

Stochastic superparameterization in a quasigeostrophic model of the Antarctic Circumpolar Current



Ian Grooms^{a,*}, Andrew J. Majda^{a,b}, K. Shafer Smith^{a,b}

^a Center for Atmosphere Ocean Science, Courant Institute of Mathematical Sciences, New York University, 251 Mercer St., New York, NY 10012, USA

^b Center for Prototype Climate Modelling, NYU-Abu Dhabi, United Arab Emirates

ARTICLE INFO

Article history:

Received 24 June 2014

Received in revised form 8 September 2014

Accepted 11 October 2014

Available online 28 October 2014

Keywords:

Stochastic parameterization

Mesoscale parameterization

Superparameterization

Southern Ocean

ABSTRACT

Stochastic superparameterization, a stochastic parameterization framework based on a multiscale formalism, is developed for mesoscale eddy parameterization in coarse-resolution ocean modeling. The framework of stochastic superparameterization is reviewed and several configurations are implemented and tested in a quasigeostrophic channel model – an idealized representation of the Antarctic Circumpolar Current. Five versions of the Gent–McWilliams (GM) parameterization are also implemented and tested for comparison. Skill is measured using the time-mean and temporal variability separately, and in combination using the relative entropy in the single-point statistics. Among all the models, those with the more accurate mean state have the less accurate variability, and vice versa. Stochastic superparameterization results in improved climate fidelity in comparison with GM parameterizations as measured by the relative entropy. In particular, configurations of stochastic superparameterization that include stochastic Reynolds stress terms in the coarse model equations, corresponding to kinetic energy backscatter, perform better than models that only include isopycnal height smoothing.

© 2014 Elsevier Ltd. All rights reserved.

1. Introduction

Mesoscale eddies play a dominant role in stirring and mixing active and passive tracers in the oceans. These eddies have typical horizontal scales between 80 and 200 km (Chelton et al., 2011), and are often not fully resolved in global coupled model simulations. The most widely used parameterizations of the effects of ocean mesoscale eddies are based on the GM parameterization (Gent and McWilliams, 1990; Gent et al., 1995); indeed, most of the IPCC AR4 climate models used ocean mesoscale parameterizations based on the GM parameterization (Kuhlbrodt et al., 2012). As the resolution of ocean models increases these eddies are beginning to be partially resolved in global coupled models; nevertheless, the need to run ensemble simulations for prediction and data assimilation, the need to run models with complex ocean–ice–land–atmosphere interactions with large numbers of organic and inorganic tracers, and the fact that partially-resolved eddies are not always superior to parameterized ones (Hallberg, 2013) implies that mesoscale eddy parameterization will remain relevant for some time to come.

Most ocean mesoscale parameterizations are developed with the goal of capturing the average feedback from the unresolved scales, and as such, model eddy fluxes of tracers and momentum as deterministic functions of the large scales. This approach is reasonable, and is justified e.g. by multiple-scales asymptotic arguments (Pedlosky, 1984; Grooms et al., 2011), when the subgridscale dynamics evolve on space and time scales much smaller than the resolved large scales. But when the coarse grid scale begins to encroach upon the mesoscale eddy range, the feedbacks from the subgridscale should no longer be expected to be purely deterministic. Instead, they include significant variability about the average value, and this variability can in principle have a pronounced effect on the resolved large-scale dynamics.

Stochastic parameterizations have been developed to include subgridscale variability in, for example, engineering-scale models (Leith, 1990; Mason and Thomson, 1992; Schumann, 1995; Marstorp et al., 2007), atmospheric models (Buizza et al., 1999; Berner et al., 2009; Lin and Neelin, 2000, 2003; Khouider et al., 2003; Frenkel et al., 2013; Khouider et al., 2010; Deng et al., Submitted for publication; Frederiksen and Davies, 1997), and ocean models (Berloff, 2005; Kitsios et al., 2013, 2014; Jansen and Held, 2014; Mana et al., 2014; Brankart, 2013; Grooms and Majda, 2013; Grooms and Majda, 2014). One typical impact of stochastic parameterizations is to increase the internal variability of a model. This leads to increased spread of prediction ensembles

* Corresponding author.

E-mail address: grooms@cims.nyu.edu (I. Grooms).

(Buizza et al., 1999), the need for which has been recently noted by Karspeck et al. (2013) in a 1° ocean model: models with too little internal variability lead to tightly clustered prediction ensembles that can cause assimilation systems to be over-confident in the model at the expense of the observations. Improved internal variability can also lead to improved long-term prediction skill. The skill of a model's long-term prediction under climate change (meaning here any change of forcing) has recently been linked to the model's 'climate fidelity', which is the accuracy of the model's unperturbed climatological statistics – including both mean and variability (Shukla et al., 2006; DelSole and Shukla, 2010; Majda and Gershgorin, 2010, 2011a,b).

Stochastic superparameterization is a framework for developing stochastic subgridscale parameterizations that has recently been developed in an idealized quasigeostrophic setting for eddy-permitting resolution (Grooms and Majda, 2013, 2014). The link to superparameterization ('SP'; Randall et al., 2003, 2013; Grabowski, 2004) is discussed by Majda and Grooms, 2014. This article develops stochastic SP for non-eddy-permitting coarse models in an idealized setting. More details on stochastic SP are given in Section 3.2. For comparison with stochastic superparameterization we implement several variations on the GM parameterization (Gent and McWilliams, 1990). In the two-layer quasigeostrophic model context the GM parameterization amounts to layer interface smoothing with coefficient κ , and we test five different schemes for setting κ as a function of the resolved large-scale flow.

The idealized setting used here is a two-layer, wind-forced quasigeostrophic channel model with complex topography: an idealized model of the Antarctic Circumpolar Current (ACC). Although they lack a variety of important physical features, few-layer quasigeostrophic models have been successfully used to study qualitative features of the Southern Ocean (e.g. Nadeau and Straub, 2009, 2012; Nadeau et al., 2013; Hutchinson et al., 2010; Hogg et al., 2008; Meredith and Hogg, 2006; Meredith et al., 2012). Mesoscale eddies have a significant impact on the large-scale dynamics in idealized wind-driven channels and in more realistic models of the ACC, e.g. by limiting the increase of transport in response to increased wind forcing. This effect is known as eddy saturation (Meredith et al., 2012; Munday et al., 2013), and was predicted by Straub (1993). Although eddies are instrumental in flattening the isopycnals, and thereby reducing the transport through thermal-wind balance, either large-scale topography or interaction with subpolar gyre dynamics is also necessary to saturate the transport (Tansley and Marshall, 2001; Hallberg and Gnanadesikan, 2001; Nadeau and Straub, 2009, 2012; Nadeau et al., 2013). Our idealized ACC includes complex topography for the purposes of limiting the net transport, and increasing the complexity of the structure of the isopycnal surface.

The fact that mesoscale eddies have a pronounced effect on the dynamics of wind-driven channels makes it a useful setting to test parameterizations. We focus in particular on the ability of parameterized coarse models to accurately represent both the time-mean and the variability of the single isopycnal surface (the layer interface) in our two-layer model. We measure the accuracy in the mean and variability separately, and also in combination using the area-averaged relative entropy, as in DelSole and Shukla (2010). The relative entropy is an information-theoretic measure of how well one statistical distribution approximates another (Kleeman, 2002; Majda et al., 2002). It is used here to measure how well the climatological distributions of the parameterized coarse models approximate the true climatology from the high resolution reference simulation. The mathematical definition of relative entropy and further discussion are provided in 4.2.

The article is organized as follows. Section 2 presents the configuration and results of the eddy-resolving reference simulation. Section 3 reviews five GM parameterizations appropriate to this

setting, and introduces stochastic SP. Section 4 presents and discusses the results of simulations using the parameterized coarse models. Conclusions are offered in Section 5. Two appendices discuss boundary conditions and barotropic transport in the quasigeostrophic channel.

2. Model configuration and eddy-resolving reference simulation

2.1. Physical parameters and model numerics

Our idealized model of the ACC consists of two-equal-layer QG dynamics in a zonally reentrant channel with topography, forced by steady zonal wind stress. The dynamics are governed by the quasigeostrophic potential vorticity (PV) in each layer, which evolves according to the following equations

$$\partial_t q_1 + J[\psi_1, q_1 + \beta y] = -\frac{2}{\rho_0 H} \partial_y \tau(y) - \nu_4 \nabla^6 \psi_1 \quad (1)$$

$$\partial_t q_2 + J[\psi_2, q_2 + \beta y] = -\frac{2f_0}{H} J[\psi_2, h_b] - r \nabla^2 \psi_2 - \nu_4 \nabla^6 \psi_2. \quad (2)$$

The PV in each layer is q_i ($i = 1$ is upper and $i = 2$ is lower), the streamfunction is ψ_i , the total depth is H , and the local Coriolis frequency is f_0 ; ν_4 is the biharmonic viscosity coefficient, r is the Ekman drag coefficient, and β is the meridional gradient of planetary vorticity. The velocity is $\mathbf{u}_i = (-\partial_y \psi_i, \partial_x \psi_i)$ and the wind stress profile is $\tau(y) = \tau_0 \sin^2(\pi y/L_y)$. The boundary conditions are stress-free $\partial_n^2 \psi_i = \partial_n^4 \psi_i = 0$. The values of the physical parameters can be found in Table 1.

The local height of the ocean floor (above its mean depth of 4 km) is h_b . It is important to include topography because zonal channels develop unrealistically large transport in response to moderate wind forcing unless either strong bottom friction or topography is included (see Treguier and McWilliams, 1990 and Appendix B). The topography is constructed similarly to Nadeau et al. (2013) using the 5-Minute Gridded Global Relief Data Collection (ETOPO5) dataset. The ETOPO5 data set is used to supply complex topography without resorting to randomly-generated topography; the model is not intended to accurately represent real Southern Ocean dynamics. The model topography is a smoothed, truncated version of the ETOPO5 data derived from a latitude band surrounding Drake Passage. The maximum amplitude of the topography is 1 km, and Gaussian buffers with height equal to the maximum topographic height and with half-width 250 km are added at the northern and southern boundaries. This is necessary to prevent the formation of viscous boundary layers where the topographic advection term $u_2 \partial_x h_b$ is balanced by viscosity. Such boundary

Table 1
Model parameters.

Parameter	Symbol	Value
Rosby deformation radius	L_d	12.25 km
Horizontal resolution	$\Delta x = \Delta y$	6.51 km
Channel length	L_x	20,000 km
Channel width	L_y	2500 km
Total depth	H	4 km
Wind stress amplitude	τ_0	0.2 N/m ²
Coriolis parameter	f_0	-10^{-4} s^{-1}
Beta parameter	β	$1.5 \times 10^{-11} \text{ m}^{-1} \text{ s}^{-1}$
Gravitational acceleration	g	9.81 m/s ²
Reduced gravity	g'	0.0015 m/s ²
Reference density	ρ_0	1027 kg/m ³
Bottom friction coefficient	r	$2 \times 10^{-7} \text{ s}^{-1}$
Biharmonic dissipation coefficient	ν_4	10 ⁹ m ⁴ /s
Harmonic dissipation coefficient	ν_2	100 m ² /s

layers were first described by Treguier (1989), and are only problematic in the coarse-resolution models which are unable to resolve the boundary layers, resulting in numerical instability. The model topography is shown in Fig. 1(a); the model's idealized version of Drake Passage appears in the rightmost quarter of the figure.

The streamfunction ψ_i is defined by the solution to the elliptic potential vorticity inversion problem

$$q_1 = \nabla^2 \psi_1 + \frac{2f_0^2}{g'H}(\psi_2 - \psi_1) - \frac{2f_0^2}{g'H} \psi_1 \quad (3)$$

$$q_2 = \nabla^2 \psi_2 + \frac{2f_0^2}{g'H}(\psi_1 - \psi_2) \quad (4)$$

where g is the gravitational acceleration, and g' is the reduced gravity. Boundary conditions are required to uniquely specify the solution ψ_i ; they are presented in Appendix A.

The model equations are simulated at high resolution on an equispaced grid of 3072×385 points, which gives a grid spacing of $\Delta x = \Delta y = 6.51$ km, approximately half the baroclinic deformation radius $L_d = 12.25$ km. Note that the baroclinic deformation radius for this model is approximately

$$L_d \approx \frac{\sqrt{g'H}}{2|f_0|}. \quad (5)$$

The barotropic deformation radius \sqrt{gH}/f_0 is finite, but at 1981 km it is too large to have a significant impact on the dynamics.

The second-order spatial discretization is described by Nadeau and Straub (2009); the jacobian advection terms are discretized using the method of Arakawa (1966). The eddy-resolving model uses third-order Adams–Bashforth time integration with a step size of 15 min, and uses a multigrid method to solve the potential vorticity inversion.

2.2. Eddy-resolving results

The eddy-resolving simulation is run from rest to a statistically steady state, as measured by the equilibration of the total energy and area-averaged zonal baroclinic shear. The simulation takes approximately 150 years to equilibrate, and statistics are computed based on daily observations over a 10 year period. The total (barotropic) volume transport equilibrates to 200 ± 9 Sv ($1 \text{ Sv} = 10^6 \text{ m}^3/\text{s}$), where 9 Sv is the standard deviation of the temporal fluctuations. The mean zonal baroclinic shear, defined as the

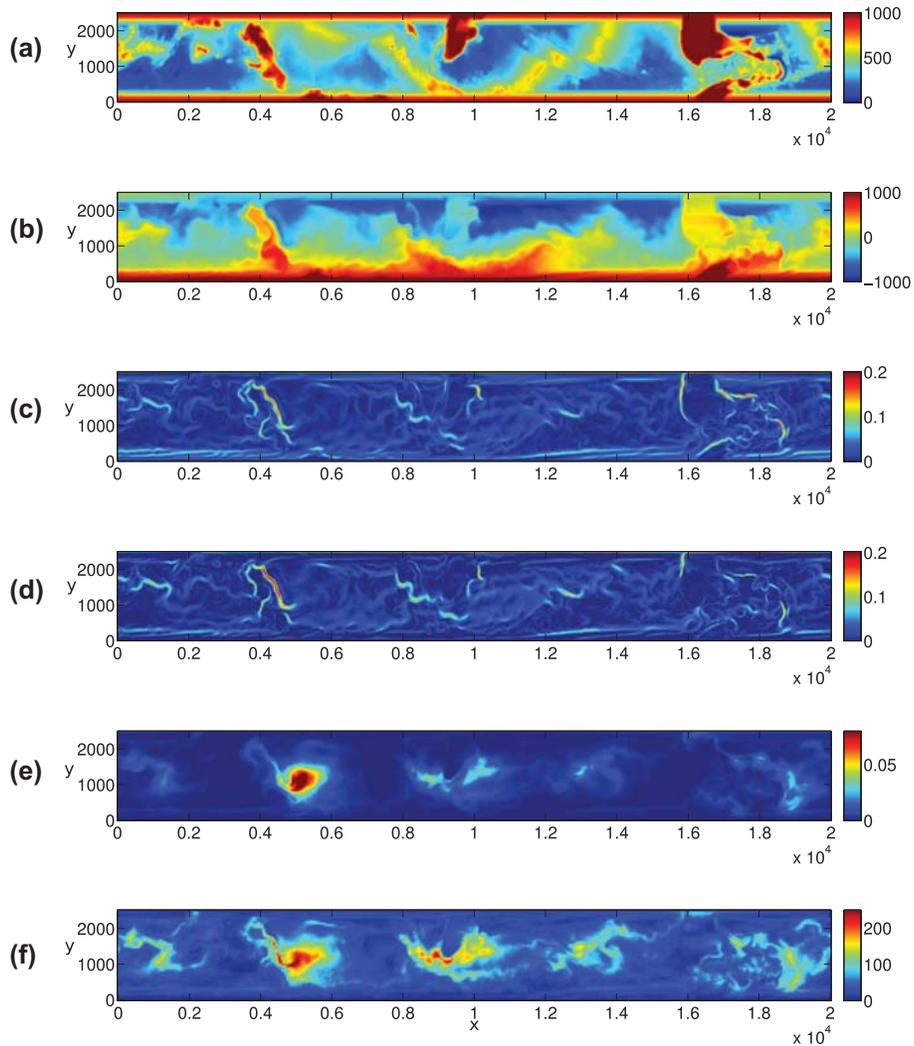


Fig. 1. Results from the eddy-resolving reference simulation. Axes are labeled in units of kilometers. Panels show: (a) topography h_b (meters), (b) mean interface height η (meters), (c) mean shear amplitude $|\mathbf{u}_1 - \mathbf{u}_2|$ (m/s), (d) growth rate of local baroclinic instability (days^{-1}), (e) eddy kinetic energy (m^2/s^2 ; color saturates at 0.08), and (f) standard deviation of coarse-grained interface height σ (meters; color saturates at 250).

area-averaged difference between the zonal velocity in the upper and lower layers, equilibrates to 0.70 ± 0.003 cm/s.

The height of the layer interface is $H/2 + \eta$ where $\eta = (f_0/g')(\psi_2 - \psi_1)$; note that the area average of η is zero because of the boundary conditions on PV inversion (Appendix A). The time average of η is shown in Fig. 1(b). It exhibits large variations up to ± 1 km, and has an RMS height of 473 m. The local mean shear ($|\mathbf{u}_1 - \mathbf{u}_2| \propto |\nabla\eta|$) shown in Fig. 1(c) reaches a maximum of approximately 20 cm/s, which implies that the slope of the layer interface remains mild despite the relatively large amplitude of η . The net baroclinic shear of 0.7 cm/s is somewhat small compared to the baroclinicity of the real ACC (see e.g. Firing et al., 2011). It could be increased by increasing the amplitude of the topography or by increasing the Ekman drag coefficient r , but the former would take the model even further from the regime of applicability of the quasigeostrophic approximation and the latter could result in unrealistic frictionally-dominated dynamics. The parameter choices here attempt to make the model as realistic as possible within the constraints of a two-layer QG channel.

Fig. 1(d) shows the growth rate of local baroclinic instability computed using the time-mean shear and PV gradient. Regions of strong growth rate largely coincide with regions of strong shear. Fig. 1(e) shows the eddy kinetic energy (EKE) based on a time average, i.e.

$$\text{EKE} = \frac{1}{2} \sum_i \overline{|\mathbf{u}_i|^2} - |\overline{\mathbf{u}_i}|^2 \quad (6)$$

where the overbar $\overline{(\cdot)}^t$ denotes a time average (the superscript t is used to distinguish the time average from the division onto large scales in Section 3.2). Regions of high EKE do not coincide with regions of strong shear and strong baroclinic instability; rather, regions of strong EKE tend to lie downstream of such regions. Significant movement of eddy energy has also been found by Grooms et al. (2013) in a QG basin model. This suggests that parameterizations that infer an eddy velocity scale from the local shear will be in error, and that it may be advantageous to develop parameterizations that account for large-scale movement of eddy energy (Eden and Greatbatch, 2008; Marshall and Adcroft, 2010; Grooms et al., 2012; Grooms and Majda, 2014). None of the parameterizations tested here incorporate such effects.

Fig. 1(f) shows the variability of the large-scale part of the interface height deviation η . Specifically, η is coarse-grained by cell-averaging from the eddy-resolving 3072×385 grid to the coarse-model 384×49 grid, and Fig. 1(f) shows the standard deviation of the time series of the coarse-grained η at each point on the 384×49 coarse-model grid. The temporal variability of the large-scale part of η is quite large, with standard deviations in excess of 200 m. This underscores the strong temporal variability of the large-scale part of η , which is particularly difficult for coarse models to reproduce. It also suggests that the time-mean EKE shown in Fig. 1(f) is not entirely due to subgrid scales.

3. Deterministic and stochastic parameterizations

3.1. Gent–McWilliams parameterizations

Mesoscale eddies and their important effects are not resolved by coarse models, and need to be parameterized. There are many approaches to parameterizing mesoscale eddies; we consider five variations on the popular GM parameterization (Gent and McWilliams, 1990), since the GM parameterization forms the basis of mesoscale eddy parameterizations in most IPCC-class climate models (Kuhlbrodt et al., 2012).

In the context of two-layer QG dynamics, the GM parameterization becomes extremely simple: subgridscale eddies smooth the

large-scale layer interface height. The interface height is $H/2 + (f_0/g')(\psi_2 - \psi_1)$, and the large-scale QG dynamics are modeled by

$$\partial_t q_1 + J[\psi_1, q_1 + \beta y] = \frac{2f_0^2}{g'H} \nabla \cdot (\kappa \nabla (\psi_2 - \psi_1)) - \frac{1}{\rho_0 H_1} \partial_y \tau^x(y) + v_2 \nabla^4 \psi_1 \quad (7)$$

$$\partial_t q_2 + J[\psi_2, q_2 + \beta y] = \frac{2f_0^2}{g'H} \nabla \cdot (\kappa \nabla (\psi_1 - \psi_2)) - \frac{f_0}{H_2} J[\psi_2, h_b] - r \nabla^2 \psi_2 + v_2 \nabla^4 \psi_2 \quad (8)$$

Note that the coarse model equations use a Laplacian vorticity diffusion instead of biharmonic, consistent with standard practice in coarse-resolution ocean modeling.

There are many ways to specify the quasi-Stokes diffusivity coefficient κ : we test five parameterizations of κ , based on schemes from the literature. The first is the simplest, namely

0. κ is a tunable constant, independent of space and time.

Since κ has dimensions of length squared over time, one might make a mixing-length approximation by setting $\kappa \propto VL$ for some velocity scale V . The simplest scheme of this form is

1. $\kappa = \alpha L_d \Delta U$

where $L_d = 12.25$ km is the deformation radius, $\Delta U = |\mathbf{u}_1 - \mathbf{u}_2|$ is the magnitude of the velocity jump across the interface, and α is a tunable constant. This scheme is loosely similar to a parameterization proposed by Stone (1972).

Scheme 1 generates a κ that is linear in ΔU , and one can construct schemes that are quadratic and cubic in ΔU . For the quadratic scheme we set

$$\mathbf{2.} \quad \kappa = \alpha \frac{(\Delta U)^2}{r}$$

In our setting with constant r this scheme is essentially just a quadratic in ΔU , but the dependence on r and quadratic dependence on ΔU are similar to the parameterization of Cessi (2008).

Held and Larichev (1996) proposed a parameterization for κ based on the phenomenology of doubly-periodic simulations of two-layer quasigeostrophic turbulence on a β -plane. Their method uses a length scale $L \propto (\beta T)^{-1}$ with $T = L_d/\Delta U$. The result is

$$\mathbf{3.} \quad \kappa = \alpha \frac{(\Delta U)^3}{\beta^2 L_d^3}$$

As suggested by Held and Larichev (1996), and as implemented in the coarse-resolution MOM3.0 ocean model (Pacanowski and Griffies, 2000), we replace β by an 'effective' value that incorporates the effect of topography

$$\beta_{\text{eff}} = \left| \beta \hat{\mathbf{y}} + \frac{f_0}{H} \nabla h_b \right| \quad (9)$$

Visbeck et al. (1997), inspired partly by Green (1970), proposed a scheme for κ based on the dimensional form $\kappa \propto L^2/T$ where the time scale T is given by $\sqrt{Ri}/|f_0|$. In the two-layer context we approximate the Richardson number $Ri = N^2/|\partial_z \mathbf{u}|^2$ using the two-layer approximation to the buoyancy frequency $N \approx \sqrt{2g'/H}$ and using $|\partial_z \mathbf{u}| \approx 2\Delta U/H$, which leads to

$$\frac{1}{T} = \frac{|f_0|}{\sqrt{Ri}} = \Delta U \sqrt{\frac{2f_0^2}{g'H}} \quad (10)$$

They proposed setting the length scale L to be the 'width of the baroclinic zone' L_z , defined as follows: In locations where $1/T$ is less than 10% of its maximum on the domain, L_z is set equal to the grid

scale, otherwise it equals the distance to the nearest point where $1/T$ is less than 10% of its maximum on the domain. We therefore consider the scheme

$$4. \kappa = \alpha \frac{L_z^2}{T}$$

where T is defined by (10). When computing T for this scheme, we set values larger than 1 day equal to 1 day, and when computing L_z values larger than 10 coarse grid points are truncated; this is qualitatively consistent with implementations in ocean general circulation models (e.g. MOM3.0: Pacanowski and Griffies, 2000). When computing distances L_z we use the standard Euclidean distance rather than the more complicated algorithm in MOM3.0.

These five schemes constitute a representative sample of existing parameterizations of the diffusivity κ ; we refer to them as GM0–GM4. The first four are essentially monomials in the baroclinic shear ΔU , and the numbering scheme is chosen to reflect the exponent so that e.g. GM0 corresponds to $\kappa \propto (\Delta U)^0$ and GM2 corresponds to $\kappa \propto (\Delta U)^2$. The GM3 scheme is based on the Held and Larichev (1996) scheme, and is the only scheme to incorporate the effect of topography, albeit in an ad hoc manner. The GM4 scheme is essentially the parameterization of Visbeck et al. (1997) applied in the two-layer QG setting. For all schemes we implement cutoffs so that values of κ less than 50 m²/s are set to 50 m²/s, and values greater than 2000 m²/s are set to 2000 m²/s; this is qualitatively consistent with typical implementations in ocean general circulation models.

3.2. Stochastic superparameterization

Stochastic superparameterization (SP) has been formulated for two-layer QG dynamics and tested in eddy-permitting scenarios by Grooms and Majda (2013, 2014). This section reviews the formulation of stochastic SP and discusses its properties.

3.2.1. Coarse grid model

The coarse model equations in stochastic SP take the form

$$\partial_t q_1 + J[\psi_1, q_1 + \beta y] = -\nabla \cdot (\overline{\mathbf{u}'_1 q'_1}) - \frac{2}{\rho_0 H} \partial_y \tau(y) + \nu_2 \nabla^4 \psi_1 \quad (11)$$

$$\begin{aligned} \partial_t q_2 + J[\psi_2, q_2 + \beta y] = & -\nabla \cdot (\overline{\mathbf{u}'_2 q'_2}) - \frac{2f_0}{H} J[\psi_2, h_b] - r \nabla^2 \psi_2 \\ & + \nu_2 \nabla^4 \psi_2 \end{aligned} \quad (12)$$

where the eddy PV flux divergence is further decomposed as

$$\begin{aligned} \nabla \cdot (\overline{\mathbf{u}'_i q'_i}) = & (-1)^i \frac{2f_0}{g'H} \nabla \cdot (\overline{\mathbf{u}'_i (\psi'_2 - \psi'_1)}) + (\partial_x^2 - \partial_y^2) \overline{u'_i v'_i} \\ & + \partial_{xy} (\overline{(v'_i)^2} - \overline{(u'_i)^2}) \end{aligned} \quad (13)$$

The overbar $\overline{(\cdot)}$ here denotes a formal projection onto large scales, and the prime $'$ denotes small-scale eddy variables. This is an exact re-writing of the standard eddy PV flux divergence that is obtained from a Reynolds decomposition, separating it into the flux of that part of PV corresponding to the interface height (e.g. $\overline{\mathbf{u}'_1 (\psi'_2 - \psi'_1)}$) and the curl of the divergence of the Reynolds stress (Marshall et al., 2012; Grooms and Majda, 2013; Grooms and Majda, 2014). The former component is conceptually related to heat flux, and will be referred to as such hereafter despite the loose physical connection. The GM parameterizations of the previous section can be written in this form by setting the Reynolds stress terms to zero ($\overline{u'_i v'_i} = 0$, $\overline{(u'_i)^2} = \overline{(v'_i)^2}$) and making the heat flux downgradient $\overline{\mathbf{u}'_1 (\psi'_2 - \psi'_1)} = -\kappa \nabla (\psi_2 - \psi_1)$. The importance of the Reynolds

stresses in addition to the heat flux was emphasized in the asymptotic analysis of Grooms et al. (2012).

3.2.2. Stochastic subgrid-scale model

Stochastic SP obtains the components of the eddy PV flux divergence from local models of the eddy dynamics. Specifically, the unresolved eddies at each location on the coarse-grid domain are modeled by QG dynamics on local, pseudo-physical subdomains embedded into the coarse grid; the dynamics on each subdomain is independent of the dynamics on all the other subdomains. This multiscale strategy (SP) has been used to parameterize ocean mixed-layer convection by Campin et al. (2011) and a variety of atmospheric phenomena (reviewed by Majda and Grooms, 2014). The SP approach is also related to the finding by Pavan and Held (1996) that the PV flux diagnosed from idealized channel simulations can sometimes be well approximated by the PV flux generated by doubly-periodic QG simulations.

A major difference between stochastic SP and conventional SP is that stochastic SP does not make use of simulations of nonlinear eddy dynamics on local, horizontally periodic subdomains. Instead, the unresolved eddies at each point of the physical domain are modeled as homogeneous random functions on infinite subdomains, obeying quasi-linear, stochastic QG equations where damping and Gaussian additive forcing replace the nonlinear advection terms. The eddy equations are of the form

$$\partial_t q'_1 = F'_1 - \Gamma q'_1 - u_1 \cdot \nabla q'_1 - u'_1 \cdot \nabla (q_1 + \beta y) - \nu_4 \nabla^6 \psi'_1 \quad (14)$$

$$\begin{aligned} \partial_t q'_2 = & F'_2 - \Gamma q'_2 - u_2 \cdot \nabla q'_2 - u'_2 \cdot \nabla \left(q_2 + \beta y + \frac{2f_0}{H} h_b \right) \\ & - r \nabla^2 \psi'_2 - \nu_4 \nabla^6 \psi'_2. \end{aligned} \quad (15)$$

Eddy variables are denoted by $'$ to distinguish them from coarse-model variables. The terms that replace the eddy-eddy nonlinear advection are F'_i , a spatially-correlated Gaussian white noise, and Γ , a positive-definite pseudodifferential operator. Similar stochastic models of quasigeostrophic turbulence are discussed by DelSole (2004) and Srinivasan and Young (2012), though not in a multiscale setting. The PV inversion (Eqs. (3) and (4)) is the same as for the coarse-grid equations, except that the barotropic deformation radius is approximated as infinite in the eddy equations for convenience (infinite barotropic deformation radius corresponds to the limit $g \rightarrow \infty$).

Although the equations are written using the same coordinates (t , x , and y) as the coarse-grid equations, each coarse-grid location actually has its own eddy microcosm where the eddies evolve on distinct space and time coordinates (Majda and Grooms, 2014). The components of the eddy PV flux are defined as space and time averages over the eddy subdomains, and because the eddies within a single subdomain are homogeneous the spatial average is equivalent to the statistical average. From the perspective of the eddy dynamics the coarse-grid variables are constant. The eddy equations are then linear in the eddy variables and have constant coefficients, so their solution can be written down in closed form using Fourier analysis (for details see Grooms and Majda, 2013, 2014). A key assumption here is that the topography h_b has no small-scale component. Topographic variation on the small scales would require additional modeling beyond the scope of the present investigation. We have chosen to use a smoothed large-scale topography, as described in Section 2.1, in order to be consistent with this assumption.

The multiscale formulation is most appropriate for situations where there is a scale separation between the coarse-grid and sub-grid scale dynamics. In such situations, multiple-scales asymptotics can be used to motivate multiscale equations coupling the large- and small-scale dynamics (e.g. Majda, 2007a,b; Grooms

et al., 2011; Grooms et al., 2012; Malecha et al., 2014; Dolaptchiev and Klein, 2013). But the scale separation between resolved and unresolved scales is often small, especially in turbulent systems with a wide range of active scales, and stochastic SP incorporates several features to minimize the severity of the scale separation imposed by the multiscale SP approach. These are discussed further below; for example, the range of active scales on the eddy domains begins seamlessly at the coarse grid scale, and the eddy dynamics are only allowed to evolve for a short time while the coarse-grid variables are held fixed.

Stochastic SP algorithm. As an algorithm, stochastic SP proceeds as follows.

1. At the beginning of each coarse-model time step, the coarse-model velocity \mathbf{u}_i and PV gradient $\nabla(q_i + \beta y + \text{topography})$ are computed at each coarse grid point for use in the eddy Eqs. (14) and (15).
2. The eddies at each coarse grid point evolve according to Eqs. (14) and (15) with the large-scale velocity and PV gradient held constant.
3. The feedback to the coarse-model equations, Eq. (13), is computed from the time- and area-mean of the eddies on each subdomain.
4. The coarse-model variables are advanced, the eddies are re-set to their initial condition, and the process repeats.

To implement this algorithm one needs an initial condition for the eddies, a model for F'_i and Γ , and a timescale over which the eddies respond to the fixed local large-scale conditions. The eddy initial condition and the stochastic model of the eddy nonlinearity (F'_i and Γ) are discussed below in this section. The length of time over which the eddies are averaged is essentially a tunable constant; the effect of this parameter is described in Section 3.2.3 and a specific choice is motivated in Section 4.1.

Initial condition. Rather than track the state of the eddies from one coarse-grid time step to the next, the eddies are re-initialized to a zero-mean Gaussian distribution at the beginning of each coarse-grid time step. The initial condition for the eddies is specified by its statistical properties, which are as follows.

1. The eddies have horizontal variation in a single, randomly-chosen direction θ (measured north from east).
2. The eddy energy spectrum is proportional to $k^{-5/3}$ for $k < L_d^{-1}$ and to k^{-3} for $k \geq L_d^{-1}$.
3. The barotropic energy is 6 times larger than the baroclinic energy at each wavenumber.
4. The upper layer has twice as much kinetic energy as the lower layer.
5. There is no heat flux (e.g. $\overline{\mathbf{u}'_1(\psi'_2 - \psi'_1)} = 0$).

These properties are the same as used by Grooms and Majda (2013, 2014), with exceptions noted below. They completely specify the second-order statistics (hence the complete Gaussian statistics) of the eddy initial condition.

The first property makes the components of the eddy PV flux stochastic, rather than being deterministic functions of the mean as in the GM parameterizations of Section 3.1. It is similar to the common practice in conventional SP of using two-dimensional small-scale domains (see, e.g. Campin et al., 2011; Majda and Grooms, 2014), the main difference here being that the orientation of the domains is random. The motivation for this choice of random directions is that the eddies that might be observed in a local patch of ocean have temporally-varying directions of anisotropy due to the random sampling of a few vortices and filaments; we attempt to account for this small-scale randomness by choosing a field of random angles θ . This field has a uniform distribution and no

spatial or temporal correlation; the addition of correlation might improve the performance but is beyond the scope of the current investigation.

The second property, the shape of the energy spectrum, is familiar from the theory of quasigeostrophic turbulence (Charney, 1971). The nondimensional proportionality constant (related to the implied local subgrid-scale energy) is denoted A , and is left as a tunable constant, similar to α in the GM parameterizations of Section 3.1. The eddies have zero energy at scales smaller than the Nyquist wavenumber of the eddy-resolving reference simulation $\pi/\Delta x$, and at scales larger than the coarse-grid Nyquist wavenumber, which in our coarse-model simulations is $8\pi/\Delta x$ (see Section 4.1); the range of subgrid-scales thus picks up seamlessly at the coarse grid scale. The third property deviates somewhat from previous work, where baroclinic energy was set equal to barotropic energy (Grooms and Majda, 2014), but is more in accord with periodic simulations of QG turbulence (Larichev and Held, 1995). Increasing the proportion of barotropic energy increases the amplitude of the Reynolds stress terms relative to the amplitude of the eddy heat flux terms. The fourth property is in reasonable accord with previous well-resolved simulations of QG dynamics (Grooms and Majda, 2014), and the fifth implies that the eddies will not generate a net heat flux in the absence of a large-scale gradient. The ratio of barotropic and baroclinic eddy energy and the ratio of upper and lower layer eddy kinetic energy (third and fourth properties) should presumably vary over the physical domain in response to topography, large-scale shear, etc. but modeling this variation is beyond the scope of the present investigation. No attempt has been made to tune the eddy initial condition in comparison with the high-resolution reference simulation, the goal being to demonstrate that good results can be achieved using simple, reasonable choices.

Stochastic model of eddy nonlinearity. The structure of the damping operator Γ is set so that it damps a wavenumber k using a timescale determined by dimensional considerations from the eddy energy spectrum. The Fourier symbol for $\Gamma q'_i$ is $\gamma_k q'_i$, and we set $\gamma_k = \gamma_0 \sqrt{k^3 E(k)/A}$ and where $E(k)$ is the one-dimensional energy spectrum described in property 2 above and γ_0 is a nondimensional tunable constant. For example, for $k \geq L_d^{-1}$ the damping rate is $\gamma_k = \gamma_0$ because $E(k) = Ak^{-3}$. The damping rate γ_k is related to the decorrelation time scale for a Fourier mode; in the absence of large-scale and viscous effects the decorrelation time is $1/\gamma_k$. The damping operator Γ is thus specified up to a tunable constant γ_0 ; the effect of this parameter is described in Section 3.2.3, and a specific choice is motivated in Section 4.1.

We choose to set the forcing terms F'_i , which are spatially correlated and white in time, so that the eddy initial condition described above is the statistical equilibrium solution of the eddy equations in the absence of large-scale and viscous effects. I.e. the eddy initial condition is the equilibrium statistical solution of the following equations

$$\partial_t q'_1 = F'_1 - \Gamma q'_1 \quad (16)$$

$$\partial_t q'_2 = F'_2 - \Gamma q'_2 \quad (17)$$

The forcing terms are thus set implicitly by the properties of the damping Γ and of the initial condition discussed above.

In summary, the eddies evolve under quasi-linear (large-scale variables are considered constants) stochastic QG dynamics on local subdomains embedded in the computational coarse grid. At the beginning of each coarse-grid time step the eddies are re-initialized, and a direction for the eddies is randomly chosen. There are three tunable parameters: the eddy amplitude A , the eddy damping rate γ_0 , and the length of time over which the eddies

are allowed to evolve while the coarse-grid variables are held fixed, denoted ϵ^{-1} . The effects of these parameters and some qualitative properties of stochastic SP are discussed in the next subsection.

3.2.3. Some qualitative properties of stochastic SP

Consider the response of the stochastic eddy model to the imposition of a zonal baroclinic shear, ignoring the effects of topography and β , and let the eddies be functions only of the horizontal coordinate $\xi = x \cos(\theta) + y \sin(\theta)$. The eddy equations take the form

$$\partial_t q'_1 = F'_1 - \Gamma q'_1 - U \cos(\theta) \partial_\xi q'_1 - \frac{4U \cos(\theta) f_0^2}{g'H} \partial_\xi \psi'_1 - v_4 \partial_\xi^6 \psi'_1 \quad (18)$$

$$\partial_t q'_2 = F'_2 - \Gamma q'_2 + U \cos(\theta) \partial_\xi q'_2 + \frac{4U \cos(\theta) f_0^2}{g'H} \partial_\xi \psi'_2 - r \partial_\xi^2 \psi'_2 - v \partial_\xi^6 \psi'_2 \quad (19)$$

The equations are linear with an additional damping and forcing, and as usual for linear dynamics there will be some critical value of $U \cos(\theta)$ dependent on the damping. For $U \cos(\theta)$ larger than the critical value the solutions will grow exponentially without saturation; for smaller values the solutions will decay to some nonzero equilibrium due to the additional forcing term. In neither case is the long-term behavior physically meaningful because it is unrealistic to keep the large-scale shear fixed for an infinite amount of time. The range of stable shear is largely controlled by the constant γ_0 : for large γ_0 the eddy dynamics will be linearly stable for a wide range of shear.

The components of the eddy PV flux are computed from the behavior of the eddy equations over a fixed amount of time of length ϵ^{-1} . Since a spatial average is equivalent to an integral over the wavenumbers, all wavenumbers contribute to the eddy PV flux. However, for sufficiently long averaging times the most unstable modes will dominate, which makes it similar to parameterizations that compute the subgridscale flux based on the most unstable linear modes (e.g. Killworth, 1997; Eden, 2011, 2012), a major difference being that in stochastic SP all scales in a randomly-chosen direction are allowed to contribute, not just the most unstable mode.

Consider the behavior of the eddy heat flux $\overline{u'_1(\psi'_2 - \psi'_1)}$. Since the velocity is orthogonal to the gradient of the streamfunction, the eddies can only flux orthogonal to θ : $u'_1 = -\partial_y \psi'_1 = -\sin(\theta) \partial_\xi \psi'_1$ and $v'_1 = \partial_x \psi'_1 = \cos(\theta) \partial_\xi \psi'_1$. Since θ is sampled from a uniform distribution the eddy flux almost always has a component across the large-scale gradient. This is a marked contrast with the GM parameterizations described above, which model the eddy flux as being always purely downgradient: $-\kappa \nabla(\psi_2 - \psi_1)$. (Note that it is possible to construct anisotropic GM parameterizations that flux across the mean gradient; see Smith and Gent, 2004).

It is possible to diagnose an 'effective' κ from stochastic SP by considering the heat flux generated by the stochastic eddy model in response to a local mean shear. Although, as noted above, stochastic SP generates fluxes across the mean gradient, we define the effective κ for the case where θ is parallel to the shear (orthogonal to the gradient $\nabla(\psi_2 - \psi_1)$). In this case the stochastic SP model generates no cross-gradient flux. The effective κ is given by

$$\kappa_{\text{eff}} = -\frac{\overline{v'_1(\psi'_2 - \psi'_1)}}{2U} \quad (20)$$

It is a nonlinear function of the local shear $2U$, which equals ΔU in the notation of the foregoing section. The eddy damping constant γ_0 (which controls the amplitude of Γ) sets the range of ΔU that are linearly stable. For values of the shear that are linearly stable the effective κ is small; as the shear increases the eddies become unstable and the effective κ rises sharply. The sharpness of the rise is

controlled by ϵ^{-1} , the length of time over which the eddies are allowed to respond to the fixed shear: the rise is sharper for longer averaging times.

These properties are evident in Fig. 2, which shows the effective κ generated by stochastic SP as a function of ΔU for two averaging times: short, $\epsilon^{-1} = 14.2$ days, and long, $\epsilon^{-1} = 70.9$ days. The value of γ_0 for the short averaging time is $\epsilon/2$, and for the long time is $6\epsilon/7$. These particular choices are discussed further in the next subsection. Fig. 2 also shows κ for the first four GM parameterizations in Section 3.1 for comparison; the GM3 parameterization uses $\beta = 1.5 \times 10^{-11} \text{ s}^{-1}$ and no topography. For the short averaging time, and over the range of shear shown, the effective κ is approximately linear, similar to the GM1 scheme. For the long averaging time the effective κ increases sharply with shear, similar to the GM3 scheme. The onset of the sharp rise could be moved to larger shear by increasing γ_0 , leading to a wider range of linearly stable shear. Note that stochastic SP does not produce negative κ in this setting, although stochastic SP can generate negative κ in the presence of β and topographic slopes. The cutoff shown in the behavior of the long-average SP in Fig. 2 is similar to the upper cutoff of $2000 \text{ m}^2/\text{s}$ imposed on the GM schemes: values of ΔU greater than 5.41 cm/s are scaled back to 5.41 cm/s before being used to calculate the eddy PV flux (discussed more below). Because stochastic SP does not generate κ directly it is not possible to configure it to have a cutoff of exactly $2000 \text{ m}^2/\text{s}$ like the GM schemes.

Stochastic SP is set up so that it produces a heat flux only in response to large-scale shear or PV gradients; in the absence of these the eddies remain at their stable initial condition, which by construction generates no heat flux. In contrast, stochastic SP generates nonzero Reynolds stresses regardless of the presence or absence of large-scale shear; large-scale conditions only alter the character of the Reynolds stress. The eddy initial condition is constructed such that the Reynolds stresses average to zero when averaged over θ ; this is because an isotropic spectrum by definition has $\overline{(u')^2} = \overline{(v')^2}$ and $\overline{u'v'} = 0$. But for a single value of θ the Reynolds stress terms are quite large. For the eddy initial condition the terms $\overline{(v')^2} - \overline{(u')^2}$ and $\overline{u'_1 v'_1}$, which appear in the mean Eq. (13), are proportional to $\cos(2\theta)$ and $\sin(2\theta)$, respectively. These are

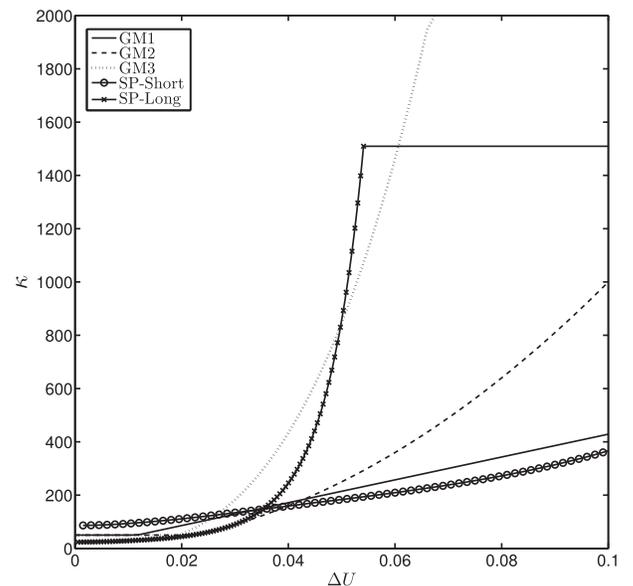


Fig. 2. Quasi-Stokes diffusivity κ and effective diffusivity (20) as functions of the baroclinic shear ΔU for GM1 (solid), GM2 (dashed), GM3 (dots), SP-short (circles), and SP-long (x). Values of α for the GM models are the optimally tuned values listed in Section 4.1; values of A for stochastic SP are the optimally tuned-values for SP-short and SP-long listed in Section 4.1. Units of ΔU are m/s, and units of κ are m^2/s .

non-Gaussian random fields with no spatial correlations, and their second derivative appears in the coarse-grid Eq. (13). These derivatives magnify the small scale features, so the effect of the Reynolds stress terms in stochastic SP is qualitatively similar to adding small-scale random forcing, and corresponds to kinetic energy backscatter. Jansen and Held (2014) used an uncorrelated (white) random forcing to simulate backscatter in an eddy-permitting QG model; the 1D spectrum of such a forcing is proportional to the wavenumber amplitude k . For comparison, the Reynolds stress terms in the mean Eq. (13) have a 1D spectrum approximately proportional to k^5 , which is more focused on scales near the coarse grid scale. A proof of the k^5 forcing spectrum is somewhat involved and will appear elsewhere.

3.2.4. Configuration of stochastic SP

There are three remaining free parameters in the configuration of the stochastic SP eddy model: the eddy amplitude A , the damping rate γ_0 , and the averaging time ϵ^{-1} . The order of magnitude for the damping rate and averaging time are set by comparison with the growth rate of local baroclinic instability induced by the time- and area-mean baroclinic shear for the channel (0.7 cm/s), which is approximately $1/141 \text{ days}^{-1}$. We consider two averaging times: short, with $\epsilon^{-1} = 14.2 \text{ days}$, and long, with $\epsilon^{-1} = 70.9 \text{ days}$. Both of these are short in comparison with 141 days, but they are long in comparison with the local growth rate of baroclinic instability in regions of larger shear. The damping rate γ_0 is set to $1/28.2 \text{ days}^{-1}$ for the short-average configuration, and to $1/82.7 \text{ days}^{-1}$ for the long-average configuration. The short-average configuration is more strongly damped, and corresponds to having eddy dynamics less strongly influenced by the large scales, whereas the long-average configuration is weakly damped, and strongly influenced by the large scales. The short-average configuration also has a shorter eddy decorrelation time, since the eddy decorrelation time is approximately one over the damping rate (the precise decorrelation time depends on the viscous, Ekman, and large-scale effects).

Comparing only two combinations of γ_0 and ϵ is clearly insufficient to determine the full sensitivity of the results to these parameters. These combinations mainly show (i) the differences between configurations strongly and weakly influenced by the large scales and, (ii) that reasonable results are obtained by setting the order of magnitude of γ_0 and ϵ in comparison with the time scale of baroclinic instability. The amplitude A is left as the primary tunable control parameter.

For both the long- and short-average configurations we test four different configurations of stochastic SP corresponding to different configurations of the coarse-to-eddy coupling and the eddy-to-coarse coupling. The minimal configuration does not include the Reynolds stress feedback terms in the coarse-grid equations, and does not include the contributions of β , topography, or coarse-grid relative vorticity in the eddy Eqs. (14) and (15). This configuration is denoted ‘NRS- f ’ for ‘no Reynolds stress, f -plane’ since the eddy dynamics are effectively on a flat-bottomed f -plane. The next configuration includes the full coarse-grid PV gradient in the eddy equations, but leaves the Reynolds stress terms out of the coarse-grid equations; this configuration is denoted ‘NRS- β ’. The remaining two configurations are the same as the first two, with the addition of Reynolds stresses to the coarse-grid equations; they are denoted ‘RS- f ’ and ‘RS- β ’.

The configuration RS- f is related to the asymptotic analysis of Grooms et al. (2012) where the eddies generate significant Reynolds stresses and do not feel β (in fact, this is among the most accurate configurations; see results in the next section), and the configuration NRS- β is related to the asymptotic analysis of Grooms et al. (2011), where the Reynolds stresses do not affect

the large-scale flow and the eddies do feel β . Testing these four configurations allows us to consider the impacts of the Reynolds stress terms in the coarse-grid equations separately from the impacts of the large-scale PV gradients on the eddy model.

In summary, we test eight configurations of stochastic SP. Four configurations use a long eddy average, and four use a short eddy average. Each set of four has two configurations where the Reynolds stress terms are included in the mean equations (RS- f and RS- β), and two where they are not included (NRS- f and NRS- β). Each set of four also has two configurations where the eddies respond only to the coarse-grid baroclinic shear (NRS- f and RS- f), and two where the eddies respond to the shear and to the full coarse-grid PV gradient (NRS- β and RS- β). We refer to these configurations as, e.g., SP-short-RS- f or SP-long-NRS- β .

As in Grooms and Majda (2013, 2014), the eddy terms are not computed on the fly during the coarse-grid simulation. Instead, they are computed over a grid of values of large-scale shear and PV gradients, and then interpolated from these pre-computed values during the coarse-grid simulation. The range of shear and vorticity gradients used to pre-compute the eddy terms is drawn from the time-mean structure of the eddy-resolving simulation. This is not fundamental to the approach, but results in increased computational efficiency in this two-layer setting.

The local mean baroclinic shear in the reference simulation reaches a maximum amplitude of 20 cm/s, almost a factor of 30 larger than the mean shear. To avoid huge eddy growth in response to these local regions of strong shear in the long-average configuration the coarse-grid shear is scaled down to a maximum amplitude of 5.41 cm/s before being used to evaluate the eddy terms. This is similar to truncating the value of κ in the deterministic parameterizations, as discussed in Section 3.1. The threshold of 5.41 cm/s has been chosen to make the cutoff in the effective κ reasonably close to $2000 \text{ m}^2/\text{s}$ as shown in Fig. 2; getting an exact match is difficult because changing the threshold requires re-tuning A , which changes the value of the effective κ .

Similarly, because of the complex nature of the topography, the topographic PV gradient in some regions can be up to 20 times larger than β . The eddy dynamics are highly unphysical for values this large, essentially because scale separation between the topography and eddies breaks down. To reduce errors due to large PV gradients, the component of the coarse-grid PV gradient associated with topography, β , and relative vorticity is scaled back to a maximum amplitude of $3 \times 10^{-11} (\text{ms})^{-1}$ before being used to evaluate the eddy terms; we also tested truncations up to $10^{-10} (\text{ms})^{-1}$. The effect of increasing the truncation level is qualitatively similar to the effect of moving from the ‘ f -plane’ configurations to the β -plane configurations, as discussed in Section 4: the accuracy of the time-mean deteriorates but the accuracy of the variability improves slightly.

4. Coarse model results

4.1. Model configuration

The non-eddy-resolving, parameterized coarse models are run on a grid 8 times coarser than the eddy-resolving model, i.e. 384×49 points with a grid spacing of 52 km. Although this might appear to be a nominally eddy-permitting grid scale, it is ‘coarse’ compared to the small deformation radius of 12.25 km, which is appropriate to an ACC model. It is also similar to the longitudinal resolution of a 1° model at 60° from the equator. The coarse grid scale has also been chosen to be relatively small so that the complex topography can be relatively well resolved.

For reasons of coding convenience, the numerical methods used in the eddy-resolving and coarse-model simulations are slightly

different: the coarse models use third-order Runge–Kutta with a step size of 3 h, and the PV inversion is solved using a direct method (Davis, 2004, UMFPAK;). The boundary conditions for PV inversion are different but mathematically equivalent to those used in the eddy-resolving model, as described in Appendix A. In all other respects the numerics are identical. As noted above, the coarse models use Laplacian instead of biharmonic vorticity diffusion, and the coefficient is fixed for all models at the relatively low value of 100 m²/s; larger values cause the barotropic transport to grow and smaller values lead to grid-scale ringing around regions of sharp topography.

All the parameterizations involve one primary tunable parameter: κ in GM0, α in GM1–4, and A in stochastic SP. In every case these were tuned such that the time- and area-mean shear was 0.7 cm/s to a tolerance of ± 0.005 cm/s; optimal values of the tunable parameters for each scheme are listed below. The coarse models are then compared based on their representation of the mean and temporal variability of the coarse-grained interface height displacement η as discussed in the following subsection.

The optimal values of the tuned coefficients for the GM parameterizations are as follows: $\kappa = 155$ m²/s for GM0, $\alpha = 0.35$ for GM1, $\alpha = 0.02$ for GM2, $\alpha = 0.0028$ for GM3, and $\alpha = 0.00605$ for GM4. The optimal values of A for stochastic SP are as follows: 6.55×10^{-6} for SP-short-NRS- f , 5.8×10^{-6} for SP-short-RS- f , 5.4×10^{-6} for SP-short-NRS- β , 4.8×10^{-6} for SP-short-RS- β , 3.6×10^{-7} for SP-long-NRS- f , 3.2×10^{-7} for SP-long-RS- f , 3.7×10^{-7} for SP-long-NRS- β , and 3.35×10^{-7} for SP-long-RS- β .

4.2. Measures of skill

The accuracy of the time-mean of η is evaluated using the RMS error, defined as the square root of the area average of the squared error in η . The pattern correlation is not a useful metric in this case since all the models have time-mean η with very good pattern correlation. The accuracy of the interface height variability is measured using the area average, the pattern correlation, and the RMS errors. Specifically, let $\sigma^2(x, y) = \overline{\eta^{2t}} - (\overline{\eta^t})^2$ denote the variance of the coarse-grained interface height displacement from the eddy-resolving simulation, and let σ_M^2 denote the variance of the interface height displacement from a coarse model where $\overline{(\cdot)^t}$ denotes the time average. The most basic measure of accuracy in the variability is the root-mean-square value of σ , i.e.

$$\text{RMS } \sigma = (L_x L_y)^{-1} \left(\iint \sigma^2 dx dy \right)^{1/2}. \quad (21)$$

The RMS value of σ for the reference simulation is 58 m; for comparison the RMS value of $\overline{\eta^t}$ is 473 m. The pattern correlation is defined as

$$\text{PC} = \frac{\iint \sigma^2 \sigma_M^2 dx dy}{(\iint \sigma^4 dx dy \iint \sigma_M^4 dx dy)^{1/2}} \quad (22)$$

where the integrals are carried out over the whole domain; it is a nondimensional number between 0 and 1. The RMS error for variability is measured in terms of the error in the standard deviation σ rather than the variance σ^2 , and thus has units of length.

It is important for a coarse model to accurately portray both the time-mean and the variability of the large-scale part of the true signal; this is known as ‘climate fidelity’. Climate fidelity is a prerequisite for accurate predictions of climate change, i.e. of the response of the system to changes in forcing. This link was suggested by Shukla et al. (2006), and has been put on a firm mathematical footing by Majda and Gershgorin (2010, 2011a,b) where other sufficient conditions for forecast skill are also explored. Climate fidelity can be measured using the relative entropy, also known as the Kullback–Leibler divergence, between the statistics

of the true system and of the coarse model. The relative entropy is defined as

$$\text{Rel. Ent.} = \int p \ln \left(\frac{p}{q} \right) \quad (23)$$

where p is the true probability distribution and q is the distribution associated with the coarse model. Relative entropy has also been used in predictability studies, for example by Kleeman (2002). It has the useful properties that it is always positive, and zero only if $p = q$, and that it is invariant under general nonlinear changes of variable (Majda et al., 2002).

It is difficult to compute the relative entropy between two high-dimensional probability distributions, partly because those distributions are often not well known. An alternative is to measure the relative entropy between the least-biased Gaussian approximations to p and q , which are simply the Gaussian distributions with the same mean and covariance as p and q (Majda and Wang, 2006). In the current setting this would still be prohibitive, since it would require estimating the full covariance matrix of η on the coarse grid, which involves estimating the correlations of η at different spatial locations. DelSole and Shukla (2010) used relative entropy to compare general circulation models by comparing the area-averaged relative entropy of the Gaussian approximations to the single-point statistics. The relative entropy between the Gaussian approximations of the true signal and the coarse model at a single point is given by

$$\text{Gaussian Rel. Ent.} = \frac{1}{2} \left(\frac{(\overline{\eta^t} - \overline{\eta_M^t})^2}{\sigma_M^2} + \frac{\sigma^2}{\sigma_M^2} - 1 - \ln \left(\frac{\sigma^2}{\sigma_M^2} \right) \right) \quad (24)$$

where $\overline{\eta_M^t}$ denotes the coarse-model’s time-mean η .

The Gaussian approximation of the relative entropy has two components: the term involving the mean is called the *signal* and the remaining terms are the *dispersion*. The signal measures the error in the mean, weighted by the coarse model’s variance; it can be made small either by making the error in the mean small, or by making the model variability large. The dispersion term measures the error in the coarse model variability; it is zero only if $\sigma_M = \sigma$, and it goes to infinity in both the limits $\sigma_M \rightarrow 0$ and $\sigma_M \rightarrow \infty$. If the mean is considered fixed then the relative entropy is minimized for a model with variance $\sigma_M^2 = (\overline{\eta^t} - \overline{\eta_M^t})^2 + \sigma^2$, which is larger than the true variance. This is a result of the identification of temporal variability with uncertainty when using the relative entropy to compare climatological statistics: an ‘optimal’ model should have an uncertainty estimate that encompasses the bias in the mean. This suggests that the relative entropy can sometimes be improved by inflating the model variability well beyond what is realistic. However, in our results this is never the case since all of the models have too little variability overall: for example, the model variance in the GM4 model, which has the smallest relative entropy of the GM models, is larger than the true variance in only 7% of the domain.

The relative entropy, and in particular its Gaussian approximation, is a useful way to combine estimates of the accuracy of the mean and of the variability into a single number. In the following we compute the area-averaged Gaussian approximation to the relative entropy in the single-point statistics of η as a way to compare the accuracy of different models.

The barotropic transport is not a useful measure of skill for the coarse models. As detailed in Appendix B, the lower-layer transport is set by the difference between domain-integrated wind stress and topographic form stress. There is significant cancellation between these two large numbers, making the lower-layer transport extremely sensitive to the resolution of topographic form stress in the coarse models. All of the coarse models, both stochastic SP and GM, have barotropic transport between 240 and 245 Sv, and lower-layer transport near 104 Sv. In Appendix B we show that

the coarse models represent the form stress to within 98% accuracy, but the error is enough to change the net transport by nearly 25% due to the large cancellation with wind stress input.

This difficulty motivated the use of equal layers: while it might be more realistic to use a thicker lower layer, doing so magnifies the errors in the coarse-model transport. Similarly, this difficulty partly motivated the use of relatively fine (though still non-eddy) resolution: the barotropic transport degrades even further at coarser resolutions making the model dynamics less and less accurate.

4.3. Deterministic parameterizations

The domain-averaged measures of skill for the deterministic parameterizations are reported in Table 2. The first column shows the RMS error in the time-mean interface height; there is a clear trend of improved performance as κ becomes a steeper function of the local shear. For example, the GM0 scheme has $\kappa \propto (\Delta U)^0$ and has the largest RMS error for the time mean, whereas the GM3 scheme has $\kappa \propto (\Delta U)^3$ and has the smallest RMS error for the time mean.

The second through fifth columns show different measures of skill for the coarse-model variability: the RMS of σ and the RMS errors in σ , the pattern correlation of σ^2 , and the area-averaged dispersion component of the relative entropy, respectively. In contrast to the behavior of the mean, there is a clear trend of decreased accuracy in representing the variability as κ becomes a steeper function of the local shear. In the GM4 scheme, based on Visbeck et al. (1997), κ is a nonlocal function of the large-scale shear, and exhibits poor skill in the mean, but good skill in the variability: it has the best dispersion component of relative entropy, and is second only to GM0 for RMS errors in σ and pattern correlation of σ^2 . There is thus a trade-off between accuracy in the mean and accuracy in the variability where the model with the best mean (GM3) has the worst variability and the models with the best variability (GM0 and GM4) have the worst mean.

The sixth column shows the area-averaged relative entropy, which is essentially a weighted combination of the errors in the mean and variance. In every case the relative entropy is dominated by the signal, meaning that errors in the mean are worse than errors in the variance. Furthermore, the relative entropy is dominated by the contributions from a few locations; for example the median relative entropy for the GM4 scheme is only 8% of its mean.

Since the errors in the mean dominate those in the variability it might be expected that the scheme with the best mean (GM3) would have the best relative entropy. But the scheme with the best relative entropy is actually the GM4 scheme, which has moderate errors in both the mean and the variability; the GM3 scheme has the second-best relative entropy. The GM4 scheme has by far the best overall relative entropy of the GM schemes. Since it has a relatively poor time-mean structure, this suggests that the locations where the mean is in error have large temporal variability both in the GM4 scheme and in the eddy-resolving simulation.

Table 2

Domain-averaged measures of skill for the GM parameterizations. The RMS of σ and the RMS errors are measured in meters, the remaining columns are dimensionless. The RMS of σ for the reference simulation is 58 m.

	RMS of $\bar{\eta}$	RMS of σ	RMSE of σ	PC of σ^2	Dispersion	Rel. Ent.
GM0	131	44	33	0.80	28	995
GM1	105	33	37	0.77	63	1080
GM2	95	31	37	0.77	51	719
GM3	78	26	40	0.69	95	617
GM4	112	36	35	0.77	18	362

Fig. 3 shows the spatial structure of the error measures for the GM4 scheme, which has the best overall relative entropy. The mean bias $\bar{\eta}^t - \bar{\eta}_M^t$ is shown in Fig. 3(a), the ratio σ/σ_M is shown in Fig. 3(b), and the relative entropy is shown in Fig. 3(c). There is no clear correlation between mean bias and relative entropy; instead, regions of high relative entropy are correlated with regions where the model variability is too low. The model variance is significantly smaller than the true variance over most of the domain; in fact, the model variance is only larger than the true variance over 7% of the domain. Comparison with Fig. 1 suggests that there does not appear to be a simple explanation for the errors in terms of the local topographic gradient or the local mean shear.

4.4. Stochastic superparameterization

Table 3 shows the domain-averaged measures of skill for the stochastic SP simulations. The columns are the same as in Table 2. As with the GM parameterizations the configuration where the effective κ increases sharply with shear (SP-long) has a better time-mean layer interface structure than the configuration where κ increases less sharply (SP-short). The range of accuracy of the mean, as measured by the RMS error, is similar to the range of behavior of the GM models in the previous section with the best GM model comparable to the best SP model and the worst GM model comparable to the worst SP model.

The range of accuracy of the variability for the stochastic SP schemes is also similar to the range of accuracy for the GM models of the previous section, although the best SP models have better variability than the best GM models. In particular, the SP-short configurations that include Reynolds stress feedbacks to the coarse grid equations have very accurate variability as measured by the RMS errors in σ (29 and 31 meters), pattern correlations of σ^2 (0.80 and 0.81), and relative entropy dispersion (4 for both). For comparison, the best RMS error and pattern correlation for the GM models is 33 meters and 0.80 (GM0), and the best relative entropy dispersion is 18 (GM4).

As in the GM schemes, the total relative entropy for stochastic SP is dominated by the signal – the squared error in the mean weighted by the variance. The relative entropy is also dominated by the contributions from a few localized regions: Fig. 4(c) shows the spatial structure of the relative entropy for SP-short-RS-f which is dominated by contributions from the top left (north side, just downstream of Drake Passage), and bottom right (within Drake Passage), and along the lower boundary (see Fig. 1(a) for the structure of the underlying topography). The median relative entropy for SP-short-RS-f is 12 while its mean is 62 and Fig. 4(c) shows that it rises as high as 1000, further indicating that the average relative entropy reflects the contribution from just a few locations. Although the relative entropy is dominated by the signal, the configurations with the smallest RMS error in the mean (SP-long-RS) do not have the best relative entropy. Instead, as with the GM schemes, configurations with moderate mean accuracy and good variability (SP-short-RS) have the best overall relative entropy.

For a given averaging time (SP-short or SP-long) the configurations with Reynolds stresses perform better than those without. For configurations with Reynolds stresses, the SP configurations with short averaging time have better overall performance than the SP configurations with the long averaging time because of the improved variability. In contrast, for SP configurations without Reynolds stresses the best results are obtained using the long averaging time, primarily due to improvements in the structure of the mean. The SP-* β configurations, where the eddies feel the full coarse-grid PV gradient including topography, β , and the relative vorticity gradient, typically have better variability than the SP-* f configurations where the eddies only feel the mean shear. At short

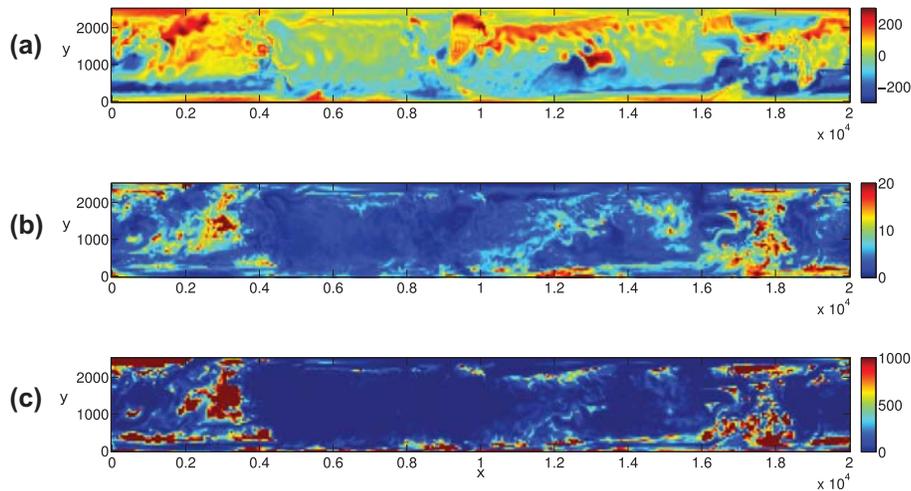


Fig. 3. Skill of the GM4 scheme: (a) time-mean bias in meters (color scale saturates at 300 m), (b) σ/σ_M , and (c) relative entropy. Axes are labeled in units of kilometers.

Table 3

Domain-averaged measures of skill for stochastic SP. The RMS of σ and the RMS errors are measured in meters, the remaining columns are dimensionless. The RMS of σ for the reference simulation is 58 m.

		RMSE of $\bar{\eta}$	RMS of σ	RMSE of σ	PC of σ^2	Dispersion	Rel. Ent.
SP-short	NRS- f	116	33	37	0.78	64	1417
	NRS- β	126	39	33	0.81	25	731
	RS- f	104	39	31	0.80	4	62
	RS- β	115	42	29	0.81	4	75
SP-long	NRS- f	87	28	39	0.76	39	455
	NRS- β	87	30	37	0.78	27	382
	RS- f	82	32	36	0.74	19	225
	RS- β	82	33	35	0.77	22	229

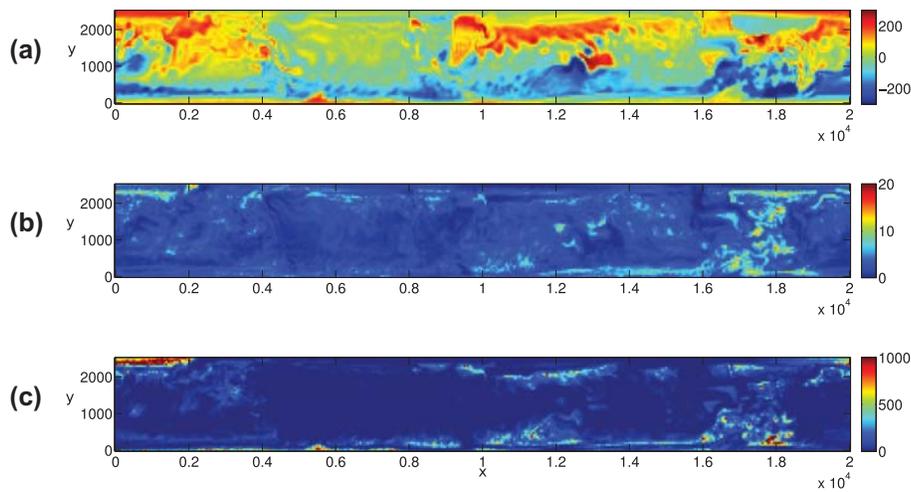


Fig. 4. Skill of the SP-short-RS- f scheme: (a) time-mean bias in meters (color scale saturates at 300 m), (b) σ/σ_M , and (c) relative entropy. Axes are labeled in units of kilometers.

averaging times the SP- β configurations have worse accuracy on the mean, but there is no difference at long averaging times.

In summary, there is a trade-off between accuracy in the mean and accuracy in the variability, as with the GM schemes. SP-long, and SP- β have better accuracy in the mean than SP-short, and SP- β , but the latter have better variability. The best overall accuracy, as measured by the relative entropy, is achieved by models with moderate accuracy in the mean and good accuracy in the variability: SP-short-RS- f and SP-short-RS- β .

Fig. 4 shows the spatial structure of the error measures for the SP-short-RS- f scheme, which has the best overall relative entropy. The mean bias $\bar{\eta}^t - \bar{\eta}_M^t$ is shown in Fig. 4(a), the ratio σ/σ_M is shown in Fig. 4(b), and the relative entropy is shown in Fig. 4(c). The results are qualitatively similar to the results for the GM4 scheme in Fig. 3, the primary difference being that the variability of the stochastic method is increased, leading to decreased relative entropy. Nevertheless, the variability is still too small over most of the domain: the model variability is too large on only 8% of the domain.

4.5. Discussion

There are several lessons to be learned from the results of the coarse model simulations. Perhaps the most surprising result is stochastic Reynolds stress feedbacks in the coarse model equations have a significant positive impact on both the mean and variability of the layer interface height. Asymptotic analysis (Pedlosky, 1984; Grooms et al., 2011) suggests that the primary feedback mechanism from mesoscale eddies to large-scale dynamics is through fluxes of heat and salinity (analogous to flux of layer interface height in the current setup), not through Reynolds stresses, and this is reflected in the widely-adopted GM mesoscale parameterizations. More recently Grooms et al. (2012) used asymptotic analysis to study the interactions at a smaller range of scales, where the large scales are closer to the mesoscale range. Their analysis indicates the importance of Reynolds stresses, although the analysis was assumed to apply to eddy-permitting situations unlike the current coarse-model configuration. The current setup has a relatively small coarse grid scale, and the impacts of the Reynolds stresses might be less important or less beneficial at coarser resolutions, but their importance is undeniable in our results: SP simulations with Reynolds stresses have more accurate mean and variability than SP simulations without, and the SP simulations with Reynolds stress terms have better relative entropy than all the other SP and GM simulations.

Another striking trend of our results is the trade-off between accuracy in the time-mean and accuracy in the variability. Deterministic GM schemes where the diffusivity κ is a sharply increasing function of the local baroclinic shear have more accurate time-mean interface height than schemes where κ is a shallow or constant function of the shear, but the trend is reversed when considering accuracy of the temporal variability. The same is true of stochastic SP when considering the effective κ (see Section 3.2.3), and similarly the SP- \ast - f schemes have better mean and worse variability than the SP- \ast - β schemes.

The accuracy of the mean and of the variability can be combined into a single measurement using the Gaussian approximation to the relative entropy, Eq. (24). The relative entropy is partitioned into the signal, which is the squared bias weighted by the model variance, and the dispersion, which measures errors in the variance. In all models the net relative entropy is dominated by the signal, which means that improvements can be had either by improving the bias or by increasing the variability. As noted above, there is a trade-off between improvements in the bias and improvements in the variability (increases in variability are generally improvements, since all models have too little); thus, the models with the best relative entropy strike a balance between bias and variability. Although the best SP schemes are not able to improve the bias compared to the best GM schemes, the SP schemes are able to strike the best balance between mean and variability as measured by the relative entropy. This implies that stochastic SP is improving the climate fidelity in comparison with the GM methods primarily by increasing the variability. Nevertheless, the variability of all models including stochastic SP is still too low.

While it is in principle possible to assess the accuracy of the parameterizations by diagnosing the eddy terms from the high resolution reference simulation for comparison, e.g. using a spatial low-pass filter (Nadiga, 2008; Grooms et al., 2013) or the more sophisticated approach of Berloff (2005), such an undertaking is computationally demanding and is beyond the scope of the present investigation. Since Figs. 3 and 4 do not suggest obvious relations between the model errors and the local topography, local mean shear, etc. we can only speculate on potential improvements to the schemes.

As noted above, we have not fully explored the impact of the eddy initial condition (see Section 3.2.2), the damping coefficient γ_0 , or the averaging time ϵ^{-1} on the performance of the stochastic SP scheme, and these could account for some of the remaining imperfections. In addition to these aspects of stochastic SP, we note two general areas for potential improvement to mesoscale parameterizations. First, all of the parameterizations considered here generate the subgridscale terms in the mean equations as (possibly random) functions of the local shear, topographic slope, etc., the only exception being the GM4 scheme, based on Visbeck et al. (1997). This latter scheme is non-local because the diffusivity κ is a function of the global structure of the baroclinic shear through the ‘width of the baroclinic zone.’ We find evidence from our eddy-resolving simulation that eddy kinetic energy (EKE), which is important for the accurate modeling of subgridscale terms (Bachman and Fox-Kemper, 2013; Marshall et al., 2012), is not well predicted by the local baroclinic shear or by the growth rate of local baroclinic instability, with regions of high EKE instead lying downstream of regions of high baroclinicity. This is in accord with the results of Grooms et al. (2013) who found direct evidence of eddy energy nonlocality in quasigeostrophic model simulations. Mesoscale parameterizations in general may benefit from incorporation of nonlocal effects, e.g. through prognostic eddy energy equations (Eden and Greatbatch, 2008; Grooms and Majda, 2014; Marshall and Adcroft, 2010).

Second, the only conventional GM parameterization considered here that incorporates the effects of topography is the GM3 scheme, which follows an ad hoc suggestion of Held and Larichev (1996). In contrast to most GM schemes, stochastic SP incorporates topographic effects directly, though the effects of small-scale topography have been avoided in the present investigation by using a smoothed topography. Topography, and in particular large-amplitude topography, can have a profound influence on eddy dynamics, and further research into its effects on parameterizations seems warranted. It is widely held that the quasigeostrophic approximation, an indispensable tool for modeling mesoscale eddies, requires small-amplitude topography, and this would seem to inhibit investigations of the impacts of large-amplitude topography on mesoscale parameterizations. Indeed, the topography-specific results of investigations that use both quasigeostrophic dynamics and large-amplitude topography, such as the present one, should be extrapolated with caution; the ‘Neptune effect’, (Holloway, 1992), for example, is a parameterization of the joint effects of eddies and topography that is motivated in the quasigeostrophic framework of small-amplitude topography. However, it is not true that quasigeostrophic dynamics are incompatible with large-amplitude topography. This has been demonstrated by Dewar and Leonov (2004), Leonov (2005) and Leonov and Dewar (2008), and means that the quasigeostrophic framework may yet be applied to inform the effects of large-amplitude topography on mesoscale eddies and their parameterizations.

5. Conclusions

Stochastic superparameterization (SP) has been tested in an idealized quasigeostrophic two-layer model of the Antarctic Circumpolar Current (ACC) at coarse, non-eddy-permitting resolution, and compared with five variants of the Gent and McWilliams (1990) parameterization. The framework for stochastic SP developed by Grooms and Majda (2013, 2014) has been used with minimal modifications, though the framework was previously applied in an eddy-permitting scenario. The primary differences between stochastic SP and the GM parameterizations considered

here are that stochastic SP is stochastic rather than deterministic, and that it parameterizes kinetic energy backscatter by momentum fluxes in addition to thickness fluxes.

Stochastic parameterizations in general are particularly useful in situations where the scale separation between resolved and unresolved dynamics is not large, because in such cases the feedbacks from the unresolved scales are not deterministic functions of the resolved flow. Stochastic parameterizations typically increase the variability of the resolved flow, although they can also affect the mean of the resolved flow because of nonlinearity. The primary improvement from using stochastic SP instead of GM is increased variability, although both the GM and stochastic SP models still have too little variability; the accuracy of the time-mean in the best stochastic SP model is comparable to the accuracy of the best GM model.

It is important to tune coarse-resolution models to have appropriate mean and variability (Shukla et al., 2006; DelSole and Shukla, 2010; Majda and Gershgorin, 2010; Majda and Gershgorin, 2011a; Majda and Gershgorin, 2011b). The agreement between the distribution of a coarse-resolution model's state and the distribution of the large-scale part of the real system is called climate fidelity and is measured by the relative entropy (Kullback–Leibler divergence), Eq. (23). In our results there is a trade-off between realistic variability and realistic mean: the models with the most accurate mean do not have the most accurate variability. Stochastic SP results in much better climate fidelity than the GM schemes tested here, primarily because of increased variability.

The increased climate fidelity of the stochastic SP models is largely due to the inclusion of stochastic Reynolds stress terms, which model kinetic energy backscatter. While the GM parameterizations considered here do not include eddy momentum fluxes, many authors have advocated parameterizations based on potential vorticity mixing, which does include eddy momentum fluxes (e.g. Marshall, 1981; Treguier et al., 1997; Marshall et al., 2012; Eden, 2010; Ringler and Gent, 2011; Olbers et al., 2000). These parameterizations are deterministic and diffusive, and though they may improve the large-scale mean it is not clear whether they might improve the variability of the large-scale dynamics.

The idealized model used here is a long way from global ocean models, and there is much to be done to prepare stochastic SP for that setting. The effects of varying (and possibly convectively unstable) stratification and of unbalanced dynamics near the equator need to be addressed, and efficient implementations need to be designed. But the results here suggest an intermediate step with reduced complexity. The minimal configuration of stochastic SP, where the coarse-model equations ignore the eddy Reynolds stresses and where the eddy equations feel only the coarse-model baroclinic shear (but not, e.g. β), is comparable to a stochastic GM method where the tracer flux has a random direction and amplitude. This configuration of stochastic SP has relatively high accuracy in both the time mean and in the variability, and its properties might be mimicked by a stochastic GM method that does not rely on the full complexity of the stochastic SP framework. The authors are actively pursuing such a strategy.

Acknowledgments

This research of A.J.M. is partially supported by the Office of Naval Research Grants N00014-11-1-0306 and ONR-DRI N00014-10-1-0554. This research of I.G. and K.S.S. is supported by the NSF Collaborations in Mathematical Geosciences program, Grant DMS-1025468. The authors thank L.-P. Nadeau for the use of his eddy-resolving QG model, and Malte Jansen and two anonymous reviewers whose suggestions strengthened the presentation.

Appendix A. Boundary conditions for QG PV inversion in a channel

This section discusses the different boundary conditions on ψ_i for the QG PV inversion problem specified by Eqs. (3) and (4) used in the coarse-grid and eddy-resolving simulations. In order for the velocity normal to the boundaries to be zero, ψ_i must be constant on each boundary segment. Thus, to uniquely specify the solution of the inversion problem, one must specify four constants $\Psi_{i,j}$ corresponding to the value of ψ_i at the southern ($j = 1$) and northern ($j = 2$) boundaries. If the external deformation radius were infinite ($2f_0^2/(gH) \rightarrow 0$) then one could add an arbitrary constant to ψ_i without changing q_i and without changing the velocity field, but with a finite external deformation radius this is no longer the case and all four boundary conditions need to be specified.

The first set of boundary conditions corresponds to mass conservation, and can be derived from the relationship between ψ_i and the layer thicknesses

$$\iint_{\Omega} \psi_i = 0 \quad (\text{A.1})$$

The double integral is taken over the full horizontal domain, denoted $(x, y) \in \Omega$.

We write the QG equations in the following general form

$$\partial_t q_1 = -\frac{2}{\rho_0 H} \partial_y \tau + \nabla \cdot \mathbf{F}_1 \quad (\text{A.2})$$

$$\partial_t q_1 = -\frac{2}{\rho_0 H} \partial_y \tau + \nabla \cdot \mathbf{F}_2 \quad (\text{A.3})$$

where the terms $\nabla \cdot \mathbf{F}_i$ include advection, β , topography, bottom friction, subgridscale parameterizations, and viscosity.

Following McWilliams (1977) we consider $(x, y) \in \omega \subseteq \Omega$ to be a rectangular region spanning the length of the channel bounded to the south by the southern boundary of the domain $\partial\omega_1 = \partial\Omega_1$ and to the north by a latitude line denoted $\partial\omega_2$. McWilliams (1977) provides many options for the additional boundary conditions, including

$$\oint_{\partial\omega_j} (\partial_{tn} \psi_i - \mathbf{F}_i \cdot \hat{\mathbf{n}}) = -\delta_{i,1} \frac{2L_x}{\rho_0 H} (-1)^j \tau_j \quad (\text{A.4})$$

where $\delta_{i,1}$ is the Kronecker delta, τ_j denotes the value of τ on $\partial\omega_j$. One value of j is chosen for each layer i . This choice can be motivated by appeal to the next-order corrections to QG theory, and can be shown to guarantee energy conservation for the inviscid problem.

Under the assumption of mass conservation we may integrate the QG equations over ω to give

$$\sum_j I_{ij} = \delta_{i,1} \frac{2L_x}{\rho_0 H} (\tau_1 - \tau_2) \quad (\text{A.5})$$

where

$$I_{ij} = \oint_{\partial\omega_j} (\partial_{tn} \psi_i - \mathbf{F}_i \cdot \hat{\mathbf{n}}) \quad (\text{A.6})$$

and ∂_n denotes the outward normal derivative. Note that this does not constitute a new, independent set of boundary conditions since it relies on mass conservation and the QG equations; however, it does imply that if the boundary condition (A.4) is imposed using $j = 1$ then it must also be true for $j = 2$ and vice versa. This proves that the boundary conditions (A.4) are equivalent whether the integrals are evaluated on either boundary, or along a line through the channel. It is more convenient to impose them at the boundary because the flux \mathbf{F} across the boundary typically is simpler than the flux across a line in the middle of the domain. The eddy-resolving

simulation imposes the extra boundary conditions (A.4) along a line three grid spaces (19.5 km) north of the southern boundary, whereas the coarse-grid simulations impose (A.4) along the southern boundary.

The eddy contribution to the flux F_i is $(-1)^i \kappa \nabla(\psi_1 - \psi_2)$ for the GM parameterizations. To impose no eddy flux normal to the wall one might impose either $\kappa = 0$ at the wall, or $\nabla(\psi_2 - \psi_1)$ at the wall. Neither condition is desirable since setting $\kappa = 0$ at the wall generates a discontinuity in the flux, and setting $\nabla(\psi_2 - \psi_1) = 0$ generates a strong constraint on the interior flow. Instead, we simply incorporate the GM component of F_i into the boundary conditions (A.4). The eddy fluxes normal to the boundary in the stochastic SP method are treated similarly.

Appendix B. Zonal stress balance

This section analyzes the stress balance of the eddy-resolving simulation. The time- and zonal-averaged QG equations, ignoring the hyperviscous terms, are

$$\partial_y \overline{\tilde{v}_1 \tilde{q}_1}^{tx} = -\frac{2}{\rho_0 H} \partial_y \tau \quad (\text{B.1})$$

$$\partial_y \overline{\tilde{v}_2 \tilde{q}_2}^{tx} = -r \partial_y^2 \overline{\tilde{\psi}_2}^{tx} - \frac{2f_0}{H} \partial_y \overline{v_2 h_b}^{tx}. \quad (\text{B.2})$$

The overbar $\overline{(\cdot)}^{tx}$ denotes the time- and zonal-mean and tildes $\tilde{\cdot}$ denote deviations from the zonal and time mean. Integrate these northwards from the southern boundary using impenetrability and the fact that $\tau = 0$ at the boundary to get

$$\overline{\tilde{v}_1 \tilde{q}_1}^{tx} = -\frac{2}{\rho_0 H} \tau \quad (\text{B.3})$$

$$\overline{\tilde{v}_2 \tilde{q}_2}^{tx} = r \overline{u_2}^{tx} - \frac{2f_0}{H} \overline{v_2 h_b}^{tx} \quad (\text{B.4})$$

Note that the time-averaged boundary conditions in Appendix A imply that $\overline{u_2} = 0$ at the boundary (when ignoring hyperviscosity). We add these together and average over latitude to arrive at the following balance of wind, Ekman, and topographic form stress for the channel

$$r \langle u_2 \rangle = \frac{2}{\rho_0 H} \langle \tau \rangle + \frac{2f_0}{H} \langle v_2 h_b \rangle \quad (\text{B.5})$$

where brackets denote the domain and time average. Cancellation of the volume integrated meridional eddy PV flux is well-known and can be derived using integration by parts. The lower-layer transport is simply $L_y H \langle u_2 \rangle$, and it is clearly set by the difference between wind stress and topographic form stress, scaled by the Ekman drag coefficient. We note that this is not a general result, but is specific to QG dynamics and linear drag. The lower layer transport equals $1217 - 1134 = 83$ Sv, where the positive term comes from wind stress and the negative one from topographic form stress.

The lower-layer transport in the coarse-grid models is set in the same way, with the contribution from wind stress remaining unchanged. The coarse-grid models all have lower-layer transport of approximately 104 Sv, which implies that the coarse-grid topographic form stress makes a contribution of 1113 Sv to the lower-layer instead of 1134. This amounts to less than 2% error in the coarse-grid representation of topographic form stress, but results in approximately 25% error in the lower-layer (and total) transport.

References

- Arakawa, A., 1966. Computational design for long-term numerical integration of the equations of fluid motion: two-dimensional incompressible flow. Part I. *J. Comput. Phys.* 1, 119–143.
- Bachman, S., Fox-Kemper, B., 2013. Eddy parameterization challenge suite. I: Eddy spindown. *Ocean Modell.* 64, 12–28.
- Berloff, P.S., 2005. On dynamically consistent eddy fluxes. *Dyn. Atmos. Oceans* 38, 123–146.
- Berloff, P.S., 2005. Random-forcing model of the mesoscale oceanic eddies. *J. Fluid Mech.* 529, 71–95.
- Berner, J., Shutts, G., Leutbecher, M., Palmer, T., 2009. A spectral stochastic kinetic energy backscatter scheme and its impact on flow-dependent predictability in the ecmwf ensemble prediction system. *J. Atmos. Sci.* 66.
- Brankart, J.M., 2013. Impact of uncertainties in the horizontal density gradient upon low resolution global ocean modelling. *Ocean Modell.* 66, 64–76.
- Buizza, R., Milleer, M., Palmer, T., 1999. Stochastic representation of model uncertainties in the ECMWF ensemble prediction system. *Q. J. R. Meteorol. Soc.* 125, 2887–2908.
- Campin, J.M., Hill, C., Jones, H., Marshall, J., 2011. Super-parameterization in ocean modeling: application to deep convection. *Ocean Modell.* 36, 90–101.
- Cessi, P., 2008. An energy-constrained parameterization of eddy buoyancy flux. *J. Phys. Ocean.* 38.
- Charney, J.G., 1971. Geostrophic turbulence. *J. Atmos. Sci.* 28, 1087–1095.
- Chelton, D.B., Schlax, M.G., Samelson, R.M., 2011. Global observations of nonlinear mesoscale eddies. *Progr. Oceanogr.* 91, 167–216.
- Davis, T.A., 2004. Algorithm 832: Umfpack v4. 3—an unsymmetric-pattern multifrontal method. *ACM Trans. Math. Softw. (TOMS)* 30, 196–199.
- DelSole, T., 2004. Stochastic models of quasigeostrophic turbulence. *Surv. Geophys.* 25, 107–149.
- DelSole, T., Shukla, J., 2010. Model fidelity versus skill in seasonal forecasting. *J. Climate* 23.
- Deng, Q., Khouider, B., Majda A.J., Submitted for publication. The MJO in a coarse-resolution GCM with a stochastic multcloud parameterization. *J. Atmos. Sci.*
- Dewar, W.K., Leonov, D.A., 2004. Variability on steep, confined topography. *Deep Sea Res. Part II: Top. Stud. Oceanogr.* 51, 2973–2993.
- Dolaptchiev, S.I., Klein, R., 2013. A multiscale model for the planetary and synoptic motions in the atmosphere. *J. Atmos. Sci.* 70.
- Eden, C., 2010. Parameterising meso-scale eddy momentum fluxes based on potential vorticity mixing and a gauge term. *Ocean Modell.* 32, 58–71.
- Eden, C., 2011. A closure for meso-scale eddy fluxes based on linear instability theory. *Ocean Modell.* 39, 362–369.
- Eden, C., 2012. Implementing diffusivities from linear stability analysis in a three-dimensional general circulation ocean model. *Ocean Modell.* 57–58, 15–28.
- Eden, C., Greatbatch, R.J., 2008. Towards a mesoscale eddy closure. *Ocean Modell.* 20, 223–239.
- Firing, Y.L., Chereskin, T.K., Mazloff, M.R., 2011. Vertical structure and transport of the antarctic circumpolar current in drake passage from direct velocity observations. *J. Geophys. Res.: Oceans* 1978–2012, 116.
- Frederiksen, J.S., Davies, A.G., 1997. Eddy viscosity and stochastic backscatter parameterizations on the sphere for atmospheric circulation models. *J. Atmos. Sci.* 54, 2475–2492.
- Frenkel, Y., Majda, A.J., Khouider, B., 2013. Stochastic and deterministic multcloud parameterizations for tropical convection. *Climate Dyn.* 41, 1527–1551.
- Gent, P.R., McWilliams, J.C., 1990. Isopycnal mixing in ocean circulation models. *J. Phys. Ocean.* 20, 150–155.
- Gent, P.R., Willebrand, J., McDougall, T.J., McWilliams, J.C., 1995. Parameterizing eddy-induced tracer transports in ocean circulation models. *J. Phys. Ocean.* 25, 463–474.
- Grabowski, W.W., 2004. An improved framework for superparameterization. *J. Atmos. Sci.* 61.
- Green, J., 1970. Transfer properties of the large-scale eddies and the general circulation of the atmosphere. *Q. J. R. Meteorol. Soc.* 96, 157–185.
- Grooms, I., Majda, A.J., 2013. Efficient stochastic superparameterization for geophysical turbulence. *Proc. Nat. Acad. Sci.* 110, USA.
- Grooms, I., Majda, A.J., 2014. Stochastic superparameterization in quasigeostrophic turbulence. *J. Comput. Phys.* 271.
- Grooms, I., Julien, K., Fox-Kemper, B., 2011. On the interactions between planetary geostrophy and mesoscale eddies. *Dyn. Atmos. Oceans* 51.
- Grooms, I., Smith, K.S., Majda, A.J., 2012. Multiscale models for synoptic-mesoscale interactions in the oceans. *Dyn. Atmos. Oceans* 58.
- Grooms, I., Nadeau, L.P., Smith, K.S., 2013. Mesoscale eddy energy locality in an idealized ocean model. *J. Phys. Ocean.* 43.
- Hallberg, R., 2013. Using a resolution function to regulate parameterizations of oceanic mesoscale eddy effects. *Ocean Modell.* 72, 92–103.
- Hallberg, R., Gnanadesikan, A., 2001. An exploration of the role of transient eddies in determining the transport of a zonally reentrant current. *J. Phys. Ocean.* 31.
- Held, I.M., Larichev, V.D., 1996. A scaling theory for horizontally homogeneous, baroclinically unstable flow on a beta plane. *J. Atmos. Sci.* 53, 946–952.
- Hogg, A.M., Meredith, M.P., Blundell, J.R., Wilson, C., 2008. Eddy heat flux in the southern ocean: response to variable wind forcing. *J. Climate* 21.
- Holloway, G., 1992. Representing topographic stress for large-scale ocean models. *J. Phys. Ocean.* 22, 1033–1046.
- Hutchinson, D.K., Hogg, A.M., Blundell, J.R., 2010. Southern ocean response to relative velocity wind stress forcing. *J. Phys. Ocean.* 40.

- Jansen, M.F., Held, I.M., 2014. Parameterizing subgrid-scale eddy effects using energetically consistent backscatter. *Ocean Modell.*
- Karspeck, A.R., Yeager, S., Danabasoglu, G., Hoar, T., Collins, N., Raeder, K., Anderson, J., Tribbia, J., 2013. An ensemble adjustment kalman filter for the CCSM4 ocean component. *J. Climate* 26.
- Khouider, B., Majda, A.J., Katsoulakis, M.A., 2003. Coarse-grained stochastic models for tropical convection and climate. *Proc. Nat. Acad. Sci. USA* 100, 11941–11946.
- Khouider, B., Biello, J., Majda, A.J., 2010. A stochastic multicloud model for tropical convection. *Commun. Math. Sci.* 8, 187–216.
- Killworth, P., 1997. On the parameterization of eddy transfer Part I. Theory. *J. Marine Res.* 55, 1171–1197.
- Kitsios, V., Frederiksen, J., Zidikheri, M., 2013. Scaling laws for parameterisations of subgrid eddy–eddy interactions in simulations of oceanic circulations. *Ocean Modell.* 68, 88–105.
- Kitsios, V., Frederiksen, J., Zidikheri, M., 2014. Scaling laws for parametrizations of subgrid interactions in simulations of oceanic circulations. *Philos. Trans. R. Soc. A: Math. Phys. Eng. Sci.* 372, 20130285.
- Kleeman, R., 2002. Measuring dynamical prediction utility using relative entropy. *J. Atmos. Sci.* 79.
- Kuhlbrodt, T., Smith, R., Wang, Z., Gregory, J., 2012. The influence of eddy parameterizations on the transport of the antarctic circumpolar current in coupled climate models. *Ocean Modell.* 52, 1–8.
- Larichev, V.D., Held, I.M., 1995. Eddy amplitudes and fluxes in a homogeneous model of fully developed baroclinic instability. *J. Phys. Ocean.* 25, 2285–2297.
- Leith, C., 1990. Stochastic backscatter in a subgrid-scale model: Plane shear mixing layer. *Phys. Fluids A: Fluid Dyn.* 2, 297–299 (1989–1993).
- Leonov, D., 2005. Effects of finite amplitude bottom topography on ocean variability (Ph.D. thesis). Florida State University.
- Leonov, D., Dewar, W., 2008. On nonlinear baroclinic trapped waves over abrupt topography. *Deep Sea Res. Part I: Oceanogr. Res. Pap.* 55, 1428–1437.
- Lin, J.W.B., Neelin, J.D., 2000. Influence of a stochastic moist convective parameterization on tropical climate variability. *Geo. Res. Lett.* 27, 3691–3694.
- Lin, J.W.B., Neelin, J.D., 2003. Toward stochastic deep convective parameterization in general circulation models. *Geo. Res. Lett.* 30.
- Majda, A.J., 2007a. Multiscale models with moisture and systematic strategies for superparameterization. *J. Atmos. Sci.* 64, 2726–2734.
- Majda, A.J., 2007b. New multiscale models and self-similarity in tropical convection. *J. Atmos. Sci.* 64, 1393–1404.
- Majda, A.J., Gershgorin, B., 2010. Quantifying uncertainty in climate change science through empirical information theory. *Proc. Nat. Acad. Sci. USA* 107, 14958–14963.
- Majda, A.J., Gershgorin, B., 2011a. Improving model fidelity and sensitivity for complex systems through empirical information theory. *Proc. Nat. Acad. Sci. USA* 108, 10044–10049.
- Majda, A.J., Gershgorin, B., 2011b. Link between statistical equilibrium fidelity and forecasting skill for complex systems with model error. *Proc. Nat. Acad. Sci. USA* 108, 12599–12604.
- Majda, A.J., Grooms, I., 2014. New Perspectives on Superparameterization for geophysical turbulence. *J. Comput. Phys.* 271.
- Majda, A., Wang, X., 2006. *Nonlinear Dynamics and Statistical Theories for Basic Geophysical Flows*. Cambridge University Press.
- Majda, A., Kleeman, R., Cai, D., et al., 2002. A mathematical framework for quantifying predictability through relative entropy. *Methods Appl. Anal.* 9, 425–444.
- Malecha, Z., Chini, G., Julien, K., 2014. A multiscale algorithm for simulating spatially-extended langmuir circulation dynamics. *J. Comput. Phys.* 271, 131–150.
- Marshall, J.C., 1981. On the parameterization of geostrophic eddies in the ocean. *J. Phys. Ocean.* 11, 257–271.
- Marshall, D.P., Adcroft, A.J., 2010. Parameterization of ocean eddies: Potential vorticity mixing, energetics and arnolds first stability theorem. *Ocean Modell.* 32, 188–204.
- Marshall, D.P., Maddison, J.R., Berloff, P.S., 2012. A framework for parameterizing eddy potential vorticity fluxes. *J. Phys. Ocean.* 42, 539–557.
- Marstorp, L., Brethouwer, G., Johansson, A.V., 2007. A stochastic subgrid model with application to turbulent flow and scalar mixing. *Phys. Fluids* (1994–present).
- Mason, P., Thomson, D., 1992. Stochastic backscatter in large-eddy simulations of boundary layers. *J. Fluid Mech.* 242, 51–78.
- McWilliams, J.C., 1977. A note on a consistent quasigeostrophic model in a multiply connected domain. *Dyn. Atmos. Oceans* 1, 427–441.
- Meredith, M.P., Hogg, A.M., 2006. Circumpolar response of southern ocean eddy activity to a change in the southern annular mode. *Geo. Res. Lett.* 33.
- Meredith, M.P., Naveira Garabato, A.C., Hogg, A.M., Farneti, R., 2012. Sensitivity of the overturning circulation in the southern ocean to decadal changes in wind forcing. *J. Climate* 25.
- Munday, D.R., Johnson, H.L., Marshall, D.P., 2013. Eddy saturation of equilibrated circumpolar currents. *J. Phys. Ocean.* 43.
- Nadeau, L.P., Straub, D.N., 2009. Basin and channel contributions to a model antarctic circumpolar current. *J. Phys. Ocean.* 39.
- Nadeau, L.P., Straub, D.N., 2012. Influence of wind stress, wind stress curl, and bottom friction on the transport of a model antarctic circumpolar current. *J. Phys. Ocean.* 42.
- Nadeau, L.P., Straub, D.N., Holland, D.M., 2013. Comparing idealized and complex topographies in quasigeostrophic simulations of an antarctic circumpolar current. *J. Phys. Ocean.* 43.
- Nadiga, B., 2008. Orientation of eddy fluxes in geostrophic turbulence. *Philos. Trans. R. Soc. A* 366, 2489–2508.
- Olbers, D., Wolff, J.O., Völker, C., 2000. Eddy fluxes and second-order moment balances for nonhomogeneous quasigeostrophic turbulence in wind-driven zonal flows. *J. Phys. Ocean.* 30, 1645–1668.
- Pacanowski, R., Griffies, S., 2000. MOM 3.0 Manual. Technical Report. NOAA/Geophysical Fluid Dynamics Laboratory.
- Pavan, V., Held, I.M., 1996. The diffusive approximation for eddy fluxes in baroclinically unstable jets. *J. Atmos. Sci.* 53, 1262–1272.
- Pedlosky, J., 1984. The equations for geostrophic motion in the ocean. *J. Phys. Ocean.* 14, 448–455.
- Porta Mana, P., Zanna, L., 2014. Toward a stochastic parameterization of ocean mesoscale eddies. *Ocean Modell.* 79, 1–20.
- Randall, D., Khairoutdinov, M., Arakawa, A., Grabowski, W., 2003. Breaking the cloud parameterization deadlock. *Bull. Am. Meteorol. Soc.* 84, 1547–1564.
- Randall, D., Branson, M., Wang, M., Ghan, S., Craig, C., Gettelman, A., Edwards, J., 2013. A community atmosphere model with superparameterized clouds. *EOS* 94, 221–228.
- Ringler, T., Gent, P., 2011. An eddy closure for potential vorticity. *Ocean Modell.* 39, 125–134.
- Schumann, U., 1995. Stochastic backscatter of turbulence energy and scalar variance by random subgrid-scale fluxes. *Proc. R. Soc. Lond. A* 451, 293–318, Osborne Reynolds Centenary Volume.
- Shukla, J., DelSole, T., Fennessy, M., Kinter, J., Paolino, D., 2006. Climate model fidelity and projections of climate change. *Geo. Res. Lett.* 33.
- Smith, R.D., Gent, P.R., 2004. Anisotropic Gent–McWilliams parameterization for ocean models. *J. Phys. Ocean.* 34.
- Srinivasan, K., Young, W., 2012. Zonostrophic instability. *J. Atmos. Sci.* 69.
- Stone, P.H., 1972. A simplified radiative-dynamical model for the static stability of rotating atmospheres. *J. Atmos. Sci.* 29, 405–418.
- Straub, D.N., 1993. On the transport and angular momentum balance of channel models of the antarctic circumpolar current. *J. Phys. Ocean.* 23, 776–782.
- Tansley, C.E., Marshall, D.P., 2001. On the dynamics of wind-driven circumpolar currents. *J. Phys. Ocean.* 31.
- Treguier, A., 1989. Topographically generated steady currents in barotropic turbulence. *Geophys. Astrophys. Fluid Dyn.* 47, 43–68.
- Treguier, A., McWilliams, J., 1990. Topographic influences on wind-driven, stratified flow in a β -plane channel: an idealized model for the antarctic circumpolar current. *J. Phys. Ocean.* 20, 321–343.
- Treguier, A., Held, I., Larichev, V., 1997. Parameterization of quasigeostrophic eddies in primitive equation ocean models. *J. Phys. Ocean.* 27, 567–580.
- Visbeck, M., Marshall, J., Haine, T., Spall, M., 1997. Specification of eddy transfer coefficients in coarse-resolution ocean circulation models. *J. Phys. Ocean.* 27, 381–402.