Introduction to Physical Oceanography
GEF 2610

Pål E. Isachsen and Kai H. Christensen
University of Oslo

December 13, 2017

Contents

1 Introduction

1.1 The role of the ocean in the climate system

The world oceans are of course the habitat of exuberant amounts of life, possibly even dominating land areas in terms of total biomass. But here, in this course, we will focus on the physical aspects of the ocean and, in particular, on the ocean circulation itself. The natural tendency for flows in Earth's atmosphere and ocean (and on any other planet in the universe, as far as we know) is for light fluid to spread out on top of heavier fluid. The process lowers the center of mass and converts gravitational potential energy into kinetic energy. The kinetic energy is then, eventually, dissipated via friction to heat. Nature steadily works towards a state of increased entropy! Since this large-scale gravitational adjustment is typically *slanted* so that also horizontal motions are involved, the end result is also a reduction in the equator-to-pole density gradient.

In the ocean, warm and fresh waters are lighter than cold and salty waters. Leaving salt aside for now, we can thus say that the ocean circulation tends to spread warm waters above cold waters. So the waters warmed up by the sun in the tropics tends to flow polewards to displace the the cold surface waters there. The cold waters duck underneath and flow equatorward (Figure 1). The net effect of the lateral component of these flows is a poleward heat transport that helps moderate the climate on the planet. The same can be said for the atmospheric flow. Were it not for this tendency, the equator-to-pole temperature contrast on Earth would be much much larger and only a very narrow band of latitudes would be inhabitable.

But the ocean has more roles to play in the climate system. The heat capacity of the oceans is huge compared to that of the atmosphere. So the oceans act as a *buffer* or *integrator* (essentially a low-pass filter) of any atmospheric temperature variations. The ocean is therefore particularly useful as an indicator of long-term global warming. Figure 2, for example, shows that the ocean surface temperature as well as the depth-integrated ocean heat content have risen gradually over the last decades. And Figure 3 shows estimates of the global-mean sea level height. The long-term rise in sea level is attributed to a combination of melt from land glaciers (particularly from Antarctica and Greenland) as well as water expansion due to higher temperatures.

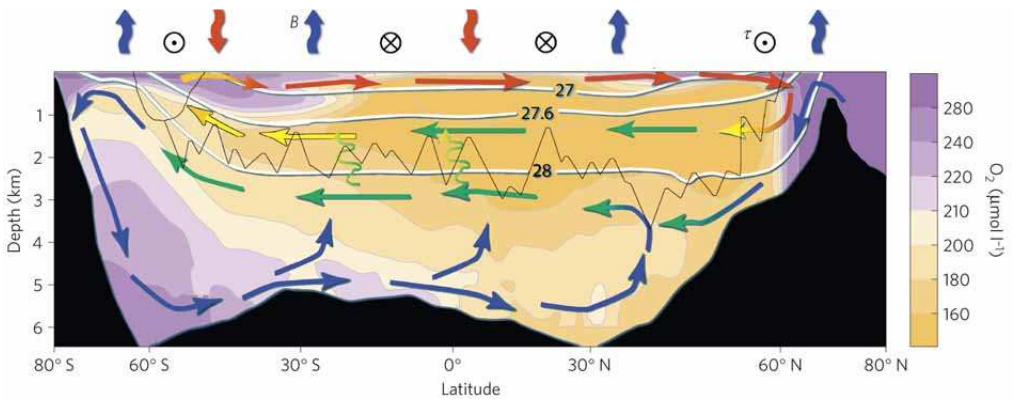

Figure 1: A 2D representation of the oceanic meridional overturning circulation. The color of the arrows indicate temperature. Warm waters flow poleward near the surface while colder waters sink at high latitudes and flow equatorward at depth. The background color indicates the dissolved oxygen concentration. As indicated, the water sinking at high latitudes is also the most rich in oxygen (since it has recently been in contact with the atmosphere).

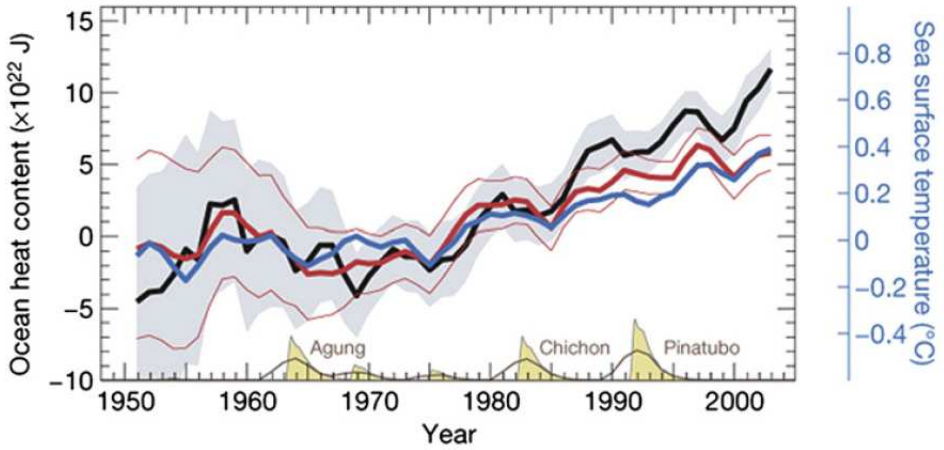

Figure 2: Global ocean heat content and sea surface temperature (SST) over the last decades. (Source: Talley et al., 2011, Fig. S15.17)

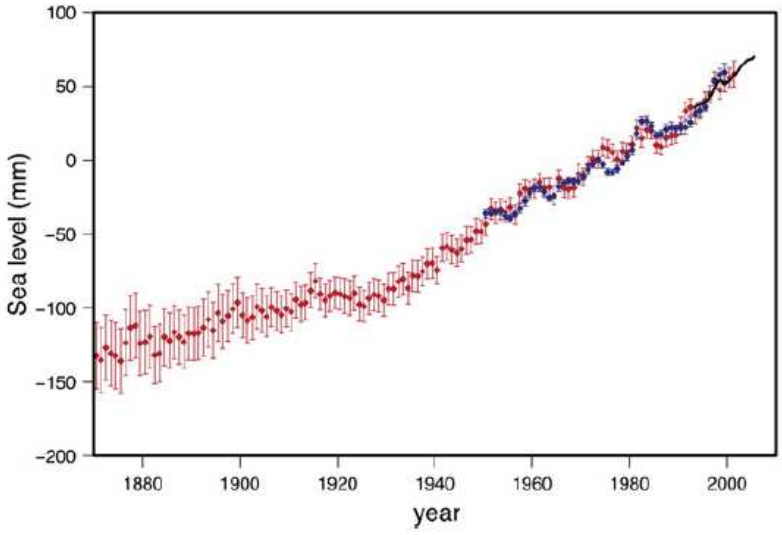

Figure 3: Global sea level from tide gauges (red and blue) and from satellite observations (black). (Source: Talley et al., 2011, Fig. S15.21)

1.2 History of exploring the ocean

It is fair to say that dedicated and systematic large-scale observations of the ocean hydrography (the ocean composition, like temperature, salinity and other chemical properties) and circulation begun with the *H.M.S. Challenger* expedition in 1872–1876. The expedition had multiple purposes, but the ship crossed all the world oceans except for the very highest latitudes (for good reasons) and collected observations of both physics, chemistry and biology (Figure 4).

Since then there have been a number of systematic observational campaigns and programs, the largest of them all taking place during the International Geophysical Year in 1957–58 and during the World Ocean Circulation Experiment (WOCE) in 1990–2002. WOCE involved observations of both currents and hydrography over an extensive 'grid' of observation sections covering the world oceans. The purpose for WOCE was not only to map out the hydrography of the world oceans but also to make *quantitive* estimates of transport of mass and water properties (e.g. heat, freshwater and nutrient transport) between the various boxes defined by this grid of sections (Figure 5). So during WOCE observations were made of property concentrations and about flow velocities (to make transport estimates).

The logistical challenges of *in situ* (on the spot) ocean observations are daunt-

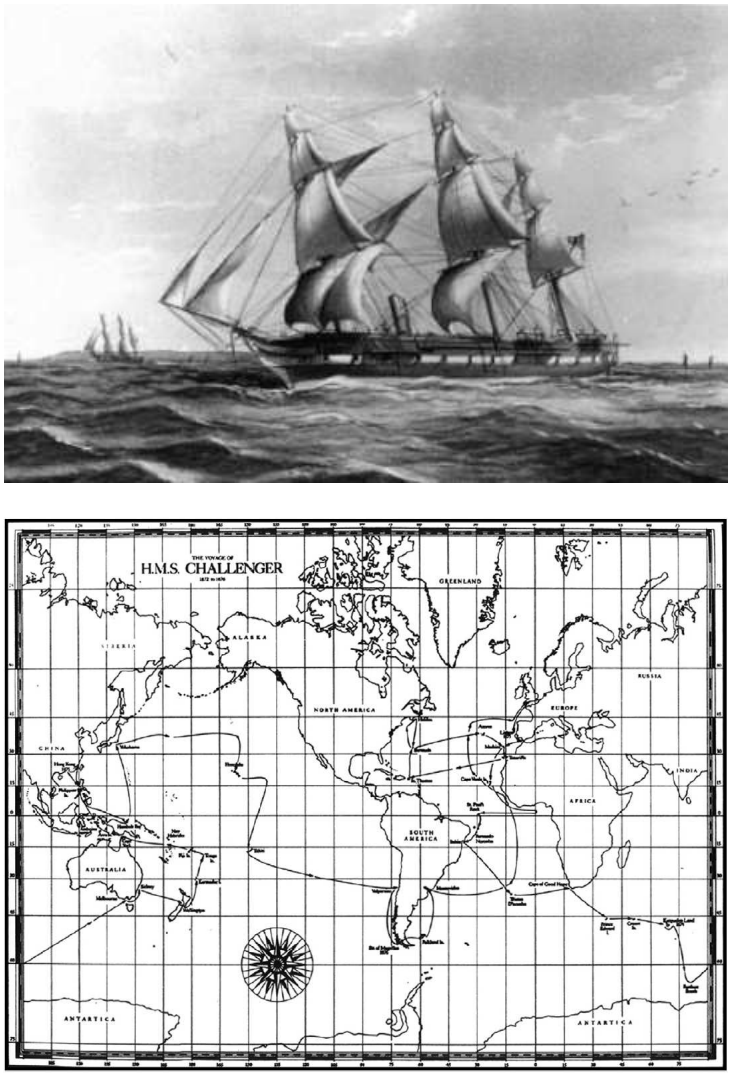

Figure 4: The H.M.S. Challenger expedition, 1872–1876. (Sources: Wikipedia; Talley et al., 2011, Fig. S1.1)

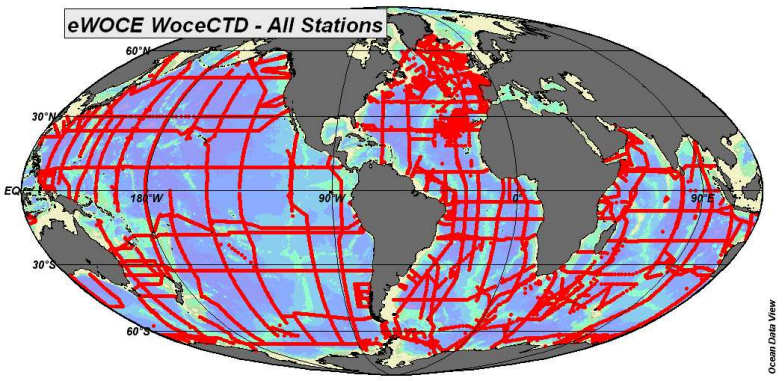

Figure 5: Hydrographic sections of the World Ocean Circulation Experiment (WOCE). (Source: http://ewoce.org)

ing. Just think of the cost of ship time, easily running into tens of thousands of dollars per day. A true revolution in the observation of the oceans therefore came with the advance of satellite remote sensing. Satellites today can give unpresedented observational coverage of the sea surface, including observations of temperature and salinity (both of which determine ocean density), ocean color (which give information about nutrient and sediment concentrations) and, importantly, about sea surface height. As mentioned above, observations of sea surface height give information about the heat content of the ocean. But as we will see later they also give invaluable information about the large-scale ocean circulation itself (specifically, about so-called *geostrophic* currents).

1.3 A first quick look

1.3.1 Bathymetry

As Figure 7 illustrates, the position of the continents and thus the shape of ocean basins have constantly changed over geological time due to the process of plate tectonics, i.e. the large-scale motion of Earth's lithosphere (the outer crust). So the ocean currents and their role in tempering Earth's climate has changed over geological time. The tectonic plates keep moving today as well (Figure 8), but other than providing volcanic and seismic activity (including the generation of tsunamis), the process is too slow to have any pragmatic impact on our view of the oceans. So for our purposes, in this course, we'll stick with the ocean basins as they are today.

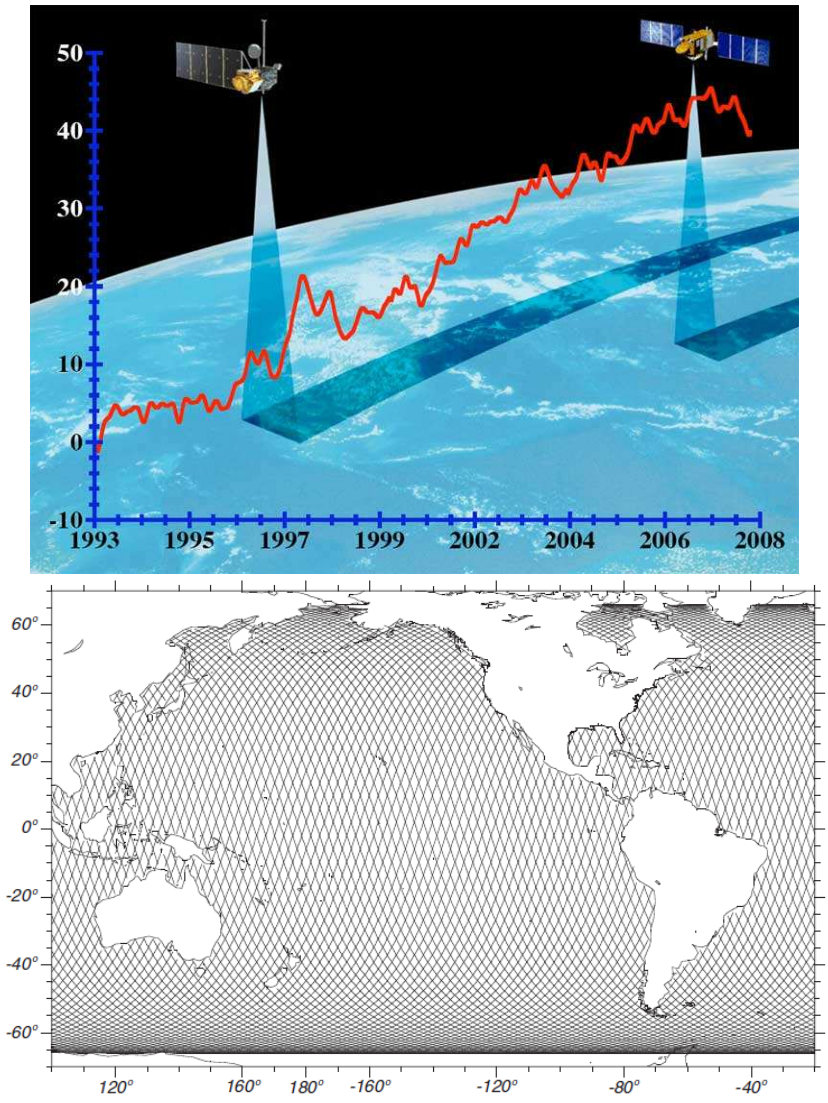

Figure 6: Satellite (Topex/Poseidon) observations of sea surface height, including orbital tracks. (Sources: Wikipedia; Stewart, 2008, Fig. 2.6[Stewart(2008)])

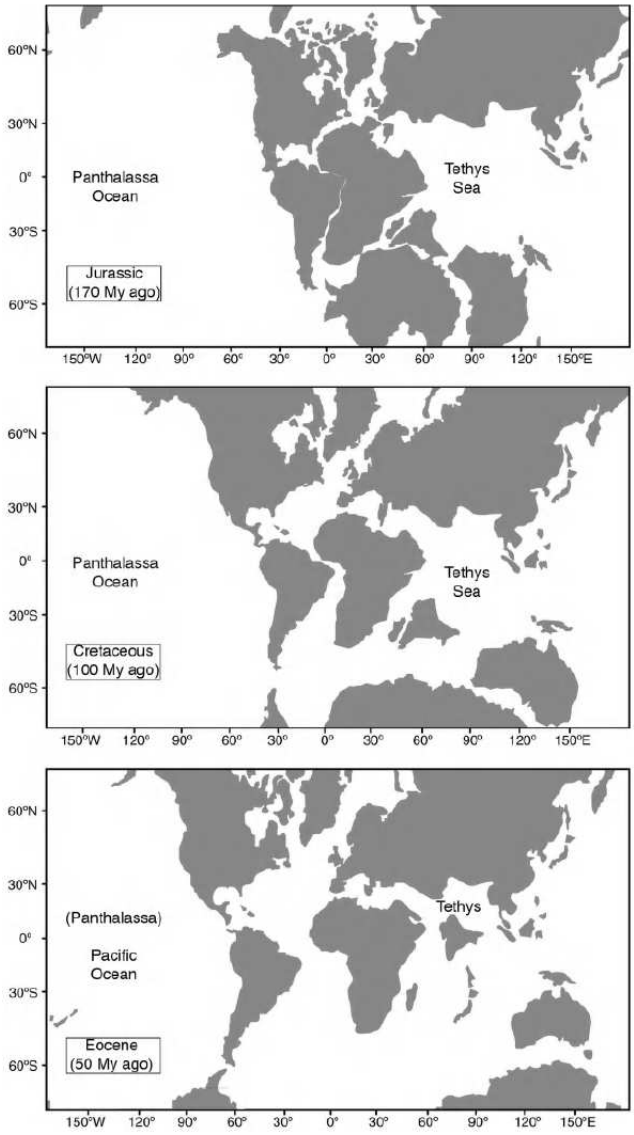

Figure 7: Paleo reconstrictions of the continents as they have evolved over the last 170 million years. (Source: Marshall and Plumb, 2008, Fig. 12.15)

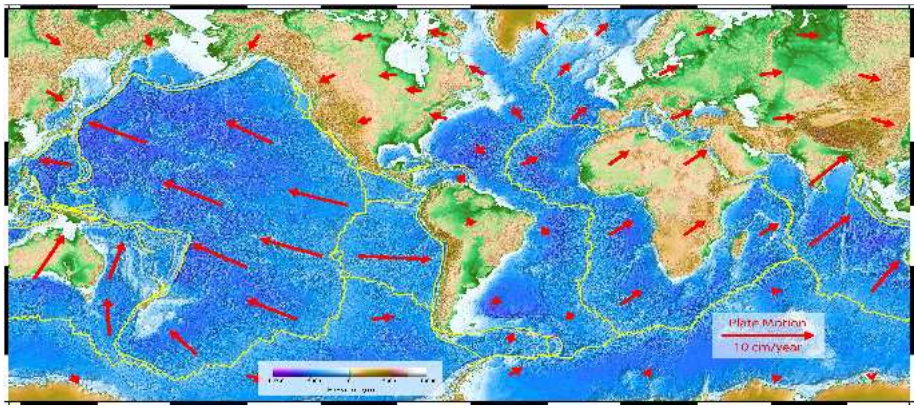

Figure 8: The shape of Earth's continental plates today and the direction at which they are moving. (Source: https://whybecausescience.com)

The world ocean today consists of five major oceans: the Pacific, Atlantic, Indian, Arctic and Southern oceans (Figure 9). As we will discuss later, the presence of continents to the east and the west of the Pacific, Atlantic and Indian oceans make the dynamics of ocean currents there quite distinct from the dynamics governing large-scale atmospheric flows (which experience no such hard boundaries). The Arctic and Southern oceans are less bounded in the east and the west and therefore have large-scale currents that more resemble atmospheric flows. In addition to the major ocean basins there are several smaller seas which are typically shallower and are also to a great extent surrounded by land areas; examples include the Medeterranean Sea, the Gulf of Mexico and the Nordic Seas.

A typical cross section of an ocean (as shown in Figure 10) reveals a number of different bathymetric 'regimes', including the *shore* and shallow *shelf* regions, then a steep *continental slope* which leads out to *abyssal ocean basins*, possibly intersected by very *deep trences* created where two tectonic plates meet. The deep basins may also be separated by *mid-ocean ridges* that are created by underwater volcanic erruptions where tectonic plates separate. Finally, as the figure illustrates, the abyssal basins may be littered with *seamounts* that may even extend through the sea surface (like Hawaii). The ocean bathymetry is every bit as complex as the topography of the land continents.

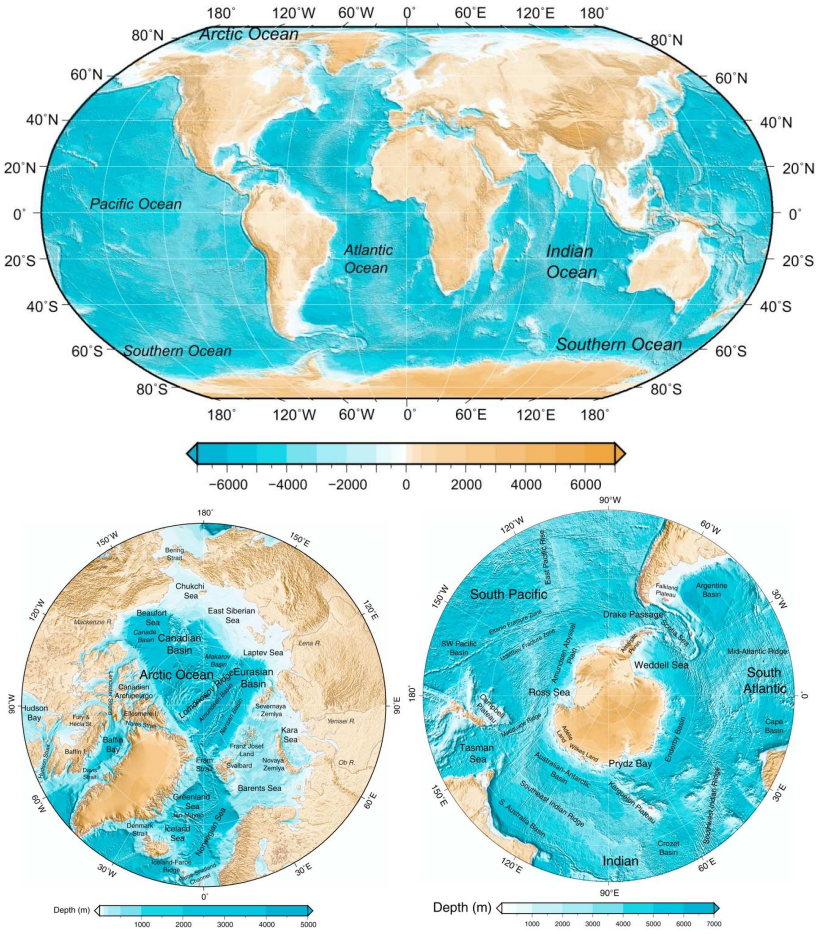

Figure 9: The bathymetry of the world oceans. (Source: Talley et al., 2011, Figs. 2.1, 2.11 and 2.10)

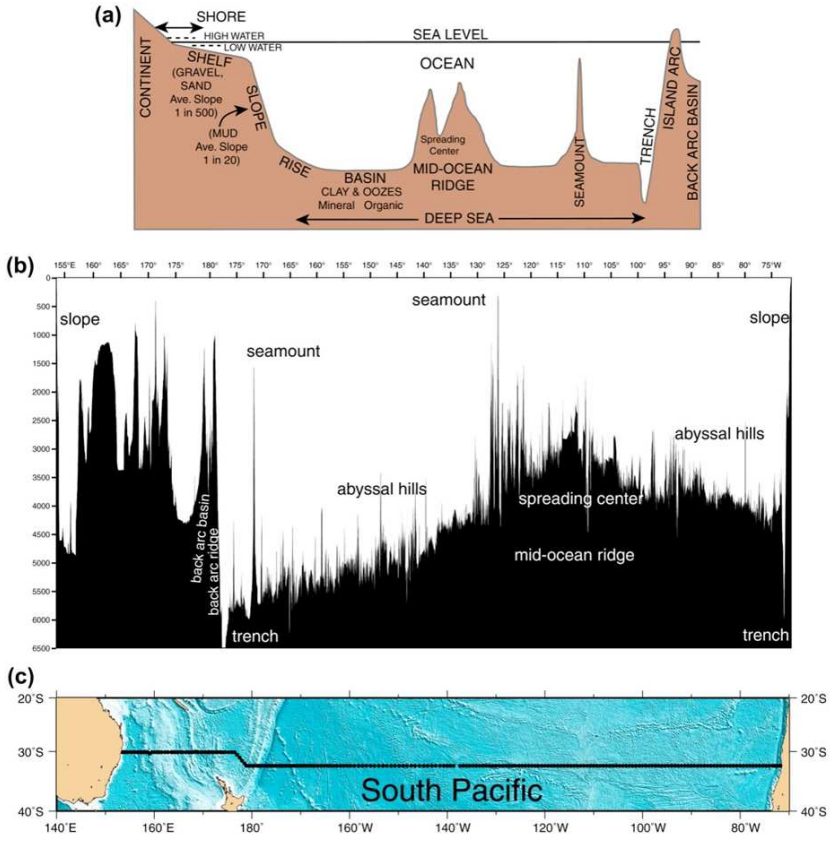

Figure 10: Schematic (top) and actual (middle) bottom bathymetry along an east-west section crossing the South Pacific (bottom). (Source: Talley et al., 2011, Fig. 2.5)

1.3.2 Hydrography

Ocean water is salty. This 'salt' is really dissolved non-organic and non-volatile material (compounds that don't easily vaporize) in the water. The salts consists of all possible types of compounds, basically all that can be transported into the oceans by e.g. rivers that bring with them erroded material. But sodium chloride dominates, making up about 87% of the total. The *salinity* is a measure of the concentration or, more precicely, the mass fraction of these salts. It is defined as the mass in grams of disolved material per kilogram of water. So where one kilogram of water contains 35 grams of salts, we give it a salinity of 35 (with units g/kg).

Figure 11 shows satellite observations of the sea surface salinity (SSS) from 14 November 2012. The observations both large-scale and smaller-scale structure. But most obvious is a tendency for surface waters in the tropics to be salty, a result of exessive evaporation there which removes fresh water while leaving behind the salts. The water salinity is *dynamically* important to the ocean circulation since it, along with temperature, determines the density of water. And, as we have mentioned above, much of the ocean circulation arrises because light waters tend to float on top of denser waters. Salty waters are dense waters and would tend to sink underneath fresher waters if temperature effects on density could be ignored. So from the figure one could be lead to think that the tropical waters, salty as they are, should dive underneath the waters at higher latitudes. But this is clearly not the case, the reason being that water temperature plays a (big) role.

Because, of course, warm waters are lighter than cold waters. We will study the contribution to the ocean temperature budget later, but clearly ocean waters are colder at high latitudes than what they are in the tropics. Figure 12 shows an example of sea surface temperatures (SST) that have been observed by satellites. We see the expected large-scale latitudinal gradients and also smaller gradients, both latitudinal and longitudinal, that are actually due to ocean dynamics itself rather than solar forcing.

The oceans also have complex vertical temperature and salinity structures, as illustrated in Figure 13. The oceans are generally warmer and fresher near the surface since these are the lightest waters. But whether temperature or salinity dominates density actually depends on the temperature itself. So one may actually encounter regions where warm and salty waters overlie cold and fresh waters—and vice versa. Still, temperature typically dominates in setting the water density except for at very low temperatures, i.e. at high latitudes. So the ocean is typically *temperature-stratified*, meaning that it gets progressively denser with

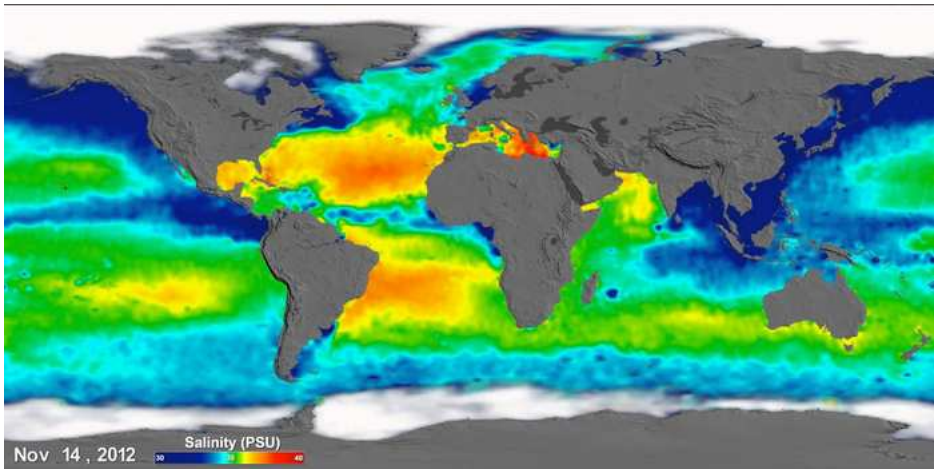

Figure 11: Sea surface salinity (SSS) observed by satellite. (Source: https://svs.gsfc.nasa.gov/cgi-bin/details.cgi?aid=4233)

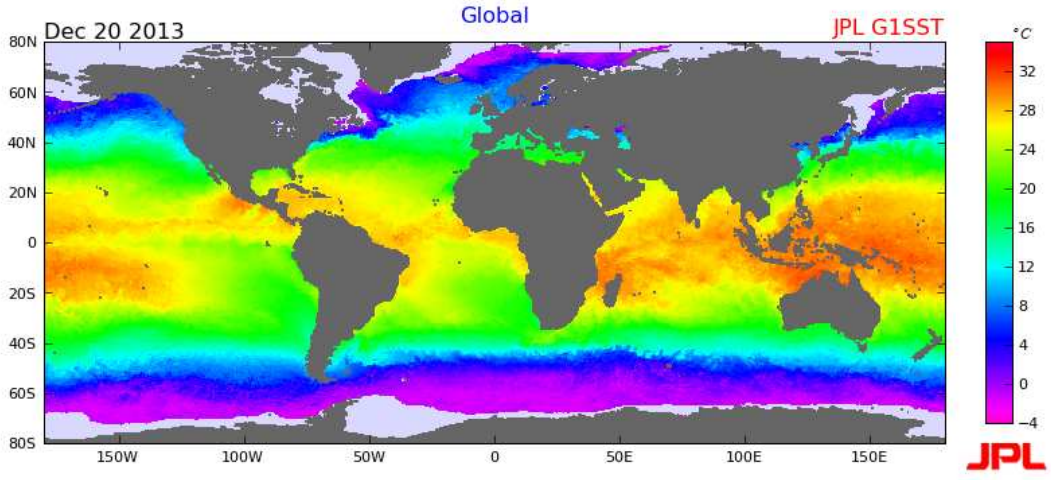

Figure 12: Sea surface temperature (SST) observed by satellite. (Source: Wikipedia)

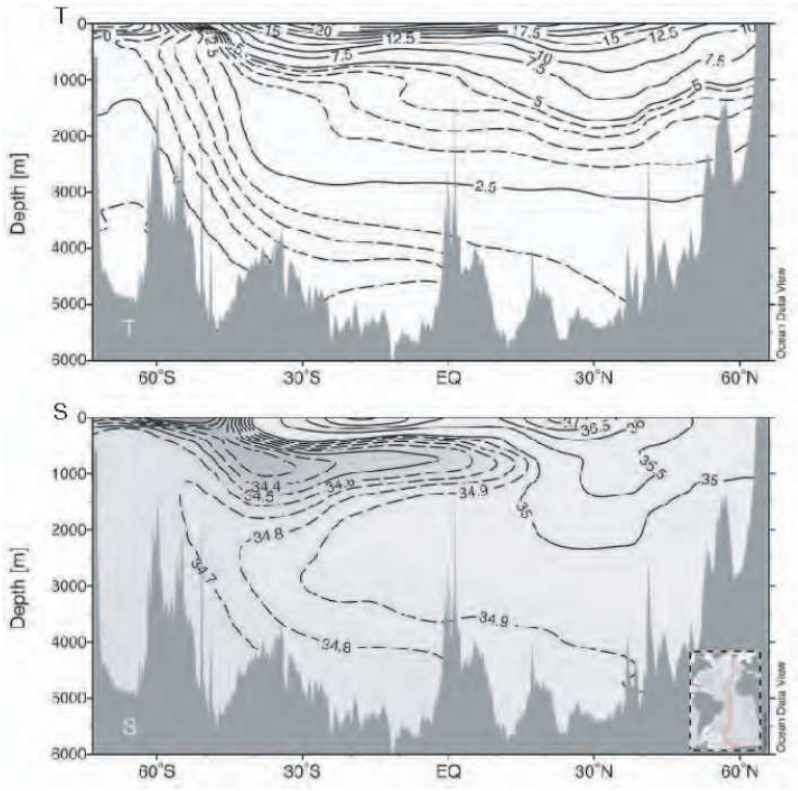

Figure 13: Temperature and salinity along a meridional section through the Atlantic Ocean. (Source: Marshall and Plumb, 2008, Fig. 9.9)

depth because it gets colder with depth. And, so, Figure 13 shows large regions where the water column appears to be *statically unstable* (with heavy waters residing above light waters) due to its salinity structure. But in reality, these regiosn are stable due to the temperature structure.

1.3.3 Currents

Ocean currents are the equivalent of atmospheric winds. As we will discuss at length later, the currents are driven either by the direct frictional 'push' of the winds or, more typically, due to pressure gradients (since water, just like air, tends to flow down the pressure gradient—from high to low pressure). More on this later.

Two schematic representations of the time-mean large-scale surface currents

of the world oceans are shown in Figure 14. In the top panel which is meant to illustrate horizontal currents we see that the currents in the major ocean basins are forming large-scale *gyres* that seem to be constrained in their extents by the presence of eastern and western boundaries (the continents). As we will learn later these are wind-driven gyres that indeed are constrained by the continental boundaries. Some gyres are rotating clockwise while others are rotating counter-clockwise. The flow in the Southern Oceans, which we call the Antarctic Circumpolar Current (ACC), is a notable exception in that it doesn't seem to encouter any notable east-west obstructions. The colored arrows in the figure illustrate how these currents transport waters having different tempereatures around. Generally, as briefly discussed above, currents tend to do their job at moderating Earth's climate by transporting or *advecting* cold waters towards the equator and warm waters towards the poles. In the bottom panel an attempt has been made to illustrate the vertical flow of large-scale currents, showing how warm currents generally flow poleward near the surface while cold currents return towards the equator at depth. Such simplified descriptions are often called 'plumbing diagrams' by the sceptics who feel that they foreshaddow important dynamical aspects (i.e. the governing physical laws) of the flow.

Figure 15gives a better illustration of what real ocean currents look like. If anything, the time-mean currents are hard to pick out from what appears to be a rather *chaotic* or *turbulent* ocean. The reality is that both the atmospheric and oceanic circulation are turbulent. There are such things as large-scale and time-mean currents (and wind systems). These definitely have a role to play, for example in equalizing the meridional temperature contrast of the planet. But the 'macroturbulence' which is so evident in Figure 15, and even more so in the close-up shown in Figure 16, is also there for a reason. As we will look more into later, the large-scale currents are hugely constrained in what they can do on a roating planet like Earth. In fact, the ambient rotation tends to produce east-west currents, and it is really only the presence of continental boundaries that allow for large-scale medirional ocean currents that can bring warm waters poleward and cold waters southward. But, importantly, the ocean macroturbulence is less constrained by Earth's rotation and can therefore also help transport heat poleward. More generally, it helps spread the cold and warm waters away from the mean currents, essentially enlarging the surface area they cover and thus making exchanges with the atmosphere more efficient.

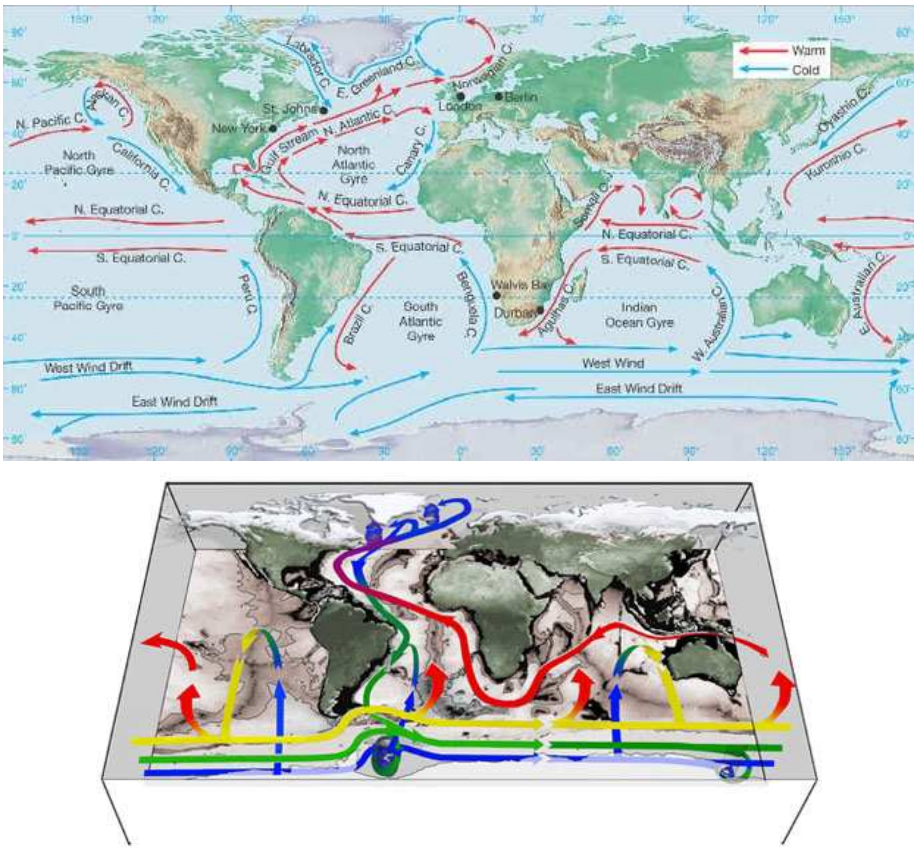

Figure 14: Two different schematic representations of the time-mean large-scale surface currents. (Sources: http://minesto.com; http://scitechdaily.com)

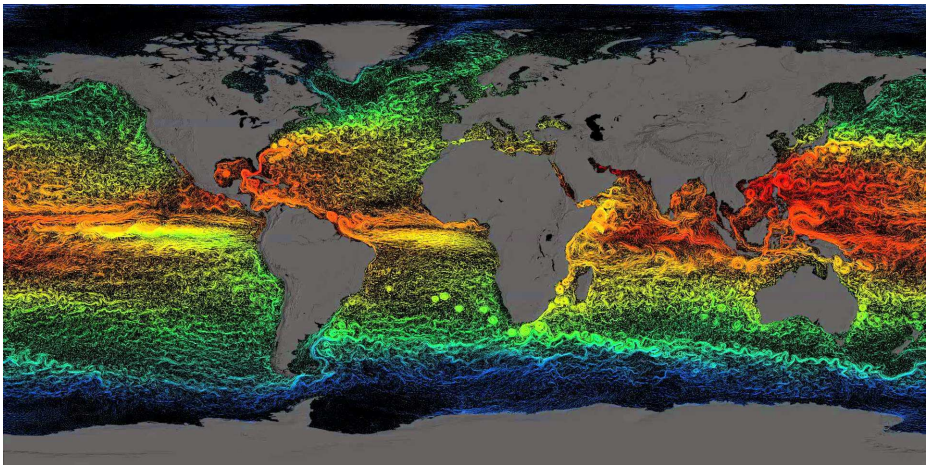

Figure 15: Numerical model simulations of real ocean surface currents. The color gives an indication of current strength, with red indicating high speeds. (Source: https://www.youtube.com/watch?v=sgOgXL4GVwA)

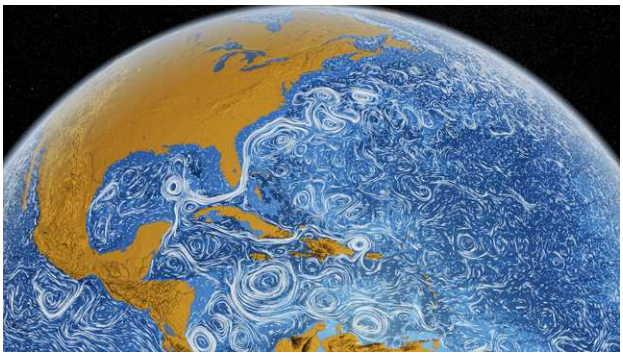

Figure 16: A close-up snapshot of currents off the east coast of North and Central America. (Source: https://svs.gsfc.nasa.gov/10841; also search for 'perpetual ocean' on youtube)

1.3.4 Waves

The purpose of waves in nature is to transmit energy from one place to another. With waves, energy is transmitted through a medium which itself doesn't need to move much. So Earth's oceans (and atmosphere) are full of waves of all possible frequencies and wavelengths. Waves also transmit information; they are nature's way of letting one place of the ocean know what happens somewhere else.

Some waves are easily observed by us, for example the high-frequency *surface gravity waves* driven by winds. They come in the from of what is called *wind sea* which are locally generated or in the form of *swell* which are also wind-driven waves but waves that may have travelled hundreds if not thousands of kilometers away from their generation region—before they eventually break on a beach (Figure 17). We also now know that the periodic rise and fall of the sea surface once or twice a day is due to *tides*, waves waves that are generated by the gravitational forces from the moon and the sun and which travel around the planet endlessly and, to their credit, in an orderly fashion. But there are also waves that normally escape our immediate attention, simply because their frequencies are so low or wavelengths so long that we simply are not able to detect them as we look out over the ocean from the beach. The so-called *planetary Rossby waves* (Figure 18) are perhaps the most peculiar waves found on our planet as they owe their very existence to the rotation of the planet. They are so huge and have such low frequencies (or long periods) that they are really just observable by satellite (Figure 18).

Yet other waves escape our immediate attentcan because they don't travel on the sea surface itself but rather at depth—as so-called *internal waves*. The phenomenon of 'dead waters' was studied at the turn of the last century by a Swedish oceanographer named Vagn Walfrid Ekman. He received reports from Fritjof Nansen that his ship Fram, on which he tried to cross the Arctic Ocean to reach the north pole, had experienced a mysterious drag force when sailing through particularly brackish (very low salinity) waters. Ekman did laboratory experiments that revealed that the drag was real and due to waves forming by the boat disturbing the interface between the fresh and therefore light surface layer and denser water layers underneath (Figure 19). Internal waves radiated energy away from where the boat disturbed the interface, and it was this radiative loss of energy that caused a drag on the boat.

Figure 17: Examples of wind-generated surface gravity waves: (top) locally-generated wind sea and (bottom) remote-generated swell. (Sources: http://pnwcirc.org; Stewart, 2008, Fig. 17.4)

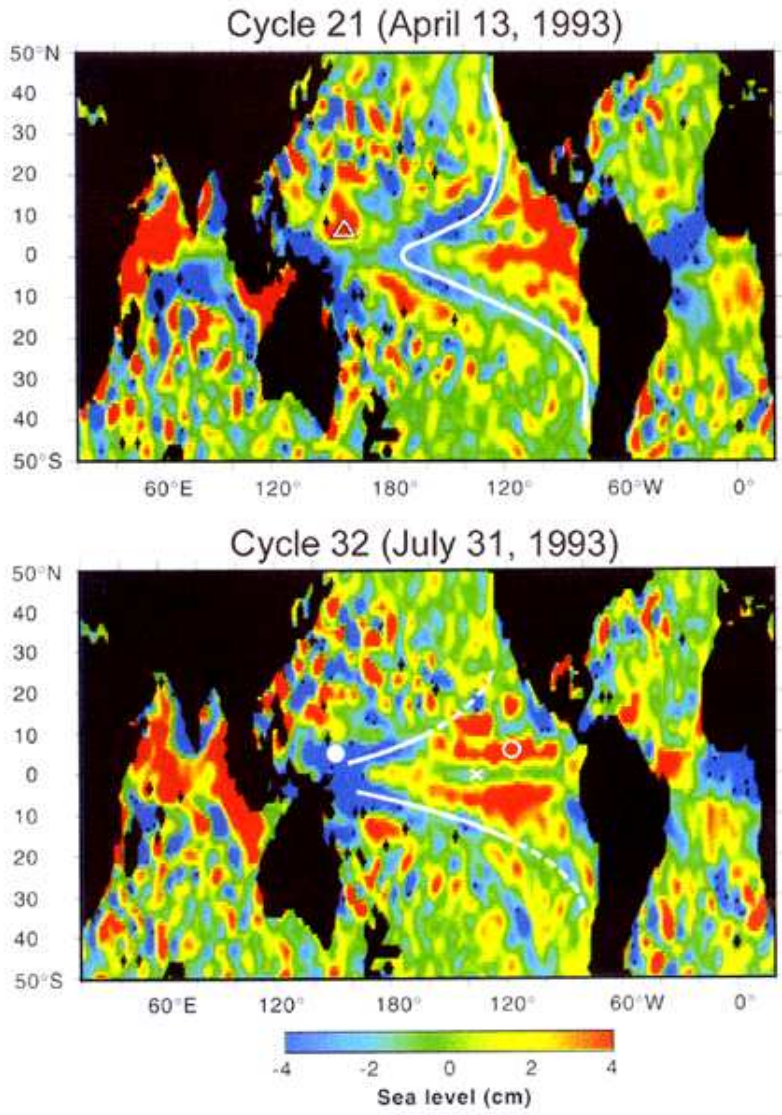

Figure 18: A planetary Rossby wave travelling from east to west at low latitudes, as captured by satellite observations of sea surface height. (Source: http://www-po.coas.oregonstate.edu/research/po/research/rossby_waves/chelton.html)

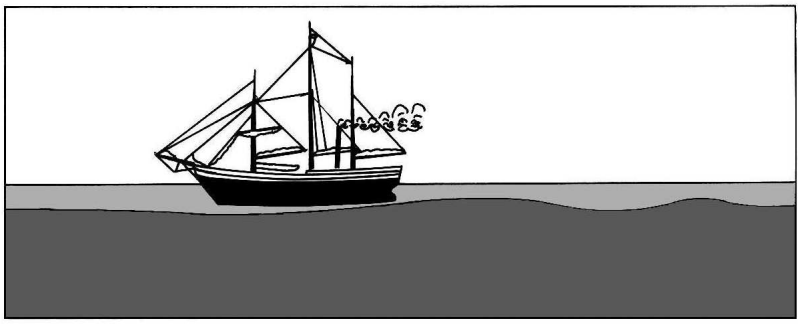

Figure 19: The phenomenon of 'dead waters', internal waves created when a ship sails through a light surface layer overlying denser waters. (Source: Cushman-Roisin and Beckers, 2011, Fig. 1.4[Cushman-Roisin and Beckers(2011)])

2 The stratified ocean

After a brief introduction to some of the main concepts related to the oceans and their circulation, we will now step back and start to better define what we really mean by some of the concepts raised above. What we call *geophysical flows*, i.e. flows pertaining to the environment of a planet like the Earth, have at least two characteristics that distinguish them from flows, say, in blood vessels, in our bath tub or around an airplane wing. These two distinguishing characteristics are that 1) the fluid is density stratified and 2) the flow is influenced by the rotation of the planet itself. We will come back to the effect of rotation later and start here with the concept of density stratification and how it is related to both potential energy of a water column and, ultimately, to ocean flows themselves. We will see that much of the large-scale flow on the planet can be understood as a result of an uneven density field.

2.1 Static stability

A water column which is made to be lighter at depth and denser above, for example by warming at depth and cooling above, becomes *statically unstable*. The most familiar example to most is water warmed in a pot on the stove top (Figure 20). Under such conditions, if a dense *fluid parcel*[1] from near the surface gets a tiny kick downwards, it will soon find itself surrouned by ligther fluid parcels than before and will hence continue to sink since it is denser than these other parcels. And conversely with a light fluid parcel from depth which gets a small kick upwards. The whole fluid column will spontaneously and quickly *overturn* so that light water rises up and dense water sinks down. The vertical overturning motion is called *convection* and the end result is a water column with lower center of mass and lower gravitational potential energy (we'll get back to this below).

After this convective adjustment we can repeat the same thought experiment. If a fluid parcel from near the surface is now displaced downwards it will find itself lighter than its new surroundings and, as a consequence, it will rise back towards its original position. Instead of growing, vertical disturbances of any kind are damped out. The fluid is now *statically stable*.

In fact, when a displaced parcel in a stably-stratified fluid returns back to its original position it will typically overshoot that position slightly. Then, since the

[1] What we call a fluid 'parcel' is a bunch of molecules of gas or liquid that we think of as flowing together as a unit. A fluid parcel can expand, contract and deform, but it has a fixed mass.

Figure 20: Convection in the kitchen.

overshooting brings it in contact with waters of still different densities, it will turn and flow the other way again. The result is a vertical oscillation, and the frequency of this oscillation increases as the vertical density gradient, the *stratification*, increases. This *buoyancy frequency* is

$$N = \left(-\frac{g}{\rho_0} \frac{\partial \overline{\rho}}{\partial z} \right)^{1/2},$$

where g is the gravitational accelleration, ρ_0 is some reference density and $\partial\overline{\rho}/\partial z$ is the background (unperturbed) vertical density gradient. For stably-stratified conditions $\partial\overline{\rho}/\partial z < 0$ and the buoyancy frequency is a positive real number—the natural frequency of oscillation if the density stratification is disturbed. If instead $\partial\overline{\rho}/\partial z > 0$, i.e. if we have dense fluid on top of light fluid, N is becomes imaginary (from taking the square root of a negative number). Mathematically, this indicates that disturbances don't oscillate but will in fact grow. We have an instability which grows into the convective overturning motion. So some of the vertical motions we can observe in the upper parts of the oceans are driven by such unstable vertical density stratifications—set up, for example, by cooling of the ocean by the atmosphere.

2.2 Stratification and potential energy

The *gravitational potential energy*[2] of an individual water parcel of mass $m = \rho dV$, where ρ is its density and δV its volume, is

$$\begin{aligned} pe &= mgz \\ &= \rho g z \delta V, \end{aligned}$$

Note that z is the height above some arbitrary reference level (let's put it to the bottom of the ocean). Then the total potential energy (PE), summed over a bunch of water parcels, is

$$PE = \sum_i \rho_i g z_i \delta V_i,$$

or, in the continuous limit of infinitesimal volume elements,

$$PE = \iiint \rho g z \, dV,$$

where the integral is over all space, say the entire volume of the ocean. So the geographic distribution of the height of the density field determines the PE. The higher up dense waters are and the lower down light waters are, the higher is the PE.

2.2.1 Potential energy of a stratified water column

Figure 21 shows three water columns that all have the same height and the same total mass. But the mass is distributed differently with height, i.e. the water density is distributed differently with height, in the three cases. The gravitational potential energy of each water column (per unit horizontal area) is

$$PE = \int \rho g z \, dz,$$

where the integral is now taken over the height of the colum. It can be shown (try it out!) that the 'well-mixed' column in the figure, where the density is the same from top to bottom, $\rho = \overline{\rho}$, has the highest PE. The column where density changes

[2] A water parcel also has *internal potential energy* caused by the pressure field, but we ignore this here.

linearly with height, from $\rho_1 = \overline{\rho} - \Delta\rho$ at the top to $\rho_2 = \overline{\rho} + \Delta\rho$ at the bottom, has a somewhat lower PE. In other words, the gravitational center of mass of this column, defined by

$$\bar{z} = \frac{\int \rho z \, dz}{\int \rho \, dz},$$

is lower than in the well-mixed column. Finally, if the total mass is separated into two well-mixed slabs of equal thickness and with densities ρ_1 and ρ_2, the center of mass and PE are even lower. The stronger the stratification, i.e. the stronger the density jump, the lower the PE.

2.2.2 Energetics of a slanted density stratification

We have seen that a water column with dense water lying on top of light water is statically-unstable and will overturn until it is again statically stable. In the process potential energy is converted into kinetic energy (the convective motion). When all settles down in the end the kinetic energy has been lost into heat (created by friction). So the end state is one of a water column with lower PE than it had to begin with. Conversely, if one starts with a statically stable water column in which the density field is also completely flat, then reorganizing water parcels in the vertical, i.e. replacing lighter parcels near the top with heavier parcels from deeper down, will raise the PE. In other words, stirring or mixing the water column vertically is energetially costly, requiring an external source of mechanical energy that can do work against gravity. In contrast, horizontal stirring of water parcels in such a flat density field does not raise the PE and is therefore easily done—the only work done is that fighting friction.

When the density field is *slanted* or *tilted*, as illustrated in Figure 22, the situation becomes a little bit more complex. Moving water parcels around can either raise PE (requiring an external mechanical energy source to do so) or it can release PE and set off motion.

Let's consider the three fluid parcels A, B and C shown in Figure 22. Parcel C is the lightest one, parcel A has an intermediate density while parcel B is the heaviest one. Exchanging parcels A and C means lifting a relatively dense parcel (A) while lowering a relatively light parcel (C). The end result is a heightening of the center of mass and an increased PE. Work must be done on the fluid to acheive this. If instead exchanging parcels A and B, a parcel with relatively low density (A) is lifted and a parcel with relatively high density is lowered. The end result is a lowering of the center of mass and a decreased PE. Finally, if parcel A was exchanged with another parcel having the same density, i.e. being situated in the

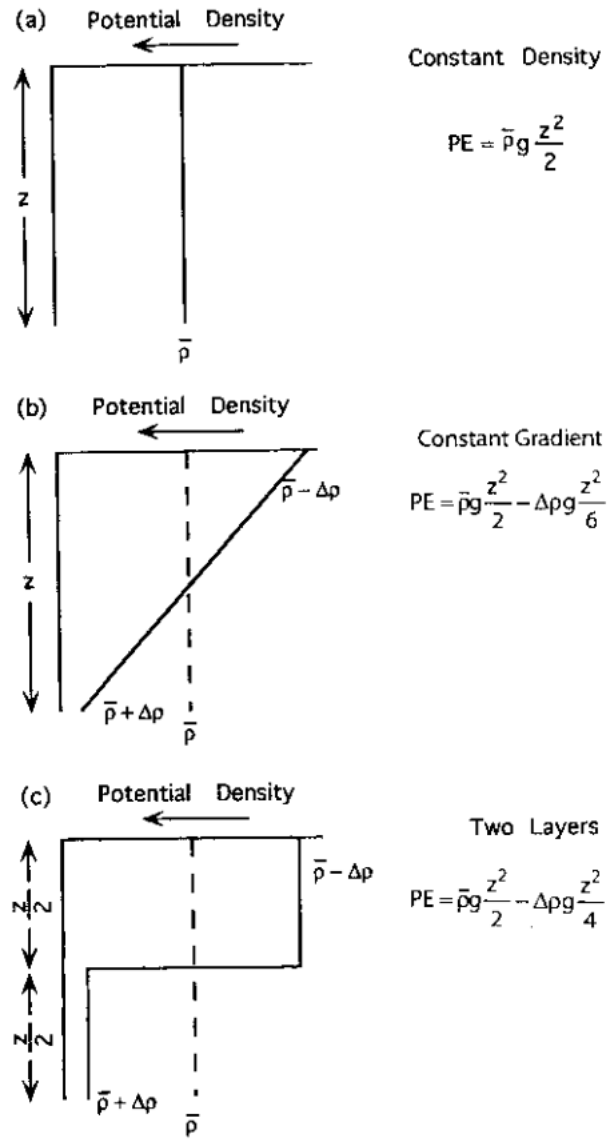

Figure 21: The gravitational potential energy for three different density stratifications. (Source: Knauss, 2005, Fig. 2.6[Knauss(2005)])

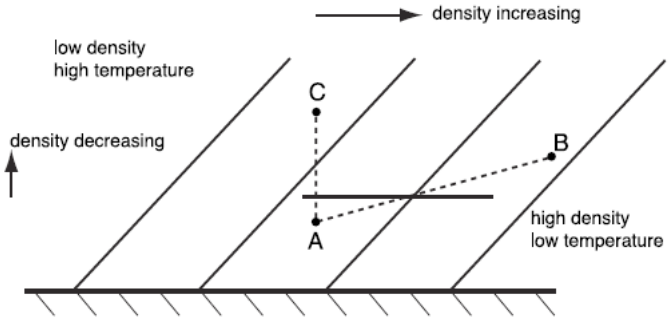

Figure 22: The exchange of fluid parcels, either between points A and C or between points A and B. (Source: Vallis, 2006, Fig. 6.9[Vallis(2006)])

same layer in the figure (not shown in the figure), the net effect would be a zero change in PE.

As seen, the outcome, i.e. whether PE is increased or decreased, depends on the angle of the movement relative to the angle of tilt of the density field. If the angle of the exchange (relative to the horizontal) is bigger than the angle of the stratification, then PE is increased. If, however, the angle is smaller than that of the stratification (but still larger than zero with respect to the horizontal) PE is released and motion can be induced. This type of motion is sometimes called *slantwise convection*.

A little extra:

Let's quantify the change in PE when parcels A and C are interchanged. It is the PE after the exchange minus the PE before the exchange. If ρ_A and ρ_C are the densities of the two parcels (let's assume that the volumes are the same) and z_A and z_C are the heights before the exchange, then the change in potential energy (per unit volume) is

$$\begin{aligned} \Delta PE &= \overbrace{(g\rho_A z_C + g\rho_C z_A)}^{\text{after}} - \overbrace{(g\rho_A z_A + g\rho_C z_C)}^{\text{before}} \\ &= -g\left[(\rho_C - \rho_A)(z_C - z_A)\right] \\ &= -g\Delta\rho\Delta z. \end{aligned}$$

For small displacements Δz we can write

$$\Delta\rho = \frac{\partial\overline{\rho}}{\partial z}\Delta z,$$

where $\partial\overline{\rho}/\partial z$ is the background vertical density gradient (the density gradient before we start exchanging fluid parcels). So we get

$$\Delta PE = -g\frac{\partial\overline{\rho}}{\partial z}(\Delta z)^2,$$

and for $\Delta z > 0$ (since point C is higher than point A initially) and $\partial\overline{\rho}/\partial z < 0$ (since the vertically density stratification is stable) we get $\Delta PE > 0$. So potential energy increases, as we argued informally above.

When considering the exchange between particles A and B, we also have to account for density changes in the horizontal direction. For small displacements Δx and Δz, the change in PE becomes

$$\Delta PE = -g\left(\frac{\partial\overline{\rho}}{\partial x}\Delta x + \frac{\partial\overline{\rho}}{\partial z}\Delta z\right)\Delta z.$$

If we now introduce the *slope* of the exchange path

$$s_{ex} = \frac{\Delta z}{\Delta x}$$

and the slope of the background density field

$$s_\rho = -\frac{\partial\overline{\rho}/\partial x}{\partial\overline{\rho}/\partial z},$$

the expression becomes

$$\begin{aligned}\Delta PE &= -g\left(-s_\rho\frac{\partial\overline{\rho}}{\partial z}\Delta x + \frac{\partial\overline{\rho}}{\partial z}s_{ex}\Delta x\right)s_{ex}\Delta x \\ &= -g\frac{\partial\overline{\rho}}{\partial z}\left(\Delta x\right)^2\left(s_{ex} - s_\rho\right)s_{ex}.\end{aligned}$$

For a stable density stratification ($\partial\overline{\rho}/\partial z < 0$) the sign of ΔPE depends on the size of s_{ex} relative to that of s_ρ. For $s_{ex} > s_\rho$ we get $\Delta PE > 0$,

an increase of PE. As with the exchanges between A and C, work must be done on the fluid to acheive this. When $s_{ex} = s_\rho$ there is no change in PE, as discussed above. Fluid parcels can easily (at low energetical cost) travel along layers of constant density. Finally, and this is the more interesting situation, for $0 < s_{ex} < s_\rho$ we get $\Delta PE < 0$, i.e. a lowering of PE. This exchange is hydrodynamically unstable (parcel A will continue to rise and parcel B will continue to sink), so the process will speed up.

2.2.3 Available potential energy—APE

So we have seen that it is possible to release PE from a tilted density stratification under certain types of exchange of fluid parcels. Let's look again at this again, but now in a bulk or integral sense. Consider the two configurations of a two-layer stratification in Figure 23. The two cases have exactly the same volumes of fluid with density ρ_1 and ρ_2. But a relatively straightforward calculation will show that the configuration where the interface separating the two density layers is flat has a lower total gravitational potential energy than the configuration where the interface is tilted. Getting from the highest to the lowest PE configuration requires exchanges of water parcels within each density layer, *exchanges that must obey the relationship between slopes that we discussed above*.

The kinetic energy of these flows is equal to the PE which is released by the flattening. So if friction is negligible

$$\Delta KE + \Delta PE = 0.$$

This potential energy that can be extracted and converted to kinetic energy by slantwise convection is termed *available potential energy* (APE). So, to use the terminology introduced above, slantwise convection releases APE.

Most of the large-scale motions in the atmosphere and also much of the motions in the ocean are forms of slantwise convection. Just think about how the uneven warming from the sun creates large-scale latitudinal density gradients. Vertical convection ensures that the vertical density gradient is almost always stable and, as a result, we are left with a tilted density field—which nature tries to flatten out to lower APE. The end result, if this process gets to act alone, is a flat density field. Such a flat field has, by definition, zero APE.

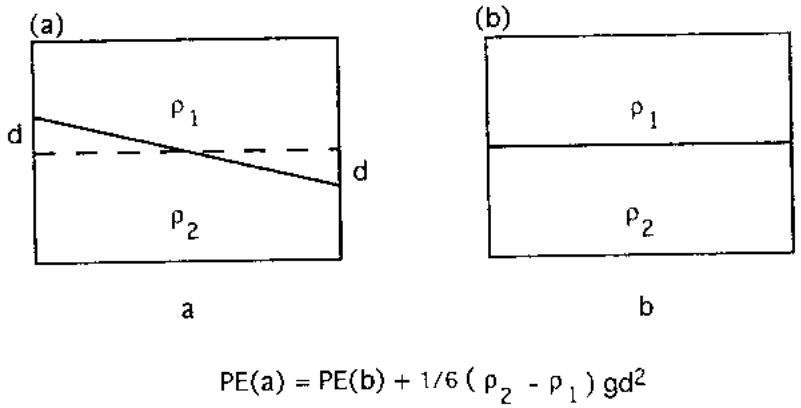

Figure 23: The concept of Available Potential Energy (APE) as the difference between the arrengement of the two fluid layers in panel a and b. (Source: Knauss, 2005, Fig. 2.7)

2.3 The oceanic equation of state

Now that we've been introduced to the importance of density variations in influencing the ocean energetics and even the ocean circulation, we will finally move on to density itself. In the discussions above we have repeatedly mentioned how warm waters are light waters. But it's now time to get more precise and acknowledge that density of sea water is a complex function of temperature, salinity and pressure. So we write

$$\rho = \rho(T, S, p).$$

Roughly speaking, density decreases with rising temperature while it increases with rising salinity and pressure. But unlike for the atmosphere, where the ideal gas law can often be used, we only have imperically-derived polynomial expressions for the oceanic equation of state. So water density as a function of temperature, salinity and pressure is measured carefully in laboratory experiments and the data are then fit in some least-squares sense to polynomial functions. The internationally agreed definition have changed over time. Currently it is termed the *Thermodynamic Equation Of Seawater - 2010 (TEOS-10).*

As it turns out, ocean water has a density that is always higher than $1000\,\mathrm{kg\,m^{-3}}$. So to save a little bit of space and time oceanographers typically subtract one thou-

sand and report ocean densities in terms of the *density anomaly*

$$\sigma(T, S, p) = \rho(T, S, p) - 1000\,\mathrm{kg\,m^{-3}}.$$

However, one often just says or writes 'density' when in fact one means density anomaly.

When studying the distribution of ocean density to assess whether the water column is statically stable or not, special care must be taken. When doing the thought experiments mentioned above of moving a water parcel vertically to see whether it becomes denser or lighter than the new surrounding parcels, we can assume that the parcel keeps its original salinity and, approximately, its original temperature (for *adiabatic* motion with no heat exchange). But it will experience a different pressure at the new level, and the density adjustment to this new pressure—it will contract, increasing its density, if moved downward—will be instantaneous, moving with the speed of sound. The point is that all other water parcels that our parcel encounters at the new level will also have adjusted to the pressure at that level. So the direct pressure effect on density will be the same for all parcels.[3]

If we want to compare the density of two water parcels, to see which is lighter than the other, we therefore need to compare them as if they were situated at the same pressure. In practice, what this means when we wish to study the static stability of an entire water column (from observations of salinity and temperature as functions of pressure), is that we need to compare the densities as if all parcels were situated at the same pressure. We simply ignore the pressure effect on density and ask "what density would all these parcels have if they were brought to some common pressure, say to the surface?". So instead of comparing $\sigma(T, S, p)$ one would compare $\sigma(T, S, p_0)$ where the *reference pressure* p_0 is fixed. The surface pressure is the most common reference pressure, and the density anomaly referenced to the surface has been termed "sigma-tee", $\sigma_t = \sigma(T, S, 0)$.

Slight additional complications arise since the temperature of a water parcel is itself a function of pressure. The first law of thermodynamics dictates that an increase in pressure in an adiabatic process (no heat flow) will do work on a water parcel and therefore cause an increase in the internal energy of the parcel. The

[3]In an ocean (or bathtub) where all the water has the same salinity and temperature, the water at the bottom will still be denser than the water on the top due to the higher pressure at depth. But if we moved a water parcel from the top down, it would not bounce up again. It would instead (nearly) instantaneously adjust its density to the new pressure to attain the same density as the parcels further down. We would not see convection, neither waves, but steady motion of our parcel (given by the initial push we gave it) until it got slowed down by friction.

internal energy is proportional to the temperature (a measure of the intensity of random movement of molecules) and, so, the temperature will increase. Hence, if a water parcel originally found near the sea surface is moved adiabatically down to greater depts, where the pressure is also greater, its temperature will rise. One may therefore encounter situations where the *in situ* temperature (the temperature measured by a probe lowered down through the water column) increases with depth—just from this pressure effect. One therefore gets the impression that density decreases with depth and that the water column is unstable. Not necessarily so! The adjustment to a change in pressure is instantaneous (just as the density adjustment itself is), so if a water parcel is displaced vertically it imediately adjusts its temperature to the pressure at the new depth—and so have all the other water parcels at that same depth done. To assess the real static stability of a water column, or part of a water column, one has to remove the pressure effect on temperature and density.

To remove this additional pressure effect we introduce *potential temperature*[4] $\theta(T, p, p_0)$ which is the temperature the water would have if moved adiabatically from pressure p to reference pressure p_0. Then the corresponding *potential density* and potential density anomaly is a function of potential temperature, salinity and the (fixed) reference pressure, i.e.

$$\sigma_\theta = \sigma(\theta, S, p_0). \tag{1}$$

So plotting the potential temperature and potential density of a set of measurements is like plotting the temperature and pressures the various parcels would have if they were all brought (adiabatically) to the same pressure level where they could meet and compare temperature and density. Only then, with the pressure effect removed, can we assess the true static stability properties of the water column. An example is shown in Figure 24. We see, from the data plotted, that the *in situ* temperature increases at great depths and that σ_t decreases accordingly, giving the impression of an unstable water column at depth. Plotting instead potential temperature and potential density—using potential temperature—shows that the water column is in fact stable.

Normally one choses the sea surface as reference pressure (and potential density is then termed σ_0 or simply σ_θ), but it is also possible to use other reference pressures, e.g. 1000 dbar (approximately 1000 m depth[5]), 2000 dbar or 5000 dbar

[4]The newest thermodynamic equation of sea water, TEOS-10, uses the term *conservative temperature* instead.

[5]The unit of pressure commonly used in oceanography is the decibar ($1\,\mathrm{dbar} = 0.1\,\mathrm{bar} =$

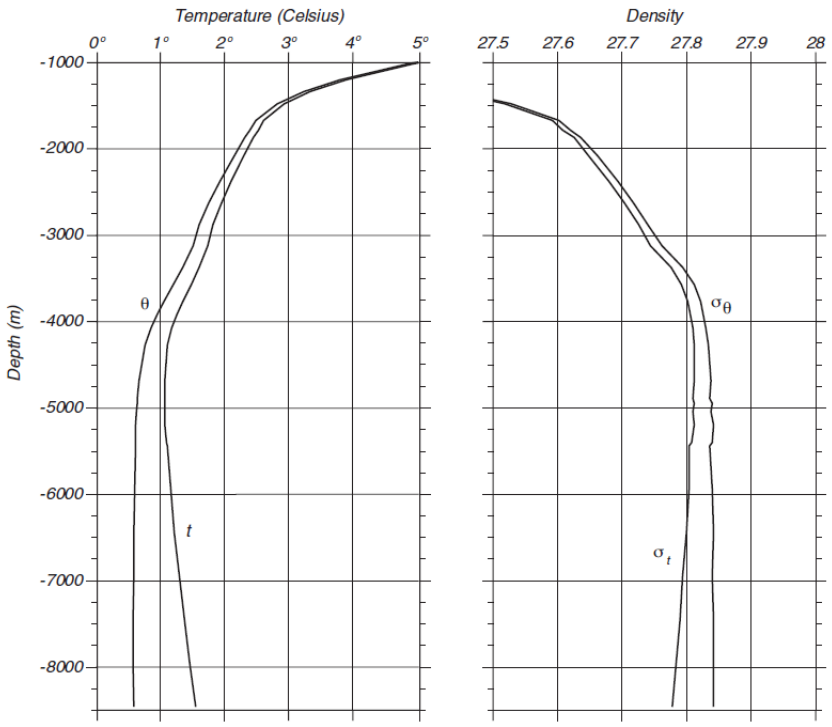

Figure 24: Deep temperature profiles (left) from the Pacific Ocean, showing both in situ temperature (t) and potential temperature (θ). Also shown (right) are profiles of potential density anomaly, using in situ temperature (σ_t) and potential temperature (σ_θ). (Source: Stewart, 2008, Fig. 6.9)

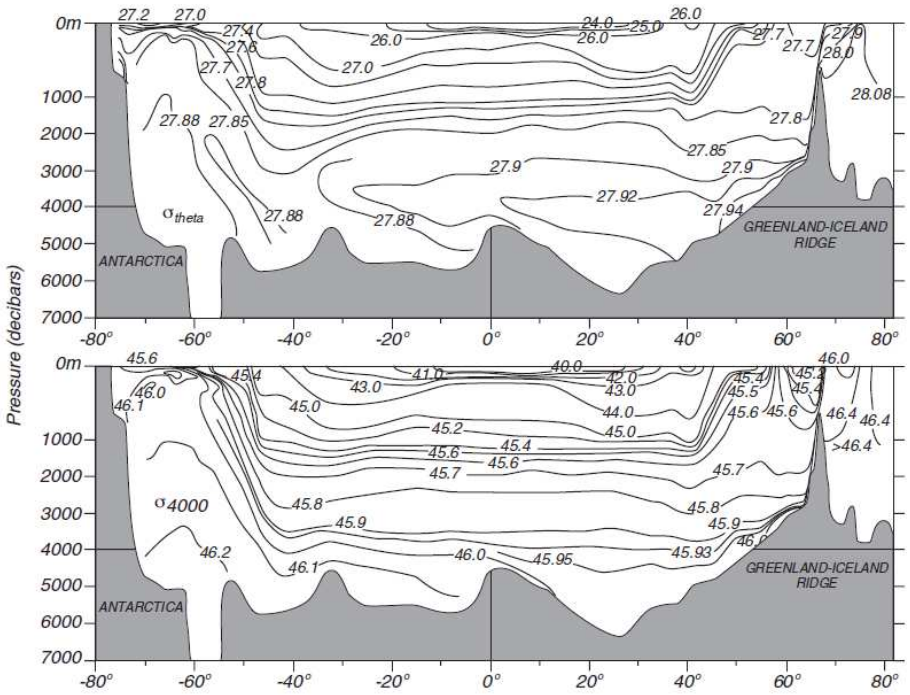

Figure 25: Potential density in the western Atlantic, referenced to two different pressures, 0 dbar (top) and 4000 dbar (bottom). (Source: Stewart, 2008, Fig. 6.10)

$(\sigma_1,\, \sigma_2,\, \sigma_5)$, etc. In fact, when assessing the stability of the water column at great depths, a shallow reference pressure like 0 dbar may give wrong answers. An example of this is shown in Figure 25. Plotting potentidal density referenced to the sea surface, 0 dbar, gives the impression that waters below about 3000–4000 m are statically unstable. Plotting the same data but now referenced to 4000 dbar shows that deep waters are indeed stable.

As mentioned above, the equation of state is complicated. But for some applications where only small density deviations are of intereste, e.g. in the study of internal waves where the background density stratification moves up and down with the waves, a a linear equation of state can be used for potential density. We then write

$$\rho = \rho_0 \left[1 - \alpha \left(T - T_0\right) + \beta \left(S - S_0\right)\right],$$

where ρ_0 is a reference density and α and β are the *thermal expansion coefficient*

$10^4\mathrm{Pa}$). As it turns out, the pressure increases about 1 dbar for each additional meter deeper one goes down into the water column. So the pressure at 1000 m depth is around 1000 dbars.

and the *haline contration coefficient*, respectively. Both are positive numbers that themselves depend on temperature, salinity and pressure. This linear fit to the full equation of state shows us what we expect, namely that water density decreases with temperature and increases with salinity.

2.4 Water types and T-S diagrams

Salinity and temperature (or, more frequently, potential temperature) observations are sometimes plotted in so-called *T-S diagrams*, with salinity on the x-axis and temperature on the y-axis (e.g. Figure 26). So these are scatter plots of temperature vs. salinity. Such a plot can often be used to identify *water masses* in the data set. A water mass is defined by a rather narrow range of temperature and salinity that was set when the water in question was exposed to the atmosphere before it sunk deeper into the ocean. So the properties of water masses are primarily set by air-sea fluxes and, as one can imagine, the T-S signature of water masses formed at high latitudes is different from that of water masses formed at lower latitudes. Waters modified by heat and freshwater fluxes at the sea surface in the Arctic certainly take on different T-S properties than waters modified by air-sea fluxes in the Mediterranean Sea.

So water masses often show up as extrema in T-S diagrams. The gradual mixing that occurs in the ocean interior then connects these extrema (as shown in Figure 26). When isolines of (potential) density are also added to the T-S diagram, one can also see which water masses are denser than others and whether the mixing between water masses is primarily *isopycnal* (taking place while conserving potential density) or *diapycnal* (associated with a density change). These properties of the T-S relation can then be tied to discussions on the energetics of mixing, as discussed in the sections above.

Note also that if one has obtained a lot of T-S observations and then keep track of the volume of different T-S classes, one can make so-called *T-S-V plots* which show the relative volume of various T-S classes. An example, based on a global temperature and salinity dataset, is shown in Figure 27. The plot illustrates that the range of T-S values in the world oceans is relatively modest and that each of the big oceans have their own T-S signature. The Pacific Ocean dominates in terms of volume, simply because it is the biggest ocean of them all. But we also see that the Atlantic and Southern/Indian oceans also have their distinct positions in T-S space. This indicates that air-sea fluxes as well as internal mixing processes in the different oceans are distinct.

Note, incidentally, that the Atlantic Ocean water masses are saltier than those

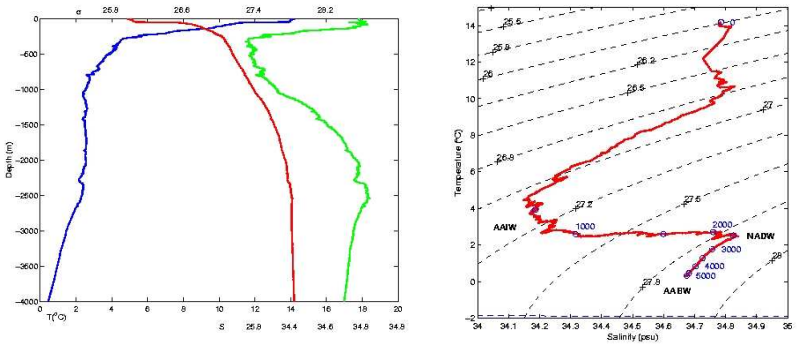

Figure 26: Temperature and salinity observations from a hydrographic profile plotted (left) as a function of depth and (right) in a so-called T-S diagram. Whereas potential density is also plotted as a function fo depth to the left it is instead contoured in the T-S diagram. (Source: http://www.soes.soton.ac.uk/teaching/courses/oa631/hydro.html)

in the Pacific Ocean[6]. Because of the higher salt content in the Atlantic, there is a higher production of dense waters there. In fact, the large-scale vertical overturning circulation through the Atlantic Ocean, the Atlantic Meridional Overturning Circulation (AMOSC) is much stronger than the corresponding circulation in the Pacific Ocean (PMOC).

[6]Why this is so is an active topic of research in the climate community. It may simply be because the Pacific Ocean is larger and can therefore get more dilluted by rain

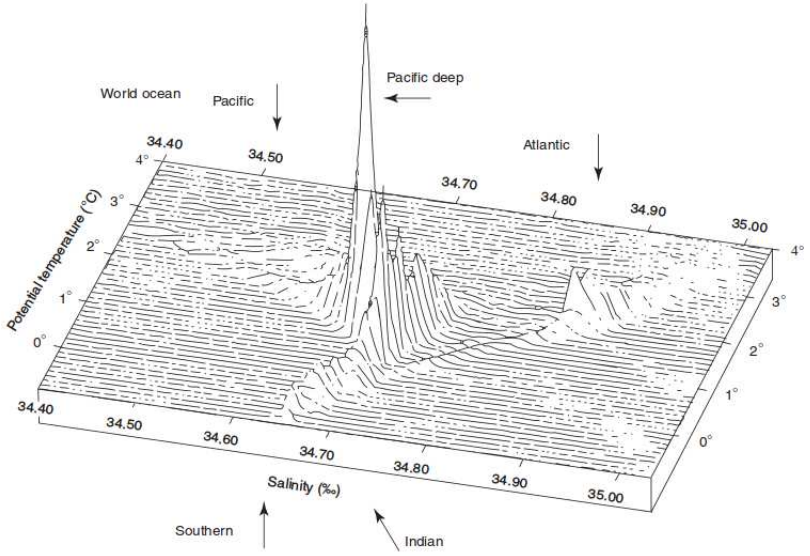

Figure 27: A T-S-V plot, illustraing the volume of various T-S classes found in the world oceans. (Source: Stewart, 2008, Fig. 6.1)

3 Fluxes through the sea surface

The ocean circulation is driven by heat and freshwater fluxes through the sea surface and by the wind stress. An uneven distribution of heat and freshwater fluxes sets up temperature and salinity gradients, i.e. density gradients that can drive flows, as discussed in the previous section. And the wind stress sets up a drag on the ocean surface that can drive the circulation directly. What is really going on is a little more complicated, involving horizontal pressure gradients, as we will see later. But we now first have a look at the various fluxes themselves.

3.1 Heat and freshwater fluxes

The total or net heat flux into the ocean through the ocean surface is the sum of four contributions: shortwave radiation from the sun (always into the ocean, so positive), longwave radiation (can go both ways), a 'latent' heat loss when water evaporates (always out of the ocean, so negative) and a 'sensible' heat flux due to temperature differences between ocean and atmosphere (can go both ways). So we write

$$\sum Q = R_{sw} + R_{lw} + Q_l + Q_s. \tag{2}$$

The two first fluxes are purely radiative while the two last involve turbulent fluid motions in *boundary layers* at the bottom of the atmosphere and the top of the ocean.

The global-mean ocean temperature doesn't change much from year to year. So, to a first approximation there is a global and yearly mean balance between incomming short wave radiation from the sun and a heat loss from the other terms, i.e.

$$< \overline{R_{sw}} >= - \left(< \overline{R_{lw}} > + < \overline{Q_l} > + < \overline{Q_s} > \right),$$

where $< \cdot >$indicates the spatial (global) average whereas $\bar{\cdot}$ indicates the time (yearly) average. But, of course, there need not be a balance at any one location, not even when averaged over a year. In the yearly mean, low latitudes receive more heat by shortwave radiation than what is lost via local vertical fluxes, and the situation is opposite at high latitudes. This is where the ocean circulation and its poleward heat transport comes into play.

Shortwave radiation (and a bit of marine optics) Shortwave radiation is the radiation emitted by the sun. The incoming solar radiation that hits Earth's surface a given latitude varies during the year because of the inclination of the Earth's

axis of rotation, as shown in Figure 28. The incoming radiation in the southern hemisphere during austral summer is slightly stronger than that in the northern hemisphere during boreal summer due to the eccentricity of Earth's orbit around the sun. Other asymmetries between the radiation that reaches the ocean surface is largely due to asymmetries in the distribution of land masses and atmospheric absorption.

The sun emits electromagnetic radiation as a black body[7] at a temperature of 5800–5900 K. As a result, and according to Plank's law for such black bodies (describing the energy emitted as a function of frequency or wavelength), most of this radiated energy is in what we call the 'visible band', having wavelengths of 400–700 nm. Absorption and scattering in Earth's atmosphere reduces the energy density that reaches the ocean and land surface, but the *irradiance spectrum*, or downward energy flux as a function of wavelength (having units of $W\, m^{-2}\, m^{-1)}$, still resembles that of the original black body (Figure 29). Of course, the shortwave flux intensity that eventually hits the ocean in a particular position on one given day is a function of the local cloud cover. A dense cloud cover increases both atmospheric absorption and scattering, leaving less energy to reach the surface.

The shortwave radiation that penetrates through the sea surface is of course also scattered and absorbed by the molecules in the sea water. So the downward irradiance is attenuated with depth. The absorption is in fact much more severe in water than in air and, as shown in Figure 30, at 100 m depth there is not much downward energy flux left.

The irradiance Γ decays approximately as

$$\frac{\partial \Gamma}{\partial z} = -\varepsilon \Gamma,$$

where ε is called the *attenuation coefficient* (having units of m^{-1}). So for a constant ε the irradiance decays exponentially with depth, and the relationship between the irradiance at depths z_1 and z_2 becomes

$$\Gamma(z_2) = \Gamma(z_1) e^{-\varepsilon(z_2 - z_1)}.$$

Both absorption and scattering are wavelength-dependent. So the attenuation coefficient ε is also wavelength-dependent. For clear sea water, water having very little particulate matter in it, the absorption has a minimum around 450 nm, in the

[7] A black body, precicely defined, is a body which absorbs absolutely all radiation it receives and which then emits radiation out again according to its own temperature.

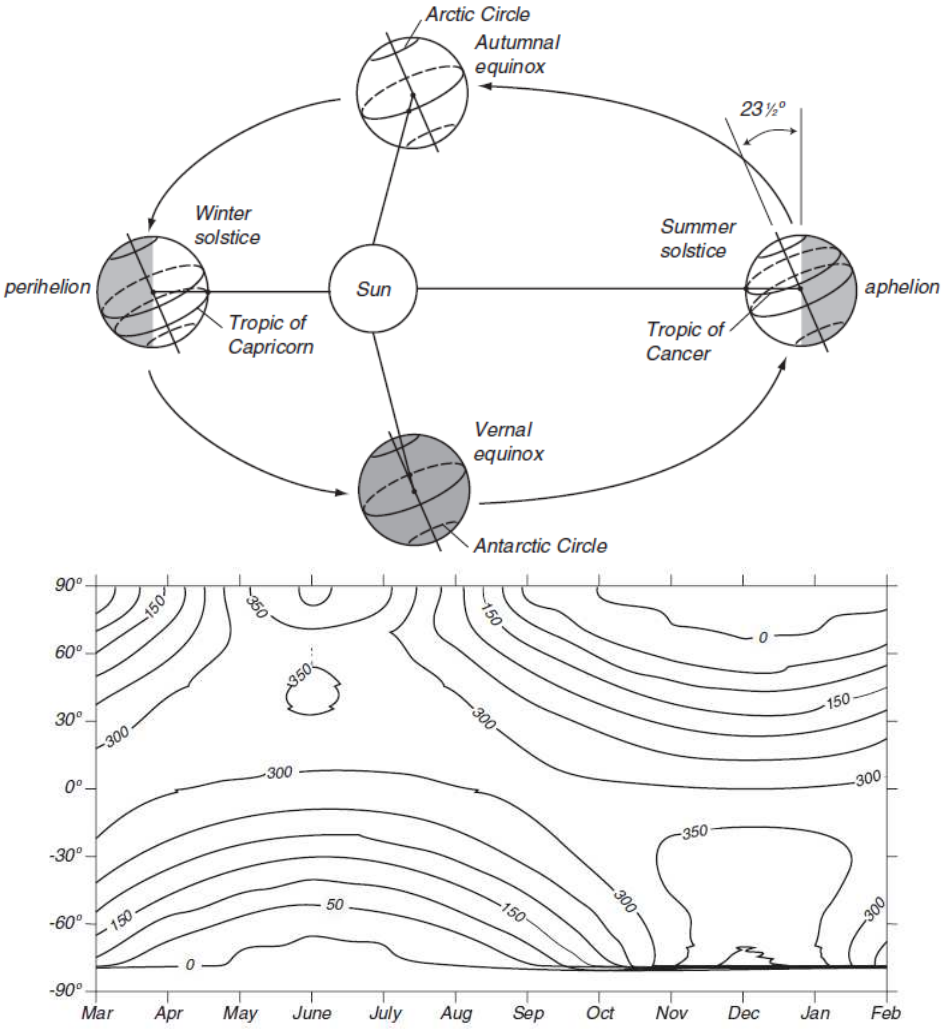

Figure 28: The seasonal variation in incomming solar radiation due to Earth's inclined rotation axis relative to the orbital plane around the sun. The values contoured in the lower panel are the downward energy flux (having units $\mathrm{W\,m^{-2}}$) into the ocean as a function of month of the year and latitude (Source: Stewart, 2008, Figs. 4.1 and 5.3)

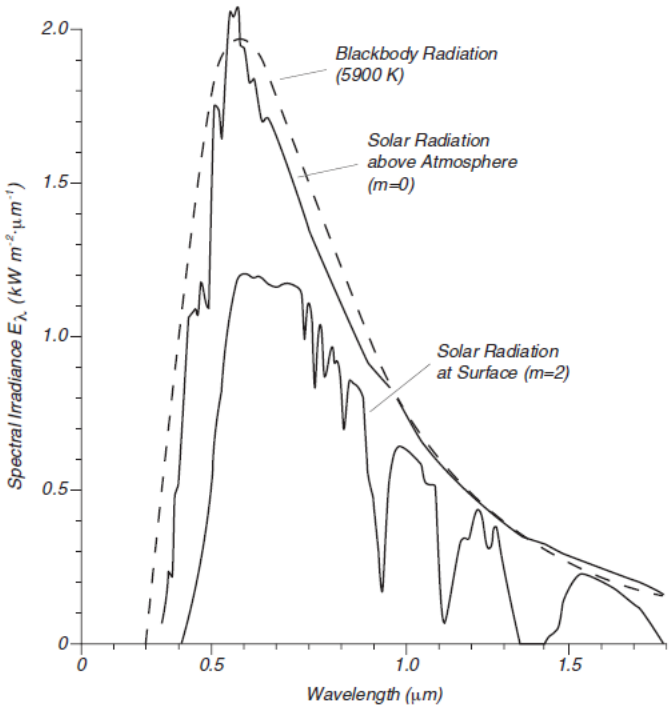

Figure 29: The irradiance spectrum at the top of the atmosphere and at Earth's surface. The theoretical black body spectrum, given a solar temperature of 5900 K is also shown. (Source: Stewart, 2008, Fig. 5.2)

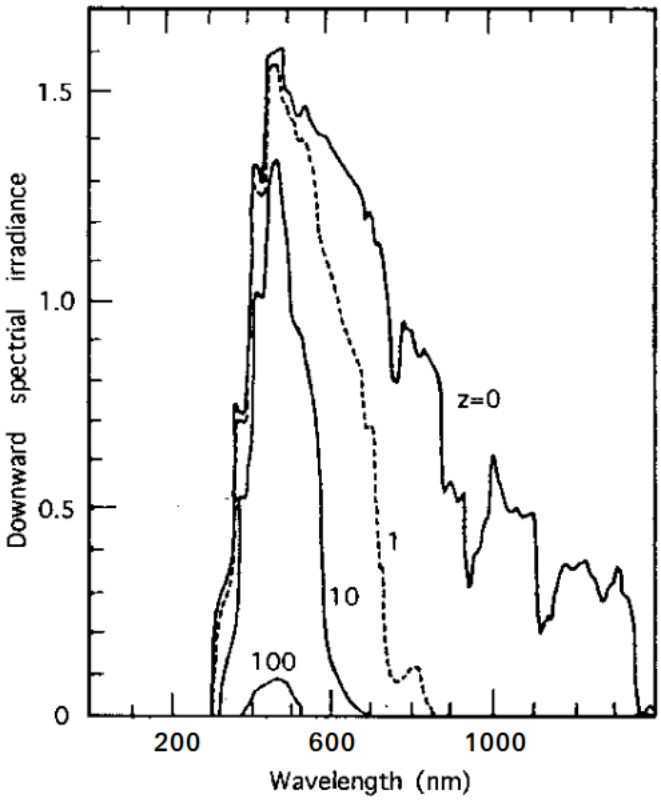

Figure 30: Observations of the irradiance as a function of wavelength at 0, 1, 10 and 100 m in clear sea water. (Source: Knauss, 2005, Fig. 12.13)

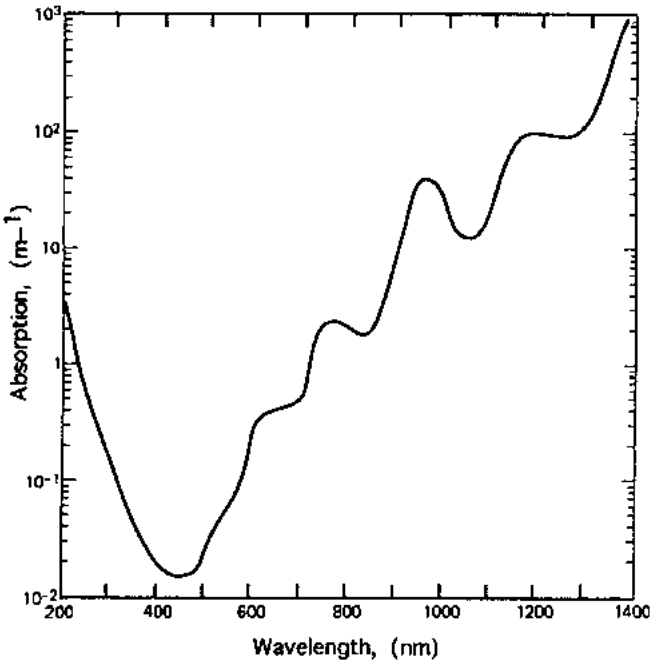

Figure 31: The absorption, measured by the attenuation coefficient ε, as a function of wavelength for clear ocean water. (Source: Knauss, 2005, Fig. 12.11)

blue range of the visible spectrum, as illustrated in Figure 31. This is why the only sign of downward irradiance reaching as far down as 100 m in Figure 30 is found at those wavelengths. It is also why is why clear ocean water looks blue to us looking at it from land. What we observe is light that has been scattered back to us. All wavelengths are scattered back towards the surface, but the blue light is what 'survives' without being absorbed.

Finally it should be mentioned that particles in the water, both phytoplankton, dead organic material and suspended sediments, absorbs short wavelengths (like blue) more efficiently than long wavelengths. So in waters full of organic material or full of sediments, like what is typically found in the coastal zone, the water color we observe moves from blue towards green and even yellow-brown. This is illustrated in Figure 32 which shows the spectral intensity of backscattered light from the sea surface as a function of chlorophyll (and thus of phytoplankton concentration). As the chlorophyll concentration increases, the backscatter from short wavelengths is drastically damped (absorbed) and the peak backscatter moves towards longer wavelengths. An increased particle concentration also causes more

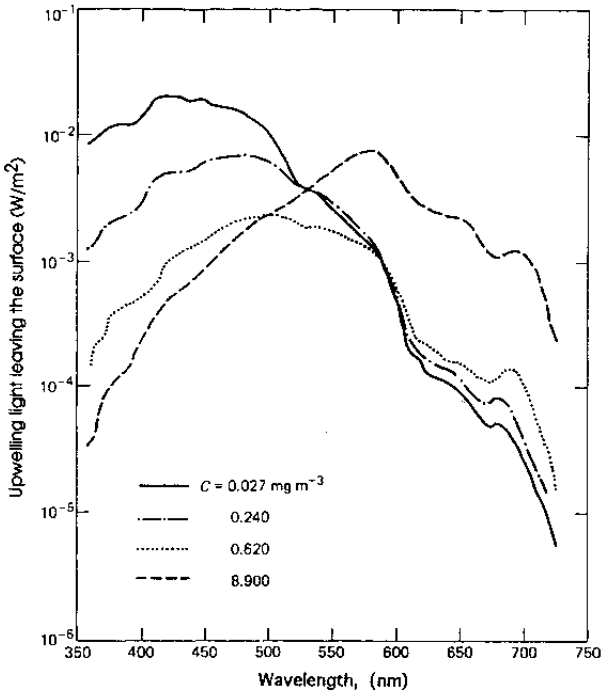

Figure 32: The intensity of light backscattered through the sea surface as a function of wavelength for various chlorophyll concentrations. (Source: Knauss, 2005, Fig. 12.16)

total absorption (integrated over all wavelengths). But at very high chlorophyll concentrations the backscatter intensity at very long wavelengths increase since scattering then dominates over absorption; at such high particle concentrations more of the incoming light is scattered back than what is being absorbed.

Details of scattering and absorption makes up the field called *marine optics*. It also includes the concept of light refraction which explains why, for example, the oar sticking into the water from a rowing boat appears to 'break' and take on a different angle in the water than what it had in the air. Marine optics is of course hugely important for life in the ocean. We will, however, not pursue such details here in this course on ocean physics and dynamics. From the above observations we should nevertheless recall that the incoming shortwave radiation penetrates some ways down into the water column. And the absorption at the various depths ultimately acts as an energy source that can warm up the water there.

Long-wave radiation The sea surface also emits black body radiation—just as the sun does—but at much longer wavelengths (in what we call the infrared part of the spectrum). The total energy flux emitted, integrated over all wavelengths, waries with temperature as

$$R_{lw} = c_s T_K^4,$$

where T_K is the ocean temperature on the Kelvin scale and c_s is the Stefan-Boltzmann constant, $5.67 \times 10^{-8}\,\mathrm{W\,m^{-2}\,K^{-4}}$. For a global-mean temperature of the ocean of around $18°\mathrm{C}$, the total outgoing longwave flux should be about $400\,\mathrm{W\,m^{-2}}$, i.e. much more than the global-mean incoming shortwave raditaion (see Fig. 28). But the ocean surface also receives black body radiation from the lower atmosphere (it too has a non-zero absolute temperature), and it is the *difference* between these two fluxes that makes up the *net* loss that goes into eqn. (2).

Latent heat flux The ocean is cooled during evaporation since the phase change from a liquid to a glass state requires energy. A simplest possible model of the resulting heat flux—defined to be positive when pointing into the ocean—is

$$Q_l = \rho_a c_e \left(e_a - e_w\right) |\boldsymbol{U}|. \tag{3}$$

Here e_w and e_a are the specific humidities of the air at the sea surface and at some height above (typically at 10 m), $|\boldsymbol{U}|$ is the wind velocity magnitude (also typically taken at 10 m height), ρ_a is the density of the air and c_e is a transfer coefficient. A lower humidity at 10 m than at the sea surface (this is the typical situation) results in a negative flux, i.e. a latent heat loss from the ocean. Why should the flux be proportional to the strength of the wind? The wind speed is really meant as an indication or a surrogate of the turbulence level in the atmospheric boundary layer; and higher tubulence levels mean faster vertical transport of the newly-formed moist air away from the sea surface as well as faster resupply of drier airs—ready to pick up and bring away some more water molecules.

Sensible heat flux The sensible heat flux is simply due to a temperature contrast between the sea surface and the air masses above. This is the good old tendency for heat to flow down the temperature gradient. So the sensible flux is typically modeled as

$$Q_s = \rho_a c_s \left(T_a - T_w\right) |\boldsymbol{U}|, \tag{4}$$

where T_w and T_a are the air temperatures at the sea surface and at some height above (10 m again) and c_s is yet another transfer coefficient. Again the strength of the flux is thought to be proportional to the wind speed, for the same reason: stronger winds mean higher turbulence levels, and turbulence is a way of enhancing the vertical transport of properties (temperature) towards or away from the surface.

Freshwater fluxes Salts are brought into the ocean by drainage (rivers) from land, not by air-sea fluxes. But the salinity near the surface can change by either removal or addition of *freshwater*, i.e. water that has no or very little salt in it. Rain is a source of freshwater to the sea surface, and this addition of freshwater dillutes the ocean surface layers to lower the salinity. Conversely, when water evaporates the salt molecules are left behind—and the salinity increases. Note that the evaporative freshwater fluxes are related to latent heat fluxes whereas the fluxes due to rain are not.

3.2 Momentum fluxes

The winds in the lower atmosphere excerts a frictional drag on the ocean surface which may accelerate the ocean (and, conversely, the ocean surface excerts friction which decelerates the winds). This *wind stress* is often modelled as

$$\boldsymbol{\tau} = c_D \rho_a \boldsymbol{U} |\boldsymbol{U}|,$$

where c_D is a drag coefficient. Notice how the expression has the same form as that of fluxes of latent and sensible heat (eqns. 3 and 4). Actually, the wind stress is a *vertical flux of horizontal momentum* ($\rho_a \boldsymbol{U}$) into the ocean from above. In practice, the drag coefficient is not taken to be constant, but varying with the wind speed and some times also with the sea state (wave height). The drag coefficient is in general assumed to increase with wind speed, although the increase is reduced or vanish altogether for wind speeds in excess of 25-30 m/s. The reason why the air-sea drag coefficient c_D saturates for high wind speeds is still debated, but one suggestion is that it is associated with breaking waves and sea spray.

3.3 The effect of sea ice

The presence of sea ice modifies air-sea fluxes at high latitudes significantly. Let's look at momentum fluxes first and how they vary as a function of the sea ice con-

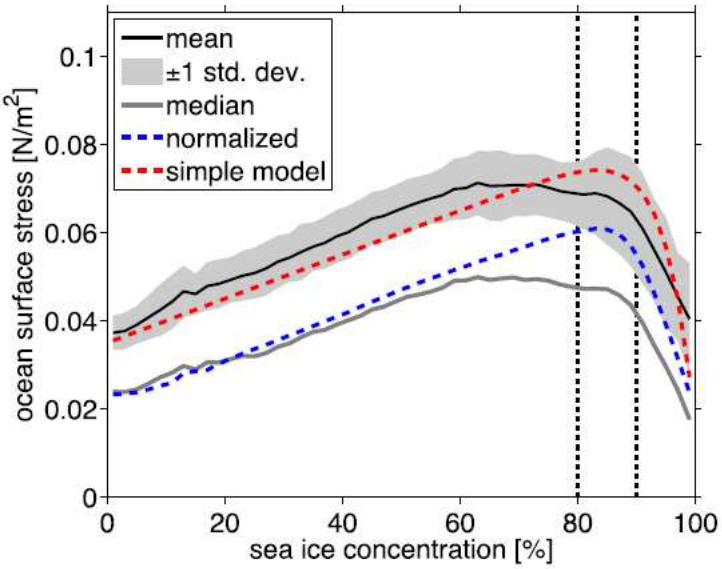

Figure 33: Estimates of the wind stress on the ocean surface as a function of sea ice concentration (Source: Martin et al., 2016, Fig. 9)

centration, in other words the fraction of ocean surface covered by ice. At low sea ice concentrations any increase in the amount of sea ice will make the surface feel rougher or bumpier to the winds. So the frictional coupling between atmosphere and ocean is enhanced and the effective wind stress increases. At very high sea ice concentrations however, the ice flows start bumping into eachother and are no longer freely responding to the winds. As the ice concentration approaches one (the entire surface is covered by ice), and in particular if the ice is also thick, the wind can blow all it wants and the ice will still not move...and neither will the ocean currents. There is, in other words, an optimal sea ice concentration where the wind stress (for any given wind speed) is optimal, as shown in Figure 33.

So momentum fluxes from the atmosphere to the ocean can either be enhanced or reduced by sea ice. In contrast, a sea ice cover always reduces air-sea heat fluxes. Since the *albedo* (the fraction of reflected shortwave radiation) of sea ice is higher than that of the ocean surface (30–95% vs. about 8%), a much larger fraction of the incoming shortwave solar radiation is simply reflected back to the atmosphere over sea ice. And the radiation which is absorbed by the ice only reaches a few millimeters into it. Any heat transport through the ice itself has to take place my molecular diffusion or *conduction*, a slow process compared to

turbulent transport. In fact, the bulk of the direct heat flux between ocean and atmosphere in ice-covered regions takes place in 'leads' or 'polynyas', openings in the sea ice created when for example the wind blows sea ice away from land. In such openings the heat fluxes may be particularly intense, reaching more than $250\,\mathrm{W\,m^{-2}}$ in winter.

The *formation* and *melting* of sea ice is intimately related to both heat and freshwater fluxes through the ocean surface. Clearly, sea ice can be melted by heat fluxes from the ocean. But sea ice formation is also associated with heat fluxes from the ocean. During the polar night, the ocean is cooled by sensible and latent heat fluxes to the atmosphere—and the atmosphere is warmed up and also picks up moisture. When the surface waters have reached freezing temperatures, about -1.9°C, the heat transfer to the atmosphere is maintained by the energy release as liquid water freezes to a solid. The details are subtle, and some of the *latent heat of fusion* (the energy released by the phase change from liquid to solid state) is also sent back to the upper ocean. But one thing it is easy to agree on is that the direction of net heat fluxes is everywhere upwards, from the ocean, via a phase transition from liquid water to ice, and eventually to the atmosphere.

When sea ice forms it is primarily pure water that freezes. Some salts are trapped in 'brine pockets' inside the ice, but eventually most is rejected into the water. So sea ice formation causes a *virtual salinity flux* into the ocean. At the near-freezing temperatures the salt injection creates very dense waters which then sink down into the water column in convective plumes. The fresh water, now in the form of sea ice, is left behind at the very surface. And when the sea ice eventually melts, for example next summer, the fresh water remains at the surface. It is light, after all. So, in fact, the net effect of the seaonal cycle of sea ice formation and melting is to *distill* the water, i.e. to remove the salt (sent to deeper layers of the ocean) from the fresh water (remains at the surface).

4 The language of nature: Conservation equations

The mathematical treatment of fluid dynamics relies on conservation equations that all basically say that what goes in of a quantity minus what goes out of a fixed control volume either has to balance (be of equal size) or result in an increase or decrease of the quantity within the control volume. It makes sense, doesn it?

4.1 Eulerian and Lagrangian descriptions

The fluid conservation laws are framed as partial differential equations that involve time derivatives and spatial derivatives. As it turns out, we can study study these laws either with respect to freely-moving fluid parcels or with respect to control volumes fixed in space. To see this dual possibility, consider any property of the fluid, having a concentration (amount of the property per unit volume) which is a function of both space and time, $c = c(x, y, z, t)$. The change of c, as one allows the independent variables t, x, y and z to change, is given by the partial derivatives:

$$dc = \frac{\partial c}{\partial t}dt + \frac{\partial c}{\partial x}dx + \frac{\partial c}{\partial y}dy + \frac{\partial c}{\partial z}dz,$$

where dt, dx, dy and dz are small ("differential") increments in time and space. So what is the total time rate of change of c experienced by a fluid parcel (of unit volume) as it flows around? We divide by dt and get

$$\begin{aligned}\frac{dc}{dt} &= \frac{\partial c}{\partial t} + \frac{\partial c}{\partial x}\frac{dx}{dt} + \frac{\partial c}{\partial y}\frac{dy}{dt} + \frac{\partial c}{\partial z}\frac{dz}{dt} \\ &= \frac{\partial c}{\partial t} + u\frac{\partial c}{\partial x} + v\frac{\partial c}{\partial y} + w\frac{\partial c}{\partial z},\end{aligned}$$

where u, v and w are the velocity components in the x, y and z directions, respectively. So the fluid parcel can experience a change in c due to a real temporal change where it happens to be located, i.e. the $\partial c/\partial t$ term, but also due to itself moving around through spatial gradients of c. The total or *Lagrangian* rate of change experienced by the moving parcel is therefore the sum of the temporal change at any one point, called the *Eulerian* rate of change, and the rate of change due to its moving through a spatial concentration gradient, the *advective* rate of change.

In fluid dynamics many like to use D/Dt (instead of d/dt) for the Lagrangian

rate of change, so we write

$$\underbrace{\frac{Dc}{Dt}}_{Lagrangian} = \underbrace{\frac{\partial c}{\partial t}}_{Eulerian} + \underbrace{u\frac{\partial c}{\partial x} + v\frac{\partial c}{\partial y} + w\frac{\partial c}{\partial z}}_{advective},$$

or, in vector notation,

$$\underbrace{\frac{Dc}{Dt}}_{Lagrangian} = \underbrace{\frac{\partial c}{\partial t}}_{Eulerian} + \underbrace{\boldsymbol{v}\cdot\nabla c}_{advective},$$

where $\boldsymbol{v} = u\boldsymbol{i} + v\boldsymbol{j} + w\boldsymbol{k}$ is the three-dimensional velocity vector and ∇ is the three-dimensional gradient operator

$$\nabla = \frac{\partial}{\partial x}\boldsymbol{i} + \frac{\partial}{\partial y}\boldsymbol{j} + \frac{\partial}{\partial z}\boldsymbol{j}.$$

The Lagrangian derivative is also sometimes called the material derivative. When introducing the various concervation equations below we will sometimes start in the Eulerian reference frame (considering budgets at a fixed point in space) and other times start in the Lagrangian frame (following a fluid parcel).

4.2 Coordinate system

The Earth is approximately a sphere, so the conservation equations should really be studied in a spherical coordinate system (Figure 34). If r is the radius of the sphere and ϕ and λ denote latitude and longitude (both in radians), then differential (small) displacements in the zonal, meridional and radial directions will be

$$\begin{aligned}\delta x &= r\cos\phi\delta\lambda \\ \delta y &= r\delta\phi \\ \delta z &= \delta r,\end{aligned}$$

and the conservation equation above would read

$$\frac{Dc}{Dt} = \frac{\partial c}{\partial t} + u\frac{1}{r\cos\phi}\frac{\partial c}{\partial\lambda} + v\frac{1}{r}\frac{\partial c}{\partial\phi} + w\frac{\partial c}{\partial r},$$

where u, v and w are the zontal, meridional and radial velocities, respectively.

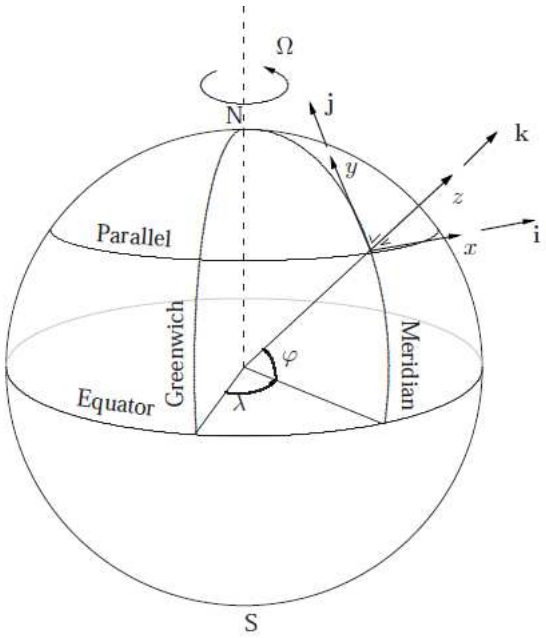

Figure 34: Spherical and local Cartesian coordinate systems for use in oceanography. (Source: Cushman-Roisin and Beckers, 2011, Fig. 2.9)

But working in a spherical coordinate system is cumbersome. For motions and displacements that are much smaller than the radius of Earth, a local flat Cartesian coordinate system is much easier to use and, for most purposes, accurate enough. What we do is simply set up a local (x, y, z) coordinate system centered on the latitude and longitue coordinates of the region of ocean we wish to study (see Figure 34). Then we do our calculations on this plane, simply ignoring the curvature of the planet.

4.3 Conservation of mass

4.3.1 The full equation

Consider a small box of volume V and density ρ, so that the mass of the box is

$$m = \rho V.$$

If the box is fixed in space (so here we are in the Eulerian reference frame) and its volume is constant, then the time rate of change of mass in the box due to flow

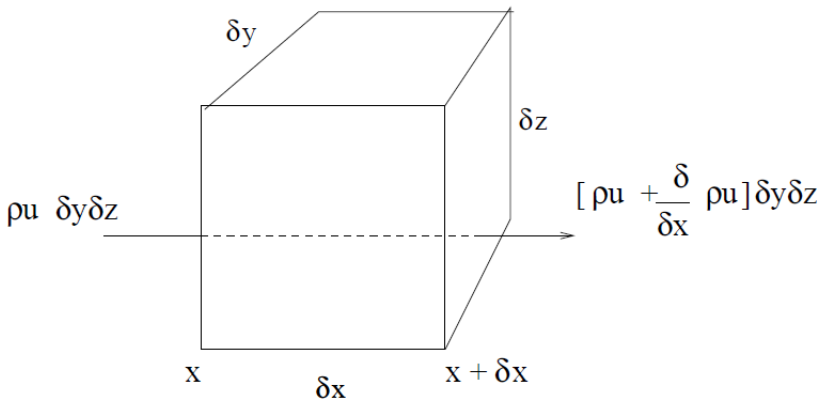

Figure 35: Concervation of mass in a box of volume $V = \delta x \delta y \delta z$. (Source: LaCasce, 2015, Fig. 1.1)

through the sides is

$$\begin{aligned} V\frac{\partial \rho}{\partial t} &= [u(x)\rho(x) - u(x+\delta x)\rho(x+\delta x)]\,\delta y \delta z + \\ &\quad [v(y)\rho(y) - v(y+\delta y)\rho(y+\delta y)]\,\delta x \delta z + \\ &\quad [w(z)\rho(z) - w(z+\delta z)\rho(z+\Delta z)]\,\delta x \delta y. \end{aligned}$$

Dividing both sizes by the volume $V = \delta x \delta y \delta z$ gives

$$\frac{\partial \rho}{\partial t} = -\frac{\delta(u\rho)}{\delta x} - \frac{\delta(v\rho)}{\delta y} - \frac{\delta(w\rho)}{\delta z},$$

where $\delta(u\rho)$, $\delta(v\rho)$ and $\delta(w\rho)$ are the differences in density fluxes between the sides of the cube in the x, y and z directions, respectively. In in the limit of a control volume of infinitesimal size, we arrive at the differential equation form:

$$\frac{\partial \rho}{\partial t} = -\frac{\partial(u\rho)}{\partial x} - \frac{\partial(v\rho)}{\partial y} - \frac{\partial(w\rho)}{\partial z},$$

or, in vector notation,

$$\frac{\partial \rho}{\partial t} = -\nabla \cdot (\boldsymbol{v}\rho)\,.$$

So the time rate of change of density in an infinitesimal control volume is given by the *convergence* (negative divergence) of the density transport into it. Essentially,

as we outlined at the beginning, the mass in the box increases if there is more mass flowing into it than leaving it (density is just mass per unit volume).

Note that if we split up the spatial derivaties by the product rule, we can write

$$\frac{\partial \rho}{\partial t} = -\left(u\frac{\partial \rho}{\partial x} + v\frac{\partial \rho}{\partial y} + w\frac{\partial \rho}{\partial z}\right) - \rho\left(\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} + \frac{\partial w}{\partial z}\right),$$

or

$$\frac{\partial \rho}{\partial t} + \left(u\frac{\partial \rho}{\partial x} + v\frac{\partial \rho}{\partial y} + w\frac{\partial \rho}{\partial z}\right) = -\rho\left(\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} + \frac{\partial w}{\partial z}\right).$$

In vector notation this becomes

$$\frac{\partial \rho}{\partial t} + \boldsymbol{v} \cdot \nabla \rho = -\rho \nabla \cdot \boldsymbol{v},$$

or, remembering our definition of the Lagrangian or material derivative,

$$\frac{1}{\rho}\frac{D\rho}{Dt} = -\nabla \cdot \boldsymbol{v}. \tag{5}$$

So for a fluid parcel which is being advected around by the flow around it, the fractional change of its density is given by the convergence of the flow field. This makes sense: a converging flow field compresses the fluid and raises the density.

4.3.2 The Boussinesq approximation

The density of air can change considerably througout the atmosphere, say between the bottom and top of the troposphere. But water density changes very little from a value just above one thousand kilos per cubic meter ($\rho \sim 1027\,\mathrm{kg\,m^{-3}}$). The velocity field however can change by 100% over relatively small distances or over relatively short time scales. So for most applications in oceanography, the left hand side of Eqn. (5) is so small compared to the right hand side that it can be ignored. This is called *the Boussinesq approximation*, and the mass budget then reduces to

$$\nabla \cdot \boldsymbol{v} = 0$$

or

$$\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} + \frac{\partial w}{\partial z} = 0,$$

meaning that the three-dimensional flow field is (approximately) non-divergent. If the horizontal velocity field is convergent at some point in space, the vertical flow

there needs to be divergent, and vice versa. So concervation of mass has instead turned into an expression for conservation of volume. In the remainder of these notes we will often call this expression the *continuity equation*.

Making the Boussinesq approximation has several other implications which we will see below. In deriving these implications we will use the assumption that density is a constant, say $\rho_0 = 1027\,\mathrm{kg\,m^{-3}}$, plus a much smaller deviation that can change in space and time, i.e.

$$\rho = \rho_0 + \rho'(x, y, z, t),$$

where

$$\rho' \ll \rho_0.$$

4.4 Conservation of salt

Recall that to estimate density (which we have seen to be central to understand the motion of the oceans) we need to find the salinity and temperature of the water. To make predictions we therefore need mathematical expression for the evolution of ocean salinity and temperature. Again, the basic principle is that the difference between what comes in minus what goes out of a control volume leads to a change in the property concentration there.

Consider again the box above but now set up a budget for the amount of salt flowing in and out. The property to be conserved is the mass of salt per unit volume ρS (recall that salinity S is the mass of salt per unit mass of water). The procedure is the same as above and gives the result,

$$\begin{aligned} \frac{\partial(\rho S)}{\partial t} &= -\frac{\partial(u\rho S)}{\partial x} - \frac{\partial(v\rho S)}{\partial y} - \frac{\partial(w\rho S)}{\partial z} \\ &= -\nabla\cdot(\boldsymbol{v}\rho S)\,. \end{aligned}$$

We can get to an equation for conservation of salinity by invokinig the Boussinesq approximation, i.e. by assuming that density is approximately constant so that it drops out (it's the same constant that can be taken out of all derivatives). The resulting equation then becomes

$$\begin{aligned} \frac{\partial S}{\partial t} &= -\left(\frac{\partial uS}{\partial x} + \frac{\partial vS}{\partial y} + \frac{\partial wS}{\partial z}\right) \\ &= -\nabla\cdot(\boldsymbol{v}S)\,. \end{aligned}$$

But there is one additional transport term that can be added to this equation. In contrast to mass itself, salt molecules can *diffuse* across the walls due to random molecular motion. The molecules move back and forth randomly, exchanging properties (salinity) as they bump into each other. Since the motion of the molecules is random, back and forth, there is no net mass transport but a transport of salt *down the concentration gradient* (from high to low concentration). In fact, this salt transport is proportional to the actual strength of the conentration gradient, so that the diffusive flux can be written

$$\begin{aligned} \boldsymbol{F}_S &= -\kappa_S \nabla S \\ &= -\kappa_S \left(\frac{\partial S}{\partial x}\boldsymbol{i} + \frac{\partial S}{\partial y}\boldsymbol{j} + \frac{\partial S}{\partial z}\boldsymbol{k} \right), \end{aligned}$$

where κ_S is the molecular diffusion coefficient or *diffusivity* for salt.

Hence, the total transport of salt is the sum of the advective component (involving also a net flow and mass transport) and the diffusive component (involving no net mass transport), or

$$\frac{\partial S}{\partial t} = -\nabla \cdot (\boldsymbol{v} S - \kappa_S \nabla S),$$

The molecular diffusivity for salt is constant, so one can take it outside of the derivatives to give

$$\frac{\partial S}{\partial t} = -\nabla \cdot (\boldsymbol{v} S) + \kappa_S \nabla^2 S,$$

where $\nabla^2 = \partial^2/\partial x^2 + \partial^2/\partial y^2 + \partial^2/\partial z^2$ is the Laplace operator.

Note, finally, that under the Boussinesq approximation (where $\nabla \cdot \boldsymbol{v} = 0$), the salinity equation can also be written

$$\frac{\partial S}{\partial t} + \boldsymbol{v} \cdot \nabla S = \kappa_S \nabla^2 S,$$

or

$$\frac{DS}{Dt} = \kappa_S \nabla^2 S.$$

So the salinity of a parcel moving with the fluid flow changes only by diffusion.

4.5 Conservation of thermal energy

An equation for temperature stems from the first law of thermodynamics which states that the internal energy of a system can increase if heat flows into it or

pressure work is exterted on it. The derivation is rather complicated and makes use of several assumptions. But an approximate and often useful final expression in terms of potential temperature (where we don't have to worry about the pressure effect on temperature) is

$$\frac{\partial \theta}{\partial t} = -\nabla \cdot (\boldsymbol{v}\theta - \kappa_T \nabla \theta + \boldsymbol{J}_R) \,.$$

where κ_T is the molecular diffusion coefficient for temperature and and $\boldsymbol{J}_R$ is a radiative temperature flux (remember, shortwave radiation is for example able to penetrate some ways into the water column). So the expression looks similar to that of salinity except for the extra radiative flux term. Finally, as for salinity, the Boussinesq approximation allows one to rewrite the advective flux to give

$$\frac{\partial \theta}{\partial t} = -\boldsymbol{v} \cdot \nabla \theta - \nabla \cdot (-\kappa_T \nabla \theta + \boldsymbol{J}_R) \,,$$

or

$$\frac{D\theta}{Dt} = -\nabla \cdot (-\kappa_T \nabla \theta + \boldsymbol{J}_R) \,.$$

Again, the molecular diffusion coefficient for temperature is constant (but different than the one for salinity[8]), so a final expression can be written

$$\frac{D\theta}{Dt} = \kappa_T \nabla^2 \theta - \nabla \cdot \boldsymbol{J}_R.$$

4.6 The momentum equations

What we call the momentum equations are really Newton's second law which states that mass times acceleration of a particle is given by the sum of forces applied to it. On a rotating planet a fluid parcel will experience a set of real forces and in addition *virtual* forces or, actually, accelerations that come about just because of the rotation. We will look at the real forces first and now put ourselves in the reference frame of the moving parcel, i.e. in the Lagrangian reference frame.

[8]The different molecular diffusivities for temperature and salt (heat diffusion is faster) can cause some very interesting phenomena, like "double deffusive layering" and "salt fingering". Look it up!

4.6.1 Real forces

Newton's second law applied to a moving fluid parcel of mass $\delta m = \rho \delta V$ is

$$\rho \delta V \frac{D\boldsymbol{v}}{Dt} = \sum \boldsymbol{F},$$

where $D\boldsymbol{v}/Dt$ is the (Lagrangian) acceleration of the parcel and $\sum \boldsymbol{F}$ is the sum of forces. In terms of the three components of velocity,

$$\begin{aligned} \rho \delta V \frac{Du}{Dt} &= \sum F_x, \\ \rho \delta V \frac{Dv}{Dt} &= \sum F_y, \\ \rho \delta V \frac{Dw}{Dt} &= \sum F_z, \end{aligned}$$

where F_x, F_y and F_z are forces in the x, y and z directions. The forces experienced by a fluid parcel are gravity, pressure gradients, and frictional stresses. Below, we go through each in turn.

Gravity Gravity is a so-called body force, working on every mass element that makes up the parcel. It points downward, in the negative z direction, or

$$F_z^g = -\rho \delta V g.$$

So if gravity were the only player in town, the z momentum equation (per unit volume) would become

$$\rho \frac{Dw}{Dt} = -\rho g.$$

Pressure gradient Presssure is a so-called surface force, acting on the surface of a fluid parcel. At every point of the surface the pressure force from the surrounding fluid is pointing normal to the surface, into the parcel. Consider again the fluid parcel shown in Figure 35. The net pressure force in the positive x-direction (to the right) is the difference between the pressure force acting on the left face (pushing the parcel to the right) and the pressure force on the right face (pushing the parcel to the left). Pressure is a force per unit area, so the total force on each surface is the pressure times the area of that surface. The net force on the parcel is therefore

$$F_x^p = p(x)\delta y \delta z - p(x + \delta x)\delta y \delta z,$$

and setting this into the x component of the force balance gives

$$\rho\delta V\frac{Du}{Dt} = p(x)\delta y\delta z - p(x+\delta x)\delta y\delta z.$$

Dividing by the unit volume $\delta V = \delta x\delta y\delta z$ gives

$$\rho\frac{Du}{Dt} = \frac{p(x) - p(x+\delta x)}{\delta x},$$

or, as we let the size of the control volume become infinitesimally small,

$$\rho\frac{Du}{Dt} = -\frac{\partial p}{\partial x}.$$

The pressure forces in the other two directions take the same form,

$$\begin{aligned}\rho\frac{Dv}{Dt} &= -\frac{\partial p}{\partial y},\\ \rho\frac{Dw}{Dt} &= -\frac{\partial p}{\partial z}.\end{aligned}$$

Frictional or viscous stresses Viscous stresses also act on the surface of the fluid parcel. We distinguish between *normal stresses*, which act perpendicular to the surface, and *shear stresses*, which act parallel to the surface. The stresses are often denoted τ_{ij}, where the subscripts indicate in what direction the stress is acting (i), and the orientation of the surface the stress is acting on (j), see Fig. 36. Two important relations hold for such fluids as we consider here: (1) we have $\tau_{i(-j)} = -\tau_{ij}$, that is, stresses on the opposite side of the same surface are equal in magnitude and oppositely directed (cf. Newton's third law), and (2) there is a symmetry with $\tau_{ij} = \tau_{ji}$, which expresses that the total moment about an arbitrary point in the fluid must be zero.

To derive the net force due to viscous stresses, let us first consider the shear stresses in the x-direction that acts on surfaces oriented in the positive and negative z-directions (i.e., the "top" and "bottom" of the fluid parcel depicted in Fig. 37). To obtain the force, we must multiply by area, hence we have

$$F_x^{(xz)} = \tau_{xz}(z+\delta z)\delta x\delta y + \tau_{x(-z)}(z)\delta x\delta y.$$

Now, since $\tau_{x(-z)}(z) = -\tau_{xz}(z)$, we get

$$F_x^{(xz)} = \tau_{xz}(z+\delta z)\delta x\delta y - \tau_{xz}(z)\delta x\delta y.$$

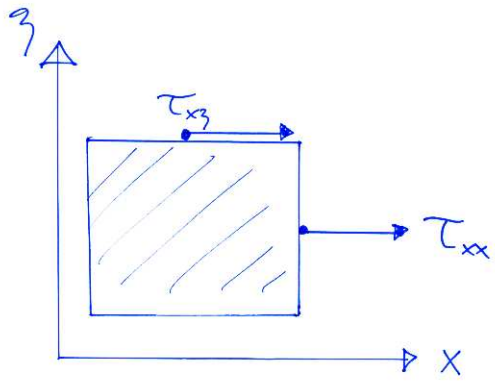

Figure 36: Normal and shear stresses acting on a volume of fluid. The direction normal to the surface, pointing away, defines the orientation of the surface. Hence we see here examples of (1) a normal stress, τ_{xx}, directed along positive x and also acting on a surface oriented in the same direction, and (2) a shear stress, τ_{xz}, which is also in the positive x-direction but acting on a surface oriented in the direction of positive z.

We find similar relations for the other stresses in the x-direction:

$$\begin{aligned} F_x^{(xx)} &= \tau_{xx}(x+\delta x)\delta y\delta z - \tau_{xx}(x)\delta y\delta z, \\ F_x^{(xy)} &= \tau_{xy}(y+\delta y)\delta x\delta z - \tau_{xy}(y)\delta x\delta z. \end{aligned}$$

Dividing by volume, as done before, and letting the size of the parcel go to zero gives

$$\rho\frac{Du}{Dt} = \frac{\partial\tau_{xx}}{\partial x} + \frac{\partial\tau_{xy}}{\partial y} + \frac{\partial\tau_{xz}}{\partial z}.$$

We now need to relate the stresses to the motion of the fluid using a *constitutive law*. Here we will assume that the fluid is *Newtonian*, which means that the stresses are taken to be linear in the velocity gradients. Because of the symmetry

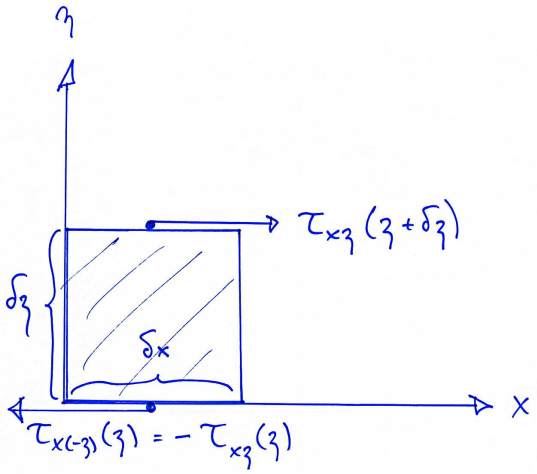

Figure 37: The shear stresses in the x-direction acting on surfaces oriented with the negative and positive z-direction.

condition $\tau_{ij} = \tau_{ji}$ we only need to specify six stresses:

$$\begin{aligned}
\tau_{xx} &= 2\mu\frac{\partial u}{\partial x}, \\
\tau_{yy} &= 2\mu\frac{\partial v}{\partial y}, \\
\tau_{zz} &= 2\mu\frac{\partial w}{\partial z}, \\
\tau_{xy} &= \tau_{yx} = \mu\left(\frac{\partial u}{\partial y} + \frac{\partial v}{\partial x}\right), \\
\tau_{xz} &= \tau_{zx} = \mu\left(\frac{\partial u}{\partial z} + \frac{\partial w}{\partial x}\right), \\
\tau_{yz} &= \tau_{zy} = \mu\left(\frac{\partial v}{\partial z} + \frac{\partial w}{\partial y}\right),
\end{aligned}$$

where μ is called the *dynamic viscocity*. It should be noted that in the absence of fluid deformation, there are no viscous stresses in the fluid. This fact is intuitive: consider a specific volume of fluid which is either at rest, or passively flowing along in a current *without changing its shape*, then there are no viscous stresses acting on (or within) the volume.

Since the dynamic viscosity is constant, as for the molecular transfer coeffi-

cients for salinity and temperature, the final expression becomes[9]

$$\begin{aligned}\rho\frac{Du}{Dt} &= \mu\left[\frac{\partial^2 u}{\partial x^2}+\frac{\partial^2 u}{\partial y^2}+\frac{\partial^2 u}{\partial z^2}\right]\\ &= \mu\nabla^2 u.\end{aligned}$$

The same argumentation of course applies in the y- and z-directions. So, to sum up, the momentum equations, when adding up all three forces, become

$$\begin{aligned}\rho\frac{Du}{Dt} &= -\frac{\partial p}{\partial x}+\mu\nabla^2 u\\ \rho\frac{Dv}{Dt} &= -\frac{\partial p}{\partial y}+\mu\nabla^2 v\\ \rho\frac{Dw}{Dt} &= -\frac{\partial p}{\partial z}-\rho g+\mu\nabla^2 w.\end{aligned}$$

4.6.2 The Boussinesq approximation

We will now check whether the momentum equations can be simplified a bit under the Boussinesq approximation of very small density changes. As mentioned earlier, to do this we substitute in $\rho = \rho_0 + \rho'(x, y, z, t)$ where $\rho' \ll \rho_0$. In the two horizontal momentum equations it is safe to ignore ρ' compared to ρ_0 on the left hand side. If we then divide by ρ_0 we get

$$\begin{aligned}\frac{Du}{Dt} &= -\frac{1}{\rho_0}\frac{\partial p}{\partial x}+\nu\nabla^2 u\\ \frac{Dv}{Dt} &= -\frac{1}{\rho_0}\frac{\partial p}{\partial y}+\nu\nabla^2 v,\end{aligned}$$

where $\nu \equiv \mu/\rho_0$ is called the kinematic viscocity. Not tremendously interesting, perhaps.

In the vertical momentum equation we can also ignore the perturbation density term ρ' on the left hand side but, as it turns out, not on the right hand side where it is multiplied by the gravitational acceleration. To show this we will compare the typical size of some of the different terms. After ignoring ρ' on the left hand side and then dividing by ρ_0, we have

$$\frac{Dw}{Dt} = -\frac{1}{\rho_0}\frac{\partial p}{\partial z}-\left(1+\frac{\rho'}{\rho_0}\right)g+\nu\nabla^2 w.$$

[9] Here we use the continuity equation $\partial u/\partial x = -\partial v/\partial y + \partial w/\partial z$.

What we will do now is compare the two gravity terms with the acceleration terms, assuming that they are all approximately the same size as $\partial w/\partial t$. To create some whooping vertical accelerations we assume that vertical velocities can reach ten centimeters per second, $W \sim 10^{-1}\mathrm{m\,s^{-1}}$, and that they can change by this amount (i.e. change by 100%) over a time scale of a couple of minutes, $T \sim 10^2\,\mathrm{s}$. This gives a vertical acceleration term of approximate size

$$\frac{\partial w}{\partial t} \sim \frac{W}{T} \sim \frac{10^{-1}}{10^2} \sim 10^{-3}\mathrm{m\,s^{-2}}.$$

Then we estimate the size of the two gravity terms. The first one is simply

$$g \sim 10\,\mathrm{m\,s^{-2}},$$

in other words four orders of magnitude larger than acceleration! To estimate the size of the second gravity term we take $\rho_0 \sim 1000\,\mathrm{kg\,m^{-3}}$ and $\rho' \sim 1\,\mathrm{kg\,m^{-3}}$. This gives

$$\frac{\rho'}{\rho_0} g \sim \frac{1}{10^3} \cdot 10 \sim 10^{-2}\,\mathrm{m\,s^{-2}}.$$

So even the gravity term involving ρ' is larger than the acceleration term. This is a hint of the *hydrostatic approximation* which we will look more closely at later. But, for now, we have to conclude that ρ' simply cannot be ignored when it stands side by side with g, at least not if we want to also keep the vertical acceleration terms. So, in summary, under the Boussinesq approximation we replace ρ with ρ_0 in all terms of the momentum equations (in both the horizontal and the vertical components of the equations) *except* where it it multiplied by the gravitational acceleration g.

4.6.3 The ficticious (!) Coriolis and centrifugal forces

Newton's second law applies in a fixed or an *inertial* reference frame. That's a reference frame which does not accellerate. But when we observe the ocean flow, say from a ship, we are standing on a rotating planet. And rotation is acceleration. So we are observing the response of the ocean to forces in a rotating reference frame, not a fixed one. The forces themselves are the same whether measured in a fixed or rotating reference frame. What we have to be careful about is our interpretation of the motions and accelerations we observe. What we will find when studying the relationship between accelerations in the fixed and rotating reference frame is of course the famous Coriolis force or, more correctly, the

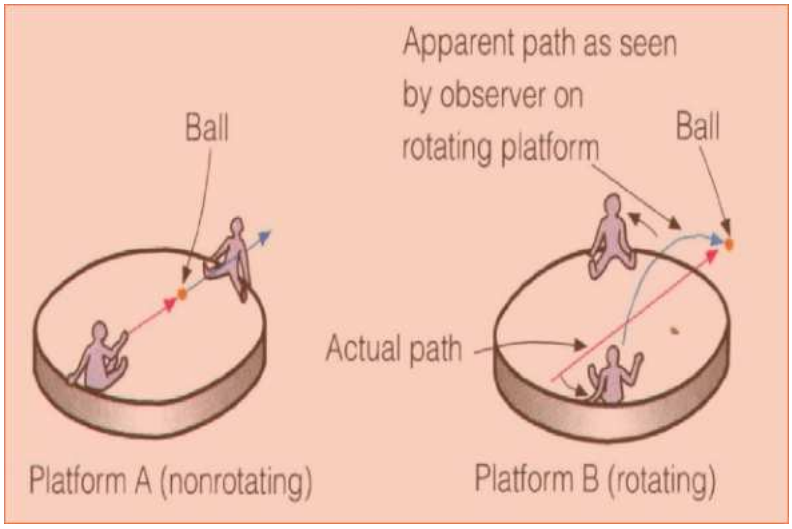

Figure 38: Two friends throwing a ball at each other, sitting on a (left) non-rotating and a (right) rotating disk.

Coriolis acceleration. The centrifugal force, or what is actually more correctly termed the centripetal accelleration, can be imbedded into the gravity force term, as we'll soon see.

The Coriolis accellertion would be easily understood by two friends sitting on each side of a rotating disk (see Figure 38). If one throws a ball directly at the other she would probably miss. After being released the ball would move in a straight line—observed from someone who looks at the rotating disk from a distance. But by the time the ball reaches the other side of the disk, what has happened? The disked has rotated a bit...and so has the friend on that other side. To the two friends rotating with the disk it looks as if it is the ball that acts funnily, moving along a curved path. Not so! The aparent curved path is a *virtual* accelleration that is observed only by the two friends that are themselves rotating. To them it seems that the ball experiences a mysterious force (the Coriolis force!) that curves its path.

On a rotating planet (rather than a rotating disk) we need to use some vector calculus to capture the Coriolis accelleration (and the centripetal accelleration) properly. We'll loose sight of the rotating disk for a while, but will see how it shows up in the math again towards the end.

Imagine a particle situated on the surface of a planet which rotates around its own north-south axis with angular speed Ω. The situation is illustrated schemat-

ically in Figure 39. Even if the particle is still (not moving) with respect to the rotating planet itself it is certainly moving (in a circle, right?) with respect to the distant stars. If $\boldsymbol{r}$ is the position vector of the particle, measured from the center of the planet, and $\boldsymbol{\Omega}$ is the rotation vector of the planet itself (the planet rotates around $\boldsymbol{\Omega}$ with rotation speed $\Omega = |\boldsymbol{\Omega}|$), then the time rate of change of its position in the inertial reference frame[10] is given by the cross product

$$\left(\frac{D\boldsymbol{r}}{Dt}\right)_I = \boldsymbol{\Omega} \times \boldsymbol{r}.$$

This relationship holds for *any* vector $\boldsymbol{C}$, not just the position vector, as illustrated in Figure 39. But here we are interested in the position vector, so that $\boldsymbol{C} = \boldsymbol{r}$. And the time rate of change of position is of course the velocity of the particle. So the velocity in the inertial frame is

$$\boldsymbol{v}_I = \boldsymbol{\Omega} \times \boldsymbol{r}.$$

Note that the velocity is at right angles to both $\boldsymbol{\Omega}$ and $\boldsymbol{r}$ and that its direction is given by the right-hand rule for the cross product. So our particle is moving along a latitude circle on the planet.

If the particle is also moving relative to the planet itself, e.g. if it is a boat steaming across the ocean at velocity $\boldsymbol{v}_R$, then its total velocity with respect to distant stars is the sum of this *relative* velocity and its velocity due to the rotation of the planet, or

$$\boldsymbol{v}_I = \boldsymbol{v}_R + \boldsymbol{\Omega} \times \boldsymbol{r}.$$

So, if we again write the velocity in terms a time derivative, we have

$$\left(\frac{D\boldsymbol{r}}{Dt}\right)_I = \left(\frac{D\boldsymbol{r}}{Dt}\right)_R + \boldsymbol{\Omega} \times \boldsymbol{r},$$

where the R subscript means relative to the rotating planet and the subscript I means relative to the fixed stars (the inertial reference frame). Now, as said, the relationship between time derivatives in the inertial and rotating frame holds for any vector, not just $\boldsymbol{r}$ itself. So we can apply it once more, now to the velocity vector, to arrive at an expression for the acceleration. This gives (try it!),

$$\left(\frac{D\boldsymbol{v}_I}{Dt}\right)_I = \left(\frac{D\boldsymbol{v}_I}{Dt}\right)_R + \boldsymbol{\Omega} \times \boldsymbol{v}_I.$$

[10]Here we will ignore the contribution to the total velocity from the planet's orbiting around a sun.

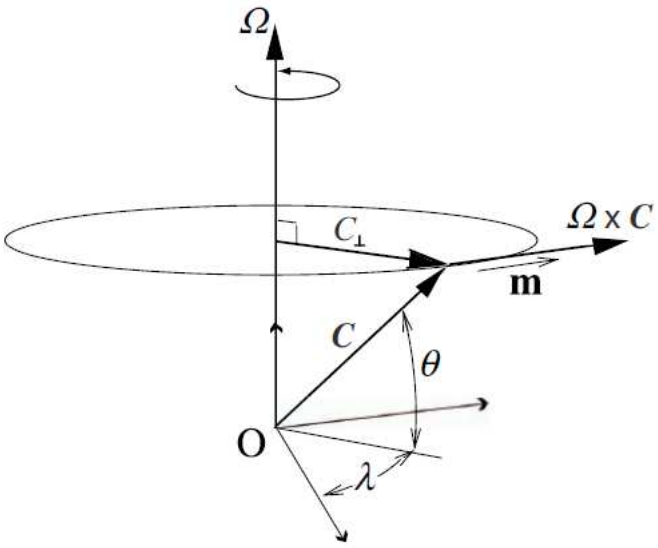

Figure 39: The rate of change of any vector $\boldsymbol{C}$ (here we take $\boldsymbol{C} = \boldsymbol{r}$, the position vector) with respect to a fixed or inertial reference frame when $\boldsymbol{C}$ is rotating around the north-south axis of a planet at angular speed Ω. (Source: Vallis, 2006, Fig. 2.1)

Notice that the left hand side here is the acceleration in the inertial reference frame, i.e. the acceleration that go into Newton's second law. The trick now is to substitute in our previous expression for for $\boldsymbol{v}_I$ on the right hand side. This gives

$$\begin{aligned}\left(\frac{D\boldsymbol{v}_I}{Dt}\right)_I &= \left(\frac{D\left(\boldsymbol{v}_R+\boldsymbol{\Omega}\times\boldsymbol{r}\right)}{Dt}\right)_R+\boldsymbol{\Omega}\times\left(\boldsymbol{v}_R+\boldsymbol{\Omega}\times\boldsymbol{r}\right)\\ &= \left(\frac{Dv_R}{Dt}\right)_R+\left(\frac{D\left(\boldsymbol{\Omega}\times\boldsymbol{r}\right)}{Dt}\right)_R+\boldsymbol{\Omega}\times\boldsymbol{v}_R+\boldsymbol{\Omega}\times\left(\boldsymbol{\Omega}\times\boldsymbol{r}\right)\\ &= \left(\frac{Dv_R}{Dt}\right)_R+\left(\frac{D\boldsymbol{\Omega}}{Dt}\right)_R\times\boldsymbol{r}+2\boldsymbol{\Omega}\times\boldsymbol{v}_R+\boldsymbol{\Omega}\times\left(\boldsymbol{\Omega}\times\boldsymbol{r}\right),\end{aligned}$$

where, in the end, we have applied the product rule to the second derivative of the second line. The second term of the final expression measures the time rate of change of the rotation of the planet, and this we can safely ignore for our pursposes. So we have

$$\left(\frac{D\boldsymbol{v}_I}{Dt}\right)_I=\left(\frac{Dv_R}{Dt}\right)_R+2\boldsymbol{\Omega}\times\boldsymbol{v}_R+\boldsymbol{\Omega}\times\left(\boldsymbol{\Omega}\times\boldsymbol{r}\right),$$

or, in terms of accelerations,

$$\boldsymbol{a}_I=\boldsymbol{a}_R+2\boldsymbol{\Omega}\times\boldsymbol{v}_R+\boldsymbol{\Omega}\times\left(\boldsymbol{\Omega}\times\boldsymbol{r}\right).$$

The first term on the right hand side is the acceleration in the rotating reference frame (the one we would observe standing on the planet). The second term is the *Coriolis acceleration*; it is at right angles to both the earth's rotation vector and to the particle velocity. This term will follow us in our continued study of the ocean circulation. Finally, the third term is the *centripetal acceleration*; it is only a function of the planet's rotation rate and the position of the particle. It points inwards towards the planets rotation axis in a plane perpendicular to the planetary rotation vector itself.

The centripetal acceleration (or centrifugal force) As it turns out, we will not have to worry about the centripetal acceleration term since it can actually be absorbed into the gravity force term. To see this, we move this acceleration term to the right hand side of the momentum equations so that it appears as an additional force, the 'centrifugal force'. The sum of this (virtual) force and the gravitational force is shown in Figure 40. The thing to note is that the Earth is not a perfect

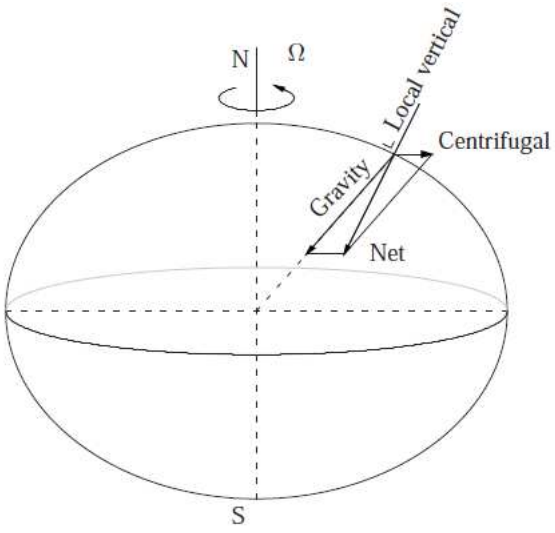

Figure 40: The balance of gravity and the (virtual) centrifugal force on the oblique Earth. The net effect is a modified gravity force which points normal to Earth's surface. (Source: Cushman-Roisin and Beckers, 2011, Fig. 2.2)

sphere but what's called an oblique ellipsoid; it is bulging out a bit at the equator because of its rotation. On this oblique ellipsoid the gravity force vector points to the center of the planet and the centrifugal force points outwards, as shown in the figure. The sum of these two forces is a 'modified gravity' force

$$\boldsymbol{g}' = \boldsymbol{g} - \boldsymbol{\Omega} \times (\boldsymbol{\Omega} \times \boldsymbol{r})$$

which points not exactly to the center of the Earth but perpendicularly to the Earth's surface at the point in question. This is perfect if we want to apply a local cartesian coordinate plane there since the gravity will then point in the negative local z direction. The actual bulging of the planet is tiny (it is exaggerated greatly in the figure), so the correction to the actual value of the gravitational acceleration is also small. In the following, we drop the prime for this modified gravity and set its magnitude to $g = |\boldsymbol{g}| \sim 9.8\,\mathrm{m\,s^{-2}}$.

The Coriolis acceleration (or Coriolis force, for those who insist) The Coriolis acceleration, however, will stay with us. So it's worth looking a bit more into this one, and we will do that by examining its components in the x, y and z directions on our cartesian plane coordinate system put down tangentially on a given

point on Earth's surface. If the tangent plane is put at a point with latitude θ and longitue ϕ, then the Earth's rotation vector will have local components

$$\boldsymbol{\Omega} = \Omega\cos\theta\boldsymbol{j} + \Omega\sin\theta\boldsymbol{k},$$

so that the three components of the Coriolis acceleration there become

$$\begin{aligned} 2\boldsymbol{\Omega}\times\boldsymbol{v} &= \begin{vmatrix} \boldsymbol{i} & \boldsymbol{j} & \boldsymbol{k} \\ 0 & 2\Omega\cos\theta & 2\Omega\sin\theta \\ u & v & w \end{vmatrix} \\ &= (2\Omega\cos\theta\, w - 2\Omega\sin\theta\, v)\,\boldsymbol{i} + 2\Omega\sin\theta\, u\,\boldsymbol{j} - 2\Omega\cos\theta\, u\,\boldsymbol{k}. \end{aligned}$$

If we now, simply for convenience, introduce the notation $f = 2\Omega\sin\theta$ and $f_* = 2\Omega\cos\theta$, the three momentum equations to be used in our rotating reference frame become

$$\begin{aligned} \frac{Du}{Dt} + f_* w - fv &= -\frac{1}{\rho_0}\frac{\partial p}{\partial x} + \nu\nabla^2 u \\ \frac{Dv}{Dt} + fu &= -\frac{1}{\rho_0}\frac{\partial p}{\partial y} + \nu\nabla v \\ \frac{Dw}{Dt} - f_* u &= -\frac{1}{\rho_0}\frac{\partial p}{\partial z} - \left(1 + \frac{\rho'}{\rho_0}\right) g + \nu\nabla^2 w, \end{aligned}$$

where g is really the gravitational acceleration which has been slightly modified by the presence of the centrifugal force as discussed above. Note that when the Coriolis acceleration is put into the equations of motion, like here, oceanographers call it the 'Coriolis force'. This is, as we now know, somewhat misleading since it is really a virtual acceleration term that pops up since we are doing our calculations in a non-inertial reference frame (our rotating planet). Even in these lecture notes, or in class, the term 'Coriolis force' may sneak itself in (and you should then point your finger at the instructore for using such sloppy language).

So where's the rotating disk we talked about in the begining? Let's look only at the two horizontal momentum equations and see what happens if only the Coriolis terms are present. Ignoring the Coriolis term involving the verticl velocity (vertical velocities are tiny compared to horizontal velocities, as we'll see later), the equations become

$$\begin{aligned} \frac{Du}{Dt} &= fv \\ \frac{Dv}{Dt} &= -fu. \end{aligned}$$

So the accellerations in the x and y directions are given by fv and $-fu$, respectively. You'll see that this is exactly the behavior observed by the two friends sitting on a rotating disk throwing a ball at each other. They are in fact sitting on a local coordinate system that is rotating around its vertical axis at angular speed $f/2$. If the ball they're throwing has a very low speed, they would have a hard time hitting each other with it. The amount of deflection of the ball, from that of a straight line (remember, as *they* observe it), depends on how much the rotating plane they're sitting on has been able to rotated in the time it takes the ball to cover the distance between the two. Of course, a ball thrown at a speed of around 10 m/s between two friends standing about 10 meters apart, the deflection won't be very big. We only need to worry about the Coriolis term for motions whose time scales are about the same as or longer than the rotation period of the planet. Ocean currents definitely operate at time scales of days and longer. So they are seriously affected!

4.7 Turbulent mixing and Reynolds fluxes

Molecular diffusion of salt and heat is an extremely slow process. One can also show that molecular diffusion only extends over distances of a few centimeters at the most. So when we model geophysical flows, like the ocean circulation or ocean waves, we typically neglect the molecular diffusion terms shown above.[11]

The same can be said for any other tracer. If we juxtaposed two fluids of different color in a lab tank, say red and blue, and then let the tank in peace, the two would eventually mix by molecular diffusion until the entire fluid takes on the color of violet. But we would have to wait a *very long* time. If we instead, in our boredom, started to stirr the two fluids by irregular movements with a paddle wheel, a soup of violet fluid would emerge rather quickly. The stirring creates *turbulence* in the fluid, chaotic motion which enhances the effective mixing of the fluid parcels. What really goes on is that the turbulent stirring stretches fluid elements into intertwined filaments and sheets of red and blue fluid. Continued stretching makes the sheets progressively thinner, down to thickness scales over which molecular diffusion actually act, and it also enhanes the area of contact between red and blue fluids. In essence, the turbulent stirring efficiently increases the surface area—by several orders of magnitude—over which molecular diffusion acts.

[11] One process in which molecular diffusion *is* important is double diffusion, or salt fingering, as we've already mentioned above.

Chaotic or turbulent motion in fluids doesn't require vigurous stirring by a paddle wheel to emerge. Large-scale fluid flows are turbulent by nature due to *nonlinear* terms in the governing equations. Nonlinear terms are those that consist of a product of two (or more) of the dependent variables. So, for example, the advection term $\nabla\,(\boldsymbol{v}\cdot S)$ in the salt equation is nonlinear since it involves a product between velocity and salinity which are both variables that we solve for.

Turbulent motions are extremely difficult or even impossible to model accurately. They are just too complex, and possess an extreme "sensitive dependence on initial conditions" (in the words of Edward Lorenz, the father of modern chaos theory). So we normally just give up on modeling turbulent motions in detail and instead focus on their *net effects* on the the flow as a whole. This typically means trying to predict the the net effects on scales of the flow that are larger than the typical scales of the turbulent motion itself.

Figure 41 can illustrate this point. It shows a jet of fluid shooting out from the left. Let's assume it's a river flowing out from a coast and that the color shown indicates the salinity. Freshwater (red color) shoots out into a salty (blue) ocean. The flow is clearly chaotic, or turbulent, with lots of small-scale structure in it. But if we average in space with our eye we also see a systematic large-scale evolution: the river plume gets wider and more salty as it extends from the mouth. Time-averaging would also work. If we showed a movie of the jet we would see that the turbulent wiggles move around chaotically. A great mess! But if we then averaged over a long time, a time much longer than the typical time scales of each wiggle, each individual turbulent feature would be averaged out and we would be left with the impression of a rather smooth jet which widens and gets saltier as it extends to the right in the figure. This graduel 'spreading' of the fresh water, away from the jet axis, is the net effect of turbulence which we wish to capture (having given up on getting every little whirl right).

So if we want to try to understand the large-scale evolution of the river plume we need equations that focus on the larger scales and treat the turbulent motions only to the extent that they impact on the larger scales. And if we wish to model the large-scale flow with a numerical model it may even be that our model isn't powerful enough to *resolve* all the small turbulent wiggles and whirls .

Regardless of the motivation, so whether we are really not interested in the details of all the turbulent whirls or we simply don't have a big enough computer to resolved them, what we often do is this: we use conservation equations that have been *averaged* over either the time or space scales of the turbulence (or preferably both), so that they only describe the time evolution of motions that are slower and larger than the turbulence itself.

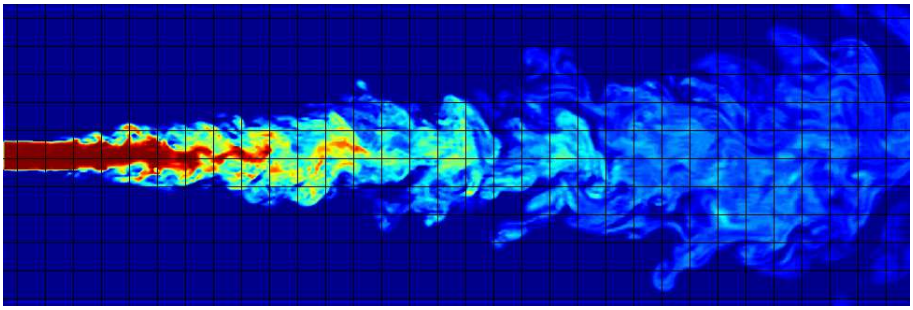

Figure 41: A turbulent jet of freshwater (red color) shooting out into a salty (blue) ocean region. The grid shown can be thought of as the numerical mesh of a numerical model used to model the jet.

Let's go back to the turbulent plume in Figure 41 and imagine how it could be represented in a numerical ocean model. All such models solve the governing differential equations on a *computational mesh* that divides the total domain into small cells. Such a mesh is shown in the figure. The conservation equations are written for each cell that make up the mesh. Since the cells each have a finite size, so that we're not dealing with infinitesimal but rather with finite distances, the differential equations become *difference equations*. We'll discuss numerical models in more detail later, but for now the point is that the flow field can only be resolved down to a certain scale, i.e. the scale of the grid cells. And, as in the case of the mesh shown in Figure 41, there will typically be turbulence that is smaller than the size of the grid cells. How can the *net* effects of the unresolved turbulence be represented in such a model?

What we do is to decompose all the dynamical variables pertaining to a given grid cell into *mean* and *perturbation* parts. The mean of a variable is a spatial average of that variable (say, salinity) over the size of the grid cell.[12] The perturbation is then meant to represent all the deviations of the variable in the grid cell from its grid-cell mean. So, for example, the flow and salinity field pertaining to the flow are written as

$$\begin{aligned} \boldsymbol{v} &= \overline{\boldsymbol{v}} + \boldsymbol{v}' \\ S &= \overline{S} + S' \end{aligned}$$

[12]Strictly speaking it is also an 'ensemble' mean over many realizations of the turbulence that could take place within such a grid cell.

where $\overline{\boldsymbol{v}}$ and $\overline{S}$ are the means taken over area A of the grid cell,

$$\begin{aligned}\overline{\boldsymbol{v}} &= \frac{1}{A}\iint \boldsymbol{v}\, dA \\ \overline{S} &= \frac{1}{A}\iint S\, dA,\end{aligned}$$

and the perturbations are then simply defined from

$$\begin{aligned}\boldsymbol{v}' &= \boldsymbol{v} - \overline{\boldsymbol{v}} \\ S' &= S - \overline{S}.\end{aligned}$$

This type of averaging, over the scales of the turbulent motion (here we assume the turbulence is taking place at scales smaller than the size of a grid cell in our numerical model), is called *Reynolds averaging*. And the separation of a variable into mean and perturbation parts is called a *Reynolds decomposition*.

Notice one important thing: by definition, the averages of the perturbations are zero, or $\overline{\boldsymbol{v}'} = 0$ and $\overline{S'} = 0$. Let's show that explicitly for salinity:

$$\begin{aligned}\overline{S'} &= \overline{S - \overline{S}} \\ &= \overline{S} - \overline{\overline{S}} \\ &= \frac{1}{A}\iint S\, dA - \frac{1}{A}\iint \overline{S}\, dA \\ &= \frac{1}{A}\iint S\, dA - \overline{S}\,\frac{1}{A}\iint dA \\ &= \overline{S} - \overline{S} \\ &= 0.\end{aligned}$$

Note that the mean salinity $\overline{S}$ is a constant over the grid cell (by how we've defined it) so that it could be taken outside the last averaging integral. Essentially, the average of a constant is just the same constant.

But let's get to the point and observe what happens when we Reynolds average the entire salinity equation. Plugging our two-component representations of velocity and salinity into the equtation and then averaging the equation itself gives

$$\frac{\partial \overline{S}}{\partial t} = -\nabla \cdot \left[\overline{(\overline{\boldsymbol{v}} + \boldsymbol{v}')\left(\overline{S} + S'\right)} - \kappa_S \nabla \overline{S}\right],$$

where we have used the facts that $\overline{S'} = 0$ and $\overline{\boldsymbol{v}} = 0$. The advection of salinity by the flow field in this expression consists of the four terms

$$\overline{(\overline{\boldsymbol{v}} + \boldsymbol{v}')\left(\overline{S} + S'\right)} = \overline{\boldsymbol{v}}\,\overline{S} + \overline{\overline{\boldsymbol{v}}\,S'} + \overline{\boldsymbol{v}'\,\overline{S}} + \overline{\boldsymbol{v}'S'}.$$

But, as it turns out, $\overline{\overline{\boldsymbol{v}}S'} = \overline{\boldsymbol{v}}\overline{S'} = 0$ and $\overline{\boldsymbol{v}'\overline{S}} = \overline{\boldsymbol{v}'}S = 0$ (show this!). So we are left with

$$\frac{\partial \overline{S}}{\partial t} = -\nabla \cdot \left(\overline{\boldsymbol{v}}\overline{S} + \overline{\boldsymbol{v}'S'} - \kappa_S \nabla \overline{S} \right).$$

So, in the Reynolds-averaged salinity equation there is an advective transport of the mean salinity field (Reynolds-averaged) by the mean velocity field. But there is also an extra transport term, the the *Reynolds flux* of salinity, $\overline{\boldsymbol{v}'S'}$. This is a net transport of salinity due to the turbulent motion. The turbulence which we thought we had averaged out (remember, $\overline{\boldsymbol{v}'} = 0$ and $\overline{S'} = 0$) can actually impact the evolution of larger scales, and this is what we see in Figure 41. It's pretty apparent there that the turbulent motions tend to transport the fresh water out from the plume or, alternatively, salty water into the plume. The time-mean velocity is predominately directed along the jet axis, so the mean-flow advection of mean salinity $\overline{\boldsymbol{v}}\overline{S}$ is probably not responsible for what actually looks like outward diffusion of the fresh water. It's the turbulent Reynolds fluxes that are responsible.

The turbulent transport in Figure 41 seems to be acting like enhanced molecular diffusion, transporting properties (salinity here) down the mean concentration gradient. Precicely for this reason, the most typical way to model, or *parametrize*, turbulent fluxes is as *enhanced diffusion*. So we write

$$\overline{\boldsymbol{v}'S'} = -K\nabla S,$$

where K is a turbulent diffusion coefficient or a *turbulent diffusivity* which is orders of magnitude larger than the molecular diffusivity. So much bigger in fact that the molecular diffusion term is simply ignored in the equations and replaced by the turbulent equivalent.

Note that, whereas the molecular diffusion coefficients are intrincic *properties of the fluid*, the turbulent diffusivities are *properties of the flow*, varying in time and space depending on the evolution of the flow field (i.e. the intensity of the turbulence). And, importantly, in a stably-stratified fluid vertical turbulent diffusion is much more difficult, or energetically costly, than horizontal diffusion[13]. That means that horizontal turbulent diffusivities should be much larger than vertical diffusivities. So in terms of components,

$$\overline{\boldsymbol{v}'S'} = -K_H \frac{\partial S}{\partial x}\boldsymbol{i} - K_H \frac{\partial S}{\partial y}\boldsymbol{j} - K_V \frac{\partial S}{\partial z}\boldsymbol{k},$$

[13]Recall our discussion about energetics from Chapter 2: Strictly speaking, when the density stratification is tilted it is only mixing at angles steeper than this tilt that is costly. But the tilt of the stratification is very slight, so we usually make the distinction between vertical and horizontal mixing.

one typically uses K_H of size 10^2–$10^3\,\mathrm{m^2 s^{-1}}$ (often assuming horizonal isotropy, i.e. no difference between the x and y directions) and K_V of size 10^{-5}–$10^{-3}\,\mathrm{m^2 s^{-1}}$ (so five to eight orders of magnitude smaller than the turbulent horizontal diffusivity!). With this, the Reynolds-averaged salinity and temperature equations, with molecular diffusivities replaced by their turbulent conterparts, become

$$\begin{aligned}
\frac{DS}{Dt} &= \nabla_H \cdot (K_H \nabla_H S) + \frac{\partial}{\partial z}\left(K_V \frac{\partial S}{\partial z}\right) \\
\frac{D\theta}{Dt} &= \nabla_H \cdot (K_H \nabla_H \theta) + \frac{\partial}{\partial z}\left(K_V \frac{\partial \theta}{\partial z}\right) - \nabla \cdot \boldsymbol{J}_R,
\end{aligned}$$

where, in light of the differences between horizontal and vertical directions, we have introduced the horizontal gradient operator

$$\nabla_H = \frac{\partial}{\partial x}\boldsymbol{i} + \frac{\partial}{\partial y}\boldsymbol{j}.$$

Note that we have also assumed that the turbulent diffusivities of salt and heat are the same since they are flow-dependent rather than being intrinsic properties of the property being transported (heat or salt). And since they are flow-dependent and may therefore vary from place to place, we have had to move them back inside the outer derivative.

The Reynolds-averaged momentum equations are derived using the same arguments used above, giving

$$\begin{aligned}
\frac{Du}{Dt} + f_* w - f v &= -\frac{1}{\rho_0}\frac{\partial p}{\partial x} + \nabla_H \cdot (A_H \nabla_H u) + \frac{\partial}{\partial z}\left(A_V \frac{\partial u}{\partial z}\right) \\
\frac{Dv}{Dt} + f u &= -\frac{1}{\rho_0}\frac{\partial p}{\partial y} + \nabla_H \cdot (A_H \nabla_H v) + \frac{\partial}{\partial z}\left(A_V \frac{\partial v}{\partial z}\right) \\
\frac{Dw}{Dt} - f_* u &= -\frac{1}{\rho_0}\frac{\partial p}{\partial z} - \left(1 + \frac{\rho'}{\rho_0}\right) g + \nabla_H \cdot (A_H \nabla_H w) + \frac{\partial}{\partial z}\left(A_V \frac{\partial w}{\partial z}\right),
\end{aligned}$$

where the use of A_H and A_V instead of K_H and K_V acknowledges that turbulent diffusion of momentum may actually be distinct from diffusion of tracers like salinity and temperature.

A little extra:

Exactly how turbulent diffusivities vary with the flow field is a topic of very active oceanographic research today, and we can only scratch at the surface here. But it is worth mentioning that the diffusivities are typically

related to the level of turbulence in the flow and, more specifically, to how *unstable* the flow is (it is instability that creates turbulence, just as static instability lies behind vertical convection). Here we will only look at one parameter of the flow which is thought to impact the level of vertical turbulent transport, i.e. vertical diffusivities. The *gradient Richardson number* is a relative measure of the strength of the vertical velocity shear $\partial|\boldsymbol{u}|/\partial z$ (where $|\boldsymbol{u}| = \sqrt{u^2 + v^2}$) to the strength of the vertical density stratification $\partial\rho/\partial z$. The stratification is written in terms of the buoyancy frequency (we introduced this in Chapter 2), giving

$$\begin{aligned} R_i &= \frac{N^2}{(\partial|\boldsymbol{u}|/\partial z)^2} \\ &= -\frac{g}{\rho_0}\frac{\partial\rho/\partial z}{(\partial|\boldsymbol{u}|/\partial z)^2}. \end{aligned}$$

So what does this ratio tell us about the turbulence level? Well, the vertical velocity shear is a source of instability and hence of turbulence. Think about how just about any velocity shear, e.g. the sheared flow near a wall, tends to create turbulent whirls. Plainly, the stronger the shear, the stronger the tendency for turbulence. The vertical density stratification, in contrast, can both stabilize and destabilize the flow. For $N^2 > 0$ (implying $\partial\rho/\partial z < 0$), the fluid is stably stratified—which, as we have discussed before, tends to inhibit vertical exchanges. So for stably-stratified flows, there is a competition between the vertical velocity shear which tends to create turbulence and the stratification which tends to suppress turbulence. But for statically unstable flows, i.e. for $N^2 < 0$, the top-heavy stratification is itself a source of instability and turbulence.

Laboratory experiments have shown that flows become turbulent approximately when $R_i < 0.25$, and many turbulent mixing parametrizations hence make the vertical diffusivities a function of R_i. Essentially, the lower the gradient Richardson number is the higher can one expect vertical diffusivities to be.

5 Observing and modeling the ocean

The previous chapters have identified the dynamical equations and physical variables that define the ocean state. The dynamical variables are density, pressure and the three velocity components. Density itself is, as we discussed in Chapter 2, calculated via an equation of state from salinity, temperature and pressure. Finally, as we saw in Chapter 3, the temperature and salinity of the oceans are partically determined by fluxes of heat and freshwater through the sea surface.

So there is a lot to keep track of! Our knowledge and understanding of the ocean state and its evolution rely on a combination of observing and modelling the relationship and interplay between these variables. Observations are the fundamental starting point. They show us what the 'solution' of the dynamical equations should look like. So even if obtaining ocean observations is pretty expensive, it is also extremely important. But modeling is also key. It is our way of showing that we understand what's going on. But the the dynamical equations are complicated (they are coupled and they are nonlinear), so obtaining analytical solutions is next to impossible unless the problems and equations are simplified dramatically. For *realistic* modelling, we have to rely on computers that can solve the equations by brute force.

5.1 Observation techniques

5.1.1 Temperature, salinity and pressure

Historically, water temperature was measured with reversing mercury thermometers and salinity was measured by chemical titration of water samples collected by botteles lowered into the sea. The thermometers and bottles were clamped to wires, and as these were lowered into the sea the thermometers were in a state where mercury could flow freely, expanding and contracting according to temperature, and the bottles were open so sea water could flow through (Figure 42). At the required depth, and after a period allowing for the thermometer to equlilibrate, signals were sent down to reverse the thermometers (to close the mercury flow) and close the bottles (to encapsule the water sample). The thermometers and water samples were then brought up to the ship for analysis. The pressure at the observation depth was typically not known with great accuracy and the depth was estimated by keeping track of the amount of wire sent out.

Today temperature, salinity and pressure are measured to great accuracy by electrical instruments bundled together in so-called CTD (conductivity, tempera-

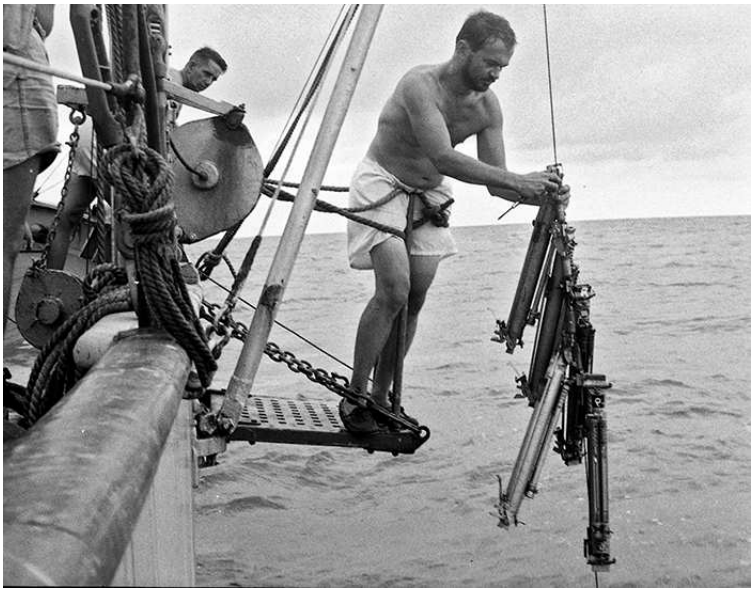

Figure 42: The deployment of a set of 'Nansen bottles' with reversing thermometers...from back in the days.

ture and depth) sensors. Temperature is measured directly using thermistors that contain materials whose electrical resistance is temperature-dependent. Salinity is measured indirectly, via the electrical conductivity of the sea water. Plainly, salty water is a better conductor than fresh water. And, finally, pressure is measured via piezoelectric elements, materials whos resistivity is pressure-dependent. Figure 43 shows the deployment of a 'rosette' of bottles that are used to collect water samples (for analysis of water chemical properties other than salinity). Several CTD sensors are also stripped on to the lower parts of the rosette frame.

So CTD sensors can be lowered into the oceans from ships. By doing repeated CTD profiles as the ship slowly progresses from position to position one can thus obtain information about lateral as well as vertical temperature and salinity gradients. Alternatively, CTD instruments can be moored at fixed locations (see below) to obtain time series of temperature and salinity. Finally, hybrid observational methods exist, as with so-called profiling floats that drift freely around with the ocean currents while collecting CTD profiles at pre-programmed intervals. CTD sensors can even be attached to large animals, like seals (see Figure 44). Utilizing animals has proven useful to obtain hydrographic observations from places that are otherwise difficult to reach, for example the marginal ice zone at high

Figure 43: A rosette of sampling bottles with CTD sensors also attached to the rosette frame.

Figure 44: A seal with a small CTD package (including satellite transmitter) attached to its fur. (Source: http://www.afsc.noaa.gov/quarterly/ond2011/divrptsNMML3.htm)

latitudes (Figure 45).

Finally, today the sea surface temperature (SST) and sea surface salinity (SSS) of the world oceans can be measured by remote-sensing instruments attached to orbiting satellites. The satellites measure the radiation emitted from the ocean surface at various wavelengths, and it is from this radiation that SST and SSS (and other properties of the ocean surface, for example its color) is deduced. But the radiation eventually received by the satellites has also passed through an atmosphere, and this 'contamination' needs to be subtracted from the signals. The *radiative transfer models* used for such corrections are based on many assumptions about the composition of the atmosphere (e.g. its water vapor content) and therefore have many potential sources of errors. But calibration of the satelite data against *in situ* observations (e.g. from ship campaigns) are constantly improving the satellite products. Besides, the unprecedented spatial and temporal coverage offered by satellite observations relative to *in situ* observations make them tremendously valuable. Examples of satellite-derived SST and SSS has already been shown in Figures 12 and 11.

5.1.2 Velocity

Observations of ocean currents can be obtained by several methods. One is to simply drop things into the ocean and observe in what direction and with what speeds they drift. This Lagrangian approach sounds simple but has actually pro-

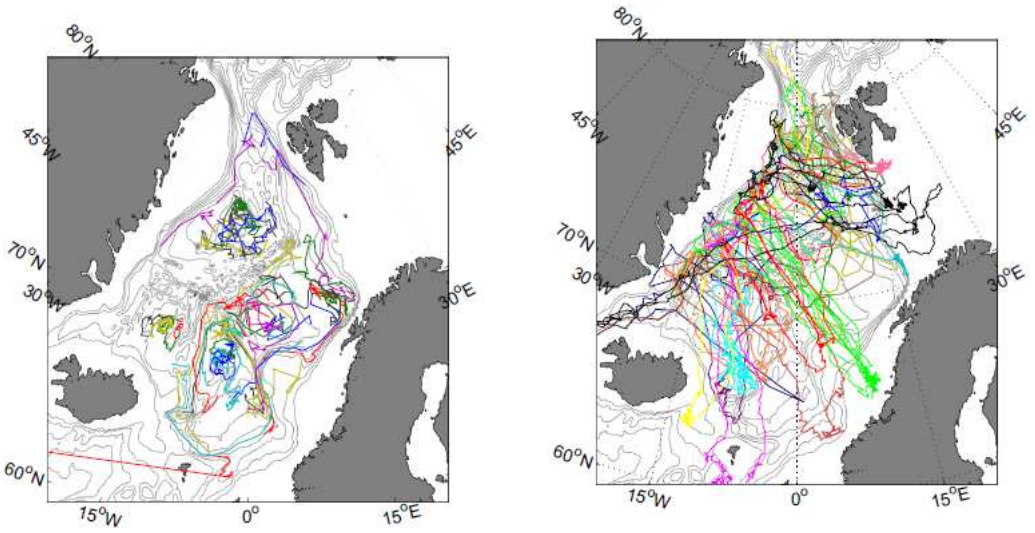

Figure 45: Tracks of profiling floats (left) and instrumented seals (right) collected by the multinational collaborative MEOP project. (Source: Isachsen et al., 2014, Figs. 5 and 6)

vided invaluable observations of currents in the world oceans via for example the World Ocean Drifter Programme (Figure 46). Ocean drifters contain a floatation device and often a drogue or sail at depth, typically at 15 meters, to make sure the drifter is following the ocean currents and not the wind.

Another way to observe currents is by deploying current meters at fixed locations to obtain Eulerian observations. Current speeds were traditionally measured by the rotation rate of propellors, and current directions were measured by the rotation angle of the instruments as they were allowed to adjust to the flow. Modern-day instruments measure the flow velocity by doppler methods, i.e. by the phase shift induced in acoustic waves as they travel from a transmitter to a receiver through the water that flows in between the two. Such current meters are typically attached, along with other instruments like CTDs, to oceanographic *moorings* (Figure 47). This approach gives continuous observations (at least as long as batteries last), but only for a limited number of fixed points in space, one for each instrument. Continuous vertical profiles of velocity can also be obtained from so-called Acoustic Doppler Current Profilers (ADCP). These consists of strong sound transmitters (transducers) that emit sound waves into the ocean. Currents can then be estimated from the doppler shift of the sound waves reflected from various depths of the water column.

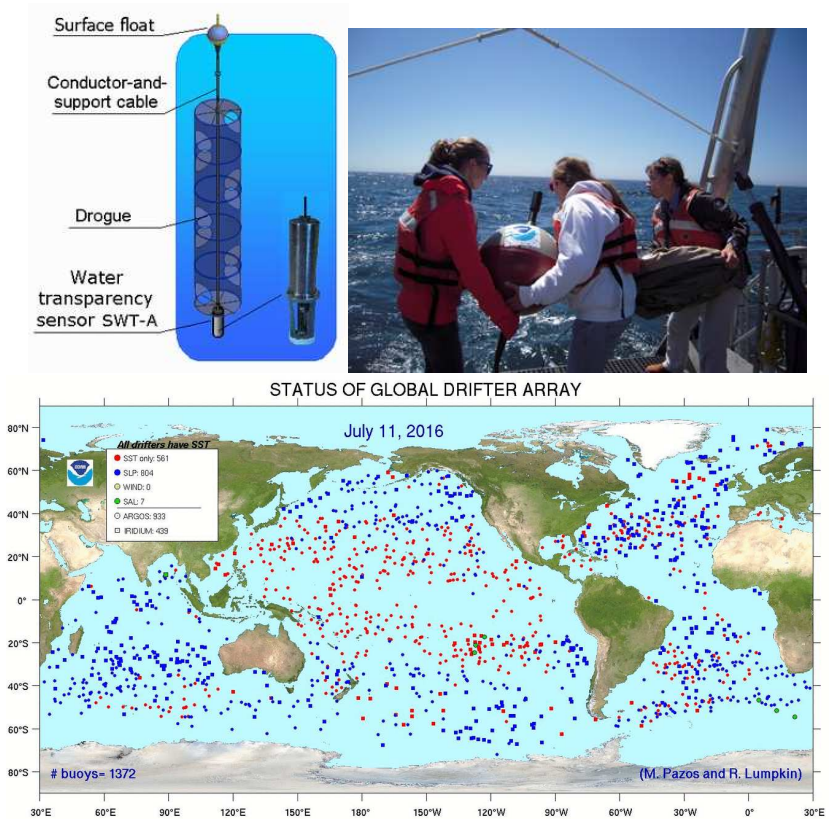

Figure 46: The global surface drifter program: (top) a typical surface drifter and its deployment, and (bottom) position of drifters at one given day in 2016. (Source: http://www.aoml.noaa.gov/phod/dac)

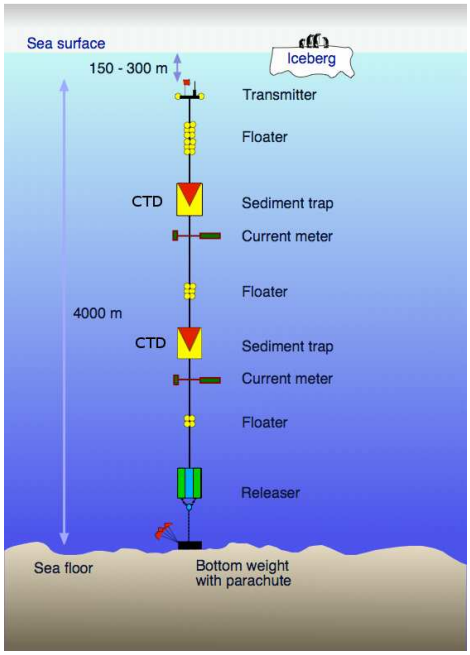

Figure 47: A modern current meter (top) and a schematic of an oceanographic mooring (left) containting current meters, CTDs and (in this particular case) sediment traps. A heavy anchor keeps the mooring in place while floatation devices keep it upright. Just above the anchor is a remote-controlled release mechanism which allows the rest of the mooring to be recovered for data retrieval and instrument reuse.

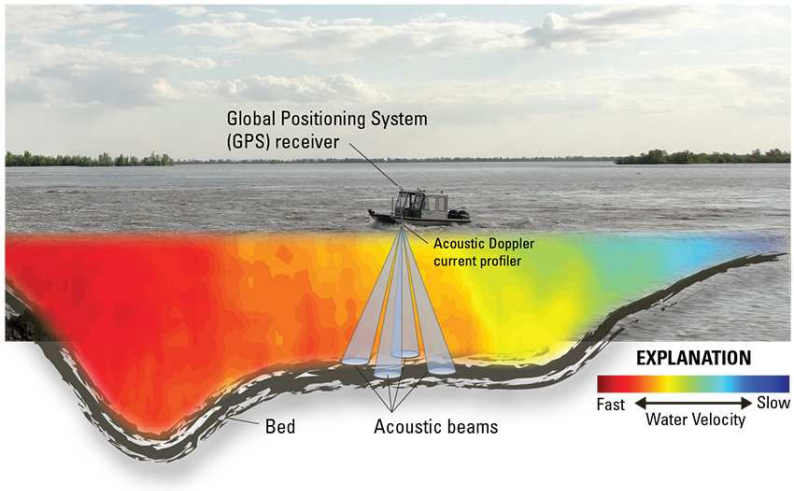

Figure 48: Acoustic Doppler Current Profiler (ADCP) observations of ocean currents.

5.1.3 Sea level

Observations of the sea level is not only important in studies of long-term changes like those associated with climate variability and climate change. Historically speaking, sea level observations were always made to calibrate tide prediction charts (the theory of tides is described in a later chapter). The measurement technology is simple in principle, involving some kind of floatation device which is allowed to move up and down while also recording its vertical position. But such observations have, for obvious reasons, always been limited to coastal stations. Today we obtain observations about sea level variability also in the open oceans from satellite *altimeter* instruments. In fact, satellite observations of sea level has in many ways revolutionized oceanographic research since such techniques became available in the late 1970s. So it's worth our while to take a slightly closer look at the basis for such observations. Altimeter satellites have tremendously good positioning systems that give their positions relative to the fixed stars throughout their orbits around Earth. They also have very good radar sensors that can measure their distance from the sea surface. Combining these two pieces of observations it is possible to deduce the sea surface height with a precision of a couple of centimeters.

As it turns out, large parts of the spatial variations of the sea surface height that the satellites observe are arrise only from local variations in the effective

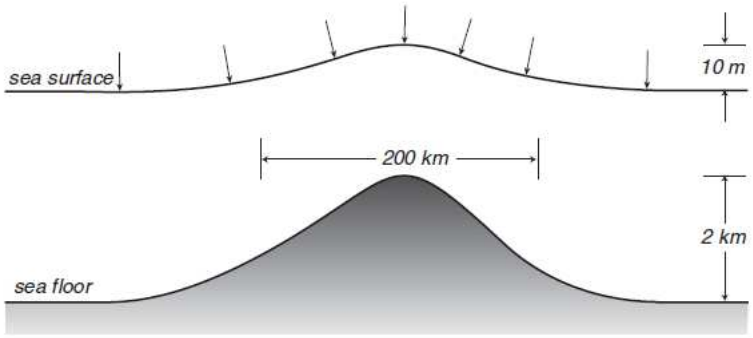

Figure 49: The sea level variation on top of a seamount due to gravitational variations.

gravitational field and do not tell us anything about the ocean state. This is the ocean *geoid*:

> "The geoid is the shape that the surface of the oceans would take under the influence of Earth's gravitation and rotation alone, in the absence of other influences such as winds and tides. All points on the geoid have the same gravity potential energy (the sum of gravitational potential energy and centrifugal potential energy). The force of gravity acts everywhere perpendicular to the geoid, meaning that plumb lines point perpendicular and water levels parallel to the geoid." (Source: https://en.wikipedia.org/wiki/Geoid)

As illustrated in Figure 49, a sea mount deep down in the ocean will create a bulge in the sea surface above it. But a particle residing on the geoid, even if the geoid is inclined like shown in the figure, feels no net force in any direction.

So the geoid is *dynamically* irrelevant (it has no impact on the ocean circulation) and must actually be subtracted from the altimeter sea level height observations before these are used to estimate for example sea level gradients that go into the momentum equations. In other words, the *dynamic topography* which actually impacts the flow—as water will tend to flow down gradients of dynamic topography—is given by

$$\eta = \eta_a - \eta_g,$$

where η_a is the 'raw' sea level observed by the satellite and η_g is the geoid (as illustrated in Figure 50).

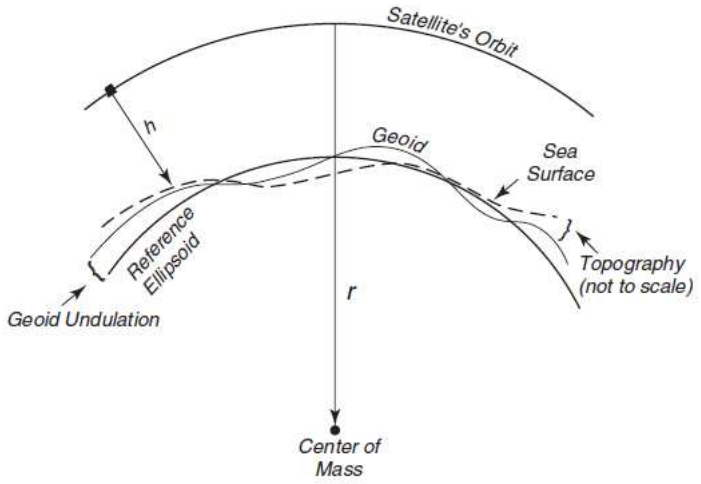

Figure 50: Satellite altimeters measure the height of the sea surface relative to a reference ellipsoid. This sea surface height contains a large contribution from the geoid which is caused by geographical mass variations and which is dynamically (Source: Stewart, 2008, Fig. 3.13)

So to make good estimates of the time-mean dynamic topography and hence time-mean ocean currents we need a good knowledge of Earth's geoid. As it turns out, observing and modelling the geoid is a hole research field in itself. But to us oceanographers it is worth remembering that the geoid doesn't change over time (at least not over the time scales that we humans normally care about). So whereas we have to be careful about subtracting the geoid to make reliable estimates of time-mean currents, we don't have to do so if we are only interested in observing how these currents change with time.

5.1.4 Air-sea fluxes

Air-sea fluxes of importance for the dynamics include radiative and turbulent heat fluxes, freshwater fluxes and momentum fluxes. Most of these fluxes depend on observations of temperature, humidity, wind, precipitation and radiation collected in the lower atmosphere. They are typically collected by instrument placed very near the sea surface and, in some cases when vertical gradients are needed, also some meters above. Then the fluxes, calculated from these atmospheric observations, are applied as *boundary conditions* at $z = 0$ to the equations of motion for the ocean.

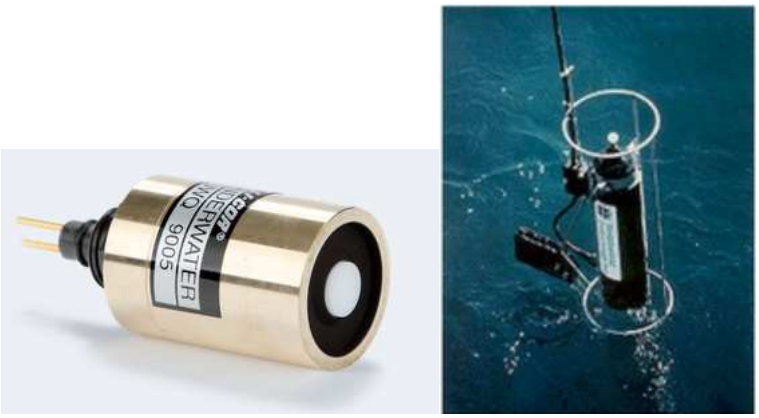

Figure 51: Sensor for measuring quantum irradiance or photosynthetically active radiation (PAR).

The detailed 'fate' of shortwave raditaion as it enters the ocean may sometimes be of interest. We have seen how such radiation can penetrate some tens of meters into the water column and that this depth impacts both the upper ocean heat budget and biological production. As it turns out, estimates of the shortwave penetration depth can be obtained with both advanced and less advanced methods. Very accurate measurements of downward shortwave irradiance or *photosynthetically active radiation* (PAR) can be obtained by a sensor lowered down through the water column (Figure 51). It 'looks up' and detects downward shortwave radiation, integrating over the wavelengths from about 400 to 750 nanometers, and reports back either an energy flux density (in $\mathrm{W\,m^{-2}}$) or a light quantum flux density (in $\mu\mathrm{mol\,s^{-1}m^{-2}}$).

A somewhat less fancy method is based on lowering a metalic circular disk, a *Secci disk* (Figure 52), and then observing, from the deck of the ship, the depth at which the disk cannot be seen anymore. A rule of thumb is then that the downward irradiance is 1% of the surface value at a depth of approximately two 'secci depths'.

5.2 Numerical ocean modeling

5.2.1 From differential equations to difference equations

Numerical ocean models solve *discrete* versions of the governing equations. This means that both time and space are discretized, and the continuous derivatives

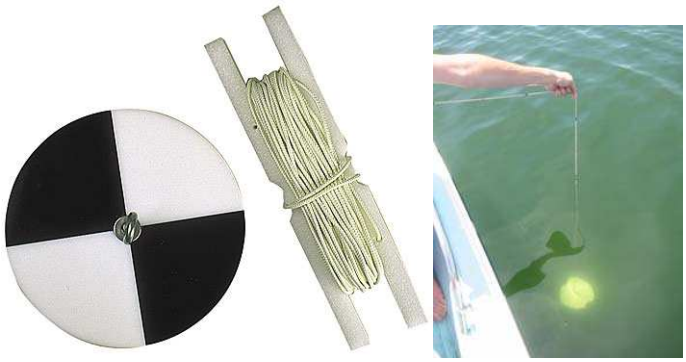

Figure 52: The Secci disk: a slightly less accruate way to measure shortwave penetration depth.

become discrete differences. The differential equations derived in the last chapter become *difference equations*. And these are solved on computational grids like those shown in Figure 53.

Time itself is discretized. So the time at step n is

$$t_n = t_0 + n\Delta t,$$

where t_0 is a start time and Δt is a time step. Hence a variable in the ocean which is actually continuous in time will, in the computer, be represented by a finite set of points, as illustrated in Figure 54.

So a continuous time derivative, which is formally defined as

$$\frac{du}{dt} = \lim_{\Delta t \to 0} \frac{u_{n+1} - u_n}{\Delta t},$$

can be approximated by the *time difference* on the right hand side—for some small time step Δt. But on a computer Δt can not shrink to zero and will have to take on some finite value. Clearly the discrete approximation of the continuous derivative becomes better the smaller Δt is. We can quantify the error by applying a Taylor series expansion to the function at t_n (t^n in the figure):

$$u_{n+1} = u_n + \frac{du}{dt}\Delta t + \frac{1}{2!}\frac{d^2u}{dt^2}\Delta t^2 + \frac{1}{3!}\frac{d^3u}{dt^3}\Delta t^3 + \ldots$$

where higher-order terms have been omitted. Putting u_n on the left hand side and then dividing by Δt gives

$$\frac{u_{n+1} - u_n}{\Delta t} \quad = \quad \frac{du}{dt} + \mathcal{O}\left(\Delta t\right).$$

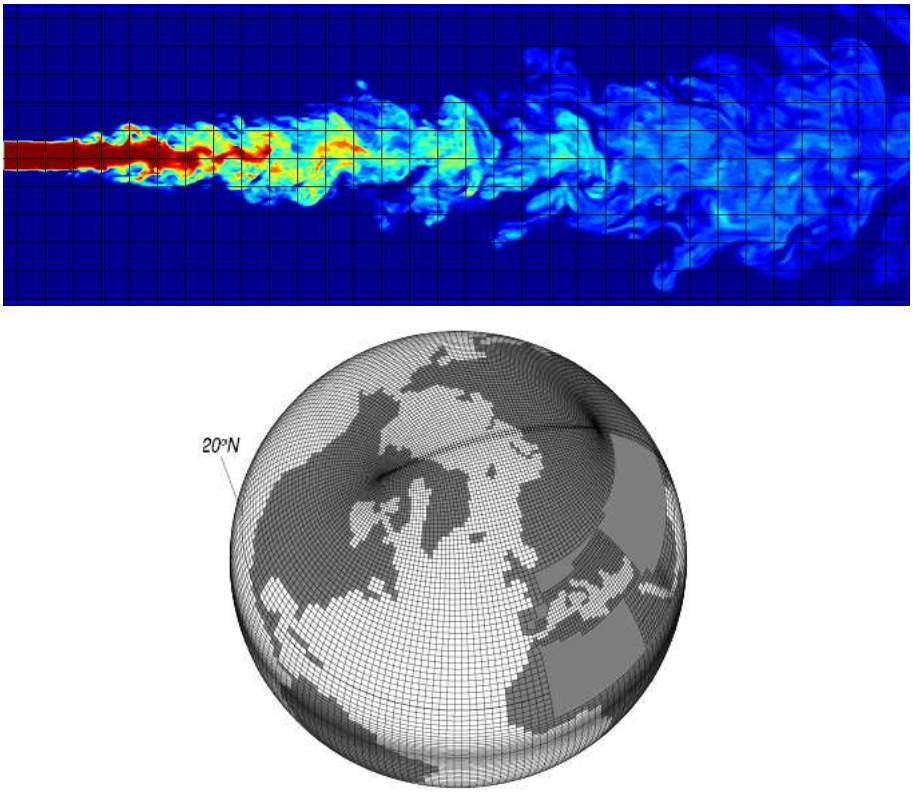

Figure 53: Examples of computational grids used to model (top) a turbulent jet and (bottom) the global ocean circulation.

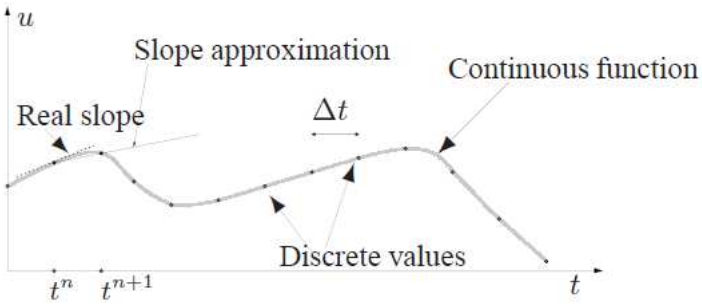

Figure 54: Time discretization in a numerical ocean model. (Source: Cushman-Roisin and Beckers, 2011, Fig. 1.12)

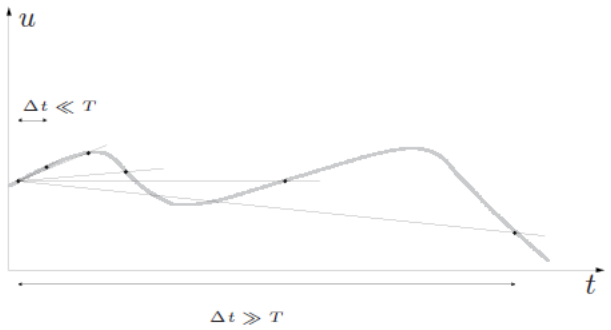

Figure 55: The discrete model representation of the derivative of a continuous function for different time steps Δt (in relation to the intrincic time scale T of the process). (Source: Cushman-Roisin and Beckers, 2011, Fig. 1.13)

So the error is proportional to the time step Δt. If the process in the ocean we wish to study has an intrincic time scale T associated with it, then it's crucial that the model time step is much smaller than this time scale i.e. that $\Delta t \ll T$, in order to model the process correctly. A failure to do so may cause a complete wrong representation of the process, as illustrated in Figure 55.

It is possible to achieve better accuracy for a given time step by creating *higher-order difference schemes* for the approximation to the derivatives. If one, for example, writes the Taylor series expansion for both u_{n+1} and u_{n-1} (backwards in time)

$$\begin{aligned} u_{n+1} &= u_n + \frac{du}{dt}\Delta t + \frac{1}{2!}\frac{d^2u}{dt^2}\Delta t^2 + \frac{1}{3!}\frac{d^3u}{dt^3}\Delta t^3 + \ldots \\ u_{n-1} &= u_n - \frac{du}{dt}\Delta t + \frac{1}{2!}\frac{d^2u}{dt^2}\Delta t^2 - \frac{1}{3!}\frac{d^3u}{dt^3}\Delta t^3 + \ldots \end{aligned}$$

and then subtract these two equations, one gets (after dividing by Δt),

$$\frac{u_{n+1} - u_{n-1}}{2\Delta t} = \frac{du}{dt} + \mathcal{O}\left(\Delta t^2\right).$$

So the error is now proportional to the square of the time step. Thus, for small time steps the error of this 'second-order' scheme is smaller than for the previous ('first-order') scheme.

Here we have discussed finite differencing with respect to time. But, of course, the same applies to spatial derivatives. So, if we have a spatial grid in the x, y and z

directions, then a second-order accurate spatial difference in the x direction would be

$$\frac{u_{i+1} - u_{i-1}}{2\Delta x} = \frac{du}{dx} + \mathcal{O}\left(\Delta x^2\right),$$

where subscript i indexes the x position.

A little extra:

5.2.2 Data assimilation: combining observations and model

The ability of numerical models to provide correct forecasts is limited by many factors, such as insufficient resolution of the numerical grid, artefacts introduced by the discretization of the governing equations, inaccurate description of the physics, and so on. One of the major limitations to predictability is, however, inaccurate initial conditions: we need to accurately know what the state of the ocean is now if we want to accurately predict the state of the ocean in the future.

Data assimilation is the process of combining observations and a model to provide a best estimate of the initial conditions for subsequent model runs. There are many different data assimilation techniques, with widely varying levels of complexity, and we will not go into any details here. The concept will be illustrated using a simple example of a model for the temperature T at a single point.

We denote the observation value T_o and the model value T_m. What we want to find is the best possible estimate of the true temperature T_t. This estimate is called the *analysis*, T_a. Unfortunately, there are no perfect observations and no perfect model. The implication here is that we will never know *exactly* what the true temperature is. What we can do is to estimate the errors in the observations, in the model values, and the analysis, and use these estimates to find an expression for T_a.

If we let $< >$ denote an average, then we can define error variances as

$$\begin{aligned} e_o^2 &= < (T_o - T_t)^2 >, \\ e_m^2 &= < (T_m - T_t)^2 >, \\ e_a^2 &= < (T_a - T_t)^2 > . \end{aligned}$$

We now assume that the observations, the model values and the analyses are

unbiased, that is, $< T_o - T_t >=< T_m - T_t >=< T_a - T_t >= 0$. This assumption implies that the model and the observations are *on average* correct, which may be far from the truth in realistic applications! We also assume that the errors in the observations and the model values are uncorrelated, that is, $< (T_o - T_t)(T_m - T_t) >= 0$. The latter assumption makes sense, there should not be any reason for an error in the observation to be connected to the model, or vice versa. It can then be shown that in order to minimize the analysis error variance e_a^2, we must obtain the analysis from the following equation:

$$T_a = T_m + K(T_o - T_m), \text{ where } K = \frac{e_m^2}{e_m^2 + e_o^2}.$$

We can consider the two extreme cases where (i) the model error is expected to be large compared to the observation error, and (ii) the other way around. In the first case we have $e_m^2 \gg e_o^2$, and hence $K \approx 1$. This is consistent with giving more weight to the observation, and according to the analysis equation we have $T_a \approx T_o$. In the second case, with $e_m^2 \ll e_o^2$, we have much less confidence in the observation compared to the model and we find that $K \approx 0$ and the analysis is $T_a \approx T_m$ as expected.

One of the main challenges in data assimilation is to provide reasonable estimates of the error variances—a nontrivial task since the true state in any geophysical system is unknown.

6 Simplified equations valid for large-scale flows

There should be little doubt that the equations derived in Chapter 4 are complicated. They are coupled and they are nonlinear. Finding solutions to them will rely on numerical methods (using computers), and even then the equations are challenging to handle. To make progress, *especially* if we wish to seek analytic solutions that give us intuition about the ocean circulation, we need to simplify them drastically. We basically need to get rid terms, like we did for the mass equation when making the Boussinesq approximation. The trick is to assess which terms are much smaller than other terms and can therefore be thrown out. What terms are expected to be small depends on the situation or process we wish to study, i.e. on what temporal and spatial scale characterize our problem of interest. In this and the following two chapters we will be interested in large-scale flows, and this regime allows certain simplifications that makes both numerical modelling and theoretical 'playing around' somewhat easier.

6.1 Defining large-scale geophysical flows

Below we will show that additional simplifications can be made if we are interested in large-scale geophysical flows. By this we mean flows for which:

1. The horizontal scales of the flow are much larger than the vertical scale.

2. The rotation of the planet (the Coriolis accelleration) is important, meaning that it is at least as important as the other acceleration terms in balancing the horizontal pressure gradient.

Let's look into the first condition. Figure 56 shows a cross section of ocean bathymetry across the South Atlantic using two different exaggerations of vertical scales. Already at a vertical exaggeration of 30:1 it is pretty clear that horizontal scales are much larger than vertical scales. Ocean circulation features, e.g. ocean gyres, can have horizontal scales of up to thousands of kilometers but vertical scales of only hundreds or, at a maximum, thousands of meters. So the aspect ratio, i.e. the ratio of vertical scale D to the horizontal scale L, is indeed very small for such large-scale flows:

$$\delta \equiv \frac{D}{L} \ll 1.$$

To begin to see what this implies, we scale the terms of the continuity equation (the Boussinesq version of the mass conservation equation). We use U and W as

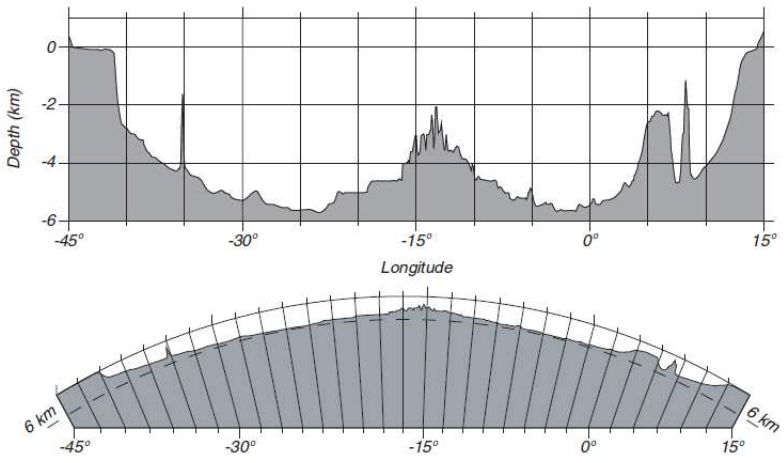

Figure 56: Cross section of ocean bathymetry in the South Atlantic along 25°S. The vertical exaggeration is 180:1 (top) and 30:1 (bottom). (Source: Stewart, 2008, Fig. 3.4)

scales for horizontal and vertical velocities, L and D for horizontal and vertical scales. This gives

$$\begin{aligned} \nabla_H \cdot \boldsymbol{u} &= -\frac{\partial w}{\partial z} \\ \frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} &= -\frac{\partial w}{\partial z} \\ \frac{U}{L}, \quad \frac{U}{L} &\sim -\frac{W}{D}, \end{aligned}$$

so that

$$\frac{W}{U} \sim \frac{D}{L} \equiv \delta \ll 1.$$

So for geophysical flows vertical velocities are typically much smaller than horizontal velocities. There are of course exceptions to this, but we'll come back to such exceptions later.

Then on to the second condition, that the Coriolis accelleration is a major player in balancing the horizontal pressure gradients. Our intuition says that pressure gradients should be important to drive any kind of flow—down the gradient. But as we have become accustomed to, the winds on the weather map are typically flowing around the high and low pressures, rather than across them. This is also the case for large-scale flows in the ocean, as illustrated in Figure 57. So when we

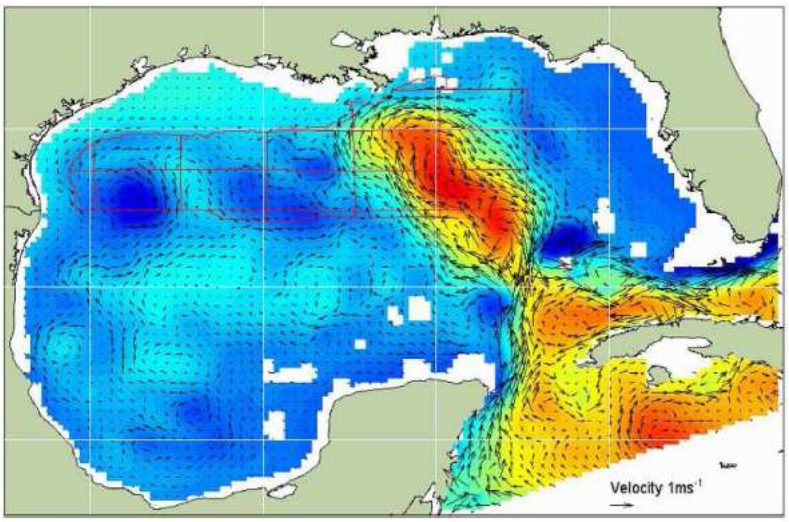

Figure 57: Sea surface height (color) and surface currents (arrows) in the Gulf of Mexico from a numerical ocean model simulation. (Source: Cushman-Roisin and Beckers, 2011, Fig. 4.5)

study large-scale motions, the Coriolis accelleration definitely needs to be taken seriously.

Now onto implications of these two requirements on the momentum equations. Here we use scales L, D, U and W as before but also introduce T for a time scale and P for a typical pressure scale. The sizes of the various terms in the horizontal momentum equations become

$$\begin{aligned} \frac{\partial u}{\partial t} + \boldsymbol{u}\cdot\nabla_H u + w\frac{\partial u}{\partial z} + f_* w - fv &= -\frac{1}{\rho_0}\frac{\partial p}{\partial x} - \nabla_H\cdot(-A_H\nabla_H u) - \frac{\partial}{\partial z}\left(-A_V\frac{\partial u}{\partial z}\right) \\ \frac{U}{T}, \quad \frac{U^2}{L}, \quad \frac{U^2}{L}, \quad f_*W, \quad fU &\sim \frac{P}{\rho_0 L}, \quad A_H\frac{U}{L^2}, \quad A_V\frac{U}{D^2} \end{aligned}$$

and

$$\begin{aligned} \frac{\partial v}{\partial t} + \boldsymbol{u}\cdot\nabla_H v + w\frac{\partial v}{\partial z} + fu &= -\frac{1}{\rho_0}\frac{\partial p}{\partial y} - \nabla_H\cdot(-A_H\nabla_H v) - \frac{\partial}{\partial z}\left(-A_V\frac{\partial v}{\partial z}\right) \\ \frac{U}{T}, \quad \frac{U^2}{L}, \quad \frac{U^2}{L}, \quad fU &\sim \frac{P}{\rho_0 L}, \quad A_H\frac{U}{L^2}, \quad A_V\frac{U}{D^2}, \end{aligned}$$

where we have already used the scaling of the continuity equation to write $W/D \sim U/L$. If we now assume that $f_* \sim f$ (this is true everywhere except for very

near the equator), we imediately see that $f_* w$ is small compared to fv in the x-momentum equation and can therefore be ignored. To make further simplifications we need to bring in the second requirement for large-scale flows, namely that Earth's rotation is important. We formalize this by expecting that the Coriolis accelleration is as large as the pressure gradient term, i.e.

$$fU \sim \frac{P}{\rho_0 L}$$

and that all other terms in the equation are either of the same size as Coriolis or smaller. In other words,

$$\frac{U}{T} \lesssim fU, \quad \frac{U^2}{L} \sim \frac{UW}{D} \lesssim fU, \quad A_H \frac{U}{L^2} \lesssim fU, \quad A_V \frac{U}{D^2} \lesssim fU.$$

But anticipating that they might be as large as Coriolis in some circumstances, we keep them all.

Scaling the vertical momentum equation gives

$$\begin{array}{rcl} \dfrac{\partial w}{\partial t} + \boldsymbol{u} \cdot \nabla w + w \dfrac{\partial w}{\partial z} - f_* u & = & -\dfrac{1}{\rho_0} \dfrac{\partial p}{\partial z} - \left(1 + \dfrac{\rho'}{\rho_0}\right) g - \nabla \cdot (-A_H \nabla w) - \dfrac{\partial}{\partial z} \left(-A_V \dfrac{\partial w}{\partial z}\right) \\ \dfrac{W}{T}, \quad \dfrac{UW}{L}, \quad \dfrac{UW}{L}, \quad f_* U & \sim & -\dfrac{P}{\rho_0 D}, \quad \left(1 + \dfrac{\rho'}{\rho_0}\right) g, \quad A_H \dfrac{W}{L^2}, \quad A_V \dfrac{W}{D^2}. \end{array}$$

The first thing to note is that since $W \ll U$, then all the three first accelleration terms as well as the two diffusion terms are all much smaller than the Coriolis term. But the Coriolis term itself turns out to be small compared to the vertical pressure gradient. This is because we have already assumed that the Coriolis accelleration approximately balances the *horizontal* pressure gradient. So the ratio of the Coriolis term to the vertical pressure gradient is

$$\frac{f_* U}{p/(\rho_0 D)} \sim \frac{D}{L} \ll 1.$$

After all these cancellations the only term left to balance the vertical pressure gradient is the gravity term. So the entire vertical momentum equation has been reduced to

$$0 = -\frac{1}{\rho_0} \frac{\partial p}{\partial z} - \left(1 + \frac{\rho'}{\rho_0}\right) g$$

or,

$$\frac{\partial p}{\partial z} = -\rho g,$$

where we still mean $\rho = \rho_0 + \rho'$.

What we have found is called the *hydrostatic approximation*—and we already saw it coming when studying the vertical momentum equation under the Boussinesq approximation in Chapter 4. Note that if the fluid is absolutely still at all times, then all the other terms in the vertical momentum equation are *identially* zero. We then have an exact *hydrostatic balance* between the vertical pressure gradient (pointing upward since the pressure is higher at depth) and the gravitational force (pointing downward). What we have just shown here is that at large scales this balance also approximately holds *even* if the fluid is moving. The pressure at any depth is basically the total weight of the fluid above.

6.2 The primitive equations

In summary, the equations relevant for large-scale flows are the three momentum equations (with the hydrostatic approximation applied to the z component),

$$\begin{aligned}
\frac{\partial u}{\partial t} + \boldsymbol{u}\cdot\nabla_H u + w\frac{\partial u}{\partial z} - fv &= -\frac{1}{\rho_0}\frac{\partial p}{\partial x} - \nabla_H\cdot(-A_H\nabla_H u) - \frac{\partial}{\partial z}\left(-A_V\frac{\partial u}{\partial z}\right)\\
\frac{\partial v}{\partial t} + \boldsymbol{u}\cdot\nabla_H v + w\frac{\partial v}{\partial z} + fu &= -\frac{1}{\rho_0}\frac{\partial p}{\partial y} - \nabla_H\cdot(-A_H\nabla_H v) - \frac{\partial}{\partial z}\left(-A_V\frac{\partial v}{\partial z}\right)\\
\frac{\partial p}{\partial z} &= -\rho g,
\end{aligned}$$

and the continuity equation,

$$\nabla_H\cdot\boldsymbol{u} + \frac{\partial w}{\partial z} = 0.$$

In a stratified ocean we also need the equations for conservation of salinity and temperature,

$$\begin{aligned}
\frac{DS}{Dt} &= -\nabla_H\cdot(-K_H\nabla_H S) - \frac{\partial}{\partial z}\left(-K_V\frac{\partial S}{\partial z}\right)\\
\frac{D\theta}{Dt} &= -\nabla_H\cdot(-K_H\nabla_H \theta) - \frac{\partial}{\partial z}\left(-K_V\frac{\partial \theta}{\partial z}\right) - \nabla\cdot\boldsymbol{J}_R,
\end{aligned}$$

and, finally, an equation of state which gives water density as a function of salinity and (potential) temperature,

$$\rho = \rho(S,\theta,p_0).$$

These equations, relevant when 1) horizontal scales are much larger than vertical scales and 2) the Coriolis accelleration is a key player, are called *the primitive equations*. They are the equations used by most numerical ocean models today and are also the starting point for most analytical treatments of ocean circulation problems.

If, however, we wish to model flows in which we believe the vertical scales of motions are comparable to the horizontal scales and flows for which we don't think Earth's rotation is a dominant player, we need to go one step back to use the full 'non-hydrostatic' version of the z-momentum equations. A good example of a non-hydrostatic flow is vertical convection. The vertical overturning in convection is so fast that vertical accelleration terms simply cannot be ignored. Note, however, that if the non-hydrostatic equationis are to be used, we also need, for consistency, to keep the $f_* w$ term in the x-momentum equation.

6.3 Estimating the hydrostatic pressure

Note that none of our equations so far have been an explicit equation for how pressure changes in time or space. But pressure can be determined in various ways, and with the primitive equations it is found by integrating the hydrostatic approximation in the vertical. We set our reference level $z = 0$ at the level of an imaginary flat sea surface. Then we let actual sea surface height variations around this reference be denoted by $\eta(x, y, t)$. Now integrating the hydrostatic pressure equation from any depth z below the reference level (remember that z is then a negative number) up to the sea surface at $z = \eta$ gives

$$p(z) = p(\eta) + \int_z^\eta \rho g \, dz$$

where $p(\eta)$ is the atmospheric pressure at the surface. So the pressure at any depth z is simply the weight of the water above plus the pressure at the sea surface (which in turn is often well approximated by the weight of the atmosphere above). If we ignore the atmospheric pressure for now and also expand the density $\rho = \rho_0 + \rho'$ into its two components (and noting that ρ_0 is depth-independent), we get

$$p(z) = \rho_0 g \, (\eta - z) + \int_z^\eta \rho' g \, dz.$$

Note now that the horizontal pressure gradients at level z (to be used in the horizontal momentum equations) then become

$$\begin{aligned} -\frac{1}{\rho_0}\frac{\partial p}{\partial x} &= -g\frac{\partial \eta}{\partial x} - \frac{g}{\rho_0}\frac{\partial}{\partial x}\int_z^\eta \rho'\, dz \\ -\frac{1}{\rho_0}\frac{\partial p}{\partial y} &= -g\frac{\partial \eta}{\partial y} - \frac{g}{\rho_0}\frac{\partial}{\partial y}\int_z^\eta \rho'\, dz. \end{aligned}$$

Finally, since the distance from $z = 0$ to $z = \eta$ is normally a very small part of the total vertical integral in the term involving ρ', this last bit is normally ignored, giving

$$-\frac{1}{\rho_0}\frac{\partial p}{\partial x} = -g\frac{\partial \eta}{\partial x} - \frac{g}{\rho_0}\frac{\partial}{\partial x}\int_z^0 \rho'\, dz \tag{6}$$

$$-\frac{1}{\rho_0}\frac{\partial p}{\partial y} = -g\frac{\partial \eta}{\partial y} - \frac{g}{\rho_0}\frac{\partial}{\partial y}\int_z^0 \rho'\, dz. \tag{7}$$

So the horizontal pressure gradient at depth is due to 1) the sea surface tilt and 2) horizontal gradients in the weight due to the perturbation density. The sea surface tilt term is depth-independent while the second term changes with depth and reflects horizontal density gradients, i.e. tilted or *inclined* isopycnals (lines of constant density). The two are often called the *barotropic* and *baroclinic* pressure gradients, although these two terms are mathematically defined by whether isopycnals are parallel to pressure surfaces (rather than z-surfaces) or not.

6.4 The shallow-water equations

For many applications of modelling or theorizing about the ocean circulation we can start by ignore density variations. So we essentially study a (thin) homogeneous layer where the density is constant, so that $\rho' = 0$. Then we don't need to carry along salinity and temperature equations (nor an equation of state), and we are left with three momentum equations (one of them is the hydrostatic approximation of the vertical component) and the continuity equation. Ignoring turbulent

now just for convenience, we have

$$\begin{aligned}
\frac{\partial u}{\partial t} + u\frac{\partial u}{\partial x} + v\frac{\partial u}{\partial y} + w\frac{\partial u}{\partial z} - fv &= -g\frac{\partial \eta}{\partial x} \\
\frac{\partial v}{\partial t} + u\frac{\partial v}{\partial x} + v\frac{\partial v}{\partial y} + w\frac{\partial v}{\partial z} + fu &= -g\frac{\partial \eta}{\partial y} \\
\frac{\partial p}{\partial z} &= -\rho_0 g \\
\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} + \frac{\partial w}{\partial z} &= 0.
\end{aligned}$$

But, in fact, the hydrostatic equation is no longer useful for us. We have already exploited it to give us horizontal pressure gradients—depth-independent (barotropic) pressure gradients. Let's look instead at the horizontal momentum equations and, specifically, how the horizontal velocities vary with depth. Taking the vertical derivative of the x-momentum equation gives

$$\frac{\partial}{\partial t}\left(\frac{\partial u}{\partial z}\right) + \frac{\partial}{\partial z}\left(u\frac{\partial u}{\partial x} + v\frac{\partial u}{\partial y} + w\frac{\partial u}{\partial z} - fv\right) = 0,$$

where we have flipped the order of differentiation in the first term. Note how the vertical derivative of the (barotropic) pressure gradient has vanished. The other thing to note now is that if the flow started off without any vertical shear, i.e. if $\partial u/\partial z = 0$ and $\partial v/\partial z = 0$ at $t = 0$, then all of the advection terms and the Coriolis term would also be zero at $t = 0$. So we must conclude that if $\partial u/\partial z$ and $\partial v/\partial z$ are zero at $t = 0$, then $\partial u/\partial z$ will remain zero at all subsequent times. The same argument can be used for the y momentum equations. So 1) the horizontal velocities are independent of depth and 2) the $w\,(\partial u/\partial z)$ and $w\,(\partial v/\partial z)$ terms drop out of the equations.

Next, let's integrate the continuity equation through the layer. Assuming that the entire fluid depth is

$$H = D - h(x, y) + \eta(x, y, t),$$

where D is some typical depth, $h(x, y)$ is the height of the bottom topography above $z = -D$ and, finally, $\eta(x, y, t)$ is our sea surface height above $z = 0$. For depth-independent horizontal velocities we get

$$H\left(\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y}\right) + w(\eta) - w(-D + h) = 0.$$

Now, kinematic boundary conditions dictate that one cannot have flow through the bottom, nor through the sea surface. At the bottom this condition implies a vertical velocity

$$w_{bottom} = w(-D + h) = u\frac{\partial h}{\partial x} + v\frac{\partial h}{\partial y}.$$

And at the top, allowing for a sea surface which may move up and down, we get

$$w_{top} = w(\eta) = \frac{\partial \eta}{\partial t} + u\frac{\partial \eta}{\partial x} + v\frac{\partial \eta}{\partial y}.$$

Plugging these expressions in gives

$$\frac{\partial \eta}{\partial t} + H\left(\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y}\right) + u\frac{\partial H}{\partial x} + v\frac{\partial H}{\partial y} = 0,$$

where we have used the fact that $\partial D/\partial x = \partial D/\partial y = 0$. Moving the second and third terms to the right hand side and then combining them gives

$$\frac{\partial \eta}{\partial t} = -\left[\frac{\partial (uH)}{\partial x} + \frac{\partial (vH)}{\partial y}\right],$$

which gives the intuitive result that the time rate of change of the sea surface height is given by the convergence of the depth-integrated horizontal transport.

In summary, what we call *the shallow-water equations* are

$$\begin{aligned}
\frac{\partial u}{\partial t} + u\frac{\partial u}{\partial x} + v\frac{\partial u}{\partial y} - fv &= -g\frac{\partial \eta}{\partial x} \\
\frac{\partial v}{\partial t} + u\frac{\partial v}{\partial x} + v\frac{\partial v}{\partial y} + fu &= -g\frac{\partial \eta}{\partial y} \\
\frac{\partial \eta}{\partial t} &= -\left[\frac{\partial (uH)}{\partial x} + \frac{\partial (vH)}{\partial y}\right],
\end{aligned}$$

where all variables are now 2-dimensional. These are three equations in three unknowns (u, v and η). The name, shallow-water, refers to the underlying assumption that horizontal scales are much larger than the maximum vertical scale, namely the thickness of the layer. So the layer itself is 'shallow' compared to the horizontal extent of the flow. This is, of course, essentially related to our earlier assumption of a small aspect ratio $\delta = D/L$ and arrival at the hydrostatic approximation. Note also that by integrating in the vertical our previous fluid parcel has now become a *fluid column* of infinitesimal horizontal extent but a finite vertical extent. And, as shown above, the velocities throughout the column are depth-independent. The ambient rotation (and the lack of stratification) has essentially caused a *vertical rigidity* to the flow, as illustrated in Figure 58

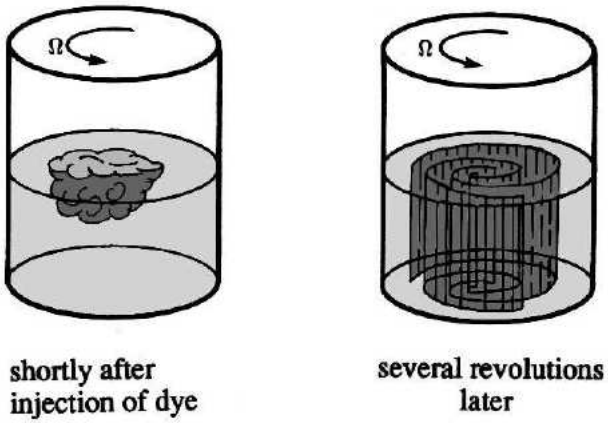

Figure 58: The motion in an unstratified rotating fluid layer tends to take on the behavior of vertically-rigid columns. (Source: Cushman-Roisin and Beckers, 2011, Fig. 1.3)

A little extra:

6.4.1 Stacked shallow-water layers

By putting two shallow-water layers on top of each other, each with a different density, we get the simplest possible model for a stratified ocean. The three equations above are applied to each layer, but we need to see what the pressures and pressure gradients become.

Let's look into the two-layer case. The hydrostatic pressure at any depth z in layer 1 (the top layer) becomes

$$p_1(z) = p(\eta) + g \int_z^\eta \rho_1 dz.$$

If we ignore the sea level air pressure and let ρ_1 be a constant, we get

$$p_1(z) = g\rho_1 (\eta - z),$$

and from this we can write the horizontal pressure gradient terms in layer 1

as

$$\begin{aligned} -\frac{1}{\rho_0}\frac{\partial p_1}{\partial x} &= -\frac{\rho_1}{\rho_0} g \frac{\partial \eta}{\partial x} \\ -\frac{1}{\rho_0}\frac{\partial p_1}{\partial y} &= -\frac{\rho_1}{\rho_0} g \frac{\partial \eta}{\partial y}, \end{aligned}$$

or, since under the Boussinesq approximation $\rho_1/\rho_0 \sim 1$,

$$\begin{aligned} -\frac{1}{\rho_0}\frac{\partial p_1}{\partial x} &= -g \frac{\partial \eta}{\partial x} \\ -\frac{1}{\rho_0}\frac{\partial p_1}{\partial y} &= -g \frac{\partial \eta}{\partial y}. \end{aligned}$$

To find the pressure at any level z in layer 2 we need to integrate first from the sea surface down to interface level Z_1, using density ρ_1, and then down to z using density ρ_2. This gives

$$p_2(z) = g\rho_1 \left(\eta - Z_1\right) + g\rho_2 \left(Z_1 - z\right),$$

and horizontal pressure gradients

$$\begin{aligned} -\frac{1}{\rho_0}\frac{\partial p_2}{\partial x} &= -\frac{\rho_1}{\rho_0} g \frac{\partial \eta}{\partial x} + \frac{\rho_1}{\rho_0} g \frac{\partial Z_1}{\partial x} - \frac{\rho_2}{\rho_0} g \frac{\partial Z_1}{\partial x}, \\ -\frac{1}{\rho_0}\frac{\partial p_2}{\partial y} &= -\frac{\rho_1}{\rho_0} g \frac{\partial \eta}{\partial y} + \frac{\rho_1}{\rho_0} g \frac{\partial Z_1}{\partial y} - \frac{\rho_2}{\rho_0} g \frac{\partial Z_1}{\partial y}, \end{aligned}$$

or, again using $\rho_1/\rho_0 \sim 1$,

$$\begin{aligned} -\frac{1}{\rho_0}\frac{\partial p_2}{\partial x} &= -g \frac{\partial \eta}{\partial x} - g \frac{(\rho_2 - \rho_1)}{\rho_0} \frac{\partial Z_1}{\partial x} \\ -\frac{1}{\rho_0}\frac{\partial p_2}{\partial y} &= -g \frac{\partial \eta}{\partial y} - g \frac{(\rho_2 - \rho_1)}{\rho_0} \frac{\partial Z_1}{\partial y}. \end{aligned}$$

At this stage it is useful to introduce the concept of *reduced gravity*

$$g' \equiv g \frac{(\rho_2 - \rho_1)}{\rho_0} = g \frac{\Delta\rho}{\rho_0}$$

so that the expressions become

$$\begin{aligned} -\frac{1}{\rho_0}\frac{\partial p_2}{\partial x} &= -g\frac{\partial \eta}{\partial x} - g'\frac{\partial Z_1}{\partial x} \\ -\frac{1}{\rho_0}\frac{\partial p_2}{\partial y} &= -g\frac{\partial \eta}{\partial y} - g'\frac{\partial Z_1}{\partial y}. \end{aligned}$$

The pressure gradient term in layer two are thus due to the sea surface tilt (this term is felt throughout the water column) and also augmented by the tilt of the interface. Not that since $g' \ll g$, the tilt of the interface has to be much much larger than the sea surface tilt to make a comparable impact on the pressure gradient in layer 2.

6.5 Geostrophic currents and the thermal wind

In addition to the condition of a small aspect ratio $\delta \equiv D/L$, the primitive equations (PE) and the shallow-water equations (SWE) that we use to describe large-scale flows required that the Coriolis accelleration is as big as the other accelleration terms. As it turns out, at length scales larger than a few kilometers and time scales longer than a few days, the Coriolis term is *much bigger* than all other accelleration and friction terms. Then there is a relatively tight balance—*the geostrophic balance*—between the Coriolis accelleration and the horizontal pressure gradient. This is the balance we can see in the model simulation of currents in the Gulf of Mexico: surface currents are not primarily down the pressure gradient (the sea surface tilt) but rather around it. Large-scale flows both in the oceans and atmosphere are basically geostrophic and hydrostatic.

Let's scale the terms in the x momentum equations again (the argumentation will be the same for the y momentum equation):

$$\begin{aligned} \frac{\partial u}{\partial t} + \boldsymbol{u}\cdot\nabla_H u + w\frac{\partial u}{\partial z} - fv &= -\frac{1}{\rho_0}\frac{\partial p}{\partial x} - \nabla_H\cdot(-A_H\nabla_H u) - \frac{\partial}{\partial z}\left(-A_V\frac{\partial u}{\partial z}\right) \\ \frac{U}{T},\quad \frac{U^2}{L},\quad \frac{UW}{D},\quad fU &\sim \frac{P}{\rho_0 L},\quad \frac{A_H U}{L^2},\quad \frac{A_V U}{D^2}. \end{aligned}$$

Now, to assess how big the various terms are compared to the Coriolis term, we divide all scaling terms by fU.This gives

$$\frac{1}{fT},\quad \frac{U}{fL},\quad \frac{U}{fL},\quad 1 \sim \frac{P}{fU\rho_0 L},\quad \frac{A_H}{fL^2},\quad \frac{A_V}{fD^2},$$

where we have used the previous result $W/D \sim U/L$ to scale the vertical advection term (the third term). These are non-dimensional numbers that even have names (except for the scaled pressure term). They are the temporal and advective Rossby numbers,

$$\begin{aligned} \epsilon_T &\equiv \frac{1}{fT} \\ \epsilon &\equiv \frac{U}{fL}, \end{aligned}$$

and the horizontal and vertical Ekman numbers,

$$\begin{aligned} E_H &\equiv \frac{A_H}{fL^2} \\ E_V &\equiv \frac{A_V}{fD^2}. \end{aligned}$$

Let's estimate the size of these terms for flows that can safely be considered large-scale. We assume horizontal and vertical length scales $L \sim 10^5\mathrm{m}$ and $D \sim 10^3\mathrm{m}$, a horizontal velocity scale $U \sim 10^{-1}\mathrm{m\,s^{-1}}$ and a time scale $T \sim 10^6\mathrm{s}$ (about ten days). Assuming that we are in mid-latitudes, with $f \sim 10^{-4}\mathrm{s}$, then results in Rossby numbers $\epsilon_T \sim 10^{-2}$ and $\epsilon \sim 10^{-2}$. So the accelleration terms are about one hundred times smaller than Coriolis. Estimating the Ekman numbers depends on good guesses for the diffusion coefficient. Fairly reasonable values are $A_H \sim 10^3\mathrm{m^2s^{-1}}$ and $A_V \sim 10^{-2}\mathrm{m^2s^{-1}}$, giving Ekman numbers $E_H \sim 10^{-3}$ and $E_V \sim 10^{-2}$. So the accelleration and diffusion terms are indeed small compared to Coriolis.

The consequence of small Rossby and Ekman numbers is that the horizontal velocities are nearly geostrophic, meaning they are governed by a very tight balance between the Coriolis accelleration and the horizontal pressure gradient. So we write

$$fv_g = \frac{1}{\rho_0}\frac{\partial p}{\partial x} \tag{8}$$

$$fu_g = -\frac{1}{\rho_0}\frac{\partial p}{\partial y} \tag{9}$$

or, in vector form,

$$f\boldsymbol{k} \times \boldsymbol{u}_g = -\frac{1}{\rho_0}\nabla_H p. \tag{10}$$

Hence, geostrophic currents flow *along* pressure contours, with high pressures to their right in the northern hemisphere and to their left in the southern hemisphere (where $f < 0$).

Since the pressure at $z = 0$ is given by the sea surface tilt (as discussed above), we have that

$$v_g(0) = \frac{g}{f}\frac{\partial \eta}{\partial x} \tag{11}$$

$$u_g(0) = -\frac{g}{f}\frac{\partial \eta}{\partial y}. \tag{12}$$

This supports what we see in Figure 57, namely a surface flow with high sea level to its right in the northern hemisphere. The numerical simulation shown in the figure is based on the full primitive equations, so it contains all accelleration terms and also parametrized turbulent momentum fluxes. And still the model results show that for flows with horizontal scales of tens of kilometers and larger the geostrophic balance is overwhelming.

Note that for geostrophic flows that don't have too large meridional extents, $f = \mathrm{const.}$ and the horizontal divergence of the flow is

$$\frac{\partial u_g}{\partial x} + \frac{\partial v_g}{\partial y} = 0.$$

So geostrophic flows that are not too extensive horizontally also have a zero horizontal divergence. This, in turn, implies that the vertical stretching of the associated vertical veloctiy is zero, i.e. $\partial w_g/\partial z = 0$. So the vertical velocity is constant with depth.

What about geostrophic currents further down in the water column? To look into this we take the vertical derivative of the geostriphic relations. For a shallow-water layer the horizontal pressure gradient doesn't change within the layer, so $\partial u_g/\partial z = \partial u_g/\partial z = 0$. In other words, the geostrophic flow in such a layer definitely act as rigid vertical columns—even without the previous requirement of having a zero vertical shear initially. If we now also consider $f = \mathrm{const.}$, the columns of geostrophic flow cannot stretch vertically. Then, since the vertical velocities at the sea surface are very small (because the sea surface is nearly flat) they also need to be very small at the bottom. This means, in practice, that the bottom flow needs to flow *around* obstacles rather than flow over them (flow over a bumpy bottom would produce large vertical velocities at the bottom). And since the vertical shear of the horizontal flow is zero, the entire column need to do the

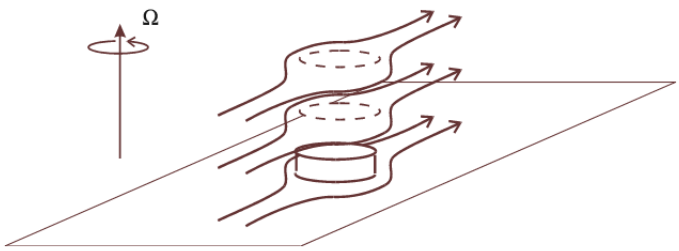

Figure 59: Taylor columns: geostrophic flows in unstratified fluids that tend to flow around bottom obstacles...throughout the entire water column.

same thing (Figure 59). So geostrophic currents in an unstratified ocean tend to act like vertical columns—so-called *Taylor columns*—that follow the bottom topography. This effect is real and can be observed many places in the oceans, especially at high latitudes where the vertical density stratification is low due to buoyancy loss (cooling) at the sea suface and extensive convective mixing. The large-scale flows there tend to follow the continental slopes that separate the shelf regions from deep off-shore basins (Figure 60).

But when the fluid is stratified the geostrophic flow can indeed be vertically sheared. Taking the vertical derivative of the geostrophic version of the primitive equations gives

$$\begin{aligned}
\frac{\partial v_g}{\partial z} &= \frac{1}{f\rho_0}\frac{\partial}{\partial z}\left(\frac{\partial p}{\partial x}\right) = \frac{1}{f\rho_0}\frac{\partial}{\partial x}\left(\frac{\partial p}{\partial z}\right) \\
\frac{\partial u_g}{\partial z} &= -\frac{1}{f\rho_0}\frac{\partial}{\partial z}\left(\frac{\partial p}{\partial y}\right) = -\frac{1}{f\rho_0}\frac{\partial}{\partial y}\left(\frac{\partial p}{\partial z}\right).
\end{aligned}$$

But we already have an expression for the vertical pressure gradient, from the hydrostatic approximation. Plugging this in gives

$$\frac{\partial v_g}{\partial z} = -\frac{g}{f\rho_0}\frac{\partial \rho}{\partial x} \tag{13}$$

$$\frac{\partial u_g}{\partial z} = \frac{g}{f\rho_0}\frac{\partial \rho}{\partial y}. \tag{14}$$

So whereas geostrophic currents are given by horizontal pressure gradients, their changes with depth are given by horizontal density gradients. We call these vertical derivatives the *thermal wind shear*. The wording stems from the atmosphere

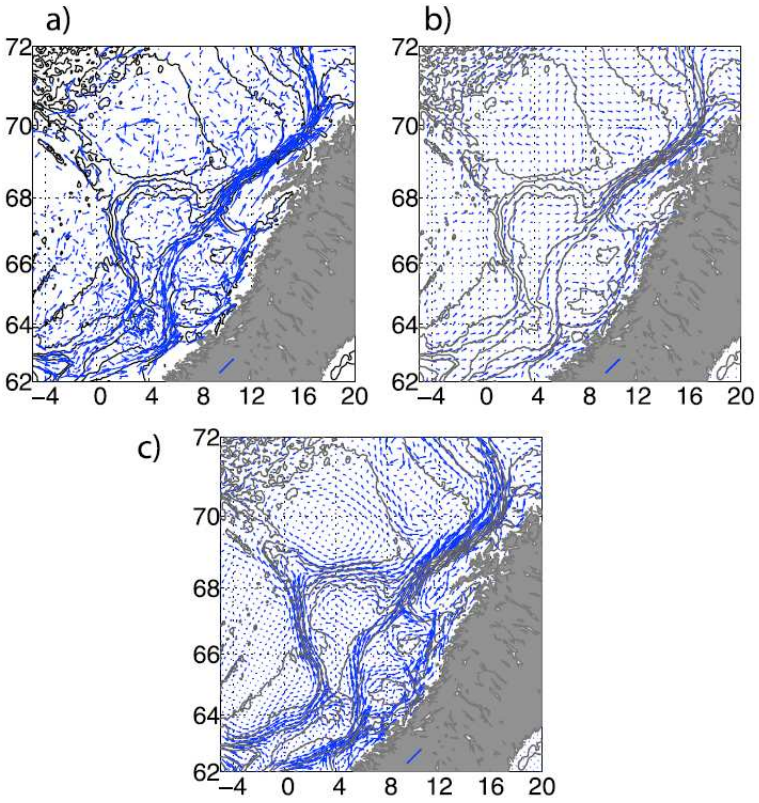

Figure 60: Time-mean surface currents off the Norwegian coast estimated by a) surface drifters, b) satellite altimeter and c) a numerical ocean model. (Source: Isachsen et al., 2012, Fig. 2)

where the vertical shear of the geostrophic winds is associated with horizontal temperature gradients (if humidity effects of air density are neglected).

The importance of the thermal wind equations for our historical understanding of oceanography cannot be overemphasized. The key thing to note is that, historically and even up to today, obtaining observations of temperature and salinity (from which we can estimate density) has been much easier than obtaining direct velocity observations. If one has, say, a hydrographic section with vertical profiles of temperature and salinity (and hence density), one can integrate the thermal wind equations from some depth level z_{ref} to any other level z to obtain

$$v_g(z) = v_g(z_{ref}) - \frac{g}{f\rho_0}\int_{z_{ref}}^{z} \frac{\partial \rho}{\partial x}\, dz \tag{15}$$

$$u_g(z) = u_g(z_{ref}) + \frac{g}{f\rho_0}\int_{z_{ref}}^{z} \frac{\partial \rho}{\partial y}\, dz. \tag{16}$$

So if one knows the geostrophic flow at any level z_{ref} one can find it at any other level given information about horizontal density gradients. In these expressions $v_g(z_{ref})$ and $u_g(z_{ref})$ are called *reference-level velocities*, and an enormous historical effort has gone into making inference about these. A classical approach has been to assume a *level of no motion* at some great depth, say at 2000 m, where one expects the flow to be very small. So one assumes $v_g(z_{ref}) = u_g(z_{ref}) = 0$ and integrates the horizontal density gradients vertically to give an estimate of the geostrophic flow at any depth. An alternative today is to set the reference level to the sea surface, make estimates of the surface geostrophic flow from altimeter-derived sea surface height gradients, and then integrate downwards from these.

A little extra:

A thermal wind shear can also be calculated in a stacked shallow water model. In the last section we found the expressions for horizontal pressure gradients in the top and second layer:

$$-\frac{1}{\rho_0}\frac{\partial p_1}{\partial x} = -g\frac{\partial \eta}{\partial x}$$

$$-\frac{1}{\rho_0}\frac{\partial p_1}{\partial y} = -g\frac{\partial \eta}{\partial y}$$

and

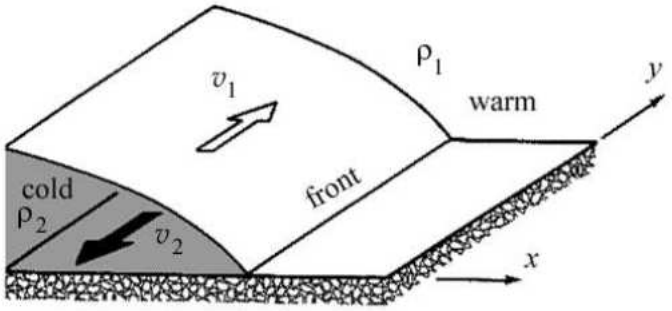

Figure 61: Thermal wind shear in a two-layer shallow-water model. (Source: Cushman-Roisin and Beckers, 2011, Fig. 15.2)

$$
\begin{aligned}
-\frac{1}{\rho_0}\frac{\partial p_2}{\partial x} &= -g\frac{\partial \eta}{\partial x} - g'\frac{\partial Z_1}{\partial x} \\
-\frac{1}{\rho_0}\frac{\partial p_2}{\partial y} &= -g\frac{\partial \eta}{\partial y} - g'\frac{\partial Z_1}{\partial y}.
\end{aligned}
$$

So from these we get expressions for the difference in geostrophic velocity between the two layers

$$
\begin{aligned}
v_{g1} - v_{g2} &= -\frac{g'}{f}\frac{\partial Z_1}{\partial x} \\
u_{g1} - u_{g2} &= \frac{g'}{f}\frac{\partial Z_1}{\partial y}.
\end{aligned}
$$

6.6 Geostrophic degeneracy and vorticity dynamics

The fact that the large-scale flow in both the atmosphere and ocean are approximately governed by the geostrophic balance is both helpful and problematic at the same time. It gives us the advantage of estimating winds and currents without having to actually make direct measurements of these hard-to-obtain and often noisy quantities. But the expression describing the geostrophic balance contains no time derivatives, so it says absolutely nothing about how the flow evolves in time. Given observations of the pressure field one can say what the large-scale

currents are there and then, but the expression tells us nothing about the future flow. This is what we call *geostrophic degeneracy*.

Even when we do keep all the terms in the primitive or shallow-water equations and try to integrate these forward in time, the huge size difference between the Coriolis and pressure gradient terms and all the other terms causes practical problems. All terms in the equations of motion contain errors, either observational errors or numerical truncation errors. And the errors in the Coriolis and pressure gradient terms may be as big as the true values of the accelleration terms. Say that we think we know the error of all terms in the primitive equation momentum equation to within 5%, individually. Now, if the Rossby numbers are 1/100, then the error in the Coriolis term alone will be 500% of the size of the accelleration terms. This does not allow for a very accurate estimate of the time-evolution of the velocity. Modern computer codes that time-step the primitive equations need to operate with very good numerics to reduce finite-difference truncataion errors to very low values.

So the geostrophic balance is a great help in *diagnosing* the velocity field from the pressure field. But to tell us anything really interesting, i.e. about how the flow evolves in time, it would be useful have a set of equations where the all-dominating geostrophic balance were somehow 'hidden'. We can actually obtain such equations by a bit of manipulation. Let's start with the shallow-water momentum equations (also now ignoring Reynolds momentum fluxes):

$$\frac{\partial u}{\partial t} + u\frac{\partial u}{\partial x} + v\frac{\partial u}{\partial y} - fv = -g\frac{\partial \eta}{\partial x} \tag{17}$$

$$\frac{\partial v}{\partial t} + u\frac{\partial v}{\partial x} + v\frac{\partial v}{\partial y} + fu = -g\frac{\partial \eta}{\partial y}. \tag{18}$$

We now take the x-derivative of (18) and subtract the y-derivative of (17). This is equivalent to taking the curl of the vector form of the momentum equation. After some algebra we obtain

$$\frac{\partial \zeta}{\partial t} + u\frac{\partial \zeta}{\partial x} + v\frac{\partial \zeta}{\partial y} + v\frac{\partial f}{\partial y} + (f+\zeta)\left(\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y}\right) = 0, \tag{19}$$

where we have introduced the 'relative vorticity' as

$$\begin{aligned} \zeta &\equiv \frac{\partial v}{\partial x} - \frac{\partial u}{\partial y} \\ &\equiv \nabla_H \times \boldsymbol{u}, \end{aligned}$$

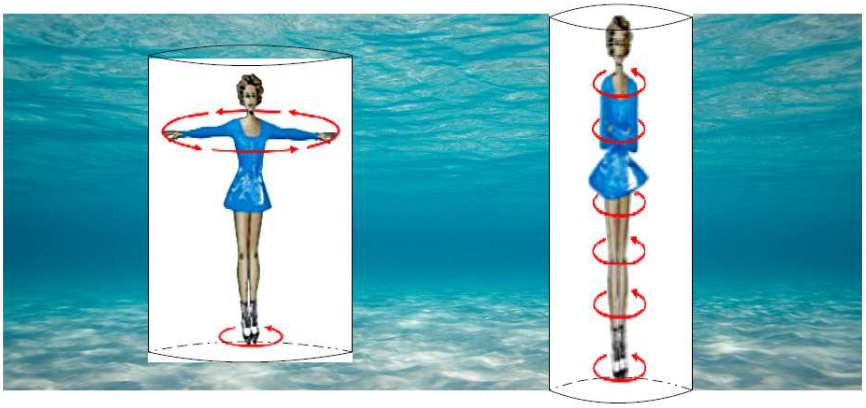

Figure 62: The effect of a horizontal convergence on a water column—via the analogy of the figure skater.

i.e. the curl of the horizontal velocity field. This curl is basically the rotation of the flow. Counter-clockwise or *cyclonic* flows have $\zeta > 0$ while clockwise or *anti-cyclonic* flows have $\zeta < 0$. Notice how the pressure gradient term has completely vanished. So this equation shows no sign of the geostrphic balance, no matter how dominant it is.

This is the shallow-water vorticity equation. But what is it saying? The first thing to note is that the total or absolute vorticity of the flow is the sum of the planetary vorticity f and the relative vorticity ζ. The second thing to note is that the planetary vorticity on our tangent-plane coordinate system is only a function of the y-coordinate (remember that $f = 2\Omega \sin\theta$ and that our y-coordinate is alligned in the meridional direction). So we have $f = f(y)$ and can therefore rewrite the vorticity equation as

$$\frac{D}{Dt}(f+\zeta) = -(f+\zeta)\left(\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y}\right). \tag{20}$$

We can now see what's going on: the time rate of change of total vorticity of a fluid column, as it moves with the flow, is set by the horizontal convergence of the flow times the absolute vorticity itself. An analogy may be the figure ice skater who is able to increase her rotation rate by pulling the arms in towards the body—a convergence (see Figure 62).

A water column starting off at rest on the northern hemisphere has some positive absolute vorticity, namely that given by the planetary rotation. A convergence of the flow can increase the absolute vorticity, either by moving northward to in-

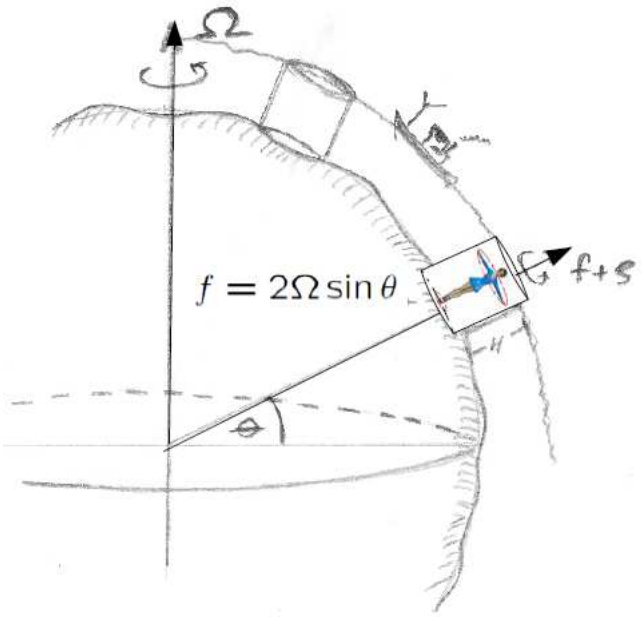

Figure 63: Two water columns whose absolute vorticity $(f + \zeta)$ will change if they experience a convergent or divergent horizontal flow. They can do so either by moving north or south (to change planetary vorticity f) or by changing their spin relative to Earth itself (to change relative vorticity ζ).

crease the planetary vorticity or by starting to spin cyclonically (Figure 63. A horizontal divergence does the opposite. But notice what happens if the absolute vorticity starts to approach zero (by lots of divergence). Since the divergence is multiplied by the absolute vorticity itself, it impacts the rate of change of itself less and less. In fact, the equation shows that a water column—if it is not forced by friction—can never obtain negative absolute vorticity. The relative vorticity must obey $\zeta > -f$. Sounds strange? But think back to the figure skater. She too cannot change her direction of rotation, no matter how hard she tries to extend or pull in her arms.

We can go one step further than this by bringing in the shallow-water version of the continuity equation. We had

$$\frac{\partial \eta}{\partial t} = -\left[\frac{\partial}{\partial x}(Hu) + \frac{\partial}{\partial y}(Hv)\right],$$

but this this can also be written

$$\frac{\partial H}{\partial t} = -\left[\frac{\partial}{\partial x}(Hu) + \frac{\partial}{\partial y}(Hv)\right]$$

since the total vertical thickness is $H = D - h + \eta$ where the sea surface elevation η is the only time-variable contribution. Splitting up the spatial derivatives and rearranging gives the expression

$$\frac{DH}{Dt} = -H\left(\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y}\right), \tag{21}$$

which says that the time rate of change of the thickness of the fluid column, as it moves with the flow, is given by the convergence of the flow (times the thickness itself). This makes sense: compressing the column horizontally elongates it vertically, and vice versa.

Now we'll get to the key point. Note that we now have two expressions that contain the horizontal divergence (or convergence) of the flow. Substituting (20) into (21) and rearranging gives the final result:

$$\frac{D}{Dt}\left(\frac{f+\zeta}{H}\right) = 0. \tag{22}$$

Notice what has happened here. Taking the curl of the shallow-water momentum equations has given us a vorticity equation. Combining this with the shallow-water continuity equation has resulted in an equation which says that in the absence of friction (and, as it turns out, the absence of volume sources or sinks) the quantity

$$q = \frac{f+\zeta}{H}$$

is conserved following the flow. We call this conserved quantity the *potential vorticity* of the fluid column. The equation for consevation of potential vorticity tells us that if a fluid column is, say, squished vertically then it has to reduce its total vorticity (rotation). It can do so by either moving southward (to reduce f) or by spinning more clockwise (to reduce ζ), or both. If the fluid column is elongated vertically, the absolute vorticity has to increase instead. Again, it may be useful to think of what happens if our figure skater stretches or squeezes. The vertical stretching/squeezing is related to a horizontal convergence/divergence.

Notice also that our three original equations (two horizontal momentum equations and one depth-integrated continuity equation) have been reduced to one.

This seems like progress...until we realize that we are now left with one equation in three unknowns (u, v and η). That doesn't seem very promising in terms of solving anything. But remember that if the velocities components u and v are nearly geostrophic then we can, to a good approximation, write these in terms of sea surface height η. And if we do exacxtly that, plug the expressions for the geostrophic balance into the potential vorticity equation, we are essentially dealing with one equation for one unknown. So this equation, i.e. the statement of conservation of potential vorticity, turns out to be extremely useful in oceanography (and in meteorology) provided that we study flows that are nearly geostrophic. In fact, most theories of the large-scale ocean and atmosphere circulation—also some that we will be discussing shortly—are based on the principle of potential vorticity convervation. So if we start loosing track of what goes on, when we simply can't see the forest for all the trees (or math), it may help to think of the figure skater.

7 The large-scale wind-driven circulation

With quite a bit of mathematical machinery in place we are now ready to revisit both the wind-driven and buoyancy-driven ocean circulation. In this and in the next chapter we will look at very simplified theories of large-scale ocean flows. The theories, even in their extreme simplification, explain key aspects of large-scale flows on a rotating planet. The most prominent feature, perhaps, is the explanation for why mid-latitude ocean flows show clear east-west asymmetries and contain western boundary currents (see e.g. Figs. 15 and 16).

The winds can 'pull' the ocean along with it because of friction, what we call the wind stress. This stress is a vertical flux of horizontal momentum through the ocean surface. But the momentum flux only reaches down to a few tens of meters and can thus only acellerate the very surface of the ocean. And yet, wind-driven ocean currents have been observed to reach all the way to the ocean bottom, to thousands of meters of depth. How is this possible? It turns out that the wind-driven flow at depth is not driven by the surface momentum fluxes directly but rather by horizontal pressure gradients created by the winds removing surface waters from some regions and piling them up in other regions. Where the wind-driven surface flow is convergent there will be a pile-up of mass, actually making the sea surface higher than elsewhere. And waters residing a thousand meters below such a pile-up region will experience a higher pressure (from the weight of the extra water above) than waters some distance away. The waters will start to accellerate down the pressure gradient, away from the pile-up region. But soon the Coriolis accelleration kicks in and steers it to the right in the northern hemisphere (or to the left in the southern). As we will see in this chapter, horizontal motion in the ocean is *primarily* driven by such pressure gradients. So most of what we call the wind-driven ocean circulation is only *indirectly* driven by the wind stress. And, as alluded to above, then Earth's rotation complicates things. In fact, as we'll soon see, the surface flows that feel the winds directly genereally don't allign up with the winds. But the deep flows, driven by the pressure gradients due to the uneven piling up of wind-driven flows, do.

7.1 Ekman transport

Consdier the steady version of the horizontal momentum equation in which the nonlinear advection terms and all but the vertical derivative of horizontal Reynolds stresses have been neglected. We keep this stress term since this gives the horizontal momentum input by the winds. The resulting equations are

$$-fv = -\frac{1}{\rho_0}\frac{\partial p}{\partial x} + \frac{1}{\rho_0}\frac{\partial \tau_{xz}}{\partial z} \quad (23)$$

$$fu = -\frac{1}{\rho_0}\frac{\partial p}{\partial y} + \frac{1}{\rho_0}\frac{\partial \tau_{yz}}{\partial z}. \quad (24)$$

Since the equations are linear we can split the velocity into two parts, one *geostrophic* and one *ageostrophic*. So we write

$$u = u_g + u_a$$
$$v = v_g + v_a,$$

where

$$-fv_g = -\frac{1}{\rho_0}\frac{\partial p}{\partial x}$$
$$fu_g = -\frac{1}{\rho_0}\frac{\partial p}{\partial y}$$

and

$$-fv_a = \frac{1}{\rho_0}\frac{\partial \tau_{xz}}{\partial z} \quad (25)$$

$$fu_a = \frac{1}{\rho_0}\frac{\partial \tau_{yz}}{\partial z}. \quad (26)$$

In the ocean interior the geostrophic balance is completely dominating. But near the top and bottom boundaries momentum fluxes are important and the ageostrophic flow non-negligible. Integrating the ageostrophic relationships vertically, from the surface to some depth z_0 below the direct influence of the winds (to a depth where we assume stresses are zero), we get

$$V_E \equiv \int_{z_0}^{0} v_a dz = -\frac{\tau_x^w}{f\rho_0}$$
$$U_E \equiv \int_{z_0}^{0} u_a dz = \frac{\tau_y^w}{f\rho_0},$$

where τ_x^w and τ_y^w are the zonal and meridional windstress components. We call these depth-integrated ageostrophic volume transports the *Ekman transports*[14].

[14] After V. W. Ekman who first examined this problem in 1902 using observations collected by Nansen's *Fram* expedition.

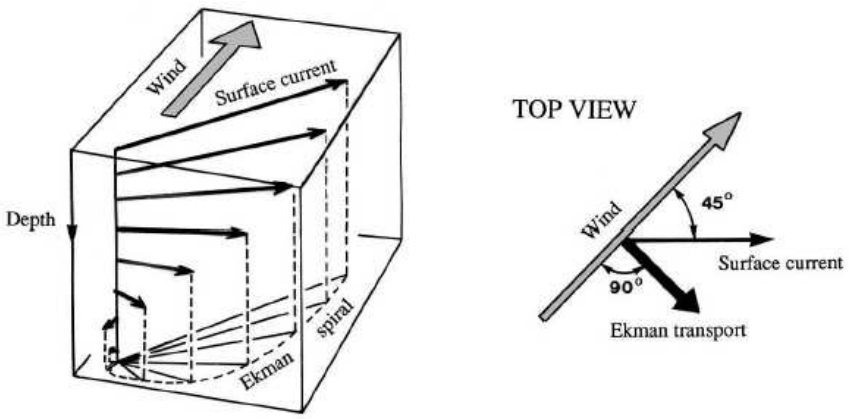

Figure 64: (Source: Cushman-Roisin and Beckers, 2011, Fig. 8.7)

Written in vector form, the Ekman transport is

$$\boldsymbol{U}_E = -\frac{1}{f\rho_0}\boldsymbol{k} \times \boldsymbol{\tau}^w \tag{27}$$

where $\boldsymbol{\tau}^w$ is the wind stress vector. So, as hinted at in the introductory paragraph above, the depth-integrated Ekman transport is *not* in the same direction as the winds but at right angles to it, to the right in the northern hemisphere and to the left in the southern hemisphere (where $f < 0$).

By assuming, as we have done earlier, that the Reynolds stress below the sea surface goes as $\tau_{xz} = \rho_0 A_z \partial u/\partial z$ (and similarly for the y component), one can study the actual vertical distribution of the Ekman transport, i.e. the horizontal velocities profile as a function of depth. We will not go through the calculations here, but the result reveals an 'Ekman spiral' in which the surface currents, at $z = 0$, are at 45% to the right of the winds in the northern hemisphere (to the left in the southern hemisphere) and then continue to spiral clockwise with depth (counter-clockwise south of the equator). Figure 64 illustrates what this looks like.

The full calculations show that the momentum stresses and hence the Ekman velocities decay exponentially with depth, with an e-folding scale of

$$d_E = \sqrt{\frac{2A_V}{f}}.$$

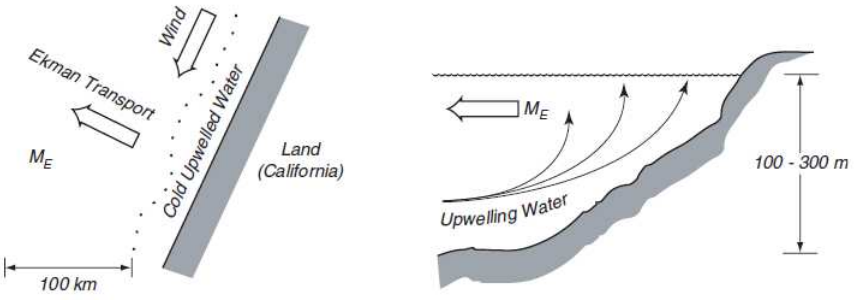

Figure 65: The process of coastal upwelling. (Source: Stewart, 2008, Fig. 9.8)

After a few tens of meters (assuming a turbulent vertical viscocity $A_v \simeq 10^{-2}\,\mathrm{m^2 s^{-1}}$ and $f \simeq 10^{-4}\,\mathrm{s^{-1}}$) the Ekman currents are vanishingly small. And yet, as we'll see below, the Ekman transport in this thin surface layer has profound imacts on the ocean circulation as a whole.

7.2 Ekman-induced upwelling and downwelling

The key process, which impacts everything else, is the horizontal *convergence or divergence* of the surface Ekman transport. Imagine a wind blowing along the eastern margin of an ocean in the northern hemisphere, i.e. along the west coast of some continent. If the wind blows from the north ("northerlies"), the Ekman transport will be away from the coast, to the west. But because of the presence of the coast, there is a divergence in the horizontal flow. There are two consequences: 1) the sea surface near land drops a bit and 2) after some time vertical flow, from depth, has to replenish the divergence in the surface layer. This is what lies behind the phenomenon of *coastal upwelling* (Figure 65).

The upward vertical flow which compensates the Ekman divergence in the surface layer advects whatever properties are at depth up to towards the surface. So in temperature-stratified mid-latitude oceans, coastal upwelling will bring cold and nutrient-rich waters up to the surface—bringing dismay to swimmers and surfers but happniess to fishermen. In addition, the drop in the sea surface near the coast and the resulting cross-shore sea surface tilt will eventually be balanced by a southward-flowing geostrophic current (in the northern hemisphere). Strange as it may sound to a non-oceanographer (but not to you anymore!), the depth-integrated surface flow directly impacted by the winds are in the cross-wind direction. But the flow deeper down, driven by the horizontal pressure gradient linked to the sea

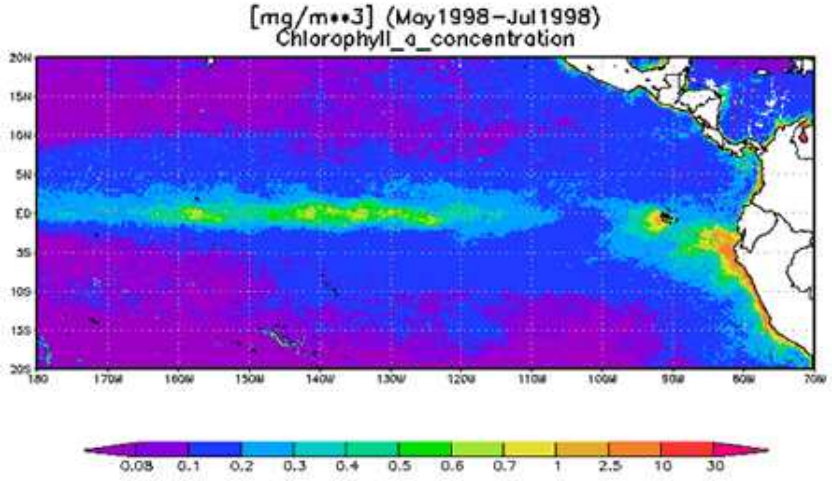

Figure 66: Surface chlorophyl concentrations in the tropical Pacific Ocean. The enhanced chlorophyl along the equator and along the Peruvian coast is due to Ekman-induced upwelling of nutrient-rich waters (Source: https://en.wikipedia.org/wiki/Upwelling)

surface tilt, is in the same direction as the winds.

A similar thing happens at the equator. There the 'easterly trade winds' (blowing westward) cause a diverging surface Ekman transport due to the latitudional variations of the Coriolis parameter. Exactly at the equator $f = 0$ and the wind-driven transport is actually in the direction of the winds, i.e. westward. But a few degrees to the north and the south there are Ekman transports away from the equator, causing a divergence and accompanying upwelling of colder, nutrient-rich, waters. The result is an increased phytoplankton production which can be observed from space (Figure 66).

Finally, wind-induced up and downwelling can happen in the open ocean, away from land boundaries and from the equator, if there is a curl, i.e. a rotational component, in the wind stress. This is relatively easy to visualize. An atmospheric low pressure system in the northern hemisphere is associated with cyclonic (counter-clockwise) winds. If the low pressure system resides over an open ocean region, the upper-ocean Ekman transport will be to the right, meaning away from the low pressure center. So the surface Ekman transport is divergent and causes a depression in the sea surface height and upward vertical velocities underneath. The vertical velocity at the base of the Ekman layer, i.e. on top of the geostrophic interior, is found by depth-integrating the continuity equation over the

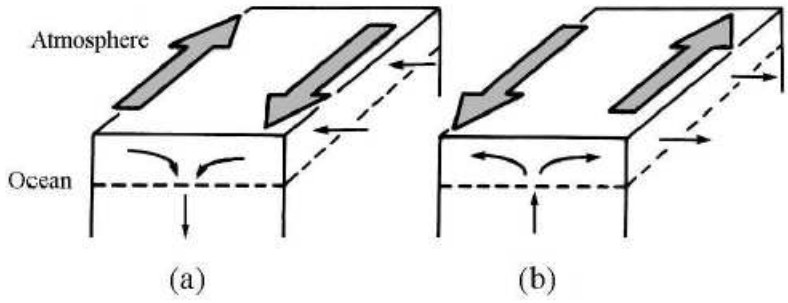

Figure 67: Vertical velocities at the base of the surface Ekman layer for winds with a) $\partial v/\partial x < 0$ (negative curl) and b) $\partial v/\partial x > 0$ (positive curl). (Source: Cushman-Roisin and Beckers, 2011, Fig. 8.8)

Ekman layer. This gives

$$w(z_0) = \frac{\partial U_E}{\partial x} + \frac{\partial V_E}{\partial y},$$

where z_0, as before, is the z level at the bottom of the Ekman layer. Substituting in from (27) gives

$$\begin{aligned} w(z_0) &= \frac{1}{\rho_0}\left[\frac{\partial}{\partial x}\left(\frac{\tau_y^w}{f}\right) - \frac{\partial}{\partial y}\left(\frac{\tau_x^w}{f}\right)\right] \\ &= \frac{1}{\rho_0}\nabla_H \times \left(\frac{\boldsymbol{\tau}^w}{f}\right) \end{aligned}$$

meaning that the vertical velocity is given by the curl of the wind stress divided by f.

As Figure 67 illustrates, a positive wind stress curl (cyclonic winds in the northern hemisphere) causes positive vertical velocities at the base of the surface Ekman layer, what we call 'Ekman suction', whereas a negative curl (anticyclonic winds) causes negative vertical velocities ('Ekman pumping'). As we will see next, such open ocean wind stress curl is key to understand the large mid-latitude gyres in all the major oceans.

7.3 Wind-driven mid-latitude ocean gyres

Figure 68 show the Mean Dynamic Topography (MDT), i.e. the time-mean sea surface height over the geoid as measured by satellite altimetry. We know that

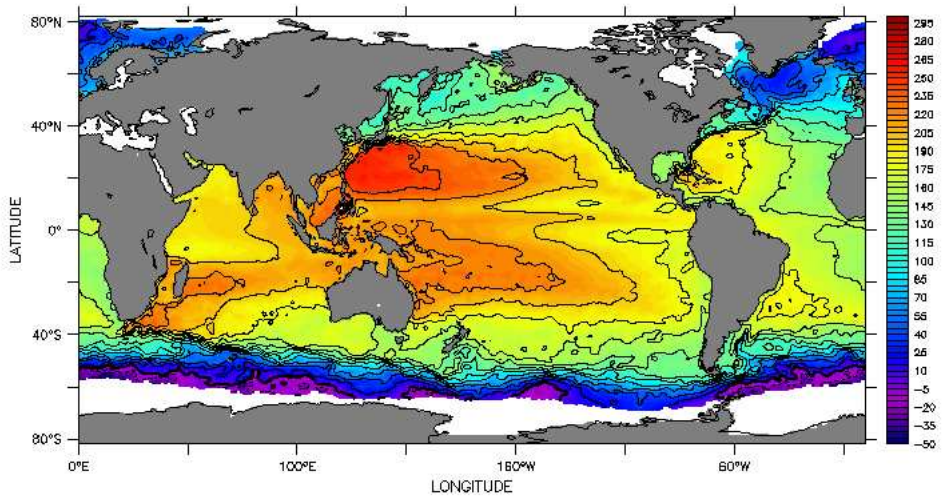

Figure 68: The time-mean sea surface height, as measured by satellites (Source: http://www.aviso.altimetry.fr)

the steady large-scale and geostrophic ocean currents flow along isolines of MDT, with high sea level their right in the northern hemisphere and to their left in the southern hemisphere. What the plot indicates is that the mid-latitudes are dominated by large-scale gyres, at least in the Pacific and Atlantic oceans. One of the most elegant and successful achievements of theoretical oceanography was the explanation for the existence of such mid-latitude gyres that were driven by winds.

A first heuristic attempt at explaining what is observed can be made from the steady and linearized momentum equations discussed above and the concept of the wind-driven surface Ekman transport. The mid-latitude gyres are in fact driven by the joint effect of westerly winds (blowing eastward) centered around 40° north and south and the easterly trade winds centered around 10°. These wind systems cause an Ekman convergence in the latitudes in between and a build-up of the sea surface height there (in addition to downward Ekman pumping, as seen above). The resulting sea level gradients drive geostrophic currents that are consistent with the observed gyres. In the North Pacific and North Atlantic, for example, the currents are eastward in the latitude band 30–40°N and westward in the band 10–20°N. One could then argue that the presence of continental barriers force these flows to turn and form closed gyres.

But what about the east-west asymmetry of the gyres, as revealed by the satel-

lite observations? The sea level gradients are much stronger along the western boundaries than anywhere else in these gyres. So the geostrophic velocities are also a lot higher along these western boundaries. These are the famous 'Kuroshio' and 'Gulf Stream' currents in the Northern Pacific and Atlantic oceans and the 'East Australian' and 'Brazil' currents in the southern Pacific and Atlantic oceans. As revealed by the rather famous oceanographers Harald Sverdrup and Henry Stommel (and others) in the 1940s and 1950s, the east-west asymmetry and the presence of these western boundary currents can only be properly explained by pulling the vorticity equation out of the hat rather than the momentum equation.

7.3.1 Interior Sverdrup balance

Harald Sverdrup studied the vorticity balance in the interior ocean, away from both vertical and horizontal boundaries where turbulent momentum fluxes may be important. Crucially, he suspected that the latitudional variation of the Coriolis parameter might be important to larges-scale dynamics. The easiest model which takes this into account is the *beta-plane model* in which the Coriolis parameter in our tangent coordinate system, centered at latitude θ, is a linear function of y:

$$f = f_0 + \beta y,$$

where, as before,

$$f_0 = 2\Omega \sin\theta$$

and

$$\beta \equiv \frac{\partial f}{\partial y} = \frac{2\Omega}{R}\cos\theta.$$

Sverdrup looked at steady geostrophic flows on this beta plane, i.e. at the equations

$$\begin{aligned} -fv_g &= -\frac{1}{\rho_0}\frac{\partial p}{\partial x} \\ fu_g &= -\frac{1}{\rho_0}\frac{\partial p}{\partial y} \end{aligned}$$

where f is allowed to vary in the y-direction. He then took the curl ($\partial/\partial x$ of the second equation minus $\partial/\partial y$ of the first) to give

$$\beta v_g + f\left(\frac{\partial u_g}{\partial x} + \frac{\partial v_g}{\partial y}\right) = 0.$$

Note that this is the steady and linear (i.e. geostrophic) version of the previous vorticity equation

$$\frac{D}{Dt}\left(f+\zeta\right) = -\left(f+\zeta\right)\left(\frac{\partial u}{\partial x}+\frac{\partial v}{\partial y}\right).$$

So for purely geostrophic flows (remember, they are slowly-evolving and large-scale), a horizontal convergence or divergence has to be balanced by a north-south translation, to change the planetary vorticity f.

Rewriting the horizontal convergence, using the continuity equation, gives

$$\beta v_g = f\frac{\partial w}{\partial z},$$

If we now assume that vertical velocities are negligible at the ocean bottom or at some other great depth and integrate this equation up vertically from there, up to the level z_0 of the bottom of the surface Ekman layer, we get

$$\beta V_g \;\; = \;\; f w(z_0),$$

where V_g is the depth-integrated meridional geostrophic flow. Plugging in the expression for the surface Ekman pumping velocity gives

$$V_g = \frac{f}{\beta\rho_0}\left[\frac{\partial}{\partial x}\left(\frac{\tau_y^w}{f}\right)-\frac{\partial}{\partial y}\left(\frac{\tau_x^w}{f}\right)\right].$$

If we now add the contribution to the vertical transport from the Ekman flow,

$$V_E = -\frac{\tau_x^w}{f\rho_0},$$

we get that the *total* depth-integrated meridional flow is

$$\begin{aligned} V &= V_g + V_E \\ &= \frac{1}{\beta\rho_0}\left[\frac{\partial\tau_y^w}{\partial x}-\frac{\partial\tau_x^w}{\partial y}\right], \end{aligned}$$

or, in vector form,

$$V = \frac{1}{\beta\rho_0}\nabla_H\times\boldsymbol{\tau}^w.$$

As it turns out, we could have reached this relationship more directly by *not* separating the ocean into a surface Ekman layer and a geostrophic interior below. Taking the curl of the original momentum equations (23) and (24) gives

$$\beta v = f\frac{\partial w}{\partial z} + \frac{1}{\rho_0}\frac{\partial}{\partial z}\left(\frac{\partial \tau_{yz}}{\partial x} - \frac{\partial \tau_{xz}}{\partial y}\right),$$

where we have switched the order of the derivatives in the last term. If we now vertically integrate this from some great depth (where both the vertical velocity and friction are negligible) and up, and not only to the bottom of the Ekman layer but all the way to the sea surface η, we get

$$\beta V = fw(\eta) + \frac{1}{\rho_0}\left(\frac{\partial \tau_y^w}{\partial x} - \frac{\partial \tau_x^w}{\partial y}\right).$$

Assuming now that the steady vertical velocity at the sea surface is very small (if not, the sea surface would rise or drop indefinitely), we end up with the same result as above, namely

$$\beta V = \frac{1}{\rho_0}\left(\frac{\partial \tau_y^w}{\partial x} - \frac{\partial \tau_x^w}{\partial y}\right),$$

or,

$$V = \frac{1}{\beta\rho_0}\nabla_H \times \boldsymbol{\tau}^w.$$

So, either way, the total vertically-integrated meridional transport is dictated by the curl of the wind stress. Where the wind stress curl is positive (giving Ekman suction), this *Sverdrup relation* dictates that the total depth-integrated transport is northward. And this holds regardless of which hemisphere we are on. Where there is a negative wind stress curl (Ekman pumping), the tranport is southward. Figure 69 shows the wind stress curl over the world oceans, estimated by observations and models. We see that the curl is negative in the bands between approximately 20 and 40 degrees north, in both the Pacific and Atlantic oceans. This is in general agreement with the meridional component of surface velocities in the gyres at these latitutdes shown in Figure 68.

Note that the Sverdrup relation only gives the meridional component of the depth-integrated transport. But if we assume that the total depth-integrated transport is divergence-free,

$$\frac{\partial U}{\partial x} + \frac{\partial V}{\partial y} = 0,$$

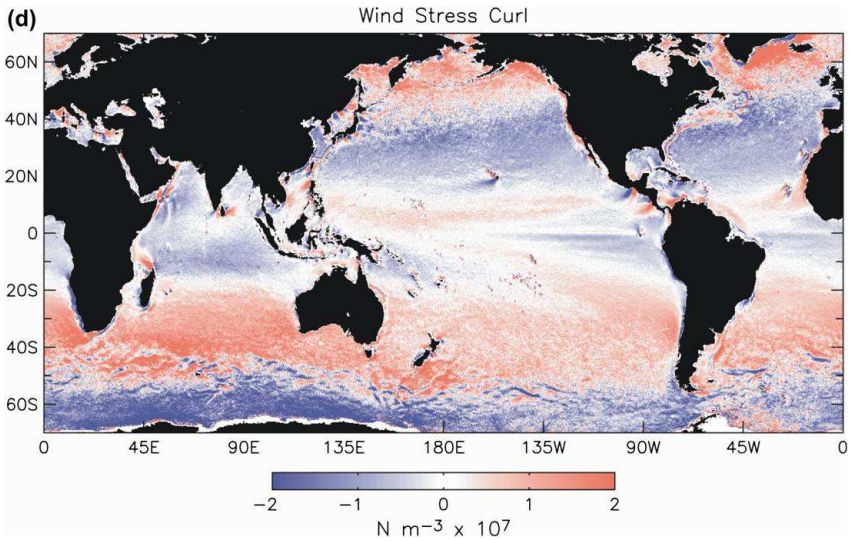

Figure 69: The time-mean wind stress curl over the oceans. (Source: Talley et al., 2011, Fig. 5.16[Talley et al.(2011)Talley, Pickard, Emery, and Swift])

and this is a pretty safe bet at long time scales (otherwise, as mentioned above, the sea surface would rise or drop forever), then we can get the zonal transport by simple integration of the depth-integrated continuity equation:

$$U(x) = U(x_0) - \int_{x_0}^{x} \frac{\partial V}{\partial y}\, dx'.$$

A natural choice is to integrate from a continental boundary where the kinematic boundary condition is zero flow into or out of that boundary. For an eastern or western boundary which is alligned north-south at x_0 we would simply have $U(x_0) = 0$. And for a boundary of arbitrary angle, $U(x_0)$ can also be found from the same kinematic boundary condition applied to the flow-component normal to the boundary. But there is a catch. Most oceans have two continental boundaries, one in the east and another in the west. But the equation cannot satisfy two such kinematic boundary conditions (it is a first-order differential equation, and such equations require exactly one boundary condition). One must either integrate from the eastern or from the western boundary.

The two different 'Sverdrup flow' solutions are shown in Figure 70. We see that if the no-normal-flow boundary condition is applied at the wall where the integration starts from, then the flow will 'run into' the boundary at the other side of the ocean. We can't have that! Simply put, the Sverdrup model cannot apply

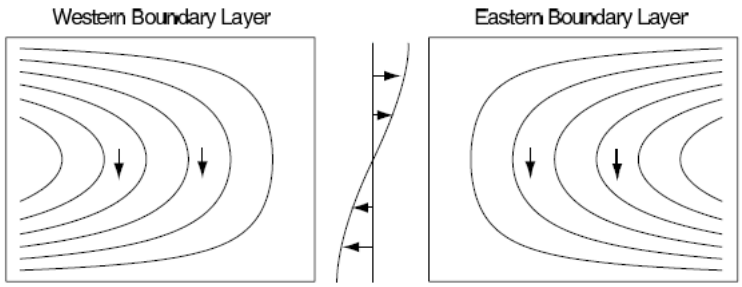

Figure 70: The Sverdrup flow resulting from an anti-cyclonic wind stress, either integrated from the eastern boundary (left panel) or from the western boundary (right panel). (Source: Vallis, 2006, Fig. 14.3)

near the opposite boundary. We need additional dynamics there to close the circulation. But which boundary do we start the integration from, the western or the eastern one? The answer to this question is directly tied to the east-west asymetry in the shape of these wind-driven gyres and the fact that they have western rather than eastern boundary currents, as seen in Figure 68. It was Henry Stommel who first extended Sverdrup's model to offer an explanation for such western boundary currents.

7.3.2 Western boundary currents

What Stommel did was introduce additional friction to the model. This does two things: 1) it allows the energy which is input by the winds to be dissipated somewhere and 2) it raises the order of the differential equation so that it will need two boundary conditions (one at each coast). The effect of introducing friction to the Sverdrup model can be illustrated in various ways, but Stommel considered bottom friction specifically (because it gives a particularly simple solution) in an ocean without density stratification.

As before, the geostrophic momentum equations with vertical friction added are

$$\begin{aligned} -fv &= -\frac{1}{\rho_0}\frac{\partial p}{\partial x} + \frac{1}{\rho_0}\frac{\partial \tau_{xz}}{\partial z} \\ fu &= -\frac{1}{\rho_0}\frac{\partial p}{\partial y} + \frac{1}{\rho_0}\frac{\partial \tau_{yz}}{\partial z}. \end{aligned}$$

Taking the curl (to obtain a vorticity equation) and integrating vertically through

the entire water column, and now also allowing for bottom friction, gives

$$\begin{aligned}\beta V &= \frac{1}{\rho_0}\left(\frac{\partial \tau_y^w}{\partial x} - \frac{\partial \tau_x^w}{\partial y}\right) - \frac{1}{\rho_0}\left(\frac{\partial \tau_y^b}{\partial x} - \frac{\partial \tau_x^b}{\partial y}\right) \\ &= \frac{1}{\rho_0}\left(\nabla \times \boldsymbol{\tau}^w\right) - \frac{1}{\rho_0}\left(\nabla \times \boldsymbol{\tau}^b\right),\end{aligned}$$

where $\boldsymbol{\tau}^b$ is the bottom friction. Stommel assumed that this is proportional to the velocity at the bottom, i.e. $\boldsymbol{\tau}^b/\rho_0 = R\boldsymbol{u}_b$, where $\boldsymbol{u}_b$ is the bottom velocity and R is a bottom friction coefficient. But note that in an ocean without a vertical density stratification, the horizontal velocities are depth-independent. Stommel assumed this. So we just write $\boldsymbol{u}_b = \boldsymbol{u}$ (the same at all depths) and get

$$\begin{aligned}\beta V &= \frac{1}{\rho_0}\left(\frac{\partial \tau_y^w}{\partial x} - \frac{\partial \tau_x^w}{\partial y}\right) - R\left(\frac{\partial v}{\partial x} - \frac{\partial u}{\partial y}\right) \\ &= \frac{1}{\rho_0}\left(\nabla \times \boldsymbol{\tau}^w\right) - R\nabla \times \boldsymbol{u},\end{aligned}$$

showing that the bottom friction is given by the curl of the bottom velocity, i.e. the relative vorticity of the bottom flow.

In fact, bottom friction creates a bottom Ekman layer, just as friction (the wind stress) creates one near the sea surface. It can be shown that the geostrophic vertical velocity out of the bottom Ekman layer is

$$\begin{aligned}w(z_{0,b}) &= -\left(\frac{\partial U_{E,b}}{\partial x} + \frac{\partial V_{E,b}}{\partial y}\right) \\ &= R\left[\frac{\partial}{\partial x}\left(\frac{v}{f}\right) - \frac{\partial}{\partial y}\left(\frac{u}{f}\right)\right] \\ &= R\nabla \times \left(\frac{\boldsymbol{u}}{f}\right).\end{aligned}$$

So vertical flow out of the bottom Ekman layer, in other words bottom Ekman pumping, can also influence the interior geostrophic flow, just as surface Ekman pumping can. The difference is that whereas the vertical flow out of the surface layer is dicated by the curl of the winds, so *forced* by the winds, the bottom Ekman pumping is a result of the ocean flow itself—which is what we are trying to solve for.

We are now in a position to argue why boundary currents have be on the western rather than the eastern sides of these wind-driven gyres. The key assumption of Stommel was that the flow in the interior of the gyres is so sluggish that bottom friction is negligible there (remember that in his model the friction is proportional to the strength of the flow). So, in the 'Sverdrup interior' we get a meridional flow dictated by the sign of the wind stress curl. Thus, where the wind stress curl is negative there will be a southward flow. But all this water eventually has to return to the north again, and it does so in a narrow and swift boundary current—either along the eastern or the western boundary. And there, because of the swift speeds in such a boundary current, the bottom friction cannot be ignored. So energy is input by the winds everywhere but only dissipatied via bottom friction in the boundary current.

But why must the boundary current be in the west and not in the east? Well, assuming the wind stress curl is the same everywhere, i.e. negative in our example here, then this can not balance βV in the boundary current since this term is positive there for a northward return flow. So only internal friction can balance βV in the boundary current, and for $\beta V > 0$ (a boundary current returning water to the north) we must have $(\partial v/\partial x - \partial u/\partial y) < 0$. Now, it is clear that relative vorticity in such a meridional boundary layer is dominated by $\partial v/\partial x$ and that the other component, $-\partial u/\partial y$, can safely be neglected. Figure 71 then shows the relative vorticities in boundary layers that are either on the eastern or the western sides of an ocean confined between two continents. We see that $\partial v/\partial x > 0$ in an eastern boundary current and $\partial v/\partial x < 0$ in a western one. A western boundary current is the only possibility which produces a closed circulation!

Figure 72 shows an estimate of the depth-integrated wind-driven transport that results from integrating the Sverdrup transport from the eastern boundaries of all continents. The product is made using a wind-stress curl pattern similar to that shown in Figure 69. There is clearly a qualitative resemblence with the direct estimate of the geostrophic flow shown in Figure 68, a sign that the fantastically simplified (!) theory of Sverdrup and Stommel contains some key dynamical elements of large-scale wind-driven flows on a rotating planet.

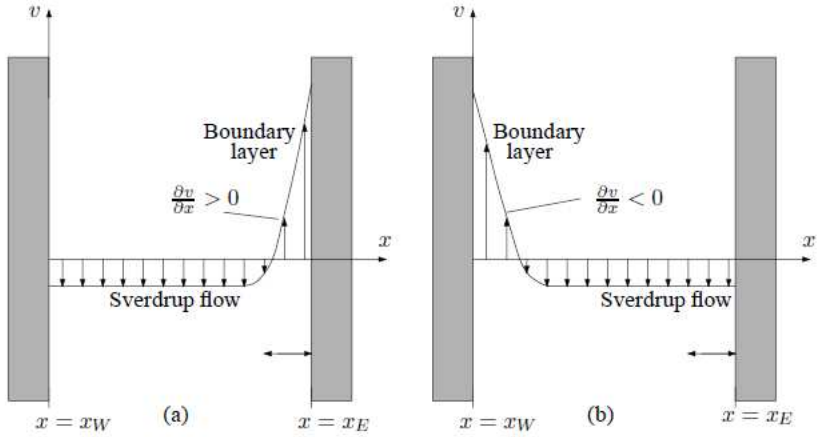

Figure 71: (Source: Cushman-Roisin and Beckers, 2011, Fig. 20.7)

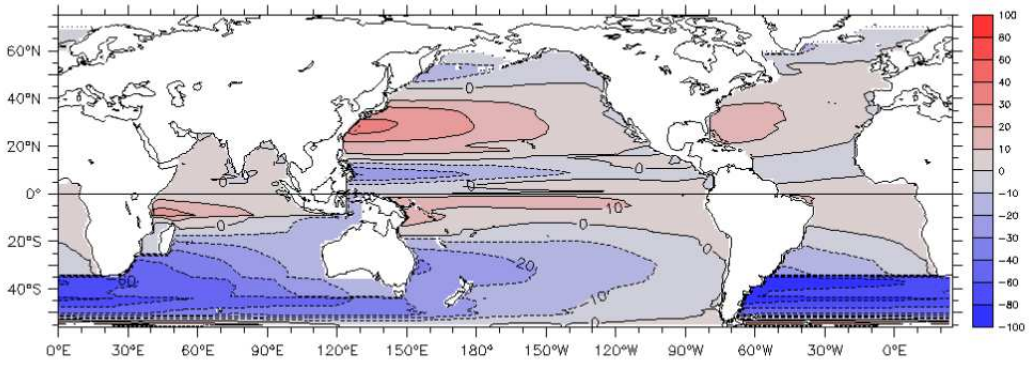

Figure 72: The global wind-driven circulation estimated by integrating the Sverdrup relation from real winds then and assuming western boundary currents.

8 The large-scale buoyancy-driven circulation

8.1 The need for both surface fluxes and turbulent vertical mixing

Imagine a tank of isothermal (constant temperature) water. Then imagine that we start cooling the sea surface in one end of the tank, say in the 'northern' end. The water there becomes colder and denser and starts to sink and slide underneath the warmer water and 'southward'. And warmer water flows northward along the surface to replace the cold water which has sunk. A buoyancy-driven overturning circulation is formed and the circulation extends to the bottom of the tank since nothing prevents the cooled water from sinking all the way down (it is the densest water after all). Now repeat almost the same thought experiment, but imagine instead warming the sea surface in the southern end of the tank. What happens? The warmed-up water gets more buoyant and flows northward to spread itself on top of the rest of the water in the tank. But how deep does the circulation reach now? If the warming in the south only acts on the water molecules at the very surface, then the northward flow will be super-thin. In fact, if one warms the surface in the south and cools the surface in the north of the tank (as we observe on a real planet) the northward surface flow would still be super thin. But when oceanographers go out and observes the poleward flow of buoyant water in the real oceans they find that it extends over hundreds of meters of depth.

J. W. Sandström, a Sweedish oceanographer, actually observed such buoyancy-driven flows in his laboratory in the first decade of last century. After studying a set of experiments where buoyancy was controlled by heating and cooling, he concluded that "a circulation [of any significant strength] can develop from thermal causes *only* if the level of the heat source lies below the level of the cold source". Surely, if the heating in the south took place through the bottom of the tank while the cooling in the north continued to take place at the surface, a vigurous circulation could be maintained. But on the real Earth heating of the ocean at low latitudes does take place through the surface (heating from geothermal vents is tiny in comparison). So what is going on?

As we have discussed earlier, the *thermohaline circulation* is essentially a form of slantwise convection. Buoyant water (the warm water in our thought experiment and in SandstrÃ¶m's laboratory experiment) needs to rise and dense water needs to sink. If the buoyant water is all produced near the sea surface, it can only move laterally and no available potential energy (APE) can be released. Some mechanism needs to get the buoyant water down (deeper) in low latitudes, so that

it can rise on its way towards high latitudes. We have seen how shortwave radiation can penetrate a few meters, perhaps a few tens of meters, into the water column. But buoyant waters in the real oceans are found much deeper than this, down to several hundred meters. As it turns out, on Earth (and probably on other planets) buoyant waters are created at the sea surface by air-sea fluxes (heating or precipitation) but are then is transported to greater depths. This downward transport of buoyant waters can be done by Ekman pumping or by vertical turbulent mixing. But in both cases the processes increases the gravitational potential energy locally (forcing buoyant water downward in a stably-stratified fluid). So the transport process requires an external input of mechanical (kinetic) energy. This can come from the winds (which can drive both Ekman pumping and also generate turbulence) or it can come from other sources of turbulence, most notably from energetic tidal currents.

So the thermohaline or buoyancy-driven circulation requires two things: 1) uneven buoyancy fluxes between low and high latitudes (a buoyancy loss at high latitudes) and 2) a mechanical energy source at low latitudes for 'pumping' buoyant waters there down. From this depth, the buoyant waters can move poleward and upward, in other words, participate in global-scale slantwise convection. The entire circulation relies on a mechanical energy source (the one pumping buoyant waters down in low latitudes), but the end result is a huge transport of thermal energy (many orders of magnitude larger, in terms of Watts, than the mechanical energy cost) to high latitudes. The global thermohaline circulation is, in fact, acting much like a refrigirator or heat pump. A small amount of mechanical energy is required to drive the pump/engine that transports heat from a warm to a cold reservoir.

8.2 Deep western boundary currents

The 2D plots we have seen of the thermohaline circulation so far give little clue about the horizontal structure of the flow. What does this look like on a rotating planet? Is the thermohaline circulation also associated with east-west asymmetries and even western boundary currents? As a matter of fact, it is. And again it was Henry Stommel with colleagues who proposed the first and simplest models for this.

We will study the horizontal structure of the lower-limb of the thermohaline overturning circulation. The basic assumption of Stommel was that the sinking of dense water happens in very localized convection regions at high latitudes but that, in contrast, the mixing-induced upwelling of this dense water takes place over the

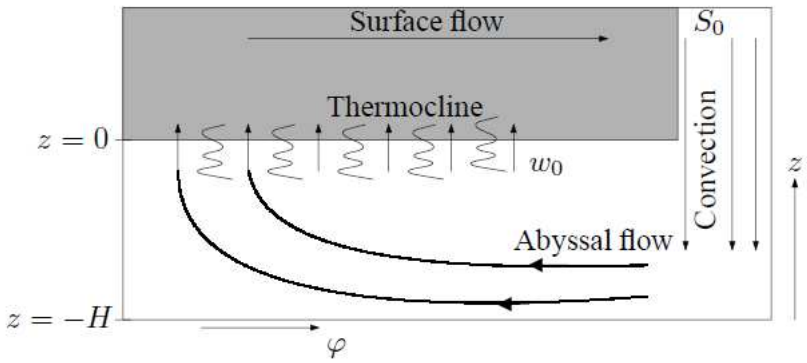

Figure 73: Schematics of buoyancy-driven flows, requiring uneven surface buoyancy fluxes, a small region where the dense waters sink and a broad region where the dense waters can upwell by mechanically-induced turbulent mixing. (Source: Cushman-Roisin and Beckers, 2011, Fig. 20.10)

world ocean as a whole. So the sinking is localized while the upwelling is broad, as illustrated in Figure 73. Stommel then proceeded to study the vorticity budget of the lower layer of this overturning circulation.

The vorticity equation is vertically integrated from the bottom $z = -H$ to the top of the lower layer (this level is defined as $z = 0$ in the figure). We allow for bottom friction but now leave out surface Ekman transport (which doesn't reach the lower layer). At the top of the layer we instead introduce a vertical velocity $w(0)$ which represents upwelling to the layers above. This gives

$$\beta V = f w(0) - R \left(\frac{\partial v}{\partial x} - \frac{\partial u}{\partial y} \right). \tag{28}$$

Here we haven't restricted our analysis to the beta plane, but β is still the meridional gradient of the Coriolis parameter at any given latitude. If we now assume that internal friction is negligible in the the inerior flow, away from boundary currents, we get

$$V = \frac{f}{\beta} w(0).$$

Note that the meridional flow in the interior (away from the boundary currents) can not cross the equator where f vanishes. But everywhere else the full circulation, including the zonal component, can be found just as for the Sverdrup interior

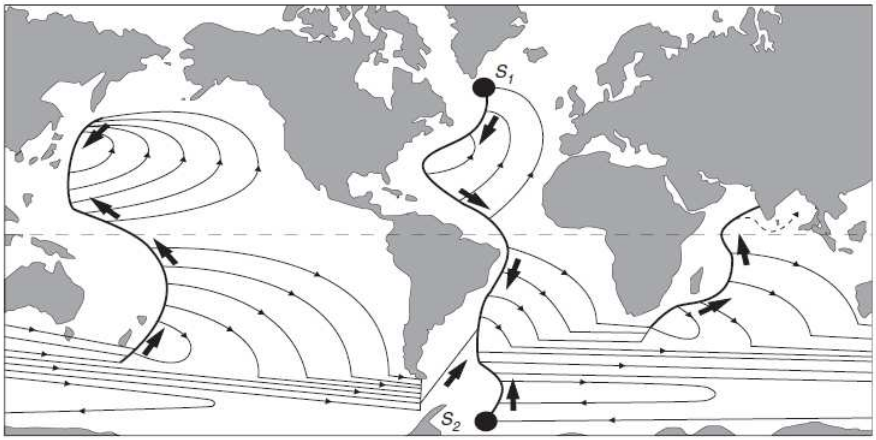

Figure 74: Stomme's simplified estimate of lower-layer flows for the large-scale buoyancy-driven circulation. The convective source regions are indicated by fat black dots. (Source: Stewart, 2008, Fig. 13.4)

discussed above (including the integral which gives the zonal component of the flow).

So for upwelling in the interior we have $w(0) > 0$ and hence $V > 0$. But this seems strange since our intuition from the 2D plots is that the deep flow is away from the poles. Here vorticity considerations suggest that the flow is poleward. The aparent paradox is easily resolved, however, by introducing a boundary current where friction enters the vorticity budget. Using similar arguments as for wind-driven gyres, it can be shown that equatorward boundary currents can only be on the western side of the oceans. And since f doesn't enter the frictional term, at least not as specified in (28), the boundary currents can cross the equator.

Figure 74 shows Stommels estimate of the deep flow given two high-latitude sources, one in the North Atlantic and another in the Weddel Sea north of Antarctica. We see that the interior flows in the lower layer is everywhere poleward and that the 'freshly-made' dense water only travels equatorward in deep western boundary currents (DWBC). Whether these boundary currents cross the equator or not depends on the relative strength of the high-latitude sinking to the upwelling in the rest of the ocean. We see for example that the North Atlantic DWBC extends into the southern hemisphere and that waters formed in the Weddel Sea crosses the equator and flow into the North Pacific (which doesn't have its own deep water production site).

There is no doubt that Stommel's model for the abyssal circulation is overly

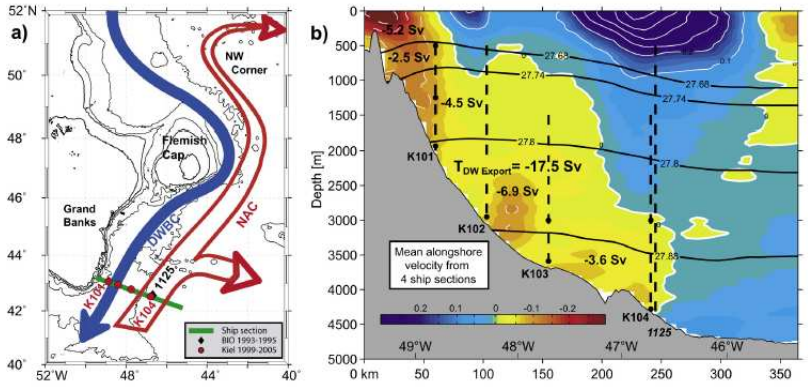

Figure 75: Observations of along-shore flow off the Canadian east coast, i.e. along the western boundary of the North Atlantic ocean. The left panel shows schematic currents and the section where currents were measured. The right panel shows along-shore flow, where yellow and red colors indicate southward flow while blue colors indicate northward flow. (Source: Schott et al., 2006, Fig. 1)

simplistic. But the presence of *deep western boundary currents* is an observational fact. Figure 75 shows velocity measurements through a transect which crosses the western margins of the North Atlantic ocean at about $43°\mathrm{N}$, off the Canadian east coast. The observations clearly show a northward-flowing surface current, thought to represent the wind-driven western boundary current, but also a southward-flowing boundary current at depth. An additional surface-intensified southward flow is associated with light waters that exit the Arctic Ocean due to sea ice melt and excessive river run-off there.

Hydrographic sections off the western boundary of the South Atlantic at about $30°\mathrm{S}$ (Figure 76) also show evidence of southward-flowing North Atlantic Deep Water (NADW) and, below this, northward-flowing Antarctic Bottom Water (AABW). Both of these deep western boundary currents were predicted by Stommel's model although in the model AABW did not reach this latitude.

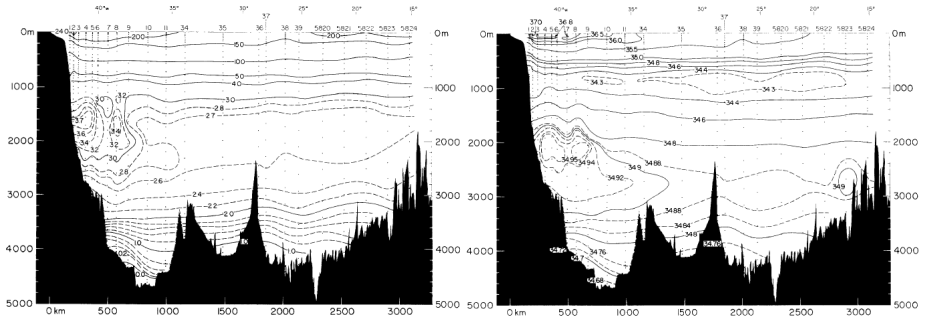

Figure 76: Hydrographic observations from a section along 30°S in the South Atlantic: potential temperature (left) and salinity (right). (Source: Warren, 1981, Fig. 1.10)

9 Ocean waves

We have all seen waves on the ocean surface. But what are waves and what is their purpose? Wikipedia says:

> "A wave is an oscillation accompanied by a transfer of energy that travels through a medium."

And one could add that it is a transfer of energy which is not primarily associated with a transfer/translation of the medium itself. So a wave is an oscillation that travels through a medium, like water, without moving the water much. The water parcels move up and down (acutally in ellipses, as we'll see), but they basically end up at their original position again. And still, energy can be transfered by the wave from one place to another, even to the other side of the planet.

In the following we will be looking at ocean waves that can exist in water whose density is constant, so basically waves that are associated with oscillations of the sea surface (in the last section we will have a quick look at so-called internal waves that exist in a denisty-stratified fluid).

9.1 Wave kinematics

A wave doesn't have to be a sinusoid, but this is the canonical form we will use. Specifically, a *monochromatic* (one single wavelength) *plane wave* traveling in the x-direction, along the sea surface (Figure 77), may be written as

$$\eta(x,t) = a\cos\left(kx - \omega t\right),$$

where the amplitude a is the height of the wave (half the height from through to crest) and *wavenumber* k and *angular frequency* ω are the reciprocals of the wavelength λ and wave period T, respectively:

$$\begin{aligned} k &= \frac{2\pi}{\lambda} \\ \omega &= \frac{2\pi}{T}. \end{aligned}$$

To see what the wavelength is, imagine freezing time and then study the shape of the wave in space. The wavelength is then the physical distance (in meters) between two neighboring wave crests or between two neighboring wave troughs. What about the wave period? Well, now instead imagine observing a wave as it

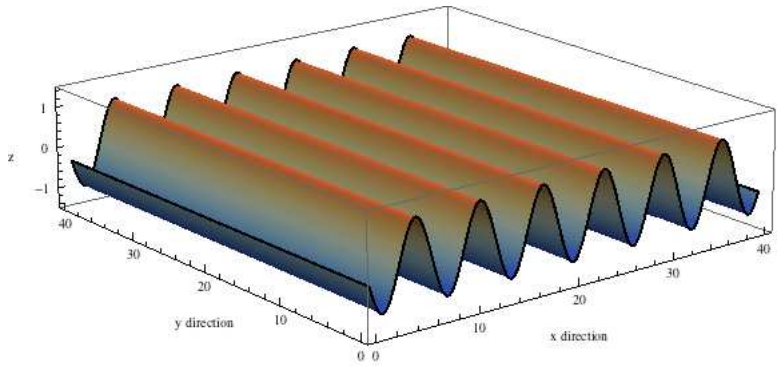

Figure 77: A monochromatic plane wave travelling in the x-direction. (Source: Wikipedia)

travels past a fixed point in space. The period is the time (in seconds) between the passing of two wave crests or two troughs.

A wave that travels in an arbitrary direction in the x-y plane can be represented as

$$\eta(x, y, t) = a \cos(kx + ly - \omega t).$$

The propagation direction is given by the *wavevector*

$$\boldsymbol{K} = k\boldsymbol{i} + l\boldsymbol{j},$$

where k is the wavenumber in the x-direction and l the wavenumber in the y-direction. But in most of what follows we will simplify by looking at waves that travel in the x-direction. Or, alternatively, study waves after we have rotated our coordinate system such that the (rotated) x-axis is directed along the wave vector.

Phase velocity The *phase speed* of a wave is the speed at which the crests (or troughs) move through the medium. A crest has moved by one wavelength in one period, so

$$\lambda = cT,$$

where c is the phase speed. Substituting in the expressions for k and ω above gives

$$c = \frac{\omega}{k}.$$

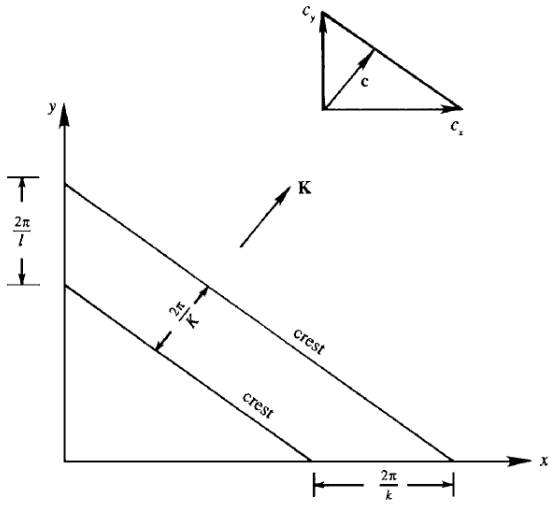

Figure 78: The relationship between wave vector, wavelength and phase velocity for a plane wave in the x-y plane. (Source: Kundu and Cohen, 2004, Fig. 7.3)

And for a wave that travels in the direction of wavevector $\boldsymbol{K} = k\boldsymbol{i} + l\boldsymbol{j}$, the phase speeds in the x and y directions are

$$\begin{aligned} c_x &= \frac{\omega}{k} \\ c_y &= \frac{\omega}{l}. \end{aligned}$$

So c_x is the speed at which a crest of a wave advances in the x-direction and c_y the speed that the crest advances in the y-direction.

Waves whose phase speed is a function of the wavelengths (or wavenumbers) are called *dispersive* waves. This is because waves with different wavelengths will then separate or disperse due to the different travel speeds. Waves we observe when we throw a rock into a lake are dispersive. We easily see that long waves travel faster away from the impact than shorter waves.

Group velocity As it turns out, the wave energy does not travel with the phase velocity but rather with what is called the *group velocity*. To see what is meant by a *wave group* and its velocity, imagine the sum of two waves, each with slightly different frequency and wavenumber (but same amplitude, which we here set equal to one). So instead of having only one wave with wavenumber and frequency k and ω, we have one wave with wavenumber and frequency $k_1 = k - \Delta k$ and

$\omega_1 = \omega - \Delta\omega$ and another with $k_1 = k + \Delta k$ and $\omega_1 = \omega + \Delta\omega$, where Δk and $\Delta\omega$ are small compared to k and ω. The sum of the two waves becomes

$$\eta = \cos\left[(k - \Delta k)\, x - (\omega - \Delta\omega)\, t\right] + \cos\left[(k - \Delta k)\, x - (\omega - \Delta\omega)\, t\right].$$

Now there is a trigonometric identity which says that

$$\cos(a \pm b) = \cos a \cos b \mp \sin a \sin b,$$

which, when applied, gives

$$\begin{aligned} \eta &= \cos(kx - \omega t)\cos(\Delta kx - \Delta\omega t) - \sin(kx - \omega t)\sin(\Delta kx - \Delta\omega t) \\ &\quad + \cos(kx - \omega t)\cos(\Delta kx - \Delta\omega t) + \sin(kx - \omega t)\sin(\Delta kx - \Delta\omega t) \\ &= 2\cos(kx - \omega t)\cos(\Delta kx - \Delta\omega t). \end{aligned}$$

So this looks like the product of two waves. Actually what we have is a wave with wavenumber k and ω whose amplitude is *modulated* by a 'wave envelope' having wavenumber and frequency Δk and $\Delta\omega$, as shown in Figure 79. As before, the speed of the fast wiggles (this is really our wave) is $c = \omega/k$, but the envelope of the wave, the "wave group", is moving with speed $\Delta\omega/\Delta k$. In the limit of very small changes to the wavenumber and frequency, we get the *group speed*

$$c_g = \frac{\partial\omega}{\partial k}.$$

The energy of the wave is proportional to the square of the wave amplitude. And, as discussed above and as show in in Figure 79, this amplitude is modulated by the wave envelope, the group. So the movement of this envelope gives the movement of the wave energy. Where the envelope is zero there is still our basic wave, $\cos(kx - \omega t)$, but nobody would be able to feel it—the wave energy there is zero. The group velocity doesn't need to be the same as the phase velocity, but for some waves it is, as we'll see examples of below.

As for phase velocity, when the wave has an arbitraty wave vector $\boldsymbol{K} = k\boldsymbol{i} + l\boldsymbol{j}$, there is an x and a y component of the group velocity:

$$\begin{aligned} c_{g,x} &= \frac{\partial\omega}{\partial k} \\ c_{g,y} &= \frac{\partial\omega}{\partial l}. \end{aligned}$$

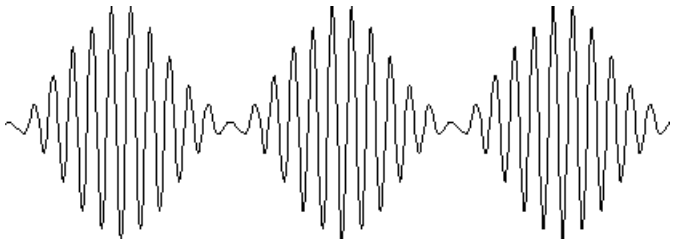

Figure 79: Wave groups. (Source: Wikipedia)

The wave spectrum Ocean waves of course don't only come with one frequency and one wavenumber. If one measures the evolution of sea surface height at one fixed location, waves of a range of frequencies will be observed, each one with its own amplitude. So the total time series can be written as a sum of waves:

$$\eta(t) \;=\; \frac{a_0}{2} + \sum_{n=1}^{\infty} \left[a_n \cos\left(\omega_n t\right) + b_n \sin\left(\omega_n t\right) \right],$$

where the very first term allows for a time-independent component to the sea surface (the mean sea level). Note that this can also be written

$$\eta(t) = \sum A_n \cos\left(\omega_n t + \phi_n\right),$$

for amplitude

$$A_n = \sqrt{a_n^2 + b_n^2}$$

and phase

$$\phi_n = \tan^{-1}\left(b_n / a_n\right).$$

What we call the *wave spectrum* is a plot of these coefficients, A_n or ϕ_n, as a function of frequency. Most often studied, the magnitude spectrum (A_n vs. ω_n) basically shows how wave energy is distributed over the various frequencies, in other words, which waves (frequencies) are energetic and which are not. Conversely, by freezing time and studying the spatial scales of waves, one can also form the wavenumber spectrum:

$$\eta(x) = \frac{a_0}{2} + \sum_{n=1}^{\infty} \left[a_n \cos\left(k_n x\right) + b_n \sin\left(k_n x\right) \right],$$

where the coefficients are now different from those forming the frequency spectrum.

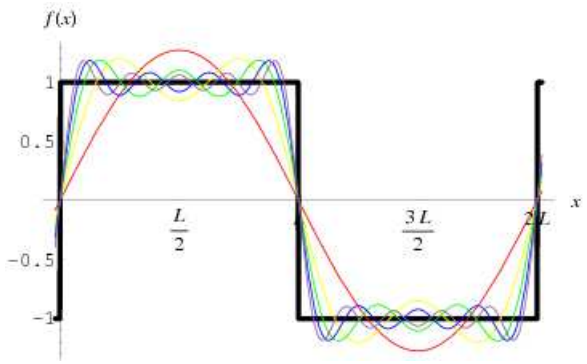

Figure 80: The Fourier series representation of a square wave (thick black line), using the sum of either a single sinusoid (blue line), having the same wavelength as the square wave, or the sum of two to five sinusoids, each of different wavelength. (Source: http://mathworld.wolfram.com)

Finally, it is useful to keep in mind that *any* period signal, not just those that look like sinusoids, can in fact be written as sums of sines and cosines. This was proven by Joseph Fourier (1768–1830), and such representations are therefore called *Fourier series*.

9.2 High-frequency ocean waves

In the following we will distinguish waves into *high-frequency waves* whose frequency ω is much higher than the Coriolis parameter f and *low-frequency waves* whose frequency is comparable to or lower than f. In both cases we will only look at *linear waves*, that is waves whose sea surface height amplitude is small compared to the ocean depth and whose water velocity (the velocity with with actual water parcels associated with the wave are moving) is small compared to the phase velocity.

For high-frequency *surface gravity waves*, the starting point is the linear Boussinesque but non-hydrostatic equations for a constant-density, non-viscous (we ignore friction) fluid. For waves having $\omega \gg f$, we can drop the Coriolis acceleration from the momentum equations. If we wish to study waves that travel in the

x-direction, the governing equations become:

$$\begin{aligned} \frac{\partial u}{\partial t} &= -\frac{1}{\rho_0}\frac{\partial p}{\partial x} \\ \frac{\partial w}{\partial t} &= -\frac{1}{\rho_0}\frac{\partial p}{\partial z} - g \\ \frac{\partial u}{\partial x} + \frac{\partial w}{\partial z} &= 0, \end{aligned}$$

with kinematic boundary conditions at the bottom ($z = -H$) and at the sea surface ($z = \eta$). The wave motions—oscillations—arrise when a displaced sea surface is restored back due to the force of gravity...and overshoots.

We won't solve these equations here but go straight to the results. If we assume a wave form for the free surface and, for now, limit ourselves to a single wave that travels in the x-direction, so

$$\eta(x,t) = a\cos\left(kx - \omega t\right),$$

then the equations give us what's called the *dispersion relation* for the wave. This is a functional relationship between frequency and wavenumber, and for such a fast surface gravity wave it is

$$\omega^2 = gk\tanh\left(kH\right).$$

This expression is slightly complicated, but two limiting cases are very easy to deal with. For waves whose wavelength is much smaller than the ocean depth—*deep-water waves*—we have $kH \gg 1$ and $\tanh\left(kH\right) \sim 1$. This gives dispersion relation

$$\omega^2 = gk.$$

The phase speed is therefore

$$\begin{aligned} c &= \omega/k \\ &= \sqrt{g/k}, \end{aligned}$$

and the waves are dispersive (since the speed is wavenumber-dependent). Long waves travel faster than short waves, as illustrated in Figure 81. The group velocity of these deep-water waves is

$$\begin{aligned} c_g &= \frac{\partial}{\partial k}\left(\sqrt{gk}\right) \\ &= \frac{c}{2}. \end{aligned}$$

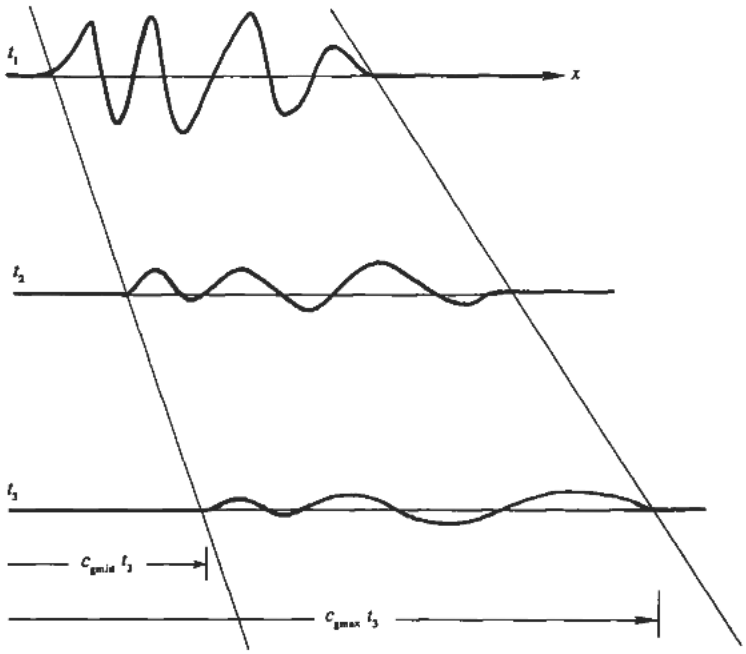

Figure 81: The sea surface at three consecutive times after throwing a rock into a pond. Deep-water waves are dispersive, with long waves travelling faster than short waves. (Source: Kundu and Cohen, 2004, Fig. 7.19)

So the group speed is half the phase speed! If one observes a group of waves, one would see individual waves passing through the group, leaving the group behind.

In the other limit, that of waves whose wavelength is much larger than the ocean depth, i.e. for $kH \ll 1$, we get that $\tanh(kH) \sim kH$. This gives

$$\omega^2 = gk^2H,$$

so that the phase speed is

$$\begin{aligned} c &= \omega/k \\ &= \sqrt{gH}. \end{aligned}$$

Hence, these waves are non-dispersive, and their phase speeds only depend on the water depth. The waves essentially travel faster as the depth increases. For these waves the group speed is

$$\begin{aligned} c_g &= \frac{\partial}{\partial k}\left(k\sqrt{gH}\right) \\ &= c. \end{aligned}$$

The group velocity is exactly the same as the phase velocity. How convenient!

So the waves travel with speeds c and c_g, but what about actual water parcels? Again, we will not go through the algebra, but for the same assumed cosine wave at the sea surface the horizontal and vertical velocity field is

$$\begin{aligned} u(x,z,t) &= a\omega\frac{\cosh k(z+H)}{\sinh kH}\cos(kx-\omega t) \\ w(x,z,t) &= a\omega\frac{\sinh k(z+H)}{\sinh kH}\sin(kx-\omega t). \end{aligned}$$

So these are the velocity components that water parcels are exposed to. As we can see, they have nothing to do with the phase or group speeds of the wave phenomenum itself. To find the time-evolving position of a fluid parcel centered around 'resting position' $(x_0,\, z_0)$ we integrate these expressions in time. This gives

$$\begin{aligned} x'(t) &= -a\frac{\cosh k(z+H)}{\sinh kH}\sin(kx-\omega t) \\ z'(t) &= a\frac{\sinh k(z+H)}{\sinh kH}\cos(kx-\omega t). \end{aligned}$$

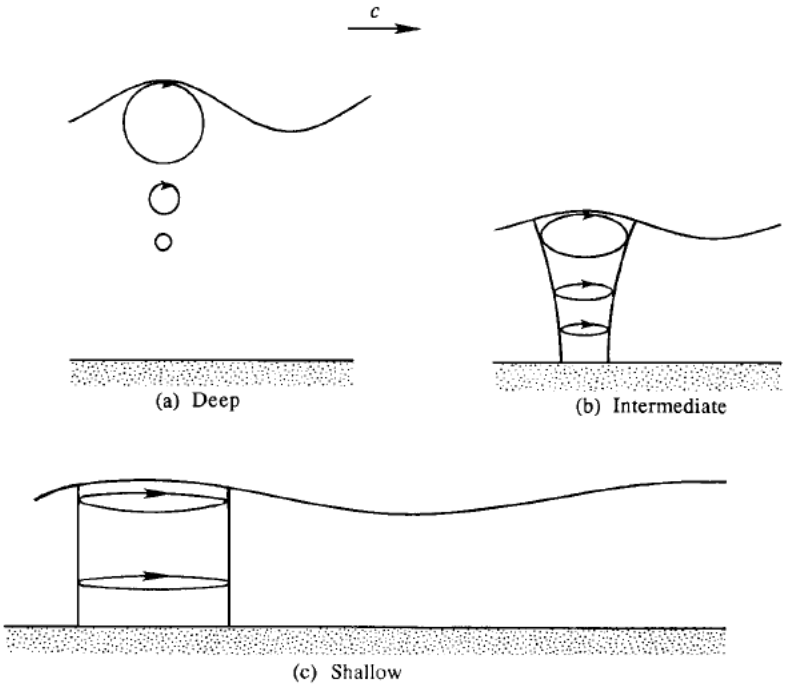

Figure 82: Particle trajectories of surface gravity waves for a) deep water, b) intermediate water and c) shallow water waves. (Source: Kundu and Cohen, 2004, Fig. 7.6)

So the fluid parcels move in ellipses, as shown in Figure 82. The detailed behavior depends on the ratio of wavelength to ocean depth. For deep-water waves, i.e. for $kH \gg 1$, the ellipses are nearly circles that decay exponentially with depth. For waves of intermediate wavelength compared to the depth, so for which $kH \sim 1$, there water parcels trace out clear ellipses. And, finally, for shallow-water waves, $kH \ll 1$, the ellipses are almost completely squished and don't change their size with depth.

9.2.1 Wind-driven surface gravity waves

The ocean waves most familiar to us are surface gravity waves generated by the winds. If the winds start blowing on a very calm sea surface it is not actually gravity waves that first get generated but *capillary* waves in which the restoring force is surface tension rather than gravity. But these capillary waves grow fast, also in size, and eventually turn into gravity waves.

Growth As the winds blow the ocean surface gets rougher and rougher, in other words the amplitude of the waves grow. We can understand this conceptually by realizing that a wavy ocean surface is more 'rough', such that the frictional coupling between atmosphere and ocean increases. It's like increasing the drag coefficient, and the result is even stronger forcing of the ocean wave field. As just about anybody has observed, the wave field grows faster with stronger winds. It also grows over time, as long as the winds blow, until there is some form of balance between the energy input by the winds and energy loss from dissipation, e.g. by wave breaking and the generation of unorganized motion—turbulence. When such a balance kicks in we have what's called a *fully-developed sea.*

Since the winds in reality are not steady either but consist of velocity fluctuations of different time and spatial scales, the ocean waves also consists of a range of different frequencies and wavelengths (or wavenumbers). So we get a broad spectrum of ocean waves. The spectrum is a directional spectrum since waves can travel in all the compass directions (Figure 83). Non-linear terms in the wave equations also cause a transfer of energy between waves of different frequencies and wavenumbers, and the end result is that wave spectra, at least for fully-developed seas, tend to have characteristic shapes. Generally the non-linear terms tend to transfer energy toward lower frequencies, so the stronger the winds the bigger (and more nonlinear) the waves become...and the larger is the shift towards lower frequencies. Finally, the horizontal length, the *fetch,* over which the winds blow also matters. Essentially, the longer the fetch the longer time the waves get to develop (Figure 84).

Propagation The waves, being waves (!), travel at speeds set by their dispersion relationship. In deep waters the short waves travel relatively slowly and are constantly forced by the winds. This is the *wind se*a, and it is choppy because of the turbulence in the winds. But longer waves travel faster and can actually outpace the winds (especially towards the outer reaches of a storm). These long waves that travel faster than the winds are the majestic *swell* (Figure 85). They, in fact, are often the messengers telling the tale that a storm has taken place somewhere far away. So an observer (a surfcer?) waiting on the beach, perhaps hundreds or even thousands of kilometers away from a storm, will observe the long waves, the swell, first. As it turns out, he or she may not even get a chance to observe any shorter waves generated by the far-away storm. Not only because it takes them much longer to cross the ocean but also because they dissipate faster (the viscous stress terms are scale-selective).

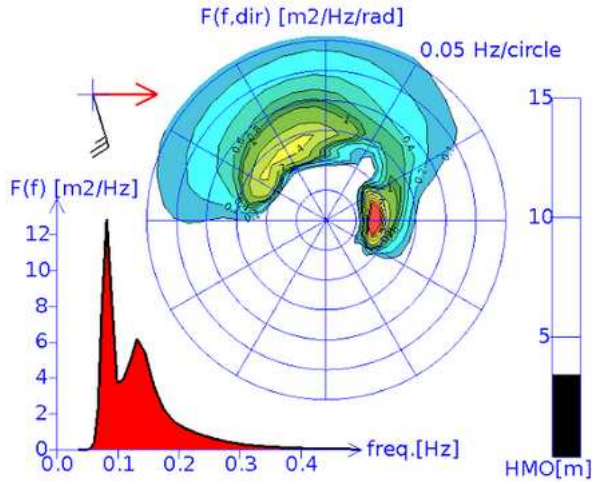

Figure 83: A directional wave spectrum from a location in the Norwegian Sea, estimated by a numerical wave model. Black arrow shows the local winds at the time of estimation. The spectrum shows two 'blobs' of waves. One is local wind sea, forced by the wind, while another is a longer-wavelength swell propagating in from some other region. (Source: Norwegian Meteorological Institute)

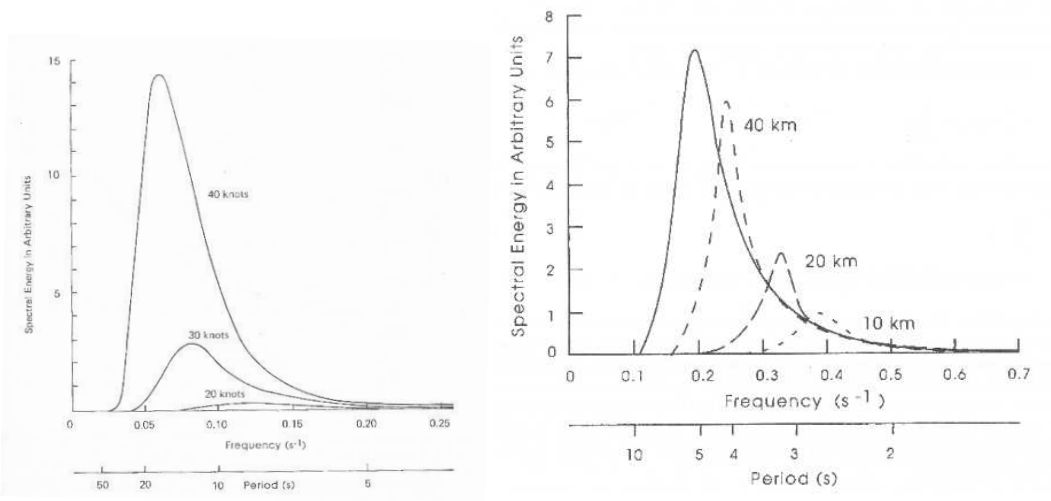

Figure 84: Idealized wave spectra (integrated around all directions): (left) for fully-developed seas for winds 20, 30 and 40 knots (about 10, 15 and 20 m/s), and (right) for about 15 knots but for various lengths of fetch. (Source: Knauss, 2005, Fig. 9.11)

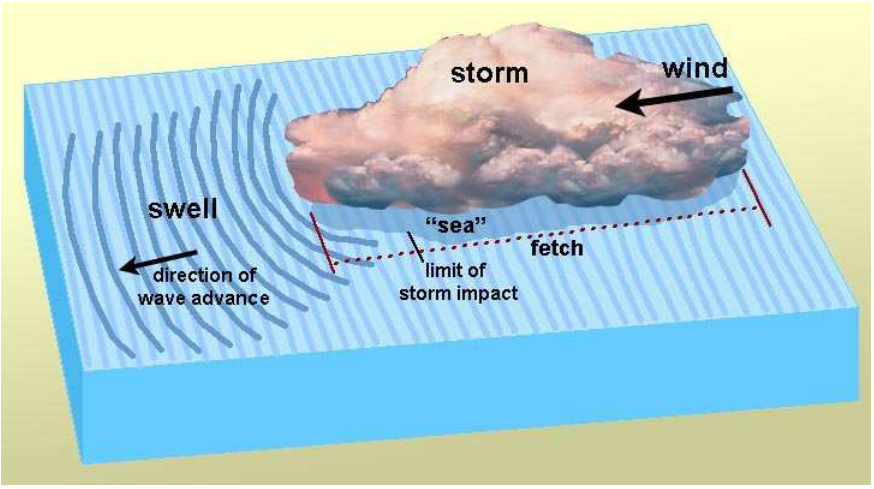

Figure 85: The swell eminating from a storm. (Source: http://geologycafe.com/oceans/chapter10.html)

A spectacular example of the journey of the swell is from the Pacific ocean, from the permanent stormy region surrounding Antarctica and all the way to the western coast of north and central America. In the event shown in Figure 86 the swell traveled around 12000–25000 km in 10 to 13 days with an average speed of 26 m/s.

As the waves hit the coast As the swell approaches land they go from being deep-water waves to being shallow-water waves, with a new dispersion relationship, i.e. not being dispersive at all. If such shallow-water waves approach the coast at some oblique angle, the part of the wave crest closest to the coast will experience a lower phase speed than the same crest further off the coast. This is called *wave refraction* and causes the wave to turn, as shown in Figure 87. The wave thus tends to hit the coast with the crests parallel to the beach. The wave energy travels perpendicular to the wave crests, so this implies that wave energy enters almost normal to the beach. Where the coast is uneven, as in the figure, the wave energy flux is also even, and this tends to cause an uneven rate of errosian along the coast—actually tending to flatten the coastline.

As the waves continue to progress into shallower waters the phase and group speed continually decrease (remember, they are now shallow-water waves). So it's as if the waves are breaking towards the beach, and this causes the waves to both get shorter wavelength and higher amplitudes. The wave height grows!

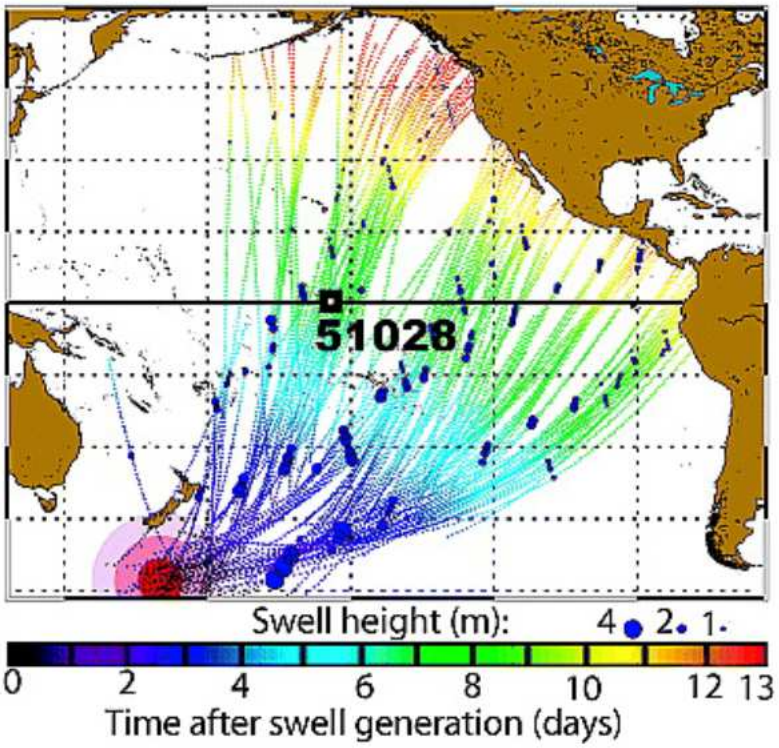

Figure 86: Estimates of paths taken by swell generated by a storm that took place in the Southern Ocean, south of New Zealand. The color on the lines show the 'age' of the swell along the journey.

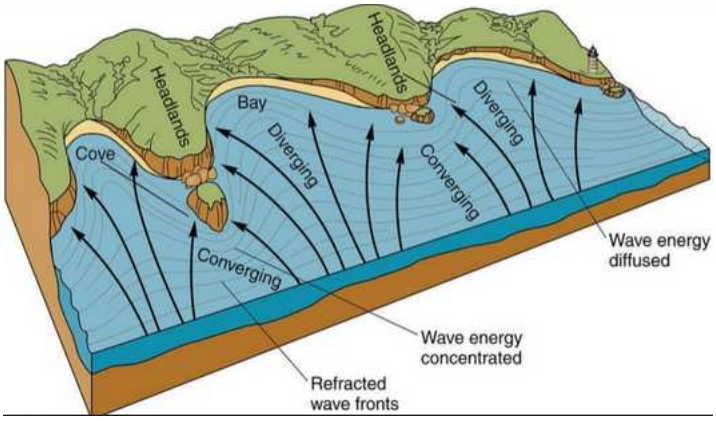

Figure 87: Refraction of shallow-water waves as they approach the coast and the uneven wave energy propagation in towards an uneven coastline.

When the wave height starts to become comparable to the wavelength the waves are no longer linear and don't looko like sinusoids anymore. Finally, when bottom friction starts to kick in, the wave water parcels near the bottom are slowed down even more. So water in the upper part of the wave travels faster than water in the lower part and the wave 'tips over' and breaks (Figure 88).

9.2.2 Tsunamis

Tsunamis are generated by off shore seismic activity. There can be multiple causes, e.g. an underwater earthquake, but all are associated with an abrupt vertical movement of the ocean bottom which displaces water vertically over a limited region. This displaced water then travels away from the generation region (in all directions). Tsunamis have wavelengths of tens to hundreds of kilometers, so they are always shallow-water waves, moving at speed $c = \sqrt{gH}$. So once detected, the travel times to the surrounding coasts is relatively easy to calculate. But for ocean depths of, say 3000 meters, the wave speed will be around 175 m/s or 625 km/hr, so large ocean basins can be crossed in only a few hours.

In the open ocean the sea surface amplitude of tsunamis may be less than a meter, too small to be felt by crew or passengers on ships. But as they approach land the wave height can increase dramatically. The key reason is that the wave speed decreases in shallower water. And since the wave period (or frequency) does not change, the wavelength then has to decrease. So water is piling up from behind the wave and the only way to conserve volume is for the wave height to

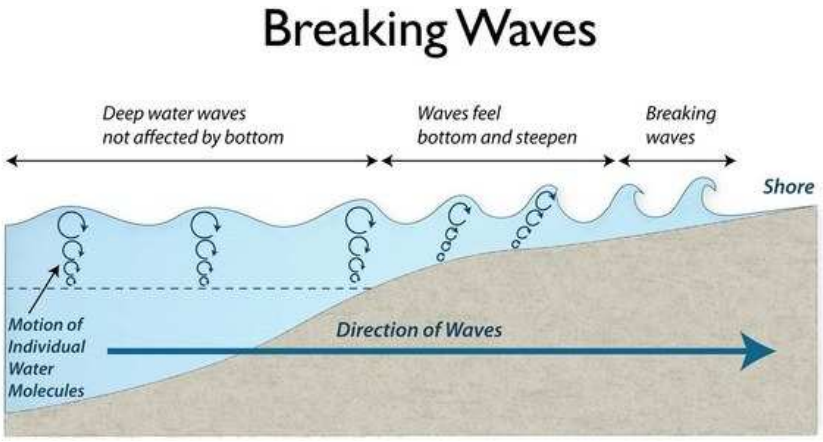

Figure 88: The behavior of the swell as it approaches the beach, starting as linear deep-water waves far off-shore, then transitioning to shallow-water waves as they enter, and finally steepening and breaking on the shores.

increase. Finally, when the wave height becomes comparable to or smaller than the water depth the wave becomes unstable and breaks.

9.3 Ocean waves impaced by Earth's rotation

When the wave frequency becomes comparable to the local Coriolis parameter we can no longer ignore the impact of Earth's rotation on the dynamics of the waves. These 'low-frequency' waves also typically have large wavelengths and are hence deep-water waves.

At this point it is rather useful to realize that the shallow-water limit, $kH \sim H/\lambda \ll 1$ is essentially another way to say that the aspect ratio which we've seen before, $\delta = H/L$, is small. This was one of the key requirements behind the hydrostatic approximation. So shallow-water waves are governed by hydrostatic dynamics. Hence, low-frequency shallow-water waves in a constant-density ocean are governed by the *shallow-water equations* that we have seen before.

If we ignore all nonlinear terms in the shallow-water equations (remember, we are after linear waves here), and we also consider only the flat-bottom case, we get

$$\begin{aligned}
\frac{\partial u}{\partial t} - fv &= -g\frac{\partial \eta}{\partial x} \\
\frac{\partial v}{\partial t} + fu &= -g\frac{\partial \eta}{\partial y} \\
\frac{\partial \eta}{\partial t} &= -D\left(\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y}\right),
\end{aligned}$$

where D is the mean water depth (ignoring the contribution from η. These equations have wave solutions that depend on boundary conditions and, as we will see later, on what time scales (or frequencies) we consider.

In studying these waves we will principally be looking at their dispersion relation, the functional relationship between frequency and wavenumber. And to do this we *try* a wave solution for the unknown variables (knowing that it'll work) and see what happens. So we will assume sine and cosine solution and insert these into the equations. Actually, in most cases we will use Euler's identity

$$e^{i\Theta} = \cos\Theta + i\sin\Theta$$

and just say that the wave is the real part of the exponential. So if we assume that, say, the sea surface height is a cosine, we write

$$\eta(x, y, t) = Re\left\{e^{i(kx+ly-\omega t)}\right\}.$$

9.3.1 PoincarÃ© and Kelvin waves

PoincarÃ© waves The most general wave solution to the linear shallow equations can be found by taking the shallow water equations at face value and inserting wave forms for the three unknowns, u, v and η. But here we will instead reorganize the equations into one single (third-order) equation for η. The procedure involves taking both the divergence and the curl of the two momentum equations, and it eventually allows us to eliminiate u and v. The final expression for η is

$$\frac{\partial}{\partial t}\left\{\left[\frac{\partial^2}{\partial t^2} + f^2 - gD\left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}\right)\right]\eta\right\} = 0,$$

which, upon inserting our exponential wave solution for η, gives

$$-i\omega\left[-\omega^2 + f^2 + gD\left(k^2 + l^2\right)\right] = 0.$$

So we can have two solutions, either

$$\omega = 0$$

or

$$\omega^2 = f^2 + gD\left(k^2 + l^2\right).$$

The first solution is a steady solution ($\omega = 0$ implies $\partial/\partial t = 0$, so what we get is geostrophy!) which we will not pursue here. But the second solution is a proper dispersion relation, stating what the frequency is for a given set of wave numbers (and Coriolis parameter and water depth).

Consider a plane wave with the wave vector aligned in the x-direction, so that $l = 0$ and

$$\eta(x,t) = A\cos\left(kx - \omega t\right).$$

The dispersion relation becomes

$$\omega^2 = f^2 + gDk^2,$$

or

$$\omega = \pm\sqrt{f^2 + gDk^2}.$$

Dividing by f gives

$$\frac{\omega}{f} = \pm\sqrt{1 + L_d k},$$

where

$$L_d = \frac{\sqrt{gD}}{f}$$

is what's called the *Rossby radius of deformation*. The importance of this length scale in the dynamics of the ocean (and atmosphere) cannot be overemphasized. Its meaning is illustrated in Figure 89 which shows the dispersion relationship (here only for posivite frequencies). For high wavenumbers, meaning wavelengths much smaller than L_d, the PoincarÃ© waves behave just like the high-frequency surface gravity waves studied above. The frequencies are so high and time scales so short that Earth's rotation doesn't come into play. But at lower wavenumbers, for wavelengths comparable to the deformation radius, Earth's rotation becomes important and the dispersion relation bends away from that of non-rotating waves. The waves actually attain a minimum freqency of f as the wavelength goes to infinity. So the Rossby radius is *the lateral scale at which Earth's rotation becomes important*.

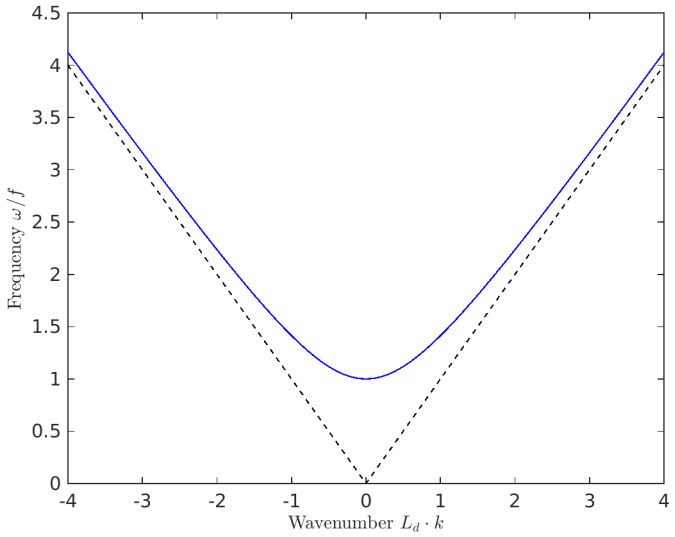

Figure 89: The dispersion relation for a PoincarÃ© wave traveling in the x-direction. Also shown (dashed lines) is the dispersion relation for shallow-water gravity waves on a non-rotating planet. The frequency and wavenumber have been nondimensionalized by the Coriolis parameter f and by the Rossby deformation radius L_d, respectively.

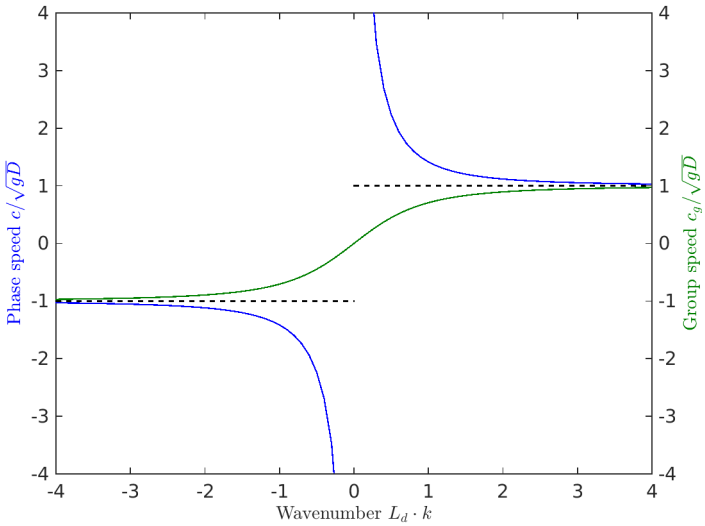

Figure 90: Phase (blue) and group (green) speeds as function of wavenumber for plane PoincarÃ© waves with positive frequency travelling in the x-direction. Both speeds assume positive frequencies (the positive root) and are also normalized by the non-rotating speed $\sqrt{gD}$, and the wavenumber has been normalized by deformation radius L_d.

The phase and group speeds, both divided by the non-rotating gravity wave speed $\sqrt{gD}$, are shown in Figure 90. We see, as expected, that both phase and group speeds take on the non-rotating speed for large wavenumbers, i.e. for small scales. But the two behave differently at large scales: the phase speed becomes infinite while the group speeds goes to zero.

What is the particle motion in such a wave? If we again consider the plane wave that travels in the x-direction, its crests and troughs are aligned in the y-direction. So there are no variations in the y-direction, and the equations become

$$\begin{aligned} \frac{\partial u}{\partial t} - fv &= -g\frac{\partial \eta}{\partial x} \\ \frac{\partial v}{\partial t} + fu &= 0 \\ \frac{\partial \eta}{\partial t} &= -D\frac{\partial u}{\partial x}. \end{aligned}$$

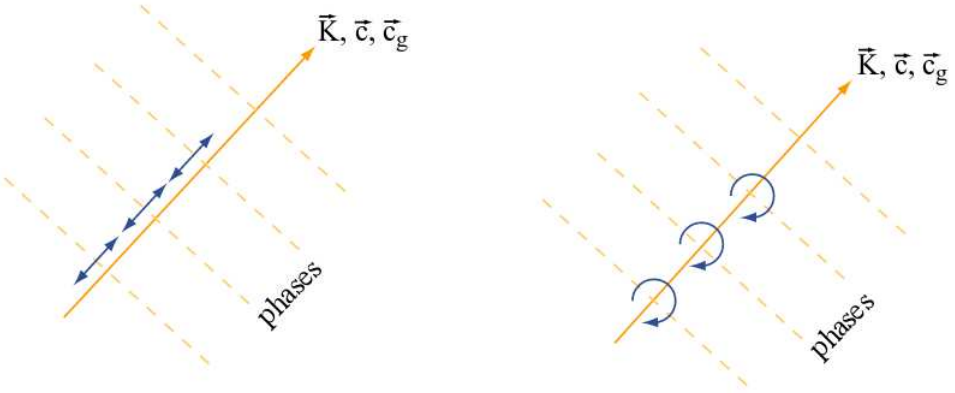

Figure 91: The fluid motion in plane (left) non-rotating and (right) rotating (PoincarÃ©) waves propagating in the direction of wavevector $\boldsymbol{K}$.

Hence, for our cosine sea-surface height field

$$\eta(x,t) = A\cos\left(kx - \omega t\right),$$

we can get an expression for the u-velocity field from the third equation. This becomes

$$u(x,t) = \frac{\omega}{kD} A\cos\left(kx - \omega t\right).$$

And the v-velocity field can be found from the second equation, giving

$$v(x,t) = \frac{f}{kD} A\sin\left(kx - \omega t\right).$$

These solutions are shown in Figure 91, now for a wave propagating in an arbitraty direction given by wave vector $\boldsymbol{K} = k\boldsymbol{i} + l\boldsymbol{j}$. For positive frequencies the fluid parcels move in clockwise ellipses in the horizontal plane that have their semi-major axis in the direction of phase propagation. For comparison, the figure also shows the water parcel motion in shallow-water gravity waves not influenced by Earth's rotation. For those, the water motion is strictly back and forth along the direction of wave propagation.

Kelvin waves Now consider the situation near a continental boundary, say a southern boundary at $y = 0$. We anticipate that a wave could be propagating along the boundary, so along the x-direction. But how does the boundary condition requiring no normal flow at $y = 0$ impact the solution? William Thompson (better

known as Lord Kelvin, 1824–1907) investigated a solution which assumed $v = 0$ everywhere and not just at the boundary. The governing equations then become

$$\begin{aligned} \frac{\partial u}{\partial t} &= -g\frac{\partial \eta}{\partial x} \\ fu &= -g\frac{\partial \eta}{\partial y} \\ \frac{\partial \eta}{\partial t} &= -D\frac{\partial u}{\partial x}. \end{aligned}$$

Taking $\partial/\partial t$ of the third equation and subtracting $D \cdot \partial/\partial x$ of the first equation gives

$$\frac{\partial^2 \eta}{\partial t^2} = gD\frac{\partial^2 \eta}{\partial x^2}$$

which is a classical wave equation. Plane wave solutions look like

$$\eta(x, y, t) = \tilde{\eta}_1(y)e^{i(kx-\omega t)} + \tilde{\eta}_2(y)e^{i(kx+\omega t)}$$

or

$$\eta(x, y, t) = \tilde{\eta}_1(y)e^{ik(x-ct)} + \tilde{\eta}_2(y)e^{ik(x+ct)},$$

with $c = \sqrt{gD}$, as before, and with the north-south structure of the pressure field contained in $\tilde{\eta}_1$ and $\tilde{\eta}_2$.

So it appears that we can have waves travelling both eastward and westward. But this is only until we also apply the remaining equation which says that the zonal velocity is in geostrophic balance with the meridional pressure gradient (even if it is also associated with wave motion). Using this equation it can be shown that the north-south structure of the wave is an exponential. But the exponential representing $\tilde{\eta}_2$ grows out of bounds away from $y = 0$, so this solution is physically impossible. The only possibility is an exponential representing $\tilde{\eta}_1$ which decays towards the north, and the final solution is

$$\eta(x, y, t) = \tilde{\eta}_{1,0}e^{-y/R_d}e^{ik(x-ct)},$$

or, taking only the real part in the end,

$$\eta(x, y, t) = \tilde{\eta}_{1,0}e^{-y/R_d}\cos\left[k\left(x - ct\right)\right],$$

where, as before, $R_d = \sqrt{gD}/f = c/f$ is the Rossby radius of deformation. So these waves, called *Kelvin waves*, propagate eastward along a southern ocean

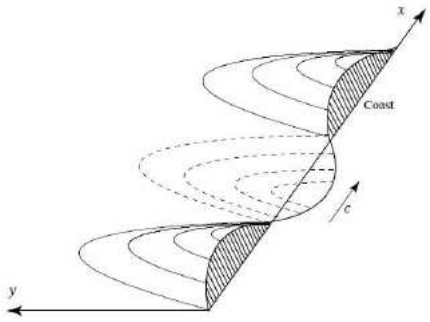

Figure 92: The eastward propagation of a kelvin wave along the southern boundary of an ocean basin in the northern hemisphere. Countours show the sea surface height.

boundary with phase and group speeds $c = c_g = \sqrt{gD}$. Their amplitude decays exponentially away from to the north with an e-folding scale set by the Rossby radius. The more general result, allowing for a coastal boundary of any orientation, is that such Kelvin waves travel with the coast to their right in the northern hemisphere (and to their left in the southern hemisphere). And even though rotation is important (the Coriolis parameter enters into the spatial decay scale), they travel with the non-rotating shallow-water gravity wave speed and are non-dispersive.

9.3.2 Tides

The tides are waves propagating through the oceans, waves forced by the gravitational forces of the Sun and the Moon. Before looking at the wave behavior of the tides, we'll have a quick look at this gravitational forcing and the response we would have seen if waves could travel infinitely fast.

Equilibrium tides The gravitational force by the moon on a water parcel is

$$F = G\frac{mM}{P^2},$$

where M and m are the mass of the moon and the water parcel, respectively, P is the distance between the water parcel and the center of mass of the moon and, finally, G is the gravitational constant. The force is a rather strong function of the distance, and it is quite clear that a water parcel on the side of Earth facing the moon feels a stronger force than another parcel on the other side of the Earth. The

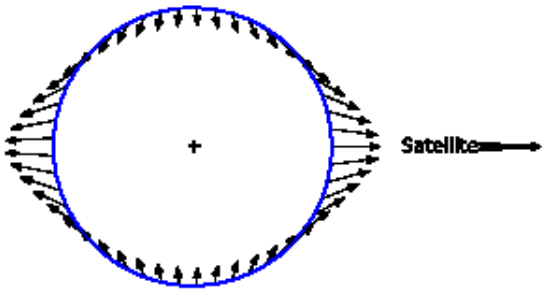

Figure 93: The net tidal forces on water parcels residing on the surface of the Earth. The moon ("satellite") is located to the right of the picture. (Source: Wikipedia)

tide generating force is this 'pull' on water parcels by the moon *relative to the pull felt at the center of Earth itself*. This 'residual' force is shown in Figure 93.

The *equilibrium response* to these forces, the steady-state response if the water experiences no inertial (i.e. it has no mass) and no friction, is two bulges of water, one on the side of Earth facing the moon and one on the other side. There are also two depressions in sea water, as shown in Figure 94. As Earth rotates around its own axis, once every 24 hours, an observer anywhere of the planet would observe the equilibrium tide passing. One would expect two bulges to pass per day, a *semi-diurnal* tide. And this is indeed the case, most of the time. But since the moon's orbit around Earth is not in the equatorial plane but rather at an angle with respect to it, a *diurnal* equilibrium tide (once per day) can also be found at high latitudes when the moon happens to be far off Earth's equatorial plane (as shown in the second panel of the figure).

Both the moon and the sun exert tidal forces on the Earth's ocean and atmosphere (yes, there are atmospheric tides!). But since the moon and sun have moved by different amounts each time Earth has rotated around its own axis, the periods of the lunar and solar equilibrium tides are slightly different. So we have *two principal semi-diurnal tidal constituents*, M2 and S2 (having periods 12.42 hours and 12 hours, respectively). Then there are diurnal constituents associated with both, the strongest one being the K1 lunar constituents (period 23.93 hours). In addition, the interaction or interference between these frequency components create new tidal frequencies that are either sums of the original frequencies or differences between them (remember our discussion earlier on the adding up of two waves of slightly different frequencies). The result is a *huge* range of equilibrium tidal frequencies, including high frequencies (up to six oscillations per day) and

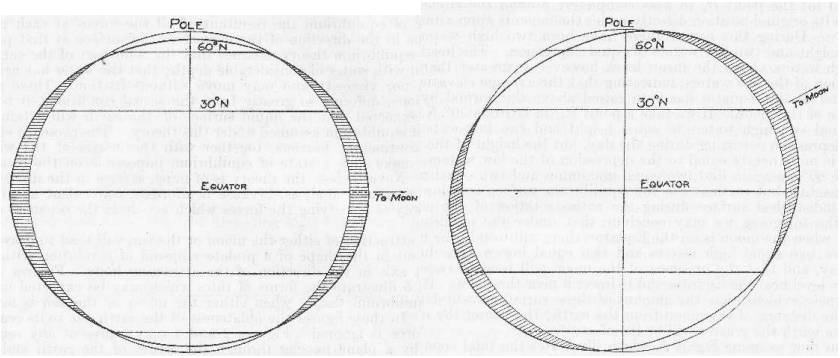

Figure 94: The equilibrium response due to the gravitational attraction of the moon, for two different positions of the moon with respect to Earth's equatorial plane. (Source: U.S. Dept. of Commerce, 1924)

very low frequencies (years!).

The most known 'interaction' tide is perhaps the spring-neap cycle which occurs due to Moon's orbiting around Earth. When the sun and the moon are aligned in a line, we have extra strong tidal bulges along that line and extra strong depressions at right angles—*spring tides* (Fig. 95). And when the Earth–moon–sun system is oriented at right angles, the tide generating forces of the sun and the moon oppose eachother, giving rise to weak bulges and depressions—*neap tides*. The spring-neap cycles has a period of 28 days.

Dynamic tides As it turns out, the equilibrium tide is only a first qualitative approximation of the real tides. In the real ocean, the 'tidal bulges' have to move around as waves, constantly trying to keep up with the astronomical (gravitational) forcing. The wavelengths are huge (some sizable fraction of the circumference of the planet), so these are shallow-water waves that are influenced by the rotation of Earth around its own axis and also by the presence of continents. So, from what we have learned above, tides can propagate as both PoincarÃ© or Kelvin waves.

But remember that PoincarÃ© waves have to have frequencies higher than the Coriolis frequency. And since the Coriolis freqency increases with latitude, there exist latitudes above which certain tides cannot propagate as PoincarÃ© waves. This is called the *critical latitude*. For the diurnal tide this latitude is about 30 degrees whereas for the semidiurnal tide it is around 75 degrees. Tides *can* propagate beyond these latitudes, but that will have to take place in the form of Kelvin waves, hugging some coast. In fact, because of the many coastal boundaries tides

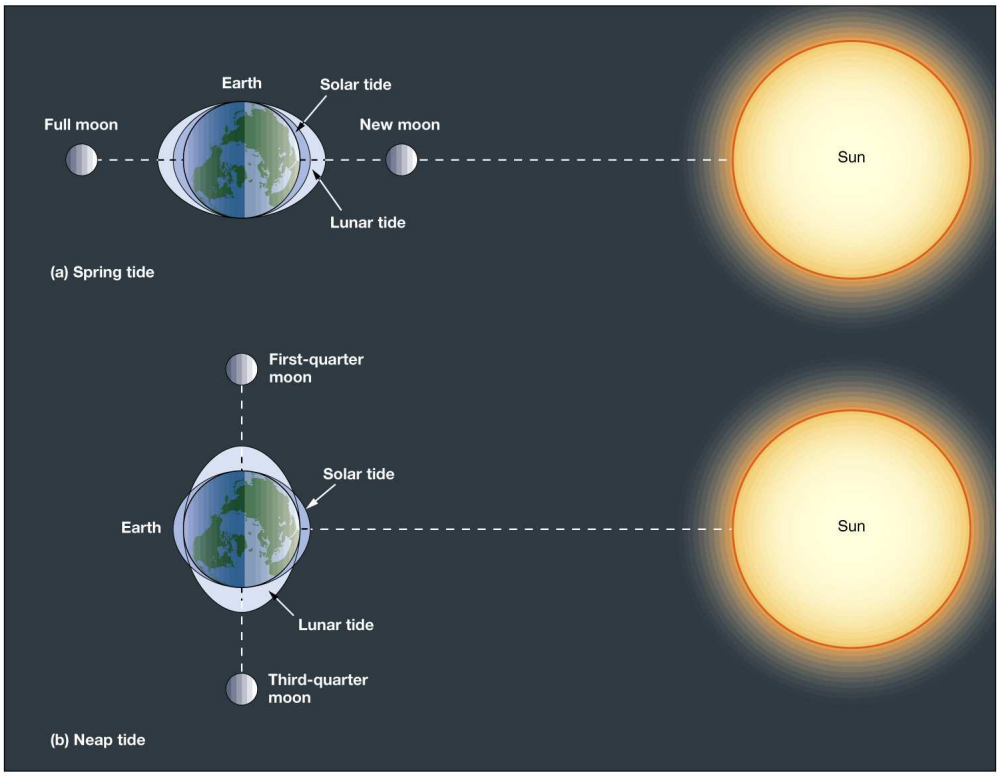

Figure 95: The spring-neap tides. (Source: Wikipedia)

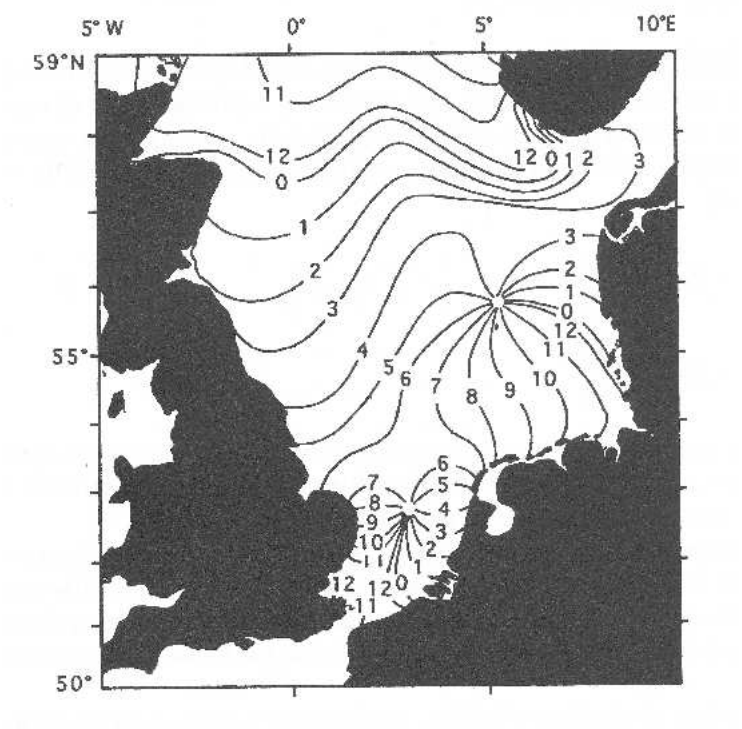

Figure 96: The propagation of the semidiurnal lunar tide (M2) through the North Sea and British Channel. The contours show the time of high tide in hours, in reference to Greenwich. (Source: Knauss, 2005, Fig. 10.5)

mostly behave like Kelvin waves everywhere, constantly trying to keep up with the forcing by the moon and the sun. So the phase speed of the tide is typically $c = \sqrt{gD}$. But now, in reality, this propagation speed varies from place to place—by the square root of the ocean depth.

An example of the propagation of the semidiurnal lunar tide through the North Sea and British Channel is shown in Figure 96. The plot shows the phase of the tide, in hours, relative to some reference time (here the time in Greenwich, England). We see that the tide propagates counter-clockwise in these seas, with the coast to its right as expected from Kelvin waves. The tide here is actually a combination of a wave entering from the north and one from the south. Shown in the figure are two points where the phase lines meet. At these locations, *amphidromic points*, the tidal amplitude is zero since this is the only way the wave can have all possible phases at the same time.

9.4 Very low frequency (Rossby) waves

We have seen that Poincar´e waves only have frequencies higher than the Coriolis parameter. We call them *super-inertial* waves. Kelvin waves can have lower frequencies, at wavelengths longer than the Rossby radius, but these can only travel along a coast. So does very low frequency waves not exist in the ocean interior, away from boundaries? Actually they do, but they require that we add one more aspect of the dynamics on a rotating planet, namely *the meridional variation of the Coriolis parameter*.

Let's start with the linear shallow-water equations, again for a flat bottom but now allowing for a weak change in the Coriolis parameter. We write

$$f = f_0 + \beta y$$

where f_0 is the Coriolis parameter at the mean latitude in the region we're interested in and $\beta = df/dy$, i.e. the meridional gradient of f. Then the equations are

$$\begin{aligned}
\frac{\partial u}{\partial t} - (f_0 + \beta y)\, v &= -g\frac{\partial \eta}{\partial x} \\
\frac{\partial v}{\partial t} + (f_0 + \beta y)\, u &= -g\frac{\partial \eta}{\partial y} \\
\frac{\partial \eta}{\partial t} &= -D\left(\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y}\right),
\end{aligned}$$

where we assume $\beta y \ll f_0$. We want to study the behavior of these equations at very long time scales, i.e. for $T \gg 1/f$, such that the temporal Rossby number $\varepsilon_T \ll 1$. When the Rossby number is small we expect the motion to be primarily geostrophic.

But now we allow for a weak *ageostrophic* component to the flow, so we write

$$\begin{aligned}
u &= u_g + u_a \\
v &= v_g + v_a
\end{aligned}$$

where $u_a \ll u_g$ and $v_a \ll v_g$. If we now plug these velocities into the momentum equations, the balance between the largest terms become,

$$\begin{aligned}
-f_0 v_g &= -g\frac{\partial \eta}{\partial x} \\
f_0 u_g &= -g\frac{\partial \eta}{\partial y}.
\end{aligned}$$

So the flow is nearly geostrophic. When we now look into the balance for the not-so-big terms, we *substitute in* the expressions for the geostrophic velocities. This gives

$$\begin{aligned}
\frac{\partial}{\partial t}\left(-\frac{g}{f_0}\frac{\partial \eta}{\partial y}\right) - f_0 v_a - \beta y\left(\frac{g}{f_0}\frac{\partial \eta}{\partial x}\right) &= 0 \\
\frac{\partial}{\partial t}\left(\frac{g}{f_0}\frac{\partial \eta}{\partial x}\right) + f_0 u_a + \beta y\left(-\frac{g}{f_0}\frac{\partial \eta}{\partial y}\right) &= 0 \\
\frac{\partial \eta}{\partial t} &= -D\left(\frac{\partial u_a}{\partial x} + \frac{\partial v_a}{\partial y}\right),
\end{aligned}$$

where we have ignored even smaller terms, like $\partial u_a \partial t$ and $\beta y v_a$ since these are even smaller. Note that the divergence of the geostrophic flow (the one involving f_0) is zero and therefore drops out of the continuity equation. So the third equation shows that the sea surface goes up and down due to divergence in the ageostrophic flow field. And we can find that from the two momentum equations. Rearranging these gives

$$\begin{aligned}
u_a &= -\frac{g}{f_0^2}\frac{\partial^2 \eta}{\partial x \partial t} + \beta y \frac{g}{f_0^2}\frac{\partial \eta}{\partial y} \\
v_a &= -\frac{g}{f_0^2}\frac{\partial^2 \eta}{\partial y \partial t} - \beta y \frac{g}{f_0^2}\frac{\partial \eta}{\partial x},
\end{aligned}$$

so that the divergence becomes

$$\frac{\partial u_a}{\partial x} + \frac{\partial v_a}{\partial y} = -\frac{g}{f_0^2}\frac{\partial}{\partial t}\left(\frac{\partial^2 \eta}{\partial x^2} + \frac{\partial^2 \eta}{\partial y^2}\right) - \beta \frac{g}{f_0^2}\frac{\partial \eta}{\partial x}.$$

Plugging this into the continuity equation then gives

$$\frac{\partial \eta}{\partial t} = \frac{gD}{f_0^2}\left[\frac{\partial}{\partial t}\left(\frac{\partial^2 \eta}{\partial x^2} + \frac{\partial^2 \eta}{\partial y^2}\right) + \beta \frac{\partial \eta}{\partial x}\right],$$

or

$$\frac{\partial}{\partial t}\left[\frac{f_0^2}{gD}\eta - \left(\frac{\partial^2 \eta}{\partial x^2} + \frac{\partial^2 \eta}{\partial y^2}\right)\right] - \beta \frac{\partial \eta}{\partial x} = 0.$$

Is this an equation describing linear waves? To find out, let's assume a wave solution

$$\eta(x, y, t) = \tilde{\eta} e^{i(kx + ly - \omega t)}$$

and plug in. This gives

$$-i\omega\left[\frac{f_0^2}{gD}+\left(k^2+l^2\right)\right]-ik\beta=0$$

or

$$\omega=-\frac{\beta k}{(k^2+l^2)+1/L_d^2},$$

where, as before, $L_d=\sqrt{gD}/f_0$ is the Rossby radius of deformation. This does indeed look like a dispersion relation.

What about phase speeds? They are

$$\begin{aligned} c_x &= \frac{\omega}{k}=-\frac{\beta}{(k^2+l^2)+1/L_d^2} \\ c_y &= \frac{\omega}{l}=-\frac{\beta k/l}{(k^2+l^2)+1/L_d^2}. \end{aligned}$$

So the meridional phase speed can be both northward and southward, but the zonal phase speed is always negative! These waves can only travel westward, as can be observed e.g. in satellite observations of sea surface height (Figure 97).

Rossby waves are extremely important for the dynamics of the oceans *and* the atmosphere. Did you know, for example, that the "synoptic" high and low pressure systems that give us weather variations on daily to weekly time scales are atmospheric Rossby waves? They all move intrincically westward, and if they are actually observed to move eastward it is only because they are advected by eastward winds that are stronger than the westward phase velocity of the waves themselves.

The mechanism responsible for Rossby waves: Rossby waves are strange waves indeed. And their restoring mechanism is not the gravitational force, as it is for surface gravity waves. To understand the Rossby wave we instead need to look to the potential vorticity equation and the principle of concervation of PV, stated as

$$\frac{D}{Dt}\left(\frac{f+\zeta}{H}\right)=0.$$

Let's ignore variations in bottom depth, as we have done in all of the wave discussion above, and consider what happens when a water parcel is displaced northward from its resting latitude. The situation is illustrated in Figure 98. So water parcel

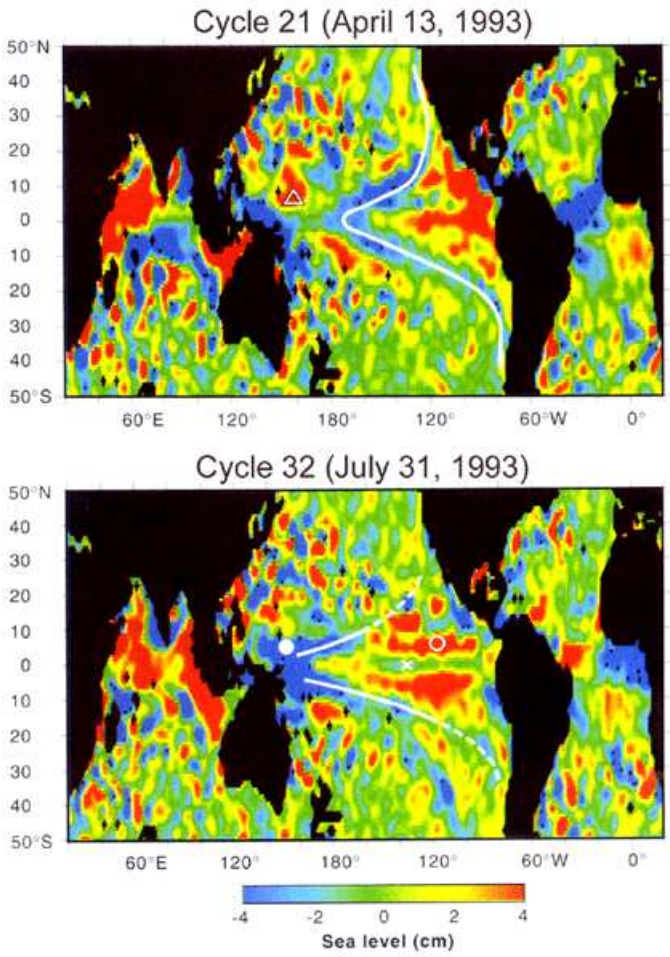

Figure 97: The sea surface height, three and a half months apart, as observed by satellites. Rossby wave patterns are seen to move westward. (See also https://www.youtube.com/watch?v=F8zYKb2GoR4)

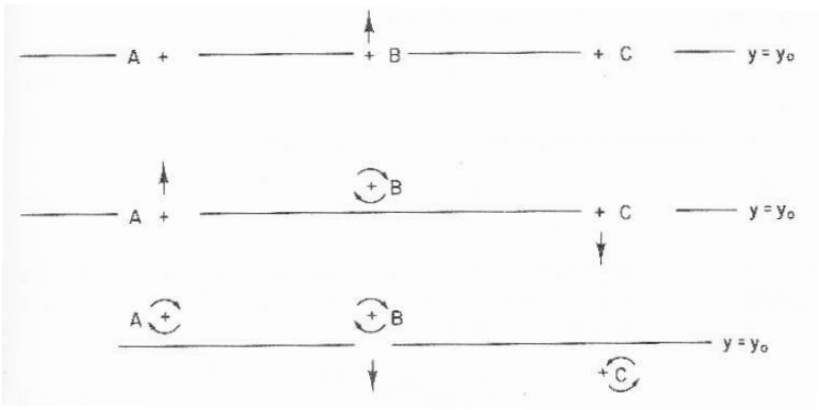

Figure 98: The mechanism for a Rossby wave. Shown is the motion of three water parcels, each of which conserve their potential vorticity. The three parcels originally reside at 'latitude' y_0. But a perturbation, starting when water percel B is displaced northward, ends up propagating westward. (Source: Pedlosky, 1987, Fig. 3.16.1)

B (in the figure) is displaced northward where it attains a larger planetary vorticity (larger Coriolis parameter). To conserve its total PV it has to reduce its relative vorticity, i.e. it has to start spinning clockwise (anticyclonically). The velocity field from this spin then moves neighbouring water columns, so a column A to the west of B is moved northward while a column C to the east of B is moved southward. When column A is displaced northward it too starts to spin clockwise, and the process is repeated. The end result is that the perturbation which started with column B moves progressively westward—with the phase speed we deduced above.

A Appendix: The flow in estuaries

Now that we are equipped with all the basic conservation equations and also with practical constructs like turbulent diffusivities, we will look into some different types of ocean flows, both relatively small-scale (where the effect of Earth's rotation is only slightly important) and really large-scale where the ambient rotation affects everything. We start with the flows very close to land, in what we all *estuaries*. A typical definition of an estuary is

> "a semienclosed body of water having a free connection with the open sea and within which the seawater is measurably diluted with freshwater deriving from land drainage."

So a Norwegian fjord (or Chilean, Canadian or New Zeeland fjord) is definitely an estuary, but there are also other kinds, as will become apparent below. Understanding the flow and hydrography in estuaries is important since estuaries are the homes of diverse ecosystems as well as a considerable fraction of Earth's human population. So an understanding of the processes impacting the transport and dispersion of pollutants in estuareis is of real practical utility. As we will see here, for example, many silled fjords have problems with water quality at depth.

A.1 Estuarine circulation

Estuaries typically link one or more rivers to the open ocean. One could imagine that the river water flows "down the hill" and out of the estuary on top of a stagnant layer of saltier and denser ocean waters. But this is hardly ever observed. Instead, estuaries are almost always associated with a net inward flow at depth and a net outward flow near the surface—opposing flows whose individual volume transports are much larger than the freshwater transport in the river itself.

Let's set up steady-state (meaning here $\partial/\partial t = 0$) budgets of volume and salt for an estuary consisting of a river flowing into a layer of *brackish water* (less salty) lying on top of another layer of saltier ocean water (Figure 99). The steady-state volume budget for the entire estuary is

$$T_o = R + T_i,$$

where V is the total volume, R is the volume transport by the river, T_i is the volume transport of salty water into the estuary at its outer boundary and T_o is the volume transport of brackish water out of the estuary. If turbulent (diffusive) lateral transports are ignored, the time-mean salt budget is

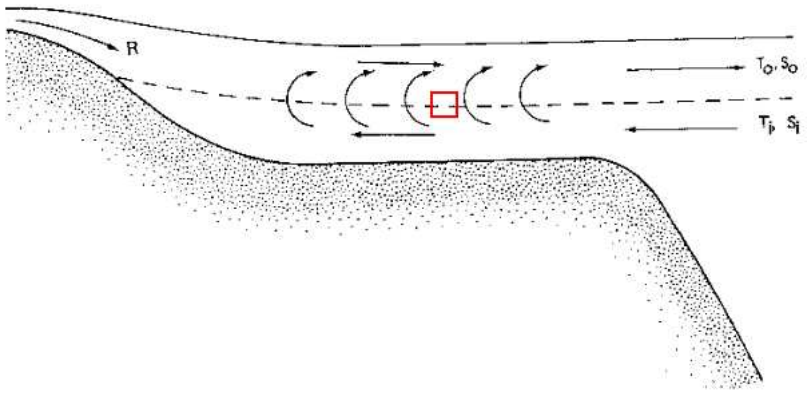

Figure 99: Two-layer schematic of the estuarine circulaton. The net flow out of the estuary at the surface, T_o, is the sum of the river flow, R, and a net flow in at depth, T_i. The red box indicates a small control volume around the interface where we study vertical salt fluxes. (Source: Knauss, 2005, Fig. 11.2)

$$T_o S_o = T_i S_i,$$

where S_i and S_o are the salinities of the in and outflow, respectively (we assume that the river water has a salinity of zero). Assuming that we know the river volume transport and the two salinities from measurements, we can solve for the in and outflows at the boundary to the open ocean:

$$\begin{aligned} T_i &= R\frac{S_o}{S_i - S_o} \\ T_o &= R\frac{S_i}{S_i - S_o}. \end{aligned}$$

For $S_i > S_o$ (the inflowing layer being saltier than the outflowing layer) we have that both T_i and T_o are larger than R and also that $T_o = R + T_i > T_i$. So the salty water which flows into the estuary at depth must upwell or be *entrained* into the upper outflowing layer. The equations show that the weaker the vertical salinity stratifiction $(S_i - S_o)$, the stronger is the flow in the lower and upper layers and, hence, the entrainment.

This entrainment of lower-layer water into the outflowing top layer is driven by turbulent salt exchanges between the layers. To see this let's set up the steady-state salt salt budget for a small control volume residing on the interface between

the two layers (the red box in the figure). If we ignore lateral transports here, we get a balance between vertical advection and vertical turbulent transport, i.e.

$$\frac{\partial S}{\partial t} = -w\frac{\partial S}{\partial z} - \frac{\partial}{\partial z}\left(-K_v\frac{\partial S}{\partial z}\right)$$

(where we have put the advection term also on the right-hand side). For $w > 0$, as our setup requires (upwelling of lower-layer water), the vertical advection term would tend to increase the salinity in the box. In other words, there is a convergence of salinity in the box by vertical advection. For this to be balanced to maintain the steady state, we require the vertical turbulent salinity transport to be divergent, i.e. to lower the salinity in the box. Alternatively, one could say that the vertical advection brings salt water up while the vertical turbulent diffusion brings freshwater down and that in steady state the two transports balance. It is easy to see from this expression that larger turbulence levels, i.e. higher values of K_v, can lead to larger w, hence larger vertical transports and larger horizontal layer transports T_i and T_o.

A.2 Types of estuaries

Estuaries are typically classified by their vertical stratification, ranging from 'well-mixed' estuaries that have little or no vertical stratification and all the way to 'salt wedge' estuares that have strong vertical stratifications (Figure 100). This vertical stratification, in turn, is tightly related to the relationship between the strength of the river flow (tending to create a vertical stratification) and mechanical sources of turbulent mixing energy (tending to eradicate this stratification).

Tidal currents (we'll discuss tides in more detail in a later chapter) are usually the main source of turbulent mixing energy, so one very common parameter used for this classification is the ratio of the tidal flow (the volume of water brought into the estuary during half a tidal cycle) to the river flow. So three types of estuaries are approximately defined by this ratio as

Salt wedge	≤ 1
Partially-mixed	1–10^3
Well-mixed	$\geq 10^3$.

But this kind of classification is very crude, and other factors also come into play, especially the estuary topography. *Fjords*, in particular, are complex types of estuaries where the circulation depend strongly on topography. They exist only

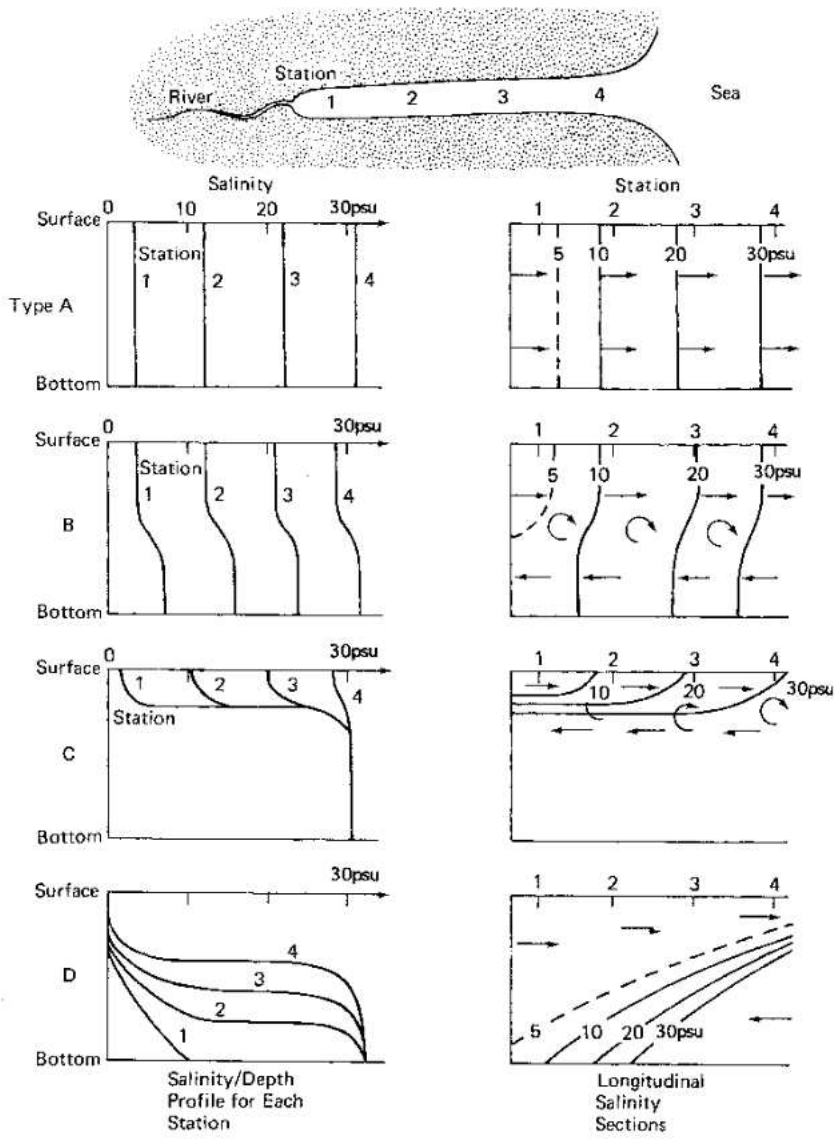

Figure 100: The horizontal structure of salinity and flow structure in four types of estuaries: A: well-mixed, B: partially-mixed, C: fjord-type and D: salt wedge estuaries. (Source: Knauss, 2005, Fig. 11.4)

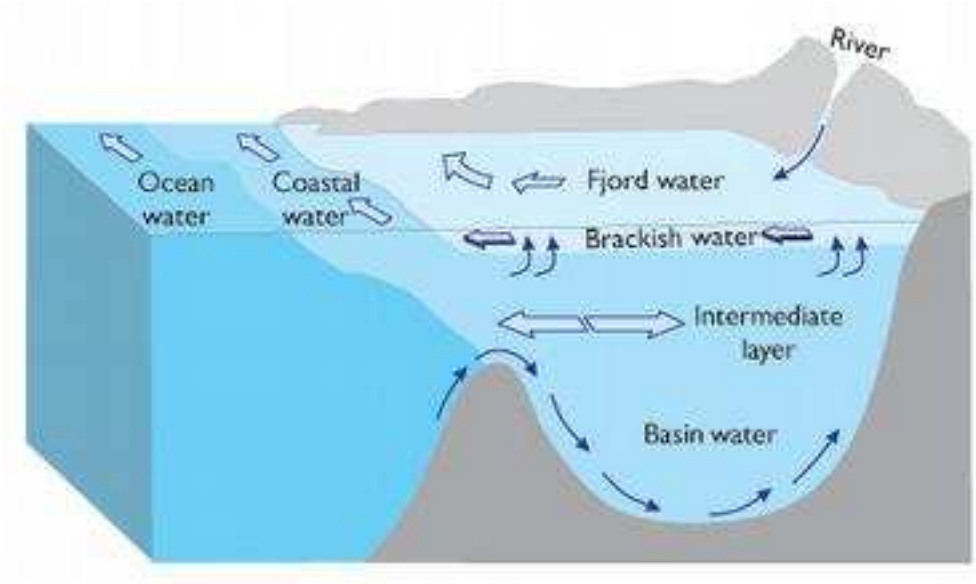

Figure 101: The circulation associated with a silled fjord.

in mid- to high latitude bands that sustain active land glaciers or have sustained such glaciers in past climate periods. The glaciers carve troughs or valleys in the terrain as they advance, and the underwater valleys are what we call fjords. The carved soil and rock pushed ahead at the front of the glacier as it advanced is left behind when the glacier eventually retreats—evidence of how far the glacier once reached. This terminal moraine or *sill* forms a dynamic boundary between the inner fjord and the coastal or open ocean circulation outside and can significantly impact the circulation and hydrography of the fjord (Figure 101).

Typically, silled fjords have a layer of dense stagnant water below the depth of the sill. This water easily gets anoxic, and hence toxic, since biological activity easily uses up the available oxygen faster than the water can be replaced. The sill, plainly speaking, is a barrier to renewal of the bottom waters. The basin bottom water renewal is erratic and is often dependent on the interplay between 1) the availability of denser coastal waters outside of the sill, 2) vigorous tidally-driven turbulent mixing in the sill region and 3) a background turbulent vertical mixing at depth in the fjord. Turbulent mixing at depth makes the bottom water gradually lighter as time passes since the last deepwater renewal event. And at some point the tides will bring outside waters to the sill that are denser than the fjord bottom water. But too strong turbulent mixing in the sill region (it is strong since the

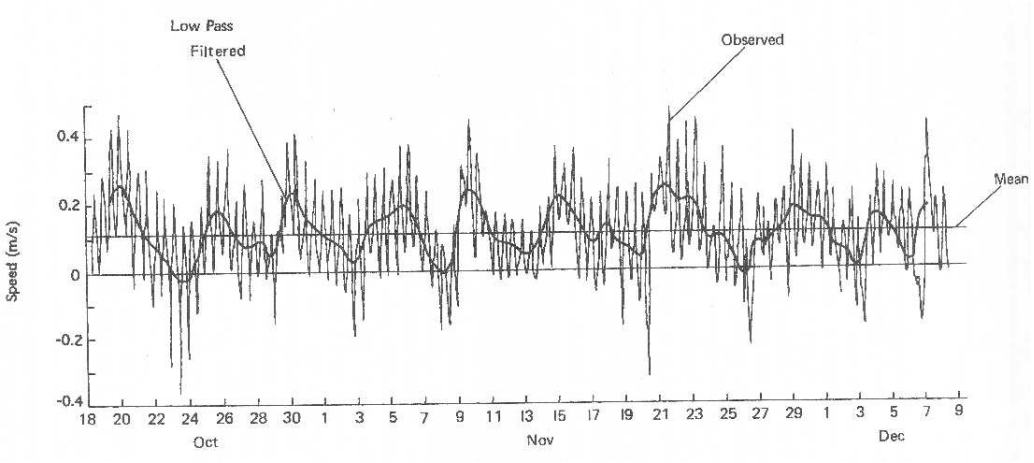

Figure 102: Observations of currents in a real estuary (Narragansett Bay in Rhode Island, USA). The unfiltered observations are dominated by diurnal and semi-diurnal tides while low-pass filtered currents show wind-driven currents. Only when a very long time mean is taken can one detect a clear background estuarine flow. (Source: Knauss, 2005, Fig. 11.7)

currents there are high) can make the water that finally enters over the sill be too fresh and light to replace the bottom water. It is typically during *moderately strong tides* that dense waters from outside pass through the sill and into the fjord, still relatively unmodified by mixing and therefore dense enough to replace the bottom water.

A.3 Real flows in estuaries

The 2D steady-state estuarine flow illustrated in Figure 99 is of course a serious simplification of what real estuarine flows look like. In real estuaries the steady-state estuarine circulation (dense water flowing in at depth, light water flowing out at the surface) is typically small compared to both tidal currents and to wind-driven currents—at least in the so-called partially-mixed and well-mixed estuaries. In fact, the estuarine background circulation may be hard to detect in real observations of the flow, as indicated in Figure 102, and it is only after *low-pass filtering* the observations over time scales longer than both the tides (half a day to a day) and the wind-driven currents (from fractions of a day to several days) that one is able to discern the background flow.

In broader estuaries that have widths larger than, say, a kilometer, Earth's ro-

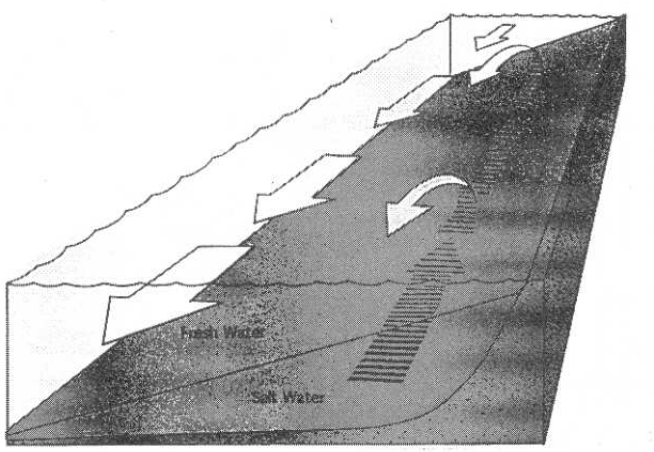

Figure 103: The effect of the Coriolis accelleration on the flow in broad estuaries. The tendency for an accelleration to the right of the flow (in the Northern hemisphere) will cause the in and outflow to hug opposite sides of the estuary. (Source: Knauss, 2005, Fig. 11.5)

tation—the Coriolis accelleration—also has an impact on the flow. In the northern hemisphere the Coriolis accelleration will tend to turn flows to the right. So the outflowing upper layer and the inflowing lower layer will tend to be squeezed up against opposite sides of the estuary, as illustrated in Figure 103. In addition to the sea surface tilt down the axis of the estuary there will also be a surface tilt from the freshwater side to the salt water side. And the interface between the two layers will be tilted the opposite direction. These tilts are a reflection of a near balance between the along-estuary flow and cross-estuary pressure gradients—the *geostrophic balance* which we will get much more aquainted with soon.

B Wind-driven flows in equatorial and high-latitude regions

Seemingly strange things happen at very low and very high latitudes. Near the equator the Coriolis parameter f tends to zero (recall that it is proportional to the sine of the latitude) and one has to expect that the flow is no longer primarily geostrophic. Something other than the Coriolis accelleration has to balance horizontal pressure gradients. In contrast, at very high latitudes the Coriolis parameter is 'alive and well', but its meridional gradient β (proportional to the cosine of latitude) becomes small. Among other things, this should make us suspect that western boundary currents are less important at high latitudes.

B.0.1 Equatorial dynamics

Figure 104 gives a schematic of winds and surface currents near the equator. The *North and South Equatorial Currents* flow westward, in the same direction as the trade winds. The trade winds move north and south with the seasons, but in the annual mean they are shifted slightly north with respect to the equator. And so are the North and South Equatorial Currents. Between these two currents, centered at around 5–7 degrees north, is the *North Equatorial Counter Current*, flowing eastward. It is situated in the Intertropical Convergence Zone (ITCZ), the latitude where the northeast and southeast trade winds meet.

Are these currents ageostrophic since they are so near the equator? Actually it can be shown that geostrophy breaks down only *very* close to the equator (closer than about 2.5°), so most of the flow shown in Figure 104 is geostrophic. And, in fact, a comparison with Figures 69 and 72 suggest that they are probably driven by Sverdrup dynamics!

But there is a current which is situated perfectly on the equator, the *Equatorial Undercurrent* (Figure 105). This eastward-flowing subsurface current can be easily understood if one accepts that at the equator itself the horizontal pressure gradient needs to be balanced by something else than the Coriolis accelleration. What happens here is that the westward-blowing trade winds pile up water on the westward ocean boundary until the resulting pressure gradient due to the sea surface tilt is balanced by something. Very near the surface, down to a few tens of meters, the pressure gradient can be balanced by the wind stress itself. But below this depth the flow will accellerate down the pressure gradient, i.e. to the east, until some form of internal friction balances the pressure gradient. So the Equatorial

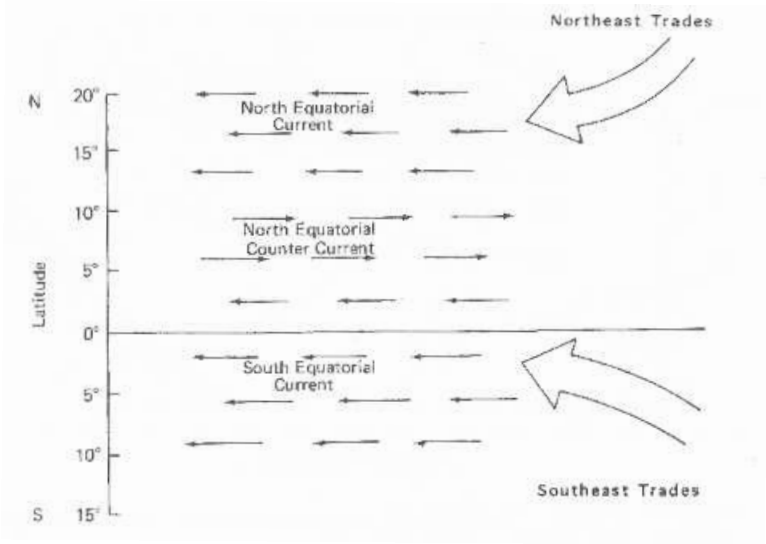

Figure 104: Schematic of winds and surface currents in equatorial regions. (Source: Knauss, 2005, Fig. 7.11)

Undercurrent is a truly ageostrophic phenomenon.

B.0.2 High-latitude dynamics

At high northern and southern latitudes the planetary vorticity gradient becomes small, so we can expect it to have less of an influence on the potential vorticity dynamics than it does at lower latitudes. Also, at high latitudes, constant cooling through the sea surface with the ensuing vertical convective mixing makes the water column very weakly stratified. So it's appropriate to think of the ocean

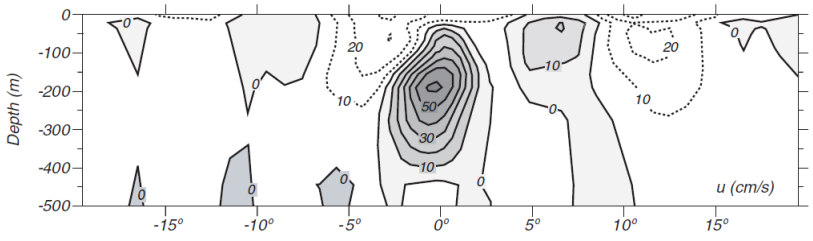

Figure 105: Vertical cross section of equatorial currents in the Pacific Ocean. (Source: Stewart, 2008, Fig. 14.4)

there as consisting of a nearly unstratified layer—at least more so than just about anywhere else in the oceans. And in an unstratified ocean, the bottom topography becomes very important.

In fact, there is little if any evidence that flows are stronger on the western sides of the ocean basins in the high north. Instead, the currents appear to tighly follow topographic features, particularly at high northern latitudes, as indicated by the sea surface field shown in Figure 106. So the Sverdrup and Stommel theories which ignored topography altogether and relied on balances between meridional flows (the βV term) and surface Ekman pumping (or friction in the western boundary currents) are clearly not appropriate here.

Where 1) the flow is constrained to follow bottom topography because of the weak stratification and 2) where this bottom topography forms closed basins, as it does several places in the Nordic Seas and Arctic Ocean, a new possible dynamical balance for the large-scale flow arises. To look at such balance, we need to look not at the vorticity equation but at the potential vorticity (PV) equation. Adding a wind-stress forcing term and also bottom friction to this equation gives

$$\frac{D}{Dt}\left(\frac{f+\zeta}{H}\right) = \frac{1}{H}\left(\nabla \times \frac{\boldsymbol{\tau}^w}{\rho_0} - R\nabla \times \boldsymbol{u}\right).$$

If we then look at steady-state solutions in which advection of relative vorticity is ignored, this becomes

$$H\boldsymbol{u} \cdot \nabla \frac{f}{H} = \nabla \times \frac{\boldsymbol{\tau}^w}{\rho_0} - R\nabla \times \boldsymbol{u},$$

which states that the flow up or down the gradient of the quantity f/H (the 'large-scale' potential vorticity) is given by the difference in rotational frictional forces ("torques") at the top and the bottom. Or, alternatively, by the difference between the vertical pumping out of the top and bottom Ekman layers.

If now, as the observations suggest, the flow follows isolines of bottom topography or, actually, isolines of f/H (the differences is small since f changes very little at high latitudes), then we can write

$$H\boldsymbol{u} \cdot \nabla \frac{f}{H} = 0,$$

the mathematical way of saying that there is no (or, realistically, very little) flow across contours of f/H. But how do we get the direction of the flow and the

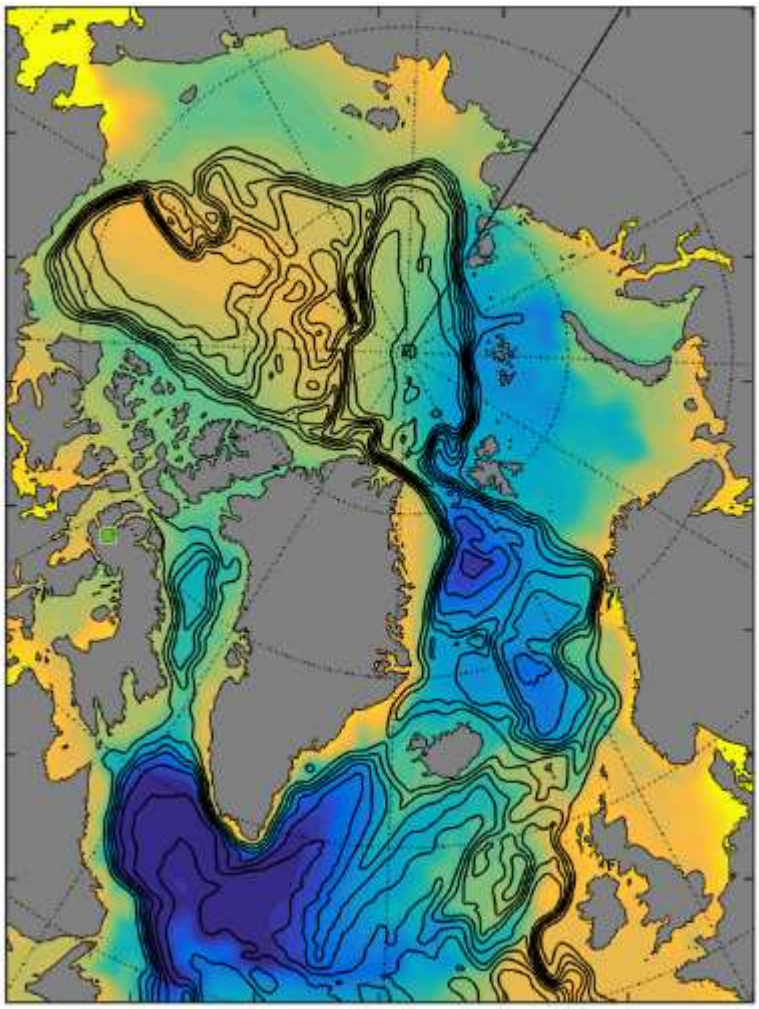

Figure 106: The bottom topography (black lines) and time-mean sea surface height (color shading) from a numerical model of the northern North Atlantic, Nordic Seas and Arctic Ocean.

strength of the flow? Well, if the left-hand side of the PV equation then the right-hand side must also be zero, so that we have

$$\nabla \times \frac{\boldsymbol{\tau}^w}{\rho_0} = R\nabla \times \boldsymbol{u}.$$

Here's how we can find the flow direction and strength: where the f/H contours are closed, we can integrate this last expression over the area of such a contour,

$$\iint \nabla \times \frac{\boldsymbol{\tau}^w}{\rho_0}\, dA = \iint R\nabla \times \boldsymbol{u}\, dA$$

or, using Stoke's theorem,

$$\oint \frac{1}{\rho_0}\boldsymbol{\tau}^w \cdot \hat{\boldsymbol{t}}\, dl = \oint R\boldsymbol{u} \cdot \hat{\boldsymbol{t}}\, dl.$$

Here $\hat{\boldsymbol{t}}$ is a unit tangent vector along the contour so that $\boldsymbol{\tau}^w \cdot \hat{\boldsymbol{t}}$ and $\boldsymbol{u} \cdot \hat{\boldsymbol{t}}$ are the components wind stress and water velocity along the contour. From this expression we at least have the average (the contour integral) of the flow along the contour, both its sign (direction) and its strength.

Essentially, what we have here is a balance between the flow in the top and bottom Ekman layers (compare with the expressions given earlier in this chapter). The situation is illustrated in Figure 107. So, given a convergent surface Ekman transport (and Ekman pumping out of the top layer), the flow along the rim of such a closed f/H region needs to spin up until it has reached speeds exactly large enough that the divergence in the bottom Ekman layer balances the convergence at the top.

The surface flow field estimated from such assumptions, i.e. 1) flows that follow contours of f/H and 2) a balance between top and bottom Ekman layers where f/H contours are closed, is shown in Figure 108 (upper panel). The estimate actually only assumes that the bottom flow follows f/H contours strictly, and that flows higher up in the water column can change via the thermal wind shear due to lateral density gradients. So the model doesn't assume a zero density stratification, but the essential dynamics is governed by the balance discussed above. As seen, the resemblence with velocities deduced from surface drifters (lower panel) is rather good. Presently such 'closed-*f/H* dynamics' is thought to be the best lowest-order description of the wind-driven large-scale flow in both the Nordic Seas and the Arctic Ocean—within regions where the f/H contours close on themselves. Obtaining the flow over the 'open contours', those that don't close on themselves but instead continue down into the North Atlantic (you can see these 'shallow' contours in Fig. 106) require another solution approach.

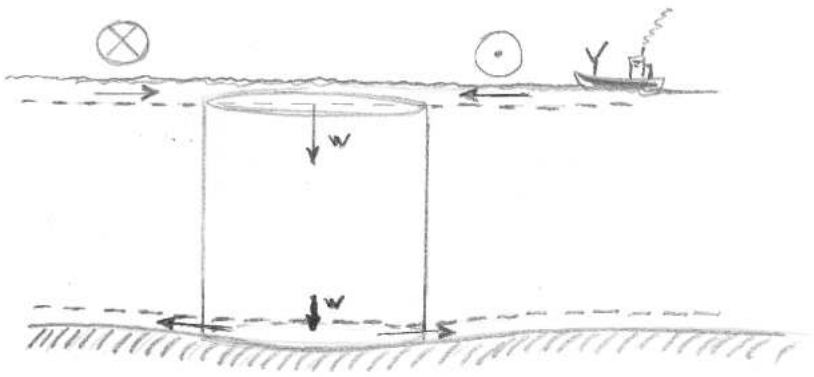

Figure 107: Sketch of closed-f/H dynamics in which there is a balance between top and bottom Ekman layers everywhere.

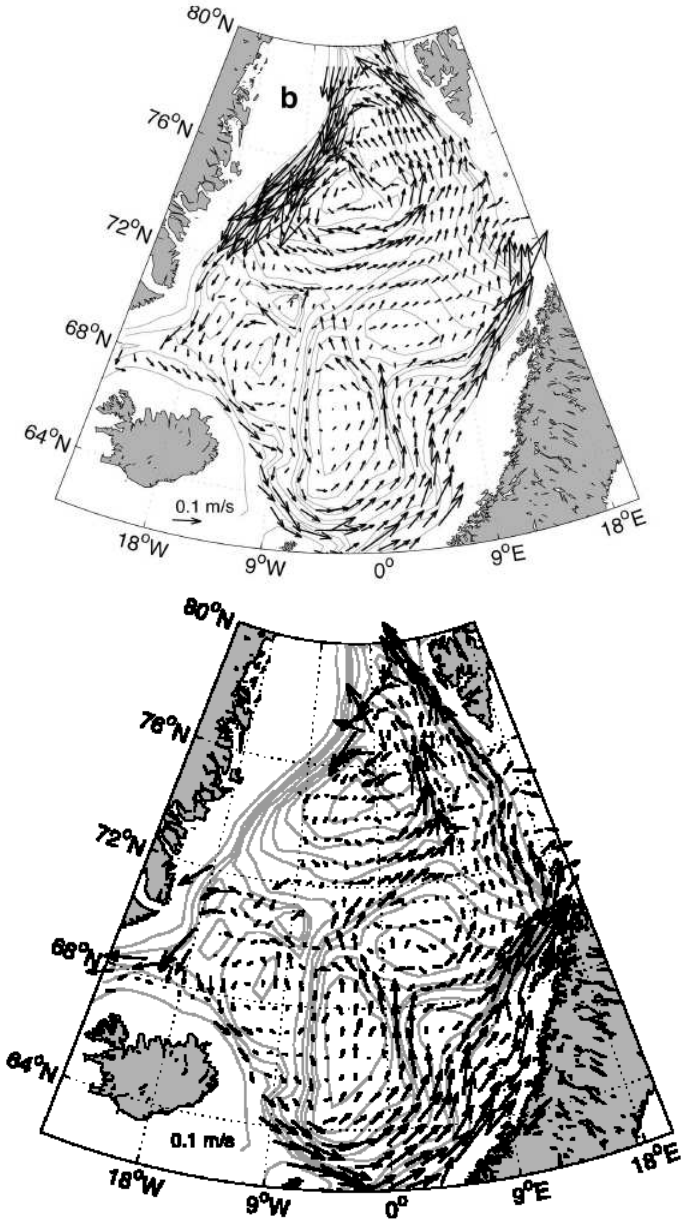

Figure 108: The time-mean surface circulation in the Nordic Seas (bottom) estimated by closed-f/H dynamics and (top) observed by surface drifters (drifter observations of currents in the west are lacking due to the presence of sea ice). (Source: Nøst and Isachsen, 2003, Figs. 11b and 10)

References

[Cushman-Roisin and Beckers(2011)] Cushman-Roisin, B., Beckers, J.-M., 2011. Introduction to Geophysical Fluid Dynamics. Academic Press.

[Knauss(2005)] Knauss, J. A., 2005. Introduction to Physical Oceanography. Waveland Press, Inc.

[Stewart(2008)] Stewart, R. H., 2008. Introduction to Physical Oceanography. Texas A&M University.

[Talley et al.(2011)Talley, Pickard, Emery, and Swift] Talley, L. D., Pickard, G. L., Emery, W. J., Swift, J. H., 2011. Descriptive Physical Oceanography - an introduction, 6th Edition. Elsevier.

[Vallis(2006)] Vallis, G. K., 2006. Atmospheric and Oceanic Fluid Dynamics - Fundamentals and Large-scale Circulation. Cambridge University Press.