\mathbf{k}}\sigma_{\mathbf{k}}^* = 2\gamma_{\mathbf{k}}C_{\mathbf{k},eq}. \tag{16}$$

In the deterministic closure of superparameterization, the covariance equation is solved for and is integrated over all wave vectors. To reduce the cost and impose randomness of the subgrid-scale effect, the stochastic closure [16, 17, 14] introduces randomness in the direction θ of the wave vector $\mathbf{k} = (k\cos\theta, k\sin\theta)$ and calculate the subgrid effect by integrating the covariance

components along the radial direction. That is, the covariance equation is solved for each radial wavenumber $k = |\mathbf{k}|$

$$\frac{d}{d\tau}C_k = \mathbf{L}_k C_k + C_k \mathbf{L}_k^* + \sigma_k \sigma_k^*. \tag{17}$$

Using the equilibrium covariance as an initial value, the stochastic superparameterization solves the above ODE for a duration of $\frac{1}{\epsilon}$ and estimates each term of (11) using

$$\overline{u_1'(\psi_1' - \psi_2')} = \int_{k_0}^{k_{max}} k^2 \sin(\theta_1) \left(\epsilon \int_0^{1/\epsilon} \mathbb{E}[imag(\hat{\psi}_1 \hat{\psi}_2^*) \right) dk, \tag{18}$$

$$\overline{v_1'(\psi_2' - \psi_1')} = -\int_{k_0}^{k_{max}} k^2 \cos(\theta_1) \left(\epsilon \int_0^{1/\epsilon} \mathbb{E}[imag(\hat{\psi}_1 \hat{\psi}_2^*) \right) dk, \tag{19}$$

$$\overline{u_j' v_j'} = \frac{1}{2} \int_{k_0}^{k_{max}} k^3 \sin(2\theta_j) \left(\epsilon \int_0^{1/\epsilon} \mathbb{E}[|\hat{\psi}_j|^2] \right) dk, \tag{20}$$

$$\overline{(v_j')^2} - \overline{(u_j')^2} = \int_{k_0}^{k_{max}} k^3 \cos(2\theta_j) \left(\epsilon \int_0^{1/\epsilon} \mathbb{E}[|\hat{\psi}_j|^2] \right) dk, \tag{21}$$

$$\overline{u_2'(\psi_1' - \psi_2')} = -\overline{u_1'(\psi_1' - \psi_2')} \tag{22}$$

$$\overline{v_2'(\psi_1' - \psi_2')} = -\overline{v_1'(\psi_2' - \psi_1')} \tag{23}$$

Here the time period $1/\epsilon$ called eddy averaging time, a tunable parameter.

In the estimation of the subgrid-scale effect through stochastic modeling, the influence of the large-scale variables on the subgrid-scale is incorporated through the dependence of the linear operators on the large-scale variables (note that $\mathbf{L}_k$ in (17) depends on the large-scale variables). In a recent version of SSP [14], a temporal correlation and spatial smoothing techniques are introduced to impose a stronger correlation with the large-scale variables in the subgrid-scale parameterization. The spatial smoothing in [14] averages the subgrid-scale effect on a coarse grid point with the subgrid-scale effects on the two adjacent coarse grid points in each direction, which introduces a spatial correlation in the subgrid-scale parameterization. Regarding a correlation in time, the direction θ_j is modeled as a Wiener process on the circle

$$\theta_j = b_j dW_j, \tag{24}$$

where b_j is a tunable parameter that is set to close to the eddy turnover time $\frac{1}{Uk_d}$ (U is the RMS velocity). In [14], it is shown that these methods are essential to remove an artificial bump on the potential energy spectrum in addition to imposing large-scale correlations in the subgrid-scale effect. Combining all the techniques explained in this section, SSP requires only 96×96 spatial resolution to capture the features of quasigeostrophic turbulence, including breaking waves and jets. SSP has also proved to be an essential method in stabilizing data assimilation methods [24] in addition to an increased ensemble size due to its low computational complexity [23,15].

3.2 Tunable modeling parameters

In the stochastic superparameterization (and also in the stochastic parameterization), an important issue is how to choose the modeling parameters. SSP requires to specify six modeling parameters:

1. noise level b_j for the angle θ_j
2. base damping coefficient $\gamma_{0,j}$
3. eddy averaging time $1/\epsilon$
4. eddy amplitude A_j
5. energy ration α between layers
6. enhanced biharmonic viscosity ν_j

Physical information can help to estimate some of these parameters, such as eddy turnover time, etc, but such information often requires a high-resolution simulation result or dense observation data. Also, even if these parameters are available, some parameters (in our case, the eddy amplitude and the viscosity) are tuned to match the energy spectra. In the stochastic superparameterization, the modeling parameters, which can be different for each layer, are assumed to be equal for all layers due to the computational complexity to tune all of the parameters. The goal of the current study is to estimate the modeling parameters 1-5 (possibly for each layer), without using a high-resolution simulation result or dense observation data, while the viscosity is tuned through optimization in a low dimensional parameter space.

4 Subgrid-scale modeling parameter estimation

This paper proposes a method to estimate the modeling parameters in the stochastic superparameterization listed in section 3.2 (except the biharmonic viscosity). The key idea of the proposed method is to generate local subgrid-scale data and use its physical and statistical information to estimate the modeling parameters. As there is no scale separation between the resolved and unresolved scales, the local domain size to solve the PDE model is relatively larger than the one used in the heterogeneous multiscale method but remains small enough to be cheap in comparison with the reference simulation.

In general, the subgrid-scale parameterization depends on the large-scale variables. It is possible to generate local data as the large-scale variables change and estimate modeling parameters that also depend on the large-scale variables. This approach will definitely increase the correlations between the large-scale and the subgrid-scale effects. To reduce the cost and the complexity of the proposed method, we estimate modeling parameters that are independent of the large-scale variables. This approach does not mean that the subgrid-scale parameterization is independent of the large-scale variables. The linear operator $\mathbf{L}_k$ (10) depends on the large-scale variables and thus the subgrid-scale parameterization depends on the large-scale variables as we solve the covariance ODE for different set of large-scale variables, which is cheaper to

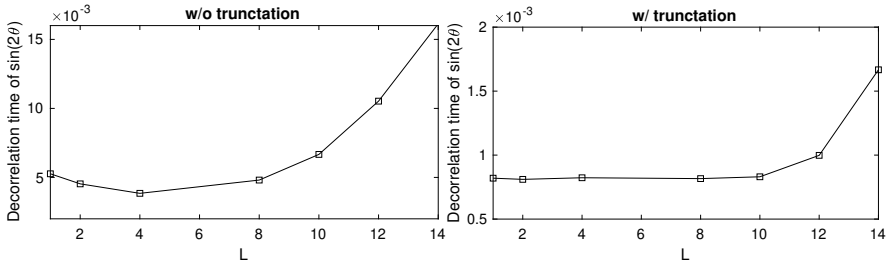
solve than the local PDE model. Additionally, to impose much stronger dependence on the large-scale variables in the subgrid-scale parameterization, the stochastic superparameterization uses the spatial smoothing method discussed in the previous section. As we look for modeling parameters independent of the large-scale variables, the local PDE simulation is run only one time to estimate the parameters.

The information to estimate the parameters are decorrelation times and energy of corresponding variables (the stream function ψ and the angle θ). As this information is available for each layer, we can estimate the modeling parameters separately for each layer. In this section, we describe the setup of a local simulation to generate subgrid-scale information and methods to estimate the modeling parameters from the local data.

4.1 Local simulation setup

The best (but not practical) way to generate the subgrid-scale information is to solve the subgrid equation (3) for each set of large-scale variables in the original domain $\Omega = (0, 2\pi)^2$. As we need only the subgrid-scale information, we solve the subgrid-scale equation in a local domain $\Omega_L = (0, \frac{2\pi}{L})^2$ where $L > 1$ is a scaling factor. In HMM, L is set to be an order of N where N is the number of the coarse grid points(thus the local domain size is comparable to the coarse mesh size). This approach is possible when there is strong scale separation between the resolved and unresolved scales. However, when there is no apparent scale separation as in our model, it is necessary to use a relatively large (but small compared to the original domain for efficiency) local domain to generate subgrid-scale information. In the current study, we choose L that covers several coarse grid points (this setup guarantees that the local domain contains coarse wave components, i.e., waves with wavenumber less than or equal to the Nyquist wavenumber of the coarse resolution).

As we assume that the modeling parameters are independent of the large-scale variables (while the overall subgrid-scale effect eventually depends on the large-scale variables through the operator $\mathbf{L}_k$), we run the local simulation only one time for a specific set of large-scale variables. Among many possibilities, we choose them to be zero. More specifically, the subgrid equation (3) (with all large-scale variables are set to zero) is solved in the local domain Ω_L with the double periodic boundary condition. An initial value is chosen as a random field from a small energy, and the simulation runs until it reaches a quasi-stationary state, i.e., the energy is saturated. Note that there is no modification for the subgrid-scale equation, except we choose zero values for the large-scale variables.



(a) raw velocity field without truncation (b) truncated velocity field

Fig. 3: Decorrelation time of upper layer $\sin(2\theta_1)$ as a function of the domain scaling factor L. Strong turbulent regime without (left) and with (right) low wavenumber truncation. Note the different vertical scales.

4.2 Data process for modeling parameter estimation

4.2.1 Noise level b_j for the angle θ_j

The noise level b_j in (24) for the angle θ_j is determined by the decorrelation time of $\sin(2\theta)$. In [14], it is shown that the decorrelation time of $\sin(2\theta_j)$ is equal to $\frac{1}{2b_j^2}$. Thus, if the decorrelation time of $\sin(2\theta_j)$ is available for each layer j, the noise level of θ_j, b_j, is estimated by

$$b_j = \sqrt{\frac{1}{2 \times \text{decorrelation time of } \sin(2\theta_j)}}. \tag{25}$$

We use the following way to estimate the decorrelation time of $\sin(2\theta_j)$. For each velocity field (u_j, v_j), the angle θ_j is estimated by $\arctan(v_j/u_j)$. From the angle information, an autocorrelation function is computed from a time series $\sin(2\theta_j)$. From the decaying rate of the autocorrelation function, the decorrelation time is estimated. This decorrelation time depends on the physical location to sample the time series. In our setup, we use four different physical locations, $(0,0), (\pi,0), (0,\pi)$, and (π,π), to estimate the noise level using the average of the decorrelation times at these locations.

Figure 3(a) shows the decorrelation time of the upper layer $\sin(2\theta_1)$ for the strong turbulent regime with varying local domain sizes. There is no consistency or convergence as the local domain size increases (i.e., $L \to 1$). This result can be explained by the relative size of the local domain compared to the local domain used in HMM. In HMM, the typical domain size is and is typically contained in a coarse grid mesh. Thus, the local simulation does not contain wave components corresponding to the coarse resolution. In our method, the local domain size is large enough that it contains a few coarse-scale wave components.

As an example, consider the case when $L = 10$ and a coarse resolution Nyquist wavenumber 48 in each direction. Considering only the zonal direction,

for example, wavenumber 1 to 4 in the local domain correspond to wavenumber 10, 20, 30, and 40 in the original domain. On the other hand, a wavenumber larger than 4 in the local domain do not have corresponding wavenumber that can be represented on the coarse grid (as there is no way to represent wavenumber 50 on the coarse resolution of Nyquist wavenumber 48). For the same Nyquist wavenumber, any L greater than 5 (we are assuming that L is an integer for simplicity) will contain two different wave groups that belong to coarse and subgrid scale components. In the estimation of the subgrid-scale modeling parameters, we need data only related to the subgrid-scale. Therefore, we truncate the wave components corresponding to the coarse resolution and use the remaining wave components to estimate the pure properties of the subgrid components. That is, we remove the low wavenumber components 1 to 4 in the above example, and use only wavenumber larger than 5 to estimate the subgrid modeling parameters. We note that the distinction between the low and the high wave groups is determined by the coarse resolution Nyquist wavenumber. Figure 1(b) shows the decorrelation time of $\sin(2\theta)$ after truncation, which shows convergence as L decreases.

4.2.2 Base damping coefficient $\gamma_{0,j}$ and eddy averaging time $1/\epsilon$

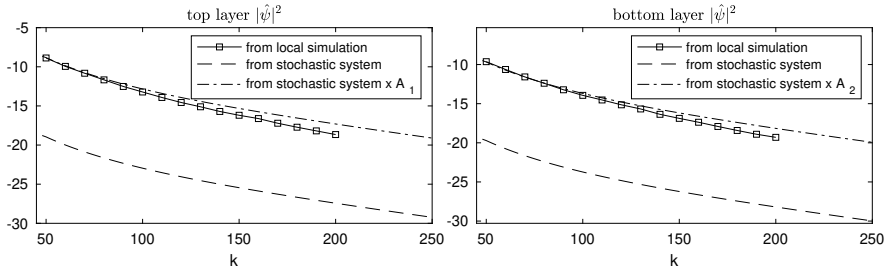
The base damping coefficient $\gamma_{0,j}$ is similarly estimated using a decorrelation time, which determines the damping coefficient γ_k through the relation (13), For a linear stochastic system for $u \in \mathbb{C}$

$$du = (-\gamma + i\omega)udt + \sigma dW,$$

where W is a Brownian motion, the autocorrelation function is proportional to $e^{-\gamma t}\cos(\omega_k t)$ where $\frac{1}{\gamma}$ is the decorrelation time (see Chapter 5 of [28] for an derivation of this relation). As the equation for $\hat{\psi}_j$ (9) is a linear stochastic system, we use the time series data of $\hat{\psi}_j$ to estimate the damping coefficient through the decorrelation time estimation. Note that ω_k is determined by the large-scale variables (10) and thus we need to estimate only γ that corresponds to $\gamma_k + \nu k^4$ in (10). For the simplicity and the efficiency of the algorithm, instead of estimating the damping coefficient for each k, we estimate the base damping coefficient γ_0 and use the model (13) to estimate γ_k from γ_0. We choose a wavenumber $\tilde{k}$ that is the minimum multiple of the scaling factor L such that $\tilde{k} \geq k_0$ (where k_0 is the Nyquist wavenumber of the coarse resolution). For this wavenumber, extract a time series of $\hat{\psi}_j(\tilde{k})$ from the local simulation data and estimate the decorrelation time of this time series, which yields

$$\gamma_{0,j} = \frac{1}{\text{decorrelation time of } \hat{\psi}_j(\tilde{k})} - \nu\tilde{k}^4. \tag{26}$$

In the estimation of the eddy averaging time that is same for both layers, the averaging time must be long enough for the system to evolve away from the initial state but not too long compared to the coarse time step. We choose



(a) Top layer (b) Bottom layer

Fig. 4: Spectra of time averaged $|\hat{\psi}|^2$. Strong turbulent regime with $L = 10$. Top (left) and bottom (right) layers.

the eddy averaging time to be two times longer than the average decorrelation time of $\hat{\psi}_j(\tilde{k})$

$$\begin{aligned}\frac{1}{\epsilon} &= 2 \times \frac{\text{decorrelation time of } \hat{\psi}_1(\tilde{k}) + \text{decorrelation time of } \hat{\psi}_2(\tilde{k})}{2} \\ &= \text{decorrelation time of } \hat{\psi}_1(\tilde{k}) + \text{decorrelation time of } \hat{\psi}_2(\tilde{k}),\end{aligned} \tag{27}$$

where $\tilde{k}$ is the smallest multiple of L that is greater than or equal to the Nyquist wavenumber.

4.2.3 Eddy amplitude A_j and energy ratio α

The two modeling parameters A_j and α in (14) are related to energy. A_j determines the actual energy scale (the modeling of the equilibrium covariance (the right hand side of (14)) does not provide any information for the actual magnitude). In our model (1), the effect of topography is described as a drag (corresponding to the term $-r\nabla^2\psi_2$) and thus the bottom layer has less energy than the upper layer. Parameter α is essential to determine the energy ratio between different layers. In the stochastic superparameterization, where A_j is assumed to be same for both layers, specifying α is crucial as it is the only way to impose different energy scale for each layer. In the current study, we estimate A_j separately for each layer and thus the effect of α is not important. Thus, we set $\alpha = 1$ and estimate the eddy amplitudes for each layer using the following covariance model

$$C_{k,eq} = n_k \begin{pmatrix} A_1(2k^2 + k_d^2) & \sqrt{A_1}\sqrt{A_2}k_d^2 \\ \sqrt{A_1}\sqrt{A_2}k_d^2 & A_2(2k^2 + k_d^2) \end{pmatrix}. \tag{28}$$

As the covariance matrix C_k is initialized proportional to the equilibrium covariance and the noise covariance is also proportional to the equilibrium covariance, the variances $\mathbb{E}(|\hat{\psi}_1|^2)$ and $\mathbb{E}(|\hat{\psi}_2|^2)$ are proportional to A_1 and A_2

respectively. Thus, if $\mathbb{E}(|\hat{\psi}_j|^2)_{true}$ and $\mathbb{E}(|\hat{\psi}_j|^2)_{sde}$ represent the variances of $\hat{\psi}$ from the true signal and from the SDE system, they satisfy

$$\mathbb{E}(|\hat{\psi}_j|^2)_{true} = A_j \mathbb{E}(|\hat{\psi}_j|^2)_{sde} \tag{29}$$

Instead of the true variance of $\hat{\psi}$ (which is not available), we use the variance from the local simulation, $\mathbb{E}(|\hat{\psi}_j|^2)_{local}$, to estimate the eddy amplitude

$$A_j = \frac{\mathbb{E}(|\hat{\psi}_j|^2)_{local}}{\mathbb{E}(|\hat{\psi}_j|^2)_{sde}}. \tag{30}$$

Figure 4 shows the spectra of time averaged $|\hat{\psi}_j|^2$ for both layers from both the local simulation (squared) and the stochastic system (dash line) for the strong turbulent regime with $L = 10$. As the damping coefficient γ_k is modeled as (13) based on qualitative information of two-dimensional geostrophic turbulent flow while the local domain does not cover the original domain, two spectra have slightly different decay rates. The lowest wavenumber has the largest magnitude that has the most significant impact on the coarse scale. Therefore, we estimate the eddy amplitude by matching the time averaged $|\hat{\psi}|^2$ at the smallest wavenumber (the dash-dot line in Figure 4 is the spectra multiplied by the estimated eddy amplitude).

4.2.4 Enhanced biharmonic viscosity ν_j

The remaining modeling parameter is the enhanced biharmonic viscosity ν_j. This parameter is estimated by matching the energy spectra of the system as in the standard stochastic superparameterization. In contrast to matching the energy spectra over all modeling parameters in the stochastic superparameterization, we match the energy spectral using only two scalar values, ν_1 and ν_2 and thus this process is significantly efficient than the original stochastic superparameterization.

It is shown in [20] that the order of the viscosity term must increase with resolution and the strength of the turbulence to correctly match the energy spectra without artificial dissipation regime of the coarse model. That is, it is more appropriate to use hyperviscosity instead of biharmonic viscosity in the current setting. This approach, however, requires non-constant viscosity varying over wavenumber to capture other features, such as time averaged zonal jets [21]. Instead of estimating the spectral slope and use this information to determine the varying viscosity coefficients, we restrict our interest to the biharmonic viscosity and constant viscosity coefficients, which still captures zonal jets and large-scale energy spectra at the cost of artificial dissipative regime.

5 Numerical tests

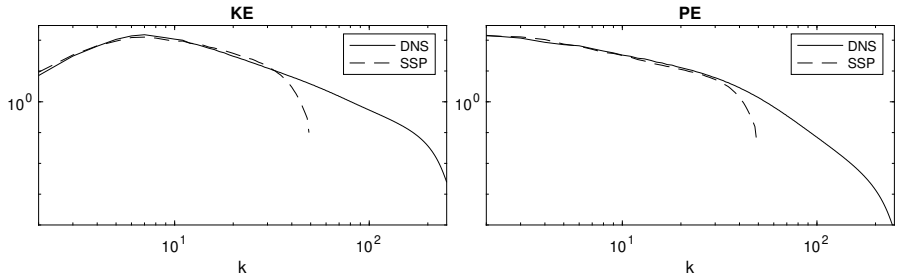
We test the stochastic superparameterization for the two test regimes of the model equation (1), strong and moderate turbulent regimes, using the modeling parameter values estimated by the proposed method. First, the coarse model (2) is resolved by 96×96 grid points for each layer using the same numerical method as the reference simulation, energy and enstrophy conserving Arakawa finite difference [2] for the advection term and the Fourier spectral method for the linear terms. Also, we use the same fourth-order time integrator for the coarse model. The derivatives related to the subgrid terms are approximated by the second-order finite difference with the spatial smoothing [14]. The computational cost of the stochastic superparameterization is 280 times cheaper than the reference simulation considering the time step 2×10^{-4} for the coarse model. Regarding the subgrid-scale covariance, the covariance matrix (17) is solved using a fourth-order Runge-Kutta method for wavenumber $k = 48, 49, ..., 256$ where 48 and 256 correspond to the Nyquist wavenumbers of the coarse and the reference resolutions. Instead of solving this equation at each coarse time step when a new set of large-scale variables is available, we precompute the system for possible combinations of large-scale variables. In the stochastic superparameterization, the actual subgrid-scale terms are interpolated from the precomputed values.

To estimate the modeling parameters, we use a local domain with scaling factor $L = 10$, which is the smallest local domain size to get converged modeling parameters as L changes (see Figure 3 for the estimated decorrelation time of $\sin(2\theta_j)$ as a function of L). This local domain covers only $\frac{1}{100}$ of the original domain Ω, and thus the local simulation can be done efficiently in comparison with the reference simulation. For the local simulation, we use a slightly finer resolution than the reference simulation using 64×64 grid points in each layer (note that the number of grid points for $L = 10$ maintaining the same resolution of the reference simulation is 52×52). The numerical scheme to solve the subgrid-scale equation (3) is the same as the reference simulation, while all large-scale variables are set to zero. The system is initialized from a random field with small energy and runs until the energy is saturated. This system is solved only one time for each test regime. The estimated modeling parameters are shown in Table 2.

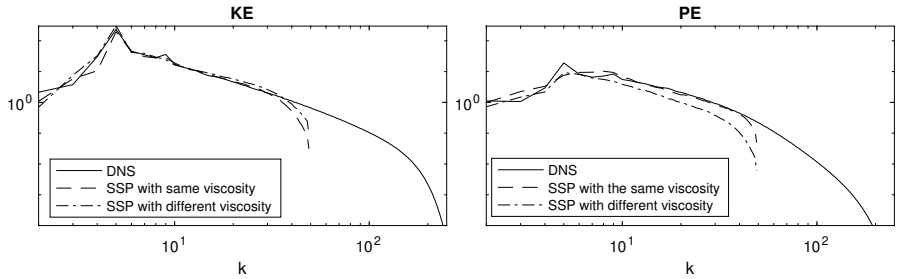
The angle θ_j changes more randomly in the strong regime than in the moderate regime, i.e., b_j is larger in the strong case, which complies with the physics, although the difference is marginal. There is a significant difference in other parameters between the regimes. The strong turbulent regime has a strong damping coefficient for the subgrid-scale, which is about 2.7 times stronger than the moderate regime. The energy in the strong regime subgrid-scale is about seven times larger than in the moderate case. The viscosity also follows the same relation, stronger viscosity for the strong regime than for the moderate regime. This result conforms to the study in [25]. The study shows that the eddy viscosity in the cascading inertial range is proportional to the cube root of the constant enstrophy flux, which is also shown by the scaling

	Strong turbulent		Moderate turbulent	
	upper layer	bottom layer	upper layer	bottom layer
b_j	2.47×10^1	2.38×10^1	2.17×10^1	2.04×10^1
$\gamma_{0,j}$	8.06×10^1	7.75×10^1	3.03×10^1	2.94×10^1
$1/\epsilon$	2.53×10^{-2}		6.69×10^{-2}	
A_j	3.31×10^4	1.57×10^4	4.63×10^3	2.53×10^3
ν_j	5×10^{-5}	5×10^{-5}	1×10^{-5}	2×10^{-6}

Table 2: Estimated stochastic superparameterization modeling parameters. The local domain Ω_L has a scaling factor $L = 10$ that covers only $\frac{1}{100}$ of the original domain Ω.



(a) Strong turbulent regime kinetic (left) and potential (right) energy spectra

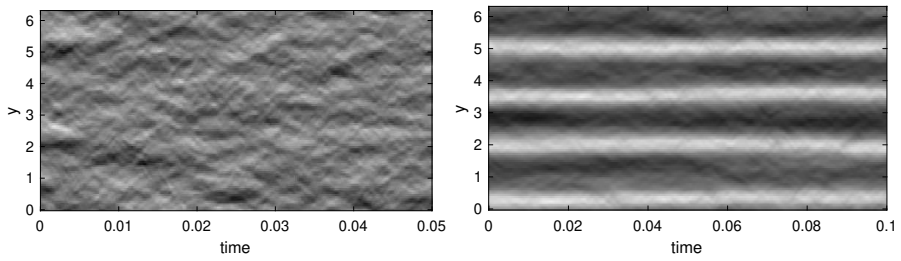


(b) Moderate turbulent regime kinetic (left) and potential (right) energy spectra with different (dash line) and same (dash-dot line) viscosity for both layers

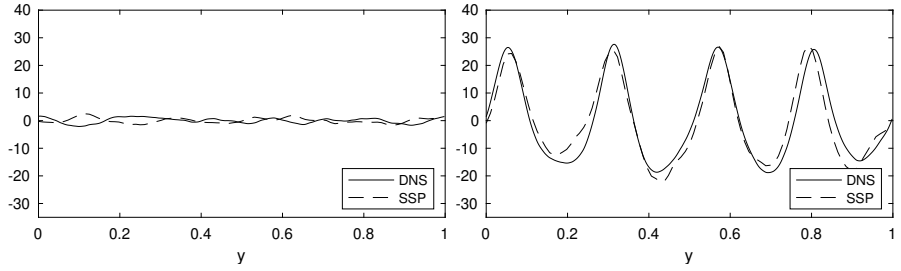
Fig. 5: Kinetic energy (KE) and potential energy (PE) spectra by SSP (dash line), along with the reference result (solid line). The dash-dot line uses different viscosity coefficients between different layers.

laws in [20]. A similar pattern is observed between different layers. The bottom layer has a drag modeling the viscous effect of the bottom layer, and this drag plays an additional stabilizer of the bottom layer. Thus, it is natural to expect that the bottom layer is less turbulent than the upper layer. The estimated modeling parameters support the relation; the bottom layer has a weaker noise for the angle, weaker damping for the subgrid-scale, and less energy than the upper layer.

The kinetic energy (KE) and potential energy (PE) spectra of the stochastic superparameterization (SSP) are shown in Figure 5 along with the reference



(a) Time series of zonal average of barotropic velocity component u_{bt}



(b) Time and zonal average of barotropic velocity component u_{bt}

Fig. 6: Top: time series of zonal average of barotropic velocity component u_{bt}. Bottom: time and zonal average of barotropic velocity component u_{bt}. Strong (left) and moderate (right) regimes.

simulation (DNS) spectra. The energy spectra of the stochastic superparameterization show a good match with the reference spectra as the parameterization accounts for the subgrid-scale effect judiciously. We note that as we have seen in Figure 2, the parameterization of the subgrid-scale is essential to obtain the correct spectrum shape using 96×96 grid points in each layer. In the standard stochastic superparameterization with all tuned parameters by matching the energy spectra, each modeling parameter is assumed to be same for both layers, which significantly reduced the dimension of the modeling parameters by a factor 2. To see the effect of the same parameter value for both layers, the dash-dot line in Figure 5(b) shows the energy spectra of the moderate regime using the same viscosity value for both layers. While the kinetic energy spectrum is not affected, the potential energy, which is related to the difference of the stream function between different layers, degrades significantly where there is no difference in the viscosity for different layers; the tail of the potential energy decreases if the same viscosity value is used for both layers.

We also check the spatial structure generated by the stochastic superparameterization through a time series of the zonal average of the barotropic zonal velocity component u_{bt} shown in Figure 6(a). In comparison with the reference result (Figure 1(a)), SSP generates similar patterns mimicking critical features of each regime; in the strong turbulent regime, the time series by

SSP shows breaking waves with no significant spatial structures. In the moderate turbulent regime that is anisotropic and inhomogeneous, SSP shows the same number of zonal jets of comparable magnitudes (see Figure 6(b) for the time and zonal average of the barotropic velocity component u_{bt}). We want to note that the stochastic superparameterization can capture the anisotropic and inhomogeneous structure of the moderate turbulent regime using modeling parameters that are independent of large-scale variables. Anisotropic and inhomogeneous characteristics in the coarse-scale are achieved by solving the ODE for the covariance (17) that depends on large-scale variables, while the modeling parameters remain independent of large-scale variables. As a quantitative performance of the proposed method, the heat flux using the estimated modeling parameters is 235 for the strong turbulent regime and 22 for the moderate turbulent regime. These values are in a 10% error range of the true values (223 and 20 for the strong and the moderate regimes, respectively).

One issue of the standard superparameterization, estimating all modeling parameters by matching the energy spectra, is that the modeling parameter set is not unique. We have already seen that we can match the kinetic energy spectra by using the same viscosity for both layers. As a more extreme case, we found the following set of parameters for the moderate turbulent regime by matching the energy spectra without using the proposed method

$$\begin{aligned} &b_1 = b_2 = 1.78 \times 10^1, \gamma_{0,1} = \gamma_{0,2} = 30, 1/\epsilon = 4.00 \times 10^{-2}, \\ &A_1 = 3330, A_2 = 1670, \nu_1 = \nu_2 = 8 \times 10^{-6}. \end{aligned} \tag{31}$$

The energy spectra and the time and zonal average of barotropic velocity component u_{bt} of the stochastic superparameterization using this parameter set are shown in Figure 7. Except for the kinetic energy spectra, all the other measures show significant differences between the stochastic superparameterization and the reference result. The potential energy is lower in the tail, and the jets are much weaker than the reference result. The heat flux is much worse, 32 by the stochastic superparameterization, compared to 20 from the reference simulation. The proposed method uses (local) information from the model to estimate the modeling parameters that are physically correct and thus provides a more accurate result than just tuning the parameters to match the energy spectra.

6 Discussion and conclusion

In geophysical fluid systems, the stochastic parameterization of unresolved subgrid-scale dynamics as a low computational cost mutiscale model involves several modeling parameters to be determined (for example, additional damping coefficient, noise level, etc.). The estimation of these parameters typically requires a specific type of information (such as dense observation data) that may not be available. We proposed an approach to estimate subgrid-scale modeling parameters in the stochastic superparameterization of quasigeostrophic turbulence. By assuming that only energy spectra are available, the proposed

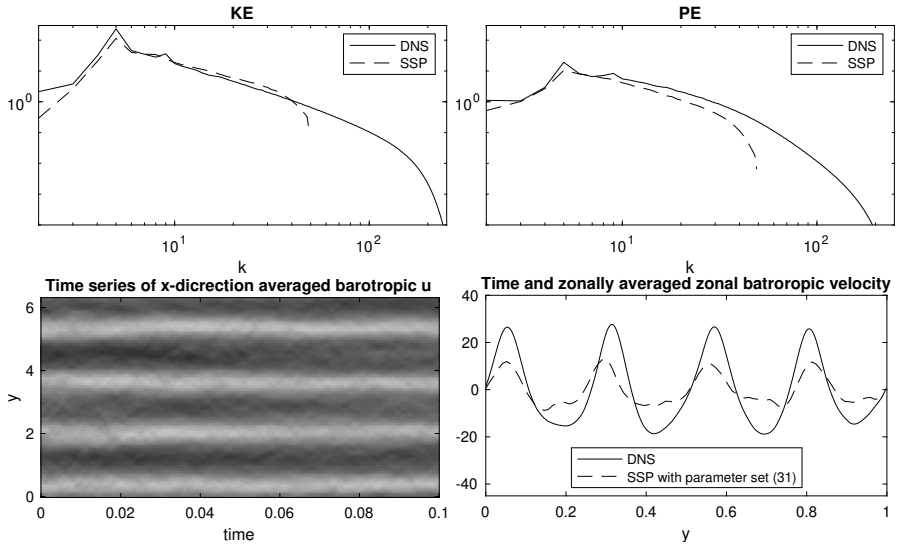


Fig. 7: Moderate turbulent regime result using the parameter set (31). This parameter set is hand tuned by matching the energy spectra without using the proposed method. The time series has the same gray scale as the other time series plots.

method directly uses the subgrid-scale PDE equation to generate local data for subgrid-scale variables. The method utilizes the physical/statistical information of the local data to estimate the modeling parameters.

In the subgrid-scale parameterization, it is natural to assume that the subgrid-scale effect depends on the large-scale variables. Instead of estimating the subgrid-scale effect on the large-scale dynamics directly, the stochastic superparameterization approximates the subgrid-scale effect by solving a cheap ODE system where large-scale variables play as input parameters of the system. The modeling parameters used in the ODE system are independent of large-scale variables. Therefore, the local PDE problem needs to be solved only one time to estimate the modeling parameters.

We considered two turbulent regimes (strong and moderate regimes) with an inverse cascade of energy to test the proposed method. The strong turbulent regime is isotropic and homogeneous with breaking wave patterns and a strong inverse cascade of kinetic energy. The moderate turbulent regime is anisotropic and inhomogeneous with a nontrivial spatial structure, zonal jets. The proposed method showed skillful prediction in capturing complicated features of the test regimes, including complicated wave patterns, turbulent energy spectra, and heat flux.

The proposed method can estimate most of the modeling parameters in the stochastic superparameterization except one parameter, the enhanced biharmonic viscosity. This parameter is tuned to match energy spectra. It is nat-

ural to investigate an estimation method of this parameter using local data or physics for the completeness of the proposed approach. Particularly, we plan to incorporate the higher-order viscosity term with viscosity coefficients varying over wavenumber [21] to remove the artificial dissipative regime and further reduce the resolution of the coarse model. For an advection-diffusion problem where the velocity field has a turbulent spectrum, there is a numerical technique to estimate the enhanced diffusion coefficient using a hierarchical computation of the subgrid-scale effect [22]. It would be interesting to apply or extend the method in [22] to the current setting as an estimation method of the enhanced viscosity.

The quasigeostrophic model in this study has several features of geophysical turbulence, but many realistic features are lacking, such as lateral boundaries, adiabatic effect, a more realistic topography, etc. Also, there are scenarios in which the subgrid-scale is also anisotropic. In the current study, we have considered isotropic subgrid-scale problems, while anisotropic and inhomogeneous characteristics appear on the large-scale. This characteristic can be a reason for the success of the current study that estimates modeling parameters independent of the large-scale variables. To validate the effectiveness and robustness of the proposed method, we plan to apply the proposed method to various test scenarios, including a case to estimate modeling parameters that strongly depend on the large-scale variables. This investigation will be reported in another place.

Acknowledgements The author would like to thank Andrew J. Majda and Bjorn Engquist for their comments and encouragement that made this work possible.

Conflict of interest

The author declares no conflict of interest.

References

1. A. Abdulle, W. E, B. Engquist, E. Vanden-Eijnden, The heterogeneous multiscale method, Acta Numer. 21 (2012) 1–87.
2. A. Arakawa, Computational design for long-term numerical integration of the equations of fluid motion: Two-dimensional incompressible flow. Part I, J. Comput. Phys., 1 (1966), pp. 119–143.
3. J. Bardina, J. H. Ferziger, and W. C. Reynolds, Improved Subgrid-Scale Models for Large Eddy Simulation, AIAA Paper 80–1357, AIAA, Reston, VA, 1980.
4. J. Berner, G. J. Shutts, M. Leutbecher, and T. N. Palmer, A spectral stochastic kinetic energy backscatter scheme and its impact on flow-dependent predictability in the ECMWF ensemble prediction system., J. Atmospheric Sci., 66 (2009), pp. 603–626.
5. G. Boffetta and R. E. Ecke, Two-dimensionalturbulence, Ann. Rev. Fluid Mech., 44(2012), pp. 427–451.
6. J.-M. Campin, C. Hill, H. Jones, J. Marshall, Super-parameterization in ocean modeling: Application to deep convection, Ocean Model. 36 (2011) 90–101.
7. J. G. Charney, Geostrophic turbulence, J. Atmospheric Sci., 28 (1971), pp. 1087–1095.

8. T. Delsole, Stochastic models of quasigeostrophic turbulence, Surv. Geophys. 25 (2004) 107–149.
9. D. Dritschel, M. McIntyre, Multiple jets as PV staircases: the Phillips effect and the resilience of eddy-transport barriers, J. Atmos. Sci. 65 (2008) 855–874.
10. W. E and B. Engquist, The heterogeneous multi-scale method, Comm. Mat, Science, 1 (2003), 87-133.
11. B.F. Farrell, P.J. Ioannou, A theory of baroclinic turbulence, J. Atmos. Sci. 66 (2009) 2444–2454.
12. W. Grabowski, P. Smolarkiewicz, CRCP: a Cloud Resolving Convection Parameterization for modeling the tropical convecting atmosphere, Physica D 133 (1999) 171–178.
13. S. M. Griffies and R. W. Hallberg, Biharmonic friction with a Smagorinsky-like viscosity for use in large-scale eddy-permitting ocean models, Monthly Weather Rev., 128 (2000), pp. 2935–2946.
14. I. Grooms, Y. Lee, A. Majda, Numerical schemes for stochastic backscatter in the inverse cascade of quasigeostrophic turbuelcne, Mult. Model Sim., 13(3) (2015), 1001–1021.
15. I. Grooms, Y. Lee, and A. Majda, Ensemble filtering and low-resolution model error: Additive inflation, stochastic parameterization, and model numerics, Monthly Weather Review, 143(10) (2015), 3912-3924.
16. I. Grooms and A. J. Majda, Efficient stochastic superparameterization for geophysical tur- bulence, Proc. Natl. Acad. Sci. USA, 110 (2013), pp. 4464–4469.
17. I. Grooms and A. J. Majda, Stochastic superparameterization in quasigeostrophic turbulence, J. Comput. Phys., 271 (2014), pp. 78–98.
18. M. F. Jansen and I. M. Held, Parameterizing subgrid-scale eddy effects using energetically consistent backscatter, Ocean Model., 80 (2014), pp. 36–48.
19. M. Khairoutdinov, D. Randall, C. DeMott, Simulations of the atmospheric general circulation using a cloud-resolving model as a superparameterization of physical processes, J. Atmos. Sci. 62 (2005) 2136–2154.
20. V. Kitsios, J.S. Frederiksen, and M.J. Zidikheri, Subgrid model with scaling laws for atmospheric simulations, Journal of the Atmospheric Sciences, 69 (2012) 1427–1445.
21. V. Kitsios and J.S. Frederiksen, Subgrid parameterization of eddy-eddy, eddy-meanfield, eddy-topographic, meanfield-meanfield and meanfield-topographic interactions in atmospheric models, Journal of the Atmospheric Sciences, 76 (2019), 457–477.
22. Y. Lee and B. Engquist, Multiscale numerical methods for advection-diffusion in incompressible turbulent flow fields, Journal of Computational Physics, 317(15) (2016), 33–46.
23. Y. Lee, A. Majda, and D. Qi, Stochastic superparameterization and multiscale filtering of turbulent tracers, Mult. Model Sim., 15(1) (2017), 215–234.
24. Y. Lee, A. Majda, and D. Qi, Preventing catastrophic filter divergence using adaptive additive inflation for baroclinic turbulence, Monthly Weather Review, 145 (2) (2017), 669-682
25. C.E. Leith, Atmospheric Predictability and Two-Dimensional Turbulence, Journal of the Atmospheric Sciences, 78(1971), 145–161.
26. C.E. Leith, Stochastic backscatter in a subgrid-scale model: Plane shear mixing layer, Phys. Fluids A, 2 (1990), pp. 297–299.
27. A. J. Majda and I. Grooms, New perspectives on superparameterization for geophysical tur- bulence, J. Comput. Phys., 271 (2014), pp. 60–77.
28. A. J. Majda and J. Harlim, Filtering Complex Turbulent Systems, Cambridge University Press, Cambridge, UK, 2012.
29. W. D. McComb, The physics of fluid turbulence, Oxford Engineering Science Series, vol. 25, Clarendon Press New York, 1991.
30. C. Meneveau and J. Katz, Scale-invariance and turbulence models for large-eddy simulation, Ann. Rev. Fluid Mech., 32 (2000), pp. 1–32.
31. P. G. L. Porta Mana and L. Zanna, Toward a stochastic parameterization of ocean mesoscale eddies, Ocean Model., 79 (2014), pp. 1–20.
32. R. Salmon, Lectures on Geophysical Fluid Dynamics, Oxford University Press, New York, 1998.
33. U. Schumann, Stochastic backscatter of turbulence energy and scalar variance by random subgrid-scale fluxes, Proc. Roy. Soc. London Ser. A Math. Phys. Sci., 451 (1995), pp. 293– 318.

34. Y. Tago and Y. Higuchi, Fluid Flow Analysis of Jets from Nozzles in Top Blown Process. ISIJ Int. 43 (2003), 209–215.
35. A.F. Thompson, W.R. Young, Two-layer baroclinic eddy heat fluxes: Zonal flows and energy balance, J. Atmos. Sci. 64 (2007) 3214–3231.
36. S Thual, A. J. Majda, and S. N. Stchmann, A Stochastic Skeleton Model for the MJO, Journal of the Atmospheric Sciences, 71 (2014), 697–715.
37. G.K. Vallis, Atmospheric and Oceanic Fluid Dynamics, Cambridge University Press, 2006.