Chapter 5
Toward Consistent Subgrid Momentum Closures in Ocean Models

Sergey Danilov, Stephan Juricke, Anton Kutsenko and Marcel Oliver

Abstract State-of-the-art global ocean circulation models used in climate studies are only passing the edge of becoming "eddy-permitting" or barely eddy-resolving. Such models commonly suffer from overdissipation of mesoscale eddies by routinely used subgrid dissipation (viscosity) operators and a resulting depletion of energy in the large-scale structures which are crucial for draining available potential energy into kinetic energy. More broadly, subgrid momentum closures may lead to both overdissipation or pileup of eddy kinetic energy and enstrophy of the smallest resolvable scales. The aim of this chapter is twofold. First, it reviews the theory of two-dimensional and geostrophic turbulence. To a large part, this is textbook material with particular emphasis, however, on issues relevant to modeling the global ocean in the eddy-permitting regime. Second, we discuss several recent parameterizations of subgrid dynamics, including simplified backscatter schemes by Jansen and Held, stochastic superparameterizations by Grooms and Majda, and an empirical backscatter scheme by Mana and Zanna.

5.1 Introduction

Climate studies require model simulations over periods from centuries to millenia, which are only affordable if ocean models are kept relatively coarse. Many of them stay at a resolution of about one degree and need to parameterize the effect of unresolved mesoscale eddies and smaller-scale motions. The issue of mesoscale eddy

S. Danilov
Alfred Wegener Institute for Polar and Marine Research (AWI), Bremerhaven, Germany
e-mail: sergey.danilov@awi.de

S. Danilov · S. Juricke · A. Kutsenko · M. Oliver (✉)
Jacobs University, Bremen, Germany
e-mail: oliver@member.ams.org; m.oliver@jacobs-university.de

S. Juricke
e-mail: s.juricke@jacobs-university.de

A. Kutsenko
e-mail: a.kutsenko@jacobs-university.de

C. Eden and A. Iske (eds.), *Energy Transfers in Atmosphere and Ocean*,
Mathematics of Planet Earth 1, https://doi.org/10.1007/978-3-030-05704-6_5

parameterization attracts continuing interest as exemplified by recent studies on eddy potential vorticity fluxes (Marshall and Adcroft 2010; Eden 2010; Ringler and Gent 2011; Marshall et al. 2012). With increasing computational power, eddy-permitting (barely resolving the first baroclinic Rossby radius) or eddy-resolving models are becoming feasible for climate studies, too, so that mesoscale dynamics will gradually be resolved. Nonetheless, as the first baroclinic Rossby radius varies widely (with values below 10 km in high latitudes), even eddy-resolving models will not necessarily represent eddy dynamics with the same skill everywhere unless their resolution is on the scale of a few kilometers. Combining resolved mesoscale dynamics in some parts of the ocean with parameterized dynamics in the other part is an interesting possibility, but cautionary results by Hallberg (2013) indicate that the transition from parameterized eddies to resolved eddies can introduce problems of its own.

Even though state-of-the-art eddy-permitting or eddy-resolving models simulate the mesoscale dynamics with a certain skill, they still use some form of explicit and/or implicit viscosity, thought to represent the effect of unresolved small-scale subgrid dynamics. The motivation is based on the picture of quasigeostrophic turbulence (Charney 1971), which indicates that the direct cascade of enstrophy has to be removed at the grid scale to prevent the enstrophy from piling up, causing code instability. Fox-Kemper and Menemenlis (2008) discuss common approaches used in oceanographic practice, in particular the Smagorinsky or Leith parameterizations in either harmonic or biharmonic implementation. While these ideas appear plausible, there are no solid theoretical arguments, especially outside the limits of applicability of quasigeostrophic theory which is questionable at grid scale.

The detailed form of the subgrid operators (e.g., Laplacian vs. biharmonic viscosity), however, is known to impact the large-scale dynamics such as the path and separation of the Gulf Stream (Hecht et al. 2008b). Moreover, removal of enstrophy at the grid scale is accompanied by energy dissipation. For example, Danilov (2005) has shown that a direct enstrophy cascade in two-dimensional turbulence is always associated with a noticeable direct energy cascade, resulting in dissipation at finite resolution. Jansen and Held (2014) point out that the popular biharmonic viscosity operator suppresses resolved eddy motion in models where the separation between the mesoscale and the grid scale is insufficient. This reduces the ability of the flow to drain eddy kinetic energy from the available potential energy (APE), thereby distorting the entire energy cycle. The effect is most pronounced for eddy-permitting models where the grid scale and the scale of APE release are not well separated. It is also important for eddy-resolving models as the first baroclinic Rossby radius may locally drop below grid scale. The remedies are less immediate and open for investigation; searching for them is the main aim of ongoing studies. There is growing interest in this topic in the community as eddy-permitting models are now beginning to be used in climate research, so that the question of how to make them more realistic becomes pressing; see, e.g., Jansen et al. (2015), Berloff (2015), Cooper and Zanna (2015), and Cooper (2017).

Analyzing the effects of spectral pileup and backscatter of eddy energy in response to common subgrid parameterizations is rather straightforward for simple two-dimensional flows with a prescribed kinetic energy production rate (Graham and

Ringler 2013), but the question remains open in the context of more realistic dynamics which includes the effects of baroclinicity and where the geometry of boundaries and topography makes spectral analysis only locally applicable. Moreover, in real flows the balance between energy production and dissipation ceases to be local, which further complicates the situation. It is not clear how the subgrid operators affect the energy exchange between balanced (quasigeostrophic) and non-balanced motions as the resolution is increasing. More broadly, the mathematical side of subgrid parameterization as used in oceanographic tasks needs a more firm basis which would dictate a scale- and frame-invariant structure for admissible parameterizations.

It is important to note that the dynamics on scales close to the grid scale is affected not only by explicit subgrid parameterizations, but also by details of the discretization of momentum advection (see Figure 5.1). For example, high-order upwind transport algorithms based on the flux form of the advection operator have implicit numerical dissipation of the same order of magnitude as typical explicit dissipation (see, e.g., Mohammadi-Aragh et al. 2015). Further, there is evidence for a numerical (Hollingsworth) instability associated with the vector-invariant form of momentum advection which creates noise in the vertical velocity field and thus influences the APE to kinetic energy conversion; see the discussion in Gassmann (2013) and Danilov and Wang (2015). Understanding the effects induced by these or

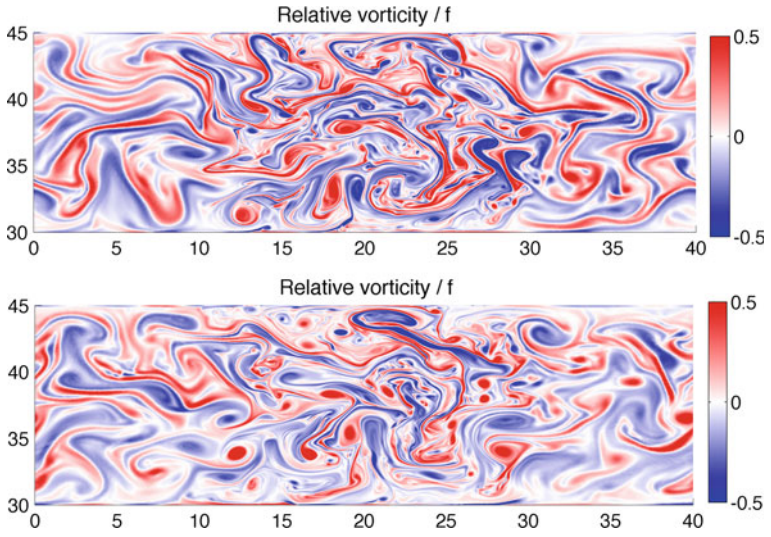

Fig. 5.1 Effect of momentum advection discretization on the relative vorticity field in a baroclinically unstable channel flow (top: vector-invariant form, bottom: flux form; near-surface snapshots are shown). Mesh resolution varies (1/36 degree in the central part and coarser elsewhere). Observed scales and amplitude of small eddies in the central part differ substantially between the two schemes due to the difference in implicit dissipation and discretization residual. The variance of vertical velocity (not shown) is substantially lower for the flux form, modifying the APE to eddy KE conversion rate. The figure is based on simulations reported in Danilov and Wang (2015)

other numerical details on the energy balance and accounting for their interplay with subgrid parameterizations is a necessary element on the road to rigorous analysis.

The need to explore the interplay between resolution, parameterized subgrid, and spurious numerical dissipation is particularly important for future earth system models employing multi-resolution technology, for example models based on FESOM (Wang et al. 2014) or ICON (Korn 2017). Recent results point to the retardation of eddy saturation when the upstream resolution is coarse (Danilov and Wang 2015). In multi-resolution models on unstructured meshes, subgrid momentum closures are also needed to stabilize against spurious numerical modes appearing on staggered triangular meshes (Danilov 2013), which adds numerical complexity.

For all the reasons mentioned, the question of how to return the overdissipated energy to the resolved scales is of central importance when working at eddy-permitting resolutions. This is known as the *energy backscatter* problem. On coarser meshes, one needs to additionally parameterize the contribution from mesoscale eddies. In both cases, there is growing interest in stochastic parameterizations. Stochastic parameterizations have been successfully used to maintain sufficient variance in ensemble forecasts (Palmer et al. 2009). However, energy and momentum consistency especially over long simulation timescales have not received as much attention (cf. the discussion in Franzke et al. 2015).

For the momentum closure problem in the ocean, stochastic parameterizations hold promise far beyond the idea of pure dissipation pursued by traditional deterministic subgrid parameterizations and also beyond downgradient parameterizations for unresolved mesoscale eddies. In the ocean context, systematic work on stochastic parameterizations is rather recent: Duan and Nadiga (2007), Mana and Zanna (2014), Jansen and Held (2014), Grooms et al. (2015b), Cooper and Zanna (2015), Cooper (2017), and Berloff (2015) all implement backscatter as stochastic forcing acting on the resolved flow, showing the potential of the approach, but also raising questions about the structure of this forcing and the choice of parameters. Stochastic backscatter can be implemented in a purely statistical way; more sophisticated approaches seek to include dynamical information, for example by shaping the backscatter forcing according to the nonlinear self-interaction derived from elementary solutions to the tangent linear equation (Berloff 2015, 2016).

Further open questions pertain to finding a more general mathematical framework and generalizations away from a quasigeostrophic setting toward the full primitive equations; work in this direction is at the very beginning. Related work on stochastic LES closures for the Navier–Stokes equations was done by San (2014) for the two-dimensional problem in vorticity form and by Xun and Wang (2014) for channel flow in three dimensions. Jansen and Held (2014) show that backscatter for the two-layer quasigeostrophic equations can be parameterized both stochastically and deterministically, with very similar results. Their approach is generalized to a simplified primitive equation isopycnal model in the presence of topography in Jansen et al. (2015). This is a very valuable step, providing one promising starting point and baseline benchmark.

Concerning subgrid dynamics, in most of the approaches cited, the model is either entirely local or is based on global energy constraints and thus couples the energy

budget over the entire domain. There are, however, early attempts at "second-order closures" by Daly and Harlow (1970) and Deardorff (1973), where the Reynolds tensor is treated as a prognostic variable and closure conditions have to be supplied for the higher-order moments. Schumann (1975) suggested a model with a single scalar transport equation for the subgrid energy. However, the algebraic closure relations for the diagnostic subgrid contributions are complicated and subject to solvability constraints; see Schmidt and Schumann (1989) and Schumann (1991). These ideas have been revisited subsequently by Schiestel and Dejoan (2005) and Chaouat (2012), but the problem remains open.

To summarize, the main open questions are:

- Find a suitable mathematical framework for subgrid momentum parameterization with minimized spurious energy dissipation.
- Implement practical backscatter algorithms in primitive equation ocean circulation models.

Any progress will have substantial impact on the energetic consistency of existing and future climate models.

This chapter aims at an elementary introduction to the circle of questions outlined above. Our intent is not to give ready and complete answers, but to highlight the issues and survey some of the emerging approaches.

We begin with a brief summary of the concept of subgrid momentum closures in Section 5.2. In Section 5.3, we review theoretical ideas on quasigeostrophic turbulence, with a brief summary on ocean mesoscale and submesoscale turbulence. Our main goal is to emphasize that the notion of "subgrid" scale, as related to ocean modeling, depends on the resolution, which complicates the question on subgrid closures.

In Sections 5.4 and 5.5, we review several proposed parameterizations. The first is the approach by Jansen and Held (2014) which is based on a local subgrid energy budget and an essentially empirical backscatter term which may be either deterministic or stochastic. The second, more sophisticated, but also more expensive and less easily generalized approach is due to Grooms and Majda (2013, 2014) who replace the Reynolds stress term with a stochastic process and explicitly evolve the local subgrid statistics in a local micro-cell attached to each grid box. The last emerging closure scheme is due to Mana and Zanna (2014); it was initially introduced semi-empirically, but later justified under precise assumptions by Grooms and Zanna (2017). The section closes with a brief review on α-models which provide a framework for regularizing fluid equations without adding dissipation which may possibly be interpreted as a nonlinear remapping of wavenumbers.

Section 5.6 offers concluding remarks and some very brief pointers to the literature for further directions beyond those covered so far.

5.2 Subgrid Momentum Closures

To fix concepts, let us focus on the momentum equations for a homogeneous incompressible or Boussinesq rotating ideal fluid,

$$\partial_t \boldsymbol{u} + \boldsymbol{\nabla} \cdot (\boldsymbol{u} \otimes \boldsymbol{u}) + 2\boldsymbol{\Omega} \times \boldsymbol{u} + \rho^{-1} \boldsymbol{\nabla} p = \boldsymbol{F} + D\boldsymbol{u}\,, \tag{5.1a}$$

$$\boldsymbol{\nabla} \cdot \boldsymbol{u} = 0\,, \tag{5.1b}$$

where $\boldsymbol{u}$ is the three-dimensional velocity field, $\boldsymbol{\Omega}$ the rotation vector, ρ the constant density, and p the pressure. All force terms are subsumed into $\boldsymbol{F}$. In particular, the system can represent the Boussinesq equations when augmented by thermodynamic equation(s) and with $\boldsymbol{F}$ representing all other forces including buoyancy. The operator D represents dissipation through physical processes or prior modeling. This equation may be read either as a partial differential equation (PDE) or as a fine-scale numerical approximation thereof.

In analogy with classical large eddy simulation, we introduce a coarsened velocity field $\overline{\boldsymbol{u}}$. We assume very little about the coarsening process other than that it is linear and commutes with time differentiation. In the classical PDE setting, $\overline{\boldsymbol{u}}$ may be obtained from $\boldsymbol{u}$ by convolution with a filter kernel. However, the more interesting point of view is that $\overline{\boldsymbol{u}}$ represents the solution of a modified numerical model at lower resolution. Then, $\overline{\boldsymbol{u}}$ satisfies the equation

$$\partial_t \overline{\boldsymbol{u}} + \overline{\boldsymbol{\nabla}} \cdot (\overline{\boldsymbol{u}} \otimes \overline{\boldsymbol{u}}) + 2\boldsymbol{\Omega} \times \overline{\boldsymbol{u}} + \rho^{-1} \overline{\boldsymbol{\nabla}}\, \overline{p} = \boldsymbol{R}(\boldsymbol{u}) + \overline{\boldsymbol{F}} + \overline{D}\overline{\boldsymbol{u}}\,, \tag{5.2a}$$

$$\overline{\boldsymbol{\nabla}} \cdot \overline{\boldsymbol{u}} = 0\,, \tag{5.2b}$$

with *eddy source term*

$$\boldsymbol{R}(\boldsymbol{u}) = \overline{\boldsymbol{\nabla}} \cdot (\overline{\boldsymbol{u}} \otimes \overline{\boldsymbol{u}}) - \overline{\boldsymbol{\nabla} \cdot (\boldsymbol{u} \otimes \boldsymbol{u})} + \overline{D\boldsymbol{u}} - \overline{D}\overline{\boldsymbol{u}} \tag{5.3}$$

where $\overline{\boldsymbol{\nabla}}$ and $\overline{D}$ denote the coarsened gradient or divergence operator and coarsened dissipation operator, respectively. Thinking of coarsening as a change of numerical resolution, we do not assume that coarsening commutes with the fine-scale operators even though this is often true for convolution coarsening on the continuum. However, we have made two minor simplifying assumptions: First, we have commuted the coarsening operation into the Coriolis term, which is exactly true on the f-plane and approximately true for a slowly varying Coriolis parameter, and second, we are assuming that the flow is incompressible at the coarse level with $\overline{p}$ denoting the implied coarsened pressure. (So, $\overline{p}$ is not obtained by convolution of p with the filter, but is chosen to enforce incompressibility of the coarse velocities.)

When considering the full Boussinesq system, the transport equations for potential temperature and other thermodynamic quantities need to be coarsened similarly (Aluie and Kurien 2011). For fully compressible flows, it is more natural to coarse-grain the product $\rho\boldsymbol{u}$, thus modifying the expression for $\boldsymbol{R}(\boldsymbol{u})$ above;

see, e.g., Aluie (2013). Additional complications arise with nonlinear equations of state; see, e.g., Eden (2016).

The modeling task is now the following: Find a *closure* or *subscale model* $\overline{\boldsymbol{R}}(\overline{\boldsymbol{u}})$ which correlates highly with the true $\boldsymbol{R}(\boldsymbol{u})$. The closure may be deterministic or include stochastic terms to reduce bias; it may also include infinitesimal or finite memory. If the momentum equation is coupled to thermodynamics, the same considerations apply to each of the prognostic equations.

One of the most elementary concerns is the distribution and flux of energy across scales, as the statistical behavior of the solution depends on it. In the next section, we illustrate the issues relevant to stratified turbulence in the ocean, concentrating mainly on the quasigeostrophic equations. They allow us to explore the main features of large-scale rotating stratified flow and are also used in most of the recent theoretical studies in the field.

5.3 Quasigeostrophic Turbulence and Ocean Eddies

In this section, we review the energetics of large-scale quasigeostrophic (QG) turbulence. In contrast to three-dimensional turbulence where energy cascades to small scales, QG turbulence is distinguished by an inverse cascade of barotropic kinetic energy to large scales and a cascade of enstrophy to small scales. Thus, it is often said that numerical schemes are required to provide a sink for QG enstrophy at grid scale without dissipating energy. In the following, we explain how and under which conditions this picture arises, but also point out its limitations when used in the context of ocean circulation models.

We emphasize that the classical picture of QG turbulence is strictly valid only under the assumption that QG enstrophy is dissipated at scales much smaller than the forcing scale and that turbulence remains geostrophic across all scales. In stratified flow, the main source of barotropic kinetic energy is the conversion of available potential energy via baroclinic instability. The most unstable baroclinic modes occur close to the first internal Rossby radius of deformation L_d. As we move to even smaller scales in a full model, the ageostrophic or non-balanced component of the flow increases and the quasigeostrophic approximation becomes inaccurate. Thus, there is only a finite small range of scales between L_d and the scale where ageostrophy starts to be important; at smaller scales, the energy cascade is direct. Moreover, since this direct cascade acts as an energy sink, there must also be some downscale flux of energy across the geostrophic range to feed it.

All of the quasigeostrophic models are able to capture only the part of the dynamics that stays close to geostrophic balance, often referred to as *mesoscale eddies*. In the real ocean, the scales at which ageostrophic effects are becoming important are rather close to L_d (see, e.g., Callies and Ferrari 2013) so that the presence of ageostrophic motions, often accompanying smaller-size *submesoscale eddies*, may have a significant impact on the direction of the energy transfer across scales. An additional complication arises from the fact that there is also a direct cascade of

available potential energy which implies that the forcing of the barotropic cascade does not only take place near the most baroclinically unstable mode, but is distributed across a wider range of scales.

In the following, we begin with the simplest concepts in a purely two-dimensional setting, then move to a two-layer model, and finally discuss the continuously stratified quasigeostrophic equations.

5.3.1 Two-Dimensional Turbulence

The very basic notions of rotating turbulence and ocean mesoscale eddies can be introduced in the framework of two-dimensional quasigeostrophic dynamics.

The barotropic quasigeostrophic equations are reviewed, e.g., in Franzke et al. (2019). In the beta-plane approximation and written in terms of the relative vorticity ζ, they read

$$\partial_t \zeta + [\psi, \zeta] + \beta\, \partial_x \psi = F + D\zeta\,, \tag{5.4a}$$

$$\zeta = \Delta \psi\,, \tag{5.4b}$$

where brackets denote the Jacobian operator $[\psi, \zeta] = \boldsymbol{\nabla}^{\perp}\psi \cdot \boldsymbol{\nabla}\zeta$ with $\boldsymbol{\nabla}^{\perp} = (-\partial_y, \partial_x)$, ψ denotes the stream function, β is the beta-parameter, D is a dissipation operator to be specified below, and F is the forcing. The forcing F is maintained by baroclinic or barotropic instabilities evolving at some intermediate scales.[1]

To begin, we set $\beta = 0$. Further, to simplify the discussion, we non-dimensionalize the horizontal length scale and consider (5.4) on the doubly periodic domain $\mathbb{T}^2 = [0, 2\pi]^2$, so that we can pass to the Fourier representation[2] where

$$\zeta_{\boldsymbol{k}} = \frac{1}{2\pi} \int_{\mathbb{T}^2} \mathrm{e}^{-\mathrm{i}\boldsymbol{k}\cdot\boldsymbol{x}}\, \zeta(\boldsymbol{x})\, \mathrm{d}\boldsymbol{x} \tag{5.5}$$

for $\boldsymbol{k} \in \mathbb{Z}^2$. It is useful to separate the dissipation operator D into "infrared" and "ultraviolet" parts that effectively act on large (D_{i}) and small (D_{u}) scales. For simplicity, we assume that these operators are diagonal in Fourier space, so that the transformed vorticity equation (5.4a) takes the form

$$\partial_t \zeta_{\boldsymbol{k}} + J_{\boldsymbol{k}} = D_{\mathrm{i}}(\boldsymbol{k})\, \zeta_{\boldsymbol{k}} + D_{\mathrm{u}}(\boldsymbol{k})\, \zeta_{\boldsymbol{k}} + F_{\boldsymbol{k}} \tag{5.6}$$

[1] For simplicity, our definition of vorticity (5.4b) does not include a barotropic stretching term modeling the effect of free surface elevation. In a geophysical context, this means that (5.4) is restricted to scales smaller than the external Rossby radius of deformation $L_{\mathrm{e}} = c/f$, where c is the speed of surface gravity waves and f is the Coriolis parameter. This is a reasonable assumption for the ocean where $L_{\mathrm{e}} \approx 2000\,\mathrm{km}$ with $c = 200\,\mathrm{m/s}$ and $f = 0.0001\ \mathrm{s}^{-1}$, but is more questionable for the atmosphere.

[2] For the purpose of this exposition, we are using the symmetric definition of the Fourier transform so that the Parseval identity holds with constant one.

where, writing $p = |\boldsymbol{p}|$ and likewise for the other wavenumber vectors, the Jacobian term is described by

$$J_{\boldsymbol{k}} = \frac{1}{2\pi} \sum_{\boldsymbol{k}=\boldsymbol{p}+\boldsymbol{q}} \frac{\boldsymbol{p}^{\perp} \cdot \boldsymbol{q}}{p^2} \, \zeta_{\boldsymbol{p}} \, \zeta_{\boldsymbol{q}} \, . \tag{5.7}$$

In the absence of dissipation and forcing, equation (5.6) conserves *energy*

$$E = \sum_{\boldsymbol{k} \in \mathbb{Z}^2} E_{\boldsymbol{k}} = -\frac{1}{2} \sum_{\boldsymbol{k} \in \mathbb{Z}^2} \psi^*_{\boldsymbol{k}} \, \zeta_{\boldsymbol{k}} \tag{5.8}$$

and *enstrophy*

$$Z = \sum_{\boldsymbol{k} \in \mathbb{Z}^2} Z_{\boldsymbol{k}} = \frac{1}{2} \sum_{\boldsymbol{k} \in \mathbb{Z}^2} \zeta^*_{\boldsymbol{k}} \, \zeta_{\boldsymbol{k}} \, , \tag{5.9}$$

with star denoting the complex conjugate.

The presence of two integrals imposes constraints on how energy and enstrophy are transferred in spectral space. The energy balance in each mode $\boldsymbol{k}$ is obtained by multiplying (5.6) by $\psi^*_{\boldsymbol{k}}$ and taking the real part, so that

$$\partial_t E_{\boldsymbol{k}} = T_{\boldsymbol{k}} + 2 \, D_{\mathrm{i}}(\boldsymbol{k}) \, E_{\boldsymbol{k}} + 2 \, D_{\mathrm{u}}(\boldsymbol{k}) \, E_{\boldsymbol{k}} + P_{\boldsymbol{k}} \, , \tag{5.10}$$

where $P_{\boldsymbol{k}} = -\Re[\psi^*_{\boldsymbol{k}} \, F_{\boldsymbol{k}}]$ is the rate of energy pumping and $T_{\boldsymbol{k}} = \Re[\psi^*_{\boldsymbol{k}} \, J_{\boldsymbol{k}}]$ is the rate of nonlinear energy transfer into mode $\boldsymbol{k}$.[3] Using (5.7), we can write

$$T_{\boldsymbol{k}} = \sum_{\{\boldsymbol{p},\boldsymbol{q}\} \colon \, \boldsymbol{k}+\boldsymbol{p}+\boldsymbol{q}=0} T(\boldsymbol{k}|\boldsymbol{p}\boldsymbol{q}) \, , \tag{5.11}$$

where

$$T(\boldsymbol{k}|\boldsymbol{p}\boldsymbol{q}) = \frac{1}{2\pi} \, \boldsymbol{p}^{\perp} \cdot \boldsymbol{q} \, (q^2 - p^2) \, \Re[\psi_{\boldsymbol{k}} \, \psi_{\boldsymbol{p}} \, \psi_{\boldsymbol{q}}] \tag{5.12}$$

denotes the rate of energy transfer into mode $\boldsymbol{k}$ from modes $\{\boldsymbol{p}, \boldsymbol{q}\}$ and the sum in (5.11) is taken over un-ordered sets $\{\boldsymbol{p}, \boldsymbol{q}\}$.

Summing up all the $T_{\boldsymbol{k}}$, we obtain the overall rate of nonlinear energy transfer T. Clearly, $T = 0$ as the rates of sending and receiving energy must balance across all modes. Following (Fjørtoft 1953), we sort the terms in this sum according to membership in resonant triads of modes

$$\mathscr{S} = \{\{\boldsymbol{k}, \boldsymbol{p}, \boldsymbol{q}\} \colon \boldsymbol{k} + \boldsymbol{p} + \boldsymbol{q} = 0\} \, , \tag{5.13}$$

[3] Expression (5.11) shows that the *instantaneous* rate of energy transfer into mode $\boldsymbol{k}$ can only be nonzero provided $\psi_{\boldsymbol{k}}$ is nonzero. This, however, does not imply that a mode which is initially zero remains zero for $t > 0$ as (5.6) allows a tendency for $\zeta_{\boldsymbol{k}}$. See, e.g., the discussion in Moffatt (2014).

so that

$$T = \sum_{\{\boldsymbol{k},\boldsymbol{p},\boldsymbol{q}\}\in\mathscr{S}} \left(T(\boldsymbol{k}|\boldsymbol{pq}) + T(\boldsymbol{p}|\boldsymbol{kq}) + T(\boldsymbol{q}|\boldsymbol{kp})\right). \tag{5.14}$$

Within each triad, $\boldsymbol{k}^\perp \cdot \boldsymbol{q} = -\boldsymbol{p}^\perp \cdot \boldsymbol{q}$. This directly implies[4] that

$$T(\boldsymbol{p}|\boldsymbol{kq}) = -\frac{q^2 - k^2}{q^2 - p^2}\, T(\boldsymbol{k}|\boldsymbol{pq}) \tag{5.15a}$$

and

$$T(\boldsymbol{q}|\boldsymbol{kp}) = -\frac{k^2 - p^2}{q^2 - p^2}\, T(\boldsymbol{k}|\boldsymbol{pq})\,. \tag{5.15b}$$

These identities constrain the transfer of energy within the triad: If $p < k < q$ and mode $\boldsymbol{k}$ loses energy by interacting with modes $\boldsymbol{p}$ and $\boldsymbol{q}$, the two other modes gain energy; vice versa, if mode $\boldsymbol{k}$ gains energy in this triad interaction, then modes $\boldsymbol{p}$ and $\boldsymbol{q}$ lose energy. The same holds true for enstrophy. In other words, nonlinear interactions between three modes always transfer energy and enstrophy either from or to the central component.

The total transfer $T_{\boldsymbol{k}}$ involves all triads this mode participates in and cannot be predicted without additional arguments. Consider first the case without forcing and dissipation, and define the *energy wavenumber* k_e as the centroid of the spectral energy density $E(k)$:

$$k_\mathrm{e} = \frac{1}{E}\sum_k k\, E(k)\,, \tag{5.16}$$

where we assume that the distribution of energy is isotropic in wavenumber space with $E(k)$ denoting the energy in the shell $k = |\boldsymbol{k}|$. The second moment

$$I = \sum_k (k - k_\mathrm{e})^2\, E(k) = Z - k_\mathrm{e}^2\, E \tag{5.17}$$

is expected to increase with time if energy spreads over wavenumbers. This is natural to expect for any energy spectrum that is initially spectrally localized. Conservation of energy and enstrophy implies that

[4] An independent, conceptual proof goes as follows. Consider a flow in the absence of dissipation and forcing where the Fourier coefficients of only a single resonant triad are nonzero. Such a flow preserves energy and enstrophy, so that $\partial_t(E_{\boldsymbol{k}} + E_{\boldsymbol{p}} + E_{\boldsymbol{q}}) = 0$ and $\partial_t(k^2\, E_{\boldsymbol{k}} + p^2\, E_{\boldsymbol{p}} + q^2\, E_{\boldsymbol{q}}) = 0$. But with only a single active triad, we can factor out $\Re[\psi_{\boldsymbol{k}}\, \psi_{\boldsymbol{p}}\, \psi_{\boldsymbol{q}}]$ from the expressions for the rate of energy and enstrophy transfer. Thus, the algebraic prefactors must already cancel, implying triadwise conservation of energy and enstrophy also for general flows; (5.15) is a direct consequence of these conservation laws.

$$\frac{\mathrm{d}I}{\mathrm{d}t} = -E\,\frac{\mathrm{d}k_\mathrm{e}^2}{\mathrm{d}t}\,, \tag{5.18}$$

so when energy spreads over wavenumbers, k_e decreases; i.e., energy moves to larger scales. Similarly, it can be shown that the enstrophy centroid moves downscale if and only if the second moment of enstrophy indicates a spread of the enstrophy distribution; see Vallis (2006) for details. This consideration indicates that if two-dimensional freely evolving turbulence develops cascades, we should expect an inverse energy cascade and a direct enstrophy cascade. It does not mean that there is no energy transfer to small scales or enstrophy transfer to large scales, it only means that on average energy tends to go upscale and enstrophy tends to go downscale.

In practice, turbulent flows are forced–dissipative systems.[5] They can reach a statistically steady state if dissipation is present at both spectral ends, as is envisioned in (5.6). Although the cause of infrared dissipation is not immediately apparent, in many cases its role can be efficiently played by bottom friction.[6] We thus return to the forced–dissipative case and consider the idealized situation when the forcing F is spectrally localized to a small interval around a forcing wavenumber k_f, infrared dissipation is localized to wavenumbers $k < k_\mathrm{i}$, and ultraviolet dissipation is localized to $k > k_\mathrm{u}$, with $k_\mathrm{i} < k_\mathrm{f} < k_\mathrm{u}$.[7] Assuming statistical stationarity, the mean rate of energy injection ε is balanced by the mean rate of energy dissipation in the infrared ε_i and the mean rate of energy dissipation in the ultraviolet ε_u, i.e.,

$$\varepsilon = \varepsilon_\mathrm{i} + \varepsilon_\mathrm{u}\,. \tag{5.19}$$

Likewise, writing η to denote the mean rate of enstrophy injection near wavenumber k_f, we balance with the mean enstrophy dissipation rates η_i and η_u in the respective dissipation ranges, so that

$$\eta = k_\mathrm{f}^2\,\varepsilon = \eta_\mathrm{i} + \eta_\mathrm{u}\,. \tag{5.20}$$

---

[5] The real ocean is close to this idealization, but not fully, as some eddies leave their regime of creation and evolve freely before being dissipated. Thus, temporal averages of forcing and dissipation may not balance locally in space.

[6] Physically, bottom or surface drag is due to small-scale turbulence, yet modeled in the equations without horizontal differential operators. Thus, the bulk of energy dissipation will happen at the scales where energy is residing; i.e., it will be infrared.

[7] In practice, this assumption is well satisfied at the infrared end because we commonly get energy spectra that are sufficiently red (with theoretical slopes of $-5/3$ or -3), so that most energy is residing at small wavenumbers. Thus, linear (Rayleigh) friction modeling bottom drag with $D_\mathrm{i} = -\lambda$ is concentrated where the kinetic energy is concentrated—at small wavenumbers. The same holds true for quadratic friction proportional to $\boldsymbol{u}|\boldsymbol{u}|$. The situation is more subtle at the ultraviolet end. Laplacian viscosity has $D_\mathrm{u}(k) = -\nu k^2$; i.e., it is concentrated where the enstrophy is concentrated. Here, we have a formal problem, for if the spectral slope is -3, $D_\mathrm{u}(k)\,E(k)$ will have a maximum in the forcing range. The slope is frequently even steeper (when coherent vortices are formed), so Laplacian viscosity will affect the forcing range. Thus, for true ultraviolet dissipation one needs hyperviscosity (a biharmonic operator at least, but even that is insufficient if the spectra are steeper than -4 which sometimes occurs).

Noting that $\eta_\mathrm{i} \leq k_\mathrm{i}^2\,\varepsilon_\mathrm{i}$, we estimate

$$\eta \geq \eta_\mathrm{u} = k_\mathrm{f}^2\,\varepsilon - \eta_\mathrm{i} \geq k_\mathrm{f}^2\,\varepsilon - k_\mathrm{i}^2\,\varepsilon_\mathrm{i} \geq (k_\mathrm{f}^2 - k_\mathrm{i}^2)\,\varepsilon = (1 - k_\mathrm{i}^2/k_\mathrm{f}^2)\,\eta\,. \tag{5.21}$$

Thus, $\eta_\mathrm{u} \to \eta$ in the limit $k_\mathrm{i}/k_\mathrm{f} \to 0$. Similarly, noting that $\eta_\mathrm{u} \geq k_\mathrm{u}^2\,\varepsilon_\mathrm{u}$, we can show that $\varepsilon_\mathrm{i} \to \varepsilon$ when $k_\mathrm{f}/k_\mathrm{u} \to 0$.

Thus, in the asymptotic limit $k_\mathrm{i} \ll k_\mathrm{f} \ll k_\mathrm{u}$, there is only upscale energy transfer for $k < k_\mathrm{f}$ and only downscale enstrophy transfer for $k > k_\mathrm{f}$. As a result, these regimes are called *(inverse) energy range* and *(direct) enstrophy range*, respectively. The mean energy spectral density $\langle E(k)\rangle$ should be such that a constant spectral energy flux is carried across each range.[8] Assuming that the mean energy transfer is spectrally local, as well as spatially homogeneous and isotropic, one expects that, in the energy range, $\langle E(k)\rangle$ depends only on the infrared energy dissipation rate $\varepsilon_\mathrm{i} = \varepsilon$ and on k. This lead Kraichnan (1967), Leith (1968), and Batchelor (1969), hereafter KLB (following earlier arguments by Kolmogorov for classical turbulence) to conclude that the only dimensionally consistent scaling law is

$$\langle E(k)\rangle = C_E\,\varepsilon^{\frac{2}{3}}\,k^{-\frac{5}{3}}\,. \tag{5.22}$$

Likewise, in the enstrophy range, one expects that $\langle E(k)\rangle$ will depend only on the ultraviolet enstrophy dissipation rate $\eta_\mathrm{u} = \eta$ and on k, leading to the scaling law

$$\langle E(k)\rangle = C_Z\,\eta^{\frac{2}{3}}\,k^{-3}\,. \tag{5.23}$$

The picture outlined above has two important limitations. The first one relates to the KLB assumption that only local triad interactions (triads where p, k, and q are of the same magnitude) contribute to the mean transfer of energy. For an individual wave number $\boldsymbol{k}$ in the energy or the enstrophy range where forcing and dissipation are absent, we expect that the mean transfer rate $\langle T_{\boldsymbol{k}}\rangle$ is zero, which only means that some triads carry energy to $\boldsymbol{k}$ and some from it. How they do this, however, does not really agree with the KLB picture—in real forced–dissipative two-dimensional turbulence, the contribution from non-local triads is indispensable. For a mode $\boldsymbol{p}$ in the energy range, one cannot neglect triads with two long legs $\boldsymbol{k}$ and $\boldsymbol{q}$ in the forcing or enstrophy range. The local triads with legs being about $\boldsymbol{p}$ dominate locally, but

[8] To be mathematically precise, the averaging operation $\langle\cdot\rangle$ must be seen as an ensemble average, even though we have no *a priori* knowledge of the probability measure. In practice, the necessary size of the ensemble is also an issue, so one may want to resort to averaging in time: A configuration is "practically statistically stationary" if the difference of averages over an interval $[t, t+T]$ and $[t, t+2T]$ is less than some prescribed tolerance. One needs, however, to recognize that the required time interval T depends on the quantities we work with. It is relatively small for $E(k)$ and the mean spectral energy flux $\Pi(k)$, it is relatively large for $E_{\boldsymbol{k}}$ or $T_{\boldsymbol{k}}$, and it is even larger for partial transfers $T(\boldsymbol{k}|\boldsymbol{pq})$.

contrary to expectations their average effect is not leading to the inverse energy transfer. Similarly, for mode $\boldsymbol{k}$ in the enstrophy range, one cannot neglect triads with one short leg $\boldsymbol{p}$ in the forcing or energy range. The first such analysis is due to Maltrud and Vallis (1993) and was corroborated by Danilov and Gurarie (2001). This means that vortices near the forcing scale are strong enough to stir small-scale vortices in the enstrophy range, which is precisely the way these smaller vortices are formed. The presence of non-locality violates the KLB argument, for it is explicitly assumed that $\langle E(k)\rangle$ in the energy range may only depend on the mean energy flux ε and on k (and not, e.g., on the forcing range), and similarly for the enstrophy range.

The second limitation of the classical KLB picture is that in real systems, the energy and enstrophy ranges are finite. If forcing pumps energy at intermediate scale k_f with finite separation from k_i and k_u, both energy *and* enstrophy are transferred to large *and* small scales through nonlinear interactions. If the wavenumber intervals separating forcing from dissipation are sufficiently broad, most of the energy is transferred upscale and most of the enstrophy downscale. However, these intervals are never broad enough in the ocean, and the question of the amplitude of the direct energy cascade relevant to the ocean is open.

However, even on finite ranges, the classical picture is not entirely lost. The following argument due to Gkioulekas and Tung (2007) provides integral bounds on energy and enstrophy fluxes which do not depend on infinite scale separation. To ease notation, we assume a wavenumber continuum (i.e., an unbounded domain in physical space) and, as before, consider energy densities and energy transfer rate densities as a function of the wavenumber modulus k. The argument, in essence, does not depend on this assumption. In a statistically stationary state, the average $\langle \partial_t E_{\boldsymbol{k}}\rangle = 0$, so that, taking the time or ensemble mean of (5.10) and averaging over the shell $|\boldsymbol{k}| = k$, we obtain

$$\langle T(k)\rangle = -D(k)\,\langle E(k)\rangle - \langle P(k)\rangle \tag{5.24}$$

where we assume that the dissipation operators depend only on k, so that we can write $D(k) = 2D_\mathrm{i}(k) + 2D_\mathrm{u}(k)$. Retaining the assumption that D_i is dissipating at wavenumbers smaller than k_i and D_u is dissipating at wavenumbers larger than k_u, the mean spectral energy flux

$$\Pi(k) = \int_k^\infty \langle T(\kappa)\rangle \,\mathrm{d}\kappa = -\int_0^k \langle T(\kappa)\rangle \,\mathrm{d}\kappa \tag{5.25}$$

is necessarily negative for $k < k_\mathrm{i}$ and positive for $k > k_\mathrm{u}$. The equality between the two integrals in (5.25) holds for each realization pointwise in time due to conservation of energy in the inviscid unforced system. Analogous statements hold true for the spectral enstrophy flux. Then,

$$\int_0^k 2\xi\, \Pi(\xi)\, \mathrm{d}\xi = \int_0^k (\kappa^2 - k^2)\, \langle T(\kappa)\rangle\, \mathrm{d}\kappa$$
$$= -\int_k^\infty (\kappa^2 - k^2)\, \langle T(\kappa)\rangle\, \mathrm{d}\kappa$$
$$= \int_k^\infty (\kappa^2 - k^2)\, (D(\kappa)\, \langle E(\kappa)\rangle - \langle P(\kappa)\rangle)\, \mathrm{d}\kappa\,, \tag{5.26}$$

where the first identity is obtained by exchanging the order of integration, the second identity is once again based on the conservation of energy and enstrophy in the inviscid unforced case, and the last step uses the statistical stationarity relation (5.24). When $k > k_\mathrm{u}$, the rate of energy pumping $P(\kappa)$ appearing in the integrand of (5.26) is zero, while the contribution from dissipation is negative. Thus, for every $k > k_\mathrm{u}$ and, trivially, for every $k < k_\mathrm{i}$,

$$\int_0^k \xi\, \Pi(\xi)\, \mathrm{d}\xi < 0\,. \tag{5.27}$$

Due to the weight in this integral, we see that the upscale flux of energy for $k < k_\mathrm{i}$ must be typically larger than the downscale flux of energy for $k > k_\mathrm{u}$. A similar inequality shows that the enstrophy flux is predominantly downscale (Gkioulekas and Tung 2007).

Whether or not inertial ranges can be observed depends on the spectral loci of dissipation and forcing. In particular, when an inverse cascade is observed, it only means that some energy dissipation is located at smaller wavenumbers than energy forcing. As a rule, dissipation and forcing are spread over wavenumbers and may even intersect. Moreover, when forcing extends up to the spectral cutoff k_max of a simulation, the small direct cascade of energy may be partially hidden. Thus, in most cases, clean inertial ranges are absent. And even when inertial ranges void of dissipation and forcing exist, the observed spectra may deviate from the KLB predictions because non-local triad interactions are always present and may be significant for finite ranges; see the discussion and examples in Danilov (2005). We conclude that spectral slopes alone tell very little about the nature of the underlying dynamics, and one must turn to exploring the distribution of forcing and dissipation over scales.

Let us comment briefly on the case when $\beta \neq 0$. In this situation, energy is channeled into zonal modes and large-scale features become highly anisotropic: Jets appear near the Rhines scale $L_\mathrm{Rh} = E^{1/4}/\beta^{1/2}$ which is several hundred km for ocean conditions (see Rhines 1975 and the discussion in Danilov and Gurarie 2004). Smaller scales are largely unaffected.

As far as subgrid closures are concerned, the framework of two-dimensional incompressible turbulence implies that the closures should be consistent with the k^{-3} power law in the enstrophy range. The extent to which this is possible with traditional closures is explored by Graham and Ringler (2013). The degree to which this is relevant to the dynamics of the real ocean remains an open question, for

dynamics at scales smaller than the internal Rossby radius develop an ageostrophic component.

5.3.2 Two-Layer Geostrophic Flows

In the presence of stratification, the situation becomes more complex. The general picture as presented in textbooks (see Salmon 1998; Vallis 2006) is as follows. On scales larger than the first internal Rossby radius L_d, there is a direct cascade of baroclinic (available potential plus kinetic) energy and an inverse cascade of barotropic (kinetic) energy. The baroclinic cascade is maintained through instabilities that release the available potential energy from an existing pool. It feeds the barotropic cascade at scales around L_d via the mechanism of baroclinic instability. This energy is transferred upscale where it is dissipated. On scales smaller than L_d, the layers interact only weakly and behave similarly to two-dimensional turbulence discussed above. In this regime, the dynamics are governed by the direct enstrophy cascade, implying the scaling exponent -3 for the modal or layer kinetic energy spectra. We note that this implies the presence of a direct energy cascade at these scales.

In this section, we discuss these concepts in the simplest possible setting, the two-layer quasigeostrophic model. It is essential that the two-layer model allows for a coupling between the eddy potential energy dynamics and the eddy kinetic energy. In this sense, it represents a minimum model for the real dynamics in ocean and atmosphere.

The two-layer QG model introduces important corrections to the single-layer situation explained in Section 5.3.1 above. First, it shows that the concept of spectrally localized forcing does not work, for the energy is supplied to the system over a broad range of scales, with the maximum spectral density of pumping shifted toward the scale of the energy spectrum maximum. Second, the notion of cascade has to be adjusted, for predictions are made for the baroclinic and barotropic energies, not for the layers.

For simplicity, we assume the layer depths are equal. The two-layer system can then be written as

$$\partial_t q_i + [\psi_i, q_i] = F_i + \delta_{2i}\, D_\mathrm{l}\psi_i + D_\mathrm{u}\psi_i + (-1)^{i+1}\,\kappa\,(\psi_1 - \psi_2)\,, \tag{5.28a}$$

$$q_i = f_0 + \beta y + \Delta\psi_i + (-1)^i\, k_\mathrm{d}^2\,(\psi_1 - \psi_2)/2\,, \tag{5.28b}$$

where $i \in \{1, 2\}$,

$$k_\mathrm{d} = \frac{1}{L_\mathrm{d}} = \frac{\sqrt{8}\,f_0}{N_0 H} \tag{5.29}$$

is the inverse of the baroclinic Rossby radius, f_0 is the Coriolis frequency, N_0 is the typical Brunt–Väisälä frequency, and H is the total fluid depth; see, e.g., Franzke et al. (2019) for details.

We remark that when diagnosed using the leading-order per-layer geostrophic balance relation, the difference in layer stream functions, $\psi_1 - \psi_2$, is proportional to the displacement of the interface between the layers. Thus, the last term in (5.28b) can be interpreted as the contribution to potential vorticity perturbations from the layer interface and is referred to as the stretching term.

We think of infrared dissipation D_i acting as bottom drag only on the lower layer. Then, $D_\mathrm{i} = -\lambda\Delta$ with λ the bottom drag coefficient. Ultraviolet dissipation is typically modeled by hyperviscosity of some order $n \geq 2$, so that $D_\mathrm{u} = \nu(-\Delta)^n$ with hyperviscosity coefficient ν. The last term in (5.28a) models thermal relaxation of the layer interface, with $2\kappa/k_\mathrm{d}^2$ the inverse timescale. It restores interface displacement and thus enters the layer equations with the opposite sign.

Although the ocean is mainly driven by wind stress applied to the upper layer, a theoretically simpler situation occurs when the interface between layers is relaxed toward a position with a uniform slope, i.e., taking $F_i = -(-1)^i \kappa U\, y$, with y the meridional coordinate. Equation (5.28a) in this case has an equilibrium solution $\psi_1 - \psi_2 = -Uy$, which in the presence of bottom drag implies $\psi_1 = -Uy$ and $\psi_2 = 0$.[9] The velocity U defines the vertical shear and interfacial slope in the two-layer QG model. This equilibrium solution corresponds to a pool of available potential energy (APE) and can be baroclinically unstable.

Splitting the stream functions into the equilibrium stream functions and perturbation or "eddy" stream functions ψ_1^eddy and ψ_2^eddy, we write

$$\psi_1 = -yU + \psi_1^\mathrm{eddy} \qquad \text{and} \qquad \psi_2 = \psi_2^\mathrm{eddy}\,. \tag{5.30}$$

Further, it is useful to rewrite the system in terms of the *eddy barotropic stream function* ψ and the *eddy baroclinic stream function* τ, respectively, defined by

$$\psi = \frac{\psi_1^\mathrm{eddy} + \psi_2^\mathrm{eddy}}{2} \qquad \text{and} \qquad \tau = \frac{\psi_1^\mathrm{eddy} - \psi_2^\mathrm{eddy}}{2}\,, \tag{5.31}$$

and the corresponding *eddy barotropic potential vorticity* q and *eddy baroclinic potential vorticity* ω defined as

$$q = \Delta\psi \qquad \text{and} \qquad \omega = \Delta\tau - k_\mathrm{d}^2\,\tau\,. \tag{5.32}$$

We note that the stretching term from (5.28b) appears as the second term in the definition of ω.

Substituting (5.30) into (5.28), writing out the sum and the difference of the layer equations, and rewriting all expressions in terms of the modal stream functions (5.31) and their associated potential vorticities, we obtain

[9] To see this, add a small deviation from the linear dependence in the y-direction and then consider the limit when this deviation vanishes.

$$\partial_t q + [\psi, q] + [\tau, \omega] + \frac{U}{2}\,\partial_x(q + \Delta\tau) + \beta\,\partial_x\psi = \frac{1}{2}\,D_\mathrm{i}(\psi - \tau) + D_\mathrm{u}\psi\,, \tag{5.33a}$$

$$\partial_t \omega + [\psi, \omega] + [\tau, q] + \frac{U}{2}\,\partial_x(\omega + q + k_\mathrm{d}^2\,\psi) + \beta\,\partial_x\tau = -\frac{1}{2}\,D_\mathrm{i}(\psi - \tau) + D_\mathrm{u}\tau + \kappa\tau\,. \tag{5.33b}$$

In the following, we will endow the perturbation quantities with doubly periodic boundary conditions. This is possible because the forcing terms, which are non-periodic in the y-direction, drop out of the equations for the perturbation quantities. However, the information on forcing is retained in the terms proportional to U.

The barotropic equation (5.33a) contains self-advection (i.e., the advection of barotropic eddy PV by the barotropic velocity field), whereas the baroclinic equation (5.33b) is linear in the baroclinic variables. Thus, barotropic dynamics are similar to two-dimensional vorticity dynamics characterized by an inverse energy cascade, whereas baroclinic dynamics are similar to the advection of a passive tracer which possesses a direct energy cascade.[10]

As in Section 5.3.1, we consider the modal energy balances for the barotropic (kinetic) energy

$$E^\psi = \sum_{\boldsymbol{k}\in\mathbb{Z}^2} E^\psi_{\boldsymbol{k}} = -\frac{1}{2}\sum_{\boldsymbol{k}\in\mathbb{Z}^2} \psi^*_{\boldsymbol{k}}\, q_{\boldsymbol{k}} \tag{5.34}$$

and baroclinic energy

$$E^\tau = \sum_{\boldsymbol{k}\in\mathbb{Z}^2} E^\tau_{\boldsymbol{k}} = -\frac{1}{2}\sum_{\boldsymbol{k}\in\mathbb{Z}^2} \tau^*_{\boldsymbol{k}}\,\omega_{\boldsymbol{k}} = \frac{1}{2}\sum_{\boldsymbol{k}\in\mathbb{Z}^2} (k^2 + k_\mathrm{d}^2)\,|\tau_{\boldsymbol{k}}|^2\,, \tag{5.35}$$

where the contribution prefactored by k^2 is baroclinic kinetic energy and the contribution prefactored by k_d^2 is available potential energy. Taking the Fourier transform of the barotropic and baroclinic equations, multiplying with $\psi^*_{\boldsymbol{k}}$ and $\tau^*_{\boldsymbol{k}}$, respectively, and taking the real part, we obtain

$$\partial_t E^\psi_{\boldsymbol{k}} = T^\psi_{\boldsymbol{k}} + C^\psi_{\boldsymbol{k}} + D^\psi_{\boldsymbol{k}}\,, \tag{5.36a}$$

$$\partial_t E^\tau_{\boldsymbol{k}} = T^\tau_{\boldsymbol{k}} + C^\tau_{\boldsymbol{k}} + G_{\boldsymbol{k}} + D^\tau_{\boldsymbol{k}}\,. \tag{5.36b}$$

The terms

$$T^\psi_{\boldsymbol{k}} = \Re[\psi^*_{\boldsymbol{k}}\, J_{\boldsymbol{k}}(\psi, q)]\,, \tag{5.37a}$$

$$T^\tau_{\boldsymbol{k}} = \Re[\tau^*_{\boldsymbol{k}}\, J_{\boldsymbol{k}}(\tau, q)] - k_\mathrm{d}^2\,\Re[\tau^*_{\boldsymbol{k}}\, J_{\boldsymbol{k}}(\psi, \tau)] \tag{5.37b}$$

with $\boldsymbol{k} = (k_x, k_y)$ describe energy transfer within the barotropic and baroclinic modes,

[10] Scott and Arbic (2007) show that there will be baroclinic self-interactions for unequal layers, leading to an inverse cascade of baroclinic kinetic energy.

$$C_{\boldsymbol{k}}^{\psi} = \Re[\psi_{\boldsymbol{k}}^{*}\, J_{\boldsymbol{k}}(\tau, \omega)] - \frac{U}{2}\, k^2\, \Re[\mathrm{i}k_x\, \psi_{\boldsymbol{k}}^{*}\, \tau_{\boldsymbol{k}}]\,, \tag{5.37c}$$

$$C_{\boldsymbol{k}}^{\tau} = \Re[\tau_{\boldsymbol{k}}^{*}\, J_{\boldsymbol{k}}(\psi, \Delta\tau)] - \frac{U}{2}\, k^2\, \Re[\mathrm{i}k_x\, \tau_{\boldsymbol{k}}^{*}\, \psi_{\boldsymbol{k}}] \tag{5.37d}$$

describe transfer from baroclinic to barotropic modes and vice versa, respectively,

$$G_{\boldsymbol{k}} = \frac{U}{2}\, k_{\mathrm{d}}^2\, \Re[\mathrm{i}k_x\, \tau_{\boldsymbol{k}}^{*}\, \psi_{\boldsymbol{k}}] \tag{5.37e}$$

represents the generation of energy, and all dissipative terms are subsumed into $D_{\boldsymbol{k}}^{\psi}$ and $D_{\boldsymbol{k}}^{\tau}$.

One can readily see that the generation term is proportional to the meridional buoyancy flux which tends to level off the layer interface (for APE has to be released) if the system is baroclinically unstable. In this case, its mean value has to be positive definite in a statistically stationary sense. Note that $G_{\boldsymbol{k}}$ is defined by the dynamics and is not an external parameter as in 2D barotropic turbulence theory.

Since

$$\sum_{\boldsymbol{k}\in\mathbb{Z}^2} T_{\boldsymbol{k}}^{\psi} = \sum_{\boldsymbol{k}\in\mathbb{Z}^2} T_{\boldsymbol{k}}^{\tau} = 0\,, \tag{5.38}$$

these two terms only redistribute energy between scales. Likewise,

$$\sum_{\boldsymbol{k}\in\mathbb{Z}^2} C_{\boldsymbol{k}}^{\psi} = -\sum_{\boldsymbol{k}\in\mathbb{Z}^2} C_{\boldsymbol{k}}^{\tau}\,, \tag{5.39}$$

so that these terms only redistribute energy between baroclinic and barotropic modes.

In the traditional view of baroclinic turbulence (Rhines 1977; Salmon 1980), one introduces spectral energy fluxes analogous to (5.25),

$$\Pi^{\psi}(k) = -\int_0^k T^{\psi}(\kappa)\,\mathrm{d}\kappa \qquad \text{and} \qquad \Pi^{\tau}(k) = -\int_0^k T^{\tau}(\kappa)\,\mathrm{d}\kappa\,, \tag{5.40}$$

describing the redistribution of energy between scales. There are numerous publications discussing the behavior of fluxes in this situation (e.g., Scott and Arbic 2007). The barotropic flux Π^{ψ} can be shown to be negative at $k < k_{\mathrm{d}}$ corresponding to an inverse cascade of barotropic energy, while the baroclinic flux Π^{τ} is always positive corresponding to a direct cascade of full (i.e., potential and kinetic) baroclinic energy. Although there is an upscale (i.e., toward large scales) transfer of barotropic kinetic energy, there is no inertial range at $k < k_{\mathrm{d}}$ because the transfer of energy from the baroclinic into barotropic mode is spread over all wavenumbers, being stronger at smaller k.

Thus, no spectral law can be predicted for the inverse cascade in this case. In contrast, on scales smaller than L_{d} the stretching term in the expression for the

quasigeostrophic potential vorticity becomes small compared to the relative vorticity and, as already mentioned, each layer behaves as in two dimensions implying the scaling exponent -3 for the kinetic energy.

This picture relies on the fact that the assumed forcing maintains a pool of available potential energy which is then transferred to eddies through baroclinic instability, which develops into a nonlinear regime of quasistationary balance between the release of potential energy, nonlinear transfer, and dissipation. In general, forcing will drive both barotropic and baroclinic components of the mean flow. But even if forcing is only baroclinic, as is the case here, a mean barotropic flow is created in the presence of friction and/or topography. For uniform shear, the release of APE through baroclinic instability is the main source of energy for the eddies, but the kinetic energy of the mean flow may also be important in general.

The picture described so far is tied to the choice of writing the fields in terms of barotropic and baroclinic modes. Arguments will differ when looking at the transfer of energy between layers or between kinetic and potential energy. In particular, the sum of transfers between modes is zero *only* when integrated over wavenumbers. This explains why the picture of transfers will be modified if considered for layers (there will be transfers between the layers) and for the total energy (when baroclinic and barotropic kinetic energies will be combined, and potential energy split off the baroclinic energy).

The total energy at wave vector $\boldsymbol{k}$,

$$E_{\boldsymbol{k}} = \frac{1}{4}\, k^2 \left(|\psi_1|^2_{\boldsymbol{k}} + |\psi_2|^2_{\boldsymbol{k}}\right) + \frac{1}{2}\, k_\mathrm{d}^2\, |\tau|^2_{\boldsymbol{k}} \tag{5.41}$$

has a rate of change

$$\partial_t E_{\boldsymbol{k}} = T^K_{\boldsymbol{k}} + T^P_{\boldsymbol{k}} + G_{\boldsymbol{k}} + D^\mathrm{i}_{\boldsymbol{k}} + D^\mathrm{u}_{\boldsymbol{k}} \tag{5.42}$$

with transfer rates

$$T^K_{\boldsymbol{k}} = \frac{1}{2}\, \Re\left[\psi^*_{1\boldsymbol{k}}\, J_{\boldsymbol{k}}(\psi_1, \Delta\psi_1) + \psi^*_{2\boldsymbol{k}}\, J_{\boldsymbol{k}}(\psi_2, \Delta\psi_2)\right], \tag{5.43a}$$

$$T^P_{\boldsymbol{k}} = -k_\mathrm{d}^2\, \Re\left[\tau^*_{\boldsymbol{k}}\, J_{\boldsymbol{k}}(\psi, \tau)\right], \tag{5.43b}$$

a generation term $G_{\boldsymbol{k}}$ as before, and rates of frictional (infrared) dissipation $D^\mathrm{i}_{\boldsymbol{k}}$ and viscous (ultraviolet) dissipation $D^\mathrm{u}_{\boldsymbol{k}}$.

Intermodal or interlayer transfers are now included in the kinetic and potential energy transfers. The emerging picture is perhaps the most transparent; see Figure 5.2. It shows that generation is nearly compensated by large-scale dissipation, that the EPE flux is direct, for it takes the generated eddy energy $G_{\boldsymbol{k}}$ and carries it to larger wavenumbers gradually releasing it to kinetic energy, and that the EKE flux is inverse, for it takes the released potential energy and carries it back to the interval of small wavenumbers where it is dissipated. It is important to note that transfers into EKE and from EPE are centered at k_d and occupy at least one octave of wavenumbers more on

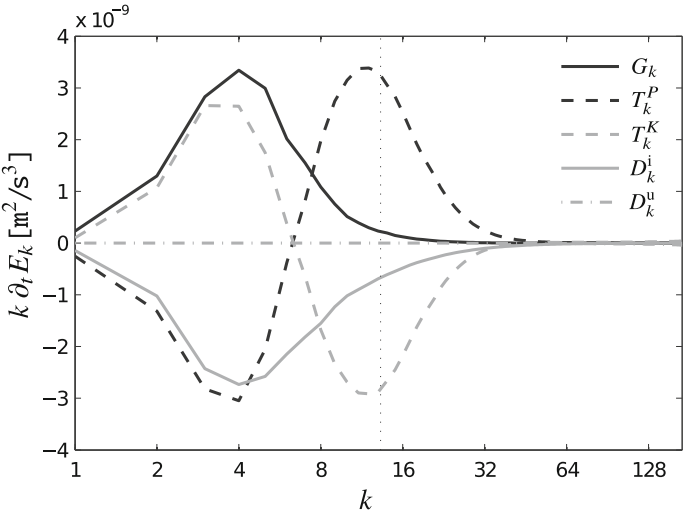

Fig. 5.2 Spectral energy fluxes corresponding to the fluxes in (5.42), integrated over the wavenumber shell $|\boldsymbol{k}| = k$. Figure adapted from Jansen and Held (2014), their Figure 4. Note that the vertical axis shows $k\,\partial_t E_k$, so that area under the curve in singly logarithmic scaling corresponds to total transfer rates. Note further that the scale of wavenumbers k shown is normalized by $2\pi/L$, where L is the domain size. The deformation scale $k_\mathrm{d} = 1/L_\mathrm{d}$ is marked by the vertical dotted line

the short-wave side. Simulations by Jansen and Held (2014) demonstrate a spectrum of barotropic EKE close to but steeper than -3 starting from k_d and toward larger wavenumbers. Yet, a substantial part of the interval where this spectrum is observed is where the transfers take place, i.e., where there cannot be an inertial range. In other words, the existence of a well-defined spectral slope is not an indicator of an inertial range, which is frequently forgotten.

Although the theoretical prediction of the inverse cascade is formally made for the barotropic kinetic energy, it is commonly observed for baroclinic kinetic energy and for layer kinetic energies. This behavior is clarified by Scott and Arbic (2007).

We see that if there is a hope for the interval of self-similar behavior in layer QG dynamics, such behavior should be on the side of small wavenumbers and be consistent with the -3 spectral law. However, the two-layer setup indicates very clearly that the transfer from EPE to EKE involves wavenumbers around k_d or larger. For this reason, this spectral law and self-similar behavior of inertial range can only be expected to hold for wavenumbers essentially larger than k_d, which come too close to the scales where ageostrophy is important in the real ocean. The wavelength associated with $k_\mathrm{d} = 1/L_\mathrm{d}$ is $2\pi L_\mathrm{d}$. On meshes with spacing $a = L_\mathrm{d}$, this wavelength is well resolved, but this extra resolution is just needed to accommodate for spectral exchanges between EPE and EKE. In practice, in ocean circulation models, the resolution of k_d^{-1} is not always (or not everywhere) achieved. In this case, eddy dynamics may suffer not only from excessive subgrid dissipation but also from the mere fact that the interval where EPE has to feed EKE is too short. The spectral

interval where most of the generation (conversion from the available potential energy to the EPE) takes place tends to be at wavenumbers smaller than k_d. Yet, as shown by Jansen and Held (2014), the generation turns out to be sensitive to dissipation in the vicinity of k_d. We propose that the ability of subgrid closures to least interfere with energy generation presents a convenient guiding principle in these cases.

An important parameterization for relatively coarse, non-eddy-permitting ocean simulations was introduced by Gent and McWilliams (1990); it is now known as the Gent–McWilliams parameterization. Here, we explain the idea in the context of the two-layer model (5.28). On scales larger than L_d, the relative vorticity is expected to be small compared to the stretching term, the last term in (5.28b), which models perturbations of the layer interface. Correspondingly, the dominant nonlinear contribution to (5.28a) is the divergence of the thickness flux

$$(-1)^i\, k_\mathrm{d}^2\, [\psi_i^{\mathrm{eddy}},\, \tfrac{1}{2}\,(\psi_1^{\mathrm{eddy}} - \psi_2^{\mathrm{eddy}})] = (-1)^i\, k_\mathrm{d}^2\, [\psi, \tau] = (-1)^i\, k_\mathrm{d}^2\, \boldsymbol{\nabla} \cdot (\tau\, \boldsymbol{\nabla}^\perp \psi)\,. \tag{5.44}$$

This term will be very small if mesoscale eddies are not well resolved. The proposal of Gent and McWilliams (1990) amounts to adding a flux divergence of the form

$$\mathscr{F}_i = (-1)^i\, k_\mathrm{d}^2\, \boldsymbol{\nabla} \cdot (\varkappa\, \boldsymbol{\nabla} \tau) \tag{5.45}$$

to the right-hand sides of the two-layer equations as a parameterization for the effect of unresolved eddies on the resolved flow. The coefficient $\varkappa$ is sometimes taken constant, more frequently, however, selected as a polynomial of the velocity difference $U = |\boldsymbol{u}_1 - \boldsymbol{u}_2|$ based on qualitative theory where degree and the coefficients of the polynomial are chosen empirically. In this case, $\varkappa$ is a measure of vertical instability in the system; see, e.g., Stone (1972), Cessi (2008), and Held and Larichev (1996).

By construction, the $\mathscr{F}_i$ model only the subgrid layer thickness flux, not the full potential vorticity flux. They provide a sink for potential energy, thus emulating the effect of baroclinic instability on the potential energy balance in a model that is too coarse to resolve this process directly. This technique prevents the buildup of an unlimited pool of available potential energy, but does not model Reynolds stresses nor does it feed energy back into the pool of resolved eddy kinetic energy.

Note that while (5.45) looks like diffusion, it acts on the *layer thickness*. Whenever the interface between two layers is inclined, thickness diffusion means that it will be leveling the interface. This implies that fluid will move in opposite direction in the layers, showing that the Gent–McWilliams parameterization creates a circulation that tends to flatten isopycnals. Thus, while thickness diffusion proceeds in two dimensions, the generated fluid motion is three-dimensional, and it is advective. See Gent (2011) for a detailed discussion.

5.3.3 Continuously Stratified and Surface QG Dynamics

Even within the quasigeostrophic family of models, the picture presented so far is not the end of the story. First, when allowing for continuous stratification, there are many baroclinic vertical modes. Second, there are surface-trapped motions that can be understood in the framework of surface geostrophic dynamics (SQG); see, e.g., a discussion and further references in von Storch et al. (2019).

For simplicity, we consider the three-dimensional QG equations on a layer of uniform depth H with rigid lid upper boundary condition at $z = 0$. The model then reads

$$\partial_t q + [\psi, q] = 0\,, \tag{5.46a}$$

$$q = f + \Delta_{\mathrm{h}}\psi + f_0^2\,\partial_z \frac{\partial_z \psi}{N^2(z)} \tag{5.46b}$$

where Δ_{h} denotes the horizontal Laplacian and brackets, as before, the horizontal Jacobian, with boundary conditions for the buoyancy b at $z = 0, -H$:

$$\partial_t b + [\psi, b] = 0\,, \tag{5.47a}$$

$$b = f_0\,\partial_z \psi\,. \tag{5.47b}$$

According to Wunsch (1997), the bulk of ocean kinetic energy is well captured by the barotropic and first baroclinic modes. For this reason, the major conclusion regarding the spectral slope -3 of the direct enstrophy cascade remains valid for the bulk of the ocean. However, the standard basis for vertical modes, as given by the eigenvalue problem

$$f_0^2\,\partial_z \frac{\partial_z \Psi_n(z)}{N^2(z)} + \lambda_n^2\,\Psi_n(z) = 0 \tag{5.48}$$

with zero boundary conditions for $\partial_z \Psi$ at $z = 0, -H$, does not take into account surface buoyancy perturbations. Baroclinic instabilities evolving as solutions of (5.46) deal with the modes of the full operator that satisfy the boundary conditions (5.47), and cannot be understood in the frame of the standard basis. Certain textbook instabilities, for example the Eady problem (see, e.g., Vallis 2006), rely entirely on surface-trapped dynamics.

Even though it is possible to reformulate the surface dynamics as δ-sheets of potential vorticity, such solutions cannot be represented in terms of the vertical eigenmode basis. Layered models, however, include the surface dynamics in their upper and lower layer potential vorticity. Therefore, the two-layer model described in Section 5.3.2 above cannot separate surface-driven instabilities from interior instability mechanisms, and the simplest model where this can be explored is the three-layer model studied in Badin (2014).

To separate the surface dynamics from the interior in the continuously stratified QG model, one considers the case where $q = \text{const}$ in an infinitely deep layer, so that only surface dynamics remains. Then, the horizontal Fourier coefficients of the stream function ψ representing the surface buoyancy perturbation satisfy

$$f_0^2 \, \partial_z \frac{\partial_z \psi_{\boldsymbol{k}}(z)}{N^2(z)} - k^2 \, \psi_{\boldsymbol{k}}(z) = 0 \tag{5.49}$$

with non-homogeneous Neumann conditions at the top and decay toward infinite depth. When $N = \text{const}$, the $\psi_{\boldsymbol{k}}$ decay with depth as $\exp(kNz/f_0)$; i.e., they decay on a vertical scale

$$H \gtrsim \frac{f_0}{kN} \, . \tag{5.50}$$

Correspondingly, in a uniformly stratified layer of depth H, only surface perturbations larger in size than the first baroclinic Rossby radius may reach through the fluid depth.

In the absence of forcing and dissipation, surface dynamics will preserve integrals of buoyancy variance and the product ψb. The latter leads to an inverse cascade at large scales with a -1 spectrum of surface kinetic energy, and the former leads to a direct buoyancy variance cascade with a $-5/3$ spectrum (see, e.g., Smith et al. 2002). Note that the prediction concerns surface kinetic energy and is valid for uniform stratification. Since in this case $|\nabla \psi_{\boldsymbol{k}}| = k \, |\psi_{\boldsymbol{k}}| \sim |b_{\boldsymbol{k}}|$, we expect the same spectral law for kinetic energy

$$E_{\text{EKE}} = \frac{1}{4} \sum_{\boldsymbol{k}} |\nabla \psi_{\boldsymbol{k}}|^2 \tag{5.51}$$

and available potential energy

$$E_{\text{APE}} = \frac{1}{4N^2} \sum_{\boldsymbol{k}} |b_{\boldsymbol{k}}|^2 \, . \tag{5.52}$$

Small scales do not penetrate deep, and spectra become steeper. Furthermore, they are modified by stratification; examples including exponential stratification and the case of a mixed upper layer are discussed in Callies and Ferrari (2013).

Instabilities in the real ocean project on both deep-ocean modes and surface modes, and depend on the structure of PV of the basic ocean state. Surface dynamics are expected to be an important contributor at locations where the interior PV gradients are weak. Since shallow surface modes are excited as a result of evolving instability, the transfer of available potential energy into eddy kinetic energy is not spectrally local. This implies that the argument in favor of precisely the -1 or $-5/3$ spectral slope at the surface is rather weak. However, it is appropriate to expect that spectral laws for near-surface velocities at small scales are shallower than the -3 prediction for the enstrophy range.

5.3.4 Ocean Models and Observational Evidence

To study ocean turbulence beyond the idealized models mentioned before, we must turn to numerical studies of the primitive equations, full ocean circulation models, and observational evidence. In this context, mesoscale or submesoscale structures which become ageostrophic in high-resolution models are of particular interest. In the following, we review a few studies which highlight these issues with the understanding that this selection is far from being complete or representative.

To begin, the recent interest in surface quasigeostrophic (SQG) dynamics was triggered by the observation that energy spectra of surface geostrophic velocities derived from altimetric data are noticeably shallower at many locations than spectra predicted by the theory of QG turbulence (Lapeyre 2009). High-resolution simulations also lend support to the relevance of the SQG concept for understanding the simulated behavior and observations. However, at scales about the first baroclinic Rossby radius and smaller, in real situations as well as in high-resolution simulations with full primitive equations, surface quasigeostrophic dynamics are accompanied by frontal and mixed-layer instabilities which deviate from geostrophy. Klein et al. (2008) and Capet et al. (2008) analyze the near-surface dynamics in ocean simulations performed at the resolution of 2 km and down to 0.75 km, respectively, and demonstrate that there is a close resemblance to SQG dynamics. The spectra of surface kinetic energy have a slope close to -2 from the spectral maximum to the spectral cutoff at large k. This is much shallower than the slope predicted by quasigeostrophic theory. The conceptual difference to SQG is, however, that the Rossby numbers of eddies at these scales are no longer small and a substantial ageostrophic flow component is generated, which modifies the turbulent energy fluxes. The presence of frontal and mixed-layer instabilities implies that the transfer of available potential energy into kinetic energy continues at rather small scales associated with these instabilities. Nevertheless, the near-surface velocities are nearly in geostrophic balance and the ageostrophic components explain only a small fraction of kinetic energy, only visible close to the high-wavenumber spectral end. Despite their smallness with respect to the dominant rotational component (computed via the Helmholtz decomposition), they are responsible for the downscale cascade of the total eddy kinetic energy. The cascade of the dominant rotational component of the velocity behaves differently: It is upscale and of smaller amplitude than the cascade of full velocity. The fact that it is upscale is perhaps not surprising: As there are transfers from the available potential energy to kinetic energy, as in QG or SQG turbulence, the flux of rotational kinetic energy proceeds to larger scales from the scale of forcing.

Callies and Ferrari (2013) discuss existing views and assess two data sets shedding light on the behavior of ocean submesoscales. They consider scales from about 200 to 1 km. For a site in the Gulf Stream, they found steep (-3) spectra of kinetic energy for scales between 200 and 20 km, and shallower spectra at smaller scales consistent with the -2 slope of the internal gravity wave spectrum. For a site in the North Pacific, they report shallower spectra whose behavior with depth, however, does not

agree with the prediction of SQG. It is proposed that the gravity wave continuum and unbalanced motions can contribute to this behavior.

To summarize, the range of submesoscale, where the subgrid scales of eddy-permitting (and eddy-resolving) ocean circulation models are located, combines features of QG and SQG turbulence but also includes ageostrophic (unbalanced) motions, depending on mesh resolution and ocean stratification. Wavenumbers larger than k_d are those of the forward cascade of EKE, but the inverse cascade can be present for the rotational component of EKE at even smaller scales if small-scale instabilities continue to transfer the available potential energy to EKE. No true slope prediction can be made on scales around k_d because of intermodal exchanges and spectrally spread dissipation.

Inertial ranges may emerge on the side of smaller scales on very high-resolution meshes, but even there one should expect a dependence on the depth and a contribution from submesoscale (frontal) instabilities. So even when they emerge, inertial ranges may deviate from the predictions of QG turbulence because of a forward energy cascade. The dominance of the rotational velocities in the energy spectra does not imply their dominance in energy transfers at large wavenumbers. One may try to draw a certain analogy with the $-5/3$ spectrum observed in the atmosphere between 500 and 10 km, which is that of stratified turbulence with forward energy cascade; see the bibliography, discussion, and analysis of high-resolution simulations in Augier and Lindborg (2013). On larger scales, it matches the dynamics predicted by QG theory.

At present, resolutions in ocean circulation models are such that the near-subgrid scales are in a range where self-similar behavior is unlikely. Subgrid closures can therefore not be universal in the range of resolutions about the Rossby radius. Hence, perhaps the guiding principle should be that of minimizing their damping effect on the rate at which energy is released from the pool of APE and the KE of the background state to the EKE at the resolved scales.

5.4 Energy Backscatter

Although most ocean circulation models used for climate research are coarse, the number of eddy-permitting models is increasing and will dominate in the future. Such models simulate eddy dynamics, but cannot resolve it fully for they suffer from overdissipation. Its origin can be explained as follows. In order to remove the variance of velocity and enstrophy at grid scales for numerical stability, they use harmonic or biharmonic dissipative subgrid viscosity operators (Fox-Kemper and Menemenlis 2008). Together with removing the grid-scale noise, these operators also dissipate energy at adjacent scales, which are in this case the scales close to the internal Rossby radius. As we have seen, these scales host exchanges between the potential and kinetic energy compartments and also determine the eddy energy release from the available potential energy. Their overdissipation is the reason why eddy-permitting flows seldom reach the observed levels of eddy kinetic energy.

The problem of overdissipation and, in more broad context, of subgrid closure that takes into account the existence of unresolved scales has been known in atmospheric sciences for a long time. First papers on this issue appeared almost simultaneously with the KLB concept of two-dimensional turbulence; see Leith (1971) and the discussion in Frederiksen and Davies (1997). It may be explained within the spectral picture of triad interactions in two-dimensional turbulence as follows: Since we can only resolve wavenumbers up to some $k_{\max}$ numerically, it is clear that we miss not only spectrally local interactions responsible for the enstrophy transfer through the boundary at $k = k_{\max}$, which lead to a net energy drain and hence behave as a form of dissipation in the ensemble mean, but also non-local triads, having two legs at $k > k_{\max}$ or on both sides of $k_{\max}$ and one leg at large scales $k_{\mathrm{LS}} \ll k_{\max}$, which might force the resolved scales. These interactions are termed *backscatter*. It is not represented by the usual dissipative subgrid operators, which is the main cause of overdissipation in conventional models. A fully deterministic representation of backscatter is impossible as the details of the state of the subgrid are in principle not available. Thus, the best we can hope for is some stochastic model of backscatter.

Theoretical developments in this direction assume as a rule QG dynamics, periodic boundary conditions or spherical geometry, ensemble averaging, and "spectral language" to come up with parameterizations. As an example, we mention the work by Kitsios et al. (2013) who derive both stochastic and deterministic closures by comparing truncated and high-resolution dynamics in a two-layer QG setup on the sphere. It is believed that both types can be equally skillful, for, in any case, useful formulas rely on ensemble averaging and thus do not describe realizations.

Although the turbulent dynamics dictates that drain and backscatter should be described as stochastic processes, additional issues such as numerical stability have to be taken into account. For deterministic parameterizations, the resulting expressions contain powers of the Laplacian, sometimes going beyond the biharmonic one. Their study shows that even in the context of two-layer QG turbulence which is statistically homogeneous in the zonal direction, the final parameterizations of drain and backscatter depend not only on k_{d} and $k_{\max}$, but also on the extent of the energy-containing range.

It will be much more difficult to propose parameterizations for domains with horizontal boundaries where spectral language and zonal homogeneity are missing. In addition, all backscatter parameterizations raise the question of numerical stability due to the effective negative viscosity of the terms providing backscatter. Finally, in addition to eddy–eddy interactions considered by Kitsios et al. (2013), contributions may come from interactions involving the unresolved mean field (in the sense of time averages) component. These examples show that progress is possible, but we can hardly expect universally valid solutions.

In the following, we review two specific backscatter parameterizations in detail. The first, due to Jansen and Held (2014), is based on a very straightforward scalar model for the subgrid energy. The second, due to Grooms and Majda (2014), uses a more sophisticated linear model for the subgrid dynamics.

5.4.1 Models with Scalar Subgrid Energy Budget

Since, as mentioned above, comprehensive first-principle models are necessarily complex, we think that simplified implementations of energy backscatter proposed by Jansen and Held (2014) and Jansen et al. (2015), who consider kinetic energy backscatter for QG and primitive equations, respectively, deserve attention. These parameterizations do not aim at mimicking missing interactions with subgrid scales, but seek instead to compensate for the overdissipated energy, which is much easier. Importantly, the amount of energy returned through the proposed backscatter parameterization can be controlled, which is a prerequisite for stability of the algorithm.

Jansen and Held (2014) study the two-layer quasigeostrophic equations with the Leith parameterization as nonlinear small-scale dissipation operator. In each layer i,

$$D_{\mathrm{u}}\psi_i = -\Delta(\nu_i\,\Delta^2\psi_i) \qquad \text{with} \qquad \nu_i = C_L\,a^6\,|\Delta^2\psi_i|\,, \tag{5.53}$$

where a denotes the grid-spacing and $C_L = 0.005$. The associated overall rate of viscous dissipation at wavenumber $\boldsymbol{k}$ is

$$V_{\boldsymbol{k}} = \frac{1}{2}\sum_{i\in\{1,2\}} k^2\,(\psi_i)^*_{\boldsymbol{k}}\,(\nu_i\,\Delta^2\psi_i)_{\boldsymbol{k}}\,. \tag{5.54}$$

(The layers are assumed to be of equal thickness, hence the additional factor of $1/2$ in the expression for $V_{\boldsymbol{k}}$ and in similar expressions below.) The rate of frictional dissipation in the bottom layer at wavenumber $\boldsymbol{k}$ is

$$F_{\boldsymbol{k}} = \tfrac{1}{2}\,\lambda\,k^2\,|\psi_2|^2_{\boldsymbol{k}}\,. \tag{5.55}$$

Summing over wavenumbers, the total rate of ultraviolet and infrared dissipation is

$$V = \sum_{\boldsymbol{k}} V_{\boldsymbol{k}} \qquad \text{and} \qquad F = \sum_{\boldsymbol{k}} F_{\boldsymbol{k}} \tag{5.56}$$

Since transfers are summed to zero, the overall balance of energy is

$$\partial_t E = G - F - V\,, \tag{5.57}$$

where G is the generation term with Fourier representation (5.37e).

To compensate for the excessive dissipation at small scales, the simplest model is to add an energy source that returns energy at a rate

$$S = (1-\varepsilon)\,V \tag{5.58}$$

so that all but a small fraction $\varepsilon \approx 0.1$ representing the physical rate of ultraviolet dissipation is returned. Jansen and Held (2014) tested two different models for this

source, one deterministic and the other stochastic. In the deterministic version, each layer potential vorticity equation is given a source term

$$s_i = -A(t)\,\Delta^2 \psi_i \tag{5.59}$$

which corresponds to negative Laplacian viscosity in the momentum equations. The amplitude $A(t)$ is set by the condition that the constraint (5.58) is satisfied at every instance in time. Since the Laplacian is less scale-selective than the biharmonic ultraviolet dissipation, energy will be returned at larger scales than those at which it is dissipated.

The second implementation is stochastic, with

$$s_i = A(t)^{1/2}\,\eta(\boldsymbol{x}, t)\,, \tag{5.60}$$

where the η is Gaussian noise, δ-correlated in space and time. The forcing is kept barotropic; i.e., the same noise process is used for both layers, to replenish the inverse cascade of barotropic kinetic energy. In this case, the ensemble mean $\langle S \rangle$ will be proportional to $A(t)$ so that the amplitude can be found at each time step from the constraint (5.58). Of course, (5.58) is satisfied only in the ensemble mean. However, it is also approximately satisfied for each realization as the stochastic forcing is distributed over a large number of spatial locations of the computational grid. The rate of backscatter energy pumping at mode $\boldsymbol{k}$ is given by

$$S_{\boldsymbol{k}} = \frac{1}{2} \sum_{i \in \{1,2\}} (\psi_i)^*_{\boldsymbol{k}}\,(s_i)_{\boldsymbol{k}}\,. \tag{5.61}$$

Thus, even when the s_i have a white noise spectrum, energy backscatter is biased toward the scales with already high energy content. In practice, this involves scales larger than those of $V_{\boldsymbol{k}}$.

Jansen and Held (2014) conclude that both parameterizations work rather similarly; however, the stochastic implementation returns energy over a broader interval of wavenumbers. The principal question here, namely how much energy has to be returned and where it should be returned, is left without answer and presents a topic for future research. The amplitude of backscatter is selected globally, which is only appropriate if flow energy is distributed uniformly. In the general case, one needs a local criterion.

A small variation of the kinematic backscatter assumption (5.58) is the introduction of a dynamic global subgrid energy budget E_{sg} via

$$\dot{E}_{\mathrm{sg}} = V - S - 2\lambda\gamma\, E_{\mathrm{sg}}\,. \tag{5.62}$$

The last term represents dissipation of subgrid energy by bottom friction, where the parameter γ is the fraction of subgrid energy residing in the lower layer and λ the

bottom drag coefficient.[11] This form of a global subgrid energy reservoir is suggested by Jansen and Held (2014) as a motivation to justify assumption (5.58), but could also be used computationally by assuming that the amplitude of backscatter $A(t)$ is proportional to the total subgrid energy in the reservoir.

At the next level of complexity, one may use a local subgrid energy budget. Jansen et al. (2015) suggest a budget for the subgrid energy density e of the form

$$\partial_t e = v - s - \nabla \cdot \boldsymbol{F} - d \,, \tag{5.63}$$

where v is the rate of viscous dissipation per unit volume of the resolved scales, s is the rate of backscatter per unit volume, $\boldsymbol{F}$ is the flux redistributing subgrid energy, and d is the rate of dissipation of subgrid energy per unit volume. Each of these terms must be modeled. Jansen et al. (2015) assume biharmonic ultraviolet dissipation

$$v = \frac{1}{H} \sum_i h_i \, \nu_i \, |\Delta \boldsymbol{u}_i|^2 \,, \tag{5.64}$$

where H is the total depth, h_i are the layer depths, and ν_i are the layer horizontal biharmonic viscosity coefficients, assumed positive. In this setting, all operators act in the horizontal only. For the backscatter source, one can take harmonic viscosity so that

$$s = -\frac{1}{H} \sum_i h_i \, \nu_{\text{bs}} \, |\nabla \boldsymbol{u}_i|^2 \tag{5.65}$$

with *negative* coefficient of viscosity

$$\nu_{\text{bs}} = -C_{\text{bs}} \, a \left(\max\{2e, 0\} \right)^{\frac{1}{2}} \tag{5.66}$$

with C_{bs} an order-one constant. If the energy to be scattered back becomes too large, e becomes negative and backscatter viscosity goes to zero. This controls the amount of energy returned back.

A major point for discussion is the choice of flux $\boldsymbol{F}$ for the subgrid energy. Jansen et al. (2015) choose the purely diffusive flux

$$\boldsymbol{F} = -k_{\text{sg}} \nabla e \,, \tag{5.67}$$

where k_{sg} is an appropriately selected constant of diffusivity. This choice is guided by the observation that the transfer from and to the subgrid can be very spatially rough so that a mechanism is needed to regularize the distribution of e horizontally.

[11] In a more realistic setting, one may wish to also model the dissipation of the fraction of subgrid energy that cascades downscale and turns into fully three-dimensional turbulence. This intention is implicit in the closure condition (5.58).

However, the question arises whether subgrid energy should not perhaps be advected by the resolved flow or be subject to some other non-local mechanism of transfer.

Finally, the dissipation rate d in (5.63) is typically small and may be neglected.

It turns out that the backscatter parameterizations by Jansen and Held (2014) and Jansen et al. (2015) lead to noticeable improvements even in situations where non-trivial bottom topography is present, and allow the mesoscale eddy dynamics in eddy-permitting simulations to approach those of high-resolution runs. On a qualitative level, the success of these simple implementations of backscatter rests on the idea that energy needs to be scattered back only in places where it is strongly dissipated. Although the vertically averaged or basin-averaged subgrid kinetic energy balance used to assess the backscatter viscosity presents an oversimplification, the energy scattered back is nevertheless modulated by the distribution of resolved energy. This also implies that the parameterization may bring improvements only in situations when an eddy-permitting model already correctly simulates the pattern of kinetic energy distribution but lacks amplitude. In realistic applications resolving the vertical structure with many more layers, the vertical distribution of backscatter viscosity may matter, since surface-trapped modes may exhibit more vertical structure, but this remains to be seen. A theory of where to return the energy scattered back presents an interesting question for further research too. Clearly, with only the harmonic operator at one's disposal, the deterministic backscatter parameterization has limited capabilities so that stochastic closures may still be needed. Furthermore, a missing point is the cascade of EPE which is dissipated too by subgrid diffusive closures or through upwind transport algorithms. Too diffusive transport schemes may result in the reduced transfer between EPE and EKE, so that the role of subgrid closures in the tracer equations should be explored. Conversely, Ilıcak et al. (2012) show that spurious mixing of transport algorithms depends on velocity variance at grid scales, so that energizing these scales above a certain level is not recommended. This issue is further explored in Klingbeil et al. (2019). This set of questions shows that even in the context of energy backscatter, the problem is far from being resolved and new ideas are required.

To apply these ideas to the full primitive equations, we face further questions. To our knowledge, this has not been pursued exhaustively, and we can only sketch a direction; more theoretical analysis and numerical experiments are needed here. To start, we may localize even further, treating the subgrid energy density e as a full three-dimensional field, so that the evolution equation (5.63) now takes the form

$$\partial_t e = v - s - \boldsymbol{\nabla}_\mathrm{h} \cdot \boldsymbol{F}_\mathrm{h} - \partial_z F_z - d \,, \tag{5.68}$$

where $\boldsymbol{F}$ is the flux redistributing the subgrid energy, taken as $\boldsymbol{F}_h = -K_\mathrm{h} \boldsymbol{\nabla}_h e$ and $F_z = -K_z \, \partial_z e$, where K_h and K_z are appropriately selected horizontal and vertical diffusion coefficients. As before, d is the rate of dissipation of subgrid energy; it is small and may be neglected. This approach is more expensive, for now the evolution equation has to be integrated in three spatial dimensions.

The contribution from ultraviolet dissipation now takes the form

$$v = -\nu_{\text{visc}} \left| \Delta_{\text{h}} \boldsymbol{u} \right|^2 , \tag{5.69}$$

where ν_{visc} is the coefficient of horizontal biharmonic viscosity. Vertical viscosity in the momentum equation would generally be provided by a vertical mixing parameterization which relies on some physics and empirical data, e.g., using a KPP closure (Large et al. 1994) or k-ε closure (Umlauf and Burchard 2003). The corresponding backscatter source term reads

$$s = -\nu_{\text{bs}} \left| \nabla \boldsymbol{u} \right|^2 \tag{5.70}$$

where ν_{bs} is again given by an expression of the form (5.66).

One may consider stochastic implementation options for the backscatter source. A caveat here is that for the primitive equations, the source must respect the divergence condition. This could be done by a simple projection. Another possibility is to write the horizontal source in the form

$$\boldsymbol{s} = \nabla \times (\Psi \boldsymbol{k}) , \tag{5.71}$$

with $\Psi(\boldsymbol{x}, t) = P(\boldsymbol{x})\, \eta(\boldsymbol{x}, t)\, A(t)$. Here, P is a spatial pattern of eddy kinetic energy (which may be modeled, inferred from high-resolution simulations, or taken from observations), $A(t)$ is the amplitude (selected to ensure subgrid energy balance), and η is a random field generated, for example, by a Markov process. Despite the presence of a differential operator, one has to introduce correlations in time and space to ensure that the resulting forcing is smooth.

Another issue is that, for the primitive equations, the concept of backscatter relates in equal measure to the momentum and to the tracer equations. Compensation for kinetic energy overdissipation is not necessarily sufficient if tracer variance is overdissipated. In principle, an approach resembling the one applied to the quasi-geostrophic potential vorticity equations can be proposed. However, there are some technical difficulties. First, in many cases dissipation is already built into the implementation of the transport operators and cannot be easily accessed. Second, even if it is not, biharmonic operators are not always available for tracers.[12] Finally, the production of tracer variance and the production of kinetic energy are linked, so that additional theoretical analysis is required.

5.4.2 *Stochastic Superparameterizations*

While the backscatter approximation by Jansen and Held (2014) and Jansen et al. (2015) seeks to return excessively dissipated energy back to the main flow, it needs

[12]For tracers, one commonly uses rotated operators that mix along isopycnals. Their biharmonic versions are more expensive and less stable. Furthermore, such operators do not directly dissipate potential energy.

an eddy-permitting model that is able to simulate a correct pattern of eddy variability. Their subgrid representation only captures the subgrid energy e and does not attempt to represent the parameterized action of the Reynolds stress. In models that are not fully eddy-permitting, this approach will not work and one needs a more sophisticated model of the subgrid. We will discuss so-called *stochastic superparameterizations* as proposed by Grooms and Majda (2013, 2014) and Grooms et al. (2015b) in the context of quasigeostrophic two-layer models.

The main difference between the stochastic parameterization (SP) and the stochastic superparameterization (SSP) is that the latter involves a prognostic fine grid equation which is motivated by the underlying physical evolution equation and involves coarse mesh quantities as well as a stochastic source term. We first explain the idea in the context of the simple single-layer model (5.4), where the essential features of the method are already visible with less notational effort.

Let us decompose the stream function ψ into a coarse mesh stream function ψ_c and a fine mesh stream function ψ', and likewise define the corresponding vorticities, so that

$$\psi = \psi_\mathrm{c} + \psi' \quad \text{and} \quad \zeta = \zeta_\mathrm{c} + \zeta' \,, \tag{5.72}$$

where it is understood that $\zeta_\mathrm{c} = \Delta\psi_\mathrm{c}$ and $\zeta' = \Delta\psi'$. We also split the forcing into a deterministic physical forcing F_c on the coarse mesh and a stochastic forcing F' on the fine mesh. Inserting this ansatz into (5.4), we obtain

$$\begin{aligned}\partial_t\zeta_c + \partial_t\zeta' + [\psi_c, \zeta_c] + [\psi_c, \zeta'] + [\psi', \zeta_c] + [\psi', \zeta'] \\ + \beta\,\partial_x\psi_c + \beta\,\partial_x\psi' = F_\mathrm{c} + F' + D\zeta_\mathrm{c} + D\zeta' \,.\end{aligned} \tag{5.73}$$

We now split this equation into an evolution equation for the coarse variables and an evolution equation for the fine mesh variables. This procedure is non-rigorous, so there is some freedom of choice. However, the fine system should be linear with constant coefficients so that it can be solved explicitly, for otherwise the combined computational cost would be higher than the cost of simulating the entire system on the fine grid.

Following Grooms and Majda (2014), we decompose the domain Ω into disjoint subdomains Ω_i. Each subdomain contains exactly one grid point of the coarse mesh, and the coarse mesh variables are assumed constant on each adjacent subdomain. The fine systems are then solved independently for one coarse time step with periodic boundary conditions on each subdomain.

The coarse system should contain all coarse terms and the divergence of the eddy potential vorticity flux[13]

$$\boldsymbol{F}_\mathrm{epv} = \overline{\zeta'\,\boldsymbol{\nabla}^\perp\psi'} \,. \tag{5.74}$$

[13]The overbar operation denotes averaging over one fine grid cell. Note that it is not possible to simply take the fine grid Jacobian and average over a coarse cell, as this expression would be identically zero. Instead, one uses the average flux and practically operates on it with the coarse grid divergence.

All the remaining terms go into the fine mesh system which is solved independently on each coarse mesh cell. The coarse system then takes the form

$$\partial_t \zeta_c + [\psi_c, \zeta_c] + \boldsymbol{\nabla} \cdot \boldsymbol{F}_{\mathrm{epv}} + \beta\, \partial_x \psi_c = F_{\mathrm{c}} + D\zeta_c\,, \tag{5.75}$$

and the fine system reads

$$\partial_t \zeta' + [\psi_c, \zeta'] + [\psi', \zeta_c] + \beta\, \partial_x \psi' = S + F' + D\zeta'\,, \tag{5.76}$$

where the nonlinear eddy–eddy interactions are represented by

$$S = \boldsymbol{\nabla} \cdot \boldsymbol{F}_{\mathrm{epv}} - [\psi', \zeta']\,. \tag{5.77}$$

The coarse system is solved on the coarse grid. The fine system is linear except for the eddy–eddy term which must be modeled. Grooms and Majda (2014) suggest to replace each Fourier mode $S_{\boldsymbol{k}}$ by the right-hand side of an Ornstein–Uhlenbeck stochastic process of the form

$$\mathrm{d}\zeta = -\gamma\, \zeta\, \mathrm{d}t + \sigma\, \mathrm{d}W\,, \tag{5.78}$$

where W is a standard Wiener process.[14] The Ornstein–Uhlenbeck process is controlled by two parameters, the inverse correlation time γ and the variance $\sigma^2/(2\gamma)$ which will later be chosen differently for different wavenumbers. They further assume that the coarse grid fields can be held constant in each fine cell and that there is no forcing on the fine scale. Then, the full fine-scale model in the Fourier representation reads

$$\mathrm{d}\zeta'_{\boldsymbol{k}} = (\ell_{\boldsymbol{k}} - \gamma_{\boldsymbol{k}})\, \zeta'_{\boldsymbol{k}}\, \mathrm{d}t + \sigma_{\boldsymbol{k}}\, \mathrm{d}W_{\boldsymbol{k}}\,, \tag{5.79}$$

where $\ell_{\boldsymbol{k}}$ is the Fourier symbol of all linear terms in (5.76) and the Wiener processes $W_{\boldsymbol{k}}$ are mutually independent.

The crucial observation is that S is quadratic in fine-scale quantities, so that a space average corresponds, by the Plancharel theorem, to an integral over $|\psi_{\boldsymbol{k}}|^2$. Averaging further over the stochastic ensemble, it is clear that it suffices to compute the evolution of $\mathbb{E}[|\psi_{\boldsymbol{k}}|^2]$. By the Itô Lemma, it is easy to derive a deterministic linear ordinary differential equation for this quantity, which can be solved explicitly and independently for each wavenumber.

The coefficients $\gamma_{\boldsymbol{k}}$ are tuned so that the equilibrium distribution without the interaction terms $\ell_{\boldsymbol{k}}$ matches a given power spectrum. Later, when initializing the second moment equation, the initial value is taken to be the equilibrium value, again without interactions. Thus, the effect of the interaction with the coarse grid quantities,

[14] The stochastic process can also be viewed as accounting for all other approximations which are implicitly made: the replacement of exact operators by coarse grid approximations in (5.75), the use of periodic boundary conditions for the fine cell dynamics, and the uncertainty in the initialization and re-initialization of the fine grid dynamics.

which are encoded in ℓ_k, is to color the fine grid statistics consistent with the coarse grid flow. In particular, when applying this method to stratified models, $\ell_k > 0$ at scales where the flow is baroclinically unstable, resulting in growth of the primed quantities.

At this point, the effective subgrid dynamics as seen from the coarse grid is purely deterministic. Grooms and Majda (2014) found that it is necessary to take a large number of modes in the subgrid to match the correct spectral decay. To keep the computational cost low, and to account for the observation that real ocean eddies have anisotropies that vary in space and time, they select at random a direction in each subgrid cell, independent for each point in coarse space–time, and choose a one-dimensional spectral decomposition in this cell.

Grooms et al. (2015b) perform a detailed computational study of their stochastic superparameterization in a two-layer zonally reentrant channel mimicking the Antarctic Circumpolar Current. They compare the model with a deterministic Gent–McWilliams (GM) parameterization in a regime where mesoscale eddies are not resolved on the coarse grid and with an eddy-resolving high-resolution simulation. Their setting is similar to the two-layer quasigeostrophic model discussed in Section 5.3.2, with vorticity equations

$$\partial_t q_1 + [\psi_1, q_1] = -\frac{2}{\rho_0 H}\,\partial_y F(y) + \nu_2\,\Delta^2\psi_1\,, \tag{5.80a}$$

$$\partial_t q_2 + [\psi_2, q_2] = -r\,\Delta\psi_2 + \nu_2\,\Delta^2\psi_2\,, \tag{5.80b}$$

where ν_2 is Newtonian viscosity, r is the Ekman drag coefficient, and the layer potential vorticities are given by

$$q_1 = f_0 + \beta y + \Delta\psi_1 + \tfrac{1}{2}\,k_\mathrm{d}^2\,(\psi_2 - \psi_1) - k_\mathrm{e}^2\,\psi_1\,, \tag{5.81a}$$

$$q_2 = f_0 + \beta y + \Delta\psi_2 + \tfrac{1}{2}\,k_\mathrm{d}^2\,(\psi_1 - \psi_2) + \frac{2f_0}{H}\,h_\mathrm{b}\,. \tag{5.81b}$$

The lateral boundary conditions are periodic in the zonal and stress-free in the meridional direction. This system is different from the two-layer equation (5.28) in the following way: The flow here is driven by a steady sinusoidal wind forcing $F(y)$.[15] It includes explicit bottom topography h_b to avoid unrealistic spin-up of the mean current. A stretching term with coefficient $k_\mathrm{e} = 1/L_\mathrm{e} = f/\sqrt{gH}$ is included in the upper layer potential vorticity so that the model is formally valid to scales up to the external Rossby radius of deformation L_e. Finally, dissipation is second order as is common for relatively coarse resolution models; the eddy-resolving simulation and the fine grid dynamics, however, are set up with fourth-order dissipation.

In the coarse model, the divergence of the subgrid potential vorticity fluxes is introduced layerwise as explained in the single-layer setting; for details, see (Grooms and

[15] The steady wind stress tilts the layer interface through Ekman pumping. When it is sufficiently tilted, the flow becomes baroclinically unstable. So here, as before, the mean forcing maintains a pool of mean kinetic and available potential energies.

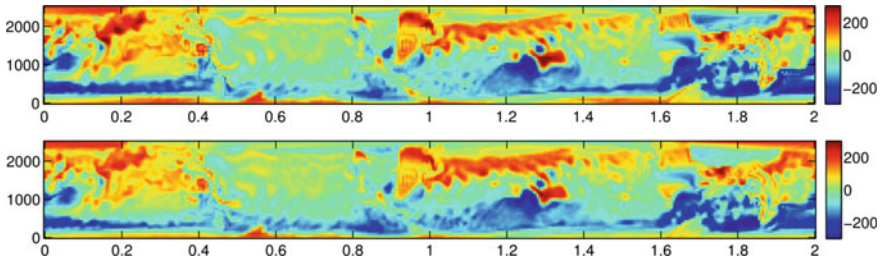

Fig. 5.3 Time mean bias of interface elevation in meters of the Gent–McWilliams parameterization (top) vs. the stochastic superparameterization (bottom). The zonal direction is shown in the horizontal in units of 10 000 km, the meridional direction in the vertical in units of kilometers. Graphs are adapted from Grooms et al. (2015b)

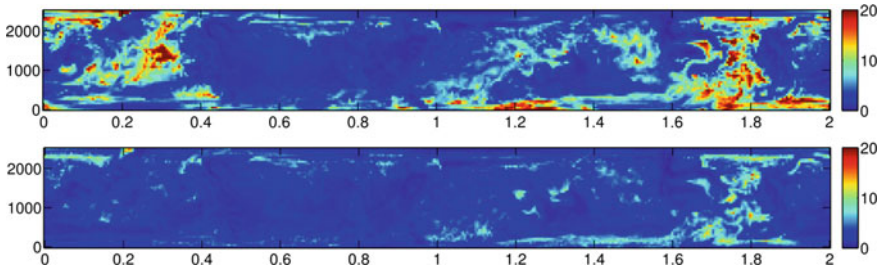

Fig. 5.4 Ratio of interface elevation time variance of the reference simulation over time variance corresponding to the data shown in Figure 5.3. The variance deficiency of the Gent–McWilliams parameterization (top) is significantly larger than the variance deficiency of the stochastic superparameterization (bottom). Graphs are adapted from Grooms et al. (2015b)

Majda 2014). In order to compare the performance of GM and SPP, the authors analyze the temporal statistics of the surface elevation at each fixed point; in particular, they compute the bias of the mean and the bias of the variance relative to the highly resolved reference simulation. While the time mean biases of both parameterizations are similar in magnitude and spatial pattern (Figure 5.3), the time variance of the stochastic superparameterization run is significantly closer to the reference simulation, even though both models are variance deficient (Figure 5.4).

While this model achieves a spectrally consistent proxy for the missing potential vorticity flux, it imposes locality and spatial uniformity of the subgrid model. In reality, however, as follows from the pattern of eddy kinetic energy in Grooms et al. (2015b), the eddy kinetic energy is (i) essentially non-uniform and (ii) does not correlate with the places of maximum instability—indeed, EKE spots are always downstream of places with maximal baroclinic instability. So, the main conceptual question is how to introduce non-trivial advection of subgrid quantities by the coarse flow.

Another question is how to make this approach practical. In essence, we need the dependence of mean subgrid fluxes as a function of velocities and quasigeostrophic

PV gradients. For a two-layer quasigeostrophic model, this can still be done. However, for continuous stratification and for the primitive equations, there is the vertical dependence which will soon require more computational effort than directly applying an eddy-resolving mesh. Another question is how to avoid overexciting gravity waves in a gravity wave-permitting model.

There may be possible simplifications. For example, since the fine grid dynamics depends only on the magnitude of resolved velocities, PV gradients, and the angle between the velocities and gradients, one may only consider a finite set of velocity values and PV gradients, and interpolate the results for the eddy flux divergence between the simulated patterns computed for this set. Such a lookup table may considerably reduce computational cost. Further, for primitive equation models, the subgrid may still be treated in quasigeostrophic approximation. Thus, it will be possible to represent PV gradients on the subgrid which is seen essential for providing a proper proxy for baroclinic instability, and cannot be easily done if the superparameterization is formulated in terms of primitive variables where information on PV gradients would be lost when going to the fine grid.

5.5 Other Closures

5.5.1 The Mana–Zanna Parameterization of Ocean Mesoscale Eddies

Mana and Zanna (2014) study the correlation of different functional forms for the eddy source term with a highly resolved direct numerical simulation, select the best candidate function, and match the remaining coefficients with the empirical data. More detailed tests in a double gyre configuration are reported in Zanna et al. (2017). In these two papers, the authors work in a 3-layer quasigeostrophic setting; possible extensions to primitive equation models are discussed and tested in Anstey and Zanna (2017). In the following, we describe the Mana–Zanna parameterization following the concise derivation later given by Grooms and Zanna (2017). We will present a slightly more general view which raises interesting possibilities for further optimization of the closure.

For simplicity, we restrict the discussion to the barotropic single-layer QG equations without β-effect. Working exclusively in the continuum setting on the plane, we define a time-independent coarsening operation via convolution with a filter kernel, i.e.,

$$\overline{\zeta}(\boldsymbol{x}) = \int_{\mathbb{R}^2} \phi_\delta(\boldsymbol{x} - \boldsymbol{y})\, \zeta(\boldsymbol{y})\, \mathrm{d}\boldsymbol{y} \tag{5.82}$$

where ϕ_δ is a radial kernel with δ referring to the width of the filter. Applying this operation to the barotropic vorticity equation (5.4a), we can write

$$\overline{D}_t\overline{\zeta} \equiv \partial_t\overline{\zeta} + [\overline{\psi}, \overline{\zeta}] = S + \overline{F} + D^*\overline{\zeta}\,, \tag{5.83}$$

where D^* is some coarsened dissipation operator and S denotes the eddy source term

$$S = [\overline{\psi}, \overline{\zeta}] - \overline{[\psi, \zeta]} + \overline{D\zeta} - D^*\overline{\zeta}\,. \tag{5.84}$$

In Mana and Zanna (2014), the authors seek to build a model for S in terms of the divergence of Rivlin–Ericksen stresses which originated in the description of non-Newtonian fluids with infinitesimal memory (see, e.g., Truesdell and Rajagopal 1999). These tensors satisfy material frame invariance and observer objectivity, properties required of a physical material law. For the barotropic vorticity equation, an exact implementation of an inviscid second-grade Rivlin–Ericksen fluid would correspond to

$$S = \alpha\,\overline{D}_t\Delta\overline{\zeta} \tag{5.85}$$

which leads to the vorticity formulation of the Euler-α model further discussed in Section 5.5.2 below. Their study, however, finds that a better correlation is obtained by using

$$S = \alpha\,\Delta\overline{D}_t\overline{\zeta} \tag{5.86}$$

which differs from (5.85) by nonlinear commutators, but preserves the property of frame invariance. They also find that the coefficient α on the right-hand side is negative, which precludes a straightforward interpretation as advection by a smoothed velocity field.

Grooms and Zanna (2017) provide *a posteriori* justification of (5.86) along the following lines. They argue that, after high-pass filtering, S is the dominant term on the right-hand side of (5.83).[16] In particular, the Laplacian of S dominates the Laplacians of the other two terms so that

$$\Delta D_t\overline{\zeta} \approx \Delta S\,. \tag{5.87}$$

They proceed to show that S is highly correlated with ΔS, so that (5.87) implies (5.86).

Let us explore such correlations from a more general perspective. We define a family of abstract approximate Laplacians which includes the usual 5-point discrete Laplacian in two dimensions. Suppose that $\{\mu_\varepsilon\}_{\varepsilon>0}$ is a family of finite positive Borel measures on $\mathbb{R}^n$ with $\operatorname{supp}\mu_\varepsilon \subset B(x,\varepsilon)$, the ball centered at x with radius ε, satisfying

[16] This assumption was tested with harmonic dissipation. It would be questionable with higher-order dissipation so that, in general, the contribution from dissipation would need to be carried explicitly.

$$\mu_\varepsilon(\{x\}) = 0\,, \tag{5.88a}$$

$$\lim_{\varepsilon\to 0}\int_{B(x,\varepsilon)} y_i \,\mathrm{d}\mu_\varepsilon(y) = 0\,, \tag{5.88b}$$

and

$$\lim_{\varepsilon\to 0}\int_{B(x,\varepsilon)} y_i\, y_j \,\mathrm{d}\mu_\varepsilon(y) = 2\,\delta_{ij} \tag{5.88c}$$

for all $i, j = 1, \ldots, n$. In particular, various normalized symmetric measures, including point measures and surface Lebesgue measures in lower dimensions, satisfy these conditions. Then,

$$\Delta S(x) = \lim_{\varepsilon\to 0} |\mu_\varepsilon|\,(\mathrm{Av}_\varepsilon(S) - S(x))\,, \tag{5.89}$$

where $|\mu_\varepsilon| = \mu_\varepsilon(\mathbb{R}^n)$ and

$$\mathrm{Av}_\varepsilon(S) = \frac{1}{|\mu_\varepsilon|}\int_{B(x,\varepsilon)} S(y)\,\mathrm{d}\mu_\varepsilon(y)\,. \tag{5.90}$$

Assume now that S is a homogeneous isotropic δ-correlated Gaussian random field with variance σ^2. Setting

$$\boldsymbol{w} = \begin{pmatrix} S(x) \\ \mathrm{Av}_\varepsilon(S) - S(x) \end{pmatrix} \tag{5.91}$$

and fixing $\varepsilon > 0$ at a small finite value, we find that the covariance matrix of $\boldsymbol{w}$ is given by

$$\Sigma = \mathrm{Cov}[\boldsymbol{w}, \boldsymbol{w}] = \sigma^2 \begin{pmatrix} 1 & -1 \\ -1 & (b+1) \end{pmatrix}\,, \tag{5.92}$$

with $b = |\mu_\varepsilon^2|/|\mu_\varepsilon|^2$, where $|\mu_\varepsilon^2| = \sum_{y\in B(x,\varepsilon)} \mu_\varepsilon^2(\{y\})$.[17]

The eigenvalue ratio corresponding to the subdominant principal component of Σ is given by

$$r \equiv \frac{\lambda_2}{\lambda_1 + \lambda_2} = \frac{b - \sqrt{b^2+4} + 2}{2\,b + 4}\,. \tag{5.93}$$

It quantifies the fraction of variance not explained by a linear relationship between the components of $\boldsymbol{w}$. When μ_ε does not have point measure components, $b = 0$ and consequently the eigenvalue ratio $r = 0$, indicating perfect correlation between the components of $\boldsymbol{w}$. The largest value of b in the class of point measures with equal weights corresponds to the 4-point Laplacian where μ_ε consists of three-unit Dirac

[17] The sum in this definition is countable and convergent as $|\mu_\varepsilon| = \mu_\varepsilon(B(x,\varepsilon)) < \infty$. Thus, $|\mu_\varepsilon^2| = 0$ if and only if the measure μ_ε has no nonzero point measure components, which is the case for measures that are absolutely continuous with respect to the Lebesgue measure in the ball or on the sphere.

masses at angles $0, 2\pi/3$, and $4\pi/3$. In this case, $b = 1/3$ and $r = (7 - \sqrt{37})/14 \approx 0.066$. For the usual 5-point stencil as considered in Grooms and Zanna (2017), $b = 1/4$ and $r = 9 - \sqrt{65}/18 \approx 0.052$.

Thus, even for relatively concentrated measures, a major fraction of the variance is explained by a linear relationship between S and the approximate Laplacian of S as defined via the right-hand expression in (5.89) for finite ε. The constant of proportionality is the ratio of the components of the principal eigenvector of Σ, i.e.,

$$S(x) \approx \tfrac{1}{2}\,(b - \sqrt{4 + b^2})\,(\mathrm{Av}_\varepsilon(S) - S(x))\,. \tag{5.94}$$

In the concrete case of the 5-point Laplacian, we find

$$S \approx -\frac{1}{8}\,(1 - \sqrt{65})\,\frac{\varepsilon^2}{4}\,\Delta S = -(c\varepsilon)^2\,\Delta S \approx -(c\varepsilon)^2\,\Delta D_t\overline{\zeta} \tag{5.95}$$

where $c \approx 0.469782$, which is close to the empirical value found by Mana and Zanna (2014).[18]

It is clear from the argument above that the eigenvalue ratio improves when the measure becomes less localized. On the other hand, the assumption of S being a δ-correlated random field must break down on small scales; we expect the decorrelation length to be at or slightly larger than the grid scale. Thus, it should be possible to replace the Laplacian in the argument above with a discrete operator Λ whose stencil nodes are at least a decorrelation length apart and which is effectively acting as a high-pass filter. The form of Λ can then be optimized for its eigenvalue ratio. In this context, we remark that the approximation made in (5.87) does not seem necessary to proceed, as $\Lambda\overline{F}$ and $\Lambda D^*\overline{\zeta}$ are readily computable. We believe that this question is worth further investigation.

A different line of reasoning might be based on a random field model for S with finite spatial correlations. Assuming a given spectrum for S, the characterization of the two-point correlation function via the Wiener–Khinchin theorem (see, e.g., Yaglom 1987) could still allow us to compute the covariance matrix Σ explicitly and subsequently optimize the filter Λ. Finally, a detailed analysis of the structure of S is warranted. Grooms et al. (2015a) provide an argument that the spectrum of S grows like k^5, which can likely only be true on a limited range of scales as a perfectly δ-correlated random field should have a flat spectrum. Thus, in particular the details of spatial correlation near the grid scale require attention.

[18] The slight discrepancy in the value of c as compared with Grooms and Zanna (2017) is due to the different normalizations of the vector $\boldsymbol{w}$. Ideally, the components of $\boldsymbol{w}$ should have unit variance, but both our choices are close enough to being normalized for the point being made.

5.5.2 α-Models

The so-called α-models initially came up in the study of nonlinear waves, not in turbulence. What is now known as the Camassa–Holm equation was first discovered by Fuchssteiner and Fokas (1981) who sought completely integrable generalizations of the Korteweg–de Vries (KdV) equation with a bi-Hamiltonian structure. It was independently re-derived by Camassa and Holm (1993)—for a more detailed exposition, see Camassa et al. (1994)—as a next order correction to the KdV equation in small amplitude expansion of unidirectional surface waves in irrotational shallow water. Camassa and Holm's work attracted a lot of attention as, in addition to integrability and bi-Hamiltonian structure, they found a family of peaked soliton solutions. Solutions of the Camassa–Holm equation can be seen as geodesics on the diffeomorphism group with respect to a right-invariant H^1-metric (Kouranbaeva 1999). The striking parallel to Arnold's (1966) view of ideal three-dimensional hydrodynamics as geodesic flow on the volume-preserving diffeomorphism group endowed with an L^2-metric was pointed out by Holm et al. (1998) who, replacing the L^2 with an H^1-metric, obtained a hydrodynamic analog to the Camassa–Holm equations which is now known as the Euler-α equations or the Lagrangian-averaged Euler equations. In velocity–momentum variables, they read

$$\partial_t \boldsymbol{v} - \boldsymbol{u} \times (\nabla \times \boldsymbol{v}) + \nabla p = 0\,, \tag{5.96a}$$

$$\boldsymbol{v} = (1 - \alpha^2 \Delta)\boldsymbol{u}\,, \tag{5.96b}$$

$$\nabla \cdot \boldsymbol{u} = 0\,. \tag{5.96c}$$

The Euler-α equations arise from the "kinetic energy" Lagrangian

$$L_\alpha = \tfrac{1}{2} \int |\boldsymbol{u}|^2 + \alpha^2 \, |\nabla \boldsymbol{u}|^2 \, \mathrm{d}\boldsymbol{x}\,, \tag{5.97}$$

which is a constant of the motion.

The connection to turbulence was made soon after its discovery, based on a number of observations. The momentum $\boldsymbol{v}$ is transported by a velocity field $\boldsymbol{u}$ which is smoother than the momentum; see (5.96b). This was seen as analogous to Reynolds averaging, even though the two operations are not equivalent; further, the non-viscous terms take the form of a Rivlin–Erikson tensor, so that, in their inviscid form, they coincide with the equations of motion for a non-Newtonian fluid of second grade (Foias et al. 2001). Analytically, the Euler-α equations possess properties which are notably lacking in ideal and Newtonian fluids: In two dimensions, the Euler-α model has unique global point vortex solutions (Oliver and Shkoller 2001), and in three dimensions, the viscous α-equations have global classical solutions (Marsden and Shkoller 2001; Foias et al. 2002). We note that it is not *a priori* clear how to add viscosity to (5.96): The references quoted so far argue that momentum should be diffused; see Chen et al. (1999a) for a discussion of this issue. The classical equations of a viscous second-grade fluid, in contrast, diffuse velocity—a mathematically

weaker form of dissipation so that, correspondingly, the global existence of solution is only known for small initial data (Cioranescu and Girault 1997), much like the situation for the Navier–Stokes equations in three dimensions.

Several authors have given derivations of the Euler-α equations as the equations of motion for some notion of a Lagrangian mean flow. Holm (1999, 2002) recognized a close connection between Lagrangian averaging and the generalized Lagrangian mean (GLM) of Andrews and McIntyre (1978). To provide closure, Holm assumes that first-order fluctuations in a small amplitude expansion are parallel-transported by the mean flow—an assumption he refers to as a *Taylor hypothesis* in analogy with G.I. Taylor's observation that turbulent fluctuations are correlated in the downstream direction of a flow (Taylor 1938). Marsden and Shkoller (2003), in contrast, assume that first-order fluctuations are transported as a vector field and that parallel transport of second-order fluctuations is, on average, orthogonal to the velocity field. Recently, Gilbert and Vanneste (2018) have pointed out that a geometric view of the Lagrangian mean fixes the higher-order closure conditions. In this framework, the Euler-α equations emerge from Lagrangian averaging under the minimal set of assumptions that (i) the averaged map is the minimizer of geodesic distance, (ii) first-order fluctuations are statistically isotropic, and (iii) first-order fluctuations are transported by the mean flow as a vector field (Oliver 2017; Badin et al. 2018).

The numerical evidence supporting the use of α-models is mixed. Early numerical studies for homogeneous turbulence were encouraging (Chen et al. 1999b; Mohseni et al. 2003). The underlying idea has also been ported to rotating geophysical fluid flow (Holm and Nadiga 2003) and used in various test cases (Hecht et al. 2008a; Aizinger et al. 2015). Careful comparative studies for two-dimensional quasigeostrophic turbulence, however, show that the α-model perturbs the dynamics of two-dimensional turbulence. In particular, it suffers from accumulation of enstrophy at small scales (Lunasin et al. 2007; Graham and Ringler 2013) and has inferior correlation with an empirically observed subgrid stress tensor (Mana and Zanna 2014), where the computationally observed behavior is close to (5.95), a relationship that is similar, but not identical to the α-model closure. In addition, as the inversion of the Helmholtz filter in (5.96b) is non-local, it is not appealing for use in a full ocean model. We note, however, that the idea of filtering in a geometrically intrinsic setting is more general than what is usually pursued and may have some merit even in the setting of the nearly geostrophic turbulence in mesoscale ocean dynamics.

In the final part of the section, we shall sketch a possible nonstandard interpretation of the α-model dynamics as a model for two-dimensional turbulence. For simplicity, we return to the barotropic vorticity equation of Section 5.3.1 with $\beta = 0$ and initially ignore forcing and dissipation. In this setting, the model coincides with the well-studied two-dimensional Euler-α equation whose vorticity dynamics reads

$$\partial_t \xi + [\psi, \xi] = 0\,, \tag{5.98a}$$

$$\xi = L_\alpha \Delta \psi\,, \tag{5.98b}$$

where (5.96b) corresponds to the choice $L_\alpha = 1 - \alpha\Delta$, but L_α could also be a more general operator defined via a Fourier symbol $\ell_\alpha(k)$. The α-energy at wavenumber $\boldsymbol{k}$ is $\mathscr{E}_{\boldsymbol{k}} = -\frac{1}{2}\,\psi^*_{\boldsymbol{k}}\,\xi_{\boldsymbol{k}}$, and the α-enstrophy is given by $\mathscr{Z}_{\boldsymbol{k}} = k^2\,\ell_\alpha(k)\,\mathscr{E}_{\boldsymbol{k}}$; system (5.98) conserves total α-energy and α-enstrophy.

Now suppose that α-wavenumber k corresponds to a different physical wavenumber $\kappa(k)$ and that there is a corresponding physical energy

$$E(\kappa) = \mathscr{E}(k)/h(k)\,. \tag{5.99}$$

A straightforward computation shows that total physical energy and enstrophy are conserved if and only if

$$\kappa^2(k) = \ell_\alpha(k)\,k^2 \tag{5.100a}$$

and

$$h(k) = \tfrac{1}{2}\,k\,\ell'_\alpha(k) + \ell_\alpha(k)\,. \tag{5.100b}$$

Looking at the detailed triad interactions of the α-model, we find transfer rate relations similar to (5.15) where the rate of nonlinear energy transfer is with respect to α-wavenumbers, whereas the prefactors on the right-hand side are satisfied with respect to physical wavenumbers. Thus, in general, it is not even approximately true that the α-triad picture corresponds to the physical triad picture under the wavenumber mapping implied by energy and enstrophy conservation. However, there is one class of triads for which this is approximately the case: when one leg of the triad is in the low wavenumbers and two legs are in the high-wavenumber range; to be definite, we take $p \ll k < q$ and set $\delta = p/k \ll 1$. We might call such interactions *catalytic triads* as (5.15) shows that there is an $O(1)$ energy exchange between modes $\boldsymbol{k}$ and $\boldsymbol{q}$ while mode $\boldsymbol{p}$ exchanges energy only at a rate $O(\delta)$. In other words, mode $\boldsymbol{p}$ takes the role of a catalyst, mediating the transfer of energy in the high-wavenumber regime while not participating in it to leading order. Provided the turbulent regime is dominated by catalytic triads (which is not the classical KLB picture, but it is likely that these triads are key players in the inverse cascade), then under mild assumptions on ℓ_α, an α-model can be interpreted as representing the physical interactions under the mappings (5.100) up to relative errors in rates and mapped wave numbers of $O(\delta)$. The details of this computation involve only elementary estimates and shall be omitted here.

Thus, to interpret the α-dynamics consistently via the remapping of wavenumbers, the same map must be applied when adding forcing and dissipation terms. Dissipation in the α-model momentum equation, in particular, should take the form $D(\kappa(k))$. This corresponds to momentum rather than velocity diffusion and thus coincides with the dissipation operator typically used in connection with α-models as reviewed earlier in this section.

Finally, to consistently interpret the energy spectrum, it must be mapped back to physical wavenumbers. In the comparison of Graham and Ringler (2013), for

example, no such map is applied. This constitutes an interesting open question as, to our knowledge, such analysis has never been done. A related open problem is to formulate the α-model subgrid closure mapped to physical wavenumbers.

5.6 Concluding Remarks

In this chapter, we have reviewed the foundations of geostrophic turbulence and its implications for ocean models in the eddy-permitting regime. In the past decade, a number of authors have looked at the problem of effective parameterizations for subgrid eddy activity and for the resulting backscatter of energy into the resolved grid. Most of the detailed testing so far has been done in the context of quasigeostrophic layer models, with increased attention to full primitive equation setups in recent years.

Our selection of parameterizations for close discussion is necessarily incomplete, highlighting recent developments in favor of older ideas, putting an emphasis on mathematical structure toward systematic, or even rigorous, analysis, and with a view toward applicability for a new generation of global circulation models featuring irregular grids with spatially varying grid resolution which rules out approaches that require explicit Fourier transforms or other constructs tied to a regular grid.

To a large extent, the ideas expressed here are exploratory. None of the parameterizations described here is widely used in operational models so that a major development cycle of introducing more energy-consistent parameterizations lies ahead. It is also not clear which of these approaches will be the most fruitful in the long run or whether some new or possibly old ideas will prevail.

Such old ideas could include the anticipated vorticity method of Sadourny and Basdevant (1985) which seeks to introduce a force $-D\boldsymbol{k} \times \boldsymbol{u}$ such that, for example, when D is chosen as an upwind estimate of the layer potential vorticity, the scheme conserves energy exactly while dissipating enstrophy. In practice, this approach is insufficient as it does not remove the component of small-scale numerical noise in $\boldsymbol{u}$ that does not project on curl as required for numerical stability. Graham and Ringler (2013) report that first-order anticipated vorticity results in either an excess of energy at all scales or dissipation of enstrophy across a too large portion of the spectrum; they suggest that applying a high-order spatial operator within the anticipated PV formalism may solve this issue, but at the expense of easy implementability in current GCMs. Yet, the underlying idea is interesting as the mathematically most direct way to reconcile energy conservation with enstrophy dissipation.

The classical development of Smagorinsky closures has been central in the modeling and simulation of turbulent flow regimes; see, e.g., the review by Meneveau and Katz (2000). However, it is not directly applicable to typical ocean regimes where, due to the scales at which forcing, instabilities, and dissipation act, one is often not in a self-similar scaling regime which is a prerequisite of LES and Smagorinsky-type closures. Dynamical Smagorinsky closures (when the subgrid viscosity is computed by applying an additional coarsening filter and fitting the difference between this and

the original filter to the simulated stresses) could be of interest, although even with these techniques the lack of self-similarity may be an issue.

Stochastic modeling of subgrid interactions and backscatter has been developed, based on the direct interaction approximation of Kraichnan (1959), by Frederiksen and co-workers (Frederiksen and Davies 1997; O'Kane and Frederiksen 2008; Kitsios et al. 2013, 2014, 2016). While their work involves a detailed analysis of unresolved eddy–eddy interactions, it also heavily relies on spectral language, so that it is not clear how applicable this approach is in the context of complex geometries and possibly non-uniform grids and what the trade-offs in terms of skill vs. computational expense are. We also remark that there is similarity between the expression for the subgrid drain dissipation matrix in Kitsios et al. (2013) and the estimation of a dynamic Smagorinsky coefficient in the spirit of Germano et al. (1991).

For systems with an explicit fast–slow scale separation, it may be possible to model the fast timescale component with a stochastic process and use *stochastic mode reduction* to reduce the system to a stochastic equation on the slow timescale. Such methods are reviewed in Section 5 of Franzke et al. (2019). However, it is completely open whether this approach is applicable to subscale modeling in geostrophic turbulence where there is no clear scale separation and whether these techniques scale up to full ocean models.

References

Aizinger, V., Korn, P., Giorgetta, M., Reich, S.: Large-scale turbulence modelling via α-regularisation for atmospheric simulations. J. Turbul. **16**(4), 367–391 (2015)

Aluie, H.: Scale decomposition in compressible turbulence. Phys. D **247**(1), 54–65 (2013)

Aluie, H., Kurien, S.: Joint downscale fluxes of energy and potential enstrophy in rotating stratified Boussinesq flows. Europhys. Lett. **96**(4), 44006 (2011)

Andrews, D.G., McIntyre, M.E.: An exact theory of nonlinear waves on a Lagrangian-mean flow. J. Fluid Mech. **89**(4), 609–646 (1978)

Anstey, J.A., Zanna, L.: A deformation-based parametrization of ocean mesoscale eddy Reynolds stresses. Ocean Model. **112**, 99–111 (2017)

Arnold, V.: Sur la géométrie différentielle des groupes de Lie de dimension infinie et ses applications à l'hydrodynamique des fluides parfaits. Ann. Inst. Fourier **16**(1), 319–361 (1966)

Augier, P., Lindborg, E.: A new formulation of the spectral energy budget of the atmosphere, with application to two high-resolution general circulation models. J. Atmos. Sci. **70**, 2293–2308 (2013)

Badin, G.: On the role of non-uniform stratification and short-wave instabilities in three-layer quasi-geostrophic turbulence. Phys. Fluids **26**(9), 096603 (2014)

Badin, G., Oliver, M., Vasylkevych, S.: Geometric Lagrangian averaged Euler–Boussinesq and primitive equations. J. Phys. A: Math. Theor. **51**(45), 455501 (2018)

Batchelor, G.K.: Computation of energy spectrum in homogeneous two-dimensional turbulence. Phys. Fluids Suppl. **12**, II:233–239 (1969)

Berloff, P.: Dynamically consistent parameterization of mesoscale eddies. Part I: simple model. Ocean Model. **87**, 1–19 (2015)

Berloff, P.: Dynamically consistent parameterization of mesoscale eddies. Part II: eddy fluxes and diffusivity from transient impulses. Fluids **1**(3), 22 (2016)

Callies, J., Ferrari, R.: Interpreting energy and tracer spectra of upper-ocean turbulence in the submesoscale range (1–200 km). J. Phys. Oceanogr. **43**, 2456–2474 (2013)

Camassa, R., Holm, D.D.: An integrable shallow water equation with peaked solitons. Phys. Rev. Lett. **71**(11), 1661–1664 (1993)
Camassa, R., Holm, D.D., Hyman, J.M.: A new integrable shallow water equation. Adv. Appl. Mech. **31**(C), 1–33 (1994)
Capet, X., McWilliams, J.C., Molemaker, M.J., Shchepetkin, A.F.: Mesoscale to submesoscale transition in the California current system. Part III: energy balance and flux. J. Phys. Oceanogr. **38**, 2256–2269 (2008)
Cessi, P.: An energy-constrained parameterization of eddy buoyancy flux. J. Phys. Oceanogr. **38**, 1807–1819 (2008)
Chaouat, B.: Simulations of turbulent rotating flows using a subfilter scale stress model derived from the partially integrated transport modeling method. Phys. Fluids **24**(4), 045108 (2012)
Charney, J.G.: Geostrophic turbulence. J. Atmos. Sci. **28**(6), 1087–1095 (1971)
Chen, S., Foias, C., Holm, D.D., Olson, E., Titi, E.S., Wynne, S.: A connection between the Camassa-Holm equations and turbulent flows in channels and pipes. Phys. Fluids **11**(8), 2343–2353 (1999a)
Chen, S., Holm, D.D., Margolin, L.G., Zhang, R.: Direct numerical simulations of the Navier-Stokes alpha model. Phys. D **133**(1–4), 66–83 (1999b)
Cioranescu, D., Girault, V.: Weak and classical solutions of a family of second grade fluids. Int. J. Non-Linear Mech. **32**(2), 317–335 (1997)
Cooper, F.C.: Optimisation of an idealised primitive equation ocean model using stochastic parameterization. Ocean Model. **113**, 187–200 (2017)
Cooper, F.C., Zanna, L.: Optimisation of an idealised ocean model, stochastic parameterisation of sub-grid eddies. Ocean Model. **88**, 38–53 (2015)
Daly, B., Harlow, F.: Transport equations in turbulence. Phys. Fluids **13**, 2634–2649 (1970)
Danilov, S.: Non-universal features of forced 2D turbulence in the energy and enstrophy ranges. Discret. Contin. Dyn. Syst. Ser. B **5**(1), 67–78 (2005)
Danilov, S.: Ocean modeling on unstructured meshes. Ocean Model. **69**, 195–210 (2013)
Danilov, S., Gurarie, D.: Forced two-dimensional turbulence in spectral and physical space. Phys. Rev. E **63**(6), 061208 (2001)
Danilov, S., Gurarie, D.: Scaling, spectra and zonal jets in beta-plane turbulence. Phys. Fluids **16**, 2592–2603 (2004)
Danilov, S., Wang, Q.: Resolving eddies by local mesh refinement. Ocean Model. **93**, 75–83 (2015)
Deardorff, J.W.: The use of subgrid transport equations in a three-dimensional model of atmospheric turbulence. J. Fluids Eng. **95**(3), 429–438 (1973)
Duan, J., Nadiga, B.T.: Stochastic parameterization for large eddy simulation of geophysical flows. Proc. Am. Math. Soc. **135**(4), 1187–1196 (2007)
Eden, C.: Parameterising meso-scale eddy momentum fluxes based on potential vorticity mixing and a gauge term. Ocean Model. **32**(1–2), 58–71 (2010)
Eden, C.: Closing the energy cycle in an ocean model. Ocean Model. **101**, 30–42 (2016)
Fjørtoft, R.: On the changes in the spectral distribution of kinetic energy for twodimensional, nondivergent flow. Tellus **5**(3), 225–230 (1953)
Foias, C., Holm, D.D., Titi, E.S.: The Navier-Stokes-alpha model of fluid turbulence. Phys. D **152/153**, 505–519. Advances in nonlinear mathematics and science (2001)
Foias, C., Holm, D.D., Titi, E.S.: The three dimensional viscous Camassa-Holm equations, and their relation to the Navier-Stokes equations and turbulence theory. J. Dyn. Differ. Equ. **14**(1), 1–35 (2002)
Fox-Kemper, B., Menemenlis, D.: Can large eddy simulation techniques improve mesoscale rich ocean models? In: Hecht, M.W., Hasumi, H. (eds.) Ocean Modeling in an Eddying Regime, pp. 319–337. American Geophysical Union (2008)
Franzke, C., O'Kane, T., Berner, J., Williams, P., Lucarini, V.: Stochastic climate theory and modelling. WIREs Clim. Chang. **6**, 63–78 (2015)
Franzke, C.L.E., Oliver, M., Rademacher, J.D.M., Badin, G.: Multi-scale methods for geophysical flows. This volume, Chapter 1 (2019)

Frederiksen, J.S., Davies, A.G.: Eddy viscosity and stochastic backscatter parameterizations on the sphere for atmospheric circulation models. J. Atmos. Sci. **54**, 2475–2492 (1997)
Fuchssteiner, B., Fokas, A.S.: Symplectic structures, their Bäcklund transformations and hereditary symmetries. Phys. D **4**(1), 47–66 (1981)
Gassmann, A.: A global hexagonal C-grid non-hydrostatic dynamical core (ICON-IAP) designed for energetic consistency. Q. J. Roy. Meteorol. Soc. **139**, 152–175 (2013)
Gent, P.R.: The Gent-McWilliams parameterization: 20/20 hindsight. Ocean Model. **39**(1–2), 2–9 (2011)
Gent, P.R., McWilliams, J.C.: Isopycnal mixing in ocean circulation models. J. Phys. Oceanogr. **20**, 150–155 (1990)
Germano, M., Piomelli, U., Moin, P., Cabot, W.H.: A dynamic subgrid-scale eddy viscosity model. Phys. Fluids A **3**(7), 1760–1765 (1991)
Gilbert, A.D., Vanneste, J.: Geometric generalised Lagrangian-mean theories. J. Fluid Mech. **839**, 95–134 (2018)
Gkioulekas, E., Tung, K.K.: A new proof on net upscale energy cascade in two-dimensional and quasi-geostrophic turbulence. J. Fluid Mech. **576**, 173–189 (2007)
Graham, J.P., Ringler, T.: A framework for the evaluation of turbulence closures used in mesoscale ocean large-eddy simulations. Ocean Model. **65**, 25–39 (2013)
Grooms, I., Lee, Y., Majda, A.J.: Numerical schemes for stochastic backscatter in the inverse cascade of quasigeostrophic turbulence. Multiscale Model. Sim. **13**(3), 1001–1021 (2015a)
Grooms, I., Majda, A.J.: Efficient stochastic superparameterization for geophysical turbulence. Proc. Nat. Acad. Sci. **110**, 4464–4469 (2013)
Grooms, I., Majda, A.J.: Stochastic superparameterization in quasigeostrophic turbulence. J. Comput. Phys. 271 (2014)
Grooms, I., Majda, A.J., Smith, K.S.: Stochastic superparameterization in a quasigeostrophic model of the Antarctic Circumpolar Current. Ocean Model. **85**, 1–15 (2015b)
Grooms, I., Zanna, L.: A note on 'Toward a stochastic parameterization of ocean mesoscale eddies'. Ocean Model. **113**, 30–33 (2017)
Hallberg, R.: Using a resolution function to regulate parameterizations of oceanic mesoscale eddy effects. Ocean Model. **72**, 92–103 (2013)
Hecht, M.W., Holm, D.D., Petersen, M.R., Wingate, B.A.: The LANS-α and Leray turbulence parameterizations in primitive equation ocean modeling. J. Phys. A: Math. Theor. **41**(34), 344009 (2008a)
Hecht, M.W., Hunke, E., Maltrud, M., Petersen, M.R., Wingate, B.A.: Lateral mixing in the eddying regime and a new broad-ranging formulation. In: Hecht, M.W., Hasumi, H., (eds.) Ocean Modeling in an Eddying Regime, pp. 339–352. American Geophysical Union (2008b)
Held, I.M., Larichev, V.D.: A scaling theory for horizontally homogeneous, baroclinically unstable flow on a beta-plane. J. Atmos. Sci. **53**, 946–952 (1996)
Holm, D.D.: Fluctuation effects on 3D Lagrangian mean and Eulerian mean fluid motion. Phys. D **133**(1), 215–269 (1999)
Holm, D.D.: Averaged Lagrangians and the mean effects of fluctuations in ideal fluid dynamics. Phys. D **170**(3), 253–286 (2002)
Holm, D.D., Marsden, J.E., Ratiu, T.S.: The Euler-Poincaré equations and semidirect products with applications to continuum theories. Adv. Math. **137**(1), 1–81 (1998)
Holm, D.D., Nadiga, B.T.: Modeling mesoscale turbulence in the barotropic double-gyre circulation. J. Phys. Oceanogr. **33**(11), 2355–2365 (2003)
Ilıcak, M., Adcroft, A.J., Griffies, S.M., Hallberg, R.W.: Spurious dianeutral mixing and the role of momentum closure. Ocean Model. **45**, 37–58 (2012)
Jansen, M.F., Held, I.M.: Parameterizing subgrid-scale eddy effects using energetically consistent backscatter. Ocean Model. **80**, 36–48 (2014)
Jansen, M.F., Held, I.M., Adcroft, A.J., Hallberg, R.: Energy budget-based backscatter in an eddy permitting primitive equation model. Ocean Model. **94**, 15–26 (2015)

Kitsios, V., Frederiksen, J., Zidikheri, M.: Scaling laws for parameterisations of subgrid eddy-eddy interactions in simulations of oceanic circulations. Ocean Model. **68**, 88–105 (2013)
Kitsios, V., Frederiksen, J.S., and Zidikheri, M.J.: Scaling laws for parametrizations of subgrid interactions in simulations of oceanic circulations. Proc. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci. 372(2018) (2014)
Kitsios, V., Frederiksen, J.S., Zidikheri, M.J.: Theoretical comparison of subgrid turbulence in atmospheric and oceanic quasi-geostrophic models. Nonlinear Proc. Geophys. **23**, 95–105 (2016)
Klein, P., Hua, B.L., Lapeyre, G., Capet, X., Le Gentil, S., Sasaki, H.: Upper-ocean turbulence from high-resolution 3D simulations. J. Phys. Oceanogr. **38**, 1748–1763 (2008)
Klingbeil, K., Burchard, H., Danilov, S., Goetz, C., Iske, A.: Reducing spurious diapycnal mixing in ocean models. This volume, Chapter 8 (2019)
Korn, P.: Formulation of an unstructured grid model for global ocean dynamics. J. Comput. Phys. **339**, 525–552 (2017)
Kouranbaeva, S.: The Camassa-Holm equation as a geodesic flow on the diffeomorphism group. J. Math. Phys. **40**(2), 857–868 (1999)
Kraichnan, R.H.: The structure of isotropic turbulence at very high Reynolds numbers. J. Fluid Mech. **5**(4), 497–543 (1959)
Kraichnan, R.H.: Inertial ranges in two-dimensional turbulence. Phys. Fluids **10**, 1417–1423 (1967)
Lapeyre, G.: What vertical mode does the altimeter reflect? On the decomposition in baroclinic modes and a surface-trapped mode. J. Phys. Oceanogr. **39**, 2857–2874 (2009)
Large, W.G., McWilliams, J.C., Doney, S.C.: Oceanic vertical mixing: A review and a model with a nonlocal boundary layer parameterization. Rev. Geophys. **32**(4), 363–403 (1994)
Leith, C.E.: Diffusion approximation for two-dimensional turbulence. Phys. Fluids **11**, 671–673 (1968)
Leith, C.E.: Atmospheric predictability and two-dimensional turbulence. J. Atmos. Sci. **28**, 145–161 (1971)
Lunasin, E., Kurien, S., Taylor, M.A., Titi, E.S.: A study of the Navier-Stokes-α model for two-dimensional turbulence. J. Turbul. **8**, N30 (2007)
Maltrud, M.E., Vallis, G.K.: Energy and enstrophy transfer in numerical simulations of two-dimensional turbulence. Phys. Fluids A **5**(7), 1760–1775 (1993)
Mana, P.P., Zanna, L.: Toward a stochastic parameterization of ocean mesoscale eddies. Ocean Model. **79**, 1–20 (2014)
Marsden, J.E., Shkoller, S.: Global well-posedness for the Lagrangian averaged Navier-Stokes (LANS-$alpha$) equations on bounded domains. R. Soc. Lond. Philos. Trans. Ser. A Math. Phys. Eng. Sci. **359**(1784), 1449–1468 (2001)
Marsden, J.E., Shkoller, S.: The anisotropic Lagrangian averaged Euler and Navier-Stokes equations. Arch. Ration. Mech. Anal. **166**(1), 27–46 (2003)
Marshall, D.P., Adcroft, A.J.: Parameterization of ocean eddies: potential vorticity mixing, energetics and Arnold's first stability theorem. Ocean Model. **32**(3–4), 188–204 (2010)
Marshall, D.P., Maddison, J.R., Berloff, P.S.: A framework for parameterizing eddy potential vorticity fluxes. J. Phys. Oceanogr. **42**(4), 539–557 (2012)
Meneveau, C., Katz, J.: Scale-invariance and turbulence models for large-eddy simulation. Ann. Rev. Fluid Mech. **32**(1), 1–32 (2000)
Moffatt, H.: Note on the triad interactions of homogeneous turbulence. J. Fluid Mech. **741**, R3 (2014)
Mohammadi-Aragh, M., Klingbeil, K., Brüggemann, N., Eden, C., Burchard, H.: The impact of advection schemes on restratifiction due to lateral shear and baroclinic instabilities. Ocean Model. **94**, 112–127 (2015)
Mohseni, K., Kosović, B., Shkoller, S., Marsden, J.E.: Numerical simulations of the Lagrangian averaged Navier-Stokes equations for homogeneous isotropic turbulence. Phys. Fluids **15**(2), 524–544 (2003)
O'Kane, T.J., Frederiksen, J.S.: Statistical dynamical subgrid-scale parameterizations for geophysical flows. Phys. Scripta **2008**(T132), 014033 (2008)

Oliver, M.: Lagrangian averaging with geodesic mean. Proc. R. Soc. A **473**(2207), 20170558, 9 (2017)
Oliver, M., Shkoller, S.: The vortex blob method as a second-grade non-Newtonian fluid. Commun. Partial Differ. Equ. **26**(1–2), 295–314 (2001)
Palmer, T., Buizza, R., Doblas-Reyes, F., Jung, T., Leutbecher, M., Shutts, G., Steinheimer, M., Weisheimer, A.: Stochastic parametrization and model uncertainty. Technical report, ECMWF (2009)
Rhines, P.B.: Waves and turbulence on a beta-plane. J. Fluid Mech. **69**, 417–443 (1975)
Rhines, P.B.: The dynamics of unsteady currents. In: Goldberg, E.A., McCane, I.N., O'Brien, J.J., Steele, J.H. (eds.) The Sea, vol 6, pp. 189–318. J. Wiley (1977)
Ringler, T., Gent, P.: An eddy closure for potential vorticity. Ocean Model. **39**(1–2), 125–134 (2011)
Sadourny, R., Basdevant, C.: Parameterization of subgrid scale barotropic and baroclinic eddies in quasi-geostrophic models: anticipated potential vorticity method. J. Atmos. Sci. **42**, 1353–1363 (1985)
Salmon, R.: Baroclinic instability and geostrophic turbulence. Geophys. Astrophys. Fluid Dyn. **15**(1), 167–211 (1980)
Salmon, R.: Lectures on Geophysical Fluid Dynamics. Oxford University Press (1998)
San, O.: A dynamic eddy-viscosity closure model for large eddy simulations of two-dimensional decaying turbulence. Int. J. Comput. Fluid D. **28**, 363–382 (2014)
Schiestel, R., Dejoan, A.: Towards a new partially integrated transport model for coarse grid and unsteady turbulent flow simulations. Theor. Comput. Fluid Dyn. **18**(6), 443–468 (2005)
Schmidt, H., Schumann, U.: Coherent structure of the convective boundary layer derived from large-eddy simulations. J. Fluid Mech. **200**, 511–562 (1989)
Schumann, U.: Subgrid scale model for finite difference simulations of turbulent flows in plane channels and annuli. J. Comput. Phys. **18**(4), 376–404 (1975)
Schumann, U.: Subgrid length-scales for large-eddy simulation of stratified turbulence. Theor. Comput. Fluid Dyn. **2**(5–6), 279–290 (1991)
Scott, R.B., Arbic, B.K.: Spectral energy fluxes in geostrophic turbulence: implications for ocean energetics. J. Phys. Oceanogr. **37**, 673–688 (2007)
Smith, K.S., Boccaletti, G., Henning, C.C., Marinov, I.N., Tam, C.Y., Held, I.M., Vallis, G.K.: Turbulent diffusion in the geostrophic inverse cascade. J. Fluid Mech. **469**, 13–48 (2002)
Stone, P.H.: A simplified radiative-dynamical model for the static stability of rotating atmospheres. J. Atmos. Sci. **29**, 405–418 (1972)
Taylor, G.I.: The spectrum of turbulence. Proc. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci. **164**(919), 476–490 (1938)
Truesdell, C., Rajagopal, K.: An Introduction to the Mechanics of Fluids. Birkhäuser (1999)
Umlauf, L., Burchard, H.: A generic length-scale equation for geophysical turbulence models. J. Marine Res. **61**(2), 235–265 (2003)
Vallis, G.K.: Atmospheric and Oceanic Fluid Dynamics: Fundamentals and Large-Scale Circulation. Cambridge University Press (2006)
von Storch, J.-S., Badin, G., Oliver, M.: The interior energy pathway: inertial gravity wave emission by oceanic flows. This volume Chapter 2 (2019)
Wang, Q., Danilov, S., Sidorenko, D., Timmermann, R., Wekerle, C., Wang, X., Jung, T., Schröter, J.: The finite element sea ice–ocean model (FESOM) v.1.4: formulation of an ocean general circulation model. Geosci. Model Dev. **7**(2), 663–693 (2014)
Wunsch, C.: The vertical partition of oceanic horizontal kinetic energy. J. Phys. Oceanogr. **27**, 1770–1794 (1997)
Xun, Q.-Q., Wang, B.-C.: A dynamic forcing scheme incorporating backscatter for hybrid simulation. Phys. Fluids **26**(7), 075104 (2014)
Yaglom, A.M.: Correlation Theory of Stationary and Related Random Functions. Basic Results, vol. 1. Springer (1987)
Zanna, L., Mana, P.P., Anstey, J., David, T., Bolton, T.: Scale-aware deterministic and stochastic parametrizations of eddy-mean flow interaction. Ocean Model. **111**, 66–80 (2017)