The Role of Mixed Layer Instabilities in Submesoscale Turbulence

Jörn Callies

1 Introduction

The upper ocean can support energetic flows at scales smaller than the order 100 km mesoscale eddies. Sharp surface fronts associated with strong along-front currents emerge in high-resolution numerical simulations [8, 16] and are observed in the wintertime mid-latitude ocean [7]. These submesoscale flows are associated with large vertical fluxes of both physical and biogeochemical tracers and may thereby regulate the uptake of heat and carbon from the atmosphere [8, 17, 11]. What drives these submesoscale flows?

Two mechanisms have been proposed: surface frontogenesis [19, 28] and mixed layer instabilities [3]. The process underlying surface frontogenesis can be understood with quasi-geostrophic (QG) dynamics [32]. In the interior of the ocean, when a strain field increases a horizontal buoyancy gradient, an ageostrophic circulation develops according to the omega equation [13] and acts to oppose the increase of the buoyancy gradient. Light water downwells on the dense side and dense water upwells on the light side of the buoyancy gradient. At the surface, however, the vertical velocity must vanish and the ageostrophic circulation cannot act to oppose the increase of the buoyancy gradient in the same way—the mesoscale strain field is left to create strong surface fronts.

The simplest model of these dynamics is the surface QG model [10, 32, 1, 12]. It assumes an infinitely deep ocean with constant interior potential vorticity (PV),

$$\nabla^2\psi + \frac{\partial}{\partial z}\left(\frac{f^2}{N^2}\frac{\partial\psi}{\partial z}\right) = 0, \tag{1}$$

where ψ is the geostrophic streamfunction, f is the Coriolis frequency, and N the buoyancy frequency, such that the dynamics are completely determined by the advection of buoyancy at the surface,

$$\frac{\partial b}{\partial t} + \mathrm{J}(\psi, b) = 0, \tag{2}$$

where $b = f\partial\psi/\partial z$ is buoyancy and

$$\mathrm{J}(a,b) = \frac{\partial a}{\partial x}\frac{\partial b}{\partial y} - \frac{\partial a}{\partial y}\frac{\partial b}{\partial x} \tag{3}$$

is the Jacobian operator. This supplies the boundary condition for the elliptic problem (1). Straining by mesoscale eddies creates sharp buoyancy gradients and strong associated flows at the surface. Secondary instabilities lead to eventually fully turbulent dynamics [12],

for which Kolmogorov-like dimensional arguments predict the energy spectrum in the submesoscale range to scale like $E_{k_\mathrm{h}} \sim {k_\mathrm{h}}^{-5/3}$ in an inertial range in which surface buoyancy variance is cascaded to small scales [1]. If non-QG dynamics are taken into account, ageostrophic advection of buoyancy further accelerates frontogenesis and leads to frontal collapse, the formation of true discontinuities [14]. In this case, the submesoscale energy spectrum is modified to $E_{k_\mathrm{h}} \sim {k_\mathrm{h}}^{-2}$ [4].

As opposed to what is assumed in surface QG dynamics, however, the upper ocean does not have a nearly constant PV. Instead, there is typically a weakly stratified mixed layer with low PV overlying a strongly stratified thermocline with high PV. There is a sharp step-like increase in PV at the base of the mixed layer. This PV step is dynamically important, because it supports edge waves that have the potential to interact with surface edge waves and thus produce a baroclinic instability in the mixed layer [3]. In winter, when mixed layers are deep, this mixed layer instability occurs at scales of 1–10 km, so it has the potential to energize submesoscale turbulence.

The importance of mixed layer instabilities is hinted at by the observation that submesoscale turbulence undergoes a seasonal cycle. Both modeling [22] and observations [7] suggest that submesoscale turbulence is energized in winter and suppressed in summer. Mixed layer instabilities are expected to undergo a strong seasonal cycle, following the seasonal cycle of mixed layer depth itself. In the frontogenetic picture, on the other hand, submesoscale turbulence is driven by mesoscale eddies that do not exhibit a strong seasonal cycle.

The goal of this report is to understand the dynamics of mixed layer instabilities and of the turbulent dynamics that emerge when these instabilities grow to finite amplitude. We formulate a simple QG model consisting of two constant-PV layers representing the mixed layer and the thermocline, coupled at a deformable interface. Despite its simplicity, this model captures both mesoscale and mixed layer instabilities and thereby a number of fundamental aspects of submesoscale dynamics. The model also allows straightforward comparison to surface QG dynamics, in which no mixed layer is present.

We use QG scaling to formulate the dynamics of our model, which requires small Rossby and Froude numbers [24]. Typical mesoscale Rossby and Froude numbers are on the order 0.1 and increase slowly with wavenumber if the submesoscale is energetic [7]. To leading order, submesoscale flows are thus expected to follow QG dynamics. Higher-order effects can become important at submesoscales, however, and we take up the discussion of non-QG effects in the conclusions.

2 Model formulation

Consider two layers of constant PV on an f-plane, with constant stratification and constant mean shear (Fig. 1). The upper layer represents the mixed layer, which has a mean depth h, stratification N_m, and mean shear Λ_m. The lower layer represents the thermocline and has stratification N_t and mean shear Λ_t. The total depth is H. The layers are coupled by a deformable interface; a rigid lid condition is applied at the surface; a flat bottom condition is applied at the bottom. The presence of a bottom at the base of the thermocline is not realistic, but we will see that the bottom layer will still represent important properties of the thermocline. The approximation that the stratification is discontinuous at the base of

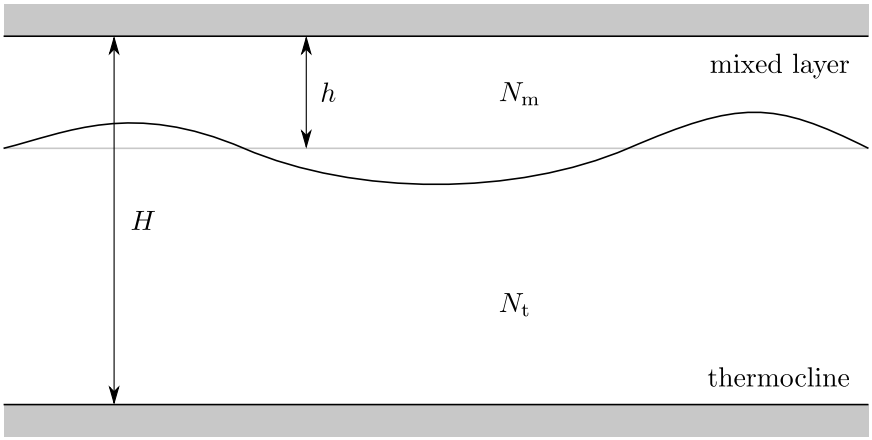

Figure 1: Schematic of the model setup in a vertical–horizontal plane. There are rigid surfaces at $z = 0$ and $z = -H$ and a deformable interface at the mean depth $z = -h$, separating layers of constant stratifications, N_m in the mixed layer and N_t in the thermocline.

the mixed layer is appropriate at horizontal scales larger than the deformation radius Nd/f associated with the transition depth d [30]. The transition at the base of the mixed layer is typically quite sharp, so this deformation radius is much smaller than the submesoscales we are interested in here.

The uniform PV within the two layers simplifies the dynamics dramatically. PV conservation within the layers is trivial, like in the classic Eady problem [10]. The flow in the interior of the layers is obtained by solving (1), with the boundary conditions supplied by the distribution of buoyancy at the surface and bottom and by matching conditions at the interface between the mixed layer and the thermocline.

At the rigid boundaries at the surface and bottom, where the vertical velocity w vanishes, the advection of buoyancy anomalies b is given by

$$\frac{\partial b}{\partial t} + \mathrm{J}(\psi, b) = 0. \tag{4}$$

To ensure that pressure is continuous at the interface, we require that the streamfunction ψ is continuous. For mass conservation, we require that the vertical velocity w is also continuous. These conditions are applied at $z = -h$, consistent with QG scaling. The conservation equations for buoyancy just above and below the interface at $z = -h$,

$$\frac{\partial b^+}{\partial t} + \mathrm{J}(\psi, b^+) + wN_\mathrm{m}^2 = 0, \qquad \frac{\partial b^-}{\partial t} + \mathrm{J}(\psi, b^-) + wN_\mathrm{t}^2 = 0, \tag{5}$$

can then be combined to eliminate w. This gives

$$\frac{\partial \theta_1}{\partial t} + \mathrm{J}(\psi_1, \theta_1) = 0, \tag{6}$$

where ψ_1 denotes the streamfunction at $z = -h$. This is a conservation equation for

$$\theta_1 = f\left(\frac{b^+}{N_\mathrm{m}^2} - \frac{b^-}{N_\mathrm{t}^2}\right). \tag{7}$$

The quantity θ_1 is nothing but the PV

$$q = \nabla^2\psi + f\frac{\partial}{\partial z}\left(\frac{b}{N^2}\right) \tag{8}$$

integrated across the interface. While there are no PV anomalies within the two layers, the displacement of the interface between the two layers of constant PV induces a PV anomaly that, according to (6) and consistent with QG dynamics, is advected by the geostrophic flow at $z = -h$. The conservation equation (6) has been used to study the dynamics of the tropopause, which is similarly an interface between weakly stratified fluid in the troposphere and strongly stratified fluid in the stratosphere [10, 27, 15, 12].

The model can be considered as consisting of three PV sheets:

$$q = \theta_0\delta(z) + \theta_1\delta(z+h) + \theta_2\delta(z+H), \tag{9}$$

where δ is Dirac's delta function and $\theta_0 = -fb/N_{\mathrm{m}}^2$ at $z = 0$ and $\theta_2 = fb/N_{\mathrm{t}}^2$ at $z = -H$. PV is advected by the geostrophic flow, so

$$\frac{\partial\theta_i}{\partial t} + \mathrm{J}(\psi_i, \theta_i) = 0, \tag{10}$$

where $i = 0, 1, 2$ and ψ_i is the streamfunction at the level corresponding to θ_i. This formulation is simply an extension of Bretherton's representation of boundary conditions [5] to include an interior PV sheet due to deflection of an interface between layers of different stratification.

Note that the statement that θ_1 is only advected by the geostrophic flow does not imply that the vertical velocity vanishes at the interface, just like the fact that PV anomalies in the QG system are advected only by the geostrophic flow does not imply that the vertical velocity vanishes. The vertical velocity is implicit in the dynamics and can be solved for using the omega equation.

To complete the dynamics, we require an inversion relation that allows us to obtain the streamfunctions ψ_i from the conserved quantities θ_i. This relation can be written as a linear equation for the Fourier coefficients, denoted by subscripts k, l:

$$\boldsymbol{\theta}_{k,l} = L\boldsymbol{\psi}_{k,l}, \qquad \boldsymbol{\theta} = (\theta_0, \theta_1, \theta_2)^{\mathsf{T}}, \qquad \boldsymbol{\psi} = (\psi_0, \psi_1, \psi_2)^{\mathsf{T}}. \tag{11}$$

The matrix L is determined by solving

$$-k_{\mathrm{h}}^2\psi_{k,l} + \frac{\partial}{\partial z}\left(\frac{f^2}{N^2}\frac{\partial\psi_{k,l}}{\partial z}\right) = 0 \tag{12}$$

in each layer. For example, the first column of L is determined by setting $\boldsymbol{\psi} = (1, 0, 0)^{\mathsf{T}}$, solving (12) for $\psi(z)$, and subsequently calculating

$$\theta_0 = -\frac{f^2}{N_{\mathrm{m}}^2}\frac{\partial\psi}{\partial z}(0), \qquad \theta_1 = \frac{f^2}{N_{\mathrm{m}}^2}\frac{\partial\psi}{\partial z}(-h^+) - \frac{f^2}{N_{\mathrm{t}}^2}\frac{\partial\psi}{\partial z}(-h^-), \qquad \theta_2 = \frac{f^2}{N_{\mathrm{t}}^2}\frac{\partial\psi}{\partial z}(-H). \tag{13}$$

This procedure gives

$$L = fk_{\mathrm{h}}\begin{pmatrix} -\frac{\coth\mu_{\mathrm{m}}}{N_{\mathrm{m}}} & \frac{\operatorname{csch}\mu_{\mathrm{m}}}{N_{\mathrm{m}}} & 0 \\ \frac{\operatorname{csch}\mu_{\mathrm{m}}}{N_{\mathrm{m}}} & -\frac{\coth\mu_{\mathrm{m}}}{N_{\mathrm{m}}} - \frac{\coth\mu_{\mathrm{t}}}{N_{\mathrm{t}}} & \frac{\operatorname{csch}\mu_{\mathrm{t}}}{N_{\mathrm{t}}} \\ 0 & \frac{\operatorname{csch}\mu_{\mathrm{t}}}{N_{\mathrm{t}}} & -\frac{\coth\mu_{\mathrm{t}}}{N_{\mathrm{t}}} \end{pmatrix}, \tag{14}$$

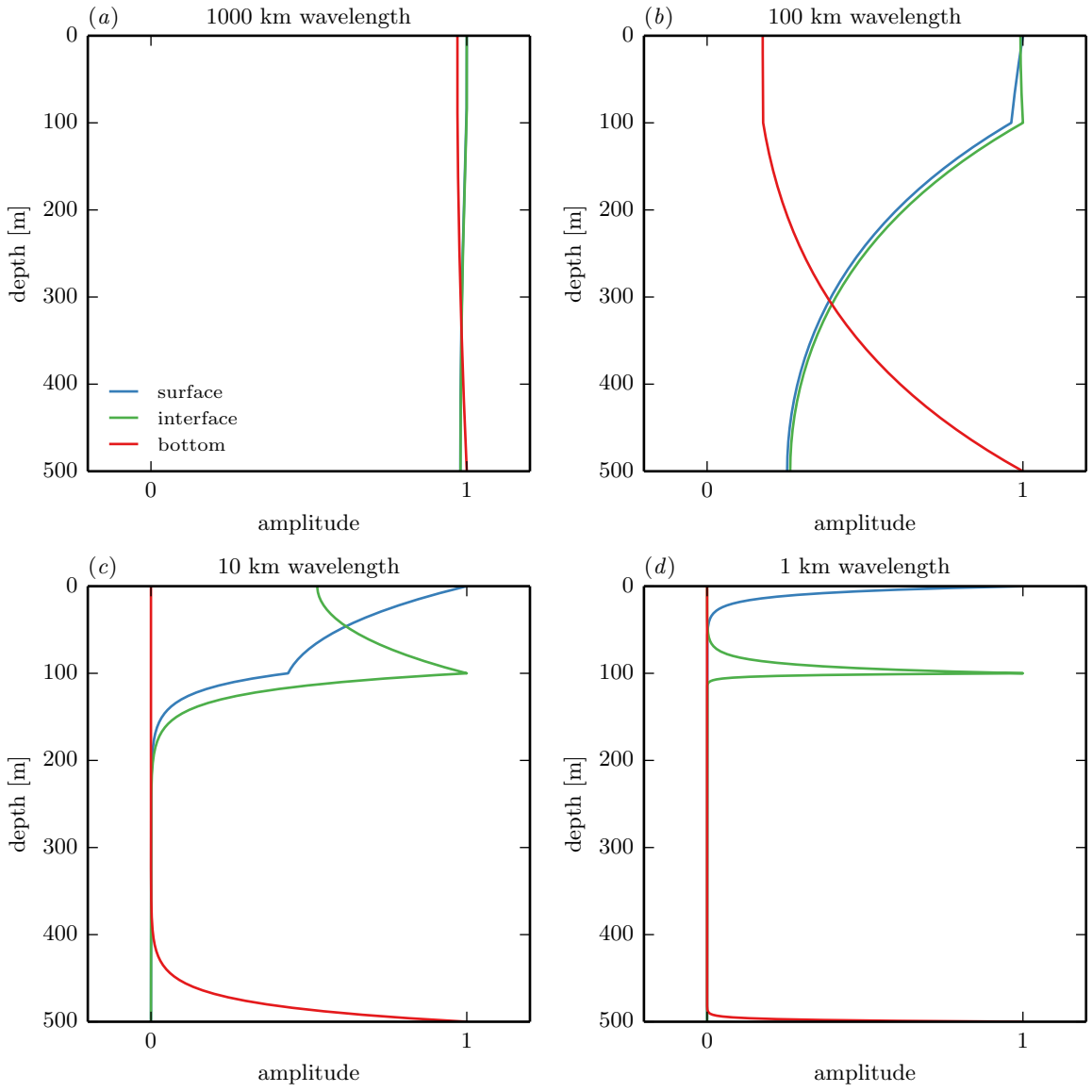

Figure 2: Vertical structure of streamfunction amplitude associated with anomalies of θ_0 (surface), θ_1 (interface), and θ_2 (bottom). Shown are the vertical profiles for θ_i anomalies with different horizontal wavenumbers $k_\mathrm{h} = 2\pi/\lambda$. The wavelength λ is given in the panel titles.

where $\mu_\mathrm{m} = N_\mathrm{m} k_\mathrm{h} h/f$ and $\mu_\mathrm{t} = N_\mathrm{t} k_\mathrm{h}(H-h)/f$ are nondimensional wavenumbers. This 3×3 matrix can easily be inverted.

This model can be generalized to an arbitrary number of layers of constant stratification and shear, which may be a useful way to approximate more realistic stratification and shear profiles. This is discussed in Appendix A. The model can also be extended to include a density jump at the interface, as is sometimes present at the base of the mixed layer. The formulation is given in Appendix B, but we here restrict ourselves to the case of a continuous density profile.

To build intuition for the dynamics of the model, we illustrate the vertical structure of flow associated with anomalies of the conserved quantities at the surface, the interface, and the bottom (Fig. 2). We apply parameters that will be used throughout the report, typical of the wintertime midlatitude ocean (Tab. 1). At the largest scales, for anomalies with wavelength $\lambda = 1000$ km or $k_\mathrm{h} \ll f/N_\mathrm{t} H$, the flow is nearly barotropic, irrespective of which level the anomaly is at. At smaller scales, $\lambda = 100$ km or $k_\mathrm{h} \sim f/N_\mathrm{t} H \ll f/N_\mathrm{m} h$, there is significant decay in the thermocline while the flow is nearly barotropic in the mixed layer. Anomalies at the surface and interface still induce significant flow at the bottom and vice versa. At $\lambda = 10$ km or $k_\mathrm{h} \sim f/N_\mathrm{m} h$, on the other hand, anomalies at the surface or interface induce very little flow at the bottom and vice versa. The flow is also not barotropic in the mixed layer anymore, but surface anomalies still induce significant flow at the interface and vice versa. At $\lambda = 1$ km or $k_\mathrm{h} \gg f/N_\mathrm{m} h$, all levels are decoupled: anomalies on any of the levels induce very little flow at the other levels.

The dependence of the vertical flow structure on the horizontal scale of the anomalies illuminates the qualitative dynamics of the model. At the smallest scales, all three levels are independent and follow surface QG dynamics. At scales $k_\mathrm{h} \sim f/N_\mathrm{m} h$, around the mixed layer deformation radius, surface and interface anomalies can interact, allowing for phase locking and instability in the mixed layer. Bottom anomalies, on the other hand are independent, so there is no deep instability at these scales. At scales $k_\mathrm{h} \sim f/N_\mathrm{t} H$, around the deep deformation radius, surface or interface anomalies can interact with bottom anomalies, so there is potential for a deep instability at these scales. At the largest scales, the flow is essentially barotropic and follows two-dimensional dynamics.

3 Linear stability analysis

We now analyse the linear stability of the model formulated above. While this linear analysis is not directly applicable to the strongly nonlinear turbulent regime, it reveals some key characteristics of the dynamics that will help us understand the nonlinear regime.

Blumen analyzed short-wave instabilities in the atmosphere also using a model consisting of two coupled constant-PV layers [2]. He performed a linear stability analysis equivalent to what will be presented in the following. For completeness, we review the linear stability in the context of the upper ocean dynamics, at the cost of being somewhat redundant with Blumen's study.

We consider the linear stability of normal-mode perturbations to a zonal flow with constant vertical shear Λ_m in the mixed layer and Λ_t in the thermocline (Fig. 3). The system is Galilean invariant, so we can arbitrarily set the mean zonal flow to zero at the surface. The linearized conservation equations for the perturbations from this mean state

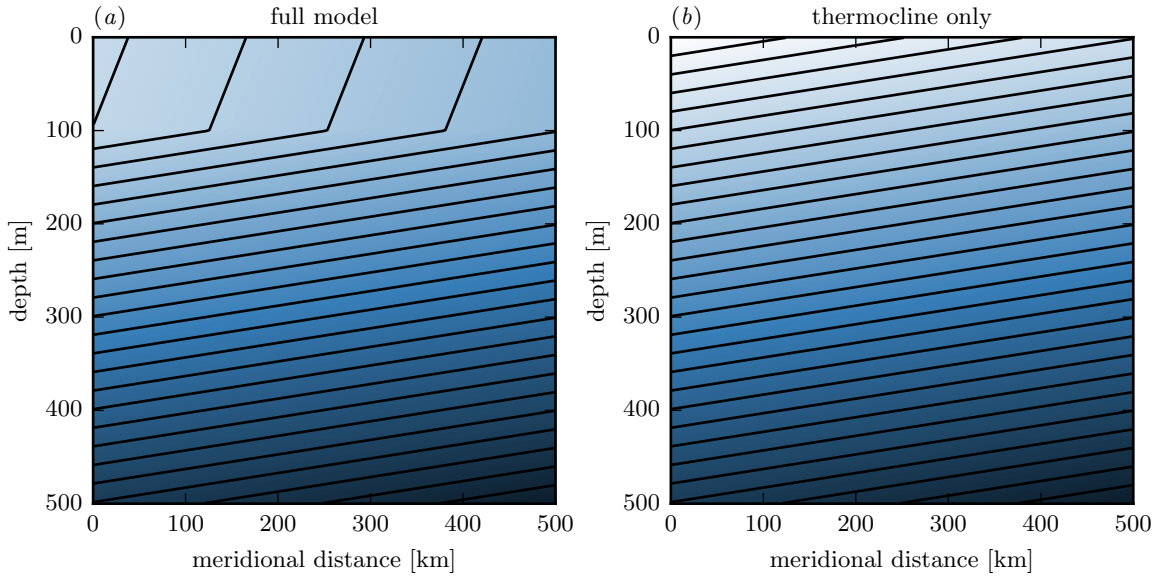

Figure 3: Mean buoyancy structure in the meridional–vertical plane for (a) the full model and (b) the thermocline only case. The contours show isopycnals; light shading indicates more buoyant fluid.

are

$$\frac{\partial \boldsymbol{\theta}_{k,l}}{\partial t} + ikU\boldsymbol{\theta}_{k,l} + ik\Gamma\boldsymbol{\psi}_{k,l} = 0, \tag{15}$$

where the mean zonal flows and mean meridional PV gradients at the respective levels are denoted by the diagonal elements of the matrices U and Γ:

$$U = \operatorname{diag}\big(0, -\Lambda_\mathrm{m} h, -\Lambda_\mathrm{m} h - \Lambda_\mathrm{t}(H-h)\big), \tag{16}$$

$$\Gamma = \operatorname{diag}\big(f^2\Lambda_\mathrm{m}/N_\mathrm{m}^2, -f^2\Lambda_\mathrm{m}/N_\mathrm{m}^2 + f^2\Lambda_\mathrm{t}/N_\mathrm{t}^2, -f^2\Lambda_\mathrm{t}/N_\mathrm{t}^2\big). \tag{17}$$

Using the inversion relation (11), we can replace the $\boldsymbol{\theta}_{k,l}$, such that

$$L\frac{\partial \boldsymbol{\psi}_{k,l}}{\partial t} + ikUL\boldsymbol{\psi}_{k,l} + ik\Gamma\boldsymbol{\psi}_{k,l} = 0. \tag{18}$$

Assuming that modes vary harmonically in time with (complex) frequency ω turns this into the generalized eigenvalue problem

$$(UL + \Gamma)\boldsymbol{\psi}_{k,l} = cL\boldsymbol{\psi}_{k,l}, \tag{19}$$

where the eigenvalue is $c = \omega/k$. The real part of c is the zonal phase speed; the imaginary part gives the growth rate $\sigma = k\,\mathrm{Im}\,c$.

Being a third-order system, (19) can be solved analytically, but the solutions are rather complicated and give little useful insight. We instead explore the characteristics of the solutions numerically for the set of parameters given above. We then explain the stability properties and parameter dependencies by considering simplifications of the model.

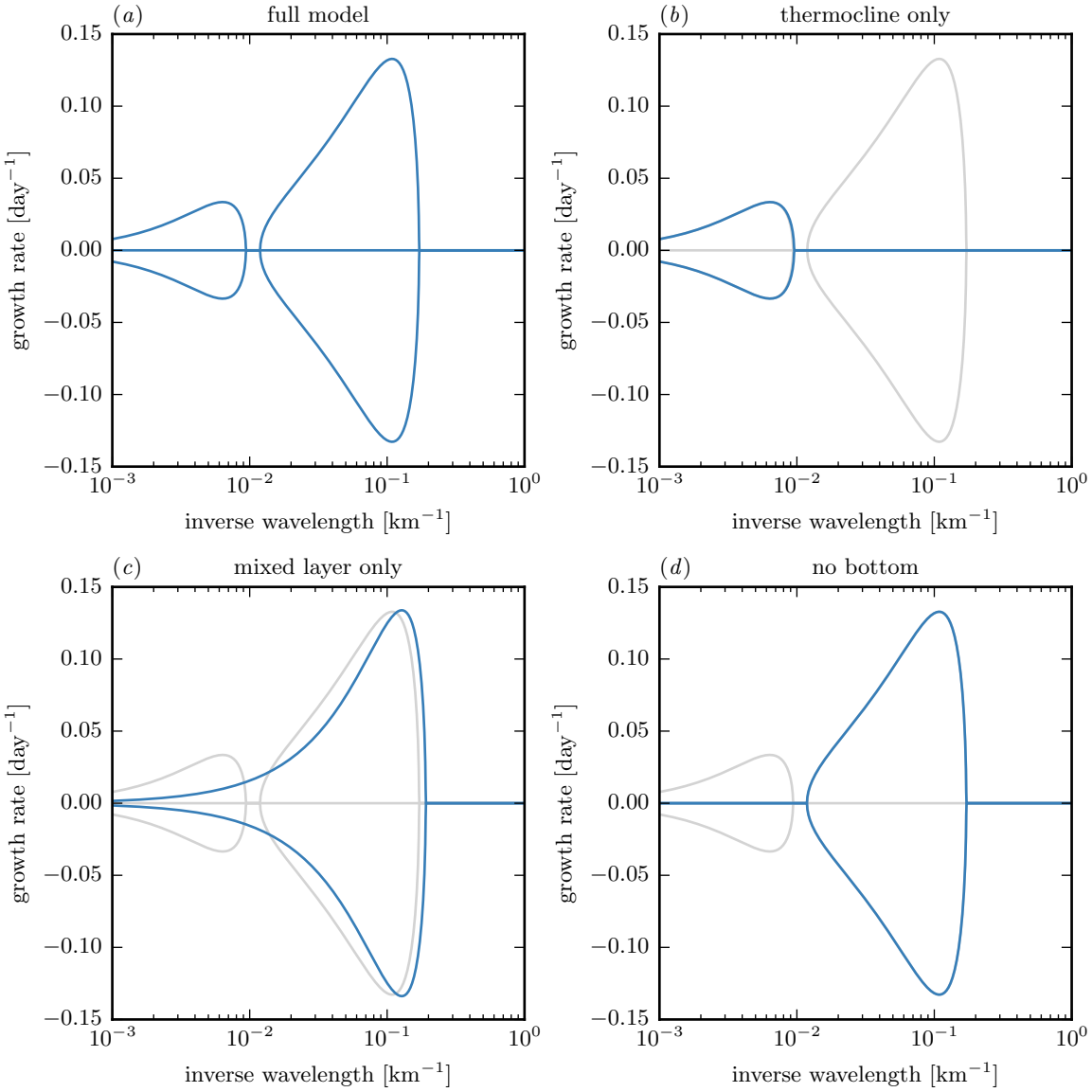

Figure 4: Growth rates from the linear stability analysis for (*a*) the full model incorporating a mixed layer, (*b*) an Eady model representing the thermocline only, (*c*) an Eady model representing the mixed layer only, and (*d*) a model like the full model but without a bottom. Growth rates are shown in blue, the growth rates of the full model are overlaid for reference in gray.

Parameter	Symbol	Value
Mixed layer depth	h	100 m
Total depth	H	500 m
Mixed layer stratification	N_m	2×10^{-3} s^{-1}
Thermocline stratification	N_t	8×10^{-3} s^{-1}
Mixed layer shear	Λ_m	10^{-4} s^{-1}
Thermocline shear	Λ_t	10^{-4} s^{-1}
Coriolis frequency	f	10^{-4} s^{-1}
Domain size	a	500 km

Table 1: Parameters used throughout this report unless otherwise noted. These are typical of the wintertime midlatitude ocean.

3.1 Full model

The linear stability analysis reveals that there are two lobes of instability: one at the mesoscale and one at the submesoscale (Fig. 4*a*). The maximum growth rates occur at $l = 0$ and zonal wavelengths of about 160 km and 10 km. The submesoscale instability has a peak growth rate much larger than the mesoscale instability in this case with equal shear in the two layers. The growing modes are conjugate to decaying modes. The growth rates are similar to what Boccaletti et al. found in a linear QG stability analysis of a realistic mean state of the wintertime eastern subtropical North Pacific [3]. The magnitudes are slightly smaller here, because the shear is slightly weaker.

The phase speeds of the linear modes give clues to the dynamics in different ranges of scales (Fig. 5*a*). For each wavenumber, there are three modes. Growing and decaying modes, being conjugate to each other, have the same phase speeds. This is the familiar phase locking of counter-propagating waves in baroclinic instability (branches 'b' and 'e'). Where these growing and decaying modes exist, there is an additional neutral mode (branches 'a' and 'd'). Where there are no growing modes, all three neutral modes have distinct phase speeds—no phase locking occurs. We will discuss the dynamics of the various branches by considering approximations to the full model.

But first consider the spatial structure of the modes corresponding to the peak growth rates. The perturbation streamfunctions show that the mesoscale mode is deep and spans the entire water column (Fig. 6*a*), whereas the submesoscale mode is almost completely confined to the mixed layer, with only weak leaking into the thermocline below 100 m depth (Fig. 6*b*). Both modes show the familiar pattern of unstable modes tilted into the shear, which is necessary to extract potential energy from the mean.

3.2 Thermocline only

We start explaining the instability properties of the full model by comparing it to the classic Eady model representing the thermocline only (Fig. 3*b*). This amounts to setting $h = 0$ or $N_\mathrm{m} = N_\mathrm{t}$ in the full model. In this case, the system reduces to two variables, the inversion

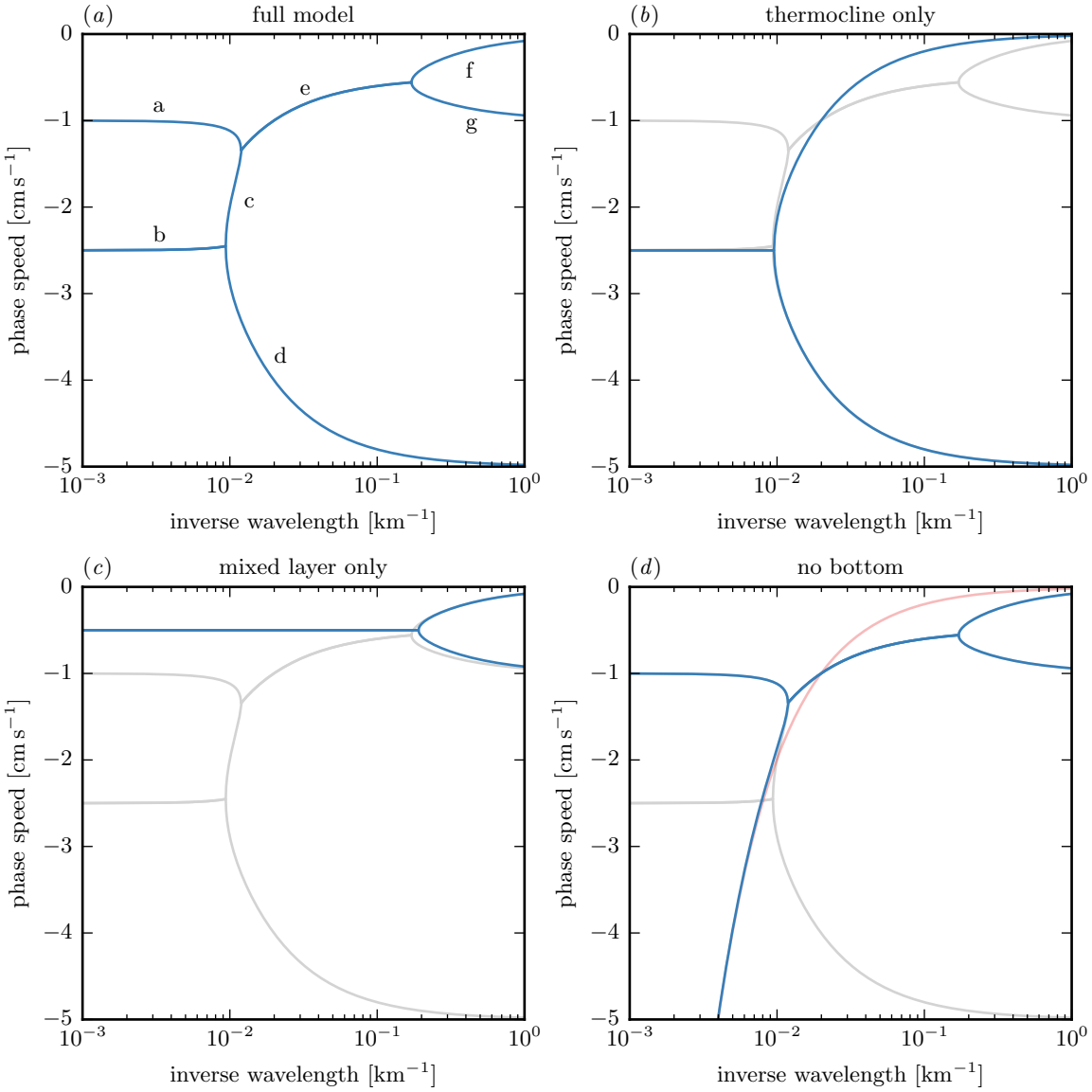

Figure 5: Phase speeds from the linear stability analysis for (*a*) the full model incorporating a mixed layer, (*b*) an Eady model representing the thermocline only, (*c*) an Eady model representing the mixed layer only, and (*d*) a model like the full model but without a bottom. Phase speeds are shown in blue, the growth rates of the full model are overlaid for reference in gray. The faint red line shows the phase speed of a surface edge wave.

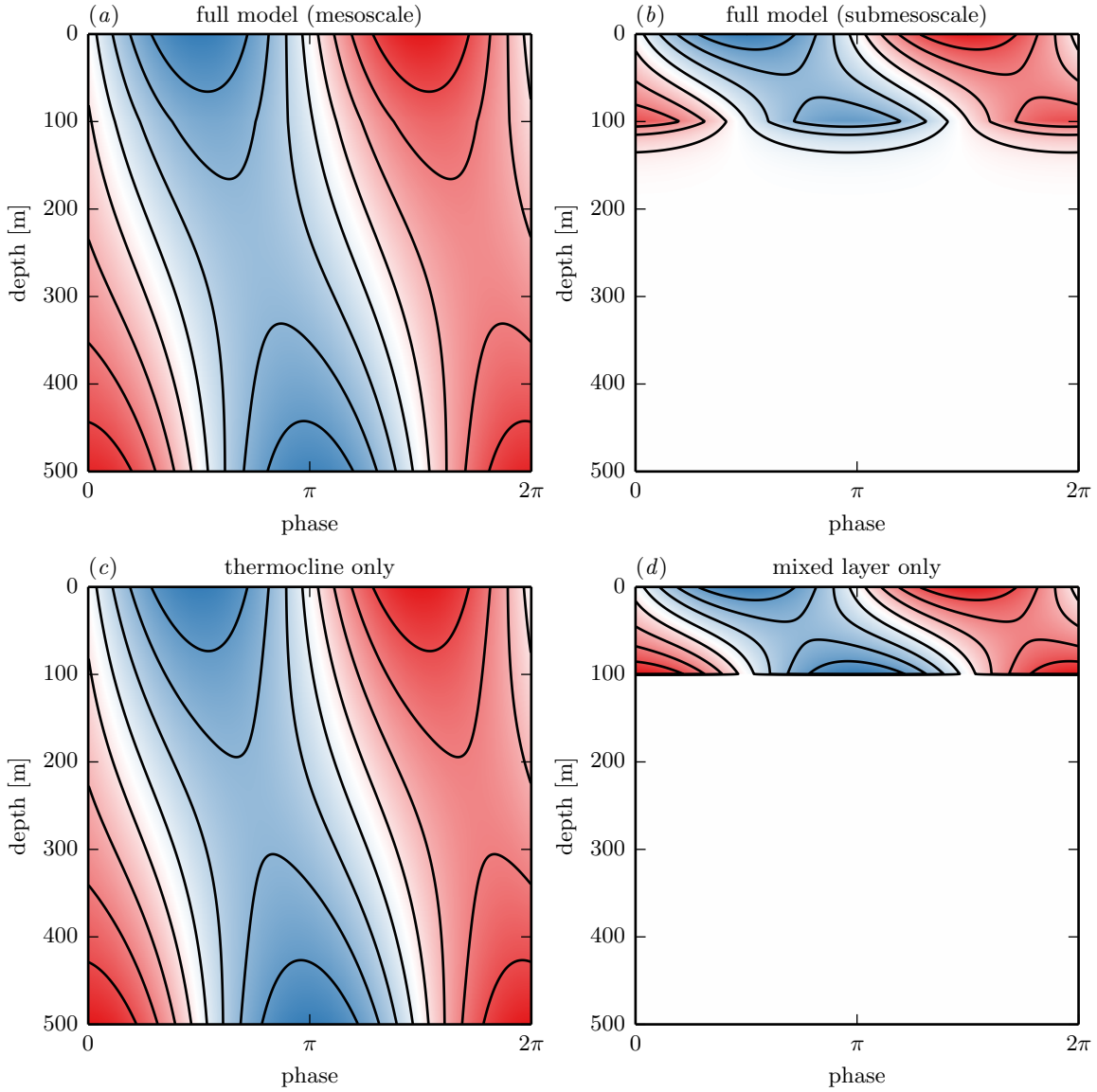

Figure 6: Perturbation streamfunction as obtained from the linear stability analysis of (a) the most unstable mesoscale mode of the full model, (b) the most unstable submesoscale mode of the full model, (c) the most unstable mode of an Eady model representing the thermocline only, and (d) the most unstable mode of an Eady model representing the mixed layer only. Red and blue shading represents positive and negative values.

matrix is

$$L = fk_{\mathrm{h}} \begin{pmatrix} -\frac{\coth \mu_{\mathrm{t}}}{N_{\mathrm{t}}} & \frac{\operatorname{csch} \mu_{\mathrm{t}}}{N_{\mathrm{t}}} \\ \frac{\operatorname{csch} \mu_{\mathrm{t}}}{N_{\mathrm{t}}} & -\frac{\coth \mu_{\mathrm{t}}}{N_{\mathrm{t}}} \end{pmatrix}, \tag{20}$$

and

$$U = \operatorname{diag}\left(0, -\Lambda_{\mathrm{t}} h\right), \tag{21}$$

$$\Gamma = \operatorname{diag}\left(f^2 \Lambda_{\mathrm{t}} / N_{\mathrm{t}}^2, -f^2 \Lambda_{\mathrm{t}} / N_{\mathrm{t}}^2\right). \tag{22}$$

The solution is [10, 33]

$$c = -\frac{\Lambda_{\mathrm{t}} h}{2} \pm \frac{i \Lambda_{\mathrm{t}} h}{\mu_{\mathrm{m}}} \left(\mu_{\mathrm{t}} \coth \mu_{\mathrm{t}} - 1 - \frac{\mu_{\mathrm{t}}^2}{4}\right)^{\frac{1}{2}}, \tag{23}$$

where $\mu_{\mathrm{t}} = N_{\mathrm{t}} k_{\mathrm{h}} H / f$ is the nondimensional wavenumber.

The Eady model has a baroclinic instability near the deformation radius $N_{\mathrm{t}} H / f$. The maximum growth rate $\sigma = 0.31 f \Lambda_{\mathrm{t}} / N_{\mathrm{t}}$ occurs at $l = 0$ and $\mu_{\mathrm{t}} = 1.6$, which corresponds to a zonal wavelength $\lambda = 3.9 N_{\mathrm{t}} H / f$. The growth curve for this thermocline only model traces out the mesoscale lobe of the full model almost perfectly (Fig. 4*b*). The short-wave cutoff in the Eady model at $\lambda = 2.6 N_{\mathrm{t}} H / f$ nearly coincides with the short-wave cutoff of the mesoscale instability in the full model. The phase speed of the phase-locked waves $-\Lambda_{\mathrm{t}} H / 2$ very nearly matches the phase speed of the unstable mesoscale mode of the full model (Fig. 5*b*). The critical level, where the phase speed matches the mean flow, is at mid-depth. The split at the short-wave cutoff into surface and bottom modes also features in the full model. In the thermocline only model, the surface and bottom modes are very nearly Eady edge waves that do not sense the other boundary. The bottom mode of the thermocline only model almost perfectly matches that of the full model (branch 'd'). The surface mode of the thermocline only model traces out branch 'c' of the full model, but then the full model transitions to dynamics associated with the mixed layer that are not present in the thermocline only model. The mesoscale instability of the full model therefore follows Eady dynamics. The presence of the mixed layer only modifies the characteristics of the instability slightly. The spatial structure of the most unstable mesoscale mode is also well captured by the thermocline only model (Fig. 6*c*).

3.3 Mixed layer only

Shifting our attention to the submesoscale instability, we now consider an Eady model representing the mixed layer with a rigid bottom at its base. A priori, we can see from (7) that the full model converges to these dynamics as the thermocline stratification goes to infinity. The thermocline then acts like a rigid bottom at the base of the mixed layer and the conservation of θ_1 turns into a conservation of the buoyancy at the base of the mixed layer $b(-h^+)$.

The Eady model for the mixed layer captures some key properties of the submesoscale instability, but misses others. The solution is the same as (23), with Λ_{t} replaced by Λ_{m} and μ_{t} replaced by $\mu_{\mathrm{m}} = N_{\mathrm{m}} k_{\mathrm{h}} h / f$. The peak growth rate $\sigma = 0.31 f \Lambda_{\mathrm{m}} / N_{\mathrm{m}}$ occurs at $\mu_{\mathrm{m}} = 1.6$, which corresponds to a zonal wavelength $\lambda = 3.9 N_{\mathrm{m}} h / f$. This is a reasonable

approximation of the maximum submesoscale growth rate of the full model (Fig. 4*c*). The mixed layer only model captures the bulk shape of the submesoscale lobe of the full model; its short-wave cutoff at $\lambda = 2.6 N_\mathrm{m} H/f$ is close to the short-wave cutoff of the full model. But the mixed layer only model does not have a long-wave cutoff, unlike the submesoscale instability of the full model. The phase speed structure of the mixed layer only model, with its two phase-locked modes in the unstable range and the split into a surface mode and one propagating on the base of the mixed layer, corresponds to a similar mode structure of the full model (Fig. 5*c*). The spatial structure of the most unstable Eady mode resembles that of the most unstable submesoscale mode of the full model (Fig. 6*d*). It tilts into the shear in the mixed layer, but does not leak into the thermocline.

3.4 No bottom

The submesoscale instability is better approximated if modes are allowed to penetrate into the thermocline. To isolate the submesoscale instability, we consider again the layered model but let the thermocline be infinitely deep. That eliminates bottom edge waves, so no mesoscale instability occurs.

Eady considered the atmospheric analogue to this system to relax the assumption that a rigid lid is placed at the tropopause [10]. The system with no bottom again reduces to two variables; the inversion matrix is

$$L = f k_\mathrm{h} \begin{pmatrix} -\frac{\coth \mu_\mathrm{m}}{N_\mathrm{m}} & \frac{\operatorname{csch} \mu_\mathrm{m}}{N_\mathrm{m}} \\ \frac{\operatorname{csch} \mu_\mathrm{m}}{N_\mathrm{m}} & -\frac{\coth \mu_\mathrm{m}}{N_\mathrm{m}} - \frac{1}{N_\mathrm{t}} \end{pmatrix} \tag{24}$$

and

$$U = \operatorname{diag}\big(0, -\Lambda_\mathrm{m} h\big), \tag{25}$$

$$\Gamma = \operatorname{diag}\big(f^2 \Lambda_\mathrm{m}/N_\mathrm{m}^2, -f^2 \Lambda_\mathrm{m}/N_\mathrm{m}^2 + f^2 \Lambda_\mathrm{t}/N_\mathrm{t}^2\big). \tag{26}$$

In the case $\Lambda = \Lambda_\mathrm{m} = \Lambda_\mathrm{t}$, the solution is [10, 2]

$$c = -\frac{\Lambda h}{2}\left(1 + \frac{\alpha}{\mu_\mathrm{m}}\right) \pm \frac{i \Lambda h}{\mu_\mathrm{m}} \left[\frac{(1-\alpha^2)(\mu_\mathrm{m} - \tanh \mu_\mathrm{m})}{\tanh \mu_\mathrm{m} + \alpha} - \frac{1}{4}(\mu_\mathrm{m} - \alpha)^2\right]^{\frac{1}{2}}, \tag{27}$$

where $\mu_\mathrm{m} = N_\mathrm{m} k_\mathrm{h} h/f$ and $\alpha = N_\mathrm{m}/N_\mathrm{t}$. This solution converges to the mixed layer only solution if $\alpha \ll 1$ and $\alpha \ll \mu_\mathrm{m}$, which is equivalent to $N_\mathrm{t} \gg N_\mathrm{m}$ and $k_\mathrm{h} \gg f/N_\mathrm{t} h$. This shows that large thermocline stratification can act like a rigid bottom, but only for scales that are not too large. Modes of large horizontal scale still manage to penetrate into the thermocline, by which the large-scale dynamics are altered.

The growth rates and phase speeds of this reduced model almost perfectly match the growth rates and phase speeds of the full model at scales smaller than about 100 km (Fig. 4*d* and 5*d*). This model now captures the long-wave cutoff of the submesoscale instability. At large scales, where $\mu_\mathrm{m} \ll \alpha$ and $\mu_\mathrm{m} \ll 1$ or equivalently $k_\mathrm{h} \ll f/N_\mathrm{t} h$ and $k_\mathrm{h} \ll f/N_\mathrm{m} h$, the dynamics split into modes that are barotropic and baroclinic in the mixed layer. The barotropic mode behaves like a surface edge wave, with phase speed $-f\Lambda/N_\mathrm{t} k_\mathrm{h}$, which does not sense the mixed layer (Fig. 5*d*). The baroclinic mode is baroclinic in the mixed layer

and remains shallow for large scales—its critical level is the base of the mixed layer, its phase speed is $-\Lambda h$. The vastly different phase speeds of these two modes prevent phase-locking, so no instability occurs at large scales. This stabilization is analogous to that by the β-effect [25, 21, 33]. For the unstable modes, this reduced model with no bottom also captures the deepening of the critical level as the scale gets larger, $-\Lambda h(1 + f/N_\mathrm{t} k_\mathrm{h} h)/2$, which is due to the increasing penetration of the unstable mode into the thermocline.

The longwave cutoff in this constant-shear case depends on the ratio $N_\mathrm{m}/N_\mathrm{t}$. In the more general case $\Lambda_\mathrm{m} \neq \Lambda_\mathrm{t}$, it also depends on the ratio $\Lambda_\mathrm{m}/\Lambda_\mathrm{t}$. No longwave cutoff occurs if $\Lambda_\mathrm{t} = 0$, as found by Rivest et al., who considered the atmospheric case with no shear in the stratosphere [27]. There is also no longwave cutoff as $N_\mathrm{m}/N_\mathrm{t} \to 0$, which is the Eady limit. The instability itself requires a reversal of the PV gradient, so the condition for instability is $\Lambda_\mathrm{m}/N_\mathrm{m}^2 > \Lambda_\mathrm{t}/N_\mathrm{t}^2$. This condition is typically satisfied in the ocean.

3.5 Summary

We are now in a position to understand all branches in the phase speed diagram of the full model. Branch 'a' is a mode that is baroclinic in the mixed layer and does not penetrate much into the thermocline. It does not sense the bottom. Branch 'b' is the unstable branch corresponding to the deep, Eady-like instability. Branch 'c' is a mode that is nearly barotropic in the mixed layer and behaves like a surface edge wave in the thermocline. It does not interact much with the bottom. Branch 'd' is a bottom edge wave that is independent of the surface and interface. Branch 'e' is the unstable branch corresponding to the mixed layer instability. The instability is significantly modified by the modes' penetration into the thermocline, but the scale and growth rate of the most unstable mode still scale with the mixed layer deformation radius and the Eady growth rate. Branches 'f' and 'g' are edge waves propagating on the surface and the interface that do not interact with any of the other edge waves.

4 Nonlinear dynamics

We now turn to the nonlinear dynamics that arise when the perturbations grow to finite amplitude. Before considering the combined effect of the deep mesoscale and mixed layer instabilities, we first consider them separately. We start with the thermocline only case, in which only the deep instability is present. We subsequently contrast that case with the case with no bottom, in which only the submesoscale instability is present. We finally consider the full model, in which both instabilities occur.

We solve the full nonlinear perturbation equations

$$\frac{\partial \boldsymbol{\theta}'}{\partial t} + U\frac{\partial \boldsymbol{\theta}'}{\partial x} + \Gamma\frac{\partial \boldsymbol{\psi}'}{\partial x} + \mathrm{J}(\boldsymbol{\psi}', \boldsymbol{\theta}') = -r\nabla^{-2}\boldsymbol{\theta}' - \nu\nabla^{2n}\boldsymbol{\theta}', \tag{28}$$

where the Jacobian operator is understood to act element-wise. We introduce hypoviscosity with coefficient r, which provides a large-scale drag, and hyperviscosity with coefficient ν and order n, which helps ensure numerical stability and absorbs enstrophy that is cascaded to small scales.

We solve these equations in a 500 km × 500 km domain that is doubly periodic in the perturbation quantities. A fully dealiased pseudo-spectral code with a resolution 512×512 is used. The time derivatives are discretized using a forth-order Runge-Kutta scheme. The hypoviscosity coefficient is $r = 10^{-16}\ \mathrm{m^{-2}\,s^{-1}}$; the hyperviscosity is of order $n = 10$ and the coefficient is $\nu = 2.5 \times 10^{46}\ \mathrm{m^{20}\,s^{-1}}$. All calculations are started with a white noise of small amplitude in $\boldsymbol{\theta}'$.

4.1 Thermocline only

We start with the familiar Eady model representing the thermocline. Since the dissipative terms are weak in the linear equations, the instability grows until it reaches finite amplitude, when the nonlinear terms become important. Secondary instabilities set in and the flow quickly evolves into a fully turbulent regime. The perturbations grow in scale until they reach a scale where hypoviscosity is strong enough to damp the flow significantly. Thereby, the flow comes into statistical equilibrium, which is the time period considered in what follows.

A snapshot from the equilibrated state exhibits a patchy surface buoyancy field (Fig. 7*a*). There are strong buoyancy gradients. The strongest vortices visible are those at a scale of about 200 km. Smaller-scale vortices are present, but successively weaker. They result from a roll-up instability that features prominently in the evolution of the flow [12]. The submesoscale dynamics of this setup are decoupled surface QG dynamics at the surface and bottom that are stirred by the meosscale thermocline instability [28].

As typical for turbulent flows, a continuum of scales is energized. This is quantified by the kinetic and potential energy spectra,

$$K_{k,l} = \frac{1}{2}\left(|u_{k,l}|^2 + |v_{k,l}|^2\right), \qquad P_{k,l} = \frac{1}{2}\frac{|b_{k,l}|^2}{N^2}, \tag{29}$$

which we average azimuthally in wavenumber space, because the statistics are very nearly isotropic. We also average in time to characterize the statistics of the equilibrated state. The surface and bottom spectra of both kinetic and potential energy peak at a wavelength of about 200 km and fall off roughly like $k_\mathrm{h}^{-5/3}$ (Fig. 8), as predicted by surface QG turbulence theory [1]. Since smaller-scale modes decay more rapidly in the vertical than larger-scale modes, the spectra are steeper in the interior. At 100 m depth, the mesoscale energy levels are similar to those at the surface, but submesoscales energy levels are much lower.

A useful diagnostic of turbulent dynamics is the spectral energy budget. While the dynamics are completely determined by the advection of conserved quantities at the surface and bottom, we first consider the energy budget over the entire depth range. We will present a lower-order energy diagnostic for the no bottom case below.

The equations for the spectral perturbation potential and kinetic energies are

$$\frac{\partial P_{k,l}}{\partial t} = \mathrm{Re}\left[\frac{f\Lambda}{N^2} v^*_{k,l} b_{k,l} - w^*_{k,l} b_{k,l} - \frac{1}{N^2} b^*_{k,l}\, \mathrm{J}_{k,l}(\psi', b')\right] - \left(r k_\mathrm{h}^{-2} + \nu k_\mathrm{h}^{2n}\right) P_{k,l} \tag{30}$$

$$\frac{\partial K_{k,l}}{\partial t} = \mathrm{Re}\left[-f\frac{\partial}{\partial z}\left(w^*_{k,l}\psi_{k,l}\right) + w^*_{k,l} b_{k,l} + \psi^*_{k,l}\, \mathrm{J}_{k,l}(\psi', \nabla^2\psi')\right] - \left(r k_\mathrm{h}^{-2} + \nu k_\mathrm{h}^{2n}\right) K_{k,l} \tag{31}$$

where P is potential energy, K is kinetic energy, the asterisks denote complex conjugates, and Re denotes the real part. The first term on the right-hand side of the potential energy

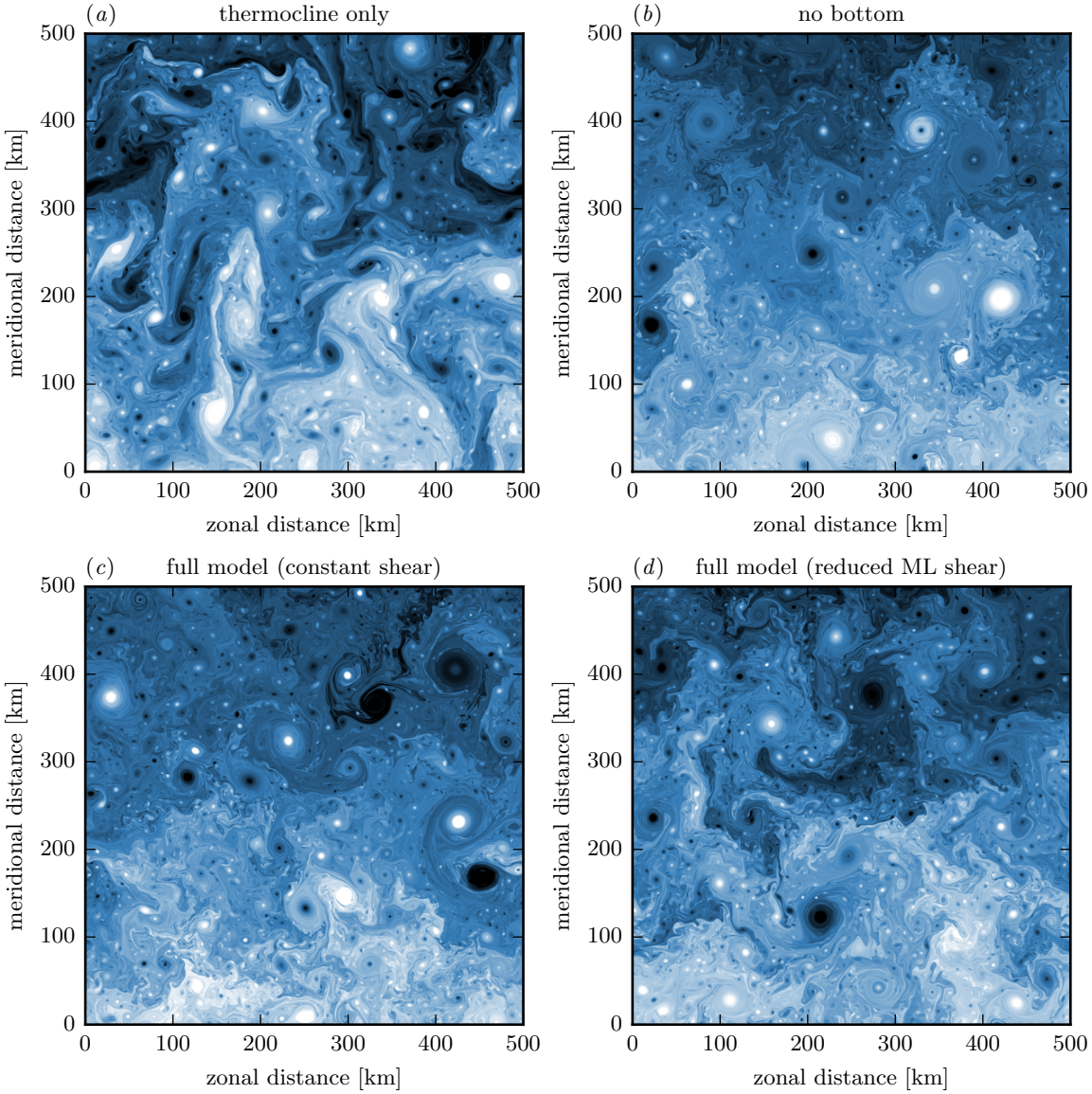

Figure 7: Snapshots of surface buoyancy (mean plus anomalies) from the equilibrated states of the (a) thermocline only simulations, (b) the no bottom simulation, (c) the full model simulation with equal mixed layer and thermocline shears, and (d) the full model with reduced mixed layer shear. The color scale extends from white (more buoyant) through blue to black (less buoyant) and extends between $\pm f \Lambda_{\mathrm{m}} L$.

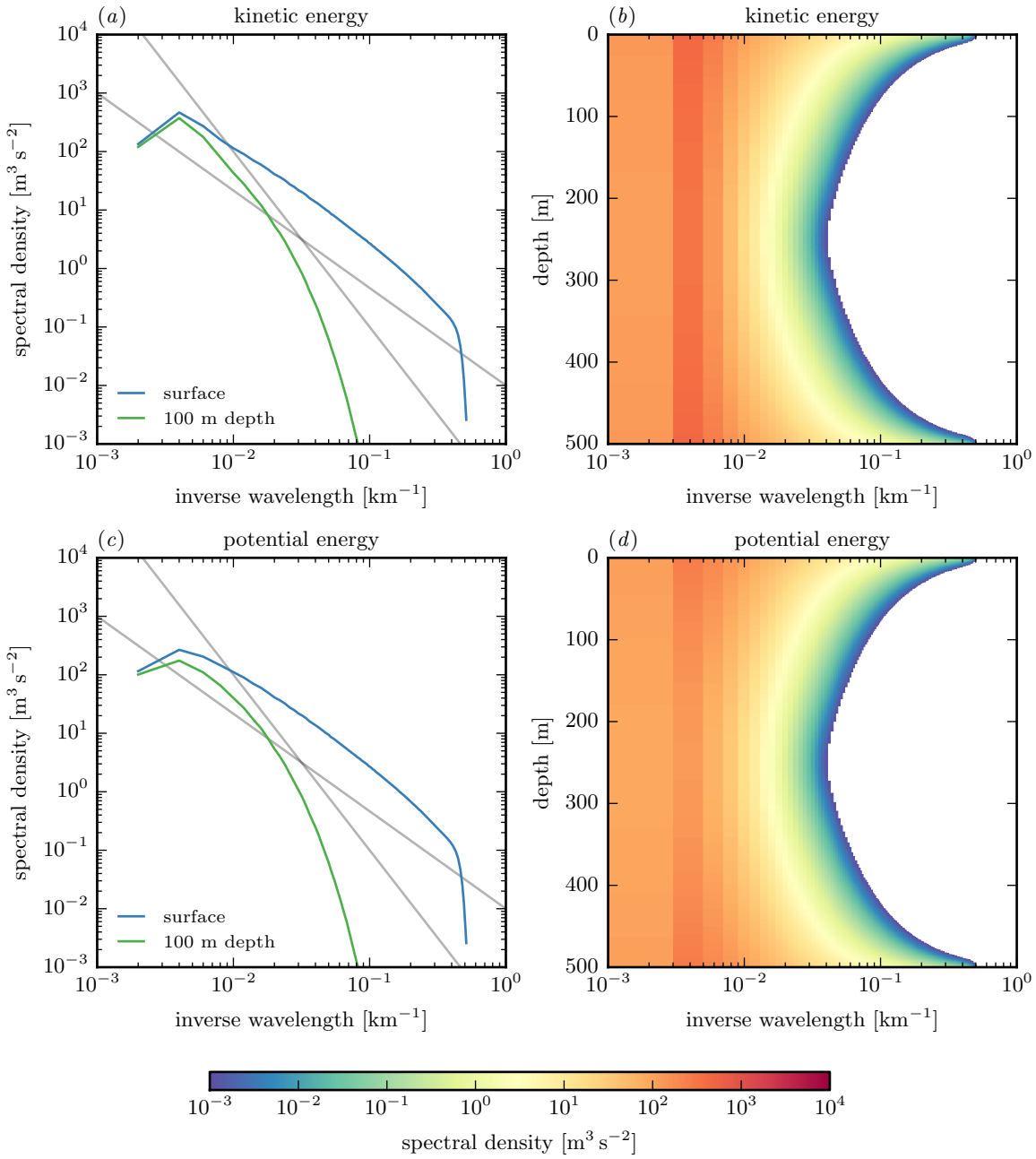

Figure 8: Wavenumber spectra of kinetic and potential energy from the thermocline only simulation. (*a*) Kinetic energy spectra at the surface and 100 m depth, (*b*) spectral density of kinetic energy in the wavenumber–depth plane, (*c*) potential energy spectra at the surface and 100 m depth, (*d*) spectral density of potential energy in the wavenumber–depth plane. In panels (*b*) and (*d*), no values below 10^{-3} m^3 s^{-2} are shown.

equation represents the extraction of potential energy from the mean flow. The second term represents the conversion from potential to kinetic energy. This term appears as a source term in the kinetic energy budget. The third term in the potential energy budget represents spectral transfer by triadic interactions. The sum of this term over all wavenumbers vanishes. An equivalent spectral transfer term appears in the kinetic energy budget (third term). Kinetic energy can also be distributed vertically by pressure fluxes, represented by the first term in the kinetic energy budget. The vertical integral of this term vanishes. The viscosity terms act as sinks for both potential and kinetic energy—hypoviscosity acting at large scales, hyperviscosity at small scales. We present these budgets again averaged azimuthally in wavenumber space and over time.

The extraction of potential energy from the mean is dominated by the largest, most energetic eddies (Fig. 9*a*). The extraction is independent of depth, because $q' = 0$ and therefore

$$0 = \mathrm{Re}\, v^*_{k,l} q_{k,l} = \mathrm{Re}\, \frac{\partial}{\partial z} \left(\frac{f}{N^2} v^*_{k,l} b_{k,l} \right). \tag{32}$$

Potential energy is transferred downscale by triadic interactions and deposited near the deformation radius and in wedges near the surface and the bottom that reach to much smaller scales (Fig. 9*b*). Where potential energy is deposited by scale interactions, it is converted into kinetic energy (Fig. 9*c*). Near the deformation radius, this conversion is due to the mesoscale instability that produces vertical buoyancy fluxes. In the wedges near the surface and bottom, the conversion is due to secondary instabilities present in the surface QG cascades, which occur independently at the surface and the bottom [28]. The kinetic energy thus created is transferred back to large scales (Fig. 9*d*). The bulk of the energy is dissipated through hypoviscosity at the scales of the largest, most energetic eddies (Fig. 9*e*). The energy dissipation through hyperviscosity is small, which reflects the fundamental property of geostrophic turbulence that energy is trapped at large scales and viscous energy dissipation vanishes as the viscosity goes to zero [18, 9].

4.2 No bottom

Now consider the case with no bottom, which has a mixed layer instability only. This instability, too, grows to finite amplitude and the flow becomes turbulent. There is a turbulent spin-up phase, in which the eddies that are initially of the size of the instability grow larger until they reach a statistical equilibrium with hypoviscosity.

A snapshot of surface buoyancy reveals that the flow's structure is quite different from the thermocline only case (Fig. 7*b*). There are many more coherent vortices. They are prominent at a scale of about 100 km, but many smaller-scale coherent vortices exist. These do not appear to be present in observations of sea surface temperature or realistic regional ocean models, a point we will come back to in the discussion. If we focus on the filamentary sea, however, there are again strong fronts, superimposed by submesoscale structure, which is realistic.

The energy spectra reflect this nearly frontal structure at the surface (Fig. 10). The kinetic energy spectra fall off slightly more steeply than $k^{-5/3}$ at the scales of the linear instability and like $k^{-5/3}$ at scales smaller than the linear short-wave cutoff, both at the surface and at the base of the mixed layer at 100 m depth (Fig. 10*a*). The mixed layer

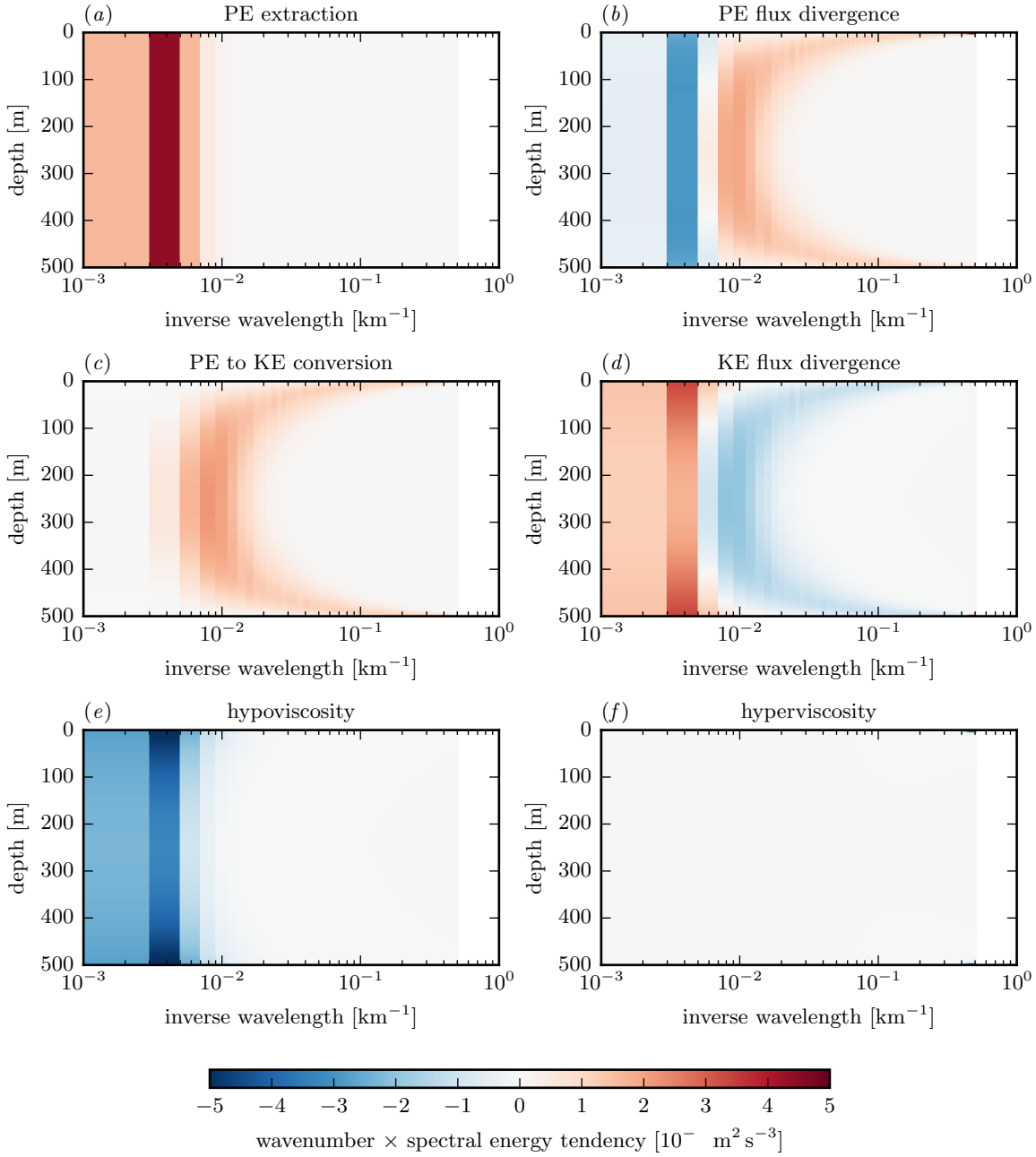

Figure 9: Spectral energy budget for the thermocline only simulation. The terms are (*a*) potential energy extraction from the mean, (*b*) spectral potential energy flux divergence, (*c*) potential to kinetic energy conversion, (*d*) kinetic energy flux divergence, including spectral flux and pressure flux, (*e*) hypoviscosity on both kinetic and potential energy, and (*f*) hyperviscosity on both kinetic and potential energy. All terms are multiplied by the wavenumber to compensate for logarithmic shrinking.

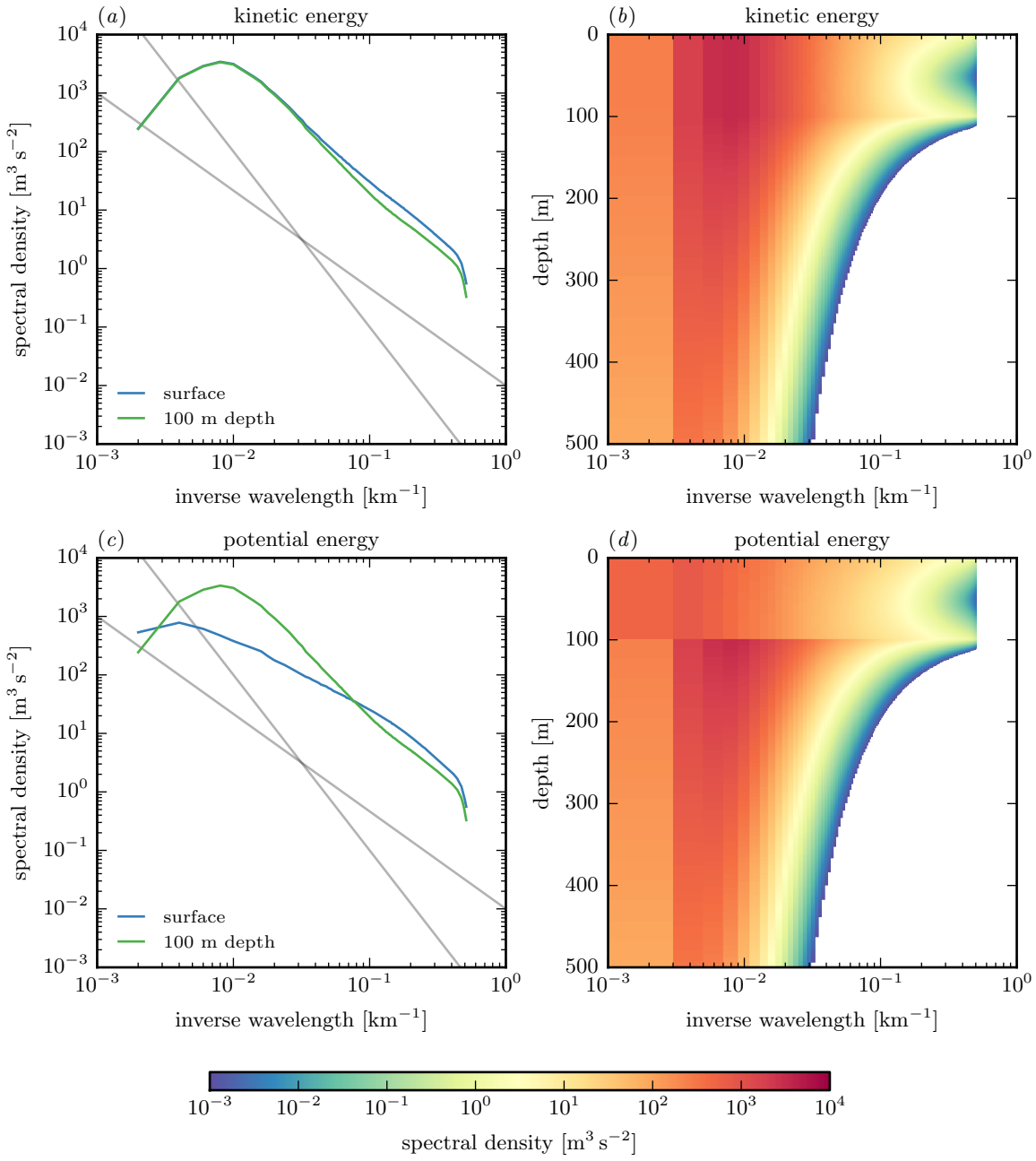

Figure 10: Wavenumber spectra of kinetic and potential energy from the no bottom simulation. (*a*) Kinetic energy spectra at the surface and 100 m depth, (*b*) spectral density of kinetic energy in the wavenumber–depth plane, (*c*) potential energy spectra at the surface and 100 m depth, (*d*) spectral density of potential energy in the wavenumber–depth plane. In panels (*b*) and (*d*), no values below 10^{-3} $\mathrm{m^3\,s^{-2}}$ are shown.

instabilities energize the entire depth of the mixed layer. This is in sharp contrast to the thermocline only simulation, in which the surface QG dynamics energize a thin wedge close to the surface only.

It should be noted that the equilibrated flow in the no bottom case is much more energetic than in the thermocline only case and more energetic than is realistic. While the equilibration by hypoviscosity is unrealistic, we will see that the enhanced energy levels are due to more efficient extraction of mean potential energy in the weakly stratified mixed layer, which is a dynamical property of the system that does not depend on how the flow is equilibrated. We will discuss reasons for this excess in energy in the conclusions.

Below the base of the mixed layer, The potential energy spectra are the same as the kinetic energy spectra (Fig. 10c). In the mixed layer, the potential energy spectra are significantly flatter than the kinetic energy spectra. This is in contrast to observations that show rough equipartition between kinetic and potential energy [6, 7].

The vertical structure of the energy shows that the mixed layer instabilities also energize the thermocline below (Fig. 10b,d). At the instability scale, the flow does not reach much into the thermocline. But as the horizontal scale of the flow increases, so does the vertical scale. The flow exhibits the familiar property of geostrophic turbulence that it barotropizes as it increases its horizontal scale [31].

The energy transfer into the thermocline can further be examined in the spectral energy budget (Fig. 11). Potential energy is again extracted at the scale of the largest, most energetic eddies, but the extraction is now confined to the mixed layer (Fig. 11a). Potential energy is transferred from the extraction scale to the scale of the mixed layer instability (Fig. 11b). The mixed layer instability converts potential energy into kinetic energy in the mixed layer, at the instability scale (Fig. 11c). The kinetic energy created by the instability again enters an inverse cascade, but now it is not only transferred to large horizontal scales, but also vertically into the thermocline (Fig. 11d). The deposition of kinetic energy at the scale of the largest eddies is well distributed across the mixed layer and upper thermocline. The vertical distribution of damping by hypoviscsosity also reflects the fact that the flow is not confined to the mixed layer at the scale of the largest eddies, where hypoviscosity acts (Fig. 11e). Hyperviscosity acts only at the smallest resolved scales (Fig. 11f). While small, it does affect the other terms in the budget. We neglect this effect, because it would tend to zero if the resolution was increased and the hyperviscosity coefficient decreased.

These energy pathways are reminiscent of the phenomenology of two-layer baroclinic turbulence. The turbulent dynamics of a two-layer system can be understood in terms of a dual cascade [26, 29]. Baroclinic energy is extracted from the mean at the scale of the largest, most energetic eddies. The barotropic flow dominates at these scales and transfers the baroclinic energy downscale. The baroclinic mode behaves like a passive tracer at these scales. Around the deformation radius, the instability converts baroclinic energy into barotropic energy. The barotropic energy then enters an inverse cascade, which gets halted at some scale by drag.

Can the turbulent dynamics induced by mixed layer instabilities be understood in similar terms? To phrase the analysis in these terms, we need a concept of modes, which the energy can be partitioned into. In our system, the vertically integrated energy can be written

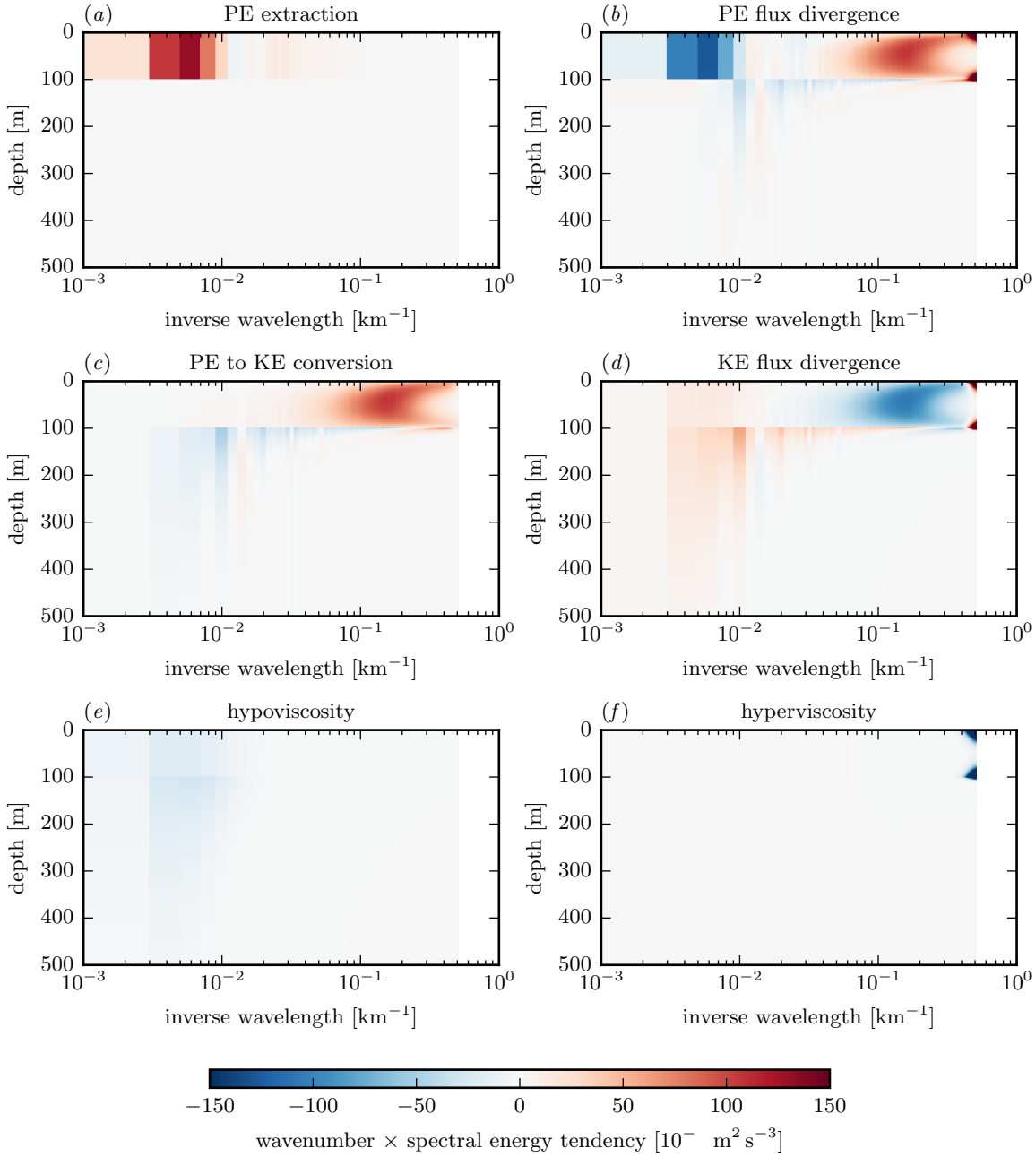

Figure 11: Spectral energy budget for the no bottom simulation. The terms are (a) potential energy extraction from the mean, (b) spectral potential energy flux divergence, (c) potential to kinetic energy conversion, (d) kinetic energy flux divergence, including spectral flux and pressure flux, (e) hypoviscosity on both kinetic and potential energy, and (f) hyperviscosity on both kinetic and potential energy. All terms are multiplied by the wavenumber to compensate for logarithmic shrinking.

entirely in terms of the quantities at the interfaces:

$$E = -\frac{1}{2}\boldsymbol{\psi}^\dagger \boldsymbol{\theta} = -\frac{1}{2}\boldsymbol{\psi}^\dagger L \boldsymbol{\psi}, \tag{33}$$

where we dropped the k, l subscripts to denote Fourier transforms. Since L is real and symmetric, it is diagonalizable by a unitary matrix,

$$L = S^\dagger D S, \tag{34}$$

where D is diagonal and consists of the real eigenvalues of L, $D_{ii} = \lambda_i$. The energy can now be written as

$$E = -\frac{1}{2}(S\boldsymbol{\psi})^\dagger D (S\boldsymbol{\psi}) = -\frac{1}{2}\sum_i \lambda_i \left|(S\boldsymbol{\psi})_i\right|^2. \tag{35}$$

This defines the modes $(S\boldsymbol{\psi})_i$ that are orthogonal with respect to the energy norm, i.e. the energy can be partitioned into contributions by these modes. The structure of the modes depends on wavenumber, because L and therefore S does.

For the no bottom case, with L given by (24), the eigenvalues of L are

$$\lambda_{0,1} = f k_\mathrm{h} \left(\frac{\coth \mu_\mathrm{m}}{N_\mathrm{m}} + \frac{1}{2N_\mathrm{t}} \pm \sqrt{\frac{\operatorname{csch}^2 \mu_\mathrm{m}}{N_\mathrm{m}^2} + \frac{1}{4N_\mathrm{t}^2}} \right) \tag{36}$$

and S, of which the columns constitute the eigenvectors, is

$$S = \begin{pmatrix} \frac{1}{\sqrt{1+\left(\cosh\mu_\mathrm{m} + \frac{N_\mathrm{m}\lambda_0}{f k_\mathrm{h}} \sinh\mu_\mathrm{m}\right)^2}} & \frac{1}{\sqrt{1+\left(\cosh\mu_\mathrm{m} + \frac{N_\mathrm{m}\lambda_1}{f k_\mathrm{h}} \sinh\mu_\mathrm{m}\right)^2}} \\ \frac{\cosh\mu_\mathrm{m} + \frac{N_\mathrm{m}\lambda_0}{f k_\mathrm{h}}}{\sqrt{1+\left(\cosh\mu_\mathrm{m} + \frac{N_\mathrm{m}\lambda_0}{f k_\mathrm{h}} \sinh\mu_\mathrm{m}\right)^2}} & \frac{\cosh\mu_\mathrm{m} + \frac{N_\mathrm{m}\lambda_1}{f k_\mathrm{h}}}{\sqrt{1+\left(\cosh\mu_\mathrm{m} + \frac{N_\mathrm{m}\lambda_1}{f k_\mathrm{h}} \sinh\mu_\mathrm{m}\right)^2}} \end{pmatrix}. \tag{37}$$

For large scales, $k_\mathrm{h} \ll f/N_\mathrm{t} h$ or $\mu_\mathrm{m} \ll N_\mathrm{m}/N_\mathrm{t}$, this reduces to

$$\lambda_0 = -\frac{f k_\mathrm{h}}{2N_\mathrm{t}}, \qquad \lambda_1 = -\frac{2f^2}{N_\mathrm{m}^2 h}. \tag{38}$$

and simply

$$S = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}. \tag{39}$$

This indicates that at large scales the first mode is barotropic in the mixed layer. It behaves like a surface QG mode penetrating into the thermocline. The streamfunction is proportional to k_h times the conserved quantity [12],

$$(S\boldsymbol{\psi})_0 = -\frac{f k_\mathrm{h}}{2N_\mathrm{t}} (S\boldsymbol{\theta})_0. \tag{40}$$

The second mode at large scales is baroclinic in the mixed layer. The relation between the streamfunction and the conserved quantity is

$$(S\boldsymbol{\psi})_1 = -\frac{2f^2}{N_\mathrm{m}^2 h} (S\boldsymbol{\theta})_1, \tag{41}$$

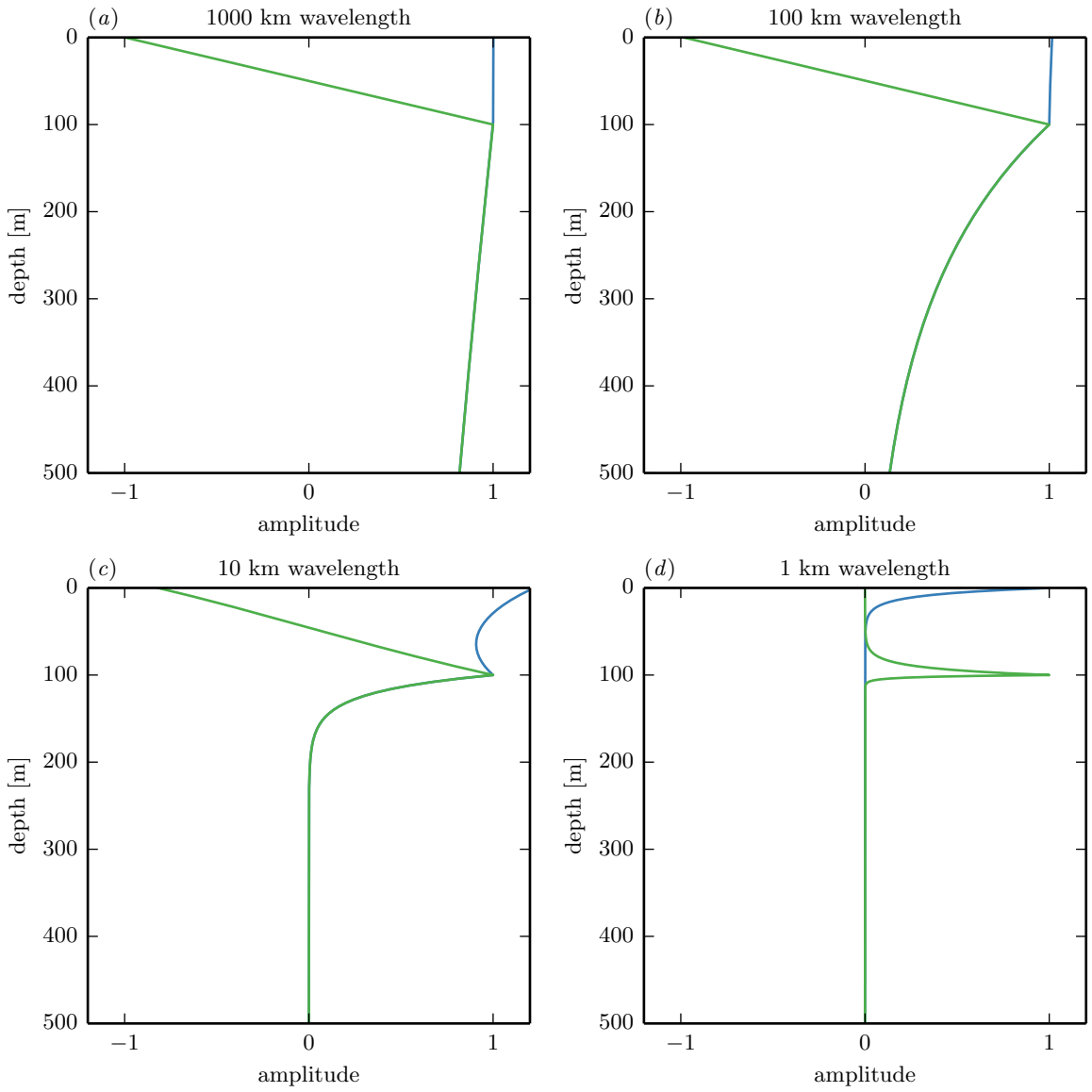

Figure 12: Vertical structure of the streamfunction corresponding to orthogonal modes in the no bottom case for different wavenumbers $k_\mathrm{h} = 2\pi/\lambda$, with the wavelength λ given in the panel titles. For panels (a), (b), and (c), the modes are normalized to unity at the interface at 100 m depth; for panel (d), the modes are normalized to have a maximum value of unity. Mode 0 is shown in blue, mode 1 in green. In panels (a), (b), and (c), the two modes coincide below the interface.

which is independent of k_h, as expected for a baroclinic mode.

This description of the modes as barotropic and baroclinic mixed layer modes only applies to large scales. At smaller scales, the modes have more complicated vertical structure (Fig. 12). At scales smaller than the mixed layer deformation radius, the modes morph into decoupled modes that are localized in the vertical at the surface and the interface. But for the cascade dynamics to be discussed, the mode structure at large scales is what is most important.

We can now consider the energy budget of these modes. We start from the vertically integrated spectral energy budget, written in terms of the conserved quantities and corresponding streamfunctions:

$$\frac{\partial E}{\partial t} = -\boldsymbol{\psi}^\dagger \frac{\partial \boldsymbol{\theta}}{\partial t}. \tag{42}$$

Again, the subscripts k, l are dropped. Using the diagonalization, this can be written as

$$\frac{\partial E}{\partial t} = -\boldsymbol{\psi}^\dagger L \frac{\partial \boldsymbol{\psi}}{\partial t} = -(S\boldsymbol{\psi})^\dagger D \frac{\partial}{\partial t}(S\boldsymbol{\psi}) = -\sum_i \lambda_i (S\boldsymbol{\psi})_i^* \frac{\partial}{\partial t}(S\boldsymbol{\psi})_i. \tag{43}$$

The budget therefore splits into

$$\frac{\partial E_i}{\partial t} = -\lambda_i (S\boldsymbol{\psi})_i^* \frac{\partial}{\partial t}(S\boldsymbol{\psi})_i. \tag{44}$$

The terms on the right-hand side of this budget can be obtained from

$$-\boldsymbol{\psi}^\dagger \frac{\partial \boldsymbol{\theta}}{\partial t} = -(S\boldsymbol{\psi})^\dagger S \frac{\partial \boldsymbol{\theta}}{\partial t}, \tag{45}$$

into which the spectral form of the evolution equation (28) is substituted. We further expand the nonlinear terms in (28) into

$$\begin{aligned} \mathrm{J}(\boldsymbol{\psi}, \boldsymbol{\theta}) &= \mathrm{J}(S^\dagger(S\boldsymbol{\psi}), S^\dagger(S\boldsymbol{\theta})) && (46) \\ &= \mathrm{J}(S^\dagger P_0(S\boldsymbol{\psi}), S^\dagger P_0(S\boldsymbol{\theta})) + \mathrm{J}(S^\dagger P_0(S\boldsymbol{\psi}), S^\dagger P_1(S\boldsymbol{\theta})) \\ &\quad + \mathrm{J}(S^\dagger P_1(S\boldsymbol{\psi}), S^\dagger P_0(S\boldsymbol{\theta})) + \mathrm{J}(S^\dagger P_1(S\boldsymbol{\psi}), S^\dagger P_1(S\boldsymbol{\theta})), && (47) \end{aligned}$$

where P_0 and P_1 are the projections onto the respective modes,

$$P_0 = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \qquad P_1 = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}. \tag{48}$$

This allows us to separate out the nonlinear interactions of the modes with themselves and with each other. The first term in (47), for example, represents the advection of the barotropic mode by the barotropic mode, to use the naming convention introduced above. The second term represents the advection of the baroclinic mode by the barotropic mode, and so on.

In terms of the orthogonal modes, the energy budget is very similar to that of a baroclinic two-layer system [20]. The extraction of potential energy from the mean flow is concentrated at the scale of the largest, most energetic eddies and creates mostly baroclinic energy (Fig. 13*b*). That input of baroclinic energy is compensated by a spectral transfer of

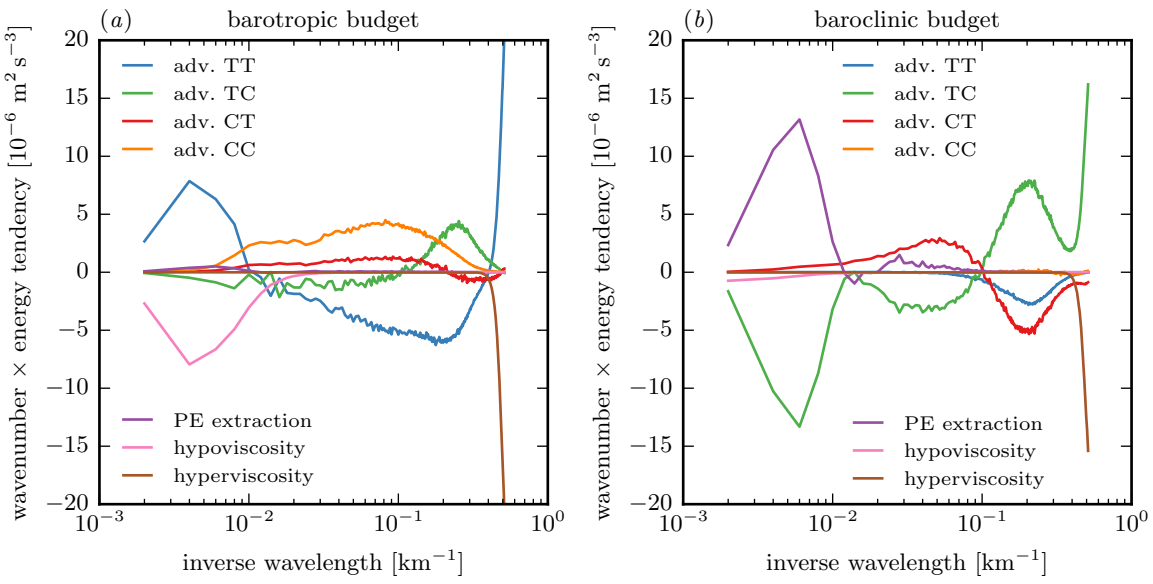

Figure 13: Modal energy budget for the no bottom case. The advective terms correspond to the contributions from the four terms in (47). The energy tendencies are multiplied by wavenumber to compensate for logarithmic shrinking.

baroclinic energy down to the instability scale, achieved by the advection of the baroclinic mode by the barotropic mode (Fig. 13*b*). The energy deposited around the instability scale is transferred to the barotropic mode by interactions between the two modes that amount to baroclinic instability (Fig. 13*b*). This energy enters the barotropic budget rather less localized in wavenumber space (Fig. 13*a*). This forcing of the barotropic mode is compensated by the upscale spectral transfer of the barotropic mode (Fig. 13*a*). The deposition of energy by the inverse cascade at the scale of the largest, most energetic eddies is compensated by hypodiffusion (Fig. 13*a*). Energy loss by hyperdiffusion again enters the budget, but is neglected because it is an artifact of finite resolution.

This model thus exhibits a dual cascade just like the classic two-layer system. Baroclinic energy is transferred downscale through advection by the barotropic mode, baroclinic instability converts baroclinic into barotropic energy, and barotropic energy is transferred back upscale in an inverse cascade. The difference is that the barotropic mode at large scales here behaves like a surface QG mode, instead of a truly barotropic or two-dimensional mode. The inverse cascade is therefore expected to yield a k_h^{-1} surface energy spectrum, which we find to emerge if the inertial range is wide enough (not shown). More importantly, the surface-QG-like behavior implies that in the inverse cascade, energy is transferred to successively larger vertical scales. This provides a pathway for mixed layer instabilities to energize the thermocline below.

4.3 Full model

We now consider the full model, which supports both deep and mixed layer instabilities. We analyze two cases, one with the same shear in mixed layer and thermocline—the setup

discussed in the linear stability analysis section—and one with reduced shear in the mixed layer.

In the constant shear case, the growth rate of the mixed layer instability is much larger than that of the deep instability. The mixed layer instability therefore grows to finite amplitude first and the evolution in the mixed layer is very similar to that of the no bottom case. Again, the eddies grow in size until they come into statistical equilibrium with hypoviscosity.

The equilibrated state of this case is very similar to that of the no bottom case, except near the bottom (Fig. 14). The energy levels and spectra at the surface and the base of the mixed layer are very similar. Near the bottom, a wedge in wavenumber–depth space is energized, just like in the thermocline only case. This is due to the surface QG dynamics at the bottom.

The energy budget is also similar to the case with no bottom (Fig. 15). The main energy pathway is again extraction of potential energy in the mixed layer, transfer to the mixed layer instability scale, conversion to kinetic energy, transfer back to large scales and into the thermocline, and dissipation by hypoviscosity. There is additional extraction in the thermocline, but that is weak compared to the extraction in the mixed layer. The dominant dynamics are therefore those described for the no bottom case. Interaction with the bottom is possible, but of secondary importance in this parameter regime.

A different picture emerges when the mixed layer shear is reduced. We choose the mixed layer shear such that the growth rates of the two instabilities are comparable, which from Eady scaling is expected to occur if $\Lambda_\mathrm{m}/N_\mathrm{m} = \Lambda_\mathrm{t}/N_\mathrm{t}$ is satisfied, so at $\Lambda_\mathrm{m} = 2.5 \times 10^{-5}$ s^{-1}. The results of the linear stability analysis for this mixed layer shear show that indeed the growth rates are comparable (Fig. 16). The horizontal scales of the instabilities and the overall structure of the dispersion curves have not changed. Reducing the mean shear by this amount means that the mean available potential energy is vertically constant.

This system with reduced mixed layer shear equilibrates to realistic and much lower energy levels than the constant shear case (Fig. 17). The energy levels are comparable to the thermocline only case. The vertical structure of energy in this case, however, is still different from the thermocline only case. The mixed layer instability, while not significantly increasing the mesoscale energy levels, does energize the mixed layer at submesoscales.

This can be explained by again considering the energy budget (Fig. 18). The potential energy extraction from the mean is now roughly constant vertically. The energy pathway induced by the mixed layer instability, however, is still active. Potential energy is transferred to the mixed layer deformation radius, where it is converted into kinetic energy and enters an inverse cascade. This energy cycle is stronger than that of surface QG turbulence in the thermocline only case (Fig. 9).

This difference between mixed layer dynamics and surface QG dynamics is also reflected in vertical velocities that are produced by the instabilities (Fig. 19). While the available potential energies are the same and the resulting surface energy levels comparable between this reduced mixed layer shear case and the thermocline only case, there are dramatically larger vertical velocities in the presence of a mixed layer. These enhanced vertical velocities extend significantly below the base of the mixed layer. The largest vertical velocities are located in the vicinity of fronts in the filamentary sea (Fig. 20). Coherent vortices, while associated with the large buoyancy gradients, induce relatively weak vertical motion. The large vertical velocities appear to be induced by the dynamic filamentary structure that

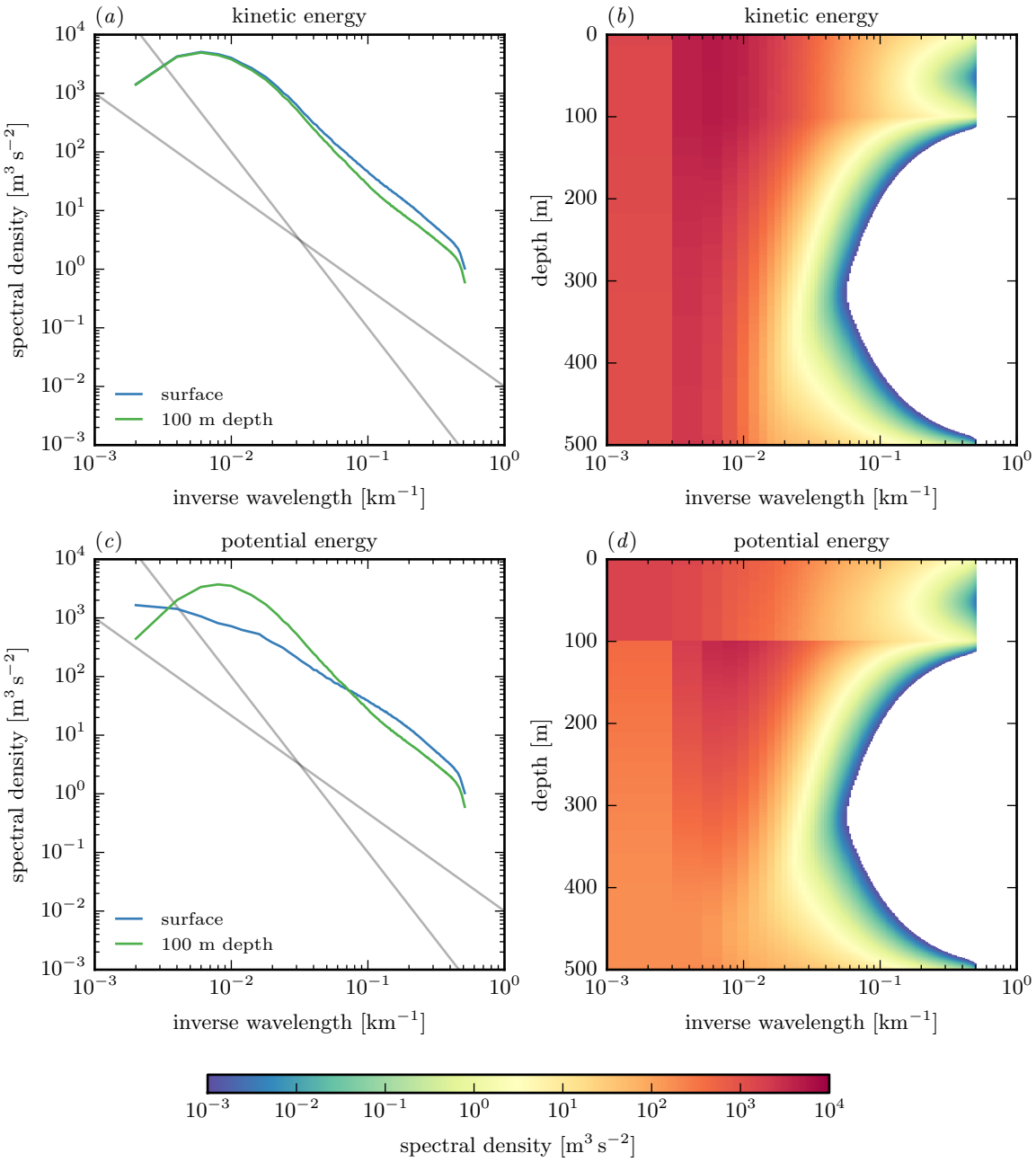

Figure 14: Wavenumber spectra of kinetic and potential energy from the full model simulation. (*a*) Kinetic energy spectra at the surface and 100 m depth, (*b*) spectral density of kinetic energy in the wavenumber–depth plane, (*c*) potential energy spectra at the surface and 100 m depth, (*d*) spectral density of potential energy in the wavenumber–depth plane. In panels (*b*) and (*d*), no values below 10^{-3} m^3 s^{-2} are shown.

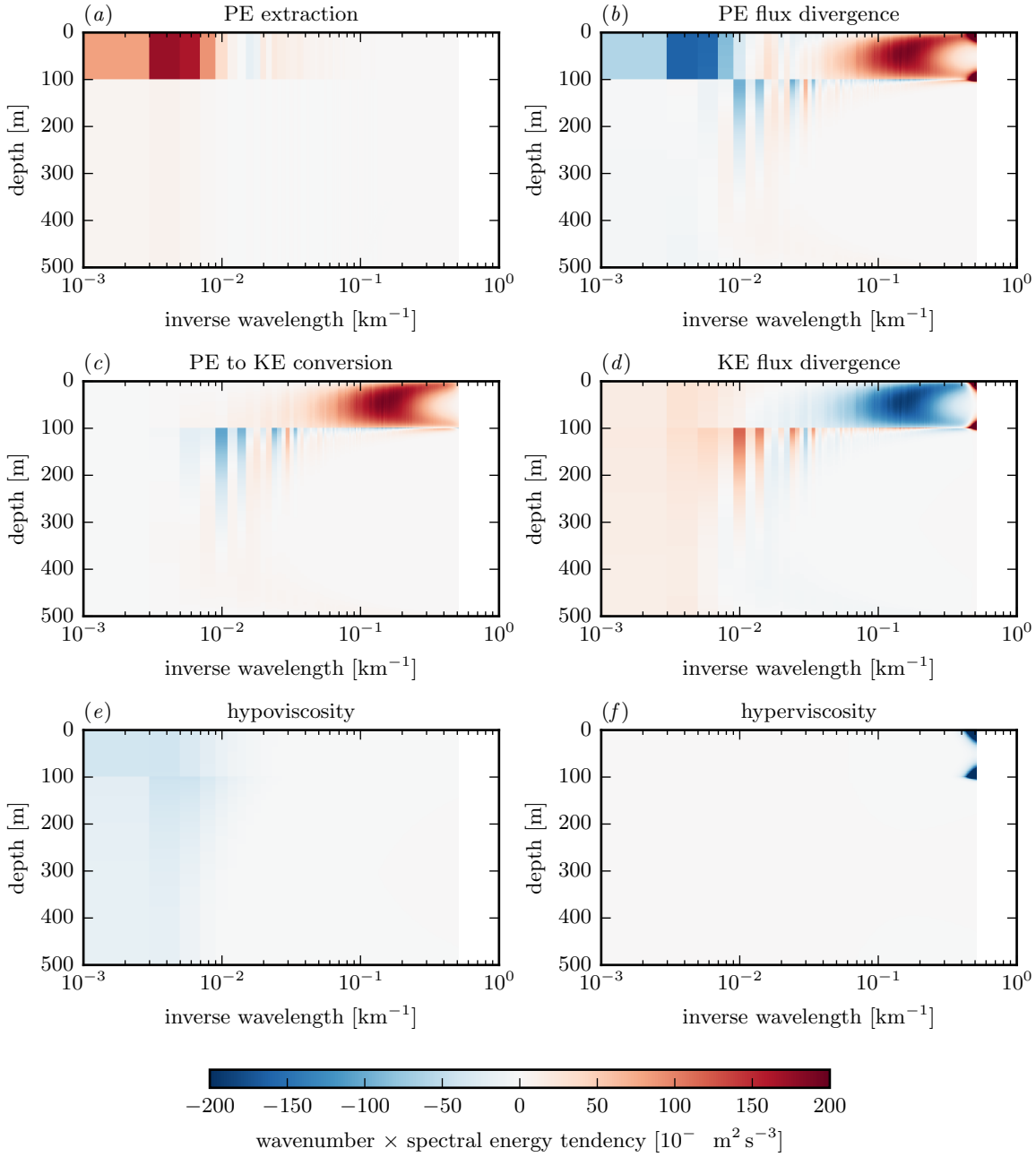

Figure 15: Spectral energy budget for the full model simulation. The terms are (a) potential energy extraction from the mean, (b) spectral potential energy flux divergence, (c) potential to kinetic energy conversion, (d) kinetic energy flux divergence, including spectral flux and pressure flux, (e) hypoviscosity on both kinetic and potential energy, and (f) hyperviscosity on both kinetic and potential energy. All terms are multiplied by the wavenumber to compensate for logarithmic shrinking.

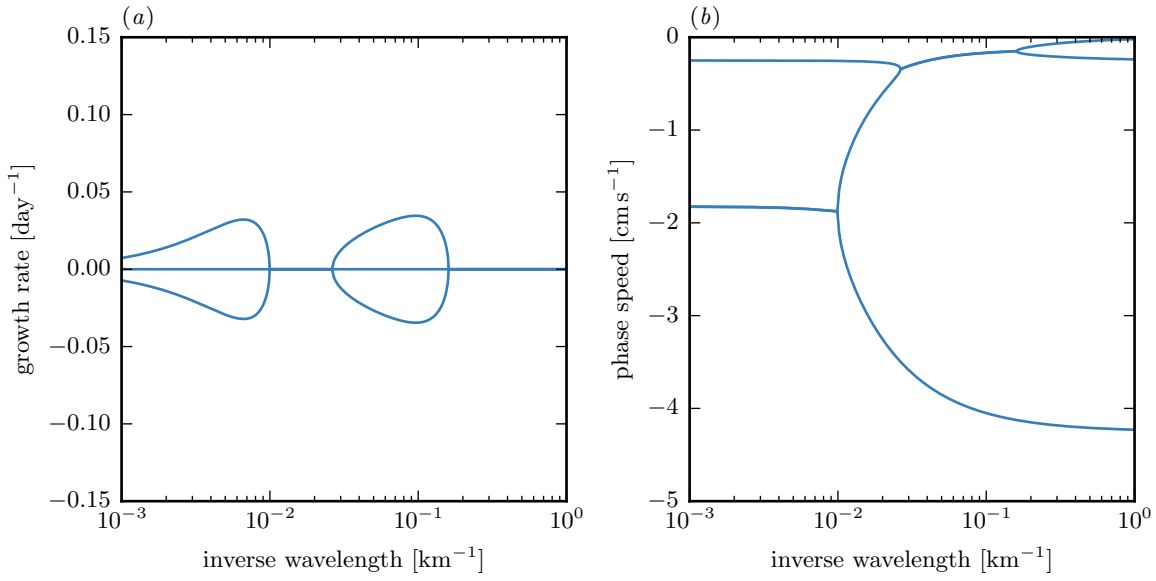

Figure 16: growth rates and phase speeds

undergoes mixed layer instabilities.

The root mean square vertical velocities are similar in structure to those found in primitive equation models [8]. A careful comparison is necessary to establish whether the QG dynamics described here reproduce the magnitude of the vertical velocities or whether non-QG effects significantly enhance vertical velocities. Such a comparison, however, is beyond the scope of the report.

5 Conclusions

Our analysis suggests that the presence of a mixed layer has a profound effect on submesoscale turbulence. The low stratification in the mixed layer, combined with geostrophic shear, provides a large amount of available potential energy that can be extracted through baroclinic instabilities in the mixed layer. The extraction is dominated by mesoscale eddies, but potential energy is subsequently cascaded down to the deformation radius of the mixed layer, where baroclinic instability converts it into kinetic energy. In the QG dynamics considered here, no energy is lost to small scales. The entire energy extracted from the mean in the mixed layer is converted to kinetic energy around the deformation radius of the mixed layer and subsequently transferred back to larger scales in an inverse cascade that also energizes the thermocline below. Through this process, mixed layer instabilities can energize the mesoscale eddy field.

These dynamics are significantly different from surface QG dynamics, which are often invoked to explain energetic submesoscales. Surface QG dynamics can only energize a thin surface layer, whereas mixed layer instabilities energize the entire mixed layer. The vertical and spectral structure found in the presence of a mixed layer is consistent with that found in wintertime observations in the Gulf Stream region, except for the potential energy spectra,

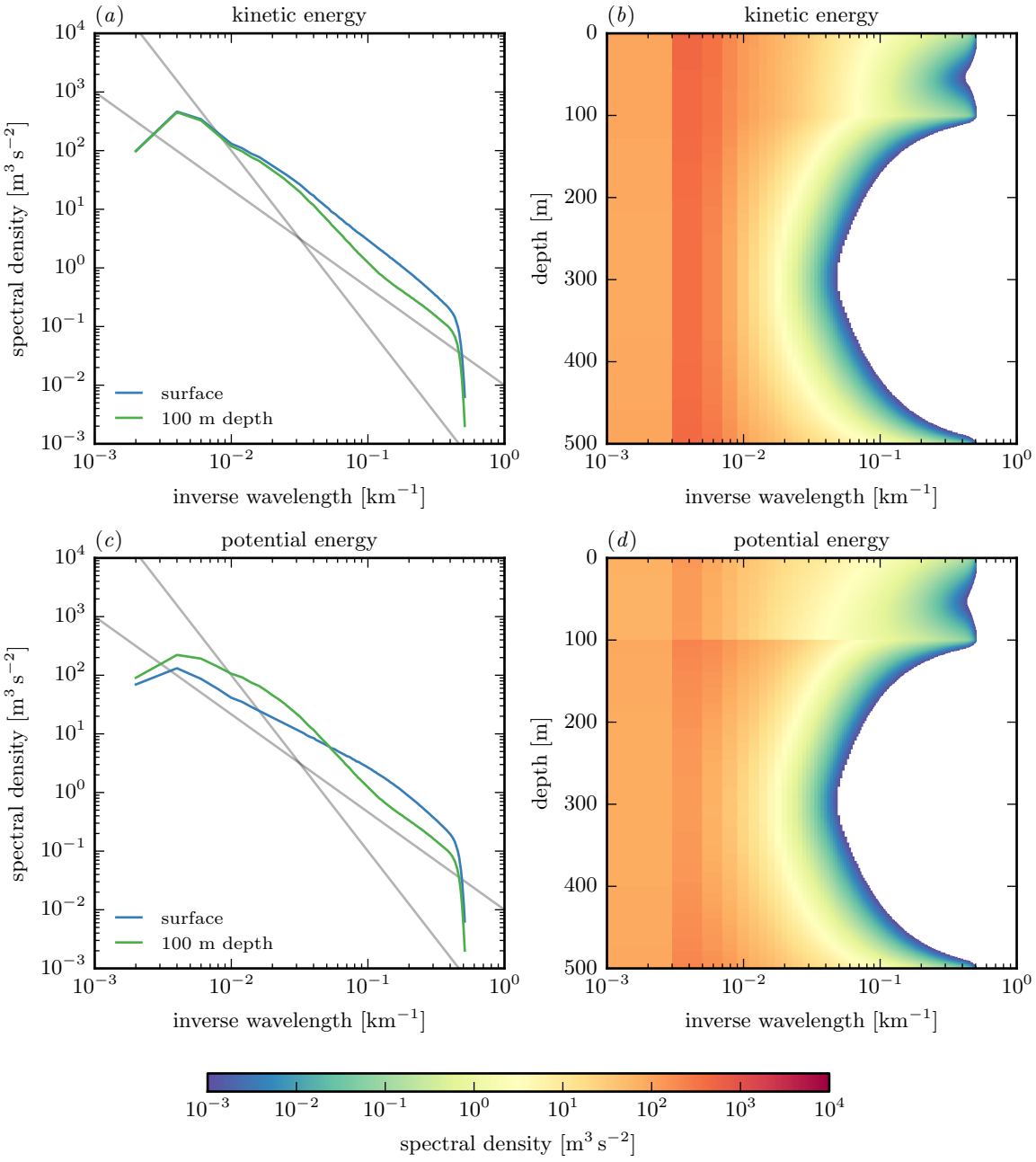

Figure 17: Wavenumber spectra of kinetic and potential energy from the full model simulation with reduced mixed layer shear. (*a*) Kinetic energy spectra at the surface and 100 m depth, (*b*) spectral density of kinetic energy in the wavenumber–depth plane, (*c*) potential energy spectra at the surface and 100 m depth, (*d*) spectral density of potential energy in the wavenumber–depth plane. In panels (*b*) and (*d*), no values below 10^{-3} m^3 s^{-2} are shown.

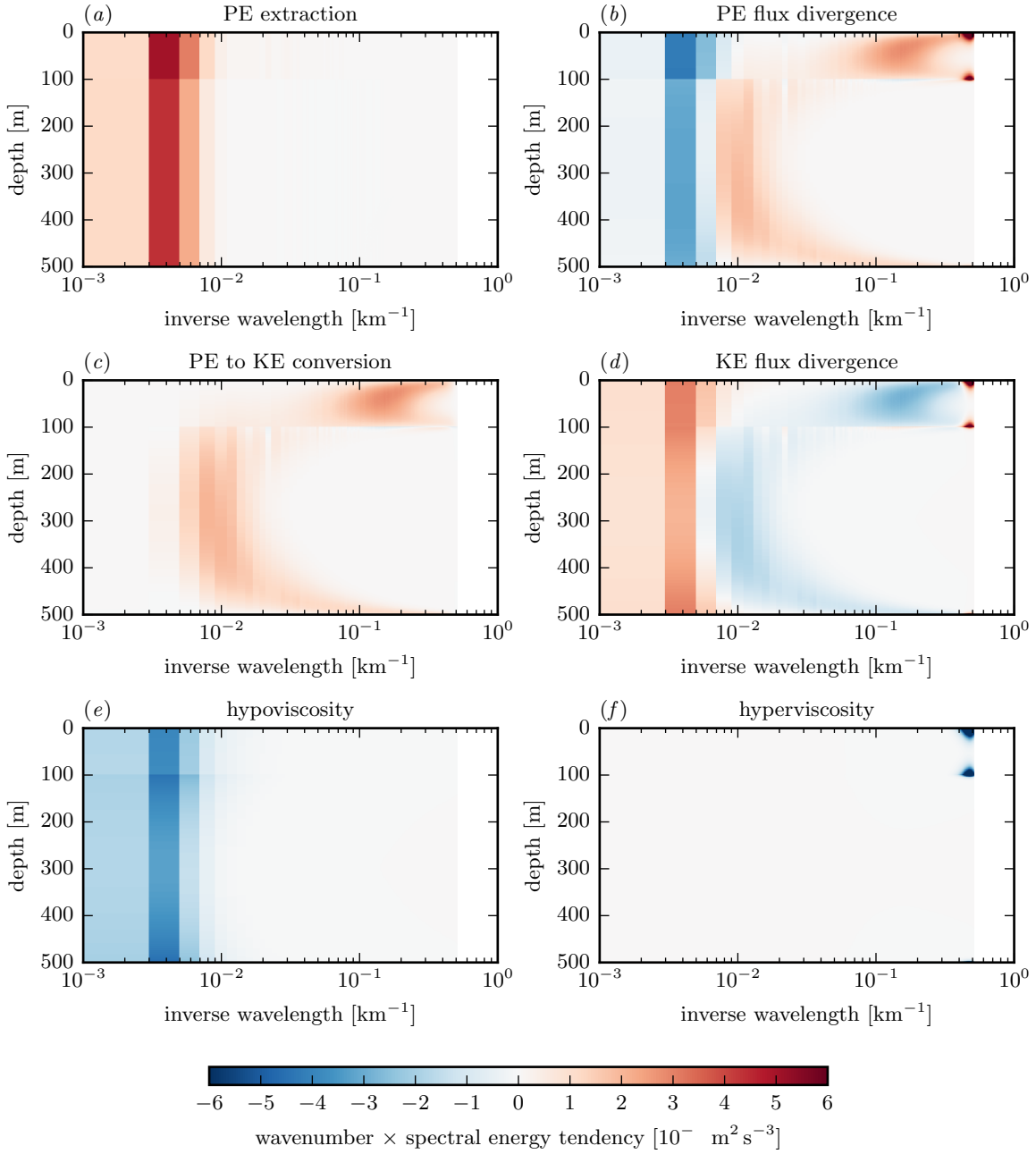

Figure 18: Spectral energy budget for the full model simulation with reduced mixed layer shear. The terms are (*a*) potential energy extraction from the mean, (*b*) spectral potential energy flux divergence, (*c*) potential to kinetic energy conversion, (*d*) kinetic energy flux divergence, including spectral flux and pressure flux, (*e*) hypoviscosity on both kinetic and potential energy, and (*f*) hyperviscosity on both kinetic and potential energy. All terms are multiplied by the wavenumber to compensate for logarithmic shrinking.

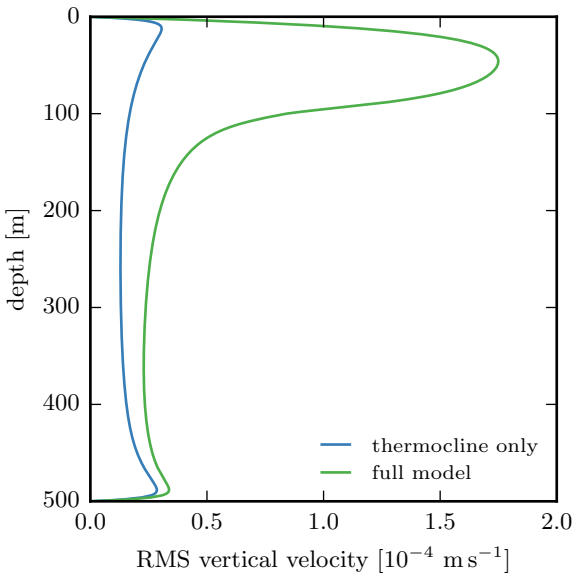

Figure 19: Profiles of root mean square vertical velocity for the thermocline only simulation and the full model simulation with reduced mixed layer shear.

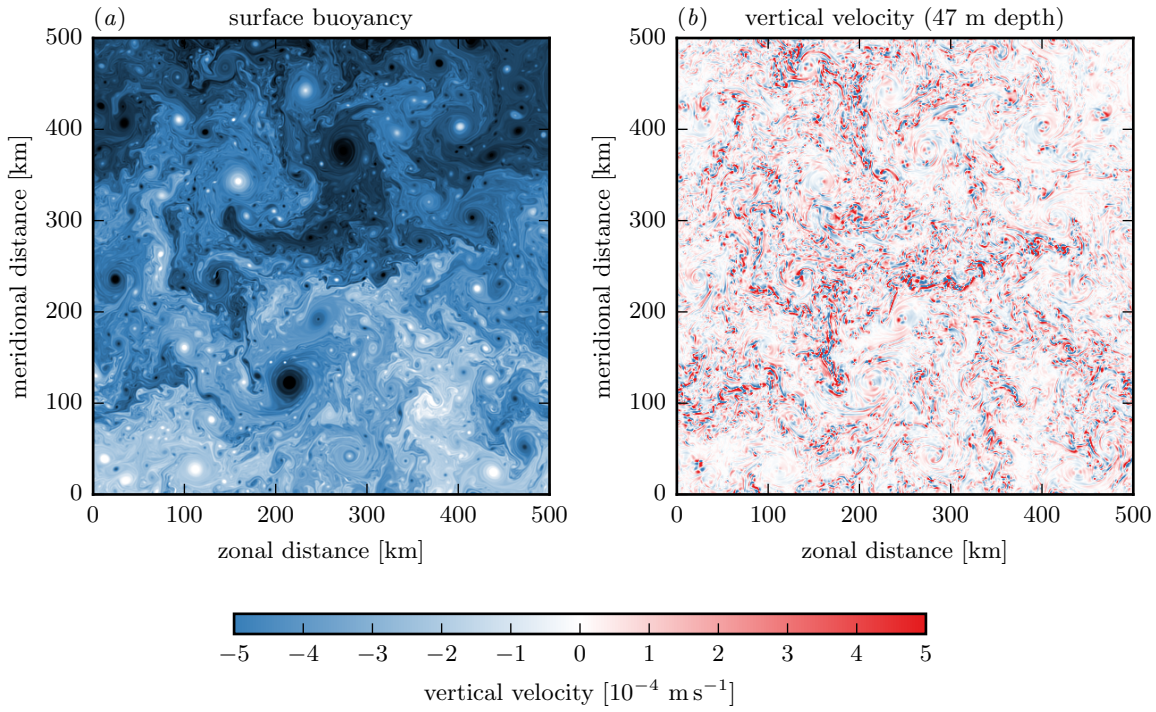

Figure 20: Concurrent surface buoyancy and vertical velocity snapshots from the reduced mixed layer shear simulation. (*a*) The surface buoyancy is the same as in Fig. 7*d*. (*b*) The vertical velocity snapshot is taken at 47 m, the depth of the maximum root mean square vertical velocity.

which are underestimated by our model.

The energization by mixed layer instabilities implies a seasonal cycle in submesoscale turbulence, which has been observed in both models and observations. Surface QG dynamics do not predict a seasonal cycle in submesoscale turbulence, because the dynamics are driven by mesoscale processes. Observational evidence for surface QG dynamics has yet to be found, but the dynamics may be important when mixed layers are shallow and surface buoyancy gradients strong.

An aspect of our model that is inconsistent with observations is that only the submesoscale appears to undergo a strong seasonal cycle, not the mesoscale, as implied by the dynamics of our simple QG model. The energization of the mesoscale by mixed layer instabilities appears to be much less effective in the real ocean. Our model is more consistent with the real ocean if the mixed layer shear is reduced. Then, the mesoscale surface energy levels do not depend on the presence of the mixed layer and the model equilibrates to realistic submesoscale energy levels.

This inconsistency with observations may point to shortcomings of the QG system in predicting the equilibrated energy levels. The lack of small-scale energy dissipation makes the mixed layer instabilities very effective in energizing the entire turbulent flow. If non-QG effects were allowed, a fraction of the energy extracted from the mean in the mixed layer would be dissipated at small scales. This energy leak to small scales is likely as the Rossby and Froude numbers can be large enough at submesoscales to only marginally allow QG scaling. Especially the host of coherent vortices forming in QG dynamics (Fig. 7) may not form in the real ocean, because they would be unstable to non-QG instabilities. The possibility of an energy leak to small scales has been demonstrated by Molemaker et al., who studied an Eady instability with $\mathrm{Ro} = \mathrm{Fr} = 0.5$ using the full Boussinesq equations [23]. While much of the energy extracted from the mean is still trapped at large scales, as predicted by QG dynamics, some is lost to dissipation at small scales. A small leak of energy in the instability may make a big difference in the cascade dynamics, because that energy is not transferred back to mesoscales, where it would further enhance the extraction of potential energy from the mean. We are currently investigating whether primitive equation models indeed equilibrate to lower energy levels than QG models at moderate Rossby and Froude numbers and will report the results elsewhere.

Another possible explanation for the lack of a strong seasonal cycle in the real ocean is that our setup overestimates the amount of mean potential energy available for extraction. QG dynamics do not allow for restratification, which would quickly increase the mixed layer stratification and thereby reduce the available potential energy. We also hold the mean shear fixed, which amounts to providing an infinite reservoir of available potential energy. It does appear, however, that strong wintertime atmospheric forcing, with heat fluxes on the order of several hundred watts per square meter, can effectively maintain both a weak mixed layer stratification and horizontal buoyancy gradients.

Acknowledgments

I would like to thank Glenn Flierl, Raf Ferrari, and Baylor Fox-Kemper, who made every imaginable effort to be available for stimulating discussions, both locally and remotely. I would also like to thank Malte Jansen, Bill Young, Ryan Abernathey, and Paola Cessi

for interesting discussions throughout the summer. I owe thanks to Antonello Provenzale, Glenn Flierl, and Raf Ferrari, our directors, for putting together such an interesting and stimulating program. Special thanks are due to Kerry Emanuel and Geoff Vallis for their engaging lectures and lunchtime discussions with us fellows. And, of course, I am grateful for the companionship of my fellow fellows who made this summer a great and memorable experience.

A Multi-layer model

A system of n layers of constant PV, of thickness h_i and stratification N_i, consists of $n+1$ conserved quantities that are advected by the geostrophic flow at their respective levels. Compared to the two-layer model considered in the main text, additional interface quantities analogous to θ_1 are present. The linear operator in the inversion relation (11) has tridiagonal structure:

$$L = fk_\mathrm{h} \begin{pmatrix} -\frac{\coth\mu_0}{N_0} & \frac{\operatorname{csch}\mu_0}{N_0} & & & & \\ \frac{\operatorname{csch}\mu_0}{N_0} & -\frac{\coth\mu_0}{N_0} - \frac{\coth\mu_1}{N_1} & \frac{\operatorname{csch}\mu_1}{N_1} & & & \\ & \ddots & \ddots & \ddots & & \\ & & \frac{\operatorname{csch}\mu_{n-1}}{N_{n-1}} & -\frac{\coth\mu_{n-1}}{N_{n-1}} - \frac{\coth\mu_n}{N_n} & \frac{\operatorname{csch}\mu_n}{N_n} & \\ & & & \frac{\operatorname{csch}\mu_n}{N_n} & -\frac{\coth\mu_n}{N_n} \end{pmatrix}, \quad (49)$$

where $\mu_i = N_i k_\mathrm{h} h_i / f$. It may be more efficient to solve the inversion relation numerically instead of calculating the inverse of this matrix, which will in general be full.

One can also include a PV gradient due to differential rotation. This can be done using a trick described by Lindzen [21]: instead of using linear shear and constant stratification in the layers, one can use parabolic shear or a modified stratification profile, which allows cancellation of the contribution from the β-effect and retaining constant PV within the layers. The PV gradient due to β is then included in the PV sheets at the interfaces.

B Density jump at layer interface

If there is a buoyancy jump g' at the interface, the matching conditions must be modified. To ensure a continuous pressure at the interface at $z = -h + \eta$, we require

$$\psi(-h^+) - \psi(-h^-) = -\frac{g'}{f}\eta. \quad (50)$$

Here, η is the perturbation of the interface between the constant-PV layers. The buoyancy equations (5) can be combined with the kinematic condition

$$w = \frac{\partial \eta}{\partial t} + \mathrm{J}(\psi, \eta), \quad (51)$$

applied at $z = -h^+$ and $z = -h^-$, to give

$$\frac{\partial \theta_1}{\partial t} + \mathrm{J}(\psi_1, \theta_1) = 0, \qquad \frac{\partial \theta_2}{\partial t} + \mathrm{J}(\psi_2, \theta_2) = 0, \quad (52)$$

where

$$\theta_1 = f\frac{b(-h^+)}{N_{\mathrm{m}}^2} + f\eta, \qquad \theta_2 = f\frac{b(-h^-)}{N_{\mathrm{t}}^2} + f\eta, \qquad \psi_1 = \psi(-h^+), \qquad \psi_2 = \psi(-h^-). \quad (53)$$

Together with the conservation of surface and bottom buoyancy,

$$\theta_0 = -f\frac{b(0)}{N_{\mathrm{m}}^2}, \qquad \theta_3 = f\frac{b(-H)}{N_{\mathrm{t}}^2}, \quad (54)$$

and the inversion relation obtained by solving (12) with the matching conditions above, the model is complete. It now consists of four conserved quantities.

References

[1] W. Blumen, *Uniform potential vorticity flow: Part I. Theory of wave interactions and two-dimensional turbulence*, J. Atmos. Sci., 35 (1978), pp. 774–783.

[2] ——, *On short-wave baroclinic instability*, J. Atmos. Sci., 36 (1979), pp. 1925–1933.

[3] G. Boccaletti, R. Ferrari, and B. Fox-Kemper, *Mixed layer instabilities and restratification*, J. Phys. Oceanogr., 37 (2007), pp. 2228–2250.

[4] J. P. Boyd, *The energy spectrum of fronts: Time evolution of shocks in Burgers' equation*, J. Atmos. Sci., 49 (1992), pp. 128–139.

[5] F. P. Bretherton, *Critical layer instability in baroclinic flows*, Q. J. R. Meteorol. Soc., 92 (1966), pp. 325–334.

[6] J. Callies and R. Ferrari, *Interpreting Energy and Tracer Spectra of Upper-Ocean Turbulence in the Submesoscale Range (1–200 km)*, J. Phys. Oceanogr., 43 (2013), pp. 2456–2474.

[7] J. Callies, R. Ferrari, J. M. Klymak, and J. Gula, *Seasonality in submesoscale turbulence*, Nat. Commun., 6 (2015).

[8] X. Capet, J. C. McWilliams, M. J. Molemaker, and A. F. Shchepetkin, *Mesoscale to submesoscale transition in the California Current System. Part I: flow structure, eddy flux, and observational tests*, J. Phys. Oceanogr., 38 (2008), pp. 29–43.

[9] J. G. Charney, *Geostrophic turbulence*, J. Atmos. Sci., 28 (1971), pp. 1087–1095.

[10] E. T. Eady, *Long waves and cyclone waves*, Tellus, 1 (1949), pp. 33–52.

[11] R. Ferrari, *A frontal challenge for climate models*, Science, 332 (2011), pp. 316–317.

[12] I. M. Held, R. T. Pierrehumbert, S. T. Garner, and K. L. Swanson, *Surface quasi-geostrophic dynamics*, J. Fluid Mech., 282 (1995), pp. 1–20.

[13] J. R. Holton, *An introduction to dynamic meteorology*, Elsevier Academic Press, Burlington, MA, 4 ed., 2004.

[14] B. J. Hoskins and F. P. Bretherton, *Atmospheric frontogenesis models: mathematical formulation and solution*, J. Atmos. Sci., 29 (1972), pp. 11–37.

[15] M. Juckes, *Quasigeostrophic dynamics of the tropopause*, J. Atmos. Sci., 51 (1994), pp. 2756–2768.

[16] P. Klein, B. L. Hua, G. Lapeyre, X. Capet, S. Le Gentil, and H. Sasaki, *Upper ocean turbulence from high-resolution 3D simulations*, J. Phys. Oceanogr., 38 (2008), pp. 1748–1763.

[17] P. Klein and G. Lapeyre, *The oceanic vertical pump induced by mesoscale and submesoscale turbulence*, Ann. Rev. Mar. Sci., 1 (2009), pp. 351–375.

[18] R. H. Kraichnan, *Inertial ranges in two-dimensional turbulence*, Phys. Fluids, 10 (1967), p. 1417.

[19] G. Lapeyre and P. Klein, *Dynamics of the upper oceanic layers in terms of surface quasigeostrophy theory*, J. Phys. Oceanogr., 36 (2006), pp. 165–176.

[20] V. D. Larichev and I. M. Held, *Eddy amplitudes and fluxes in a homogeneous model of fully developed baroclinic instability*, J. Phys. Oceanogr., 25 (1995), pp. 2285–2297.

[21] R. S. Lindzen, *The Eady problem for a basic state with zero pv gradient but $\beta \neq 0$*, J. Atmos. Sci., 51 (1994), pp. 3221–3226.

[22] J. A. Mensa, Z. Garraffo, A. Griffa, T. M. Özgökmen, A. Haza, and M. Veneziani, *Seasonality of the submesoscale dynamics in the Gulf Stream region*, Ocean Dyn., 63 (2013), pp. 923–941.

[23] M. J. Molemaker, J. C. McWilliams, and X. Capet, *Balanced and unbalanced routes to dissipation in an equilibrated Eady flow*, J. Fluid Mech., 654 (2010), pp. 35–63.

[24] J. Pedlosky, *Geophysical Fluid Dynamics*, Springer, New York, 2 ed., 1987.

[25] N. A. Phillips, *Energy transformations and meridional circulations associated with simple baroclinic waves in a two-level, quasi-geostrophic model*, Tellus, 6 (1954), pp. 273–286.

[26] P. B. Rhines, *The dynamics of unsteady currents*, in Sea, vol. VI, E. Goldberg, ed., Wiley, 1977, pp. 189–318.

[27] C. Rivest, C. A. Davis, and B. F. Farrell, *Upper-Tropospheric Synoptic-Scale Waves. Part I: Maintenance as Eady Normal Modes*, J. Atmos. Sci., 49 (1992), pp. 2108–2119.

[28] G. Roullet, J. C. McWilliams, X. Capet, and M. J. Molemaker, *Properties of steady geostrophic turbulence with isopycnal outcropping*, J. Phys. Oceanogr., 42 (2012), pp. 18–38.

[29] R. Salmon, *Two-layer quasi-geostrophic turbulence in a simple special case*, Geophys. Astrophys. Fluid Dyn., 10 (1978), pp. 25–52.

[30] K. S. Smith and E. Bernard, *Geostrophic turbulence near rapid changes in stratification*, Phys. Fluids, 25 (2013), p. 046601.

[31] K. S. Smith and G. K. Vallis, *The scales and equilibration of midocean eddies: freely evolving flow*, J. Phys. Oceanogr., 31 (2001), pp. 554–571.

[32] P. H. Stone, *Frontogenesis by horizontal wind deformation fields*, J. Atmos. Sci., 23 (1966), pp. 455–465.

[33] G. K. Vallis, *Atmospheric and Oceanic Fluid Dynamics*, Cambridge University Press, New York, 2006.