Recent Progress in Modelling Imbalance in the Atmosphere and Ocean

Bruce R. Sutherland,¹,²,* Ulrich Achatz,³
Colm-cille P. Caulfield,⁴,⁵ and Jody M. Klymak⁶,⁷

¹Department of Physics, University of Alberta,
Edmonton, Alberta, Canada T6G 2E1
²Department of Earth and Atmospheric Sciences,
University of Alberta, Edmonton, Alberta, Canada T6G 2E3
³Institut fuer Atmosphaere und Umwelt,
Goethe-Universitaet Frankfurt, Frankfurt, Germany 60438
⁴BP Institute, University of Cambridge, Cambridge, UK CB3 0EZ
⁵Department of Applied Mathematics and Theoretical Physics,
University of Cambridge, Cambridge, UK CB3 0WA
⁶School of Earth and Ocean Sciences,
University of Victoria, Victoria,
British Columbia, Canada V8W 3P6
⁷Department of Physics and Astronomy,
University of Victoria, Victoria,
British Columbia, Canada V8W 3P6

Abstract

Imbalance refers to the departure from the large-scale primarily vortical flows in the atmosphere and ocean whose motion is governed by a balance between Coriolis, pressure-gradient and buoyancy forces, and can be described approximately by quasi-geostrophic theory or similar balance models. Imbalanced motions are manifest either as fully nonlinear turbulence or as internal gravity waves which can extract energy from these geophysical flows but which can also feed energy back into the flows. Capturing the physics underlying these mechanisms is essential to understand how energy is transported from large geophysical scales ultimately to microscopic scales where it is dissipated. In the atmosphere it is also necessary for understanding momentum transport and its impact upon the mean wind and current speeds. During a February 2018 workshop at the Banff International Research Station (BIRS), atmospheric scientists, physical oceanographers, physicists and mathematicians gathered to discuss recent progress in understanding these processes through interpretation of observations, numerical simulations and mathematical modelling. The outcome of this meeting is reported upon here.
I. INTRODUCTION

With some exceptions, the atmosphere and ocean are described by the same equations of motion, the two fluids differing primarily by their equation of state. And so it is not surprising that they exhibit features in common. The most energetic motions of the atmosphere occur at synoptic and planetary scales, on the order of 1000 km and longer, being manifest as perturbations of the mean zonal (east-west) winds by cyclonic flows (rotating in the same sense of the Earth’s rotation) and anticyclonic flows (rotating in the opposite sense) respectively associated with low and high pressure systems. The analogous flows in the ocean occur at the mesoscale, on the order of 100 km, being manifest as cyclonic and anticyclonic eddies embedded in the eastward Antarctic Circumpolar Current and in western boundary currents such as the Gulf Stream, and the Kuroshio and Agulhas Currents. Flows such as these are said to be “balanced”, meaning that Coriolis forces are equal to pressure-gradient forces such that quasi-geostrophic potential vorticity is conserved as expressed through the quasi-geostrophic equations. Generally a suite of balanced models (quasi-geostrophy, semi-geostrophy, planetary geostrophy, and others) has been devised to account for large-scale motion in the tropics, dynamics leading to frontogenesis, and submesoscale motions in the ocean where isopycnals (constant density surfaces) are relatively steep. (See Klein for a recent review of scalings leading to different balance models of the atmosphere.) Common to all balance models is that they filter out smaller spatial-scale and faster time-scale motions associated in particular with internal gravity waves, which are waves that propagate horizontally and possibly vertically being driven by buoyancy forces where the effective background density decreases with height.

Constantly driven by incoming solar radiation (which itself drives winds that force the upper ocean), an outstanding question for atmospheric scientists and physical oceanographers is how energy at these large scales is transferred progressively to smaller and smaller scales until ultimately it is dissipated efficiently by viscosity, closing the global energy balance. One conduit through which these transfers can take place is through direct generation of fully nonlinear turbulence arising when balanced motions interact with solid boundaries, for example through the atmospheric boundary layer or or when ocean currents and eddies interact with continental margins and, particularly with the Antarctic Circumpolar Current, with the ocean floor. Away from such boundaries, the main conduit for energy transfer from balanced motions to dissipative scales is through internal gravity waves. Although there are similarities, the processes of energy transfer by waves differ significantly in atmosphere and ocean.
In the atmosphere, observations suggest that internal waves are generated primarily by zonal flows over topography and by non-orographic sources such as convection and the formation of fronts. As a consequence of momentum conservation, these waves grow in amplitude as they propagate upwards into thinner, less dense air. Due to such so-called “anelastic” growth as well as interaction with the large-scale winds, the waves break and deposit their momentum, so decelerating or accelerating the mean wind speeds aloft. Breaking waves also vertically mix the air tending to homogenize it. However, absorption of solar radiation by ozone and other gases readily re-establishes stable stratification.

In the ocean, internal waves are generated primarily as a result of either wind stress on the surface or through the action of tides and eddies embedded within currents that move over bottom topography. These generation and other breakdown processes are illustrated in Figure 1. The density of the ocean changes little over its depth and so upward propagating waves do not experience anelastic growth. Except in the equatorial Pacific Ocean, around Antarctica and at the western boundaries of the ocean basins, ocean currents are generally weak and have negligible influence on wave propagation and breaking. Instead, it appears that the main mechanism leading ultimately to dissipation is through nonlinear wave-wave interactions and wave-induced shear that directly results in turbulent breakdown. Whereas atmospheric scientists are most interested in momentum transport by internal waves, this is less of interest to physical oceanographers because the directionality of the waves is generally horizontally isotropic. Instead, oceanographers are most interested in energy transport by internal waves which ultimately determines the degree of mixing that occurs as they breakdown into turbulence. Because sunlight penetrates little below the top 100 m of the ocean surface, the observed stratification of the oceanic abyss can only be explained by internal wave-driven processes.

“Modelling imbalance” typically describes the endeavour to go beyond the balanced (e.g. quasi-geostrophic) equations in order to capture the physical processes by which energy is exchanged between balanced flows and smaller scale motions. Here we take a broader perspective by considering in addition the cascade of energy from inertia-gravity waves to faster and smaller scales through the internal wave spectrum and then onwards to dissipation via turbulence, which itself is typically affected non-trivially by stratification. While we do not include consideration of the direct dissipation of energy due to balanced motions interacting with solid boundaries, the reader is referred to recent reviews of research on atmospheric boundary layers and of the oceanic kinetic energy budget.
Guiding this review are the talks and discussions that took place during the Workshop on Modelling Imbalance in the Atmosphere and Ocean, which was held at the Banff International Research Station (BIRS) during February 19 to 23, 2018. Therein presenters described recent breakthroughs realized through advances in observational technologies, increased computing power, and newly developed mathematical modelling techniques. Crucially, it is through the combination of these resources that advances have taken place. For example, now observational campaigns in the ocean are often guided by...
numerical simulations that predict where dynamics of interest may be taking place. Mathematical methods are now being applied to observations to distinguish balanced vortical motions from waves. Simulations have improved to the extent that they can now begin to capture the observed spectra of internal gravity waves as well as the processes of turbulent mixing and dissipation in stratified fluids at geophysically realistic parameter values. Just as models are being created to adapt quasi-geostrophy to include unbalanced motions that are less influenced by rotation, higher resolution global simulations are now beginning to resolve such small scales that hydrostatic models need to be adapted to account for non-hydrostatic effects.

All of these advances have the ultimate goal of developing parameterizations of energy and momentum transport from large to small scales in atmospheric and oceanic general circulation models. Here again the emphasis of atmosphere and ocean modellers differ. While internal wave dynamics are important for short term weather forecasts and predictions of clear-air turbulence, climate models of the atmosphere are less concerned about the details of transport by internal waves. Typically their influence is parameterized by assuming they propagate purely in the vertical and depositing all their momentum where they break, as estimated by linear theory. That said, the cumulative impact of internal waves on the middle atmosphere is known to be important. It virtually controls the mean circulation in the mesosphere and leaves significant traces in the stratosphere, e.g. by contributing to the quasi-biennial oscillation in the equatorial stratosphere and the formation and breakdown of the ozone hole in the southern hemisphere. Since middle-atmosphere dynamics is essential for weather predictions on the seasonal time scale as well as climate simulations, middle-atmosphere internal-wave dynamics and its correct handling in models is important for these issues as well. By contrast, internal waves are thought to be of primary importance to the ocean’s long-term climatology, particularly in the abyss. With increasing observational data and through insights gained by theoretical and numerical modelling of idealized circumstances, ocean general circulation models are only now beginning to develop parameterizations of internal wave-induced mixing. But it is clear that a deeper understanding of the various breakdown mechanisms is required before such parameterization schemes can be agreed upon.

The principal aim of this review is to report recent advances, as discussed in the workshop, rather than to present a comprehensive historical survey. Therefore, the references presented herein are inevitably incomplete, yet hopefully act as a useful starting point for exploring this very important topic. We begin in Sec.II by describing recent progress in distinguishing balanced from
imbalanced flows in observations and in representing such processes by a reduced set of equations. Significant advances have been made in this area with the recognition that internal waves interacting with balanced flows may enhance the generation of internal waves through the process of stimulated imbalance. Other mechanisms for extraction and transport of energy from balanced flows through interaction with topography, convection, and turbulence are discussed in Sec.III. Primarily through numerical simulations, insights have been gained into the final stages of breakdown into turbulence, momentum deposition and mixing. These are presented for the atmosphere in Sec.IV and for the ocean in Sec.V. In light of this progress, the most promising directions for future advances are discussed in Sec.VI.

II. CHARACTERIZING BALANCE AND IMBALANCED FLOWS

Earth observing satellites readily visualize synoptic-scale cyclonic motions in the atmosphere through the structure of clouds as shown, for example, in the left image of Fig. 2. Their counterpart in the ocean, manifest as mesoscale eddies can be visualized through satellite imagery of the sea-surface temperature, as shown in the right image of Fig. 2. With some further processing that includes sea surface height data and the assumption of geostrophic balance, global surface currents at the mesoscale can be revealed as derived, for example, by the Ocean Surface Current Analyses Real-time (OSCAR) product. These large scale motions are said to be “balanced”, meaning that their motion is determined by a combination of pressure gradient and Coriolis forces as well as the influence of vortex stretching. These dynamics are captured by the so-called “quasi-geostrophic (potential vorticity) equations”. In the atmosphere, these describe the synoptic scale motions which vary over distances of about 1000 km and greater. In the ocean, these describe the mesoscale motions which vary over distances of about 100 km and greater.

While large-scale motions are clearly distinguished qualitatively from finer strained structures, it is more challenging quantitatively to partition the energy associated with the balanced motions and the unbalanced motions, which occur at smaller scales. In situ observations can be gathered in the atmosphere, for example of wind speed and temperature from airplanes and radiosonde balloons, and in the ocean, for example of speed, temperature and salinity from shipboard acoustic Doppler current profilers (ADCPs) and conductivity, temperature and depth (CTD) profiles. However, these provide limited spatial information.

One advance in the interpretation of velocity measurements along a given
FIG. 2. Images acquired from the Moderate Resolution Imaging Spectroradiometer on NASA’s Terra satellite showing (left) atmospheric cyclones between Iceland and Scotland visualized by clouds on November 20, 2006 [NASA image by Jesse Allen, Earth Observatory] and (right) anticyclonic and cyclonic eddies interacting with the Gulf Stream on May 8, 2000 and visualized by sea surface temperature [Created at the University of Miami using the 11- and 12- micron bands, by Bob Evans, Peter Minnett and co-workers]. The horizontal white bar to the lower right in the left image indicates a length of 250 km. The vertical white bar to the upper left of the right image indicates a length of 222 km.

line (e.g. taken from ship or aircraft tracks) is the use of the Helmholtz decomposition, which separates the fields in terms of their vertically rotational and horizontally divergent components. Significantly, Bühler et al. and Callies et al. showed how to separate geostrophic from internal wave motions while Bühler et al. extended the method to account for the anisotropy in the spectra resulting from eddies and waves being embedded within mean winds and currents. Through these methods, as shown in Fig. 3, it is clear that the power spectrum as a function of horizontal wavenumber, $k$, varies as $k^{-3}$ at mesoscales in the ocean, as predicted by Charney (corresponding to a downscale transfer of enstrophy), and that the energy at these scales can be attributed almost entirely to balanced (rotational) motion. At mesoscales in the atmosphere ($\lesssim 100$ km) and at the submesoscale in the oceans ($\lesssim 50$ km) the spectrum varies as $k^{-5/3}$. The $-5/3$ spectrum is predicted for the downscale cascade of energy in three-dimensional turbulence and the upscale transfer of energy in two-dimensional turbulence. If indeed this spectral slope is associated with two-dimensional turbulence, as has been proposed for the
atmosphere\textsuperscript{27,28}, this suggests that energy at mesoscale ranges in the atmosphere and submesoscale ranges in the ocean is driven “inversely” to large rather than dissipative scales. Other processes must operate in order to provide a global energy balance. These include mechanisms, discussed below, in which internal gravity waves are generated by balanced motions and, by interacting either with these motions or with themselves, energy is carried to small scale, either indirectly via weakly nonlinear interactions or directly by wave breaking. Another possible mechanism not involving waves is through vertical shear which develops naturally in stratified quasi-two-dimensional turbulence.

As a first theoretical step to understand how energy is extracted from balanced flows in the absence of boundaries, attempts have been made to extend the balanced (e.g. quasi-geostrophic) equations to introduce the influence of divergent processes associated with inertia-gravity waves\textsuperscript{29–34}. Conceptually, these processes can be broken down into triad interactions between vortical modes (V) and waves (W) acting through the nonlinear advection terms of the fully nonlinear equations of motion\textsuperscript{35}.

The quasi-geostrophic equations can be represented by the triad ‘VVV’, meaning that two vortical modes interact (first two letters) to put energy into another vortical mode (the last letter). The triad ‘VWW’ represents the equations describing the spontaneous generation of internal waves from interactions between vortical modes\textsuperscript{30}. Due to the corresponding eigenfrequencies
only non-resonant interactions are allowed. Indeed, analysis of these equations in comparison with observations suggest that this mechanism for wave generation is relatively weak. That said, a recent breakthrough has been the demonstration that stimulated emission of internal waves, represented by the triad ‘VWW’, may provide an efficient mechanism through which waves interacting with the vortical modes produce more waves. The question remains as to whether these excited waves in the ocean are able to propagate away from the generation site and so remotely cause mixing elsewhere in the ocean. A recent study by Nagai et al\textsuperscript{36} suggests this is not the case: after the waves are generated it appears they are re-absorbed into the large-scale balanced flows, as represented by the triad ‘VWV’. A particular mechanism for re-absorption of waves by large scale flows through the straining motion between vortices was formulated by Bühler and McIntyre\textsuperscript{37}. Another possibility in the ocean is for waves to become trapped in anticyclonic eddies\textsuperscript{38–40}, eventually re-depositing their energy there.

In the context of reduced models, the triad ‘WWV’ describes waves interacting with themselves without breaking in a way that transfers energy to large-scale flows. For vertically propagating waves, this is manifest as the (Eulerian) induced mean flow\textsuperscript{6,41–44} that, for a localized wavepacket, can be responsible for creating a circulation far from the wavepacket itself\textsuperscript{45}. Of course another mechanism by which waves can affect the large-scale flow, particularly in the atmosphere, is through momentum deposition caused by breaking. However, the process of breaking does not involve a resonant wave-wave interaction, as implied by the triad WWV. Instead it involves a cascade of energy to the small scales of turbulence, the detailed description of which lies outside the realm of reduced models. Some aspects of such turbulent processes are discussed below.

The last of the interaction triads, ‘WWW’, corresponds with weakly nonlinear interactions between waves. This can involve refraction of waves by other waves of different scale\textsuperscript{46}, the self-interaction of waves in a spanwise-wide wavepacket that resonantly generates long waves\textsuperscript{44,47–49} or wave-wave interactions leading to parametric subharmonic instability\textsuperscript{50–52}. In particular, it has been proposed that the last process may be important for the breakdown of the M2 internal tide at the critical latitude of 29°, where the subharmonic wave frequency equals the background Coriolis frequency\textsuperscript{53–55}. However, in situ measurements did not reveal as strong a signal of breakdown as suggested by numerical models\textsuperscript{56}. It is likely that processes described by the ‘WWW’ triad will explain the prevalence of the universal spectra for internal waves observed in the abyssal ocean\textsuperscript{57–59}, and possibly also in the middle atmosphere, whether through the downscale transport of energy through scattering and
parametric subharmonic instability or through the upscale transport of energy from small-scale wavepackets generated, for example in the atmosphere through convection\textsuperscript{27,28,60} or by localized turbulent patches associated with wave breaking\textsuperscript{61}, mechanisms that will be discussed below. While much has yet to be explored regarding upscale energy transfers in the context of two-dimensional turbulence or otherwise, numerical models are now able to achieve resolutions capable of reproducing these spectra from the balanced flows to the inertia-gravity wave (low wavenumber) end of the internal wave spectrum.

Apart from triad interactions, energy may be dissipated directly through wave breaking. This can occur due to anelastic growth of atmospheric waves propagating upward into thinner air, due to waves approaching a critical level where their horizontal phase speed matches the background flow speed, or due to other interactions that drive the waves to shear or convective instability. Observations of overturning in the oceanic thermocline suggest that wavepackets interacting with large scale flows overturn predominantly due to convective rather than shear instability\textsuperscript{62} though shear instability may play a more important role for inertia gravity waves (cf §4.6.3 of Sutherland\textsuperscript{63}).

Despite the upscale transfer of energy in two-dimensional turbulence, quasi-two-dimensional turbulence in non-rotating stratified fluid may nonetheless develop structures leading to efficient dissipation in the absence of waves. Here it is important to note that while a power spectrum of an idealized two-dimensional turbulence field inevitably and naturally reveals a spectrum, it is incorrect to interpret this spectrum physically as a pure superposition of waves: vortices develop naturally in two-dimensional turbulence, and a power spectrum loses the information of such coherent, spatially localized structures. Stratification can render incoherent the vertical structure of vortices in quasi-two-dimensional turbulence: columnar vortices naturally devolve into “pancake” eddies\textsuperscript{64}.

This suggests an alternative framework to model quasi-two-dimensional turbulent flow in stratified fluid which still leads to $-5/3$ slope of the energy spectrum by relying on strong anisotropy in the flow\textsuperscript{65} to give a forward cascade of energy consistent with numerical evidence\textsuperscript{66}. It is important to appreciate that this argument inherently assumes that rotation plays no dynamical role. As recently demonstrated by Kafiabad & Bartello\textsuperscript{67}, this places an upper bound of $O(10 \text{ km})$ in the atmosphere where this regime might be expected to be observable. Analogously in oceanographic contexts, this regime is only expected to occur at below sub-mesoscales, with horizontal scales $O(1 – 10 \text{ km})$ or less.

As demonstrated by the self-similar scaling analysis of Billant & Chomaz\textsuperscript{68}, such non-rotating ‘layered anisotropic stratified turbulence’ (LAST)\textsuperscript{69} is in-
evitably and inherently three-dimensional. Although the horizontal scales $L$ are at least in some sense large, though not too large for rotation to play a dynamical role as characterized by the horizontal Froude number $\text{Fr}_h \equiv U/(NL) \ll 1$, the vertical structure scales as $H \sim U/N$ such that the vertical Froude number $\text{Fr}_v = U/(NH)$ is order unity. In effect this introduces a different mechanism for energy dissipation whereby as two pancakes slide over each other the vertical shear over the small distance between them can result in shear-driven turbulence and mixing if the Richardson number, $\text{Ri} \equiv N^2H^2/U^2$, is sufficiently small, which is certainly possible if $\text{Fr}_v = \text{Ri}^{-1/2} \sim O(1)$.

For shear instability to occur between pancakes, the flow must be of sufficiently high Reynolds number $\text{Re} = UL/\nu$, based on the (largest) horizontal scale such that the combination $\text{Re}\text{Fr}_h^2 \gg 1$. With a further inertial scaling for the turbulent dissipation rate $\epsilon \sim U^3/L$, this scaling for the onset of instability corresponds to the requirement that the quantity conventionally referred to as the “buoyancy Reynolds number” $\text{Re}_b \equiv \epsilon/(\nu N^2)$ is much larger than unity. The quantity $\text{Re}_b$ can also be expressed in terms of the ratio of the Ozmidov scale $\ell_O \equiv (\epsilon/N^3)^{1/2}$, which is the largest vertical scale of turbulence not strongly affected by stratification, and the Kolmogorov scale $\ell_\nu \equiv (\nu^3/\epsilon)^{1/4}$, which is the smallest turbulence scale at which dissipation occurs. Explicitly, $\text{Re}_b = (\ell_O/\ell_\nu)^{4/3}$, thus implying that high $\text{Re}_b$ flows as observed geophysically, have a very high dynamic range over which the flow is inherently nonlinear and turbulent, with no significant role played by the internal wave field. Large values of $\text{Re}_b$ thus imply a five-fold hierarchy of scales for the LAST regime to exist such that $L_R \gg L \gg H \gg \ell_O \gg \ell_\nu$, where $L_R$ is an appropriate characteristic horizontal scale above which rotational effects can no longer be ignored. As a consequence, this phenomenon, if it even occurs, is difficult to replicate in the laboratory and to simulate numerically, although high-resolution simulations are beginning to access this regime. Furthermore, though there is observational evidence (see for example Falder et al) that this regime occurs, a key outstanding challenge is to understand how it is connected to, and interacts with rotationally, or indeed even wave-dominated dynamical processes. At the moment, the fluid-dynamical research in this area is inevitably somewhat disconnected from oceanic and atmospheric flows, and there is a pressing need to bridge this gap, both in terms of constructing (and testing) new theories, but also generating numerical and observational data-sets capturing the transition (or indeed transitions) in flow dynamics. As perhaps a first step on the theoretical side, some progress has been made in developing a reduced equation set that exploits the large
horizontal to vertical scale separation to allow for a “quasi-linear”, yet still
non-trivial coupling between such widely separated scales\textsuperscript{33}, constituting a
very interesting advance on classical approaches leading to “pure” hydrostatic
flow.

III. INTERNAL WAVE GENERATION BY TOPOGRAPHIC, CONVECTIVE AND TURBULENT PROCESSES

Besides spontaneous and stimulated generation of internal waves by large
scale flows, there are several other mechanisms by which internal waves can be
excited. One of the most significant of these is that of flow over topography,
although the main processes of generation and consequent evolution once again
differ qualitatively in the atmosphere and ocean.

In the atmosphere, observations suggest that the largest source of internal
waves results from the unidirectional flow of mean zonal winds over mountain
ranges\textsuperscript{3}. Whether or not internal waves are generated depends upon four fac-
tors: the characteristic flow speed at the mountain height, $U$, the characteristic
horizontal length scale of the topography, $L$, the stratification as character-
ized by the buoyancy frequency, $N$, and the Coriolis frequency $f$. From linear
theory, vertically propagating waves exist only if their intrinsic frequency, es-
timated by $U/L$, lies between $N$ and $f$, with typical values of $N \approx 10^{-2} \text{s}^{-1}$
being two orders of magnitude larger than $f \approx 10^{-4} \text{s}^{-1}$ at mid-latitudes\textsuperscript{1,63}. Waves generated on the scale of mountain ranges have frequencies close to $f$
and so should be classified as inertia-gravity waves. (Waves generated at very
large horizontal scales are also influenced by the change in the Coriolis fre-
quency with latitude, the so-called “beta effect”. Study of these waves, known
as planetary or Rossby waves, lies beyond the scope of this review.) General
circulation models of the atmosphere are now beginning to resolve such small
scales that they must begin to include nonhydrostatic effects associated with
waves having frequency moderately below $N$. Being nonhydrostatic means
that buoyancy forces are not balanced by vertical pressure gradients and so
can lead directly to vertical accelerations.

The amplitude of topographically generated internal waves is set by the
height, $H$, of the topography. This is a crucial parameter since it influences
the height at which the waves ultimately break in the absence of encounter-
ing a critical level, which is where the upper level winds change their zonal
direction so that the wind speed equals the (ground-based) zero-horizontal
phase speed of the waves. Because the density of the atmosphere decreases
approximately exponentially with height, the amplitude of the waves grow as
they propagate upward in order to conserve momentum. Due to this anelastic growth, the waves ultimately break when they grow to such large amplitudes that they overturn. Waves actually generated at large amplitude can overturn in the upper troposphere irrespective of anelastic effects, sometimes leading to downslope windstorms, and in the lower stratosphere, potentially leading to clear-air-turbulence, which is a hazard to aircraft.

A recent comprehensive observational campaign called the Deep Propagating Gravity Wave Experiment (DEEPWAVE) used ground-based lidar, radar and airglow images as well as in situ aircraft measurements to make unprecedented measurements of the life-cycle of internal waves generated by flow over mountains in southern New Zealand and extending into the ionosphere situated above 85 km from the surface. Energy fluxes by internal waves were found to be highly variable in part owing to the pre-existence of large-scale internal waves in the middle atmosphere generated by other sources. Some of the fluxes were the largest ever recorded with values over 20 times typical values. This campaign also made the first observations of small horizontal scale and small amplitude waves penetrating through the middle atmosphere to the ionosphere. In some instances wave breaking was so large as to mix sodium ions in the ionosphere downwards by 10 km. It has also been found that from the total spectrum of waves emitted by flow over topography the less energetic components are most important for the mesosphere, since the more energetic components already break at lower altitudes.

Conversely, in the ocean it is believed that most of the energy put into the internal wave field results from the oscillatory flow of the barotropic tide (i.e. the tide with virtually no vertical structure) over submarine ridges and at the continental shelves, a process sometimes described as “baroclinic conversion”. More recently, observations and modelling efforts suggest that significant internal waves may also be generated by mesoscale eddies interacting with topography or by the flow of the Antarctic Circumpolar Current over rough bottom topography resulting in lee waves. The efficiency by which energy from the barotropic tide is converted into internal waves is set by the tidal frequency relative to the buoyancy and Coriolis frequencies and the maximum topographic slope, $s^*$. Independent of their wavenumber, internal waves with frequency $\omega$ propagate with lines of constant phase forming an angle to the horizontal of $\alpha = \tan^{-1}[(\omega^2 - f^2)/(N^2 - \omega^2)]$, provided $f < \omega < N$. The barotropic tide of frequency $\omega$ efficiently generates internal waves with the same frequency if $s^* > \tan \alpha$ in which case the internal waves emanate as beams originating from the topography where its slope equals $\tan \alpha$. Such topography is said to be "supercritical".

Comprehensive observational studies of internal wave generation by the
barotropic tides in the ocean have been carried out at a few sites in the open ocean\textsuperscript{82,83}. These studies provided evidence for enhanced mixing near topography due to breaking internal tides\textsuperscript{84}. However, most energy appears to escape the hot-spots of generation. The life-cycle of these radiating waves remains poorly understood. There is no anelastic growth of waves in the ocean and, except for the transient occurrence of eddies, critical levels play little role in wave breaking. Far from topographic sources, internal waves are manifest as mode-1 internal waves, characterized by sinusoidal oscillations of the thermocline. Sometimes these mode-1 waves steepen, as observed west of Luzon Strait in the South China Sea, and form solitary wave trains\textsuperscript{85,86} that eventually shoal and break. Otherwise it remains unclear how these mode-1 waves ultimately deposit their energy into turbulence and mixing. Studies examining mode-1 waves incident upon continental shelves indicate that most of the energy is reflected back to the open ocean with dissipation recorded in a few localized regions\textsuperscript{87,88}, though the proportion reflected is thought to be
quite variable.\textsuperscript{89}

Mean and eddying flows in the ocean also generate internal waves and are thought to be a substantial sink to such flows\textsuperscript{80,90}. Linear lee-waves are believed to radiate from topographic roughness with modest topographic height, as represented by the inverse vertical Froude number or Long number $L_0 ≡ N h / U \lesssim 1$, and with horizontal scales smaller than $\approx 6$ km. These lee waves are then envisioned to undergo wave-wave interactions that drive the cascade to turbulence\textsuperscript{76}, though empirical evidence of this process has been hard to detect\textsuperscript{78,91}. Larger-scale topography is expected to exert blocking effects, a process that has been realized to be important for parameterizing mountain drag in the atmosphere\textsuperscript{92,93}, and for generating low mode internal tides in the ocean\textsuperscript{94–96}. This blocking and its effects on mean flows is only beginning to be looked at in the oceanic context\textsuperscript{79,97,98}.

Convection is understood to be another important source of atmospheric internal waves, particularly above the equatorial oceans and in the southern mid-latitudes in summer, but possibly also in northern mid-latitudes\textsuperscript{4,99,100}. The waves can be excited by vertical motions within storm clouds, induced by latent heating, flow over the cloud tops or by the vertical oscillations of the cloud tops\textsuperscript{101}. In addition, wave generation is influenced by the collective behaviour of merging convective cells, frontogenesis and the thermodynamics of moist convection\textsuperscript{102,103}. Recent numerical models that capture internal waves generated at the submesoscale by these processes demonstrate that injection of energy into waves at these small scales contributes to the observed $-5/3$ kinetic energy spectrum in the mesoscale through upscale energy transfers\textsuperscript{60,102}. One global-scale manifestation of these small-scale processes is the Quasi-Biennial Oscillation (QBO) in which stratospheric winds above the equator alternately flow eastward and westward on an approximately two-year time-scale. The QBO is understood to result from momentum deposition of vertically propagating eastward and westward Kelvin, Rossby-gravity, and internal waves where they encounter critical levels\textsuperscript{104}. While general circulation models of the atmosphere are able qualitatively to reproduce the QBO, the details of its amplitude and period differ between different models, which is an indication of the sensitivity of the results due both to the convective parameterization scheme used by each\textsuperscript{105} and to the gravity-wave parameterization\textsuperscript{15}.

The role of convection in exciting internal waves in the ocean is unclear at present. In part this is because rapid cooling of the ocean surface, as occurs beneath a hurricane or by low pressure systems passing over the Labrador Sea in winter, is associated with strong surface wind stress, which itself acts as a significant hydrodynamic rather than thermodynamic source of waves\textsuperscript{106}. Observations of processes occurring directly underneath the storms remains
a challenge, though progress may be made as more autonomous vehicles are being deployed in the world's oceans. Besides such extreme events, there is some evidence that Langmuir circulations (counter-rotating cells aligned with the wind) may generate internal waves at scales between meters and kilometers\textsuperscript{107}.

Indeed, turbulence itself can act as a source of internal waves as observed in laboratory experiments\textsuperscript{108,109} and through simulations and observations in the atmosphere of secondary generation of internal waves resulting from the breaking of primary waves\textsuperscript{61,74,110,111}. Investigations of the coupling between turbulence and internal waves remains active research.

IV. DRAG PARAMETERIZATIONS IN THE ATMOSPHERE

Climate simulations necessarily are run at relatively coarse resolution so that predictions on decadal and century time-scales can be produced in reasonable time. Although such models are beginning to capture mesoscale horizontal motions with scales on the order of 50-200 km, they are unable to capture the dynamics of the generation, propagation and breaking of mesoscale (\(\approx 100 \text{ km}\)) and submesoscale (\(\approx 10 \text{ km}\)) internal waves despite observational evidence that suggests such dynamical behaviour has a non-negligible influence upon synoptic scale motions in the middle atmosphere\textsuperscript{112–114}. By not appropriately incorporating the effects of subgrid-scale waves, general circulation models fail to reproduce realistic zonal mean winds and temperatures in the middle atmosphere\textsuperscript{112,115–117}, which affects seasonal and climate forecasts.

Guided by observations and the results of high-resolution numerical simulations, gravity wave drag parameterization schemes have progressively improved over the years through better representation of topographic and non-orographic sources. However, general circulation models continue to underpredict the temperature of the southern-hemisphere winter polar night jet in the middle atmosphere, which crucially influences accurate predictions of the ozone hole evolution over Antarctica. While the southern tip of Chile and New Zealand are significant topographic sources of internal waves, recent evidence suggests that islands in the southern ocean whose sizes lie below the climate model resolutions may nonetheless contribute non-negligibly to the total momentum flux\textsuperscript{118,119}. Moreover, especially intermittent wave generation from non-orographic sources, particularly strong winds, constitute an important source of waves\textsuperscript{120}. In particular, the comparison shown in Figure 5 between measurements of momentum fluxes inferred by radiosonde balloons and those predicted and resolved by the European Centre for Medium Range Forecasts.
FIG. 5. Comparison between simulations (left) and radiosonde balloon observations (right) of 5-month time-averaged momentum fluxes at 19 km altitude in the Southern Hemisphere. The simulations from ECMWF have had their spatial resolution downgraded to correspond to that of the balloon data and their values have been multiplied by 5. Although the spatial pattern of the fluxes is similar, the simulation underpredicts observations by a factor of 5. [© American Meteorological Society. Used with permission from Figure 1c,d of Jewtoukoff et al.]

(ECMWF) operation model indicate that modelled fluxes can be a fifth of what is observed. This hints at excessive damping of internal waves resolved in the ECMWF model at the respective altitudes in the lower stratosphere. However, it might also indicate that further improvements are necessary to account for non-orographic and transient wave sources, as well as other effects.

The discrepancies between observed and modelled internal-wave momentum fluxes seem indirectly to point at other fundamental problems of present-day gravity-wave parameterizations. A model analysis of parameterized gravity-wave drag in spring during southern stratospheric final warming showed that these events were simulated in a realistic manner though with gravity-wave momentum fluxes similar to those diagnosed from the above-mentioned ECMWF simulations. As the latter, however, are known to be too weak in comparison to measurements, only with too weak an input of internal-wave fluxes at launch-altitude can the effects at higher altitude be captured well. As well as improving parameterizations of sources for internal
waves, the representation of their propagation should be improved. Idealized theoretical models and simulations suggest that the effect of time-transient background winds, wave energetics and lateral propagation of the waves as well as weakly nonlinear effects acting upon moderately large amplitude waves may need to be incorporated into the next generation of parameterization schemes. Further corresponding indirect observational evidence is provided again from radiosonde balloons, showing a conspicuous dependence of the intermittency of internal-wave momentum fluxes on the large-scale wind strength that parameterizations cannot reproduce. Finally, in all parameterizations the handling of internal-wave breaking is still very crude, and often in considerable disagreement with findings from direct simulations of this process.

V. MIXING PARAMETERIZATIONS IN THE OCEAN

The ocean is driven at the boundaries by surface winds and tides, and some of this energy is carried into the interior by internal waves. Through most of the ocean the mean flow of currents is weak and the propagation of waves is horizontally isotropic. Thus the influence of drag is generally less important. However, there is a growing appreciation that drag may play a role where mesoscale eddies in strong currents such as the Antarctic Circumpolar Current and the western boundary currents interact with topography. Here the drag more likely acts on the eddy field itself, rather than the large-scale circulation. Because ocean climate models do not yet resolve mesoscale eddies, parameterization of drag presently is not an important issue, but likely will become one as the resolution of these models increases. Anticipating this eventuality, idealized studies of lee wave drag are being revisited in the context of oceanography.

At present general circulation models of the ocean are primarily concerned with mixing as a result of breaking waves. Only through such mixing can the observed abyssal stratification of the ocean be explained. Physical oceanographers remain actively engaged in observational and modelling efforts in order to understand the energy cascade so that better parameterizations can be formulated.

As in the atmosphere, significant advances have been made through detailed in situ and satellite observations as well as numerical models that capture both mesoscale and submesoscale dynamics. However, there are many mechanisms through which energy at large scales ultimately results in mixing at small scales, and the relative importance of these processes de-
FIG. 6. Estimates of global ocean energy flux inferred from observations. Colours represent a log scale from $10^{-4}$ W/m$^2$ (white) to $10^{-1}$ W/m$^2$ (dark red). The spidery dark red features correspond to internal modes launched by tidal flow over bottom topography; the remaining light-red zones represent inertia gravity waves launched by wind stress in the ocean. [©American Meteorological Society. Used with permission from Figure 2 of Waterhouse et al$^{134}$].

Turbulence is spatially and temporally intermittent and inhomogeneous. The most important outstanding issue in the improvement of ocean general circulation models is to develop a better understanding of the connection between the large-scale flows and stratification captured by the simulations and the subgrid-scale location and time of occurrence of turbulence, and in particular the cumulative diapycnal mixing induced by such inherently spatio-temporally intermittent turbulent events. Although it may yet be some years before observations and modelling efforts provide better predictions for the onset of turbulence, some progress has been made linking mixing to energy dissipation when turbulence actually occurs.

Most parameterizations of mixing in the ocean quantify the turbulent kinetic energy dissipation rate $\epsilon$ as a function of fine- or mesoscale parameters. (It is always important to remember that this quantity is itself highly spatially and temporally intermittent, with growing evidence that it can be elevated by orders of magnitude in the vicinity of “rough” or “ridge”-like bottom topography$^{134}$.) However, what is needed for parameterization in larger-scale simulations is a turbulent (vertical, and hence approximately diapycnal) eddy
diffusivity for density, \( \kappa_\rho \equiv B/N^2 \), in which \( B \equiv (g/\rho_0)\langle w'\rho' \rangle \) is the (vertical) buoyancy flux. Osborn\(^\text{135} \) argued that on average, particularly for quasi-stationary turbulence it was reasonable to assume that the buoyancy flux is proportional to the dissipation rate: \( B \approx \Gamma \epsilon \), and based on field and laboratory work suggested that the turbulent flux coefficient \( \Gamma \) (often referred to as the “mixing efficiency”) may be bounded by \( \Gamma \lesssim 0.2 \) for a statistically steady state. Studies have shown that during a transient mixing event \( \Gamma \) can be much larger\(^\text{136,137} \). There is also evidence that \( \Gamma \) depends on the nature of the mixing event\(^\text{138,139} \), itself being a function of the buoyancy Reynolds number, \( \text{Re}_b \), and Richardson number, \( \text{Ri} \), and perhaps other parameters (see, for example, Ivey et al\(^\text{140} \)). It remains an area of active and somewhat controversial research to capture the fundamental character of different breaking mechanisms, and in particular the effect of those mechanisms on the particular value of the turbulent flux coefficient\(^\text{141–143} \). In particular, plausible interpretation of the observational evidence in the ocean so far indicates that \( \Gamma \approx 0.2 \) is a reasonable estimate given other uncertainties in measuring and parameterizing ocean mixing\(^\text{144} \). On the other hand, there is accumulating evidence that \( \Gamma \propto \text{Re}_b^{-1/2} \) at sufficiently high values of the buoyancy Reynolds number, at least in some circumstances\(^\text{138,139,141,142} \). Clearly, much further interdisciplinary work and collaboration in this area is required, resolving apparent discrepancies between observation, experiment, simulation and theory. Furthermore, it is important to remember that \( \kappa_\rho \) is the actual quantity of interest for parameterization in larger scale models, and since \( \kappa_\rho = \nu \Gamma \text{Re}_b \), understanding whether there is dependence of \( \Gamma \) on \( \text{Re}_b \) is of leading order importance. Furthermore, though there is a large amount of research activity attempting to model the properties of \( \Gamma \), in practice the inherent uncertainty in \( \epsilon \) and \( N \), and hence \( \text{Re}_b \) plausibly swamps any robust effect of \( \Gamma \)-variability on estimation of \( \kappa_\rho \). Just to mention one (potentially very important) example of profound uncertainty, it is still an open question as to whether \( \Gamma \) and/or \( \kappa_\rho \) have qualitatively different behaviour when the flow is in the LAST regime as described in section 2 above.

Issues with the parameterization of \( \Gamma \) aside, substantial recent efforts have been made to develop better parameterizations of ocean mixing\(^\text{8} \), leading to spatially variable vertical mixing being included in ocean models in justifiable ways\(^\text{145} \). These methods usually involve as a first step computing the energy that is believed to radiate into the internal wave field either from tides\(^\text{146} \) or from mean and eddying flows\(^\text{76,147} \). The simplest of these parameterizations then assume a fixed fraction of the radiated energy goes to turbulent dissipation in a decay scale near the generation site, presumably due to wave-wave interactions.
interactions. More sophisticated versions assume horizontal isotropy and use wave-wave interaction theory to decide when the wavefield is susceptible to breaking, and deposit the radiated energy at those depths. These “local” mixing schemes have been found to matter for large-scale modelling and to deep-water buoyancy budgets.

These schemes however, often do not dissipate all their energy at the generation site, with large vertical wavelength internal waves able to escape the region of generation. This is particularly true of large obstacles like mid-ocean ridges and seamounts, where most of the internal wave energy generated goes into low-mode waves. For large obstacles, parameterizations have been developed to estimate the local mixing, but in those cases more than 90% of the energy is still believed to escape the bathymetry. As discussed above some of this breaks at remote locations. Efforts are underway to track this radiated energy in consistent manner and decide how it will be dissipated on remote shores.
VI. SUMMARY AND FUTURE DIRECTIONS

This review has shown that there remain several open questions regarding fundamental processes involved with energy transfers in the ocean and atmosphere from large scales (> 100 km) to dissipative scales (< 1 mm, though larger in the upper atmosphere). Theoretical and observational insights have guided the development of parameterizations for these phenomena in general circulation models. However, their limitations are evident.

It remains a challenge in observations, simulations and theoretical modelling to capture the dynamics occurring in the full range of scales from synoptic and mesoscale through the internal wave spectrum and downward to dissipative scales. Numerical simulations of idealized circumstances (e.g. uniform stratification and zero background flow) are still far from having the resolution to capture the full range. Simulations further suggest that nontrivial effects come into play when the stratification and background flow are not uniform\(^\text{155}\).

As well as building a better database of observations and improving the resolution of simulations there may be other processes important to understanding the global energy budget. For example, a recently proposed hypothesis is that mixing and transport in the deep ocean is driven primarily by its sloping boundaries, not least because of the substantially enhanced turbulence there.\(^\text{133,156,157}\) It is becoming increasingly evident how important it is to study key processes by idealized and semi-realistic simulations, by measurements and observations, and at all stages by critically checking how well internal-wave parameterizations respect corresponding findings. This seems to be the only way to limit the use of over-tuned parameterizations that might be insufficiently reliable in climate-change studies where the simulated conditions cannot agree anymore with those used for present-day empirical parameter optimizations. However, with increasing computing power, parameterizations will be able to incorporate and represent more detailed dynamics. More work is to be done in this direction, with interesting prospects ahead.

ACKNOWLEDGMENTS

We are grateful to all the participants of the BIRS Workshop for their stimulating presentations and discussion and whose feedback contributed to the contents of this review: Brian Arbic, Gergely Bölöni, Oliver Buhler, Greg Chini, Laura Cimoli, Dale Durran, Raffaele Ferrari, Oliver Fringer, David Fritts, Elena Gargarina, Alain Gervais, Basem Halawa, Steffen Hien, Laura
Holt, Shaun Johnston, Hossein Kafiabad, Eric Kunze, Sonya Legg, Pascale Lelong, Yongxing Ma, Richard Peltier, Rob Pinkel, Riwal Plougonven, David Randall, Richard Rotunno, Hesam Salehipour, Mark Schlutow, Kat Smith, David Straub, Ujjwal Tirwari, Jacques Vanneste, Cat Vreughenhil, Caitlin Whalen, Henrike Wilms, Greg Wagner and Qi Zhou. The two anonymous reviewers of this manuscript provided much appreciated constructive comments. We also wish to thank BIRS for their financial support and, in particular, the staff of BIRS for their excellent administration of the workshop. The authors gratefully acknowledge financial support by the following agencies: Achatz - German Research Foundation (DFG) for partial support through the research unit Multiscale Dynamics of Gravity Waves (MS-GWaves) and through Grants AC 71/8-2, AC 71/9-2, AC 71/10-2, AC 71/11-2, and AC 71/12-2; Caulfield - EPSRC Programme Grant (EP/K034529/1) entitled ‘Mathematical Underpinnings of Stratified Turbulence’; Klymak - US Office of Naval Research (N00014-15-1-2585) and Natural Science and Engineering Research Council (NSERC) Discovery Grant (327920-2006); Sutherland - NSERC Discovery Grant (RGPIN-2015-04758).

* Email: bruce.sutherland@ualberta.ca


Information and videos of several talks given during the workshop can be viewed at https://www.birs.ca/events/2018/5-day-workshops/18w5119.


The OSCAR product can be viewed at https://www.esr.org/research/oscar/.


Different groups have used other letters to represent vortical modes and waves, for example using ‘G’ for geostrophic to represent the rotational, slow time-scale flows and ‘A’ for ageostrophic to represent horizontally divergent, relatively fast
time-scale flows.

The buoyancy Reynolds numbers has also been called the “activity parameter” or the “Gibson number”.


D. S. Trossman, B. K. Arbic, D. N. Straub, J. G. Richman, E. P. Chassignet,
100 L. A. Holt, M. J. Alexander, L. Coy, A. Molod, W. Putman, and S. Pawson,
146 L. C. St. Laurent, H. L. Simmons, and S. Jayne, Geophys. Res. Lett. 29,


