MECHANISMS OF NATURAL AND FORCED VARIABILITY IN THE SOUTHERN OCEAN

Submitted for the degree of Doctor of Philosophy

Department of Mathematics and Statistics

Sebastiano Roncoroni

Supervised by David Ferreira

August 2022

Abstract

The Southern Ocean is an important regulator of global climate, and accurately predicting its future evolution under climate change constitutes a critical scientific challenge. Mesoscale eddies are key to the dynamics of the Southern Ocean, but the mechanisms and time scales of their natural and forced variability are not completely understood. Motivated by the dynamical analogy between the Antarctic Circumpolar Current and the tropospheric jet stream, the natural variability of eddy-mean flow interaction is studied by adapting a two-dimensional model of storm track variability to the oceanic case. It is found that eddies and the mean flow interact according to a predator-prey oscillatory relationship in both an idealised, eddy-resolving, channel configuration and the SOSE state estimate product of the Southern Ocean. The oscillatory nature of the dynamics reflects in the structure of the phase space diagrams, where quasi-periodic cycles with typical timescales of a few weeks can be observed. The simplified mathematical model qualitatively captures the statistical properties of the interaction well. The time scales of forced adjustment are investigated by means of an ensemble of wind step-change experiments run with the idealised channel configuration. It is found that the temperature response is driven largely, but not exclusively, by changes in the ocean's circulation, with enhanced mixing also playing an important role. Circulation changes have a rich spatial structure, and vertical/meridional displacements of the residual overturning circulation cells have a large impact on the temperature response even though the channel is strongly eddy-compensated. The time scales of the response vary across the domain, and are set by the spin-up of baroclinic eddies. The results presented in this Thesis bring the fundamental mechanisms of eddy variability into clearer focus, and inform the interpretation of more realistic numerical simulations of the Southern Ocean.

Declaration: I confirm that this is my own work and the use of all material from other sources has been properly and fully acknowledged.

Sebastiano Roncoroni

Acknowledgements

I wish to thank everybody who supported me during the PhD years, and made this journey such an amazing and formative experience. First and foremost, I thank David Ferreira for his inspiring and unfaltering mentoring. I have learnt enormously from you, both in the matters of this Thesis and about scientific research in general. I am grateful especially for the time you dedicated to our meetings and to the revision of the many drafts I have sent you. Most importantly, though, thank you for your continuous encouragement to explore and develop as a researcher. Thanks to Maarten Ambaum and to Valerio Lucarini for the instructive discussions we had and for your valuable suggestions, which deepened my understanding of a broad range of scientific subjects. I wish to thank all staff and students member of the MPE CDT, who collectively created the welcoming working environment I had the privilege of being part of. A special mention goes to the colleagues and friends from Reading's cohort Delta, with whom I shared office desks and many fantastic moments. Thank you also to all my friends in Reading, and particularly to Carlo, Stella, and Marisa, for brightening many weekends with their music. Finally, a huge thanks to my family, who gave me unconditional support and precious advice, and especially to Alice, for always being with me in the good and in the difficult days.

This work has been accomplished with the financial support of the Engineering and Physics Sciences Research Council Centre for Doctoral Training in the Mathematics of Planet Earth.

Osservare tra frondi il palpitare
lontano di scaglie di mare

E. Montale

Contents

Chapter 1

Introduction

1.1 The Southern Ocean: observations and dynamics

This work is dedicated to the study of the Southern Ocean. The Southern Ocean is a region of crucial importance for the global climate but, due to its remote location and severe environmental conditions, is only sparsely observed compared to other ocean basins. Nevertheless, a growing number of observations has become available in recent decades (particularly since the advent of satellite measurements and e.g. of the Argo and SOCCOM projects), which puts several aspects of its dynamics into clearer focus. In this introductory section, I lay out the basic facts and briefly review what is known about the Southern Ocean based on observations.

The remainder of this chapter is structured as follows: the importance of the Southern Ocean for the global climate is discussed in section 1.2. Section 1.3 focuses on the patterns of climate change detected in the Southern Ocean by examining the trends observed in recent decades. Possible explanations for the observed changes and open questions are summarised in section 1.4. Finally, in section 1.5 I explain what the purpose of this Thesis is, and how it fits into the wider research context.

The geography of the Southern Ocean

The Southern Ocean is the portion of the world ocean that encircles Antarctica: the Antarctic continental shelf marks its southern boundary. There is no such geographical boundary to the north, where the Southern Ocean communicates directly with the other ocean basins (Atlantic, Pacific, and Indian). The northern boundary is, rather, of a dynamical nature, and is commonly located around 30°S: equatorward of this latitude, sub-tropical gyres are the dominant feature of the oceanic circulation. Importantly, the longitudinal extent of the Southern Ocean is not restricted by land: Drake passage, the relatively small gap between the southern tip of South America and the Antarctic Peninsula, is where the Southern Ocean is narrowest. A latitudinal band centred at Drake passage exists with no bottom topography above approximately 2000 m (i.e., water is at least 2000 m deep at any point within the band), which gives the Southern Ocean the unique geometry of a re-entrant channel (figure 1.1).

The Antarctic Circumpolar Current

The Southern Ocean hosts the world's largest oceanic current, named the Antarctic Circumpolar Current (ACC). The ACC flows eastward and circumnavigates Antarctica. Its volume transport at Drake passage is estimated around 134 ± 13 Sv (Rintoul et al., 2001), where 1 Sv $= 10^6$ m^{3}s^{-1}. By comparison, the volume transport of the Gulf stream is approximately 30 Sv (Marshall and Plumb, 2008). The ACC has a complex structure and consists of a varying number of interacting jets (Thompson, 2008), which are associated with regions of strong horizontal gradients named fronts (Orsi et al., 1995). The climatological positions of the intense SubAntarctic Front (SAF) and Polar Front (PF) mark the northern and southern flanks of the ACC respectively (see figure 1.1). The large-scale flow of the ACC is zonal (that is, directed along a latitude circle), but the fronts display significant and persistent meridional excursions due to topographical steering (Rintoul et al., 2001), which are particularly pronounced in the lee of major topographical features (for example,

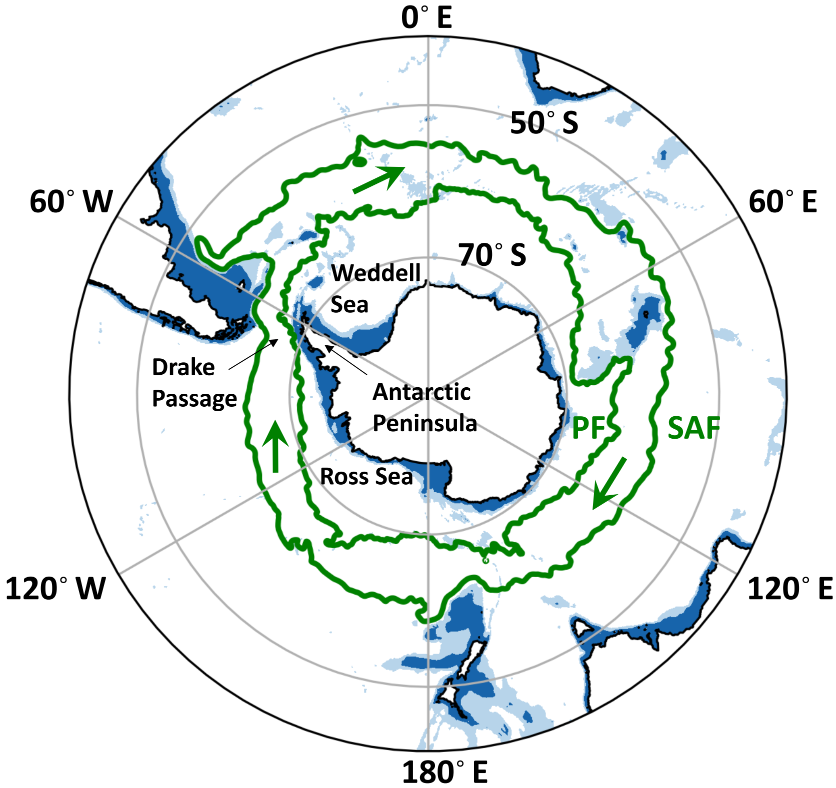

Figure 1.1: Stereographic view of the Southern Ocean region. Continuous green lines mark an estimate of the climatological position of the SubAntarctic Front (SAF) and Polar Front (PF) during the 2005-2010 period (data from SOSE: see chapter 5 for details). Green arrows indicate the sense of circulation of the Antarctic Circumpolar Current (ACC). Regions with bottom topography above 1000 m (between 2000 m and 1000 m) are shaded in dark (light) blue.

downstream of Drake passage).

Drivers of the ACC

The ACC is driven by wind stress and by buoyancy fluxes at the surface. Wind stress is the force imprinted by the wind on a surface of unit area (and is computed from the wind speed by means of semi-empirical expressions named bulk formula), while buoyancy is a quantity closely related to the density of seawater. The wind forcing is supplied by the Southern Hemisphere jet stream, and is predominantly zonal (see for example figure 4 in Marshall and Speer (2012)). The leading mode of variability of the southern winds is called Southern Annular Mode (SAM) (Swart et al., 2015), and a positive phase of the SAM index is associated with a poleward intensified atmospheric jet. The SAM index oscillates between its positive and negative phase with weekly to monthly time scale (Thompson et al., 2011). The buoyancy fluxes stem from a number of different physical mechanisms, including radiative fluxes, freshwater fluxes due to evaporation and precipitation, and sea-ice and ice-sheet processes. The resulting pattern is of buoyancy loss near Antarctica, buoyancy gain over the ACC region, and buoyancy loss north of the SAF (see for example figure 1 in Abernathey et al. (2011)).

The meridional circulation

The surface winds drive not only the strong, zonal flow of the ACC but also, due to the effect of the Earth's rotation, a weaker Ekman flow directed northward and confined to a shallow surface layer. By mass conservation, the northward flow must be accompanied by a return flow directed southward, initially thought to be located in a thin bottom Ekman layer (geostrophic flow below 2000 m depth also contributes). Closure of the circulation would be supplied by upwelling (upward movement of water) near the Antarctic shelf and downwelling (downward movement) north of the ACC: this pattern of circulation is known as the Deacon cell (Doos and Webb, 1994). However, the upper interior of the ocean is stratified, i.e. the vertical gradient of density is not zero. A problem with a theory of the meridional circulation based on the Deacon cell is then that it requires large levels of diapycnal mixing to explain the vertical motions, where diapycnal mixing is a diffusion process that acts across lines of constant density (a parcel of water that is displaced vertically in a stratified environment moves from a surface of constant density to another, and must change its potential temperature or salinity accordingly (Ledwell et al., 1998). Buoyancy sources are largely confined to the surface boundary layer, which means that changes of density in the interior must be accomplished by physical processes akin to diffusion). However, the amount of diapycnal mixing necessary to support the Deacon cell is not observed (Marshall and Speer, 2012). Hydrographical sections of the Southern Ocean offer an avenue towards the solution to this problem. The contours of temperature, salinity, density, and oxygen concentration rise poleward rather than being approximately horizontal, which represents a major difference between the Southern Ocean and the sub-tropical oceans (see figure 1.2 for a schematic and figure 4.6.3 in Rintoul et al. (2001) for observations). A simple set of equations, named the thermal wind equations, relate the negatively sloping buoyancy surfaces to the zonal flow of the ACC. Importantly, the tilted isopycnals (lines of constant density) provide a quasi-adiabatic pathway for deep water to reach the surface and close the meridional circulation. Inversion methods (Marshall and Speer, 2012) confirm that the Meridional Overturning Circulation (MOC) of the Southern Ocean is directed mostly along isopycnals in the interior (see a schematic in figure 1.2). The MOC is partitioned into two main cells, named upper and lower cell, and the 27.6 kg m^{-3} density surface represents an approximate divide between the two (Marshall and Speer, 2012). The cells are counter-rotating, in that the average transport operated by the upper cell is clockwise, while that operated by the lower cell is counter-clockwise. In the meridional plane, in between the upper and lower cell is located the upwelling branch of the MOC. In 3D, parcels follow a spiralling path modulated by topographical features (Tamsitt et al., 2017). The inferred pattern of circulation crosses isopycnal surfaces in the surface layer, where diapycnal flow is supported by buoyancy fluxes: I will examine the associated processes of water mass formation more in detail in chapter 3. A complete dynamical explanation of the MOC of the Southern Oceans requires advanced theoretical instruments, which I discuss in depth in chapter 2. In essence, though, the observed MOC is understood as a small, residual circulation from the partial compensation of two larger contributions: one is the circulation described by the Deacon cell, which acts to tilt isopycnals and is powered by wind stress at the surface. The competing circulation acts to flatten isopycnal surfaces (see figure 1.2), and is induced by baroclinic eddies.

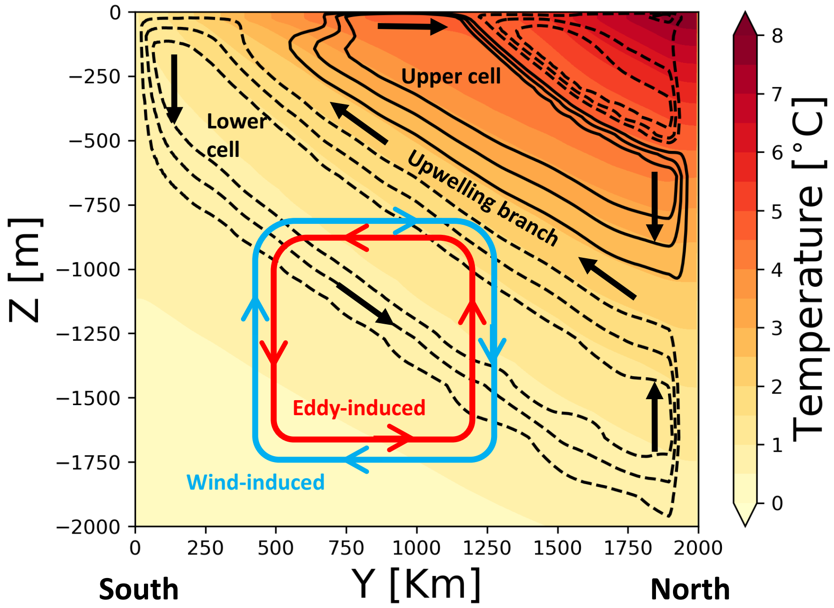

Figure 1.2: Schematic of the Meridional Overturning Circulation. Zonal-mean, time-mean temperature, colours, and time-mean residual overturning circulation, contours. Continuous (dashed) lines indicate positive (negative) values of the residual streamfunction. Data from an idealised configuration of the MITgcm, see chapter 3 for details. Black arrows highlight the sense of the circulation of the upper and lower cells. A third, counter-clockwise rotating cell is also shown near the northern boundary. The action of the wind-induced and eddy-induced circulations is schematically represented by blue and red oriented lines respectively.

Baroclinic eddies

Eddies are turbulent structures that develop in geophysical flows, and represent departures from a mean state. The precise meaning of the word ”eddy”, though, may vary greatly depending on the context, and largely depends on the exact definition of what the mean state is. A first, major distinction is between standing eddies (which are persistent features of the flow: an example is the standing meanders of the ACC, formally, time-mean deviations from a zonal average), and transient eddies (which vanish in the time mean: atmospheric storms are examples). Transient eddies are coherent, vortex-like structures, and oceanic transient eddies of typical size ranging from order ten to a hundred km are called mesoscale eddies. Mesoscale eddies are especially important because they contribute a large fraction of the ocean's total kinetic energy (Frenger et al., 2015). In the Southern Ocean, mesoscale eddies are generated via baroclinic instability (and by barotropic instability to an extent, for example in western boundary currents), which is a mechanism ubiquitous in stably stratified, rotating fluids. Baroclinic eddies tend to cluster along the strongest fronts of the ACC (Frenger et al., 2015), and their life-cycle (growing phase, propagation, and decay) and interaction with the mean jets bear resemblance with those of atmospheric storms (Williams et al., 2007). In the Southern Ocean, baroclinic eddies lived more than one month have a mean radius of approximately 40 km and a mean life span of 10 weeks, during which time they propagate an average distance of approximately 120 km (Frenger et al., 2015, Moreton et al., 2020). On average, they extend vertically down to approximately 2000 m. The distribution of eddy kinetic energy is not uniform throughout the Southern Ocean region, with eddy generation hot-spots found in the lee of topography (Rintoul, 2018). Baroclinic eddies are key to the dynamics of the Southern Ocean for a number of reasons: firstly, they counteract the wind-induced circulation and thereby set the observed MOC, as discussed above. Secondly, baroclinic eddy fluxes are the primary mechanism of near-surface poleward heat transport in the Southern Ocean (Hogg et al., 2008), which is needed to close the heat budget because the part of the Southern Ocean closer to Antarctica is exposed to buoyancy loss. The dominant mechanism by which this transport is accomplished is eddy stirring (Frenger et al., 2015): poleward anomalies of the meridional velocity tend to be co-located with warm temperature fluctuations, and *vice versa*, so that the net effect is to transport heat polewards. Finally, baroclinic eddies communicate zonal momentum from the surface (where it is replenished by wind stress) to the bottom (where it is dissipated by bottom drag and topographic form drag), and thus mediate the zonal momentum balance. In fact, the meridional transport of heat and the downward transfer of momentum are two different manifestations of the same physical mechanism (named form stress), as explained in chapter 2.

1.2 Importance of the Southern Ocean for the global climate

The ocean is a key regulator of global climate because it has the potential to exchange large amounts of heat and carbon with the atmosphere via air-sea fluxes. Due to its unique dynamics, the Southern Ocean is the place where most of these exchanges take place, and thus plays a special role in operating the ocean's climate regulating functions. The upwelling branch of the MOC is a quasi-adiabatic pathway that allows ancient, abyssal water (i.e., water that has not been in direct contact with the atmosphere since pre-industrial times) to upwell and graze the surface, where it is exposed to air-sea fluxes. The upwelling of deep water and ensuing exchanges with the atmosphere are important for a number of reasons: firstly, abyssal water is rich in dissolved inorganic carbon, also called natural carbon (Gruber et al., 2019): part of this water is advected equatorward by the upper cell of the MOC, with release of natural carbon in the atmosphere south of 44°S by outgassing. On the other hand, the Southern Ocean is a sink of natural carbon north of 44°S approximately, and the two effects compensate nearly exactly (Gruber et al., 2019). The amount of natural carbon released in the atmosphere depends critically on the rate of upwelling, and thus on the intensity of the MOC. Modulations of the MOC in glacial-interglacial cycles may have altered the natural carbon sink of the Southern Ocean, with important effects on past climates: for example, a strengthening of the MOC in the transition between glacial and inter-glacial states may have enhanced the outgassing of natural carbon, further warming the climate (Marshall and Speer, 2012). Secondly, the upwelled water is rich in nutrients (i.e., chemical elements involved in the life cycle of phytoplankton), which are (partly) transported outside of the Southern Ocean where they compensate for the downward flux due to sinking of organic matter (Rintoul, 2018). Thirdly, since it was last in contact with the surface centuries ago, the upwelled water is poor in anthropogenic heat and carbon (i.e, heat and carbon produced by human activity), of which it can uptake a large fraction. Frölicher et al. (2015) used the suite of CMIP5 climate models to estimate that, over the period 1865-2005, the Southern Ocean has absorbed up to about 40% of the carbon and up to 75% of the heat produced by human activity. Thus, the Southern Ocean acts prominently to mitigate the effects of global climate change. The uptake of anthropogenic heat and carbon, however, is regulated by the circulation and structure of the Southern Ocean, which are themselves subject to a broad range of changes induced by global warming with unclear consequences for the heat and carbon sink. Before discussing possible future scenarios and open challenges, I briefly review the trends observed in the climatology of the Southern Ocean over recents decades.

1.3 The Southern Ocean in a changing climate

The limited purpose of the following review is to provide the reader with a general sense for the recent changes observed in the Southern Ocean: I refer to the cited literature for a comprehensive illustration. Possible drivers of the observed changes and implications for the future climate are discussed in section 1.4.

Surface winds and SAM index

Surface westerly winds blowing over the ACC have strengthened on average in the recent decades. Swart and Fyfe (2012), for example, found a statistically significant positive trend of surface (zonal-mean) zonal wind stress over the period 1979-2010 (see their figure 1). The annual mean jet position has not shifted significantly over the same period, although a statistically significant poleward shift is observed in the Austral summer (DJF) position. The strengthening is more pronounced during the Austral summer months too. Stronger winds are associated with positive anomalies of the SAM index, and observations indicate that the SAM index has increased in the last few decades. For example, Doddridge and Marshall (2017) found a significant DJF positive trend of the SAM index from 1970 onwards (see their figure 1).

Circulation

Böning et al. (2008) found no significant increase in the slope of the isopycnals and by thermal wind conclude that the ACC zonal transport (and the meridional circulation) have not changed significantly in recent decades. More recently, Hogg et al. (2015) found a small decrease in the ACC circumpolar transport from 1993 onwards. On the other hand, they found a statistically significant positive decadal trend in Eddy Kinetic Energy (EKE), more pronounced in the Indian and Pacific sectors.

Sea Surface Temperature

The sea surface of the Southern Ocean, poleward of the ACC, has cooled over the last few decades (or it has warmed very weakly compared to regions equatorward of the ACC, see figure 1 in Marshall et al. (2014)). For example, Fan et al. (2014) estimate that the Southern Ocean's SST has decreased by 0.4 °C in the annual mean over the period 1979-2011 (see their figure 4), with a more pronounced decrease during the Austral summer. There are, however, important regional differences in the temperature trends: notably, the area around the Antarctic peninsula registered a warming tendency over the same time period.

Sub-surface temperature

The interior of the Southern Ocean, north or within the ACC, has warmed over recent decades. For example, Gille (2008) found that the upper 1000 m of the Southern Hemisphere Ocean warmed at all depths in the period from the 1930s to the early 2000s, particularly in the upper 200 m (see their figure 5). Estimates of interior temperature trends south of the ACC are less reliable due to the paucity of observations (Sallée, 2018).

Sea Ice Concentration

Antarctic Sea Ice Concentration (SIC) has increased on average in the last few decades. Fan et al. (2014) estimate that SIC increased by 12 % in the annual mean during the period 1979-2011. Similarly to the SST response, there are significant regional differences: for example, SIC has decreased near the Antarctic peninsula in the period of time considered by the study. A large and sudden decrease in sea ice extent was observed in 2016-2017 (Meehl et al., 2019), but the overall trend over the 40 year period 1979-2018 remains positive (Parkinson, 2019).

Salinity

The surface and the interior of the Southern Ocean have freshened over time. For example, Durack and Wijffels (2010) found a statistically significant freshening trend for the Southern Ocean region during the period 1950-2008, see their figure 6 for details. Böning et al. (2008) estimate a freshening trend of approximately 0.010 p.s.u. per decade since the 1960s.

Carbon absorption

The Southern Ocean carbon sink has reversed trend, from weakening to strengthening, over the last three decades. Specifically, Gruber et al. (2019) found that the carbon sink weakened in the period 1990-2000 and recovered in the subsequent decade to the levels expected from the atmospheric increase in carbon dioxide. See their figure 5 for details.

1.4 Challenges and open questions

Many of the most hotly debated problems regarding the dynamics of the Southern Ocean concern the physical explanation of the recent observational trends. In this section, I present an overview of the state of the affairs for a selection of these problems. In the interest of brevity, I tie the discussion up with the effect of wind stress changes on the circulation and surface properties of the Southern Ocean. The purpose is to provide the reader with a simplified summary of recent research advances, highlighting which problems have achieved a general consensus and which are still under scrutiny. In the latter case, I briefly examine the main open lines of investigation. Three themes, I believe, are recurrent in this presentation: (i) whether or not decadal trends of surface winds project onto the interannual modes of variability, (ii) the key role of baroclinic eddies for virtually all of the phenomena under consideration, and (iii) the competition between forced versus natural explanations of the observed trends. I shall discuss in section 1.5 how this Thesis fits into the discussion.

1.4.1 Drivers of observed wind changes

There is general consensus that the observed trends in the Southern Hemisphere tropospheric circulation (strengthening of the Southern Hemisphere jet stream and its poleward shift in the Austral summer months; positive trend of the DJF SAM index) have been driven mainly by ozone depletion over Antarctica, although increased concentration of greenhouse gases (GHGs) also played a role (Polvani et al., 2011). The precise mechanism by which ozone depletion induces the observed changes likely involves stratosphere-troposphere coupling, and is not entirely understood yet (Swart et al., 2015). Recovery of stratospheric ozone is expected to drive a negative trend of summertime SAM during the twenty-first century which, however, will be probably offset or overruled by a positive trend induced by increased GHGs concentration (Thompson et al., 2011). Anomalous wind stress is likely going to be influential in the dynamics of the Southern Ocean region over the next decades.

1.4.2 Response to observed wind changes: the ACC circumpolar transport

Wind stress is a primary driver of the ACC. This leads to the question of whether the observed positive trend in the SAM index has the potential to induce long-term changes in the circumpolar flow. There is now general consensus about the fact that the equilibrium ACC transport is rather insensitive to wind stress changes, a phenomenon known as eddy saturation. Note that eddy saturation is a statement about the equilibrated response of the ocean (Munday et al., 2013): the response to wind stress variability on short time scales behaves quite differently, as observations indicate that circumpolar transport is highly correlated with the interannual variability of the SAM index at lags smaller than one year (Hogg et al., 2015, Meredith et al., 2004). Rather than on observations, which (i) cover a relatively short time-span, (ii) are characterised by strong variability, and (iii) do not allow to separate easily between the effects of wind changes and other contributions, the consensus on eddy saturation is rooted in numerical simulations. Early works employing eddy-parametrising general circulation models suggested that the ACC transport would increase significantly in response to wind stress changes (Fyfe and Saenko, 2006), in agreement with simplified theoretical models based on eddy closures (e.g. Marshall and Radko (2003) find that the equilibrium transport scales linearly with wind stress). The advent of eddy-resolving numerical simulations changed this picture drastically: a plethora of studies (Farneti et al., 2010, Hallberg and Gnanadesikan, 2006, Hogg et al., 2015, Morrison and Hogg, 2013, Munday et al., 2013) employing eddy-permitting or eddy-resolving GCM configurations at varying degrees of realism has since then demonstrated that the sensitivity of the circumpolar transport to wind changes is much smaller than previously thought (albeit typically non-zero: the transport increase ranges between 0 - 20 % across models). Note that, by thermal wind, this implies that the slope of the isopycnals is only weakly sensitive to wind stress too. The additional energy input in the system by the excess surface wind stress is not transferred to the mean ACC flow but fuels eddy motions (hence the name eddy saturation), see the next section.

Evidence that the Southern Ocean is close to a state of eddy saturation is supplied by theoretical studies too. By imposing net zero meridional circulation above topographically blocked geostrophic contours, Straub (1993) found that baroclinic instability is a precondition to achieve a statistically equilibrated state: physically, the Ekman transport must be balanced by baroclinic eddies. The Charney-Stern-Pedlosky necessary condition for baroclinic instability (Vallis, 2017) can then be used to estimate the baroclinic circumpolar transport. Notably, the prediction obtained by Straub (1993) does not depend on surface wind stress.

The idea that the circumpolar transport of the ACC is governed by the baroclinic activity of the channel has recently been the object of further investigation. By making assumptions on the physical mechanisms of eddy energy production and dissipation, Marshall et al. (2017) obtained an estimate for the baroclinic circumpolar transport based on the eddy energy balance. The key point is that there must be enough eddy energy production to compensate for dissipation from bottom drag and scattering into lee waves. Since the eddy energy source is assumed proportional to the vertical shear of the zonal velocity (or, by thermal wind, to the slope of the isopycnals), the eddy energy budget is effectively a constraint on the baroclinic zonal velocity. A surprising consequence, corroborated by numerical simulations, is that the predicted baroclinic transport increases with increasing bottom drag: similarly to Straub (1993), however, it is independent on wind stress. Marshall et al. (2017) also shown that, in turn, the eddy energy of the channel is dictated by its zonal momentum balance. The fact that the eddy activity determines the zonal transport while the momentum balance sets the mean eddy energy is reminiscent of the Ambaum and Novak model of atmospheric variability (Ambaum and Novak, 2014), where the diabatic production of baroclinicity sets the mean eddy heat flux, and the dissipation of eddies sets the mean baroclinicity. The Ambaum and Novak model is studied at length later on in this manuscript, see sections 1.5 or 4.1 for an anticipation.

1.4.3 Response to observed wind changes: eddy activity

There is general consensus about the following facts: (i) Eddy activity, as measured for example by EKE, peaks with a lag of 0-3 years after a positive deviation in the wind stress (or SAM index), and (ii) at equilibrium, EKE scales approximately linearly with wind stress. The consensus on point (i) emerges from both observations (Hogg et al., 2015, Meredith and Hogg, 2006, Screen et al., 2009) and numerical simulations with GCMs (Patara et al., 2016, Wilson et al., 2015), with observations also revealing that there are important regional differences in the response. The time lag appears to be dependent on the amplitude of the peak (Meredith and Hogg, 2006), with larger wind anomalies being associated with faster response time scales. Meredith and Hogg (2006) propose that the mechanism responsible for the delayed response is a positive feedback induced by topographic steering on baroclinic generation (Hogg and Blundell, 2006). In this picture, the 0-3 years time scale corresponds to the time needed to communicate surface momentum anomalies to the bottom topography. On the other hand, Patara et al. (2016) finds that EKE correlates negatively with bottom roughness, which conflicts with the mechanism invoked by Meredith and Hogg (2006). Point (ii) is supported primarily by studies based on numerical simulations performed with eddy-resolving GCM configurations (Abernathey et al., 2011, Morrison and Hogg, 2013, Munday et al., 2013). It is possible to model the dependence of equilibrium EKE on wind stress with a (quasi-linear) power law relationship: these scaling theories, however, cannot be easily validated at present as many models have not achieved convergence under grid redefinition yet, and the equilibrium sensitivity of EKE depends on model resolution (Munday et al., 2013).

1.4.4 Response to observed wind changes: meridional overturning circulation

The equilibrium response of the Meridional Overturning Circulation to wind stress changes is less uniform across models than that of the circumpolar transport. The salient concept is that of eddy compensation: the term was originally introduced by Viebahn and Eden (2010) to denote a set of mathematical conditions describing the sub-linear scaling of equilibrium meridional transport with wind stress. Since then, though, its meaning has evolved, and the term is now more generally used to indicate a weak dependence of the MOC on wind stress. Note that eddy compensation and eddy saturation are distinct mechanisms, and an eddy saturated ocean does not necessarily imply an eddy compensated one, and *vice versa* (Morrison and Hogg, 2013). The current view, substantiated by numerical simulations performed with general circulation models, is that the real Southern Ocean is in a partially eddy compensated state. There is, however, significant disagreement between models about the structure and magnitude of the response. For example, Viebahn and Eden (2010), Abernathey et al. (2011), and Munday et al. (2013) use idealised model configurations and find qualitatively similar scaling of the upper cell with wind stress (sub-linear, but with finite sensitivity). However, the lower cell strengthens in Munday et al. (2013) and weakens in Abernathey et al. (2011), possibly due to the different treatment of the northern boundary (Munday et al., 2013). The response of the upper cell is weak in the realistic configuration employed by Farneti et al. (2010), but sizeable in those of Bishop et al. (2016) and Patara et al. (2016).

Compared to the equilibrium sensitivity, the transient response of the MOC to wind stress changes is far less studied, with only a handful of studies having addressed the problem to date (e.g. Doddridge et al. (2019) for an eddy-resolving configuration). The MOC is thought to be one of the major drivers of variability in the Southern Ocean carbon sink (Gruber et al., 2019), therefore a complete understanding of its inner workings will be critical to anticipate future trends of the Southern Ocean carbon uptake.

1.4.5 Drivers of observed SST and SSI changes

There is no clear consensus about what mechanisms drive the observed trends in surface temperature and sea ice. Here, I offer a brief review of recent progress made towards the solution of this problem. The presentation is not comprehensive and focuses on the role of wind stress changes, which is one of the main emphasis of this Thesis. I refer the reader to the cited literature for a more general discussion.

Wind changes and the two-time scale mechanism

A first, debated possibility is that surface wind stress changes (primarily induced by Antarctic ozone depletion) drive the observed SST and sea-ice trends. It is well established from both models and observations that positive anomalies in the SAM index precede surface cooling over and south of the ACC region on interannual time scales, due to anomalous northward Ekman fluxes and turbulent heat fluxes (Ciasto and Thompson, 2008). Furthermore, Doddridge et al. (2019) found that a similar relationship holds between the SAM index and the seasonal sea-ice extent. This naturally leads to the question of whether decadal modulations of the surface winds can explain the observed surface cooling and sea-ice expansion (note that this problem is similar to that discussed in section 1.4.2 for the circumpolar transport: do long-term trends of the wind forcing project on the interannual modes of variability?). Surprisingly, early modelling studies suggested that ozone depletion induces long-term warming and melting of sea ice (Bitz and Polvani, 2012).

A mechanism to reconcile this apparent paradox was put forward by Ferreira et al. (2015), who used the climate response function (CRF) formalism of Marshall et al. (2014) to show that the simulated SST response of two different general circulation models (an idealised configuration of the MITgcm and CCSM3.5) to an instantaneous reduction in ozone cover is endowed with two time scales, a fast and a slow one. The fast, short-term response is characterised by surface cooling operated by anomalous northward Ekman transport of cold water, while the slow, long-term response consists of surface warming due to enhanced upwelling of warm water from below the seasonal sea ice. In this picture, the cross-over time scale (i.e. the time at which a reversal of the response is observed, from cooling to warming) corresponds to the time needed for the sub-surface warm anomaly to entrain the mixed layer.

The two-time scale mechanism of Ferreira et al. (2015) offered an attractive framework to attribute observed SST and sea-ice trends based on physical arguments, but concomitantly opened a number of questions, including: (i) the magnitude and time scales of the SST response vary significantly between the two models considered in the study, which leaves the response of the real Southern Ocean largely unconstrained, and (ii) it is unclear to what extent the proposed mechanism can explain the variability of the observed SST time series under realistic modulations of the ozone forcing.

In order to answer (i) and place a tighter constraint on the time scales of response, Kostov et al. (2017) developed an alternative CRF framework which allows to estimate the SST response function to unit SAM deviations from the internal variability of a model's unperturbed state. This methodology was applied to the preindustrial control simulations of the GMCs included in the CMIP5 suite and revealed that, although the response of many models is indeed characterised by a fast cooling phase followed by a long-term warming, some cool monotonically following a SAM increase. Furthermore, the cross-over time varies over a broad range of values (from a few years to a few decades) for those models that do show sustained warming, with large uncertainties associated with the ensemble-mean response. The intermodel spread, the authors found, can be partially explained by differences in the models' climatologies. Subsequently, Kostov et al. (2018) demonstrated that the CRF technique allows to reconstruct the SST trends simulated in the historical runs of 19 CMIP5 models with good approximation. This technique, however, was less effective for observations: correcting for biases in the models' representation of historical SAM trends reduces the discrepancy between predicted and observed trends only marginally.

Seviour et al. (2016) addressed point (ii) by performing ozone step-change experiments with a comprehensive general circulation model, GFDL-ESM2Mc. They showed that, although the signature of the SST forced response is consistent with the two-time scale model of Ferreira et al. (2015) and statistically significant, its magnitude is small compared to that of natural variability (although Seviour et al. (2017) acknowledge that natural variability is boosted in their model by a possibly unrealistically large mode of convective variability). This casts doubts over the possibility of actually detecting the forced response to wind changes from observations. Complicating the issue, Seviour et al. (2017) found that upwelling of warm water may not be the sole driver of the warming phase of the forced response, with sea-ice induced freshwater fluxes and subsurface mixing also playing a role. Seviour et al. (2019) compared the predicted SST response to instantaneous ozone depletion from published and new model simulations and concluded that: (i) nearly all models considered in the analysis are unable to account for the observed SST trends based on their forced response to ozone changes (most models predict warming from the 1980s to date, whereas the surface cooled), and (ii) nearly all models are unable to account for the observed sea-ice trends based on their forced response to ozone changes (most models predict melting, while sea ice expanded).

Thus, evidence against the two-time scale mechanism accrues. It should be noted, however, that all the general circulation models employed in the works cited above are not eddy-resolving: given the crucial importance of resolving eddies in the similar problem of predicting the ACC's circumpolar transport response to wind stress changes, it seems that further investigation of this mechanism by means of high-resolution models is a key priority going forward. For example, Doddridge et al. (2019) found that the SST transient response to abrupt changes in the surface winds is characterised by sustained cooling (although their simulation is ten years long only and not equilibrated), which is consistent with the response required to reproduce the observed SST trends (Seviour et al., 2019). Before discussing how the problem is addressed in this Thesis, I summarise a few of the alternative mechanisms proposed in the literature to explain the recent observations.

Natural variability

A second possibility is that natural variability alone can account for the observed trends. In the author's opinion, this theory draws part of its strength from the Occam's razor principle: namely, it explains the facts with minimal ingredients. It is well known that idealised (Hogg and Blundell, 2006), fairly realistic (Le Bars et al., 2016), and comprehensive (Gnanadesikan et al., 2020) general circulation models can exhibit modes of internal variability on decadal time-scales. An early work by Polvani and Smith (2013) demonstrated that internal variability spontaneously produces sea-ice extent trends of similar magnitude or larger than the observed ones in the control simulations of four CMIP5-class models. These results were later expanded by Singh et al. (2019), who used a comprehensive climate model to show that sea-ice trends similar to the observed ones can arise concurrently to increasing GHGs concentration. Additionally, Polvani et al. (2021) argue that the observed recent sea-ice expansion was unlikely driven by decadal trends in the SAM index, as the correlation between SIE and SAM is weak, albeit significant, even on interannual scales. It is unclear to what extent natural variability can explain the observed trends in SST (Seviour et al., 2019).

Freshwater fluxes

An alternative mechanistic explanation of the observed SST and sea-ice trends hinges on freshwater fluxes by sea-ice transport. The argument is as follows: enhanced northward transport of sea ice is a major driver of observed salinity trends over the Southern Ocean region (Haumann et al., 2016). Haumann et al. (2020) argue that the anomalous freshwater transport acts to increase the stratification of the open ocean (i.e. the portion of the Southern Ocean away from the coastal region of freezing) and thus reduces upward mixing of warm water from the sub-surface, effectively cooling the surface. In their freshwater flux modulation experiments with a regional ocean model at eddy-permitting resolution, the pattern and magnitude of SST change were found to reproduce observations skilfully. A companion experiment featuring realistic modulations of the surface wind stress, on the other hand, did not result into significant SST variations. A related potential mechanism is that SST and sea-ice trends could be at least partially driven by Antarctic glacial melt, see for example Rye et al. (2020).

1.5 This Thesis

The purpose of this Thesis is to investigate the role of geostrophic eddies in setting the natural variability of the Southern Ocean and its forced response to wind stress changes.

The Southern Ocean exhibits natural variability on a wide range of time scales, from interannual to multi-decadal and longer. Natural variability on long time scales hinders the physical attribution of recent observational trends (or perhaps drives some of them), and thus hampers our ability to make confident predictions about the future state of the Southern Ocean. Variability on shorter time scales is important because the response of the Southern Ocean to modulations of the mechanical forcing at the surface may project onto interannual modes of internal variability.

The sea surface temperature response to wind stress changes is a prominent example: positive SAM fluctuations lead cool surface anomalies on interannual scales, but will decadal trends in the SAM index induce sustained cooling of the Southern Ocean's surface? Baroclinic eddies are a crucial and, yet, not fully understood component of the answer to this and similar questions. It is known that eddies may partially compensate for wind-induced changes in the meridional overturning circulation, thereby delaying or suppressing altogether the warming phase of the two-time scale mechanism proposed by Ferreira et al. (2015). Moreover, by modulating the response of the MOC to wind changes, the influence of the mesoscale eddy field extends to the salinity, carbon and heat budgets of the Southern Ocean, with implications for global climate. The mechanisms that set the magnitude and time scales of response of baroclinic eddies to wind-stress changes, however, remain elusive.

In this manuscript, I address the problem by exploring the dynamics of (i) the natural interaction between eddies and the mean flow and (ii) the forced response of eddies and of the circulation of the Southern Ocean to wind stress changes. My analysis rests primarily on numerical simulations run with an idealised, eddy-resolving configuration of a general circulation model, the MITgcm, although a realistic state estimate product, the SOSE, is considered too. Given the idealised nature of this study, the results presented here cannot be expected to contribute directly towards the problems discussed in the previous sections, namely, I do not make quantitative predictions about the future state of the Southern Ocean. However, I hope that by focussing on some of the key physical mechanisms involved, this work will trigger and feed further studies of the driving factors of the recent observed changes.

In order to investigate the interaction between eddies and the mean flow in the unperturbed numerical simulations, I exploit a well-documented dynamical analogy between the Southern Ocean's baroclinic eddy field and the atmospheric storm track (Thompson, 2008, Williams et al., 2007). Specifically, I will adapt a simplified theoretical model of atmospheric storm track variability formulated by Ambaum and Novak (2014) to the oceanic case, and I will show that the model skilfully captures the salient traits of eddy-mean flow interaction in both the idealised configuration of the MITgcm and in SOSE. The theoretical model consists of a two-dimensional dynamical system which describes the interplay between eddies and the large-scale flow in terms of a predator-prey relationship, akin to that expressed by the Lotka-Volterra equations of population growth (Trefethen et al., 2018). In this picture, baroclinic eddies behave as a population of predators feeding on a pool of preys, that is, the available potential energy stored in the mean flow: feedbacks between the two populations induce periodic oscillations. The eddy life-cycle discussed by Ambaum and Novak (2014) is depicted schematically in figure 1.3, and takes place as follows: following an increase in the mean flow, as measured for example by the slope of the isopycnals (Phase I), more energy is available to be consumed by eddies. Accordingly, the eddy activity (measured e.g. by the eddy heat flux) begins to increase too (Phase II) until, after a certain amount of time, it attains its peak value. The increased eddy activity depletes the mean flow energy reservoir, leading to a decrease in the slope of the isopycnals (Phase III). Eventually, there is no longer enough energy to sustain an intense eddy activity, and the eddy heat flux falls to a minimum (Phase IV). In the absence of eddies, the additional energy supplied by an external forcing can then be converted into baroclinicity, and the cycle repeats. In this work, I will consider a linearised version of the theoretical model, and I will improve it by explicitly accounting for stochastic fluctuations. For both the idealised MITgcm configuration and the SOSE, I find that the eddy life-cycle of figure 1.3 captures the dynamics of eddy-mean flow interaction accurately.

The study of the transient response of the Southern Ocean to wind changes is conducted by running and analysing an ensemble of wind stress step-change experiments with the idealised configuration of the MITgcm. The step-change approach has a significant provenance in the literature, and its main advantage is that linear theory (see e.g. Seviour et al. (2016) for the conditions of applicability) allows to compute the response to an arbitrary time modulation of the forcing from the step-change response economically (Hasselmann et al., 1993, Lembo et al., 2020). Alongside this ensemble of simulations, I consider an individual, long-time integration of the perturbed system, which allows to investigate the time-mean properties

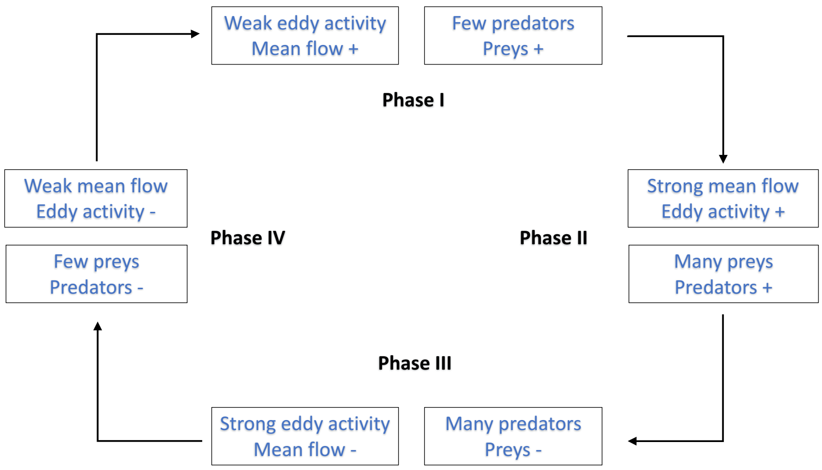

Figure 1.3: Schematic of the eddy life cycle. See main text for details.

of the equilibrated channel too. Thus, although the transient ensemble cannot be run to equilibrium due to the computational cost of the task, I can indirectly estimate how far its members are from the statistically equilibrated state. In order to determine exactly which processes drive SST and sub-surface temperature trends in the model, I diagnose the terms of the temperature budget, which I recombine in a novel way to quantify the contribution of the residual circulation to the advection terms. I find that the idealised channel configuration reacts to a wind step-change increase with a complex spatial pattern of circulation and temperature response. Although the time scales of adjustment are similar to those found by other studies (Doddridge et al., 2019), the detailed analysis presented here reveals that surface and subsurface temperature trends are nontrivially driven by a combination of anomalous circulation and mixing contributions. The meridional circulation compensates within about three years but, importantly, this time scale does not communicate straightforwardly to temperature changes.

The Thesis is structured as follows: following well-established research, in chapter 2 I discuss the dynamics of the Southern Ocean from a theoretical standpoint. In chapter 3, I introduce the MITgcm idealised channel configuration, and show that its reference state constitutes a plausible representation of the real Southern Ocean. I test the predictions of the simplified theoretical model of eddy-mean flow interaction with data from the MITgcm idealised channel and from SOSE in chapters 4 and 5 respectively. Chapter 6 is dedicated to the study of the wind step-change experiments. I offer conclusions and future perspectives in chapter 7.

Chapter 2

Theoretical background

2.1 Introduction

This chapter is dedicated to a review of the theory pertaining the dynamics of the Southern Ocean: its main purpose is to provide the unfamiliar reader with the minimal tools needed to understand the subsequent parts of this manuscript. When relevant, the material is presented under a set of restrictive assumptions, including the fact that the domain is simplified to a zonally symmetric re-entrant channel. These assumptions reproduce those of the GCM configuration used to perform numerical simulations of the Southern Ocean in the rest of this work and, at the cost of generality, facilitate comparison between the analytical equations and the numerical model. A secondary goal of this chapter (which I pursue in chapter 3 too) is thus to pinpoint to what extent existing theory explains the output of the numerical simulations considered later on. The material is potentially vast and the presentation is not comprehensive for reasons of space, but I strived to strike a balance between brevity and self-consistency. The chapter is laid out as follows: in section 2.2 I present the equations of motion, and illustrate how they can be simplified by means of scaling arguments. Section 2.3 is devoted to baroclinic instability, the primary mechanism by which turbulence is generated in the Southern Ocean. Section 2.4 explores Eulerian Mean and Transformed Eulerian Mean theory. In section 2.5, finally, I use the formal equipment introduced above to discuss a number of dynamical models specific to case of the Southern Ocean.

2.2 The equations of motion

The equations of motion for a stratified, rotating fluid are outlined in this section. Specifically, we work with the Boussinesq approximation on the β-plane: this choice is suited to the oceanic case (Vallis, 2017), popular in the literature (Cessi et al., 2006, Eden and Greatbatch, 2008, Jansen and Ferrari, 2012, Sinha and Abernathey, 2016), and common in the modelling practice (Abernathey et al., 2011). Furthermore, the general circulation model used to perform the numerical simulations presented in this study relies on the Boussinesq equations. The purpose of this introduction is to set the notation and to assemble a theoretical reference for the subsequent chapters. We do not attempt to provide a rigorous and complete presentation. The discussion follows Vallis (2017), to which we refer the reader for an extended illustration.

2.2.1 Boussinesq equations

We study the Boussinesq equations on the β-plane. The β-plane approximation consists in introducing a locally Cartesian coordinate system (tangent to the surface of the Earth at a given latitude) to avoid the complications due to the Earth's sphericity. The Boussinesq approximation hinges on the observation that, in the ocean, density variations are small with respect to the mean value of density:

$$\rho(t,x,y,z) = \rho_0 + \delta\rho(t,x,y,z), \tag{2.1}$$

with $|\delta\rho| << \rho_0$. Here, ρ_0 is the (constant) reference value of the density, $\rho_0 = 999.8$ kg m^{-3}. The reference pressure $p_0(z)$ is defined as the pressure that is in hydrostatic balance with the reference density:

$$\frac{\mathrm{d}}{\mathrm{d}z} p_0(z) = -g\rho_0. \tag{2.2}$$

Typically, pressure variations are small compared to the reference profile $p_0(z)$ too, so that we can write:

$$p(t,x,y,z) = p_0(z) + \delta p(t,x,y,z), \tag{2.3}$$

with $|\delta p| << p_0$. Informally, the Boussinesq equations are obtained by neglecting the terms proportional to $\delta\rho$, except when they are multiplied by the gravitational acceleration constant g. More rigorously, an asymptotic expansion of the dynamical variables in the small parameter $|\delta\rho/\rho_0|$ is effected.

The Boussinesq equations on the β-plane are:

$$\partial_t \mathbf{u} + (\mathbf{v}\cdot\nabla)\mathbf{u} + \mathbf{f}\times\mathbf{u} = -\nabla_z\phi + \mathbf{F} \tag{2.4}$$

$$b = \partial_z\phi \tag{2.5}$$

$$\nabla\cdot\mathbf{v} = 0 \tag{2.6}$$

$$\partial_t b + (\mathbf{v}\cdot\nabla)b = \mathscr{B}. \tag{2.7}$$

Notation: $\mathbf{v} = (u,v,w)$ is the three-dimensional velocity vector, and $\mathbf{u} = (u,v,0)$ is the horizontal velocity. $\nabla = (\partial_x, \partial_y, \partial_z)$ is the three-dimensional gradient operator, and $\nabla_z = (\partial_x, \partial_y, 0)$ is the horizontal gradient operator. It makes sense to treat the horizontal and vertical operators separately because, due to the small aspect ratio (i.e., the fluid is wider than deep), the flow scales differently in the two directions. $\mathbf{f} = (0,0,f)$ is the Coriolis parameter, which accounts for Earth's rotation. On the β-plane f is allowed to vary with latitude, namely $f = f_0 + \beta y$. Typical values of f_0 and β relevant for the dynamics of the Southern Ocean are $f_0 = -1\cdot 10^{-4}\ \mathrm{s}^{-1}$ and $\beta = 1\cdot 10^{-11}\ \mathrm{s}^{-1}\ \mathrm{m}^{-1}$. Note that f is negative in the Southern Hemisphere. ϕ is the deviation pressure field divided by ρ_0: $\phi = \delta p/\rho_0$. F is a frictional source of momentum: it may represent wind stress applied at the surface of the ocean (which inputs momentum into the fluid) or bottom drag (which removes it). b is the buoyancy:

$$b = -g\frac{\delta\rho}{\rho_0}. \tag{2.8}$$

Note that b has the units of an acceleration. A particle of fluid denser than the reference density ρ_0 is negatively buoyant, and will tend to sink if immersed in a fluid environment at the reference density . Conversely, a particle of fluid less dense than the reference density is positively buoyant. $\mathscr{B}$ is a diabatic source of buoyancy: it may represent heat fluxes at the interface between the sea and the atmosphere. Equation (2.4) is the horizontal momentum equation. The term $\partial_t \mathbf{u} + (\mathbf{v} \cdot \nabla)\mathbf{u}$ represents advection of horizontal momentum by the three-dimensional velocity $\mathbf{v}$. $\mathbf{f} \times \mathbf{u}$ represents the Coriolis effect, and $-\nabla_z \phi$ is the pressure force term. Equation (2.5) is the vertical momentum equation, simplified by hydrostatic scaling. Equation (2.6) is the mass conservation equation, expressing incompressibility. Finally, equation (2.7) is the thermodynamic equation, which is obtained by neglecting compressibility effects. The term $\partial_t b + (\mathbf{v} \cdot \nabla)b$ represents advection of buoyancy. These equations form a closed set (the simple Boussinesq equations). They are complemented by an equation of state, which relates the thermodynamical variables to each other. In the simplest case (no salinity and no compressibility effects):

$$b = g\alpha(T - T_0), \tag{2.9}$$

where a typical value for the thermal expansion coefficient is $\alpha = 2 \cdot 10^{-4}$ K^{-1}. The reference temperature value T_0 is dynamically unimportant as only buoyancy gradients enter the equations of motion.

2.2.2 Geostrophic balance and thermal wind

Problems of geophysical fluid dynamics are characterised by the typical values of the variables that describe the state of the system. These values can be postulated *a priori*, or inferred from observations, or a mix of the two. If we are interested in large scale oceanic flows, for example, we set the horizontal length scale $L = 1000$ km, and learn from observations that the appropriate horizontal velocity scale is $U = 0.1$ m s^{-1}. This can be exploited to simplify the equations of motion. Intuitively, the characteristic scales indicate which terms in the equations are larger in magnitude, and thus contribute to the dynamics in greater measure. More formally, they allow to identify a small, non-dimensional parameter which is used to effect a perturbation expansion of the state variables. The lowest order terms in the perturbation parameter express the dominant dynamical balance for the specific problem under consideration. In the case of large-scale oceanic and atmospheric flows, the relevant parameter is the Rossby number:

$$Ro = \frac{U}{f_0 L}. \tag{2.10}$$

The Rossby number represents the ratio between the typical scales of the advection and Coriolis terms: here, $Ro << 1$. A second important (dimensional) quantity is the Rossby radius of deformation L_d, which is variously defined in the literature depending on the specific setup but in the simplest case takes the form:

$$L_d = \frac{N_0 H}{f_0}, \tag{2.11}$$

where H is the vertical scale of the problem and N_0 a measure of stratification. In the Southern Ocean $L_d \approx 10 - 30$ km. The geostrophic momentum equations are derived from the Boussinesq equations when $Ro << 1$, the horizontal scale L is much larger than the deformation radius L_d, and frictional terms can be neglected:

$$\mathbf{f} \times \mathbf{u}_g = -\nabla_z \phi, \tag{2.12}$$

where $\mathbf{u}_g$ is the geostrophic velocity. Equation (2.12) expresses geostrophic balance, i.e. the fact the for large scale motions in the ocean and in the atmosphere the Coriolis term is approximately equal to the pressure force term. One of the implications of geostrophic balance is that geostrophic winds (or currents) tend to be tangent to isobars. Geostrophic balance combined with hydrostatic balance gives the thermal wind relations:

$$f \partial_z v_g = \partial_x b \tag{2.13}$$

$$f \partial_z u_g = -\partial_y b, \tag{2.14}$$

i.e., the vertical shear of the horizontal velocity is related to the horizontal gradient of buoyancy.

2.2.3 Quasi-geostrophic equations

Quasi-geostrophic equations are obtained by assuming that the Rossby number is small (as for the geostrophic equations), and that the horizontal scale of motion is not significantly larger than the Rossby radius of deformation. For a stratified fluid, this implies that variations in stratification are small compared to the mean stratification. The quasi-geostrophic equations are widespread in the literature, and have the advantage that they reduce to a single diagnostic equation for the quasi-geostrophic potential vorticity q:

$$\partial_t q + J(\psi, q) = 0. \tag{2.15}$$

Here, J is the Jacobian operator: $J(\psi, q) = -\partial_y \psi \partial_x q + \partial_x \psi \partial_y q$. Potential vorticity is related to the other variables through:

$$\begin{aligned} q &= \nabla_z^2 \psi + \beta y + \partial_z \left(\frac{f_0^2}{N^2} \partial_z \psi \right) \\ \mathbf{u} &= \mathbf{k} \times \nabla \psi \\ \phi &= f_0 \psi \\ b &= f_0 \partial_z \psi , \end{aligned} \tag{2.16}$$

where b and ϕ are the buoyancy and pressure deviations from a reference vertical profile.

2.3 Baroclinic instability

Baroclinic instability is a physical process that concerns rotating, stably stratified fluids. In essence, it is a mechanism that acts to transform available gravitational potential energy into turbulent motion, or eddies. We study baroclinic instability because it constitutes the primary route by which eddies are released in the Southern Ocean (Marshall and Speer, 2012), where they contribute to the dynamical balance at leading order. The key ingredients of baroclinic instability are a stable stratification (otherwise the fluid is convectively unstable), baroclinicity (i.e., there exists a meridional gradient of buoyancy) and rotation (necessary to support a zonal mean flow via the thermal wind equation).

Qualitatively, baroclinic instability may be illustrated by means of a simple parcel argument as follows. Our discussion is adapted from Marshall and Plumb (2008) and Vallis (2017), to which we refer for an extended treatment. Consider first a stably stratified fluid with flat isopycnals (figure 2.1 (a)). Convective stability is guaranteed provided the vertical density gradient is negative, in this case $\rho_1 < \rho_2$: to prove the point, we imagine to swap positions of two fluid parcels A and B. Parcel A finds itself at a higher height but in a lighter environment, and will thus sink under the action of a restoring buoyancy force. Conversely, parcel B ends up in a denser environment, and will be propelled upwards. Overall, the initial disturbance to the background stratification decays, and the fluid is convectively stable. In terms of energetics, the initial potential energy (per unit volume) is:

$$U_i = \rho_2 g z_A + \rho_1 g z_B \,, \tag{2.17}$$

while the potential energy immediately after the parcels have swapped places is:

$$U_f = \rho_2 g z_B + \rho_1 g z_A \,. \tag{2.18}$$

The variation of potential energy is:

$$\Delta U = -g(z_A - z_B)(\rho_2 - \rho_1) > 0 \,, \tag{2.19}$$

because $z_B > z_A$ and $\rho_2 < \rho_1$. The potential energy of the fluid increases under vertical displacement of the parcels, therefore the process must be powered externally. Due to the fact that the isopycnals are flat, the outcome is the same regardless of how one chooses A and B within layers 1 and 2, and the system does not exhibit baroclinic instability.

We now turn the attention to the case of a fluid with tilted isopycnals, as in figure 2.1 (b): note that rotation is necessary to balance the meridional buoyancy gradient via the thermal wind equations (Vallis, 2017). The fluid is still stably stratified because $\partial_z \rho < 0$, i.e., if we swap positions of parcels A and C the potential energy

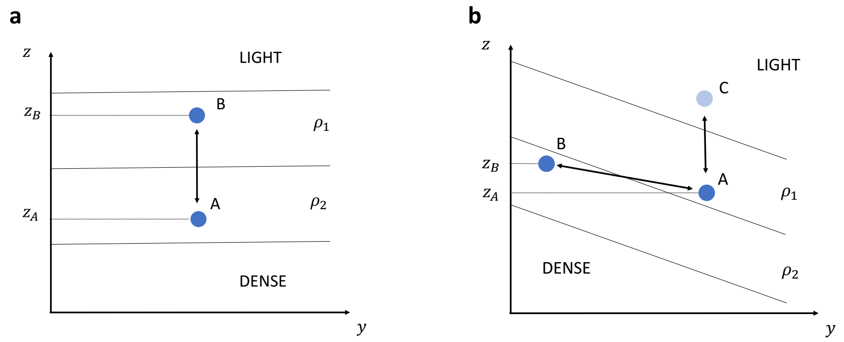

Figure 2.1: Parcel view schematic of (a) a stably stratified fluid with flat isopycnals and (b) a rotating, stably stratified fluid with tilted isopycnals. Fluid (b) is baroclinically unstable.

of the system increases as before. This time, though, the outcome does depend on the initial choice of the parcels. Consider for example an upward displacement by an angle comprised between the horizontal and the local slope of the tilted isopycnals, as for parcels A and B. Then, parcel A finds itself at a higher height and in a less buoyant environment, and will keep rising. Parcel B, on the other hand, is surrounded by lighter fluid, and will keep sinking. Overall, the initial disturbance applied to the system is amplified. In terms of the potential energy of the system, $\Delta U < 0$ because $z_B > z_A$ but $\rho_2 > \rho_1$. By acting to flatten the isopycnals, baroclinic instability releases potential energy and converts it into turbulent kinetic energy.

A variety of models have been put forward over time to describe baroclinic instability quantitatively. Usually, these models consist of the following steps: (i) choice of the equations of motion and of the boundary conditions (ii) choice of the basic state upon which the turbulence develops. This is often a zonal flow with vertical shear (iii) introduction of a small perturbation, for example in the form of a wave (iv) linearisation of the equations (v) study of the conditions under which the amplitude of the wave is allowed to grow. Here, we discuss in some detail the model of Eady (Eady, 1949), one of the pioneering works on baroclinic instability. Apart from the respect it commands due to its historical importance, it allows ourselves to introduce the Eady growth rate, a quantity that we shall use extensively in chapters 4 and 5. To complete the presentation, we also briefly overview the models of Phillips

(Phillips, 1954) and Charney (Charney, 1947). We follow the presentation of Vallis (2017).

2.3.1 Eady model

The equations of motion are the Boussinesq QG equations of section 2.2.3. The geometry of the problem is that of a zonally re-entrant channel (i.e., we impose periodic boundary conditions along x) with null lateral boundary conditions ($\psi = 0$ at the meridional boundaries). The supplementary assumptions, which suit the atmosphere better than the ocean, are that N^2 is constant (uniform stratification), $f = f_0$, and the fluid is confined between two rigid lids. Therefore, the boundary conditions at the top and the bottom are:

$$\partial_t b + (\mathbf{u} \cdot \nabla) b = 0. \tag{2.20}$$

The basic state is a zonal flow with uniform vertical shear, $U(z) = \Lambda z$, where Λ is a positive constant. The associated basic state streamfunction is then $\Psi = -\Lambda z y$, and the basic state potential vorticity is $Q = 0$. We introduce a small perturbation on top of the basic state:

$$\psi = \Psi + \psi', \tag{2.21}$$

and similarly for the other variables, so that for example $q = Q + q'$ and $u = U + u'$. We seek wave solutions of the form:

$$\psi' = \Phi(z) \sin ly e^{ik(x-ct)}. \tag{2.22}$$

We substitute equation (2.21) into the QG potential vorticity equation (2.15), and linearise it by discarding terms that are second order or higher in the perturbation. This yields:

$$(\partial_t + \Lambda z \partial_x)(\nabla_z^2 \psi' + \partial_z (\frac{f_0^2}{N^2} \partial_z \psi')) = 0. \tag{2.23}$$

The final step is to note that the initial disturbance grows when the coefficient c has non-zero imaginary part. By inserting the ansatz (2.22) into equation (2.23), one can work out exactly the values of the zonal and meridional wave numbers k and l such that this condition is satisfied. It is found that, for any given k, the most unstable mode is that associated with the smallest value of the meridional wavenumber l. For fixed $l << 1$, the instability grows the fastest when the zonal wavelength $\lambda = 2\pi/k$ is similar to the deformation radius L_d. The maximum growth rate is called the Eady growth rate, and is given by:

$$\omega = 0.31 \frac{f}{N} \Lambda = 0.31 \frac{f}{N} \partial_z U. \tag{2.24}$$

There are limitations to the Eady model: firstly, it focuses on the initial stages of the instability, but is unable to capture its decaying phase (physical solutions are bounded, and the instability cannot grow indefinitely). Secondly, due to its idealised nature it does not describe real world problems with quantitative accuracy. Nevertheless, it greatly enhances our understanding of how baroclinic instability works. The Eady growth rate is commonly used as a measure of baroclinic instability in atmospheric and oceanic problems (Ambaum and Novak, 2014, Williams et al., 2007), and we follow suit in chapter 4.

2.3.2 Phillips problem

The continuously stratified Boussinesq equations are not the minimal setup to study baroclinic instability, as the vertical structure of the problem can be further simplified by considering a layered fluid instead. The main advantage of this approximation is that it makes it easier to include the effect of β in the model. When the fluid consists of two layers only, the setup is known as Phillips problem. Specifically, the equations of motion are the two-layer QG Boussinesq equations:

$$\partial_t q_j + J(q_j, \psi_j) = 0, \tag{2.25}$$

where $j = 1, 2$ (layer 1 being on top of layer 2), and the potential vorticity in each layer is given by:

$$\begin{aligned} q_1 &= \nabla_z^2 \psi_1 + \beta y + \frac{k_d^2}{2}(\psi_2 - \psi_1) \\ q_2 &= \nabla_z^2 \psi_2 + \beta y + \frac{k_d^2}{2}(\psi_1 - \psi_2). \end{aligned} \tag{2.26}$$

Here, k_d is proportional to the inverse of the Rossby deformation radius:

$$k_d^2 = \frac{8}{L_d^2} = \frac{8f_0^2}{N^2H^2}\,. \tag{2.27}$$

These equations can be derived from the continuously stratified Boussinesq equations by assuming that the two levels are on average $H/2$ deep each, that the fluid is confined between two rigid lids, and by applying finite differencing. The boundary conditions corresponding to the rigid lids are $\partial_z\psi = 0$ at the top and bottom boundaries.

Similarly to the Eady model, the basic state is characterised by uniform vertical shear of zonal velocity, that is $\mathbf{U}_1 = (U,0)$ and $\mathbf{U}_2 = (-U,0)$. This implies $\Psi_1 = -Uy$, $\Psi_2 = Uy$, and:

$$\begin{aligned} Q_1 &= \beta y + k_d^2 U y \\ Q_2 &= \beta y - k_d^2 U y\,. \end{aligned} \tag{2.28}$$

Note that, contrary to the Eady model, the basic state potential vorticity is different from zero even when $\beta = 0$. This is a manifestation of the rigid lid boundary conditions in the layered equations. The next steps are familiar: we add a small perturbation to the basic state:

$$q_j = Q_j + q_j'\,, \tag{2.29}$$

and keep only linear terms in the equations of motion. We assume that the domain is doubly periodic, and seek wave-like solutions of the form:

$$\psi_j' = \tilde{\psi}_j e^{ik(x-ct)} e^{ily}\,. \tag{2.30}$$

The disturbance grows and the system is unstable when the phase speed c is complex. In the interest of brevity, we skip the details of the computation and focus instead on the important results. There are three main cases to consider, depending on the problem settings:

1. If there is no shear ($U = 0$) but β is different from zero the system is always stable. Mathematically, there are two solutions for c and they are both real: one of them corresponds to the phase speed of barotropic Rossby waves.

2. If the system is sheared ($U \neq 0$) but $\beta = 0$, the solution approximates that of the Eady model, and the accuracy of the approximation grows with the number of isopycnal layers. In particular, there is a high-wavenumber cutoff (i.e, the system is unstable only if the scale of the disturbance is $\lambda \gtrsim L_d$), but no critical value of the vertical shear. The growth rate of the instability, though, increases with U, therefore low-shear modes will grow slowly.

3. If the system is sheared ($U \neq 0$) and $\beta \neq 0$: (i) there is a critical value of the shear, $U_1 - U_2 > \beta L_d^2/4$. Physically, this condition guarantees that the meridional gradient of the basic state potential vorticity changes sign within the domain. (ii) There are both a high-wavenumber and a low-wavenumber cutoff to the instability. The disturbance will only grow if its scale λ is such that:

$$L_d \lesssim \lambda \lesssim L_\beta \, , \tag{2.31}$$

where $L_\beta = \sqrt{U/\beta}$ is the Rhines scale (Williams et al., 2007). The low-wavenumber cutoff is not present in the Eady model, but even there modes with $\lambda \gtrsim L_\beta$ grow slowly.

2.3.3 Charney model

It is possible to include the β effect in a continuously stratified model of baroclinic instability: this is the Charney model. Its technical aspects are far less transparent than in the previous two cases and we shall skip them altogether, bar for saying that the background state is of uniform vertical shear Λ, and the equations of motion are the QG compressible equations (depending on the problem parameters, they may simplify to Boussinesq). A critical quantity in the Charney model is the dynamical height h:

$$h = \frac{\Lambda f_0^2}{\beta N^2} \, . \tag{2.32}$$

When h is much smaller then the geometrical depth H, the equations reduce to the Boussinesq approximation and the system behaves as in the Eady model, except that the relevant height scale is h and not H. Thus, for example, an appropriate horizontal scale for the instability is Nh/f_0 rather than $L_d = NH/f_0$. These modes of instability are shallow, and require the presence of near-surface Rossby waves to develop. They cannot be represented neither in the Eady model (because there are no Rossby waves) nor in the Phillips problem (because the simplified vertical structure does not admit shallow modes). The upshot is that there is a high-wavenumber (i.e. small scale) cutoff in the Eady and Phillips models, but not in the model of Charney. When the shear is strong or β is weak, instead, the system is reminiscent of the Phillips problem. In this case, the modes of instability are deep, and the conditions for instability are analogous to those of the layered system (which only represents deep modes).

2.3.4 Comparison

We have presented a number of theoretical models of baroclinic instability. The parcel view argument is important because it constitutes the most intuitive rationalisation of the phenomenon, but it does not convey any quantitative prediction. The models of Eady, Phillips, and Charney, instead, are quantitative models based on modal growth of a perturbation from an unstable background state. They all focus on the initial stages of the instability, when the perturbation is small and the equations may be linearised. However, they differ in the choice of the physical setup for the investigation, which results in solutions with qualitatively different behaviour and varying degrees of richness. The Phillips model hinges on the strong approximation that the fluid is composed by two stacked layers only. This considerably simplifies the equations of motion, so that it is possible to include the effect of β while keeping the problem analytically transparent. The main drawback of this choice is that, by construction, the model only captures deep modes of instability. As a consequence, the model predicts a high-wavenumber cutoff. The Eady model, conversely, adopts the continuously stratified Boussinesq equations, but foregoes the β effect. Setting $\beta = 0$ allows to study the conditions for the onset of instability within a stratified setup in a mathematically appealing way, yet the Eady model too fails to capture shallow modes of instability, because Rossby waves are not represented. The Charney model affords both a continuous stratification and non-zero β, and in this respect is the most complete of the models presented here. The downside is that the technical aspects tend to take central stage, and the interpretation of the results is less intuitive than in the previous cases. In the remainder of this manuscript, the discussion around baroclinic instability is generally couched in terms of the Eady model.

2.4 Eulerian mean and transformed eulerian mean theory

2.4.1 Reynolds decomposition

Reynolds decomposition is a mathematical technique that allows to separate the average value of a variable from its fluctuations. Let c be a scalar variable and let the bar denote an averaging operator. Then, the Reynolds decomposition of c is:

$$c = \overline{c} + c'. \tag{2.33}$$

Note that this equation constitutes the definition of the deviation term c'. A couple of simple properties of Reynolds decomposition are used extensively in geophysical fluid dynamics:

$$\overline{c'} = 0, \qquad \overline{c_1 c_2} = \overline{c_1}\,\overline{c_2} + \overline{c_1' c_2'}. \tag{2.34}$$

These properties rely on the assumption $\overline{\overline{c}} = \overline{c}$, but are otherwise independent on the choice of the averaging operator.

2.4.2 Eulerian mean

While the flow of the Southern Ocean is steered by bottom topography, the ACC is approximately zonally symmetric on large scales (compared to, for example, ocean gyres), which suggests that its dynamics may be investigated effectively and conveniently by averaging fluid properties along latitude circles. The operation of taking the average of a variable along the x direction, at fixed latitude and depth, is called zonal average. The zonal average involves an integration at fixed spatial coordinates, and is thus a special instance of Eulerian mean. In this context, the application of the zonal average operator to the equations of motion (2.4)-(2.7) is called Eulerian mean theory. To study Eulerian mean theory, we make the supplementary assumption that the geometry of the domain can be reduced to that of a zonally re-entrant channel, which is motivated by the existence of a latitudinal band centred at Drake passage with no topography above 2000 meters depth. This conceptual idealisation naturally fosters the introduction of zonal averaging, is the choice of numerous recent modelling studies (Abernathey et al., 2011, Cessi et al., 2006, Doddridge et al., 2019, Ferreira et al., 2015, Viebahn and Eden, 2010, Wolfe and Cessi, 2009), and is adopted for the numerical simulations presented in subsequent parts of this work.

We demonstrate the basic ideas of Eulerian mean theory for the simple Boussinesq equations on the β plane, equations (2.4)-(2.7), with mechanical and thermodynamical forcing and periodic zonal boundary conditions. The zonal-mean equations are:

$$\begin{aligned}
&\partial_t \overline{\mathbf{u}} + \overline{(\mathbf{v}\cdot\nabla)\mathbf{u}} + \mathbf{f}\times\overline{\mathbf{u}} = -\nabla_z\overline{\phi} + \overline{\mathbf{F}} \\
&\overline{b} = \partial_z\overline{\phi} \\
&\nabla\cdot\overline{\mathbf{v}} = 0 \\
&\partial_t\overline{b} + \overline{(\mathbf{v}\cdot\nabla)b} = \overline{\mathscr{B}}.
\end{aligned} \tag{2.35}$$

Here, the bar denotes zonal average:

$$\overline{c} = \frac{1}{L_x}\int \mathrm{d}x\, c. \tag{2.36}$$

where c is an arbitrary scalar variable. An obvious property is $\partial_x\overline{c} = 0$. Note also that, thanks to the periodic boundary conditions, $\overline{\partial_x c} = 0$, and thus the zonal average commutes with the gradient operator. By incompressibility (which implies $\nabla\cdot\mathbf{v}' = 0$) and by operating a Reynolds decomposition, tracer advection terms can be decomposed as:

$$\overline{(\mathbf{v}\cdot\nabla)c} = (\overline{\mathbf{v}}\cdot\nabla)\overline{c} + \nabla\cdot\overline{\mathbf{v}'c'}, \tag{2.37}$$

where the term $\overline{\mathbf{v}'c'}$ is the eddy flux of c. Similarly, the advection term in the momentum equation can be decomposed as:

$$\overline{(\mathbf{v}\cdot\nabla)\mathbf{u}} = (\overline{\mathbf{v}}\cdot\nabla)\overline{\mathbf{u}} + \nabla\cdot\mathbf{R}, \tag{2.38}$$

where:

$$\mathbf{R} = \overline{\mathbf{v}'\mathbf{u}'}, \tag{2.39}$$

is the 2nd-rank Reynolds stress tensor. Overall, the zonal average equations of motion are:

$$\begin{aligned}
&\partial_t\overline{\mathbf{u}} + (\overline{\mathbf{v}}\cdot\nabla)\overline{\mathbf{u}} + \mathbf{f}\times\overline{\mathbf{u}} = -\nabla_z\overline{\phi} - \nabla\cdot\mathbf{R} + \overline{\mathbf{F}}\\
&\overline{b} = \partial_z\overline{\phi}\\
&\nabla\cdot\overline{\mathbf{v}} = 0\\
&\partial_t\overline{b} + (\overline{\mathbf{v}}\cdot\nabla)\overline{b} = -\nabla\cdot\overline{\mathbf{v}'b'} + \overline{\mathscr{B}}.
\end{aligned} \tag{2.40}$$

Note that the terms with zonal gradients are identically zero. We can exploit this fact to cast the equations in a simpler form. Consider for example the zonal-mean buoyancy equation:

$$\partial_t\overline{b} + \overline{v}\partial_y\overline{b} + \overline{w}\partial_z\overline{b} = -\partial_y\overline{v'b'} - \partial_z\overline{w'b'} + \overline{\mathscr{B}}. \tag{2.41}$$

There is no zonal advection term. Furthermore, the zonal-mean velocity on the meridional plane is non-divergent thanks to the zonal-mean continuity equation. Therefore, we can introduce a streamfunction $\overline{\psi}$ such that:

$$\overline{v} = -\partial_z\overline{\psi} \qquad \overline{w} = \partial_y\overline{\psi}, \tag{2.42}$$

where $\overline{\psi}$ is called the Eulerian streamfunction. By integrating both sides of the first part of equation (2.42) vertically, we have that:

$$\overline{\psi}(z) = -\int_{-H}^{z}\mathrm{d}z'\,\overline{v}(z'). \tag{2.43}$$

Overall, the buoyancy equation can be written as:

$$\partial_t\overline{b} + J(\overline{\psi},\overline{b}) = -\partial_y\overline{v'b'} - \partial_z\overline{w'b'} + \overline{\mathscr{B}}, \tag{2.44}$$

where $J(\psi, c)$ is the Jacobian operator. Note that the streamfunction $\overline{\psi}$ does not have the dimensions of a volume transport (meter cube per second). The correct dimensionality is restored by scaling equation (2.43) by the zonal extent of the domain:

$$\overline{\psi}(z) = -L_x \int_{-H}^{z} \mathrm{d}z' \overline{v}(z'). \tag{2.45}$$

This mild misnomer is innocuous and widespread in the literature, therefore we conform to the use by referring to $\overline{\psi}$ as defined by (2.43) or (2.45) indifferently as the Eulerian streamfunction (and similarly for other streamfunctions on the meridional plane). We adopt the Eulerian mean framework to study the zonal balance and the mechanical energy balance of the Southern Ocean in sections 2.5.1 and 2.5.3 respectively. In section 2.5.4, we will see that in the Southern Ocean the Eulerian streamfunction is related to the wind stress applied at the surface.

2.4.3 Transformed Eulerian Mean theory

Above, we have introduced Eulerian Mean Theory and demonstrated it in the specific case of the Boussinesq equations of motion for a zonally re-entrant channel. In this section, we highlight some of the limits inherent to the theory, and illustrate how it can be adapted to address them. To start, consider the steady-state zonal-mean buoyancy equation:

$$(\overline{\mathbf{v}} \cdot \nabla)\overline{b} = -\nabla \cdot \overline{\mathbf{v}'b'} + \overline{\mathscr{B}}. \tag{2.46}$$

In the interior layer, we can assume that the diabatic forcing term is zero, so that:

$$(\overline{\mathbf{v}} \cdot \nabla)\overline{b} = -\nabla \cdot \overline{\mathbf{v}'b'} \qquad \text{Interior.} \tag{2.47}$$

If the divergence of the raw eddy buoyancy flux $\overline{\mathbf{v}'b'}$ is not zero (as it is not in general), then the zonal-mean buoyancy balance requires $(\overline{\mathbf{v}} \cdot \nabla)\overline{b} \neq 0$. The geometrical interpretation is that, in the adiabatic interior, the zonal-mean velocity is not directed along mean isopycnals. Observational and modelling evidence, however, suggests that tracers are advected along mean isopycnals in the Southern Ocean (Marshall and Speer, 2012). Therefore, the Eulerian velocity is not the velocity that advects tracers in the Southern Ocean, and the Deacon cell (the pattern of circulation defined by the Eulerian streamfunction, Doos and Webb (1994)) is not a representation of their average circulation.

What is then the circulation on the meridional plane that on average advects tracers in the Southern Ocean? Mathematically, we seek a streamfunction such that the associated velocity $\mathbf{v}_{\text{res}}$ on the meridional plane satisfies, at steady state:

$$(\mathbf{v}_{\text{res}} \cdot \nabla)\overline{b} = 0 \qquad \text{Interior.} \tag{2.48}$$

Before, we have made the observation that the eddy buoyancy flux, on the right hand side of equation (2.47), is largely directly along mean isopycnals in the adiabatic interior. In order to meet condition (2.48), therefore, we need to find a transformation of the equations of motion such that the skew (= along mean isopycnals) component of the eddy buoyancy flux is represented by advection of mean buoyancy, so that the candidate velocity $\mathbf{v}_{\text{res}}$ represents advection by the Eulerian velocity *and* along-isopycnal advection by skew eddy fluxes. Finding such transformation is the concern of Transformed Eulerian Mean (TEM) theory. Amongst the main advantages of TEM, it simplifies the equations of motion by transferring the eddy buoyancy flux from the buoyancy equation to the horizontal momentum equation, where it can be related to the flux of of a quasi-conserved quantity, potential vorticity. The residual circulation defined in TEM theory is also related to the thickness averaged meridional transport (section 2.4.4), a fact that is systematically exploited in the modelling practice because it makes the associated streamfunction easier to diagnose. In the context of Southern Ocean dynamical theories, it solves the conundrum related to the large interior diabatic transport seemingly implied by the Eulerian circulation, and is key to understand the near balance maintained by wind forcing and baroclinic eddies. TEM theory comes with a number of drawbacks, and the treatment of horizontal boundaries is one of them (section 2.4.3). As one might expect from this preamble, TEM theory can be introduced from a variety of angles: here, we will mainly follow Plumb and Ferrari (2005), and refer to Ferreira et al. (2005), Marshall and Radko (2003), Poulsen (2018), Treguier et al. (1997) to complement the discussion.

The residual velocity

The key idea in TEM is to define a transformed velocity $\mathbf{v}_{\text{res}}$ such that the skew component of the eddy buoyancy flux is represented by mean buoyancy advection. Our definition of residual velocity is:

$$\mathbf{v}_{\text{res}} = \overline{\mathbf{v}} - \nabla \times \psi^* \mathbf{i}. \tag{2.49}$$

The notation convention is that the bar denotes zonal average, $\mathbf{i}$ is the unit vector in the x direction, and $\mathbf{v} = (0, v, w)$ is the velocity vector on the meridional plane. With this, we depart from the notation used in other sections of this manuscript, for the reason that zonal advection is unimportant in the zonal-mean buoyancy equation and we are thus allowed to concentrate on the meridional plane. The first term on the right hand side is the now familiar Eulerian velocity. The second term on the right hand side is the eddy-induced velocity:

$$\mathbf{v}^* = -\nabla \times \psi^* \mathbf{i} = (0, -\partial_z \psi^*, \partial_y \psi^*). \tag{2.50}$$

Here, ψ^* is the streamfunction that describes the eddy-induced circulation on the meridional plane, also called quasi-Stokes streamfunction. Note that the eddy-induced velocity is divergence-free by construction, and thanks to $\nabla \cdot \overline{\mathbf{v}} = 0$ so is the residual velocity.

The buoyancy equation

The zonal-mean buoyancy equation is:

$$\partial_t \overline{b} + (\overline{\mathbf{v}} \cdot \nabla)\overline{b} = -\nabla \cdot \mathscr{F}\{b\} + \overline{\mathscr{B}}, \tag{2.51}$$

where $\mathscr{F}\{b\} = \overline{\mathbf{v}'b'}$ is the raw buoyancy flux. We rewrite equation (2.51) in terms of the residual velocity $\mathbf{v}_{\text{res}}$ by adding $(\mathbf{v}^* \cdot \nabla)\overline{b}$ on both sides:

$$\partial_t \overline{b} + (\mathbf{v}_{\text{res}} \cdot \nabla)\overline{b} = -\nabla \cdot \mathscr{F}_{\text{res}}\{b\} + \overline{\mathscr{B}}, \tag{2.52}$$

where:

$$\mathscr{F}_{\text{res}}\{b\} = \mathscr{F}\{b\} + \psi^* \mathbf{i} \times \nabla \overline{b} \tag{2.53}$$

is the residual buoyancy flux. In writing $\mathscr{F}_{\text{res}}\{b\}$ we have made use of the vector identity:

$$\nabla \cdot (\mathbf{A} \times \mathbf{B}) = (\nabla \times \mathbf{A}) \cdot \mathbf{B} - (\nabla \times \mathbf{B}) \cdot \mathbf{A}, \tag{2.54}$$

with $\mathbf{A} = \psi^* \mathbf{i}$ and $\mathbf{B} = \nabla \overline{b}$, together with $\nabla \times \nabla C = 0$. The derivation of equations (2.52) and (2.53) only depends on the fact that the eddy-induced velocity is non-divergent.

The residual flux

To fully specify $\mathbf{v}_{\text{res}}$ we need to define ψ^*. Remember that $\mathbf{v}_{\text{res}}$ must be such that condition (2.48) holds, which is the case if the skew component of the raw eddy buoyancy flux $\mathscr{F}\{b\}$ is written as advection of mean buoyancy by $\mathbf{v}^*$. Equivalently, by considering the residual form of the zonal-mean buoyancy equation (2.52) we see that condition (2.48) is met provided the skew component of the residual eddy buoyancy flux $\mathscr{F}_{\text{res}}\{b\}$ is zero. This requirement constrains ψ^*, but does not identify it uniquely: we examine some of the main alternatives below. To aid our discussion, we denote by $\mathbf{i}, \mathbf{j}, \mathbf{k}$ the unit vectors in the cartesian coordinates, by $\mathbf{n} = \nabla \overline{b} / |\nabla \overline{b}|$ the unit vector in the direction of the mean buoyancy gradient and by:

$$\mathbf{s} = -\mathbf{n} \times \mathbf{i} = \left(0, -\frac{\partial_z \overline{b}}{|\nabla \overline{b}|}, \frac{\partial_y \overline{b}}{|\nabla \overline{b}|} \right) \tag{2.55}$$

the unit vector normal to the mean buoyancy gradient. Note that $\mathbf{s}$ is southward, along mean isopycnals and for a stable stratification. Equipped with these notations, we can separate the skew and diapycnal components of the raw eddy buoyancy flux as follows:

$$\mathscr{F}\{b\} = (\mathbf{s} \cdot \overline{\mathbf{v}'b'}) \cdot \mathbf{s} + (\mathbf{n} \cdot \overline{\mathbf{v}'b'}) \cdot \mathbf{n}, \tag{2.56}$$

which amounts to a change of basis, from cartesian to isopycnal-diapycnal coordinates. The residual flux is then:

$$\mathscr{F}_{\text{res}}\{b\} = (\mathbf{s} \cdot \overline{\mathbf{v}'b'}) \cdot \mathbf{s} + (\mathbf{n} \cdot \overline{\mathbf{v}'b'}) \cdot \mathbf{n} + \psi^* \mathbf{i} \times \nabla \overline{b}. \tag{2.57}$$

The first term on the right hand side is the skew component of the raw eddy buoyancy flux, therefore our definition of ψ^* must satisfy:

$$\psi^* \mathbf{i} \times \nabla \bar{b} = -(\mathbf{s} \cdot \overline{\mathbf{v}'b'}) \cdot \mathbf{s}. \tag{2.58}$$

With a little algebra one can show that the above implies:

$$\psi^* = -\frac{1}{|\nabla \bar{b}|} \mathbf{s} \cdot \overline{\mathbf{v}'b'}. \tag{2.59}$$

This definition can be generalised if one notes that, in the adiabatic interior, the diapycnal component of the raw eddy buoyancy flux is approximately zero. Thus, we are at liberty of modifying equation (2.59) by adding a term proportional to $\mathbf{n} \cdot \overline{\mathbf{v}'b'}$:

$$\psi^* = -\frac{1}{|\nabla \bar{b}|} \mathbf{s} \cdot \overline{\mathbf{v}'b'} - \frac{\alpha}{|\nabla \bar{b}|} (\mathbf{n} \cdot \overline{\mathbf{v}'b'}). \tag{2.60}$$

The choice of α determines ψ^* univocally. The residual flux takes the form:

$$\mathscr{F}_{\mathrm{res}}\{b\} = \mathbf{n} \cdot \overline{\mathbf{v}'b'}(\mathbf{n} - \alpha \mathbf{s}). \tag{2.61}$$

The quasi-Stokes streamfunction

Following Plumb and Ferrari (2005), we examine three possible choices for α:

(i) The first choice is $\alpha = 0$. It is the most natural choice in that ψ^* is proportional to the skew component of the raw eddy flux:

$$\psi^* = -\frac{1}{|\nabla \bar{b}|} \mathbf{s} \cdot \overline{\mathbf{v}'b'}. \tag{2.62}$$

The associated residual flux is:

$$\mathscr{F}_{\mathrm{res}}\{b\} = (\mathbf{n} \cdot \overline{\mathbf{v}'b'}) \cdot \mathbf{n}. \tag{2.63}$$

The geometrical interpretation is that, for a downgradient diapycnal flux ($\mathbf{n} \cdot \overline{\mathbf{v}'b'} < 0$), the residual flux is directed downgradient.

(ii) The second option is $\alpha = -\partial_y\overline{b}/\partial_z\overline{b}$, so that α is equal to the isopycnal slope. The eddy-induced streamfunction reads:

$$\psi^* = \frac{\overline{v'b'}}{\partial_z\overline{b}}, \tag{2.64}$$

and the associated residual flux is:

$$\mathscr{F}_{\text{res}}\{b\} = \frac{\nabla\overline{b}\cdot\overline{\mathbf{v}'b'}}{\partial_z\overline{b}}\hat{z}. \tag{2.65}$$

The geometrical interpretation is that for a downgradient diapycnal flux the residual flux is directed downward. This is the choice of ψ^* that emerges naturally in quasi-geostrophic theory, and the one that we shall adopt in the rest of the manuscript. The eddy-induced velocity on the meridional plane is the same as in Ferreira et al. (2005), reading:

$$v^* = -\partial_z\left(\frac{\overline{v'b'}}{\partial_z\overline{b}}\right), \qquad w^* = \partial_y\left(\frac{\overline{v'b'}}{\partial_z.\overline{b}}\right). \tag{2.66}$$

A downside of this definition is that it can be problematic in the mixed layer, as explained below.

(iii) The final option is $\alpha = \partial_z\overline{b}/\partial_y\overline{b}$. The eddy-induced streamfunction is:

$$\psi^* = -\frac{\overline{w'b'}}{\partial_y\overline{b}}, \tag{2.67}$$

and the associated residual flux:

$$\mathscr{F}_{\text{res}}\{b\} = \frac{|\nabla\overline{b}|}{\partial_y\overline{b}}\mathbf{j}(\mathbf{n}\cdot\overline{\mathbf{v}'b'}). \tag{2.68}$$

The geometrical interpretation is that for a downgradient diapycnal flux the residual flux is directed horizontally down the horizontal mean buoyancy gradient. This choice of ψ^* was notably employed in Treguier et al. (1997) and in the model of the Southern Ocean circulation of Marshall and Radko (2003), which we explore in section 2.5.5. This definition has the advantage that it applies in the mixed layer too with no need for corrections.

Note that in all cases $\mathscr{F}_{\text{res}}\{b\} = 0$ in the interior provided that $\overline{\mathbf{v}'b'}$ is along isopycnals, so that the definitions differ mainly near horizontal boundaries, at the surface and bottom.

The surface boundary condition

The treatment of horizontal boundaries represents a conceptual and practical challenge to TEM theory. For example, the definition of the eddy induced streamfunction ψ^* in case (ii) above is problematic near the surface because $\partial_z \overline{b}$ is approximately zero in the mixed layer (Treguier et al., 1997). One possible solution to the problem, introduced by Treguier et al. (1997) and adopted in Ferreira et al. (2015), is to divide the domain in an interior layer (where the flow is adiabatic and the diapycnal component of the raw eddy buoyancy flux is approximately zero) and a surface diabatic layer. In the diabatic layer isopycnals steepen under the influence of turbulent mixing and air-sea fluxes, and there is a non-zero diapycnal component of the residual flux, which in the limit of vertical isopycnals becomes parallel to the boundary (Ferreira et al., 2005). Then, one assumes that the quasi-stokes streamfunction varies smoothly from its value at the base of the diabatic layer to zero at the surface (Ferreira et al., 2005):

$$\psi^* = \psi^*|_{z=-h}\mu(z)\,, \tag{2.69}$$

where h is the depth of the diabatic layer and $\mu(z)$ is 1 at the base of the diabatic layer and 0 at the surface. Plumb and Ferrari (2005) note that, since definition (ii) is convenient in the interior (because of the relation with thickness-averaged transport, see below) and definition (iii) is convenient in the surface layer (because the residual flux is horizontal by construction, and it is easier to implement the no-normal flow condition), one can organise a smooth transition between the two as the horizontal boundaries are approached.

Summary

The circulation that on average advects tracers in the Southern Ocean is the residual circulation. This is apparent from the residual form of the zonal-mean buoyancy equation:

$$\partial_t \overline{b} + J(\psi_{\text{res}}, \overline{b}) = 0. \tag{2.70}$$

The equation above is formulated in the interior, where we assume that there are no diabatic contributions from air-sea fluxes and that the eddy buoyancy flux is directed along mean isopycnals. Here, the residual circulation on the meridional plane is described by the residual streamfunction ψ_{res}. The residual streamfunction is defined by:

$$-\nabla \times \psi_{\text{res}}\mathbf{i} = \mathbf{v}_{\text{res}}, \tag{2.71}$$

where $\mathbf{v}_{\text{res}}$ is the residual velocity in the meridional plane. Therefore:

$$\psi_{\text{res}}(z) = -\int_{-H}^{z} \mathrm{d}z' v_{\text{res}}(z'). \tag{2.72}$$

In turn, $\mathbf{v}_{\text{res}}$ is defined by $\mathbf{v}_{\text{res}} = \overline{\mathbf{v}} + \mathbf{v}^*$. By taking the vertical integral of the meridional component of this equation, we obtain:

$$\psi_{\text{res}} = \overline{\psi} + \psi^*. \tag{2.73}$$

The eddy-induced streamfunction must be chosen appropriately to guarantee that the skew component of the residual buoyancy flux vanishes in the interior. Our definition in what follows is:

$$\psi^* = \overline{v'b'}/\partial_z\overline{b}, \tag{2.74}$$

which corresponds to choice (ii) above.

2.4.4 Thickness-averaged meridional transport

The residual streamfunction on the meridional plane is defined by equation (2.73), but this definition is rarely used to compute ψ_{res} in practice, because ψ^* is difficult to diagnose and often a noisy variable. Many studies (Abernathey et al., 2011, Doos and Webb, 1994, Hallberg and Gnanadesikan, 2006, Poulsen, 2018, Viebahn and Eden, 2012, Wolfe and Cessi, 2014, 2015) prefer to take an alternative route which relies on the approximate equivalence between ψ_{res} and the thickness-averaged meridional transport, which we now examine. Our presentation is adapted from the seminal paper McIntosh and McDougall (1996) and from Poulsen (2018), Vallis (2017): the reader is referred to these works for an extended treatment.

Buoyancy and isopycnals

Let $b(t,x,y,z)$ be the buoyancy field: in this section we work at fixed time and meridional coordinates, therefore we write $b(t,x,y,z) = b(x,z)$. The same notation applies to other variables so that, for example, $v(x,z)$ is short for $v(t,x,y,z)$. Let b_1 be an arbitrary value of buoyancy. We say that the isopycnal associated to b_1 is the curve $\eta_1 = \eta_1(x)$ such that $b(x,\eta(x)) = b_1$. Note that the equation $b(x,\eta(x)) = b_1$ actually defines a two-dimensional surface in the three-dimensional (x,y,z) space. However, given that y is fixed, we treat $\eta(x)$ as a one-dimensional curve residing in the two-dimensional (x,z) space.

Isopycnal streamfunction

Consider two isopycnals η_1 and η_2, with $\eta_2 > \eta_1\ \forall x$. The thickness integrated meridional transport between η_1 and η_2 is:

$$T = \int_{\eta_1}^{\eta_2} \mathrm{d}z\, v(x,z). \tag{2.75}$$

The zonal average of the thickness integrated transport is:

$$\overline{T} = \frac{1}{L_x}\int \mathrm{d}x \int_{\eta_1}^{\eta_2} \mathrm{d}z\, v(x,z), \tag{2.76}$$

where the bar denotes zonal average. The zonal integral could in fact be replaced by a circumpolar path, but we shall skip this complication. The isopycnal streamfunction is defined by:

$$\psi_I(b_1) = -\frac{1}{L_x}\int \mathrm{d}x \int_{-H}^{\eta_1} \mathrm{d}z\, v(x,z), \tag{2.77}$$

so that $\psi_I(b_1)$ is minus the zonal average, thickness integrated meridional transport from the bottom to the isopycnal depth $\eta_1(x)$. Sometimes this expression is written in an equivalent but slightly more sophisticated way in the literature (Wolfe and Cessi, 2014, 2015):

$$\psi_I(b_1) = -\frac{1}{L_x}\int \mathrm{d}x \int_{-H}^{0} \mathrm{d}z\, v(x,z)\mathscr{H}(b_1 - b(x,z)), \tag{2.78}$$

where $\mathscr{H}$ is the Heaviside step function. We will limit ourselves to the first formulation here.

Relation between residual and isopycnal streamfunction

The statement we want to prove is:

$$\psi_{\text{res}}(z_1) \approx \psi_I(b_1), \tag{2.79}$$

i.e., the residual streamfunction at fixed depth z_1 is approximated, at leading order in a small perturbation parameter, by the isopycnal streamfunction evaluated at the buoyancy b_1 such that $\overline{\eta}_1 = z_1$. This statement is remarkable (and not intuitive) because the left hand side is expressed in terms of variables defined in height coordinates, while the right hand side resides in isopycnal coordinates. We shall expand on this at the end of the proof.

Useful Taylor expansions

We examine two Taylor expansions that are essential to the main proof, and will also be used later on in this chapter. Firstly, consider the isopycnal $\eta_1(x)$, and let $\overline{\eta}_1$ be its zonal average. The Taylor expansion of the meridional velocity $v(x,z)$ around $\overline{\eta}_1$ is:

$$v(x,z) = v(x,\overline{\eta}_1) + \partial_z v|_{\overline{\eta}_1}(z-\overline{\eta}_1) + \mathscr{O}(\alpha^2), \tag{2.80}$$

where α is the perturbation parameter of the power expansion, representing small deviations of z from the fixed depth $\overline{\eta_1}$. The expression above can be applied to a generic scalar variable q and to the case $z = \eta_1(x)$, where we assume that the fluctuations of $\eta_1(x)$ around $\overline{\eta_1}$ are small:

$$q(x,\eta_1(x)) = q(x,\overline{\eta}_1) + \partial_z q|_{\overline{\eta}_1}\eta_1' + \mathscr{O}(\alpha^2), \tag{2.81}$$

which allows to express a variable evaluated along an isopycnal contour η_1 in terms of variables evaluated at the fixed depth $\overline{\eta}_1$. Buoyancy is a special case, because by definition of η_1 we have that $b(x,\eta_1(x)) = b_1$, therefore:

$$b_1 = b(x,\overline{\eta}_1) + \partial_z b|_{\overline{\eta}_1}\eta_1' + \mathscr{O}(\alpha^2), \tag{2.82}$$

The first order term can be approximated as:

$$\partial_z b|_{\overline{\eta}_1}\eta_1' = \partial_z(\overline{b}+b')|_{\overline{\eta}_1}\eta_1' \approx \partial_z\overline{b}|_{\overline{\eta}_1}\eta_1'. \tag{2.83}$$

We can now use the buoyancy expansion to obtain an expression for η_1'. At leading order:

$$\eta_1' \approx \frac{b_1 - b(x,\overline{\eta}_1)}{\partial_z\overline{b}|_{\overline{\eta}_1}}. \tag{2.84}$$

This can still be improved. Equation (2.82) can be rewritten as:

$$b(x,\overline{\eta}_1) \approx b_1 - \partial_z b|_{\overline{\eta}_1}\eta_1', \tag{2.85}$$

and by taking zonal average we have that at leading order $\overline{b}(x,\overline{\eta}_1) \approx b_1$. Therefore, equation (2.84) becomes:

$$\eta_1' \approx -\frac{b(x,\overline{\eta}_1) - \overline{b}(x,\overline{\eta}_1)}{\partial_z\overline{b}|_{\overline{\eta}_1}} = -\frac{b'(x,\overline{\eta}_1)}{\partial_z\overline{b}|_{\overline{\eta}_1}}. \tag{2.86}$$

A lot of work with these perturbation expansions! We will meet this formula again in section 2.5.2. Finally, we are ready for the main proof.

Main proof

The thickness integrated meridional transport between isopycnals η_1 and η_2 can be decomposed as follows:

$$T = \int_{\eta_1}^{\eta_2} \mathrm{d}z\, v = \int_{\overline{\eta}_1+\eta_1'}^{\overline{\eta}_2+\eta_2'} \mathrm{d}z\, v = \int_{\overline{\eta}_1}^{\overline{\eta}_2} \mathrm{d}z\, v + \int_{\overline{\eta}_2}^{\overline{\eta}_2+\eta_2'} \mathrm{d}z\, v - \int_{\overline{\eta}_1}^{\overline{\eta}_1+\eta_1'} \mathrm{d}z. \tag{2.87}$$

We start by transforming the second and third terms on the right hand side. Substituting the expansion of v around $\overline{\eta}_2$, equation (2.80), into the η_2 integral and keeping terms up to second order yields:

$$\begin{aligned}\int_{\overline{\eta}_2}^{\overline{\eta}_2+\eta_2'} \mathrm{d}z\, v &\approx \int_{\overline{\eta}_2}^{\overline{\eta}_2+\eta_2'} \mathrm{d}z\, \left[v(x,\overline{\eta}_2) + \partial_z v|_{\overline{\eta}_2}(z-\overline{\eta}_2)\right] \\ &= v(x,\overline{\eta}_2)\eta_2' + \frac{1}{2}\partial_z v|_{\overline{\eta}_2}\eta_2'^2.\end{aligned} \tag{2.88}$$

We take zonal average, obtaining:

$$\frac{1}{L_x}\int \mathrm{d}x \int_{\overline{\eta}_2}^{\overline{\eta}_2+\eta_2'} \mathrm{d}z\, v \approx \overline{v(x,\overline{\eta}_2)\eta_2'} + \frac{1}{2}\overline{\partial_z v|_{\overline{\eta}_2}\eta_2'^2}. \tag{2.89}$$

The second term on the right hand side is actually third order in the perturbation, and can be neglected. To see why, substitute $v = \overline{v} + v'$ into $\overline{\partial_z v|_{\overline{\eta}_2} \eta_2'^2}$. The v' contribution is third order by construction, and $\partial_z \overline{v} \approx \mathscr{O}(\alpha)$ by thermal wind (in fact, $\partial_z \overline{v} = 0$ for a zonally re-entrant channel with idealised geometry), so that at leading order:

$$\frac{1}{L_x} \int \mathrm{d}x \int_{\overline{\eta}_2}^{\overline{\eta}_2+\eta_2'} \mathrm{d}z\, v \approx \overline{v(x, \overline{\eta}_2)\eta_2'}. \tag{2.90}$$

By substituting $v = \overline{v} + v'$ and using $\overline{\overline{v}(\overline{\eta}_2)\eta_2'} = 0$, we obtain:

$$\frac{1}{L_x} \int \mathrm{d}x \int_{\overline{\eta}_2}^{\overline{\eta}_2+\eta_2'} \mathrm{d}z\, v \approx \overline{v'(x, \overline{\eta}_2)\eta_2'}. \tag{2.91}$$

The last step with this integral is to replace η_2' with its representation in terms of fixed-depth variables, equation (2.86). We obtain:

$$\frac{1}{L_x} \int \mathrm{d}x \int_{\overline{\eta}_2}^{\overline{\eta}_2+\eta_2'} \mathrm{d}z\, v \approx -\frac{\overline{v'b'}|_{\overline{\eta}_2}}{\partial_z \overline{b}|_{\overline{\eta}_2}}. \tag{2.92}$$

The η_1 integral is treated similarly. Therefore, we can write the zonal average, thickness integrated meridional transport between η_1 and η_2 as :

$$\overline{T} = \frac{1}{L_x} \int \mathrm{d}x \int_{\eta_1}^{\eta_2} \mathrm{d}z\, v \approx \frac{1}{L_x} \int \mathrm{d}x \int_{\overline{\eta}_1}^{\overline{\eta}_2} \mathrm{d}z\, v + \frac{\overline{v'b'}|_{\overline{\eta}_1}}{\partial_z \overline{b}|_{\overline{\eta}_1}} - \frac{\overline{v'b'}|_{\overline{\eta}_2}}{\partial_z \overline{b}|_{\overline{\eta}_2}}. \tag{2.93}$$

The zonal average operator commutes with the vertical integral if the extremes of integration do not depend on x: therefore, we can rewrite the first term on the left hand side as:

$$\frac{1}{L_x} \int \mathrm{d}x \int_{\eta_1}^{\eta_2} \mathrm{d}z\, v \approx \int_{\overline{\eta}_1}^{\overline{\eta}_2} \mathrm{d}z\, \overline{v} + \frac{\overline{v'b'}|_{\overline{\eta}_1}}{\partial_z \overline{b}|_{\overline{\eta}_1}} - \frac{\overline{v'b'}|_{\overline{\eta}_2}}{\partial_z \overline{b}|_{\overline{\eta}_2}}. \tag{2.94}$$

Finally, we combine the second and third terms on the right hand side as follows:

$$\frac{1}{L_x} \int \mathrm{d}x \int_{\eta_1}^{\eta_2} \mathrm{d}z\, v \approx \int_{\overline{\eta}_1}^{\overline{\eta}_2} \mathrm{d}z\, \overline{v} - \int_{\overline{\eta}_1}^{\overline{\eta}_2} \mathrm{d}z\, \partial_z \left(\frac{\overline{v'b'}}{\partial_z \overline{b}} \right). \tag{2.95}$$

The right hand side is the vertical integral of the residual velocity:

$$\frac{1}{L_x} \int \mathrm{d}x \int_{\eta_1}^{\eta_2} \mathrm{d}z\, v \approx \int_{\overline{\eta}_1}^{\overline{\eta}_2} \mathrm{d}z\, v_{\mathrm{res}}. \tag{2.96}$$

Next, we let $\eta_1 = -H$ (more formally, we set $b_1 = 0$ and use representation (2.78) of the vertical integrals to extend the lower limit of integration to $-H$), obtaining:

$$-\frac{1}{L_x} \int \mathrm{d}x \int_{-H}^{\eta_2} \mathrm{d}z\, v \approx -\int_{-H}^{\overline{\eta}_2} \mathrm{d}z\, v_{\mathrm{res}}. \tag{2.97}$$

By applying the definition of the residual and isopycnal streamfunctions we obtain the desired relation, equation (2.79).

Summary

We have proved equation (2.79), which establishes a leading order correspondence between the residual streamfunction in height coordinates and the isopycnal streamfunction, in buoyancy coordinates. The small perturbation parameter whose higher order terms are neglected in the equality represents departures from zonal mean. Equation (2.79) is used in the modelling practice as follows: the value of the residual streamfunction at a given depth z_1 is well approximated by the value of the isopycnal streamfunction at buoyancy b_1, where b_1 is such that the zonal average $\overline{\eta}_1$ of the associated isopycnal curve satisfies $\overline{\eta}_1 = z_1$. Normally, though, when the z_i's are a numerical model's vertical grid levels, the buoyancy grid levels at which the isopycnal streamfunction can be diagnosed do not meet this condition. Hence, at least an interpolation is usually necessary to compute ψ_{res}. We discuss this issue further for the specific case of the MITgcm in section 3.3.9.

2.5 Dynamical models of the Southern Ocean circulation

2.5.1 Zonal balance

Wind stress imparts momentum at the surface of the Southern Ocean. Therefore, in order for a statistically equilibrated state to be attained, there must exist one or more physical mechanisms to remove momentum from the system. For the sake of simplicity, we will study the zonal momentum balance in the simplified setup of a zonally re-entrant channel with flat bottom topography, zonally symmetric wind stress forcing at the surface, and linear bottom drag (see chapter 3 for a discussion of these assumptions). Note that the periodic zonal boundary condition is a closed boundary condition, i.e. there can be no net gain or loss of zonal momentum across it. Also, lateral Reynolds stresses are ineffective at transporting zonal momentum away across meridional boundaries (Olbers et al., 2004). Thus, at leading order zonal momentum must be transported downward and dissipated. In this section, we briefly discuss how the zonal momentum balance is maintained. We refer the reader to Abernathey et al. (2011), Cessi et al. (2006), Olbers et al. (2004), Rintoul et al. (2001), Vallis (2017) for a complete discussion. We start from the zonal momentum equation:

$$\partial_t u + \nabla \cdot (\mathbf{v}u) - fv = -\partial_x \phi + F_x, \tag{2.98}$$

where $F_x = \partial_z \tau_x$ is the kinematic stress. The steady-state, zonal average, zonal momentum equation is then:

$$\nabla \cdot \overline{\mathbf{v}u} - f\overline{v} = -\overline{\partial_x \phi} + \partial_z \overline{\tau_x}. \tag{2.99}$$

The first term on the right hand side is zero because of the periodic boundary conditions along the zonal direction (this result depends on the flat bottom assumption. We shall relax this hypothesis in the next section). The first term on the left hand side can be rewritten as:

$$\nabla \cdot \overline{\mathbf{v}u} = \partial_x \overline{uu} + \partial_y \overline{vu} + \partial_z \overline{wu}. \tag{2.100}$$

However, $\partial_x \overline{uu}$ is zero, again because of the periodic boundary conditions. Therefore, equation (2.99) reads:

$$-f\overline{v} = \partial_z \overline{\tau_x} - \partial_y \overline{vu} - \partial_z \overline{wu}. \tag{2.101}$$

By vertically integrating over the water column we obtain:

$$-f\int_{-H}^{0} \mathrm{d}z\,\overline{v} = \int_{-H}^{0} \mathrm{d}z\,\partial_z \overline{\tau_x} - \int_{-H}^{0} \mathrm{d}z\,\partial_y \overline{vu} - \int_{-H}^{0} \mathrm{d}z\,\partial_z \overline{wu}. \tag{2.102}$$

The vertically integrated, zonal-mean meridional velocity must be zero due to conservation of mass (there can be no net meridional volume flow across any given latitude line). The third term on the right hand side is zero too due to no flow conditions at the horizontal boundaries (i.e. $w = 0$ at the surface and bottom). The terms that remain are thus:

$$\overline{\tau_x}(0) = \overline{\tau_x}(-H) - \int_{-H}^{0} \mathrm{d}z\,\partial_y \overline{vu}. \tag{2.103}$$

This is the vertically integrated zonal momentum balance. It can be simplified further by noting that the meridional term $\partial_y \overline{vu}$ is negligible in the Southern Ocean

(Cessi et al., 2006, Olbers et al., 2004, Rintoul et al., 2001, Vallis, 2017). Therefore:

$$\overline{\tau_x}(0) = \overline{\tau_x}(-H). \tag{2.104}$$

The kinematic stress at the surface is the zonal wind stress divided by ρ_0, while $\overline{\tau_x}(-H) = r\overline{u}_b$, where r is the linear bottom drag parameter and u_b is the bottom zonal velocity. Thus:

$$\frac{\tau_w}{\rho_0} = r\overline{u}_b. \tag{2.105}$$

Therefore, the zonal momentum imparted at the surface is dissipated by bottom drag. In the real Southern Ocean, or in numerical models with bottom topography, the dominant term on the right hand side is instead the topographic form stress, which we discuss in the next section. It follows that models with flat bottom topography require a very large bottom zonal flow to compensate the gain of momentum at the surface:

$$\overline{u}_b = \frac{\tau_w}{\rho_0 r}. \tag{2.106}$$

Since wind stress at the the surface is balanced by linear drag at the bottom, there must be a mechanism of downward transport of zonal momentum. This mechanism is supported by geostrophic eddies and is called interfacial form stress.

2.5.2 Interfacial form stress

We explore how zonal momentum is transported from the surface to the bottom by baroclinic eddies. Please refer to Olbers et al. (2004) for a wonderful and complete presentation. Ward and Hogg (2011) also offers a lucid explanation of the subject, with a special focus on the spin-up of interfacial form stress from rest in an isopycnal ocean model.

Consider two isopycnal surfaces $\eta_1(x)$ and $\eta_2(x)$, and assume for the moment that they do not intersect the surface or the bottom. Let F_p be the vertically integrated contribution of the zonal pressure force to the rate of change of zonal momentum between the two isopycnals:

$$F_p = -\int_{\eta_1(x)}^{\eta_2(x)} \mathrm{d}z\, \partial_x \phi. \tag{2.107}$$

The right hand side can be transformed by means of Leibniz integral rule:

$$\partial_x \left(\int_{\eta_1(x)}^{\eta_2(x)} \mathrm{d}z\, \phi \right) = \int_{\eta_1(x)}^{\eta_2(x)} \mathrm{d}z\, \partial_x \phi + \phi(x, \eta_2(x)) \partial_x \eta_2 - \phi(x, \eta_1(x)) \partial_x \eta_1(x), \tag{2.108}$$

so that:

$$F_p = -\partial_x \left(\int_{\eta_1(x)}^{\eta_2(x)} \mathrm{d}z\, \phi \right) - \phi(x, \eta_2(x)) \partial_x \eta_2 + \phi(x, \eta_1(x)) \partial_x \eta_1(x). \tag{2.109}$$

Now we take a zonal average of both sides (this could actually be replaced by the average along any closed circumpolar path). The first term on the right hand side vanishes, and the layer-averaged, zonal-average zonal pressure force $\overline{F}_p$ is:

$$\overline{F}_p = -\frac{1}{L_x} \int \mathrm{d}x\, \phi(x, \eta_2(x)) \partial_x \eta_2 + \frac{1}{L_x} \int \mathrm{d}x\, \phi(x, \eta_1(x)) \partial_x \eta_1(x), \tag{2.110}$$

which we rewrite as:

$$\overline{F}_p = -\overline{\phi \partial_x \eta_2}^{\eta_2} + \overline{\phi \partial_x \eta_1}^{\eta_1}. \tag{2.111}$$

Here, the bar with η denotes zonal average evaluated along the isopycnal contour. By effecting an integration by parts, we can write equivalently:

$$\overline{F}_p = \overline{\partial_x \phi \eta_2}^{\eta_2} - \overline{\partial_x \phi \eta_1}^{\eta_1}. \tag{2.112}$$

Consider for example the second form of the pressure force term, equation (2.112). It means that the layer of fluid enclosed by the isopycnals η_1 and η_2 gains momentum $\tau_2 = \overline{\partial_x \phi \eta_2}^{\eta_2}$ from the layer above, and loses momentum $\tau_1 = \overline{\partial_x \phi \eta_1}^{\eta_1}$ to the one below. The term:

$$\tau_i = \overline{\partial_x \phi \eta_i}^{\eta_i} \tag{2.113}$$

is the interfacial form stress at the interface η_i. Note that $\tau_i = \overline{\partial_x \phi' \eta_i'}^{\eta_i}$, where $\phi(x, \eta(x))' = \phi(x, \eta(x)) - \overline{\phi}(\overline{\eta})$ and $\eta' = \eta - \overline{\eta}$. Thus, the interfacial form stress can be non-zero only if all the following conditions are met: (i) the zonal profile of the isopycnal is distorted by eddies, (ii) there are fluctuations in the zonal pressure gradient (also due to eddies), and (iii) isopycnal and pressure fluctuations are correlated. It is remarkable that despite the fact that zonal gradients of pressure can only act horizontally, they are effective at transporting momentum downward by transferring it across isopycnal surfaces (Olbers et al., 2004).

The downward transfer of zonal momentum is related to the meridional eddy buoyancy flux (Olbers et al., 2004). To see why, consider that $\partial_x \phi' \approx v'_g$ by geostrophy and that $\eta'_i \approx -b'/\partial_z \overline{b}|_{\overline{\eta}_i}$ by equation (2.86). Thus:

$$\tau_i = \overline{\partial_x \phi' \eta'_i}^{\eta_i} \approx \frac{\overline{v'_g b'}}{\partial_z \overline{b}}|_{\overline{\eta}_i}. \tag{2.114}$$

The interface η_i delimiting a layer of fluid does not have to be an isopycnal line and can also be a solid interface, for example the bottom of the ocean. The associated interfacial form stress term takes then the name of bottom form stress τ_b:

$$\tau_b = \overline{\partial_x \phi \eta_b}^{\eta_b}, \tag{2.115}$$

where $\eta_b(x)$ is the bottom topography. To obtain a comprehensive picture of the downward transfer of momentum, we can imagine that the fluid is composed by N stacked shallow water layers (Vallis, 2017). If we assume steady flow (and take time average over a large interval of time), the layer-integrated, zonal average, zonal momentum equation for the surface layer is:

$$-\frac{f}{L}\int \mathrm{d}x \int_{\eta_{N-1}}^{\eta_s} \mathrm{d}z v = \frac{\tau_w}{\rho_0} - \tau_1. \tag{2.116}$$

There is no vertical gradient term because the horizontal velocity does not depend on depth within a shallow water layer. The equation for an interior layer is:

$$-\frac{f}{L}\int \mathrm{d}x \int_{\eta_i}^{\eta_{i+1}} \mathrm{d}z v = \tau_{i+1} - \tau_i, \tag{2.117}$$

and the equation for the bottom layer is:

$$-\frac{f}{L}\int \mathrm{d}x \int_{\eta_b}^{\eta_1} \mathrm{d}z v = \tau_1 - \tau_b - r u_b. \tag{2.118}$$

The vertically integrated zonal momentum balance is obtained by summing over the layers:

$$\frac{\tau_w}{\rho_0} = r u_b + \overline{\partial_x \phi \eta_b}^{\eta_b}. \tag{2.119}$$

If there is no bottom topography the second term on the right hand side is zero, and we recover equation (2.106).

2.5.3 Mechanical energy balance

We examine the mechanical energy budget in the context of an idealised Boussinesq re-entrant channel, so that the equations of motion are expressed by equations (2.4)-(2.7). The presentation is adapted from Cessi et al. (2006) and Abernathey et al. (2011), to which we refer the reader for further discussion. We take the dot product of the frictional momentum equations with the three-dimensional velocity $\mathbf{v}$, which gives:

$$\begin{aligned} &u\partial_t u + u(\mathbf{v}\cdot\nabla)u - fuv + u\partial_x\phi = u\frac{\tau_w}{\rho_0}\delta_s - ru^2\delta_b \\ &v\partial_t v + v(\mathbf{v}\cdot\nabla)v + fvu + v\partial_y\phi = -rv^2\delta_b \\ &w\partial_z\phi = bw\,, \end{aligned} \tag{2.120}$$

where δ_s and δ_b are Dirac's delta functions centred at the surface and bottom respectively. The Coriolis term drops out when effecting the sum, and we obtain:

$$\partial_t K + \mathbf{v}\cdot\nabla(K+\phi) = bw + u\frac{\tau_w}{\rho_0}H\delta_s - r\mathbf{u}^2\delta_b. \tag{2.121}$$

Here, K is the kinetic energy density (divided by ρ_0):

$$K = \frac{1}{2}(u^2+v^2). \tag{2.122}$$

The volume average of the energy density flux term is zero when computed over a closed volume with rigid (no flow) or periodic boundary conditions:

$$\frac{1}{V}\int_V \mathrm{d}V\ [\mathbf{v}\cdot\nabla(K+\phi)] = \frac{1}{V}\int_V \mathrm{d}V\ [\nabla\cdot\mathbf{v}(K+\phi)] = \frac{1}{V}\int_{\partial V}\mathrm{d}\mathbf{S}\cdot\mathbf{v}(K+\phi) = 0, \tag{2.123}$$

where we have used the divergence theorem for the last equality. Therefore:

$$\frac{\mathrm{d}}{\mathrm{d}t}\frac{1}{V}\int_V \mathrm{d}V\,K = \frac{1}{V}\int_V \mathrm{d}V\,bw + \frac{1}{V}\int_V \mathrm{d}V\left[\frac{\tau_w}{\rho_0}u\delta_s - r\mathbf{u}^2\delta_b\right]. \tag{2.124}$$

The first term on the right hand side represents the conversion between kinetic and available potential energy (Vallis, 2017). The second term represents frictional contributions to the total energy of the channel. We assume that the channel is in a statistically equilibrated state and take time average over a large time interval, so that we can discard time derivative terms. The mechanical energy balance becomes:

$$\frac{1}{L_yH}r\iint \mathrm{d}y\mathrm{d}z\,\overline{\mathbf{u}^2}\delta_b = \frac{1}{L_yH}\iint \mathrm{d}y\mathrm{d}z\,\overline{bw} + \frac{1}{L_yH}\iint \mathrm{d}y\mathrm{d}z\left[\frac{\tau_w}{\rho_0}\overline{u}_s\delta_s\right], \tag{2.125}$$

where bar denotes zonal and time average. This equation ignores viscous dissipation and side drag (Abernathey et al., 2011). The APE conversion term is also negligible (Cessi et al., 2006), and the dominant mechanical balance is:

$$r \int \mathrm{d}y \overline{\mathbf{u}_b^2} = \int \mathrm{d}y \frac{\tau_w}{\rho_0} \overline{u}_s, \tag{2.126}$$

where the vertical integrals have been eliminated by the delta functions. To tidy up notation, we write:

$$r\langle \mathbf{u}_b^2 \rangle = \langle \frac{\tau_w}{\rho_0} u_s \rangle, \tag{2.127}$$

where angular brackets denote time, zonal, and meridional average. Given that τ_w and ρ_0 do not depend on time and x (under the assumption of constant, zonally symmetric wind forcing at the surface), we can rewrite the right hand side as:

$$r\langle \mathbf{u}_b^2 \rangle = \langle \frac{\tau_w}{\rho_0} \overline{u}_s \rangle, \tag{2.128}$$

which is equation 24 in Abernathey et al. (2011). This is the mechanical balance equation: wind work input at the surface is dissipated primarily by bottom drag. We can transform equation (2.128) into an eddy energy equation by writing:

$$\begin{aligned} \langle \mathbf{u}_b^2 \rangle &= \langle \overline{\mathbf{u}}_b^2 \rangle + \langle \mathbf{u}'^2_b \rangle \\ &= \langle \overline{u}_b^2 \rangle + \langle \overline{v}_b^2 \rangle + \langle \mathbf{u}'^2_b \rangle \\ &\approx \langle \overline{u}_b^2 \rangle + \langle \mathbf{u}'^2_b \rangle, \end{aligned} \tag{2.129}$$

where the $\langle \overline{v}_b^2 \rangle$ term is negligible (Cessi et al., 2006). We obtain:

$$r\langle \mathbf{u}'^2_b \rangle = \langle \frac{\tau_w}{\rho_0} \overline{u}_s \rangle - r\langle \overline{u}_b^2 \rangle, \tag{2.130}$$

An approximate expression for the zonal-mean, time-mean bottom zonal velocity was derived in the previous sections, equation (2.106), therefore:

$$r\langle \mathbf{u}'^2_b \rangle = \langle \frac{\tau_w}{\rho_0} (\overline{u}_s - \overline{u}_b) \rangle. \tag{2.131}$$

The right hand side is the useful wind work, i.e. the amount of wind-supplied energy that survives bottom drag and is actually available to drive baroclinic eddies (Sinha and Abernathey, 2016). Equation (2.131) implies that the large bottom flow due to the absence of bottom topography is unimportant for the eddy energy cycle because the right hand side depends on the baroclinic difference $\overline{u}_s - \overline{u}_b$, and not on the surface velocity only. Note that $\overline{u}_s - \overline{u}_b$ is related to the large scale gradient of buoyancy via the thermal wind equation, hence we expect $\overline{u}_s - \overline{u}_b > 0$.

2.5.4 Zonal balance and Eulerian streamfunction

Wind stress at the surface is related to the Eulerian streamfunction defined in equation (2.43) via the zonal momentum balance. We demonstrate this relationship in the simplified setup of the previous sections. Consider the steady-state zonal balance equation (2.101):

$$-f\overline{v} = \partial_z \overline{\tau_x}. \tag{2.132}$$

where we have ignored Reynolds stress terms (Marshall and Radko, 2003). By integrating vertically from the surface to depth z in the interior we have:

$$-f\int_z^0 \mathrm{d}z'\overline{v}(z') = \overline{\tau_x}(0) - \overline{\tau_x}(z). \tag{2.133}$$

The kinematic stress is non-zero in the surface and bottom Ekman layers only, thus $\overline{\tau_x}(z) = 0$. Hence:

$$-f\int_z^0 \mathrm{d}z'\overline{v}(z') = \frac{\tau_w}{\rho_0}, \tag{2.134}$$

which can also be written as:

$$\int_z^0 \mathrm{d}z'\overline{v}(z') = -\frac{\tau_w}{\rho_0 f}. \tag{2.135}$$

Now, by mass conservation the left hand side integral is equal to $-\int_{-H}^{z} \mathrm{d}z'\overline{v}(z')$, and by applying definition (2.43) one obtains:

$$\overline{\psi} = -\frac{\tau_w}{\rho_0 f}. \tag{2.136}$$

The Eulerian streamfunction depends linearly on wind stress applied at the surface. In the interior, $\overline{\psi}$ does not depend on depth, thus Eulerian streamlines are vertical. The Eulerian circulation closes in the surface and bottom Ekman layers, where Ekman stresses support a non-zero ageostrophic component of velocity. In the Southern Ocean the Coriolis parameter is negative and therefore the Eulerian streamfunction is positive: the Eulerian circulation is thus characterised by upwelling in the southern parts of the domain, downwelling in the north, equatorward flow at the surface, and poleward bottom return flow. This circulation pattern is the Deacon cell, and is discussed further in section 3.3.8.

2.5.5 TEM and the Southern Ocean

Transformed Eulerian Mean theory can be used in conjunction with the zonal balance equation to study how wind stress and eddies compete to maintain the meridional circulation of the Southern Ocean. The physical problem is as follows: the westerly winds drive the zonal flow of the ACC, and induce northward Ekman transport in the channel. The associated wind-induced circulation (represented by the Eulerian streamfunction $\overline{\psi}$) acts to steepen isopycnals, thereby increasing the available potential energy stored in the fluid. The excess energy is transformed into turbulent motion via baroclinic instability, and the associated eddy-induced circulation (represented by the quasi-Stokes streamfunction ψ^*) acts to flatten isopycnals. The two circulations are in near balance, and it is their sum (the residual circulation represented by the streamfunction ψ_{res}) that on average advects tracers in the Southern Ocean. But how exactly does this balance set the observed structure and magnitude of the meridional overturning circulation, stratification, and zonal flow? Here, we briefly review the model proposed in the seminal work of Marshall and Radko (2003) (MR03). Since then, the argument has been refined to accommodate for more realistic boundary conditions (e.g. Marshall and Radko (2006)), but the MR03 model already elucidates the essential facts with minimal ingredients. Firstly, we gather a few useful relations we have studied in the preceding sections:

(i) $\psi_{\text{res}} = \overline{\psi} + \psi^*$.

The definition of residual streamfunction. It highlights the fact that the streamfunction that advects tracers in the SO is the result of a balance between the wind-induced and eddy-induced circulations.

(ii) $v_{\text{res}} = -\partial_z \psi_{\text{res}}$.

The relationship between the meridional residual velocity and the residual streamfunction.

(iii) $\overline{\psi} = -\tau_w / \rho_0 f$.

The relationship between the Eulerian streamfunction and the wind stress, obtained (with assumptions consistent with those below) by studying the zonal balance equation.

(iv) $\partial_t \overline{b} + (\mathbf{v}_{\text{res}} \cdot \nabla)\overline{b} = -\nabla \cdot \mathscr{F}_{\text{res}}\{b\} + \overline{\mathscr{B}}$.

The buoyancy equation in residual form. By construction of ψ^*, the right hand side is approximately zero in the interior. It can be non-zero in the mixed layer.

(v) $s = -\partial_y \overline{b} / \partial_z \overline{b}$.

The definition of isopycnal slope. s is negative in the Southern Ocean.

Next, we introduce two fundamental assumptions:

(i) The ocean can be divided into an adiabatic interior layer and a diabatic layer. This is a standard assumption in TEM theory (see section 2.4.3).

(ii) $\psi^* = -ks^2$

This is a closure relation for the eddy-induced circulation. k is a positive constant and s is the slope of the isopycnals.

The remaining assumptions are needed to characterise the diabatic layer and to clarify how we treat boundaries. They are less fundamental in nature, and indeed Marshall and Radko (2006) part with some of them:

(iii) The diabatic layer is vertically homogeneous (i.e., buoyancy does not depend on z. This means that the diabatic layer is well mixed) and has fixed depth. Also, there is no seasonal cycle.

(iv) Buoyancy is prescribed in the diabatic layer: $b_0 = b_0(y)$.

(v) The vertical integral over the diabatic layer of the right-hand side of relation (iv) is prescribed: $B_0 = B_0(y)$. This amounts to prescribing the combined effect of air-sea fluxes and lateral eddy fluxes in the diabatic layer.

(vi) Wind forcing at the surface is constant and zonally symmetric.

These are the basic ingredients. How are they put to use? The idea at the core of the MR03 model is that, since $(\mathbf{v}_{\mathrm{res}} \cdot \nabla)\overline{b} = 0$ in the steady state interior, there exists a functional relationship between ψ_{res} and $\overline{b}$. In physical terms, this amounts to saying that the residual velocity is everywhere tangent to isopycnal lines, implying that if one follows a line of constant $\overline{b}$, then ψ_{res} is constant along that line too. Mathematically, it means that one can seek a functional relationship $\psi_{\mathrm{res}} = \psi_{\mathrm{res}}(\overline{b})$. The second key point is to note that the relationship $\psi_{\mathrm{res}} = \psi_{\mathrm{res}}(\overline{b})$ must be set by the boundary conditions, in this case at the base of the diabatic layer. To see this, we take the vertical integral of the buoyancy equation in residual form over the diabatic layer and use assumption (v) to take care of the right hand side. We obtain:

$$\int_{-h}^{0} \mathrm{d}z\,(\mathbf{v}_{\mathrm{res}} \cdot \nabla)\overline{b} = B_0(y), \tag{2.137}$$

where h is the constant depth of the diabatic layer. On the left hand side, there is no time derivative because we suppose that the system is in a statistically equilibrated state (the bar denotes time and zonal average). Furthermore, the term with $\partial_z\overline{b}$ vanishes due to assumption (iii), and by virtue of the same hypothesis $\partial_y\overline{b}$ can be moved outside the vertical integral. On the right hand side, $B_0 = B_0(y)$ is prescribed. Therefore:

$$\partial_y b_0(y) \int_{-h}^{0} \mathrm{d}z\, v_{\mathrm{res}} = B_0(y), \tag{2.138}$$

where $b_0(y)$ is the prescribed buoyancy profile in the diabatic layer, assumption (iv). The vertical integral of the meridional residual velocity over the diabatic layer is minus the residual streamfunction evaluated at the base and at the top of the diabatic layer, equation (ii). But the top of the diabatic layer is the surface, where $\psi_{\mathrm{res}} = 0$ by definition. Hence:

$$\psi_{\mathrm{res}}|_{z=-h}\partial_y b_0(y) = B_0(y). \tag{2.139}$$

This equation sets the functional relationship between the residual streamfunction and buoyancy: the final step consists in exploiting it to compute $\overline{b}$ and ψ_{res}. We substitute assumption (ii) and equation (iii) into equation (i):

$$\psi_{\mathrm{res}} = -\frac{\tau_w}{\rho_0 f} - ks^2, \tag{2.140}$$

or, equivalently:

$$s = -\sqrt{-\frac{\tau_w}{\rho_0 f} - \frac{\psi_{\text{res}}}{k}}. \qquad (2.141)$$

Next, we substitute the definition of s, equation (v), yielding:

$$\partial_y \overline{b} - \sqrt{-\frac{\tau_w}{\rho_0 f} - \frac{\psi_{\text{res}}}{k}}\,\partial_z \overline{b} = 0. \qquad (2.142)$$

This is a partial differential equation for the zonal-mean buoyancy field $\overline{b}(y,z)$. ψ_{res} is a function of $\overline{b}$ in the interior and the wind stress only depends on y thanks to assumption (vi), therefore the equation is of the form $\partial_y \overline{b} + c(y, \overline{b})\partial_z \overline{b}$. Thus, it can be solved with the method of characteristics (Leveque, 2002) (once $\overline{b}(y,z)$ is computed, one obtains the residual streamfunction by exploiting the fact that ψ_{res} is constant along isopycnals). The precise form of the solution depends on the specific choice made for the prescribed variables $b_0(y)$, $B_0(y)$, and $\tau_w(y)$, but the remarkable result is that the set of idealised assumptions above produce plausible predictions for the thermocline depth, stratification, and for the structure and magnitude of the meridional overturning circulation (we refer to the paper for details). The MR03 model is a milestone in the current understanding of the dynamics of the Southern Ocean, and demonstrates how TEM theory elucidates the key role played by baroclinic eddies in its dynamical balance.

Chapter 3

The MITgcm in idealised channel configuration

3.1 Introduction

The results presented in this Thesis hinge on numerical simulations of the Southern Ocean performed with a particular configuration of a general circulation model, the MITgcm. The purpose of this chapter is to detail the modelling choices inherent to the GCM configuration, illustrate its salient physical properties, and assess what its advantages and disadvantages are relative to the general aims of this manuscript. No prior familiarity with GCMs is assumed. The model configuration is inherited from and very similar to that of Abernathey et al. (2011) and Hill et al. (2012), and readers are referred to these studies for a summarised presentation.

The configuration I employ is that of an idealised, re-entrant channel. I will henceforth refer to it as ”the idealised channel”. The *idealised* part is motivated by the simplifying assumptions involved in the setup, which I discuss in detail in the following sections. Amongst these, the channel geometry is zonally symmetric, there is no bottom topography and no continental shelf. The equation of state is linear with no salinity, and there is no sea ice. Constant, zonally symmetric wind stress and buoyancy fluxes are imposed at the surface, and diabatic processes that take place outside of the Southern Ocean are represented with a sponge layer located at the northern boundary. A number of physical processes that are known to be important for the dynamics of the Southern Ocean cannot be captured under these assumptions: the effect of topography on the upwelling branch of the MOC (Tamsitt et al., 2017), or the presence of a region of temperature inversion due to sea ice (Ferreira et al., 2015) are examples. In view of this, I do not attempt to use the results obtained with this configuration to make quantitative predictions about the real Southern Ocean (although, when possible, I will check that the model sits in a regime similar to that of observational and reanalysis products). What are then the advantages of this setup? Firstly, it can be run economically at a high horizontal resolution of 5 km, which means that baroclinic eddies are resolved and there is no need to employ an eddy parametrisation scheme. The ability to resolve baroclinic eddies has been shown to be crucial for the accurate representation of the dynamics of the Southern Ocean by a plethora of studies (e.g. Hallberg and Gnanadesikan (2006), Screen et al. (2009), Viebahn and Eden (2010)). Secondly, although some important physical mechanisms are not represented in the model, the effect of those that *are* included can emerge more clearly, complicating factors being reduced to a minimum. In the spirit of Abernathey et al. (2011), thus, rather than to make quantitative predictions my goal is to gain qualitative insights on the nature of the processes governing the physics of the Southern Ocean.

This chapter is organised as follows: in section 3.2 I give an overview of the idealised channel configuration, including equations, domain, gridding choices, forcing, boundary conditions, and numerical algorithms. Section 3.3 describes the physical properties of the idealised channel, with a particular emphasis placed on linking the structure of zonal-average temperature profiles on the meridional plane to the zonal and meridional circulation. I conclude in section 3.4.

3.2 Setting up the model

By general circulation model we mean a numerical model that simulates the state of a geophysical system. This could be the atmosphere, the ocean, or both. Here, we use a general circulation model to simulate the state of the Southern Ocean: the numerical core is the MITgcm (Marshall et al., 1997a,b).

3.2.1 Dynamical equations

The model solves the Boussinesq equations of motion in a β-plane, equations (2.4)-(2.7), with no salinity and a linear equation of state, equation (2.9). The density ρ is a function of temperature only, therefore lines of constant density (isopycnals) coincide with lines of constant temperature (isothermals): unless otherwise stated, the two terms will be used interchangeably.

3.2.2 Domain and grid

The domain is a re-entrant rectangular channel with no bottom topography and no continental shelf. The channel is $L_x = 1000$ km wide in the zonal direction, $L_y = 2000$ km wide in the meridional direction, and $H = 2985$ m deep. Periodic boundary conditions are applied in the zonal direction (which makes the channel re-entrant), and no-slip boundary conditions at the meridional boundaries. The horizontal resolution is $\Delta x = \Delta y = 5$ km, and there are 30 unevenly spaced vertical levels. The thickness of the vertical levels ranges from 10 m at the surface to 280 m at the bottom. Note that the zonal extent of the real Southern Ocean is about a factor 25 larger than in the model (the length of a latitudinal circle at 45°S is of about 28000 km): therefore, the streamfunctions defined on the meridional plane should be scaled by the same factor to compare with observational estimates of the meridional volume transport. The absence of bottom topography and continental shelf

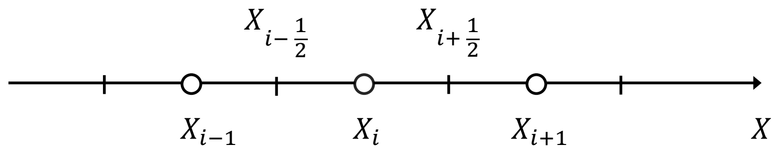

Figure 3.1: Schematic of the discrete grid for the 1-dimensional problem: the i-th cell is the interval $(x_{i-\frac{1}{2}}, x_{i+\frac{1}{2}})$, and $x_{i-\frac{1}{2}}$ and $x_{i+\frac{1}{2}}$ are called the cell edges. The cell center is x_i.

constitute idealising assumptions that favour simplicity over realism: some of their implications are discussed in section 3.3.3.

3.2.3 Spatial discretisation and the finite volume method

The model uses the finite volume method to integrate the equations of motion numerically. A basic familiarity with the finite volume method is therefore essential in order to understand how the variables are discretised, and to interpreter the output of the numerical simulations. To introduce the subject, we follow an example from Leveque (2002), to which the reader is referred for a comprehensive discussion.

Consider the 1-dimensional conservation law:

$$\frac{\mathrm{d}}{\mathrm{d}t}\int_a^b \mathrm{d}x q(t,x) = f(q(t,a)) - f(q(t,b)), \tag{3.1}$$

where q is a passive tracer and f is its flux function. The goal of the finite volume method is to divide the spatial domain into intervals, which are called cells, and to accurately approximate the time evolution of the spatial average of q over cells. In this one-dimensional example, the i-th cell is the interval $(x_{i-\frac{1}{2}}, x_{i+\frac{1}{2}})$, and $x_{i-\frac{1}{2}}$ and $x_{i+\frac{1}{2}}$ are called the cell edges. The conservation law for an individual cell reads:

$$\frac{\mathrm{d}}{\mathrm{d}t}\int_{x_{i-\frac{1}{2}}}^{x_{i+\frac{1}{2}}} \mathrm{d}x q(t,x) = f(q(t,x_{i-\frac{1}{2}})) - f(q(t,x_{i+\frac{1}{2}})), \tag{3.2}$$

where the right hand side is the flux of q across the cell edges. Integration in time from time step t_n to time step t_{n+1} yields:

$$\begin{aligned}\int_{x_{i-\frac{1}{2}}}^{x_{i+\frac{1}{2}}} \mathrm{d}x q(t_{n+1},x) = \int_{x_{i-\frac{1}{2}}}^{x_{i+\frac{1}{2}}} \mathrm{d}x q(t_n,x) + \\ - \left[\int_{t_n}^{t_{n+1}} \mathrm{d}t f(q(t,x_{1+\frac{1}{2}})) - \int_{t_n}^{t_{n+1}} \mathrm{d}t f(q(t,x_{1-\frac{1}{2}})) \right].\end{aligned} \tag{3.3}$$

To simplify the equation, we denote with Q_i^n the spatial average of q over the i-th cell at time t_n:

$$Q_i^n = \frac{1}{\Delta x} \int_{x_{i-\frac{1}{2}}}^{x_{i+\frac{1}{2}}} \mathrm{d}x q(t_n,x), \tag{3.4}$$

where $\Delta x = x_{i+\frac{1}{2}} - x_{i-\frac{1}{2}}$. Also, we denote with $F^n_{i-\frac{1}{2}}$ the time-averaged flux at the cell edge $x_{i-\frac{1}{2}}$:

$$F^n_{i-\frac{1}{2}} = \frac{1}{\Delta t} \int_{t_n}^{t_{n+1}} \mathrm{d}t f(q(t,x_{i-\frac{1}{2}})), \tag{3.5}$$

where $\Delta t = t_{n+1} - t_n$. Then, the approximate conservation equation reads:

$$Q_i^{n+1} = Q_i^n - \frac{\Delta t}{\Delta x}(F^n_{i+\frac{1}{2}} - F^n_{i-\frac{1}{2}}). \tag{3.6}$$

This is the skeleton equation for the finite volume method. The numerical method itself is defined by the specific choice of the flux, which we do not concern ourselves with here. Numerous alternatives for the treatment of the time and spatial stepping exist, and the MITgcm user guide (Adcroft et al., 2022) provides a detailed account of the options supported by the model. Our configuration is defined on a more complicated domain than the simple 1-dimensional example above, but similar ideas apply: the user guide illustrates how the spatial discretisation is generalised to the three-dimensional case. The key point is that tracers (e.g., temperature) are represented by averages over spatial cells, and are located at cell centers. Flux variables (for example, streamfunctions), on the other hand, are located at cell interfaces. When necessary, linear interpolations will be effected to combine variables defined at different spatial locations.

3.2.4 Tracer advection, time stepping, and mixed layer scheme

The model uses the Second Order Momentum scheme of Prather (Prather, 1986) with flux limiter for temperature advection. Hill et al. (2012) showed that this scheme, with no horizontal and vertical diffusivity (i.e., $k_v = k_h = 0$ m^2 s^{-1}), produces values of numerical diapycnal mixing below 10^{-5} m^2 s^{-1} for the same MITgcm configuration considered here, consistently with observational estimates. The implication is that the model respects the quasi-adiabatic nature of tracer advection in the interior. The horizontal and vertical eddy viscosity coefficients and the horizontal eddy hyperviscosity coefficient parametrise the effect of turbulence on sub-grid scale, and are set to $A_h = 12.0$ m^2 s^{-1}, $A_v = 3.0 \cdot 10^{-4}$ m^2 s^{-1}, and $A_4 = 9 \cdot 10^8$ m^4 s^{-1} respectively, see also table 3.1. The time stepping algorithm is Adam-Bashforth with staggering in time, and the model time step is 900 s. The KPP mixed layer scheme of Large et al. (1994) is used, see Damerell et al. (2020) for a comparison with observational products. Model parameters are summarised in table 3.1.

3.2.5 Forcing and boundary conditions

We force the model mechanically, by wind stress applied at the surface, and thermodynamically, by buoyancy fluxes applied at the surface and at the northern boundary. Energy is removed mechanically through linear bottom drag.

The forcing is idealised: we impose constant, zonally symmetric surface wind stress according to the formula:

$$\tau(y) = \tau_0 \sin(\pi \frac{y}{L_y}), \tag{3.7}$$

so that the surface wind stress peaks in the centre of the domain ($y = 1000$ km), as represented schematically in figure 3.2. The reference value for τ_0 is 0.1 N m^{-2} (see table 3.1). The meridional wind stress profile is chosen to mimic the Southern Hemisphere jet stream, see figure 1 in Abernathey et al. (2011) for a comparison with a coupled ocean-atmosphere product, and figure 4 in Marshall and Speer (2012) for a comparison with a reanalysis product. The wind forcing does

Symbol	Value	Description
L_x, L_y	1000 km, 2000 km	Domain size
L_{sponge}	100 km	Sponge layer size
H	2985 m	Domain depth
ρ_0	999.8 kg m^{-3}	Reference density
α_T	2 $\cdot 10^{-4}$ K^{-1}	Thermal expansion coefficient
f_0	-1 $\cdot 10^{-4}$ s^{-1}	Reference Coriolis parameter
β	1 $\cdot 10^{-11}$ s^{-1} m^{-1}	Planetary vorticity gradient
Q_0	10 W m^{-2}	Surface heat flux magnitude
τ_0	0.1 N m^{-2}	Surface wind stress magnitude
r_b	1.1 $\cdot 10^{-3}$ m s^{-1}	Linear bottom drag parameter
τ_{sponge}	7 days	Sponge layer relaxation time scale
Δx, Δy	5 km	Horizontal grid spacing
Δ_z	7 days	Vertical grid spacing
κ_v	0 m^2 s^{-1}	Vertical diffusivity
κ_h	0 m^2 s^{-1}	Horizontal diffusivity
A_v	3.0 $\cdot 10^{-4}$ m^2 s^{-1}	Vertical viscosity
A_h	12.0 m^2 s^{-1}	Horizontal viscosity
A_4	9 $\cdot 10^8$ m^4 s^{-1}	Horizontal hyperviscosity

Table 3.1: Summary of model parameters

not include the seasonal cycle, zonal asymmetries, and fast time-scale variability, although it would be possible to include these features in an idealised setup: for example, Doddridge et al. (2019) employs a similar configuration of the MITgcm, but with more realistic winds.

We prescribe fixed buoyancy fluxes at the surface, according to:

$$Q(y) = \begin{cases} -Q_0 \cos\left(3\pi \frac{y}{L_y}\right) & y \leq \frac{5}{6} L_y \\ 0 & \text{else}\,. \end{cases} \tag{3.8}$$

The convention is that the flux is positive into the ocean. Q_0 is set to 10 W m^{-2}, see table 3.1, and the meridional profile of $Q(y)$ is represented schematically in figure 3.2. The value of Q_0 and the functional form of $Q(y)$ are intended to represent the surface buoyancy fluxes in the Southern Ocean, see figure 1 in Abernathey et al. (2011) and figure 4 in Marshall and Speer (2012): cooling near Antarctica is connected to the formation of dense Antarctic Bottom Water (AABW), while cooling north of the ACC with the formation of Antarctic Intermediate Water (AAIW) and Subantarctic Mode Water (SAMW).

Previous research indicates that channel models not including heat exchanges between the Southern Ocean and the other basins tend to return an excessively weak deep residual circulation (Cessi et al., 2006). In our model configuration, diabatic processes that take place equatorward of the Southern Ocean are represented by a sponge layer located in the northernmost 100 km of the channel. Within this layer, temperature is relaxed to a prescribed exponential vertical profile $T^\star(z)$:

$$T^*(z) = \Delta T (e^{z/h} - e^{-H/h})/(1 - e^{-H/h}), \tag{3.9}$$

where $h = 1000$ m represents the vertical decaying scale and $\Delta T = 8$ °C the surface to bottom temperature difference. The exponential vertical profile is chosen to mimic the observed stratification at the northern flank of the Southern Ocean, and is shown in figure 3.2. Temperature is relaxed to the $T^\star(z)$ profile with a relaxation coefficient that varies from $\tau = 0$ (no relaxation) at southern edge of the sponge layer ($y = 1900$ km) to $\tau = 7$ day^{-1} at $y = L_y$.

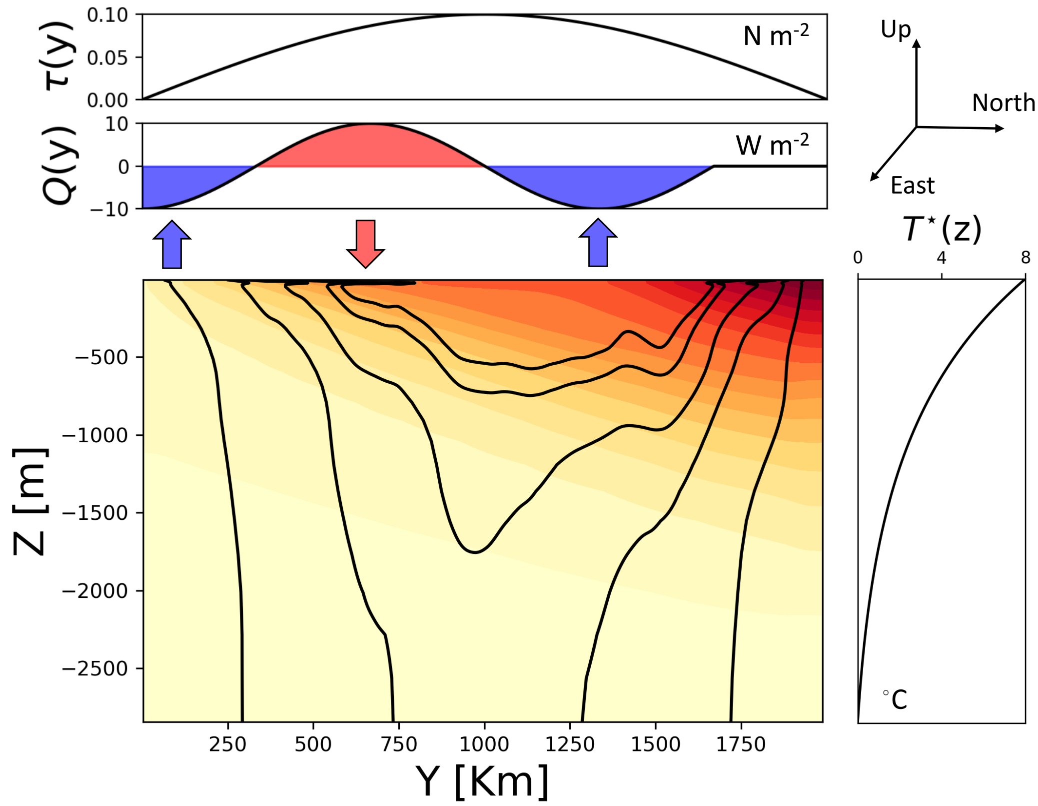

Figure 3.2: Schematic of mechanical and thermodynamical forcing applied at the model boundaries. The main panel shows time-mean, zonal-mean temperature (colours) and zonal velocity (contours). The time averages are computed over a period of 51 years. The top panels show the zonally-symmetric wind stress and buoyancy forcing profiles applied at the surface. The convention is that the heat flux is positive into the ocean. Blue (red) arrows represent cooling (warming) of the ocean due to the air-sea fluxes. The lateral panel shows the vertical temperature profile prescribed in the northern sponge layer.

Name	Unit	Description
UVEL	m s^{-1}	Zonal component of velocity
VVEL	m s^{-1}	Meridional component of velocity
THETA	°C	Potential temperature
MXLDEPTH	m	Depth of mixed layer
ETAN	m	Surface height anomaly
LaVH1TH	m^2 s^{-1}	Layer-integrated meridional transport
LaHs1TH	m	Layer thickness

Table 3.2: Summary of model diagnostics

3.2.6 Running the model

The model is run on the Imperial College High Performance Computing CX1 cluster. For the latest simulations, the horizontal model tile comprises $sNx \times sNy = 25 \times 40$ grid points, and there are $nPx = 8$ blocks along the x axis and $nPy = 10$ blocks along the y axis. The model is run on 3 nodes with 32 CPUs each. Less recent runs have slightly different data layout and nodes allocation to conform to the system's job sizing requirements, but are similar in computational demand. The model was spun up for approximately 200 years until it reached a statistically equilibrated state, as diagnosed from domain-averaged kinetic energy.

3.2.7 Model diagnostics

Table 3.2 contains a list of the model diagnostics used in the remainder of this chapter. Grid diagnostics have been omitted, and the interested reader is referred to the MITgcm user guide (Adcroft et al., 2022) for an illustration. Model diagnostics introduced in subsequent chapters will be presented separately on a case by case basis.

3.2.8 Computation of the streamfunctions

The computation of the streamfunctions describing the circulation on the meridional plane ($\overline{\psi}$, ψ_{res}, and ψ^*) is complicated by the gridding issues introduced by the model's spatial discretisation. In this section, we briefly document on how the discrete computations are carried out in practice.

Eulerian streamfunction

The Eulerian streamfunction is computed according to equation (2.43). The integration is effected using a discrete integration formula:

$$\overline{\psi}_{j+\frac{1}{2}} = -\sum_{k=j+1}^{N_z} \mathrm{drF}_j \overline{\mathrm{VVEL}}_j \qquad j = 1, \ldots, N_z - 1\,, \tag{3.10}$$

where the index $j = 1, \ldots, N_z$ labels vertical grid points (j increases from top to bottom), drF is the distance between vertical cell interfaces, and VVEL is the model diagnostic for the meridional component of velocity (see table 3.2). At the bottom (corresponding to the interface point with index $N_z + \frac{1}{2}$ on the z grid), we explicitly prescribe $\overline{\psi} = 0$.

Residual streamfunction

The residual streamfunction is the sum of the Eulerian and eddy-induced streamfunctions:

$$\psi_{\text{res}} = \overline{\psi} + \psi^*,$$

but this definition is rarely used in practice, because diagnosing the eddy streamfunction directly is complicated. Commonly, the computation of ψ_{res} is carried out by taking advantage of its first order equivalence with the thickness averaged meridional circulation. The key formula is equation (2.79), which we repeat for clarity:

$$\psi_{\text{res}}(\overline{\eta_1}) \approx \psi_I(T_1).$$

Here, $\overline{\eta_1}$ is the zonal average of the isopycnal $\eta_1(x)$, and $\psi_I(T_1)$ is given by:

$$\psi_I(T_1) = -\frac{1}{L_x}\int \mathrm{d}x \int_{-H}^{\eta_1} \mathrm{d}z\, v, \tag{3.11}$$

Formula (2.79) means that the residual streamfunction at depth z is well approximated by the value of ψ_I at the temperature T, where T is such that the zonal average depth of the associated isopycnal $\eta(x)$ is equal to z.

The steps to compute $\psi_{\text{res}}(z)$ are thus (i): compute $\psi_I(T)$ and (ii): interpolate $\psi_I(T)$ from temperature coordinates back to depth coordinates. In order to effect the first task, we diagnose the layer-integrated meridional transport LaVH1TH, where:

$$\text{LaVH1TH}_j \approx \int_{\eta_{j-\frac{1}{2}}}^{\eta_{j+\frac{1}{2}}} \mathrm{d}z\, v. \tag{3.12}$$

Here, the index j runs on temperature grid points (j increases from cold to warm): in our model configuration, there are 42 equally spaced temperature bins, with cell centers ranging from $T = -0.1$ °C to $T = 8.1$ °C and a temperature interval $\Delta T = 0.2$ °C. By writing:

$$\int_{-H}^{\eta_{j+\frac{1}{2}}} \mathrm{d}z\, v = \sum_{k=1}^{j} \int_{\eta_{k-\frac{1}{2}}}^{\eta_{k+\frac{1}{2}}} \mathrm{d}z\, v \approx \sum_{k=1}^{j} \text{LaVH1TH}_k, \tag{3.13}$$

where $\eta_{\frac{1}{2}} = -H$, we can compute the residual streamfunction in isopycnal coordinates as:

$$\psi_{I\,j+\frac{1}{2}} = -\sum_{k=1}^{j} \overline{\text{LaVH1TH}}_k. \tag{3.14}$$

As for the cold boundary condition, we explicitly set $\psi_{I\,\frac{1}{2}} = 0$.

In order to map ψ_I back to depth coordinates, we diagnose the layers' thickness LaHs1TH:

$$\text{LaHs1TH}_j \approx \int_{\eta_{j-\frac{1}{2}}}^{\eta_{j+\frac{1}{2}}} \mathrm{d}z = \eta_{j+\frac{1}{2}} - \eta_{j-\frac{1}{2}}. \tag{3.15}$$

By summing over the definition of LaHs1TH above, and using the fact that $\eta_{\frac{1}{2}} = -H$, we have:

$$\sum_{k=1}^{j} \text{LaHs1TH}_k \approx \eta_{j+\frac{1}{2}} + H. \tag{3.16}$$

Therefore, we can explicitly compute the zonal average depth of the isopycnal $\eta_{j+\frac{1}{2}}$ as:

$$\overline{\eta}_{j+\frac{1}{2}} = \sum_{k=1}^{j} \overline{\text{LaHs1TH}}_k - H. \tag{3.17}$$

By virtue of (2.79), we obtain:

$$\psi_{\text{res}}(\overline{\eta_{j+\frac{1}{2}}}) \approx \psi_I(T_{j+\frac{1}{2}}). \tag{3.18}$$

A single interpolation is needed to map ψ_{res} from the zonal average isopycnal depths $\{\overline{\eta}\}_{j+\frac{1}{2}}$ to the z-interfaces grid points $\{\text{RF}\}_{k+\frac{1}{2}}$.

It is possible to compute ψ_{res} even without diagnosing LaHs1TH, although this requires two interpolations rather than one. The first interpolation is needed to map the zonal average temperature $\overline{T}(z)$ from z-centers to z-interfaces. This corresponds to defining an approximate function $T = T(\text{RF})$. With a second interpolation, which corresponds to defining an approximate function $\psi_I = \psi_I(T)$, we compute:

$$\psi_{\text{res}}(\text{RF}_{k+\frac{1}{2}}) \approx \psi_I(T(RF_{k+\frac{1}{2}})). \tag{3.19}$$

The two methods give similar results, with the largest differences found in the diabatic layer close to the surface (not shown), where isothermals are nearly vertical and the double interpolation is not accurate. Unless otherwise stated, the residual streamfunction is computed according to the first method in the following.

3.3 Properties of the control run

We have discussed the technical details inherent to configuring and running the model. Next, we turn our attention to the physical picture of the Southern Ocean that it returns: we focus on those aspects that will recur more frequently in the subsequent chapters, and endeavour to highlight the connection between theoretical predictions and model output whenever feasible. We concentrate on the reference state of the idealised channel, which we call ”control run” and is defined by the choice of model parameters specified by table 3.1. A different state of the channel, corresponding to an ocean driven by stronger winds, will be investigated in chapter 6.

3.3.1 Geostrophic scaling

We test whether the idealised channel is in the low Rossby number regime. As discussed in section 2.2.2, $Ro << 1$ is one of the necessary conditions for geostrophic balance. The Rossby number is defined in equation (2.10), and for the idealised channel we have $U \approx 0.01 - 0.1$ m/s, $f_0 = -10^{-4}$ s^{-1}, and $L \approx 10^6$ m. Therefore, $Ro \approx 10^{-4} - 10^{-3}$, and the idealised channel is well into the low Rossby number regime. Anticipating on the fact that $L_d << L$ (as illustrated below), we can expect geostrophic balance and the thermal wind relations to hold in the channel. We will check that this is actually the case in section 3.3.6.

The first Rossby radius of deformation L_d is formally defined as the largest eigenvalue of the Sturm-Liouville problem (Chelton et al., 1998), and represents the typical horizontal length scale of geostrophic eddies (Williams et al., 2007). A simple estimate of its magnitude can be obtained via formula (2.11): for the idealised channel, $N_0 \approx 2.3 \cdot 10^{-3}$ s^{-1} and $H = 3000$ m. Therefore $L_d \approx 22$ km, which compares well with the $L_d \approx$ 10-25 km estimate of Chelton et al. (1998) for the real Southern Ocean. The horizontal grid size for the idealised channel is $\Delta x = \Delta y = 5$ km, so that $\Delta x < L_d$ by a factor 4 approximately. We thus expect that the horizontal surface corresponding to a typical geostrophic eddy is populated by at least $O(10)$ grid points, which confirms that geostrophic turbulence is resolved in the channel.

3.3.2 Zonal-mean temperature and mixed layer depth

We start our survey of this model's physical properties in earnest by looking at the structure of zonal-mean temperature. The time-averaged, zonally-averaged temperature is shown in figure 3.3 left, where the time average is computed over 51 years of simulation. For a comparison with observations, see for example figure 4.6.3 in Rintoul et al. (2001).

In the idealised channel, temperature ranges from a maximum of nearly 8 °C, at the surface and near the southern boundary of the domain, to 0 °C in the deep interior. It increases upwards and northwards everywhere: the meridional gradient is related to the vertical shear of zonal velocity via the thermal wind equation. Contrary to more complex MITgcm configurations including sea ice, for example Doddridge et al. (2019), Ferreira et al. (2015), there is no region of temperature inversion where water gets warmer with depth. Isopycnals are tilted with a negative slope: note however that the aspect ratio of figure 3.3 (left) greatly exaggerates its magnitude. In fact, isopycnals would look almost flat if plotted with equally scaled axes due to $L_y >> H$. At the northern boundary, the stratification is determined by the prescribed exponential profile $T^\star(z)$ described above. Isopycnals are not straight lines (see Marshall and Radko (2003, 2006) for an investigation of this idealised limit), therefore the stratification in the northern sponge layer is not linearly mapped in the meridional gradient at the surface.

The depth of the zonal-mean mixed layer, defined as the depth such that temperature is 0.8 °C less than at the surface, is also shown in figure 3.3 (left, dashed black line). Note that with this definition the mixed layer does not necessarily coincide with the surface diabatic layer or with the Ekman layer. The model is not forced by a seasonal cycle, hence there are no seasonal variations in the mixed layer depth. The mixed layer is a few hundred meters deep in all regions of the domain except near the southern boundary, where it attains a depth of over a thousand meters and

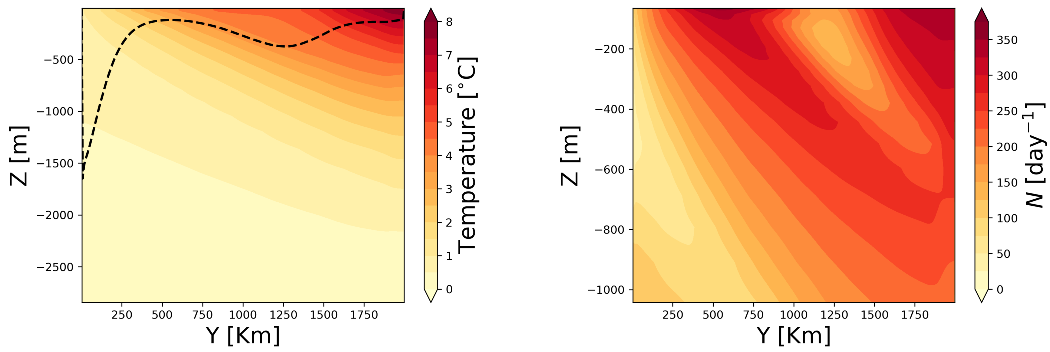

Figure 3.3: Left: time-mean, zonal-mean temperature (colours). The time mean is computed over a period of 51 years. The dashed black line marks the depth of the time-mean, zonal-mean mixed layer, defined as the depth such that temperature is 0.8 °C less than at the surface. Right: time-mean, zonal-mean buoyancy frequency for the upper 1000 m of the channel.

the stratification is weak. The stratification is computed as:

$$N^2 = g\alpha \frac{\mathrm{d}T}{\mathrm{d}z}. \tag{3.20}$$

The buoyancy frequency N is the square root of the stratification N^2, and is shown for the top 1000 meters in figure. 3.3 (right). As expected, N is larger near the surface and in the centre of the domain (where the mixed layer is shallower), and smaller at depth and near the southern boundary of the domain (where the mixed layer reaches further in depth). The minimum of N at $y = 1250$ km approximately is associated with subduction of water masses operated by the upper branch of the intermediate overturning cell (see section 3.3.9), and to the formation of Subantarctic Mode Water (SAMW) and Antarctic Intermediate Water (AAIW). The average buoyancy frequency in the top 1000 meters is of $2.3 \cdot 10^{-3}$ s^{-1}, in broad agreement with the estimates of Marshall and Plumb (2008).

3.3.3 Zonal transport

The Southern Ocean hosts the ACC, the world's largest oceanic current. How well does our model represent the important zonal flow? Traditionally, the ACC domain is partitioned by the climatological position of regions of strong horizontal gradients, named fronts (Orsi et al., 1995). More recent data, however, convey a somewhat different picture, with the ACC being composed of a greater number of relatively narrow jets, which continuously interact by splitting and merging (Thompson, 2008). Figure 3.4 (left) shows an instantaneous snapshot of zonal-mean zonal velocity, and demonstrates that in the idealised channel the zonal flow is composed by a (varying) number of narrow and intense jets. The time-averaged, zonal-mean zonal velocity is shown instead in figure 3.4 (right) where the time mean is computed over a period of 51 years: as expected, the largest average velocity is attained near the centre of the domain, where wind stress is maximum. The barotropic time-averaged, vertically-averaged zonal flow is computed as:

$$U_{bt} = \int_0^{L_y} \mathrm{d}y \int_{-H}^{0} \mathrm{d}z\,\overline{u}, \tag{3.21}$$

where $\overline{u}$ is the time-average, zonal-average zonal velocity. We obtain $U_{bt} \approx 433$ Sv, which is unrealistically large (Rintoul et al., 2001) due to the absence of bottom topography. The baroclinic zonal flow U_{bc}, however:

$$U_{bc} = \int_0^{L_y} \mathrm{d}y \int_{-H}^{0} \mathrm{d}z\,\overline{u}_{bc}, \tag{3.22}$$

where $u_{bc} = u - u_b$ is the baroclinic velocity and u_b the zonal velocity at the bottom, returns the more reasonable value of $U_{bc} = 88$ Sv. As mentioned in section 2.5.3, topographic form drag does not participate to the eddy energy cycle (see also Ferrari and Wunsch (2009)) and the mechanical energy budget depends on the baroclinic component of zonal velocity only. Thus, the unrealistically large barotropic component does not play an important dynamical role, and we can expect that results obtained with the idealised channel are relevant to more complex model configurations and to the real Southern Ocean.

3.3.4 Zonal momentum balance

The vertically integrated zonal momentum balance equation for a Boussinesq fluid in a re-entrant channel, subject to the assumptions discussed in section 2.5.1, reads

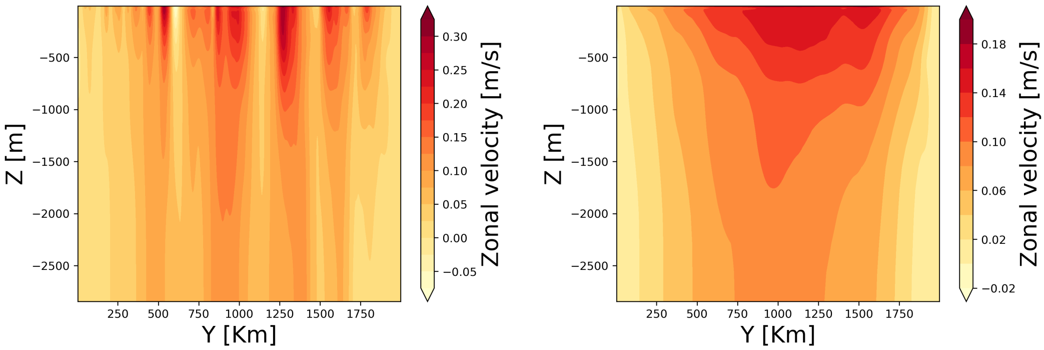

Figure 3.4: Instantaneous (left) and time-mean (right) zonal-mean zonal velocity. The time average is computed over a period of 51 years.

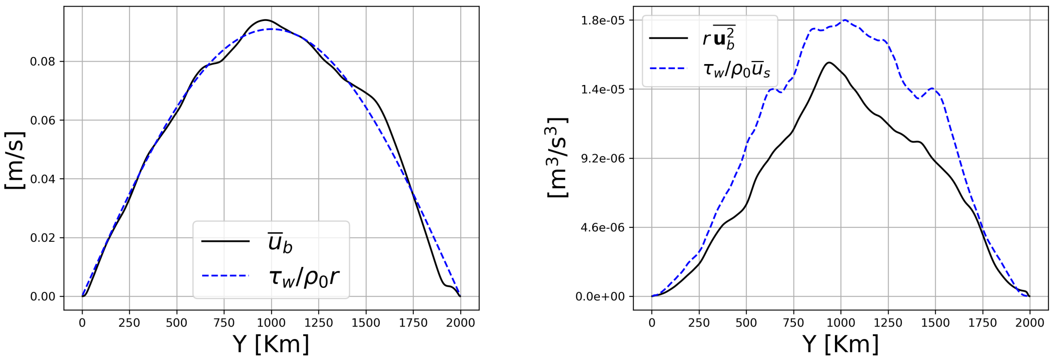

Figure 3.5: Left: meridional profile of time-mean, zonal-mean bottom velocity and the theoretical prediction from zonal momentum balance $\tau_w(y)/\rho_0 r_b$. Right: wind work at the surface (dashed blue line) and energy dissipated by bottom drag (continuous black line).

(equation (2.106)):

$$\overline{u}_b = \frac{\tau_w}{\rho_0 r},$$

which expresses the fact that zonal momentum imparted at the surface by wind stress is dissipated at the bottom by linear bottom drag. Figure 3.5 (left) shows that the meridional profile of zonal-mean, time-mean bottom zonal velocity is in excellent agreement with the theoretical prediction $\tau_w(y)/\rho_0 r_b$ computed using the model parameters shown in table 3.1. The time averages are computed over a period of 51 years.

3.3.5 Mechanical energy balance

The mechanical energy balance for the Boussinesq re-entrant channel is given by equation (2.128). Figure 3.5 (right) shows the wind work at the surface (i.e., the left hand side of equation (2.128), dashed blue line) and the energy dissipated by bottom drag (continuous black line). For this analysis, we diagnosed 5-day spaced instantaneous snapshots of the horizontal velocity, and computed time averages over a short period of 6 years (since using a shorter period gives qualitatively similar outcomes, we do not expect that the accuracy of the results depends heavily on the length of the averaging period). Figure 3.5 demonstrates that the vertically integrated mechanical energy balance holds in our model at any given latitude to within a 20% accuracy. The discrepancy between the two curves is attributed to the undiagnosed lateral stress and viscous dissipation terms (which are neglected in the theoretical derivation of equation (2.128)), and to the fact that a temporal resolution of 5 days may not adequately represent the quadratic bottom dissipation term (as this foregoes all eddy energy at time scales shorter than 5 days).

3.3.6 Thermal wind balance

We have seen in section 3.3.1 that the idealised channel is in the low Rossby number regime and satisfies geostrophic scaling. Geostrophic theory maintains that the vertical shear of zonal velocity is related to the meridional gradient of buoyancy by the thermal wind equation (geostrophy + hydrostasy):

$$\partial_z \overline{u} = -\frac{\partial_y \overline{b}}{f}, \tag{3.23}$$

where the minus sign takes care of the fact that we are in the Southern Hemisphere (f is negative), and the bar denotes time and zonal average. Therefore, we expect that in the idealised channel the baroclinic component of zonal velocity is related to the meridional gradient of buoyancy according to:

$$\overline{u}(z) = -\int_{-H}^{z} \mathrm{d}z' \, \frac{\partial_y \overline{b}(z')}{f}. \tag{3.24}$$

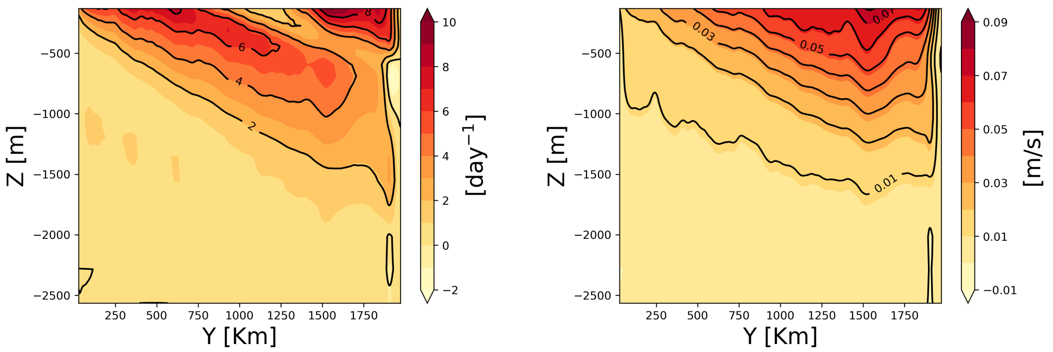

Figure 3.6: Left: vertical shear of time-mean, zonal-mean zonal velocity (colours), and the corresponding thermal wind prediction from equation (3.23) (contours). Right: baroclinic component of time-mean, zonal-mean zonal velocity (colours), and the corresponding thermal wind prediction prediction from equation (3.24) (contours)

Figure 3.6 (left) shows the vertical shear of time-average, zonal-average zonal velocity (colours) and the corresponding prediction from thermal wind (i.e., the right hand side of equation (3.23), contours). Time averages are computed over a period of 51 years. Similarly, figure 3.6 (right) shows the time-average, zonal-average baroclinic velocity (colours) and the thermal wind prediction (the right hand side of equation (3.24), contours). The qualitative agreement between the two sides of the thermal wind relation is satisfactory in both cases. For a more quantitative comparison, the baroclinic zonal flow computed with equation (3.24) amounts to approximately 93 Sv, corresponding to a 5% deviation only from the value of 88 Sv obtained using equation (3.22).

3.3.7 Ekman spirals

Geostrophic balance does not hold in the surface and bottom frictional layers, where the horizontal velocity is endowed with an ageostrophic component supported by Ekman stresses. Here, we demonstrate that the magnitude of the ageostrophic velocity in the surface layer of the idealised channel agrees well with the theoretical prediction based on Ekman theory. In the interest of brevity, we only review the essential, and refer to Vallis (2017) for the complete discussion. The time-mean, zonal-mean horizontal velocity $\overline{\mathbf{u}}$ can be decomposed into its geostrophic and ageostrophic components, $\overline{\mathbf{u}} = \overline{\mathbf{u}}_g + \overline{\mathbf{u}}_{ag}$. The geostrophic component of the zonal velocity is given by thermal wind, while the geostrophic component of the meridional velocity is zero as the zonal-mean zonal pressure gradient is zero by periodicity of the domain. Vallis (2017) shows that the ageostrophic component of the horizontal velocity in the surface Ekman layer can be modelled by the equations:

$$u_{Ek}(z) = -\frac{\sqrt{2}}{d}\frac{\tau_0}{\rho_0 f}e^{\frac{z}{d}}\cos\left(\frac{z}{d}-\frac{\pi}{4}\right) \tag{3.25}$$

$$v_{Ek}(z) = \frac{\sqrt{2}}{d}\frac{\tau_0}{\rho_0 f}e^{\frac{z}{d}}\sin\left(\frac{z}{d}-\frac{\pi}{4}\right), \tag{3.26}$$

where d is an estimate of the depth of the Ekman layer. The name Ekman spirals comes from the fact that in two dimensions $\overline{\mathbf{u}}_{ag} = (u_{Ek}, v_{Ek})$ describes a spiral as depth varies from the bottom of the layer to the surface. Figure 3.7 illustrates the separation of horizontal velocity into its geostrophic and ageostrophic components for the top few hundreds meters of depth, at $y = 1000$ km. Figure 3.7 (left) shows that the vertical profile of the baroclinic zonal velocity (continuous blue line) is well reproduced by the sum of the geostrophic and ageostrophic zonal velocity (continuous red line). The geostrophic zonal velocity alone (dashed red line), instead, is not a good approximation of $\overline{u}_{bc}$ in the Ekman layer. Figure 3.7 (right) demonstrates that the meridional velocity (continuous blue line) is accurately captured by the ageostrophic meridional velocity (continuous red line). The parameter $d \approx 25$ m was tuned so that the theoretical curves best represent the diagnosed fields. Time averages are computed over a period of 51 years.

3.3.8 The Eulerian Streamfunction

The Eulerian streamfunction describes the circulation induced by the zonal-mean velocity on the meridional plane. The theory is dealt with in sections 2.4.2 and 2.5.4, to which we refer for details.

Figure 3.8 (left) shows the time-averaged Eulerian streamfunction (colours),

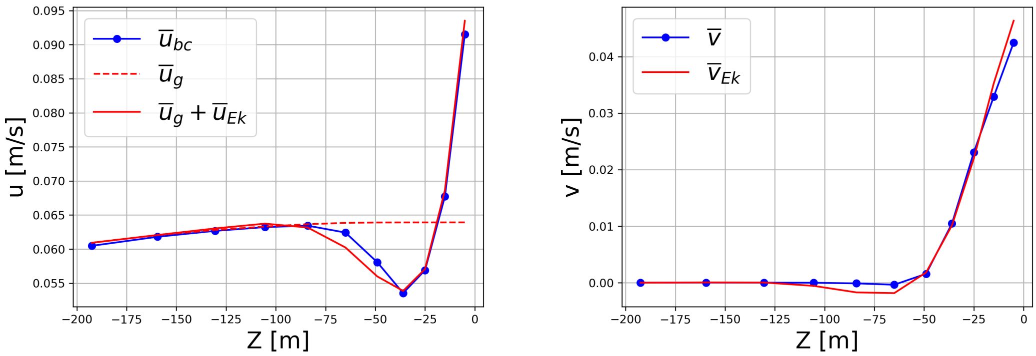

Figure 3.7: Separation of the time-mean, zonal-mean horizontal velocity into its geostrophic and ageostrophic (Ekman) components. Left: vertical profile of time-mean, zonal-mean baroclinic zonal velocity (continuous blue line), geostrophic zonal velocity (dashed red line), and geostrophic plus ageostrophic zonal velocity (continuous red line) over the top 200 m of the ocean. Right: vertical profile of time-mean, zonal-mean meridional velocity (continuous blue line) and ageostrophic meridional velocity (continuous red line) over the top 200 m.

where the time average is computed over a period of 18 years (we found that 18 years provide a sufficiently robust climatology for the Eulerian, residual, and eddy-induced streamfunction, therefore we diagnosed the layers-integrated meridional transport for 18 out of 51 years of model integration only). The iso-contours of $\overline{\psi}$ are approximately vertical away from the surface and bottom Ekman layers: this is an expression of the fact that the zonal-mean meridional velocity $\overline{v}$ is entirely ageostrophic, and that the ageostrophic velocity is non-zero only in the Ekman layers. The circulation described by $\overline{\psi}$ is the notorious Deacon cell (Doos and Webb, 1994), and is characterised by upwelling near the southern boundary, equatorward transport at the surface, downwelling at the northern boundary, and a bottom return flow. This pattern of circulation would imply a large diabatic transport in the interior of the ocean but, as seen in section 2.4.3, it is not the Eulerian circulation that on average advects tracers in the Southern Ocean. The diagnosed Eulerian streamfunction compares well with the theoretical prediction computed from the zonal balance equation (2.136) (black contours):

$$\overline{\psi} = -\frac{\tau_0}{f\rho_0}.$$

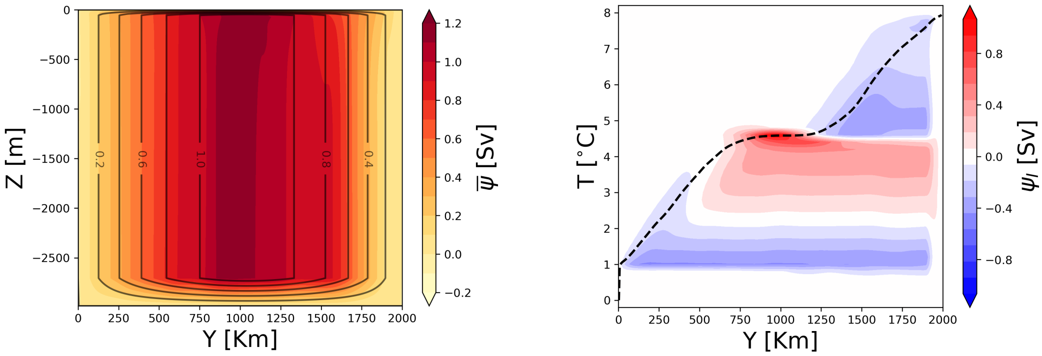

Figure 3.8: Left: Time-mean eulerian streamfunction (colours) and theoretical prediction from zonal balance, equation (2.136) (contours). For the theoretical prediction we assume that $\overline{\psi}$ decreases linearly to zero in the surface and bottom Ekman layers. Right: Residual streamfunction in isopycnal coordinates (colours). The black dashed line marks the time-mean, zonal-mean Sea Surface Temperature

Note though that this equation is only valid in the interior. A common assumption made to parametrise $\overline{\psi}$ in the surface and bottom Ekman layers is to assume that it decreases linearly from its value at the base of the layer (given by equation (2.136)) to zero at the surface and bottom.

3.3.9 The residual Streamfunction

The residual streamfunction describes the circulation that on average advects tracers in the Southern Ocean, see the theory in section 2.4.3. The time-averaged residual streamfunction in isopycnal coordinates is shown in figure 3.8 (right), where the average is computed over a period of 18 years. The iso-contours of ψ_I are approximately horizontal away from the surface and northern diabatic layers (the black dashed line marks the time-mean, zonal-mean Sea Surface Temperature), which is an expression of the fact that the residual circulation is directed along isopycnals in the interior. Iso-contours are not horizontal in the diabatic layers though, where a cross-isopycnal return flow allows closure of the meridional overturning. There are three distinct overturning cell: a lower negative cell, an intermediate positive cell, and an upper negative cell. Many studies focus on the first two cells only (e.g. Abernathey et al. (2011)), which is why the intermediate cell is sometimes referred to as ”the upper cell”. In between the lower and intermediate cell is located the important upwelling branch of the overturning circulation, which provides an adiabatic pathway for deep water to reach the surface (Marshall and Speer, 2012). Water that outcrops close to the southern boundary of the domain is exposed to buoyancy loss and transforms into very dense downwelling water, which can be thought of as our model’s representation of AABW formation. Water that outcrops at lower latitudes is exposed to northward Ekman transport and buoyancy gain, and eventually enters a region of buoyancy loss at the surface: the ensuing downwelling branch of the intermediate cell is representative of SAMW-AAIW formation (Abernathey et al., 2011). We observe that the intermediate cell peaks within the surface diabatic layer (Abernathey et al. (2011) noted that this cell tends to become confined in the surface diabatic layer in their weak wind-stress forcing experiments). The partitioning of the overturning circulation in three cells takes a more familiar form when the streamfunction is mapped back to depth coordinates, figure 3.9 (left). This representation of the streamfunction makes it even clearer that, in the interior, the residual circulation follows isothermals (shown in black contours). Overall, the idealised channel produces an overturning circulation whose structure is in good qualitative agreement with estimates for the real Southern Ocean, see for example Marshall and Speer (2012), Rintoul et al. (2001). Similarly to Abernathey et al. (2011), we estimate the interior volume transport associated with the lower and intermediate overturning cell by taking the minimum and maximum respectively of ψ_{res} below 500 m depth and at $y = 1800$ km. The precise meridional location where the extrema are taken is not important because we have seen that, in the interior, the streamfunction is approximately constant along isopycnals. We obtain a value of approximately 0.4 Sv for both cells, which corresponds to ~ 10 Sv when we take into account that the zonal extent of the idealised channel is about 25 times smaller than that of the real Southern Ocean. This compares well with estimates for the real Southern Ocean, see for example Marshall and Speer (2012).

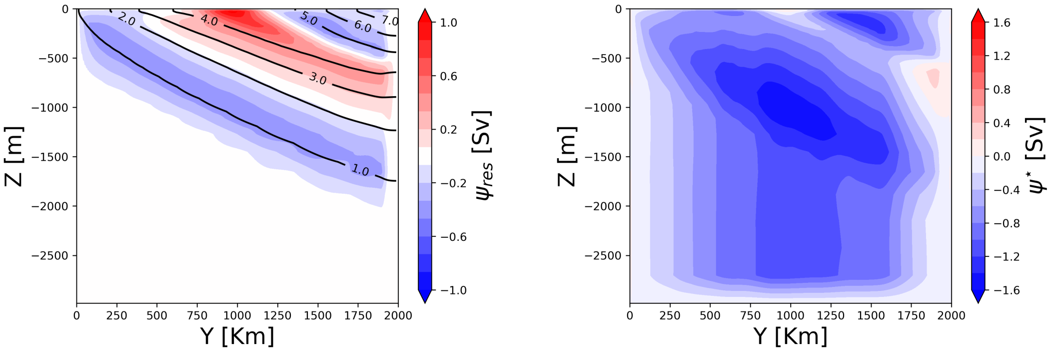

Figure 3.9: Left: Time-averaged residual streamfunction in depth coordinates (colours) and time-mean, zonal-mean temperature (contours). Right: Time-averaged eddy-induced streamfunction.

3.3.10 The eddy-induced streamfunction

The eddy-induced streamfunction describes the pattern of circulation associated with the action of baroclinic eddies. Rather than from its definition, equation (2.74), ψ^* is diagnosed as $\psi^* = \psi_{\text{res}} - \overline{\psi}$. The time-mean eddy-induced streamfunction is shown in figure 3.9 (right), where the time mean is computed over a period of 18 years. The eddy-induced streamfunction is negative over most of the domain and the overall sense of the circulation is counter-clockwise, demonstrating that ψ^* effectively opposes the wind-induced circulation and acts to flatten isopycnals in the channel. The volume transport associated with ψ^* is large, and comparable in magnitude to that of the Eulerian streamfunction, figure 3.8 (left), which is consistent with the idea that ψ_{res} is the small residual of the balance between the large wind-induced and eddy-induced circulations.

3.4 Conclusions

In this chapter, I have discussed a particular configuration of the MITgcm, the idealised channel, which I will use to simulate the dynamics of the Southern Ocean in the rest of the manuscript. I have made several assumptions regarding the geometry of the problem and the nature of the forcing applied at the boundaries, thereby choosing to work in a simplified setup. In return for the sacrificed realism, the model can be run at a high horizontal resolution, so that baroclinic eddies can be resolved throughout the domain. Consistently with this premise, my goal is to investigate the dynamical processes at play in the channel (and those driven by baroclinic eddies in particular) qualitatively, and not to make quantitative predictions for the real Southern Ocean. Nevertheless, following the lead of previous research I have explored the physical properties of a reference state of the idealised channel (the control run), and found that the model produces a plausible representation of the observed Southern Ocean. Specifically, I have demonstrated that the idealised channel has a realistic meridional structure of temperature and large scale stratification. Our estimate for the time-averaged baroclinic zonal flow sits in the range of values reported in the literature. The barotropic flow, on the other hand, is unrealistically large due to the absence of topographic drag but, as explained in chapter 2, this is unimportant for the eddy energy cycle. The model is in the low Rossby number regime, and I have ascertained that the thermal wind relation, which crucially links tilted isopycnals with the intense zonal flow, holds to a high accuracy in the channel. The Eulerian streamfunction, describing the wind-induced circulation, agrees well with the theoretical prediction. When properly scaled, the residual streamfunction has plausible magnitude and structure too, and captures the main processes of water masses formation believed to be at play in the real ocean. Overall, this analysis confirms that the idealised channel constitutes a reliable representation of the Southern Ocean, and an appropriate tool to study how baroclinic eddies affect its dynamics.

Chapter 4

Time-scales of natural variability in the Southern Ocean: part 1

4.1 Introduction

4.1.1 Motivation

Ocean mesoscale eddies enter the dynamical balance of the Southern Ocean at leading order (Marshall and Radko, 2003). It has been shown that they play a pivotal role in controlling its internal variability (Hogg and Blundell, 2006, Sinha and Abernathey, 2016, Wilson et al., 2015) and the response of its circulation to wind stress changes (Abernathey and Ferreira, 2015, Abernathey et al., 2011, Hallberg and Gnanadesikan, 2006, Viebahn and Eden, 2010). However, a broad consensus over the physical mechanisms governing their interaction with the large scale flow has not yet been established. A recent and promising research avenue towards an improved understanding of the subject builds on the dynamical similarity between the ACC and the atmospheric jet stream (Thompson, 2008). Williams et al. (2007), for example, found analogies in the patterns of eddy vorticity forcing on the mean flow in the atmospheric storm track and in the Southern Ocean. Ambaum and Novak

(2014) (AN14 hereafter) introduced a model of atmospheric storm track variability which describes the baroclinic life-cycle in terms of a two-dimensional dynamical system. The key idea is that baroclinic development can be sustained by an external forcing even in conditions of baroclinic neutrality, and that interactions between the growing instability and the mean flow may produce periodic oscillations. The overall physical picture is similar to that expressed by Lotka-Volterra models of population growth, with eddies acting as a population of predators feeding upon the mean flow, as described in chapter 1. One of the key advantages of the AN14 model is that it captures the full eddy life-cycle, whereas conventional models of baroclinic instability (e.g. Eady (1949)) tend to focus on the growing phase of the instability only.

The equations proposed by AN14 are:

$$\frac{\mathrm{d}X}{\mathrm{d}t} = F - Y \tag{4.1}$$

$$\frac{\mathrm{d}Y}{\mathrm{d}t} = 2(X - D)Y\,, \tag{4.2}$$

where X and Y represent spatial averages of baroclinicity (measuring mean flow) and eddy heat flux (measuring eddy activity) respectively. F is a diabatic forcing term, and D is a dissipation term. The AN14 model was introduced based on heuristic arguments, but Novak et al. (2017) found the model predictions in good agreement with atmospheric reanalysis and General Circulation Model (GCM) data. Yano et al. (2020) and Marcheggiani et al. (2022) provided further evidence that a similar approach based on simplified dynamical systems can be successfully employed to study the variability of the atmosphere. Here, I address the following question: to what extent is the AN14 reduced-order model of atmospheric turbulence transferable to its natural oceanic equivalent? Although Marshall et al. (2017) proposed a model of eddy saturation in the Southern Ocean inspired by AN14, the idea that this dynamical system approach can capture the physics of eddy-mean flow interaction in the Southern Ocean jets has not been tested yet. The purpose of this work is to explore this possibility by using a high-resolution dataset, the MITgcm idealised channel configuration described in chapter 3.

4.1.2 The AN14 model

We start by taking a quick look at the AN14 model in its original formulation. Firstly, note that equations (4.1) and (4.2) admit a stationary solution $(\bar{X},\bar{Y}) = (D,F)$, which is found by imposing the condition $\frac{\mathrm{d}}{\mathrm{d}t}(X,Y) = (0,0)$. Further properties of the AN14 model include:

1. It is non-linear, meaning that the terms on the right hand side of equations (4.1) and (4.2) are non-linear in the state variables X and Y.
2. It is conservative, therefore there exists a function of X and Y that is conserved under time evolution. Roncoroni (2018) explicitly calculated the Lagrangian function associated to the equations of motion.
3. It is an oscillator, i.e. solutions to the model equations are periodic.

A corollary of the last property is that when the AN14 model is linearised it reduces to a harmonic oscillator. To see this, it is sufficient to expand the right hand side of equations (4.1) and (4.2) to first order around $(\bar{X},\bar{Y})$, yielding:

$$\frac{\mathrm{d}X}{\mathrm{d}t} = F - Y \tag{4.3}$$
$$\frac{\mathrm{d}Y}{\mathrm{d}t} = 2F(X-D)\,. \tag{4.4}$$

By rescaling $X - D \to X$ and $-(Y-F) \to Y$ we obtain:

$$\frac{\mathrm{d}X}{\mathrm{d}t} = Y \tag{4.5}$$
$$\frac{\mathrm{d}Y}{\mathrm{d}t} = -2FX\,, \tag{4.6}$$

which is a harmonic oscillator of elastic constant $k^2 = 2F$. We will see more details about this linear limit in section 4.2.7.

Figure 4.1 (left) shows the numerical solution of equations (4.1) and (4.2), for $F = 1$, $D = 1$, $X(t=0) = 0$, and $Y(t=0) = 2$. The equations are integrated with the 4-step Runge-Kutta method with time step $\Delta t = 0.01$, and up to a final time

$t = 10$ (time is in arbitrary units). As anticipated, the system is an oscillator, i.e. the solution repeats itself after a certain fixed period of time. Also, consistently with the predator-prey interpretation of the dynamics, we observe that smooth peaks of the mean flow X (the prey, continuous line) lead sharper bursts of the eddy heat flux Y (the predator, dashed line).

Due to the chaotic nature of atmospheric and oceanic flows, the physical interpretation of real-world time series is more complicated than in the simple example above (see for example figure 4 in AN14). A more convenient way to visualise the time evolution of the system is by tracking its trajectory in the phase space. The phase space is the space spanned by the values of the state coordinates - in the case of the AN14 model, there are only two variables X and Y and the phase space is a plane. At any given instant of time, therefore, the state of the system is represented by a point in the plane. As time evolves, the representative points of the system define curves, also called trajectories. The structure of phase space trajectories reflects the nature of the dynamics: a rich phenomenology is observed in dynamical system theory, including trajectories that collapse to a stationary point (e.g. the damped harmonic oscillator), diverge (a free particle), or generate complicated fractal objects (the Lorenz system). The AN14 model is an oscillator, therefore phase space trajectories form closed orbits. Figure 4.1 (right) shows examples of phase space trajectories obtained by integrating equations (4.1) and (4.2) with the Runge-Kutta method for different initial conditions. Phase-space diagrams constructed with real-world data are noisy too, but we will exploit a kernel averaging technique developed by Novak et al. (2017) to smooth the raw data in a way that preserves information about the dynamics. Importantly, for the AN14 model the phase space diagram may be interpreted as a quantitative representation of the eddy-mean flow life cycle described in figure 1.3, where a complete circuit of the phase space orbit corresponds to a full cycle. The implication is that, when the phase space diagram is reconstructed from data with the kernel averaging technique, nearly-closed orbits such as those of figure 4.1 (right) constitute a strong indication that the underlying dynamics are characterised by oscillatory behaviour. In the following, we will try to detect

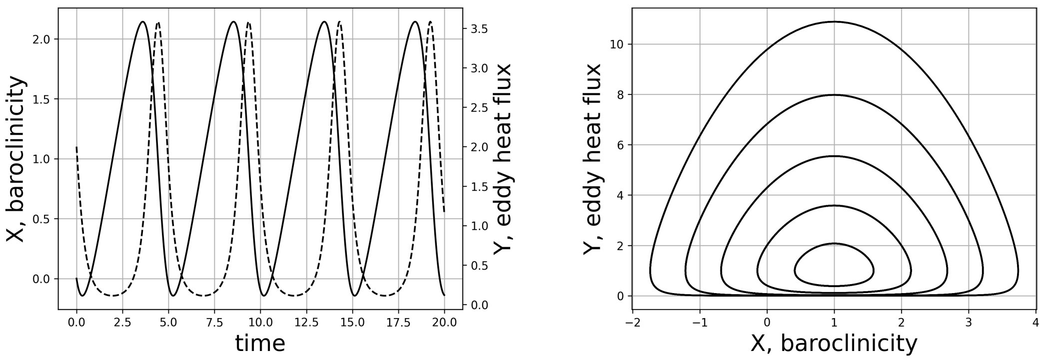

Figure 4.1: Left: time evolution of baroclinicity X (continuous line) and eddy heat flux Y (dashed line) from a numerical solution of equations (4.1) and (4.2), for $F = 1$, $D = 1$, $X(t = 0) = 0$, and $Y(t = 0) = 2$. Right: phase space trajectories obtained by integrating the AN14 model with the same parameters as before but varying the initial conditions.

such signature for the oceanic case using data from the MITgcm idealised channel configuration.

4.1.3 Structure of the chapter

The chapter is structured as follows: in section 4.2 I discuss the theoretical background necessary for the subsequent analysis. Importantly, I introduce a modified version of the AN14 model designed to take into account stochastic effects explicitly. I explain which data are used and how they are processed in section 4.3. In section 4.4, I present the method employed to fit the dynamical system to the data and evaluate the goodness of the fit. Results are illustrated in section 4.5: this includes the analysis of the idealised channel dataset and of a synthetic validation experiment. Conclusions and perspectives are offered in section 4.7.

4.2 Theoretical background

The primary goal of this work is to model the interaction between eddies and mean flow in the Southern Ocean with a two-dimensional dynamical system. As it turns out, effecting this task requires preliminary consideration of a sizeable amount of heterogeneous material. I have decided to collect as many technical details as possible in an individual section dedicated to theory rather than cluttering the illustration of key results with that of necessary but unessential subtleties. Structural homogeneity comes at the cost of a slight abstraction, so that this section presents the reader with content whose usefulness will only become manifest at a later stage of the discussion. The diverse nature of the material also made it a challenge to achieve a thematic coherence in its arrangement. Simplifying matters, however, the topics presented in this section can be divided into three main subjects:

1. Stochastic processes and their statistical properties.
2. Stochastic bivariate oscillators.
3. Kernel averaging.

The section is structured as follows: in the first part I set out the basic notation necessary to understand the mathematical properties of the stochastic dynamical system with which I intend to model eddy-mean flow interaction. This part mostly concerns stochastic processes and some of their statistical properties, such as covariance and correlation functions. The model itself is introduced and discussed in the second part. In the third and final part, I discuss a mathematical technique which allows to reduce the amount of noise contained in an individual realisation of a stochastic dynamical system. This section is almost entirely based on existing literature: the exceptions are the red noise test in section 4.2.6 and the benchmarking of the kernel averaging routine in section 4.2.9, which are the author's own work.

4.2.1 Foundations

Consider the two-dimensional autonomous dynamical system:

$$\frac{\mathrm{d}}{\mathrm{d}t}X = f(X,Y) \tag{4.7}$$

$$\frac{\mathrm{d}}{\mathrm{d}t}Y = g(X,Y)\,. \tag{4.8}$$

The state variables are named X and Y. The functions f and g on the right hand side describe the deterministic part of the dynamics, and are at present arbitrary. Equations (4.7) and (4.8) are meant to constitute a simplified description of a complex, high-dimensional geophysical flow. The variables X and Y can be interpreted as spatially averaged geophysical quantities, with the average taken over a suitable domain. A notable example is the AN14 model presented in the introduction. In this case:

$$f(X,Y) = F - Y \tag{4.9}$$

$$g(X,Y) = 2(X-D)Y\,, \tag{4.10}$$

where X is the Eady growth rate (see section 4.3.2 for the definition), Y is the eddy heat flux, and their spatial average is computed over the North Atlantic storm track region.

In general, it is not possible to derive a closed set of equations akin to (4.7) and (4.8) that predicts the evolution of X and Y for a high-order geophysical system exactly. The reason is that, even if a physical relation between the variables X and Y exists, it is likely to be mediated by the system's remaining degrees of freedom, so that equations (4.7) and (4.8) should be interpreted as an averaged and approximated representation of the dynamics (Yano et al., 2020). Under favourable circumstances, though, the effect of the neglected degrees of freedom on the evolution of X and Y can be modelled with the inclusion of stochastic terms on the right hand side of the equations. With informal notation:

$$\mathrm{d}X = f(X,Y)\mathrm{d}t + \sigma_x \mathrm{d}\xi_x \tag{4.11}$$

$$\mathrm{d}Y = g(X,Y)\mathrm{d}t + \sigma_y \mathrm{d}\xi_y\,, \tag{4.12}$$

where $d\xi_x$, $d\xi_y$ are stochastic terms (for example, Brownian motion increments) and σ_x, σ_y non-negative real numbers representing the amplitude of the noise. Note that $d\xi_x$, $d\xi_y$ do not represent any specific missing physical processes (e.g., sub-grid scale processes): rather, they are meant to model the loss of complexity inherent to reducing a comprehensive climate model to a simple system of two differential equations. Such a modelling choice can sometimes be rigorously justified by means of scale separation arguments but here, similarly to Yano et al. (2020), we simply take it as an ansatz.

The explicit inclusion of the stochastic terms in the model introduces mild constraints on the deterministic part of the dynamics. Consider as an example the case when the functions f and g are linear:

$$\mathrm{d}X = (aX + bY)\mathrm{d}t + \sigma_x \mathrm{d}\xi_x \tag{4.13}$$

$$\mathrm{d}Y = (cX + dY)\mathrm{d}t + \sigma_y \mathrm{d}\xi_y\,. \tag{4.14}$$

We will study these equations at length in section 4.2.7. For now, we mention that they have statistically stationary solutions only when the matrix with components a, b, c, and d is negative definite. The negative eigenvalues are associated to damping: in other words, the deterministic dynamics are dissipative. Intuitively, the noise terms introduce supplementary ”energy” into the system, which must be dissipated if the solutions are to remain bounded. This should be contrasted with the original AN14 model, which is a conservative system and cannot support stationary solutions if the stochastic terms are explicitly represented. In view of this, we seek to represent eddy-mean flow interaction in the Southern Ocean with a dynamical system that fulfils the following requirements:

1. It is two-dimensional.

2. It explicitly accounts for noise.

3. The deterministic part of the dynamics is oscillatory, and as in AN14 it expresses a predator-prey relationship between eddies and the mean flow.

4. The deterministic part of the dynamics is dissipative, so that stationary solutions of the stochastically forced problem exist.

This is our basic problem setup. Conditions 1 and 2 are automatically satisfied as long as the system has the form of (4.11) and (4.12). In section 4.2.7, we will see that the simplest model that can accommodate all the remaining conditions is a stochastic bivariate linear oscillator. Accordingly, the following sections are dedicated to the exploration of some mathematical properties of the system defined by equations (4.11) and (4.12), with a special focus on linear oscillatory dynamics.

4.2.2 Stochastic processes

Upon discretisation, a stochastic bivariate linear oscillator is a bivariate auto-regressive process. This family of models is best discussed with the language of stochastic processes theory. It is beyond the scope of this work to offer a detailed presentation of the subject (for that we refer to Higham (2001) and references therein), and we shall thus limit ourselves to those aspects that are strictly necessary to our analysis. In this section, we set out the notation we use in the rest of the chapter. Selected topics are explored more in depth in the following sections.

Consider a scalar stochastic differential equation (SDE) of the form:

$$\mathrm{d}X = f(X)\mathrm{d}t + \sigma_x \mathrm{d}\xi. \tag{4.15}$$

Here, $f(X)$ encapsulates the deterministic dynamics, σ_x is the noise amplitude, and $\mathrm{d}\xi$ is, with informal notation, the infinitesimal brownian motion increment (see below). A discrete version of this equation can be obtained upon finite differencing. The simplest choice is the Euler-Maruyama scheme (Higham, 2001), giving:

$$X_{n+1} = X_n + f(X_n)\Delta t + \sigma_x \mathrm{d}W_{n+1}, \tag{4.16}$$

where $t_n = n\Delta t$, $X_n = X(t_n)$, and $\mathrm{d}W_{n+1}$ is the Brownian motion increment, a normally distributed random variable with zero mean and variance Δt. By stochastic process we mean an indexed collection of random variables, where the index describes time. When the index is an integer, the stochastic process is said to be discrete. Equations of the form (4.15) and (4.16) define a continuous and a discrete stochastic process respectively. A time series is a random realisation of a discrete stochastic process. A stochastic process is said to be real valued if X_n is real for all n. The expectation value of X_n, which we denote by $\mathbb{E}[X]$, is defined in the usual sense of probability theory. A stochastic process is stationary if its stochastic properties do not depend on the time index n or, more formally, if the joint probability distribution of X_n and X_m only depends on $|n-m|$ (von Storch and Zwiers, 1999). In this work, we only consider discrete, real-valued, stationary stochastic processes. Some special cases are of particular interest. White noise is a sequence of independent and identically distributed normal random variables. Some properties of white noise will be discussed in subsequent sections. A stochastic process is linear when $f(X) = aX$. The associated SDE is:

$$\mathrm{d}X = aX\mathrm{d}t + \sigma_x \mathrm{d}\xi, \tag{4.17}$$

which, for $a < 0$, is called the Ornstein-Uhlenbeck process, see for example Vatiwutipong and Phewchean (2019). The corresponding discrete process reads:

$$X_{n+1} = (1 + a\Delta t)X_n + \sigma_x \mathrm{d}W_n, \tag{4.18}$$

which is named an auto-regressive process of order 1, or red noise (von Storch and Zwiers, 1999). We will see more properties of auto-regressive processes later on. Finally, a stochastic process is said to be multivariate when X_n is replaced with a vector-valued variable $\mathbf{X}_n$. If $\mathbf{X}_n$ only has two components, the process is bivariate. Linear bivariate processes are the focus of this chapter.

4.2.3 Covariance and correlation functions

In many practical cases, stochastic processes can be characterised by investigating a few of their statistical properties. Covariance and correlation functions are especially useful to illuminate a process' typical time scales, and a common tool in time series analysis. Our goal in this section is to outline the definitions adopted in subsequent parts of the chapter. We refer to von Storch and Zwiers (1999) for a comprehensive discussion of the subject.

Let X_n be a real-valued, stationary, univariate stochastic process, with $n = 1, ..., N$. The auto-covariance function of X_n is:

$$\mathrm{Cov}_{XX}(\tau) = \mathbb{E}[(X_n - \mu_X)(X_{n+\tau} - \mu_X)], \tag{4.19}$$

where μ_X is the expectation value of X_n (i.e, the average of X in usual language), and the integer τ is named the lag. This definition is not universal, and sometimes the convention:

$$\mathrm{Cov}_{XX}(\tau) = \mathbb{E}[(X_{n+\tau} - \mu_X)(X_n - \mu_X)], \tag{4.20}$$

is adopted instead (e.g. Frankignoul and Hasselmann (1977)). Two properties are worth remembering: (i) the auto-covariance is an even function of the lag τ:

$$\mathrm{Cov}_{XX}(\tau) = \mathrm{Cov}_{XX}(-\tau), \tag{4.21}$$

and (ii) the lag-zero auto-covariance is equal by definition to the variance of X_n:

$$\mathrm{Cov}_{XX}(0) = \mathrm{Var}[X_n]. \tag{4.22}$$

Intuitively, the auto-covariance function measures how rapidly the stochastic process X_n loses memory of its past values. Suppose, for the sake of the illustration, that the auto-covariance of a process X_n decays exponentially with the time lag τ (as is the case for a red noise process, see below), and let T be the e-folding time. For $\tau >> T$, the auto-covariance is approximately zero: this indicates that the random variables $X_n - \mu_X$ and $X_{n+\tau} - \mu_X$ are weakly related. Put in other words (von Storch and Zwiers, 1999), the quantity X_n is not a skilful predictor for the future value $X_{n+\tau}$. Conversely, $X_{n+\tau}$ will not be too dissimilar from the initial value X_n for $\tau << T$, and a prediction based on the persistence of X_n will yield reasonable success. The e-folding time T separates between the two regimes, and may be interpreted as a characteristic decorrelation time for the process. The auto-covariance function is related to the spectrum of the process via the Fourier transform (von Storch and

Zwiers, 1999), and is thus a powerful tool to detect periodic signals obscured by noise.

When there are two stationary, univariate, stochastic processes X_n and Y_n, the notion of auto-covariance can be generalised by that of cross-covariance:

$$\mathrm{Cov}_{XY}(\tau) = \mathbb{E}[(X_n - \mu_X)(Y_{n+\tau} - \mu_Y)], \tag{4.23}$$

where μ_X is the expectation value of X and μ_Y is the expectation value of Y. The cross-covariance function satisfies the property:

$$\mathrm{Cov}_{XY}(\tau) = \mathrm{Cov}_{YX}(-\tau). \tag{4.24}$$

For a multivariate stochastic process $\mathbf{X}_n$, the notions of auto- and cross-covariance functions combine into that of lagged covariance matrix:

$$\Sigma(\tau) = \mathbb{E}[(\mathbf{X}_n - \boldsymbol{\mu}_X)(\mathbf{X}_{n+\tau} - \boldsymbol{\mu}_X)^T], \tag{4.25}$$

where ^{T} denotes transposition and $\boldsymbol{\mu}_X$ is the vector expectation value of $\mathbf{X}_n$. In the special case of a bivariate process $\mathbf{X}_n$ with components X_n and Y_n, the lagged covariance matrix is a 2×2 matrix:

$$\Sigma(\tau) = \begin{bmatrix} \Sigma_{XX}(\tau) & \Sigma_{XY}(\tau) \\ \Sigma_{YX}(\tau) & \Sigma_{YY}(\tau). \end{bmatrix}. \tag{4.26}$$

Of special importance is the lagged covariance matrix at lag zero, simply called the covariance matrix and denoted by Σ^0:

$$\Sigma^0 = \mathbb{E}[(\mathbf{X}_n - \boldsymbol{\mu}_X)(\mathbf{X}_n - \boldsymbol{\mu}_X)^T], \tag{4.27}$$

The covariance functions normalised by the variance are called correlation functions. For two processes X_n and Y_n, the definition of the auto- and cross-correlation functions are:

$$\rho_{XX}(\tau) = \frac{\mathrm{Cov}_{XX}(\tau)}{\mathrm{Var}[X_n]} \tag{4.28}$$

$$\rho_{YY}(\tau) = \frac{\mathrm{Cov}_{YY}(\tau)}{\mathrm{Var}[Y_n]} \tag{4.29}$$

$$\rho_{XY}(\tau) = \frac{\mathrm{Cov}_{XY}(\tau)}{\sqrt{\mathrm{Var}[X_n]\mathrm{Var}[Y_n]}}. \tag{4.30}$$

The correlation functions convey the same information as the covariance functions, but are non-dimensional variables and are valued between -1 and 1. If it is convenient to view X_n and Y_n as the components of a bivariate stochastic process $\mathbf{X}_n$, the equations above can be re-written in the equivalent form:

$$\rho_{XX}(\tau) = \frac{\Sigma_{XX}(\tau)}{\Sigma^0_{XX}} \tag{4.31}$$

$$\rho_{YY}(\tau) = \frac{\Sigma_{YY}(\tau)}{\Sigma^0_{YY}} \tag{4.32}$$

$$\rho_{XY}(\tau) = \frac{\Sigma_{XY}(\tau)}{\sqrt{\Sigma^0_{XX}\Sigma^0_{YY}}}, \tag{4.33}$$

where the emphasis is placed on the matrix structure of the problem.

4.2.4 Estimation of the correlation functions

Consider a bivariate time series $\mathbf{X}_n = (X_n, Y_n)$, with $n = 1, ..., N$. The correlation functions can be inferred from data by means of the following estimators:

$$r_{XX}(\tau) = \frac{\frac{1}{N}\sum_{n=1}^{N}(X_n - \mu_X)(X_{n+\tau} - \mu_X)}{\mathrm{Var}[X_n]} \tag{4.34}$$

$$r_{YY}(\tau) = \frac{\frac{1}{N}\sum_{n=1}^{N}(Y_n - \mu_Y)(Y_{n+\tau} - \mu_Y)}{\mathrm{Var}[Y_n]} \tag{4.35}$$

$$r_{XY}(\tau) = \frac{\frac{1}{N}\sum_{n=1}^{N}(X_n - \mu_X)(Y_{n+\tau} - \mu_X)}{\sqrt{\mathrm{Var}[X_n]\mathrm{Var}[Y_n]}}, \tag{4.36}$$

where r_{XX}, r_{YY}, r_{XY} are the estimators of ρ_{XX}, ρ_{YY}, and ρ_{XY} respectively. In the statistical literature, the normalisation factor is commonly set to $N-1$ rather than N, but the difference is negligible if the sample size is sufficiently large, which is always the case in this work.

4.2.5 Correlation functions: examples

To familiarise with the computation and interpretation of auto- and cross-correlation functions, we consider two simple examples where analytical expressions can be derived exactly (see von Storch and Zwiers (1999) for the complete theory). In the first example we compute the auto-correlation function of an AR(1) process. The exercise is instructive in that most of the ideas introduced here apply straightforwardly to the more complicated cases considered later on. The second example consists in predicting the form of the cross-correlation function between an AR(1) process and its driving noise. It is informative because the cross correlation is not symmetric with respect to $\tau = 0$, which helps attributing the correct physical meaning to values taken at positive and negative lags.

AR(1) process

Consider the zero mean AR(1) process:

$$X_{n+1} = \alpha_1 X_n + Z_{n+1}, \tag{4.37}$$

with $0 < \alpha_1 < 1$ and $Z_{n+1} = \sigma \mathrm{d}W_{n+1}$. The analytical expression for the auto-correlation function can be obtained easily as follows. By taking expectation of the defining equation we obtain $\mathbb{E}[X_{n+1}] = \alpha_1 \mathbb{E}[X_n]$, which leads to:

$$\mathbb{E}[X_n] = 0, \tag{4.38}$$

as expected. To compute the variance, we note that:

$$\mathrm{Var}[X_n] = \mathbb{E}[X_n X_n] = \alpha_1^2 \mathrm{Var}[X_n] + \sigma_Z^2, \tag{4.39}$$

with $\sigma_Z^2 = \mathrm{Var}[Z_n] = \sigma^2 \Delta t$, yielding:

$$\mathrm{Var}[X_n] = \frac{\sigma_Z^2}{1 - \alpha_1^2}. \tag{4.40}$$

The auto-covariance function is computed similarly, obtaining:

$$\mathrm{Cov}_{XX}(\tau) = \frac{\sigma_Z^2}{1 - \alpha_1^2} \alpha_1^{|\tau|}. \tag{4.41}$$

Finally, upon normalisation by the variance we compute the auto-correlation function:

$$\rho_{XX}(\tau) = \alpha_1^{|\tau|}. \tag{4.42}$$

Thus, the auto-correlation function of an AR(1) process decays exponentially with the lag, and the rate of decay is controlled by the process parameter α_1. We generate a synthetic realisation of the process by integrating equation (4.37) with the Euler-Maruyama method (Higham, 2001), for $\Delta t = 5$, $N = 2160$ (corresponding to 30 years if the time unit is days), and $\sigma = 1$. Figure 4.2 (left) shows the auto-correlation function estimated with formula (4.34) (continuous blue line), and the theoretical predication (4.42) (dashed black line). The numerical simulation confirms the exponential profile of the auto-correlation.

AR(1) process and its driving noise

We consider a zero-mean AR(1) process as before, and we study the cross-correlation function between the driving noise and the process itself. Although we consider this problem for its instructional value, a famous application was offered in the seminal work of Frankignoul and Hasselmann (1977), where it was shown that Sea Surface Temperature anomalies in the mid-latitudes can be modelled as an AR(1) process forced by rapidly fluctuating air-sea fluxes. The cross-correlation function between SST anomalies (the process) and air-sea fluxes (the driving noise) was exploited in Figure 9 of the paper to demonstrate the agreement between model and observations. Here, we seek to understand how the form of the cross-correlation function shown in Frankignoul and Hasselmann (1977) can be predicted analytically from the defining equations of the process.

We start by computing the cross-covariance function at lag zero:

$$\mathrm{Cov}_{ZX}(0) = \mathbb{E}[Z_n X_n] = \mathbb{E}[Z_n(\alpha_1 X_{n-1} + Z_n)] = \sigma_Z^2, \tag{4.43}$$

the last equality due to $\mathbb{E}[Z_n X_{n-1}] = 0$ (i.e, the process at time $n-1$ does not depend on the realisation of the noise at time n). The cross-covariance function at non-zero lags is computed similarly, yielding:

$$\mathrm{Cov}_{ZX}(\tau) = \begin{cases} \sigma_Z^2 \alpha_1^\tau & \tau \geq 0 \\ 0 & \tau < 0\,. \end{cases} \tag{4.44}$$

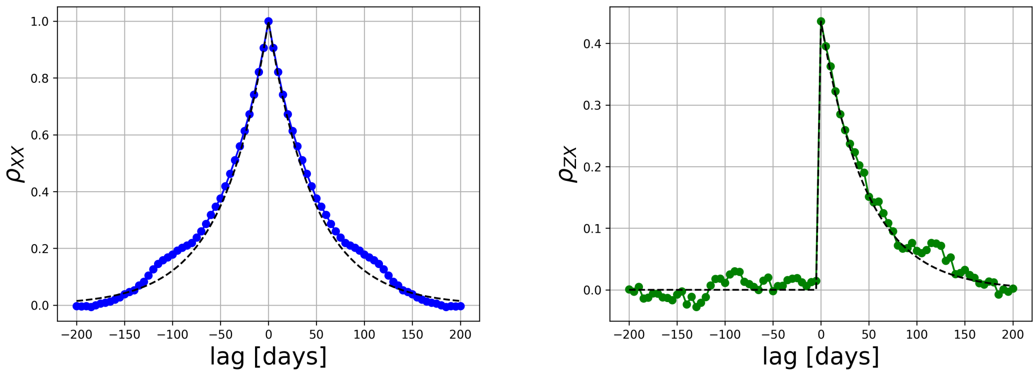

Figure 4.2: Left: estimated (continuous blue line) and theoretical (dashed black line) auto-correlation function of an AR(1) process. Right: estimated (continuous green line) and theoretical cross-correlation function between the driving noise of an AR(1) process and the process itself. See main text for interpretation.

Upon normalisation we obtain:

$$\rho_{ZX}(\tau) = \begin{cases} \sqrt{1-\alpha_1^2}\alpha_1^\tau & \tau \geq 0 \\ 0 & \tau < 0\,. \end{cases} \tag{4.45}$$

In this example, the cross-correlation function is highly asymmetric. Its form can be interpreted as follows: at positive lags, Z leads X, hence $\rho_{ZX}(\tau)$ represents the correlation between X and past values of its driving noise. This correlation is maximum at lag zero and decays exponentially with time, expressing the idea that memory of past values of the driving noise fades with time. At negative lags, X leads Z, and $\rho_{ZX}(\tau)$ represents the correlation between X and future values of its driving noise. Accordingly, the correlation is zero. To compare the theoretical prediction with data, we generate a synthetic realisation of the process (4.37) as before. The cross-correlation function estimated with formula (4.36) (continuous green line) and the prediction from equation (4.45) (dashed black line) are shown in figure 4.2 (right).

4.2.6 Statistical significance of correlation functions

In subsequent parts of this manuscript, correlation functions are estimated from data. A common task in this case is to assess, with the help of statistical devices, whether or not the estimated values are significantly different from zero. The problem is usually formulated in terms of deciding whether a null hypothesis, named H_0, should be accepted or rejected in favour of an alternative hypothesis, named H_a. The choice of the null hypothesis determines the nature of the statistical test, as described below for two simple examples.

White noise test

In the simplest case, we consider a univariate time series X_n with $n = 1,...,N$, and formulate the hypothesis that X_n is a white noise process. Since the auto-correlation function of white noise is zero at all non-zero lags (von Storch and Zwiers, 1999), this is equivalent to:

$$H_0 : \rho_{XX}(\tau) = 0 \qquad \forall \tau \neq 0. \tag{4.46}$$

The alternative hypothesis H_a is that the auto-correlation function is non-zero at some non-negative lag, and therefore that X_n is not a white noise process. We estimate the auto-correlation ρ_{XX} with r_{XX} defined in equation (4.34). It can be proven (Tsay, 2013) that if X_n is a white noise process, then (i) the expectation value of r_{XX} is zero, and (ii) the variance of r_{XX} satisfies:

$$\mathrm{Var}[r_{XX}] \simeq \frac{1}{N}. \tag{4.47}$$

Therefore, at the 95% level of significance, the interval of acceptance of the null hypothesis H_0 is $\left[-2/\sqrt{N}, 2/\sqrt{N}\right]$ (this assumes that the distribution of r_{XX} is normal). Note that, with a 95% level of significance, H_0 will be rejected on average 5% of the times even if always true: the effect is called multiplicity (von Storch and Zwiers, 1999). We use the white noise test to check that the residuals of a model fit are normally distributed, see section 4.5.

Red noise test

The assumptions of the white noise test are too restrictive to cover all the cases we are interested in. A relevant example is the following: we consider a pair of time series (X_n, Y_n) where $n = 1, ..., N$, and estimate the cross correlation function $\rho_{XY}(\tau)$ with $r_{XY}(\tau)$, as defined in equation (4.36). The task is to decide whether, at a given lag τ, the estimated value $r_{XY}(\tau)$ is significantly different from zero. The assumptions of the white noise test do not apply if X_n and Y_n are not white noise processes, which can be easily ascertained by studying the respective auto-correlation functions. The rigorous way of addressing this problem is to apply the Ljung-Box test (Tsay, 2013), which is however fairly involved and to the author's best knowledge quite uncommon in the Southern Ocean's literature. Instead, we propose a simplified approach which is inspired by the white noise test. We formulate the null hypothesis H_0 that X_n and Y_n are independent auto-regressive processes of order 1. Numerical evidence (see figure 4.3) reveals that, under this assumption, (i) the distribution of r_{XY} is approximately normal, (ii) The expectation value of r_{XY} is approximately zero, and (iii) the variance of r_{XY} approximately satisfies:

$$\mathrm{Var}[r_{XY}] \simeq \frac{1}{N_e}, \tag{4.48}$$

where N_e is the effective number of degrees of freedom (Bretherton et al., 1999, Screen et al., 2009):

$$N_e = N \frac{1 - r_X r_Y}{1 + r_X r_Y}. \tag{4.49}$$

Here, r_X and r_Y are the lag-1 auto-correlation coefficients of X_n and Y_n respectively. Therefore, at the 95% level of significance, the interval of acceptance of the null hypothesis H_0 is $\left[-2/\sqrt{N_e}, 2/\sqrt{N_e}\right]$. Screen et al. (2009) used N_e to perform a two-tailed Student-t test in the context of a univariate linear regression problem. Their approach is more general, but it is rather unclear whether the underlying assumptions would hold in the bivariate problem we consider later in this work. Bretherton et al. (1999) note that the approximation provided by equation (4.49) is accurate for $r_X, r_Y << 1$, but our numerical experiments suggest that equation (4.48) holds even for $r_X, r_Y \simeq 1$. We use the red noise test in sections 4.5.1 and 4.5.2.

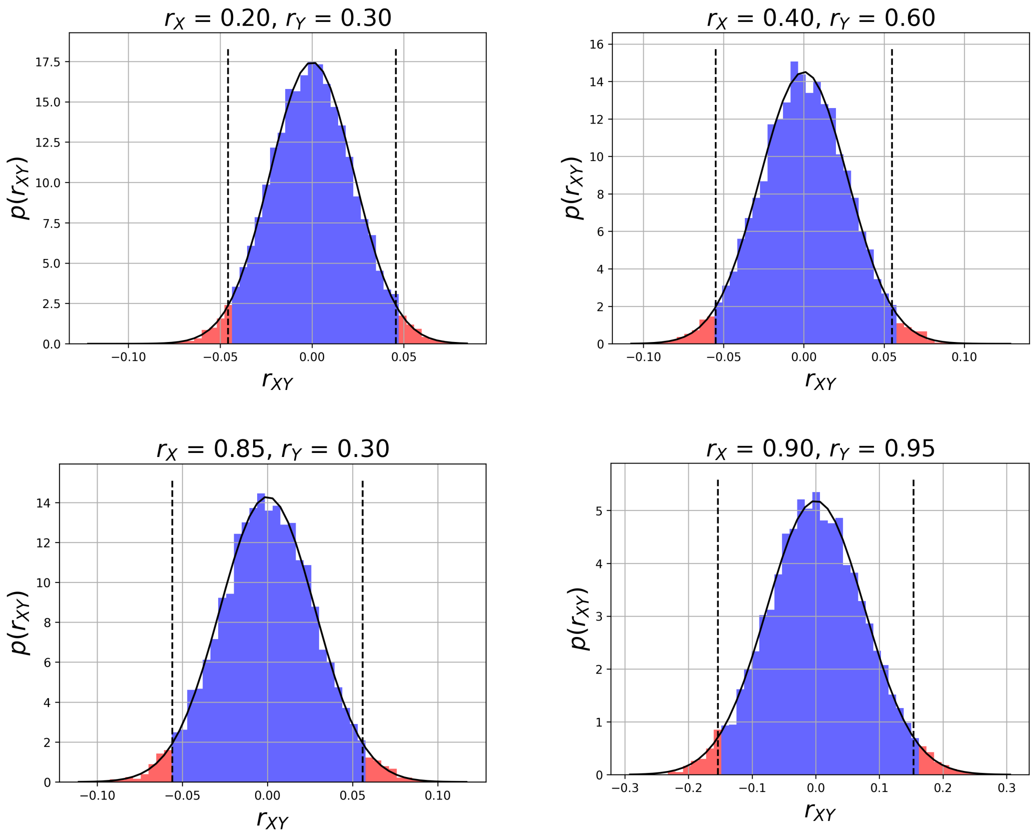

Figure 4.3: Numerical evidence supporting the results presented in the main text. For each couple (r_X, r_Y), we generated 10^4 realisations of two independent red noise processes, with $\Delta t = 5$, $N = 2160$ (corresponding to 30 years if the time unit is days), and σ is extracted from a uniform distribution with bounds zero and $\sqrt{\Delta t}$. The cross correlation function was estimated according to formula (4.36). The histograms show the sampled distribution of $r_{XY}(0)$ for different values of r_X and r_Y (see subplot titles). The distribution of $r_{XY}(\tau)$ is independent of τ (not shown), therefore we only consider the distribution of the lag-zero coefficient. Histogram bars are coloured in red if the associated bin edges are smaller than the 2.2 sample percentile or larger than the 97.8 sample percentile, and in blue otherwise. The blue region corresponds to the acceptance interval of the null hypothesis at 95% significance. The dashed black lines mark the corresponding prediction computed according to $\left[-2/\sqrt{N_e}, 2/\sqrt{N_e}\right]$. The continuous black line shows the zero mean gaussian distribution with variance $1/N_e$.

4.2.7 Linear oscillators

The AN model is a two-dimensional conservative oscillator. Upon linearisation, we obtain a two-dimensional linear conservative oscillator. If we explicitly account for the unrepresented degrees of freedom of the complete geophysical system, the equations are forced by stochastic noise. To guarantee the existence of stationary solutions, dissipation terms must also be included. The model we are looking for is thus a two-dimensional, stochastically forced, damped, linear oscillator. In this section, we discuss some of its important mathematical properties.

General case

In two dimensions, the most general SDE defining a linear oscillator is:

$$\mathrm{d}\mathbf{X} = \mathscr{A}\mathbf{X}\mathrm{d}t + \sigma \mathrm{d}\mathbf{W}, \tag{4.50}$$

where $\mathscr{A}$ and σ are 2×2 matrices, $\mathbf{X}$ and $\mathrm{d}\mathbf{W}$ are two-dimensional column vectors, $\mathrm{d}\mathbf{W}$ is Gaussian white noise, and certain conditions on $\mathscr{A}$ and σ, discussed immediately below, are assumed. The associated discrete stochastic process is obtained with the Euler-Maruyama finite differencing scheme:

$$\mathbf{X}_{n+1} = A\mathbf{X}_n + \sigma \mathrm{d}\mathbf{W}_{n+1}, \tag{4.51}$$

with $A = \mathbb{I}_2 + \Delta t \mathscr{A}$. Our assumptions are:

$$\sigma = \begin{bmatrix} \sigma_x & 0 \\ 0 & \sigma_y \end{bmatrix}, \tag{4.52}$$

with $\sigma_x, \sigma_y > 0$, and:

$$\mathscr{A} = \begin{bmatrix} \gamma_{xx} & k_{xy} \\ k_{yx} & \gamma_{yy} \end{bmatrix}, \tag{4.53}$$

with $\gamma_{xx}, \gamma_{yy} \leq 0$ and $k_{xy}k_{yx} \leq 0$. These properties ensure that the dynamical part of the process describes an oscillator, and that there exist stationary solutions. To understand why, consider the eigenvalues of the coefficient matrix $\mathscr{A}$:

$$\lambda = \frac{\gamma_{xx} + \gamma_{yy}}{2} \pm \left[\left(\frac{\gamma_{xx} - \gamma_{yy}}{2} \right)^2 + k_{xy}k_{yx} \right]^{\frac{1}{2}}. \tag{4.54}$$

The system is endowed with stationary solutions if the real part of λ is negative (Vatiwutipong and Phewchean, 2019). It is an oscillator if the imaginary part of the eigenvalues is non-zero in the limit $\gamma_{xx} = \gamma_{yy} \to 0$. Here, the second condition is automatically satisfied thanks to $k_{xy}k_{yx} \leq 0$. The first is guaranteed by combining the two assumptions $\gamma_{xx}, \gamma_{yy} \leq 0$ and $k_{xy}k_{yx} \leq 0$. Inspection of the eigenvalues reveals that the γ's represent damping terms, while the k's represent coupling terms. This becomes particularly clear when one individually investigates the limits $k \to 0$ and $\gamma \to 0$. When neither the k's nor the γ's are zero there are two cases, depending on whether the condition $(\gamma_{xx} - \gamma_{yy})^2 > 4k_{xy}k_{yx}$ is satisfied or not. The former case (with $(\gamma_{xx} - \gamma_{yy})^2 > 4k_{xy}k_{yx}$) defines the super-critical regime, when the imaginary part of λ is zero and oscillations are suppressed by damping. The latter (with $(\gamma_{xx} - \gamma_{yy})^2 < 4k_{xy}k_{yx}$) defines the sub-critical regime, where the imaginary part of λ is non-zero and oscillations are not completely suppressed. When the real part of λ is exactly zero the deterministic system is conservative, i.e. there is no dissipation but equilibrium solutions exist. The harmonic oscillator and the AN14 model are examples of deterministic conservative systems. A conservative system forced by Gaussian white noise does not enjoy stationary solutions.

Statistical properties of the stochastic oscillator can be equivalently investigated in the continuous or discrete formulation. Here we study the discrete equations, because they are analytically simpler and relate more directly to numerical simulations of the process. The expectation value of the process is $\mathbb{E}[\mathbf{X}_n] = \mathbf{0}$. The covariance matrix Σ^0, defined by equation (4.27), is computed by solving the matrix equation:

$$\Sigma^0 = A\Sigma^0 A^T + \Delta t \sigma \sigma^T. \tag{4.55}$$

This formula is obtained by substituting equation (4.51) into the definition of the covariance, equation (4.27). The Δt factor on the right hand side originates from the expectation value of the square of the Brownian motion increment. Under our set of assumptions, it can be shown that the equation above is equivalent to solving (von Storch and Zwiers, 1999):

$$\left(\mathbb{I}_3 - \begin{bmatrix} \gamma_{xx}^2 & k_{xy}^2 & 2\gamma_{xx}k_{xy} \\ k_{yx}^2 & \gamma_{yy}^2 & 2\gamma_{yy}k_{yx} \\ \gamma_{xx}k_{yx} & k_{xy}\gamma_{yy} & \gamma_{xx}\gamma_{yy} + k_{xy}k_{yx} \end{bmatrix}\right) \begin{bmatrix} \Sigma_{xx}^0 \\ \Sigma_{yy}^0 \\ \Sigma_{xy}^0 \end{bmatrix} = \Delta t \begin{bmatrix} \sigma_{xx}^2 \\ \sigma_{yy}^2 \\ 0 \end{bmatrix}. \tag{4.56}$$

The time-lagged covariance matrix, defined by equation (4.25), is computed similarly, yielding:

$$\Sigma_\tau = \Sigma^0 (A^T)^\tau, \tag{4.57}$$

where $\tau \in \mathbb{Z}$ is the lag in time steps. The correlation functions can be obtained from the lagged covariance matrix as described in section 4.2.3. Finally, the marginal Probability Distribution Functions (PDFs) of X and Y are given by:

$$p(x) \approx \frac{1}{\sqrt{2\pi\Sigma_{xx}^0}} \exp\left(-\frac{x^2}{2\Sigma_{xx}^0}\right) \tag{4.58}$$

$$p(y) \approx \frac{1}{\sqrt{2\pi\Sigma_{yy}^0}} \exp\left(-\frac{y^2}{2\Sigma_{yy}^0}\right), \tag{4.59}$$

where we have made the supplementary assumption, justified a posteriori, that $\Sigma_{xy}^0 \approx 0$.

Equations (4.50) and (4.51) define the model we shall refer to in the subsequent analysis of MITgcm data. Up to this point, we have concentrated on its mathematical properties: in order to shed some light on its dynamics, we briefly explore a few special cases.

Special case: the harmonic oscillator

The quintessential physical model: with our notation, it corresponds to $\gamma_{xx} = \gamma_{yy} = 0$ and $k_{xy} = -k_{yx} = \omega > 0$:

$$\frac{\mathrm{d}}{\mathrm{d}t} X = \omega Y \tag{4.60}$$

$$\frac{\mathrm{d}}{\mathrm{d}t} Y = -\omega X. \tag{4.61}$$

The harmonic oscillator is a conservative system, i.e. $\Re(\lambda) = 0$, therefore the associated SDE does not have equilibrium solutions (Vatiwutipong and Phewchean, 2019). The ODE solutions are well known, and satisfy:

$$X^2 + Y^2 = R^2, \tag{4.62}$$

with R a constant. Thus, trajectories in phase space are circles of radius R.

Special case: the elongated harmonic oscillator

The elongated harmonic oscillator is a linear oscillator whose trajectories in phase space are ellipses. It corresponds to (Yano et al., 2020) $\gamma_{xx} = \gamma_{yy} = 0$, $k_{xy} = \alpha\omega$ and $k_{yx} = \omega/\alpha$, with α and ω positive real numbers:

$$\frac{\mathrm{d}}{\mathrm{d}t}X = \alpha\omega Y \tag{4.63}$$

$$\frac{\mathrm{d}}{\mathrm{d}t}Y = -\frac{\omega}{\alpha}X\,. \tag{4.64}$$

The elongated harmonic oscillator is a conservative system. Following Yano et al. (2020), solutions satisfy the equation:

$$\left(\frac{X}{A}\right)^2 + \left(\frac{Y}{B}\right)^2 = 1, \tag{4.65}$$

with A and B positive real numbers such that $\alpha = A/B$. Note that the linearised AN14 model, introduced in section 4.1.2, is formally equivalent to an elongated harmonic oscillator with $\omega = \sqrt{2F}$ and $\alpha = 1/\sqrt{2F}$.

Special case: the elongated, rotated, harmonic oscillator

Starting from the elongated harmonic oscillator and with a passive rotation of the coordinate axes, one can define a linear oscillator whose trajectories in phase space are ellipses rotated by an angle θ with respect to the horizontal (Yano et al., 2020). The change of coordinates is $\hat{\mathbf{X}} = R_{-\theta}\mathbf{X}$, where R_θ is the 2×2 rotation matrix (the rotation is counter-clockwise when θ is positive) and $\hat{\mathbf{X}}$ are the rotated axes:

$$\hat{X} = \cos\theta X + \sin\theta Y \tag{4.66}$$

$$\hat{Y} = -\sin\theta X + \cos\theta Y\,. \tag{4.67}$$

As anticipated, solutions in the $(\hat{X},\hat{Y})$ coordinate system satisfy the equation of an ellipse rotated by an angle θ with respect to the horizontal. In terms of the rotated variables, the evolution equations read:

$$\frac{\mathrm{d}}{\mathrm{d}t}\hat{X} = \mu_1\hat{X} + \eta_1\hat{Y} \tag{4.68}$$

$$\frac{\mathrm{d}}{\mathrm{d}t}\hat{Y} = -\eta_2\hat{X} - \mu_2\hat{Y}\,, \tag{4.69}$$

with:

$$\mu_1 = \mu_2 = \omega\left(\alpha - \frac{1}{\alpha}\right)\sin\theta\cos\theta \tag{4.70}$$

$$\eta_1 = \omega\left(\alpha\cos^2\theta + \frac{1}{\alpha}\sin^2\theta\right) \tag{4.71}$$

$$\eta_2 = \omega\left(\alpha\sin^2\theta + \frac{1}{\alpha}\cos^2\theta\right)\,. \tag{4.72}$$

Up to a minus sign, this is the model Yano et al. (2020) tested their data against. Note that, despite appearances, the elongated, rotated, harmonic oscillator is still a conservative system because $\Re(\lambda) = \mu_1 - \mu_2 = 0$. Thus, there are no equilibrium solutions for the associated SDE problem.

Remark

Anticipating on our results, we will show that eddy-mean flow oscillations in the Southern Ocean are characterised by a balance between stochastic fluctuations and dissipation. Therefore, rather than the conservative oscillator proposed by Yano et al. (2020), we will fit our data with the forced-dissipative linear system of equation (4.51). Note that this approach involves little loss of generality, as near-conservative solutions can still be obtained from the fit in the limit of weak stochastic forcing and dissipation.

4.2.8 Kernel averaging

Kernel averaging is a mathematical technique which allows to reduce the amount of noise contained in a sequence of observations of a fluctuating dynamical system. In the following, we assume that we are given N observations (X_i, Y_i) of a dynamical system of the form (4.11) and (4.12): we use kernel averaging to investigate the mean structure of trajectories in the phase space, and to gather information on the deterministic part of the dynamics. The index i runs $i = 1, ..., N$, and observations are separated by a time interval Δt. Our presentation follows Marcheggiani et al. (2022), Novak et al. (2017), Yano et al. (2020), to which we refer the reader for the original discussion.

The kernel averaging technique in phase space consists in a weighted average, where the weight attributed to each observation is a measure of the distance between the observation itself and the point in the phase space where the average is evaluated. For instance, the phase space density at point (x_0, y_0) is defined by:

$$\mu(x_0, y_0) = \sum_{i=1}^{N} K(x_0 - X_i, y_0 - Y_i), \tag{4.73}$$

where $K(x, y)$ is the averaging kernel. Intuitively, $\mu(x_0, y_0)$ counts the number of observations that fall in the vicinity of the point (x_0, y_0). The concept of vicinity in the phase space is quantified by $K(x, y)$, which we take of the form:

$$K(x, y) = \exp\{-\frac{1}{2}\left[\left(\frac{x}{h_x}\right)^2 + \left(\frac{y}{h_y}\right)^2\right]\}, \tag{4.74}$$

so that distance is evaluated according to a bivariate Gaussian distribution. The parameters h_x and h_y control the width of the Gaussian filter. Note that observations that fall in the vicinity of a given point (x_0, y_0) may be separated by a long interval of time. In this sense, phase space kernel averaging allows to concentrate on the dynamical similarity between observations, regardless of their proximity in time. The average value of an arbitrary function $q(X, Y)$ at point (x_0, y_0) is defined by:

$$\tilde{q}(x_0, y_0) = \frac{1}{\mu(x_0, y_0)} \sum_{i=1}^{N} K(x_0 - X_i, y_0 - Y_i) q(X_i, Y_i), \tag{4.75}$$

A special case is when kernel averaging is used to reconstruct the mean phase space velocity. The raw phase space velocity vector can be estimated with the finite differencing formula:

$$u_{i+\frac{1}{2}} = \frac{X_{i+1} - X_i}{\Delta t} \tag{4.76}$$
$$v_{i+\frac{1}{2}} = \frac{Y_{i+1} - Y_i}{\Delta t}, \tag{4.77}$$

with $i = 1, ..., N-1$. Note that the velocity is defined on the interface points:

$$X_{i+\frac{1}{2}} = \frac{X_i + X_{i+1}}{2} \tag{4.78}$$
$$Y_{i+\frac{1}{2}} = \frac{Y_i + Y_{i+1}}{2}. \tag{4.79}$$

Thus, the kernel-averaged velocity at point (x_0, y_0) is:

$$\tilde{u}(x_0, y_0) = \frac{\sum_{i=1}^{N-1} K_{i+\frac{1}{2}} u_{i+\frac{1}{2}}}{\sum_{i=1}^{N-1} K_{i+\frac{1}{2}}} \tag{4.80}$$
$$\tilde{v}(x_0, y_0) = \frac{\sum_{i=1}^{N-1} K_{i+\frac{1}{2}} v_{i+\frac{1}{2}}}{\sum_{i=1}^{N-1} K_{i+\frac{1}{2}}}, \tag{4.81}$$

where $K_{i+\frac{1}{2}}$ is short for $K(x_0 - X_{i+\frac{1}{2}}, y_0 - Y_{i+\frac{1}{2}})$. The width of the Gaussian filter, controlled by the parameters h_x and h_y, can be defined based on the variability of the observations in the x and y directions. To this end, we define the mean phase space location of the observations as:

$$\overline{X} = \frac{1}{N-1} \sum_{i=1}^{N-1} X_{i+\frac{1}{2}} \tag{4.82}$$
$$\overline{Y} = \frac{1}{N-1} \sum_{i=1}^{N-1} Y_{i+\frac{1}{2}}. \tag{4.83}$$

The amplitude of fluctuations around the mean location is measured by:

$$s_x = \frac{1}{N-1} \left[\sum_{i=1}^{N-1} (\overline{X} - X_{i+\frac{1}{2}})^2 \right]^{\frac{1}{2}} \tag{4.84}$$
$$s_y = \frac{1}{N-1} \left[\sum_{i=1}^{N-1} (\overline{Y} - Y_{i+\frac{1}{2}})^2 \right]^{\frac{1}{2}}, \tag{4.85}$$

which allows to define the parameters h_x and h_y as fractions of the phase space occupied by data, that is: $h_x = f_x s_x$ and $h_y = f_y s_y$, where f_x and f_y are typically smaller than 1. Following Marcheggiani et al. (2022), Yano et al. (2020), we set $f_x = f_y = 0.35$ unless otherwise specified so that the Gaussian filter is sufficiently large to remove small scale fluctuations without obscuring the large scale dynamics. See Novak et al. (2017) for a thorough discussion of the effect of the Gaussian filter's size on the reconstructed phase space trajectories. The kernel-averaged velocity $(\tilde{u}, \tilde{v})$ is usually evaluated on a discrete grid in the phase space. Here, we adopt an $N_x \times N_y$ uniform rectangular grid, with $N_x = N_y = 30$ and grid spacing:

$$\Delta x = \frac{\max_i X_{i+\frac{1}{2}} - \min_i X_{i+\frac{1}{2}}}{N_x - 1}(1+2\alpha) \tag{4.86}$$

$$\Delta y = \frac{\max_i Y_{i+\frac{1}{2}} - \max_i Y_{i+\frac{1}{2}}}{N_y - 1}(1+2\alpha), \tag{4.87}$$

where $\alpha = 0.1$ is a parameter that controls how tightly the data-populated region is bounded by the grid.

It is possible to evaluate the significance of the kernel-averaged trajectories. Yano et al. (2020) introduced the following metrics: (i) the statistical significance, which tends to be large in densely populated regions of the phase space, and measures how reliable the estimate of the averaged trajectories is, and (ii) the signal to noise ratio, which measures the amount of fluctuations of the system around the averaged trajectories and tends to be large in scarcely populated regions. The physical interpretation is that a significance tradeoff should be expected: in densely populated regions the actual trajectories of the system depart markedly from the mean streamlines due to the elevated levels of noise (Marcheggiani et al., 2022, Yano et al., 2020). One the other hand, it is precisely in these regions that mean streamlines can be estimated with the highest statistical significance. Throughout this work, however, we prefer to adopt more conventional measures of statistical significance, such as those illustrated in section 4.2.6.

4.2.9 Kernel averaging: examples

Before turning to more complicated cases, we illustrate the phase space kernel averaging technique with two simple examples.

Van der Pol Oscillator

The Van der Pol Oscillator is a completely deterministic dynamical system defined by:

$$\frac{\mathrm{d}}{\mathrm{d}t}X = \gamma(1-Y^2)X - Y \tag{4.88}$$

$$\frac{\mathrm{d}}{\mathrm{d}t}Y = X\,. \tag{4.89}$$

The system is nonlinear and conservative: trajectories in the phase space are closed orbits. We demonstrate that, in the limiting case of a dynamical system with no superimposed stochastic noise, the phase space kernel averaging procedure does not distort the structure of the phase space diagram. We integrate the Van der Pol equations for $\gamma = 3$ with the four steps Runge-Kutta method, with $\Delta t = 0.02$ and up to a final time of $t_f = 60$. The raw phase space diagram is shown in figure 4.4 (left). The reconstructed phase space diagram, obtained with the procedure described in section 4.2.8 and for $f_x = f_y = 0.15$, is shown in figure 4.4 (right). Kernel-averaged streamlines (oriented black lines) are clustered around the deterministic trajectory, and correctly reproduce the overall structure of the phase space diagram. The width of the Gaussian kernel, shown by the green ellipse in the top left corner, controls the extent of the phase space area used for the smoothing filter, and is associated to the amount of blurring applied to the deterministic trajectory. The width of the streamlines is proportional to the phase space speed. Consistently, regions of velocity convergence are associated with the largest values of the data density $\mu(x,y)$ (colours), and vice-versa.

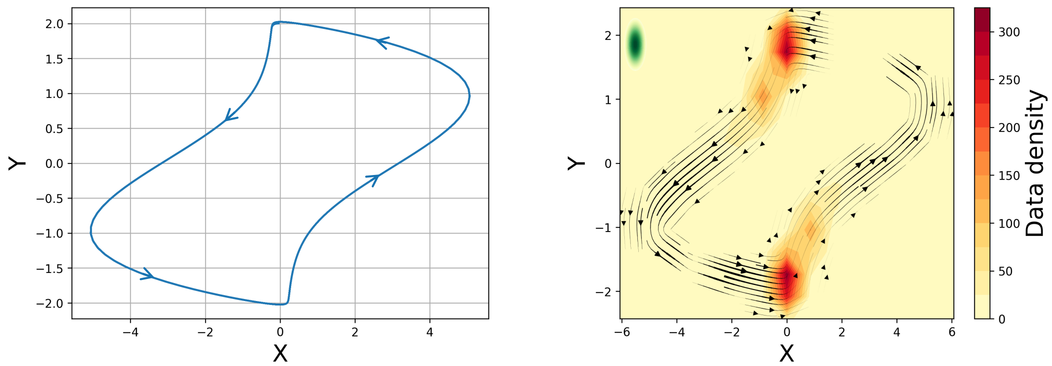

Figure 4.4: Left: raw phase space diagram for the Van der Pol oscillator with $\gamma = 3$. Arrows indicate the sense of the circulation along the trajectory. Right: reconstructed phase space trajectories with $f_x = f_y = 0.15$. Kernel-averaged streamlines are represented by oriented black lines. The width of the lines is proportional to the phase space speed. The estimated data density $\mu(x,y)$ is shown in colour shades. The width of the Gaussian kernel is represented by the green ellipse in the top left corner.

Lorenz system

The Lorenz system (Lorenz, 1963) is the deterministic dynamical system defined by:

$$\frac{\mathrm{d}}{\mathrm{d}t}X = \sigma(Y - X) \tag{4.90}$$

$$\frac{\mathrm{d}}{\mathrm{d}t}Y = X(\rho - Z) - Y \tag{4.91}$$

$$\frac{\mathrm{d}}{\mathrm{d}t}Z = XY - \beta Z\,. \tag{4.92}$$

We set the problem parameters to $\sigma = 10$, $\rho = 28$, and $\beta = 8/3$. In this regime, the equations have chaotic solutions, and almost all initial conditions tend to a strange attractor. The Lorenz system is three-dimensional, but we only consider two-dimensional sections of its phase space diagram. We use this example to test the kernel averaging technique in the case of a system endowed with deterministic chaos, when one degree of freedom is not directly represented in the phase space. We integrate the Lorenz equations with the four steps Runge-Kutta method, with $\Delta t = 0.01$ and up to a final time of $t_f = 40$. The YX section of the raw phase space

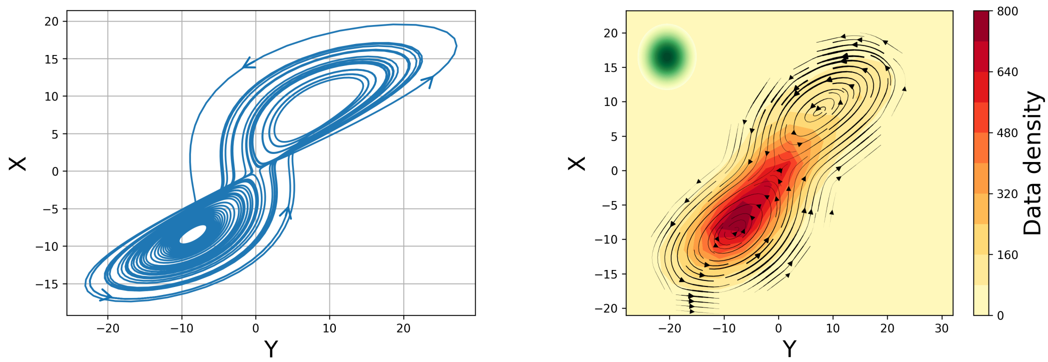

Figure 4.5: Left: raw YX phase space section for the Lorenz system with $\sigma = 10$, $\rho = 28$, and $\beta = 8/3$. Arrows indicate the sense of the circulation along the trajectory. Right: reconstructed phase space trajectories with $f_x = f_y = 0.35$. Kernel-averaged streamlines are represented by oriented black lines. The width of the lines is proportional to the phase space speed. The estimated data density $\mu(x,y)$ is shown in colour shades. The width of the Gaussian kernel is represented by the green ellipse in the top left corner.

diagram is shown in figure 4.5 (left). The reconstructed phase space diagram, obtained with $f_x = f_y = 0.35$, is shown in figure 4.4 (right). The width of the Gaussian kernel is represented by the green ellipse in the top left corner. The kernel-averaged trajectories (oriented black lines) reproduce the structure of the strange attractor satisfactorily (for example, the change of orientation of the quasi-periodic orbits at $(0,0)$ is correctly captured), and the estimated data density (colours) is consistent with that of the raw phase space diagram. The width of the averaged streamlines clearly demonstrates that different sections of the attractor are characterised by different average speed, a feature not apparent from the analysis of the raw phase space diagram only.

4.3 Data

4.3.1 Dataset

In this chapter, we analyse data from the CTRL dataset of the idealised channel. We use a total of 33 years of simulations, with 5-day averaged model output. Our working hypothesis is that eddy-mean flow interaction can be described by a simplified model of the form (4.51). Here, we detail how data from the idealised channel configuration is used to test this assumption. For a thorough discussion of the channel's setup and physics, we refer the reader to chapter 3.

4.3.2 Definition of the dynamical variables

The first task we undertake is to introduce the dynamical variables describing mean flow and eddy activity in the oceanic case. In the atmospheric case, AN14 proposed to use the Eady growth rate to measure mean flow and the eddy heat flux to measure eddy activity. With minor adjustments, these choices are relevant to the oceanic case too, as we now illustrate.

The Eady growth rate is an inverse time scale for the formation of eddies via baroclinic instability (see the theory in section 2.3). In the Southern Ocean, it is defined by (Williams et al., 2007):

$$\omega = -0.31 f \frac{\partial_z \overline{u}}{N}. \tag{4.93}$$

Here, the bar denotes zonal average, f is the Coriolis parameter, and N is the buoyancy frequency (note that N is a time-dependant variable here). The unit measure is $[\omega] = \text{day}^{-1}$. The Coriolis parameter is negative in the Southern Hemisphere and the vertical shear of zonal velocity is generally positive in the ACC region, therefore the Eady growth rate ω is a positive quantity. Note that the Eady growth rate is related to baroclinicity (the meridional gradient of buoyancy) through the thermal wind relation. Substituting equation (3.23) into the definition (4.93), we obtain:

$$\omega = 0.31\frac{\partial_y \overline{b}}{N}, \tag{4.94}$$

or:

$$\omega = -0.31Ns, \tag{4.95}$$

where $s = -\partial_z \overline{b}/\partial_y \overline{b}$ is the slope of the isopycnals (which is negative on average in the ACC region). We know from section 2.4.3 that the wind-powered Eulerian circulation and the eddy-induced circulation have competing effects on s, respectively acting to steepen and flatten isopycnals. The physical interpretation is that s is a measure of the potential energy stored in the mean flow, and available for consumption by baroclinic eddies (the larger the reservoir of available potential energy, the faster the growth of eddies). Thus, equation (4.95) highlights that the Eady growth rate in turn constitutes a metric for the mean flow. Figure 4.6 (left) shows the time-mean meridional profiles of ω in the top 1000 m of the idealised channel, where the time mean is taken over a time interval of 33 years. The spatial average of the time-mean Eady growth rate over the rectangular domain shown in figure 4.6 (left) is 0.052 days^{-1} (corresponding to an average growth period of $\omega^{-1} \approx 20$ days, broadly consistent with the estimate given by Williams et al. (2007)).

The quantity adopted by AN14 to measure eddy activity is the eddy heat flux, which only differs from the eddy buoyancy flux by a dimensional constant. The latter quantity, however, is more pervasive in the Southern Ocean literature considered for this work (see for example section 2.4.3 on TEM theory), and is therefore the preferred choice here. We define the zonal-mean meridional eddy buoyancy flux as:

$$\mathscr{F}_y = \overline{v^+ b^+}, \tag{4.96}$$

where the bar denotes zonal mean and $^+$ deviations from time average. The unit measure is $[\mathscr{F}_y] = J/\mathrm{Kg}\,\mathrm{day}$. This quantity should be contrasted with $\overline{v'b'}$, where the prime denotes deviations from zonal average. For the idealised channel, we found that the results presented in subsequent sections depend only marginally on the choice of the averaging operator (time mean, zonal mean, or time and zonal

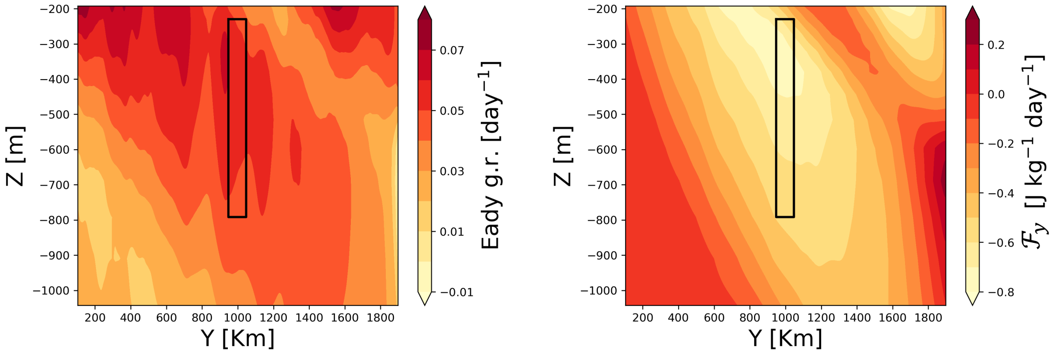

Figure 4.6: Left: time-mean meridional profiles of Eady growth rate. Right: time-mean meridional profile of zonal-mean eddy buoyancy flux. The rectangular domains show the region used to take spatial averages.

mean) with respect to which deviations are taken (not shown). This is not unexpected as the idealised channel's configuration is zonally symmetric, and zonally averaged fields approximately coincide with time averaged fields. The difference may become important in situations where the model setup is not zonally symmetric and, of course, in the real ocean. Our definition places the emphasis on the interaction between mean flow and transient eddies, as opposed to topographically driven standing meanders. We will see in chapter 5 how it can be generalised to a non-zonally symmetric geometry. Figure 4.6 (right) shows the time-mean meridional profile of the zonal-mean eddy buoyancy flux in the top 1000 m of the idealised channel, where the time mean is taken over a time interval of 33 years. Over most of the domain $\mathscr{F}_y$ is negative, which is an expression of the fact that eddies transfer heat poleward (the negative meridional direction in the Southern Hemisphere). The spatial average of time-mean $\mathscr{F}_y$ over the rectangular domain shown in figure 4.6 (right) is $-0.63 J/\mathrm{Kg}\,\mathrm{day}$.

The dynamical variables ω and $\mathscr{F}_y$ are related to the physical quantities discussed in the theoretical model of eddy saturation of Marshall et al. (2017). Under suitable assumptions, it is possible to exploit this connection to express their (average) values in terms of model parameters such as the surface wind stress or the eddy dissipation parameter, as we now briefly demonstrate. Consider the eddy buoyancy flux $\mathscr{F}_y$ first: we shall show that this quantity is proportional to the eddy energy E

defined by Marshall et al. (2017). From the definition of quasi-Stokes streamfunction (see chapter 2), we have:

$$\mathscr{F}_y = N^2 \psi^* \,. \tag{4.97}$$

The quasi-stokes streamfunction, in turn, is related to the eddy form stress as follows. Firstly, recall that:

$$\psi_{\text{res}} = \overline{\psi} + \psi^* \,. \tag{4.98}$$

Now, $\overline{\psi} = -\tau_w/\rho_0 f$ from zonal momentum balance. The Coriolis parameter f is negative, therefore:

$$\psi_{\text{res}} = \frac{\tau_w}{\rho_0 |f|} + \psi^* \,, \tag{4.99}$$

which we rewrite as:

$$\tau_w = -\rho_0 |f| \psi^* + \rho_0 |f| \psi_{\text{res}} \,. \tag{4.100}$$

Similarly to Marshall et al. (2017), we define the eddy form stress S as:

$$S = -\rho_0 |f| \psi^* \,. \tag{4.101}$$

Physically, the right hand side is (minus) the Coriolis torque on the eddy-induced meridional circulation. Note that, with this definition, S is a positive quantity. The above reads:

$$\tau_w = S + \rho_0 |f| \psi_{\text{res}} \,, \tag{4.102}$$

which, up to the convention on sign, is equation (1) in Marshall et al. (2017). From their equation (2), the eddy energy E relates to the eddy form stress as:

$$S = \frac{\alpha_1 |f|}{N} E \,, \tag{4.103}$$

where α_1 is a non-dimensional constant. Substituting equations (4.97) and (4.101) into the above gives:

$$\mathscr{F}_y = -\frac{\alpha_1 N}{\rho_0} E \,. \tag{4.104}$$

Scaling law (3) in Marshall et al. (2017) then tells us that $\mathscr{F}_y$ may be expected to increase linearly with the surface wind stress. Physically, this is an expression of the fact that stronger winds fuel a more intense eddy field, and the eddy fluxes scale accordingly.

Next, we demonstrate that, under the same assumptions of Marshall et al. (2017), the Eady growth rate is proportional to the eddy dissipation parameter λ. From the eddy energy balance, equation (6) in Marshall et al. (2017), we obtain:

$$\partial_z u = \frac{1}{\alpha_2} \frac{\lambda N}{|f|}, \tag{4.105}$$

whence:

$$\omega = \frac{0.31\lambda}{\alpha_2}. \tag{4.106}$$

Here, α_2 is a second non-dimensional parameter. If we use the value $\alpha_2 = 0.61$ appropriate for linear instability (Marshall et al., 2017), this simplifies to:

$$\omega \approx \frac{\lambda}{2}. \tag{4.107}$$

The assumptions involved in the derivation above are quite restrictive (e.g. uniform stratification and shear), and it is unclear to what extent they apply quantitatively to our model setup. Furthermore, the eddy dissipation parameter λ has units s^{-1}, and thus different dimensionality from the bottom drag r_b, which hinders a direct evaluation of equation (4.107). Nevertheless, equations (4.104) and (4.107) establish a clear, qualitative link between our variables and the theoretical model of eddy saturation of Marshall et al. (2017), and offer a roadmap to help interpret the equilibrium sensitivity of ω and $\mathscr{F}_y$ to changes in the numerical model parameters. We highlight, however, that the simplified theoretical model developed in this chapter is intended to capture the variability of the dynamical variables rather than to predict their mean values, and it is thus not straightforward to communicate the added knowledge supplied by equations (4.104) and (4.107) to our results. We return to this point in the discussion of section 4.5.3.

4.3.3 Spatial averaging

Simplified models of the form (4.51) require the dynamical variables to depend on time only. Here, the Eady growth rate and the zonal-mean eddy buoyancy flux are presently functions of y, z and time. In order to obtain a suitable pair of dynamical variables, we average ω and $\mathscr{F}_y$ over a rectangular domain Ω in the meridional plane:

$$\tilde{\omega}(t) = \frac{1}{V(\Omega)} \iint_{\Omega} \mathrm{d}y\mathrm{d}z\, \omega(y,z,t) \tag{4.108}$$

$$\tilde{\mathscr{F}}_y(t) = \frac{1}{V(\Omega)} \iint_{\Omega} \mathrm{d}y\mathrm{d}z\, \mathscr{F}(y,z,t)\,, \tag{4.109}$$

where:

$$V(\Omega) = \iint_{\Omega} \mathrm{d}y\mathrm{d}z. \tag{4.110}$$

The rectangular domain Ω is shown in figure 4.6 (continuous black lines). Vertically, it extends from $z = -792$ m to $z = -230$ m, so that the vertical average is computed below the mixed layer, where vertical gradients become vanishingly small, and within the top 1000 m (roughly, above the thermocline depth). Meridionally, it is located at the centre of the domain (where the zonal flow and the associated baroclinic activity is most intense) and is 100 km wide, comparable to the width of an individual ACC jet. We found that, while quantitative details may change, the results presented below do not depend critically on the exact size of the averaging domain or on its precise location (see section 4.6 at the end of the chapter for an overview). The quality of the agreement between the simplified model and the data, however, tends to deteriorate in the opposite limits of very large ($\approx$ 1000 km) or very small (a few grid points) meridional width of the domain. We speculate that in the former case the domain is sufficiently large to include statistics from multiple ACC jets, possibly leading to a partial cancellation of the signal. The latter case is more difficult to interpreter, but also peripheral to the main scope of this work as model (4.51) is not meant to provide an accurate representation of pointwise evolution of geophysical variables. In view of this, these limits are not explored further. The last step to obtain X and Y is to remove the time mean from the time series:

$$X(t) = \tilde{\omega}(t) - \frac{1}{T}\int_0^T \mathrm{d}t\, \tilde{\omega}(t) \tag{4.111}$$

$$Y(t) = \tilde{\mathscr{F}}_y(t) - \frac{1}{T}\int_0^T \mathrm{d}t\, \tilde{\mathscr{F}}_y(t)\,, \tag{4.112}$$

where $T = 33$ years. The dynamical variables are computed using 5-day averaged model output, which yields the discrete bivariate time series $\mathbf{X}_n = (X_n, Y_n)$, where $n = 1,...,N$, $N = 2376$, and $\Delta t = 5$ days.

4.4 Methods

We consider the following problem: we have N observations of two variables X and Y, $\mathbf{X}_n = (X_n, Y_n)$ for $n = 1, ..., N$. The observations are uniformly spaced by a known time interval Δt. We hypothesise that the time series $\mathbf{X}_n$ is a realisation of the linear stochastic process:

$$\mathbf{X}_{n+1} = A\mathbf{X}_n + \sigma \mathrm{d}\mathbf{W}_{n+1}, \tag{4.113}$$

where A and σ satisfy the conditions outlined in section 4.2.7, but are otherwise undetermined. The task we undertake is (i) to compute the best estimate for the unknown parameters A and σ, and (ii) to evaluate the goodness of the fit. This is a standard problem in time series analysis: the method of solution which we adopt follows the guidelines set out in von Storch and Zwiers (1999), and is based on the Yule-Walker equations. We introduce the Yule-Walker equations in the simpler case of a univariate process first, and from there generalise to the multivariate case.

4.4.1 Univariate Yule-Walker equations

The Yule-Walker equations are a set of relations that connect the parameters of an auto-regressive process with its lagged correlation coefficients. In this section, we study the simple case of a univariate, zero-mean autoregressive process of order 1:

$$X_{n+1} = \alpha_1 X_n + Z_{n+1}, \tag{4.114}$$

with $0 < \alpha_1 < 1$ and $Z_{n+1} = \sigma \mathrm{d}W_{n+1}$. The free model parameters to be estimated are α_1 and σ. In order to estimate α_1, we multiply both sides of the equation by X_n and take the expectation value:

$$\mathbb{E}[X_n X_{n+1}] = \alpha_1 \mathbb{E}[X_n X_n]. \tag{4.115}$$

The noise term drops out because X_n and Z_{n+1} are uncorrelated. The left hand side is the lag-1covariance, the right hand side is the lag-0 covariance multiplied by α_1:

$$\mathrm{Cov}_{XX}(1) = \alpha_1 \mathrm{Cov}_{XX}(0), \tag{4.116}$$

or:

$$\alpha_1 = \rho_{XX}(1). \tag{4.117}$$

The Yule-Walker equations assume a very simple form in this idealised example. The same result could have been obtained by substituting $\tau = 1$ into equation (4.42), but the procedure shown above is more general and applies to autoregressive processes of higher order as well. The lag-1 autocorrelation $\rho_{XX}(1)$ is estimated from the data, and yields the best estimate for α_1. The noise amplitude σ can be estimated from the analysis of the residuals, as explained in section 4.4.3 below.

4.4.2 Multivariate Yule-Walker equations

We now consider the multivariate case of equation (4.51). The model parameters to be estimated are the 2×2 matrix of coefficients A and the 2×2 noise amplitude matrix σ. The general procedure to compute the Yule-Walker equation was illustrated in the previous section, therefore here we substitute $\tau = 1$ into equation (4.25) directly, obtaining:

$$\Sigma(1) = \Sigma^0 A^T. \tag{4.118}$$

The covariance matrix is invertible, therefore:

$$A = \Sigma(1)^T (\Sigma^0)^{-T}, \tag{4.119}$$

which is the matrix equivalent of equation (4.117). It is convenient to write the above in terms of the lagged correlation functions:

$$A = \begin{bmatrix} \Sigma^0_{XX}\rho_{XX}(1) & \left(\Sigma^0_{XX}\Sigma^0_{YY}\right)^{\frac{1}{2}}\rho_{XY}(1) \\ \left(\Sigma^0_{XX}\Sigma^0_{YY}\right)^{\frac{1}{2}}\rho_{YX}(1) & \Sigma^0_{YY}\rho_{YY}(1) \end{bmatrix}^T (\Sigma^0)^{-T}. \tag{4.120}$$

The lag-0 covariance matrix and the lag-1 correlation functions are estimated from data, which yields the estimate of A. The noise amplitude matrix σ is estimated from the residuals, see section 4.4.3.

4.4.3 Goodness of fit

We assume that the estimate $\hat{A}$ of the matrix A has been computed with the Yule-Walker method described above. In order to evaluate the goodness of the fit, we define:

$$\mathbf{x}_n = \mathbf{X}_n \tag{4.121}$$

$$\mathbf{y}_n = \mathbf{X}_{n+1}\,, \tag{4.122}$$

for $n = 1,...,N-1$. Here, $\mathbf{x}_n$ and $\mathbf{y}_n$ are analogous to, respectively, the independent and dependent variables in a standard curve fitting problem. The values of $\mathbf{y}$ predicted by the model are:

$$\hat{\mathbf{y}}_n = \hat{A}\mathbf{x}_n, \tag{4.123}$$

for $n = 1,...,N-1$. The residuals, or errors, are the differences between the observed and predicted values of $\mathbf{y}$:

$$\mathbf{e}_n = \mathbf{y}_n - \hat{\mathbf{y}}_n. \tag{4.124}$$

The sum of square errors is the quantity defined by:

$$\mathbf{SSE} = \sum_{n=1}^{N-1} \mathbf{e}_n^T \mathrm{diag}(\mathbf{e}_n), \tag{4.125}$$

where $\mathrm{diag}(\mathbf{e}_n)$ is the 2×2 diagonal matrix whose diagonal elements are the components of $\mathbf{e}_n$. Note that, component-wise:

$$\mathbf{SSE}_j = \sum_{n=1}^{N-1} e_{nj}^2, \tag{4.126}$$

with $j = 1,2$. Another important measure of variability is:

$$\mathbf{SSY} = \sum_{n=1}^{N-1} \mathbf{y}_n^T \mathrm{diag}(\mathbf{y}_n). \tag{4.127}$$

The coefficient of determination R^2 is obtained combining $\mathbf{SSE}$ and $\mathbf{SSY}$. Component-wise, we have:

$$R_j^2 = 1 - \frac{SSE_j}{SSY_j}, \tag{4.128}$$

for $j = 1,2$. The coefficient of determination R^2 is valued between 0 and 1. It can be interpreted as the fraction of the variability in the values of $\mathbf{y}_j$ which is explained by the variability in the values of $\mathbf{x}_j$ (Ross, 2014). Thus, the coefficient of determination is a measure of the goodness of the fit. Finally, we can estimate the noise amplitude matrix σ from the analysis of the residuals. It can be shown that if the linear model is correct, then the errors are normally distributed:

$$\bar{\mathbf{e}}_j = \frac{\mathbf{e}_j}{\sqrt{\frac{\mathbf{SSE}_j}{N-3}}} \sim N(0,1), \tag{4.129}$$

where $j = 1,2$ and $N(0,1)$ is a mean zero normal distribution with unit variance. The proof requires independence of the $\mathbf{y}_i$'s (Ross, 2014), a condition which is not fully satisfied in our setup because the time series $\mathbf{X}_i$ has non-zero lagged correlation matrix. However, in section 4.5 below we argue that if the correlations decay fast enough, equation (4.129) still constitutes a reasonable approximation for the distribution of the residuals. The variables $\bar{\mathbf{e}}_j$ are called normalised errors. The last step is to note that, if the linear model (4.51) is correct, then:

$$\mathbf{e} = \sigma \mathrm{d}\mathbf{W} = \begin{bmatrix} \sigma_x \mathrm{d}\mathbf{W}_1 \\ \sigma_y \mathrm{d}\mathbf{W}_2 \end{bmatrix}, \tag{4.130}$$

with $\mathrm{d}\mathbf{W}_j \sim \sqrt{\Delta t} N(0,1)$ for $j = 1,2$. Comparing equation (4.129) with equation (4.130), we obtain an estimate for the components of the noise amplitude matrix σ:

$$\sigma_x = \frac{\sqrt{\frac{\mathbf{SSE}_1}{N-3}}}{\sqrt{\Delta t}} \tag{4.131}$$

$$\sigma_y = \frac{\sqrt{\frac{\mathbf{SSE}_2}{N-3}}}{\sqrt{\Delta t}}. \tag{4.132}$$

4.5 Results

Before we present our results, let's first summarise the problem here. The working hypothesis is that eddy-mean flow interaction in the Southern Ocean can be represented by a simplified two-dimensional dynamical system. The model we consider here is inspired by AN14 and consists in a linear, bivariate, damped, stochastically forced oscillator. To keep notation within reason, we shall refer to it as the "simplified model". The simplified model was introduced in section 4.2.7, and its defining equation is:

$$\mathbf{X}_{n+1} = A\mathbf{X}_n + \sigma \mathrm{d}\mathbf{W}_{n+1},$$

Here, $\mathbf{X}_n = (X_n, Y_n)$. X and Y are called state variables and represent mean flow and eddy activity respectively. Note that the simplified model is the discretised version of a time-continuous Ornstein-Uhlenbeck process, equation (4.50):

$$\mathrm{d}\mathbf{X} = \mathscr{A}\mathbf{X}\mathrm{d}t + \sigma \mathrm{d}\mathbf{W}_t,$$

The two formulations are equivalent. Depending on the specific task at hand, we may switch from one to the other, and section 4.2.7 provides all the necessary details. To test the working hypothesis, we consider two dataset:

1. The idealised channel dataset. In this case, the dynamical variables are computed from the MITgcm model output, as described in section 4.3.
2. A validation dataset. Here, the dynamical variables are computed by generating a synthetic realisation of the simplified model. The details are in section 4.5.2.

In both cases, the final product is a bivariate time series $\mathbf{X}_n = (X_n, Y_n)$. The analysis consists in fitting the simplified model to the data. It hinges on the Yule-Walker fitting method presented in section 4.4, and can be divided into the following steps:

1. Visual inspection of the time series and of the associated phase space diagram.
2. Characterisation of the time series and estimate of its statistical properties.
3. Comparison of the estimates with qualitative predictions from the simplified model.
4. Fit of the simplified model to the data.
5. Quantitative predictions and goodness of the fit.

We begin by looking at the idealised channel data.

4.5.1 Idealised channel

We analyse the discrete time series $\mathbf{X}_n = (X_n, Y_n)$ computed according to the procedure described in section 4.3. X_n is the time series of spatially-averaged Eady growth rate (representing mean flow), and Y_n is the time series of spatially-averaged eddy buoyancy flux (representing eddy activity). The individual time series comprise $N = 2376$ points each, and data are uniformly separated by a time interval of $\Delta t = 5$ days.

Visual inspection: time series and phase space diagram

Figure 4.7 (left) shows three-year excerpts from the full time series. Predictably, the excerpts are characterised by intense variability at multiple time scales, so that direct inference of dynamical relationships between the variables is impeded. To filter out part of the noise and shed light on the dynamical structure of the evolution of X and Y, we apply the phase space kernel averaging technique described in section 4.2.8. The kernel-averaged streamlines (black oriented lines) and the kernel-averaged data density (colour shading) are shown in figure 4.7 (right). The width of the averaging kernel is also shown by the green dot in the top left corner. We observe that the data density is largest around $(0,0)$ and decreases outward. This is consistent with the prediction that the distributions of X and Y have finite moments and are symmetrical with respect to their respective mean (more precisely, the simplified model predicts that the distributions of X and Y are Gaussians, as discussed below). The most remarkable feature of the phase space diagram, however, is that the kernel-averaged streamlines are approximately closed orbits. This result aligns with the findings of Novak et al. (2017) for the atmospheric storm track, and as discussed in the introduction to this chapter constitutes a strong indication that, on average, the interaction between X and Y is characterised by oscillations between the two variables. The orientation of the kernel-averaged orbits is such that positive deviations of X (intensification of mean flow) are followed on average by negative deviations of Y (intensification of eddies: the eddy buoyancy flux is negative in the Southern Ocean, section 4.3). This in turn leads to an average decrease in X (weakening of mean flow) which, in the final phase of the oscillation, is followed by positive deviations of Y (weakening of eddies). We further observe that the phase space velocity is generally larger at the outskirts of the data-populated region, implying that for large deviations dynamical effects dominate over stochastic contributions (as already noted by Marcheggiani et al. (2022) for the atmospheric case). The oscillatory life cycle is schematically illustrated in figure 4.8, and its physical interpretation is consistent with the predator-prey picture of eddy-mean flow interaction offered in AN14 and discussed in section 4.1. Based on the phase space diagram, the baroclinic life cycle can be divided into four phases: in phase I, eddies are weak (no predators), which allows the mean flow to gradually build up (the prey population increases). When sufficient reserves have been accumulated, they become available for consumption by baroclinic eddies. This is associated to a sudden increase in eddy activity, as seen in phase II. The baroclinic decrease soon depletes mean flow reserves: this is phase III of the cycle. In the final phase IV, there is no longer enough energy in the mean flow to sustain the baroclinic activity, and eddies fall back to a minimum. After that, the cycle resumes. It is necessary at this juncture to remark that kernel-averaged trajectories constitute a statistical description of the time evolution of the system, which includes contributions from both dynamical and stochastic effects. As such, they do not necessarily coincide (and should therefore not be confused with) the deterministic trajectories of the system (i.e. the paths in the phase space that would be observed in the absence of noise). We shall return to this important point later in the section.

Estimation of sample properties

Our next task is to characterise the time series of X and Y by estimating a few of their statistical properties from data. We begin from the sample marginal distributions of X and Y. The sample histograms of X (blue, left) and Y (red, right) are shown in figure 4.9. The histograms are normalised, so that the value of the integral over the bin range is 1. Visual inspection reveals that the sample PDFs are compat-

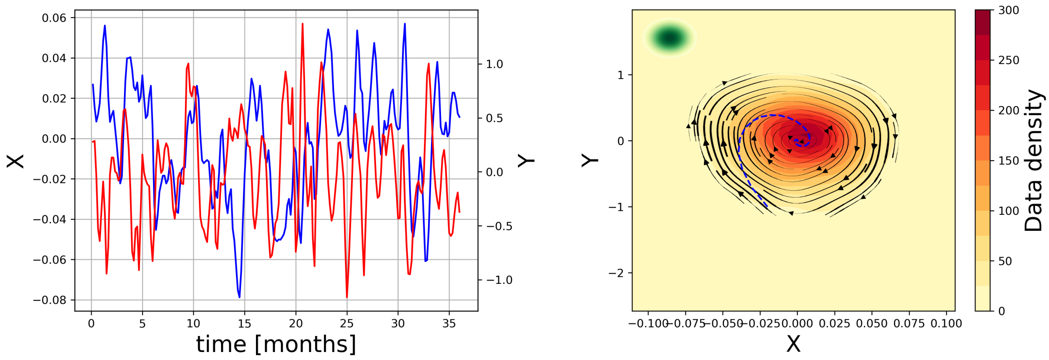

Figure 4.7: Left: three-year excerpts from the time series of X (continuous blue line) and Y (continuous red line). Right: kernel-averaged phase space diagram. Black oriented lines show kernel-averaged streamlines. The width of the lines is proportional to the phase space speed. Coloured intervals represent the density of data points. The width of the averaging kernel is shown by the green dot in the top left corner. The dashed blue line represents the deterministic trajectory computed for arbitrary initial conditions and for the value of the model parameters estimated with the Yule-Walker method. See main text for details.

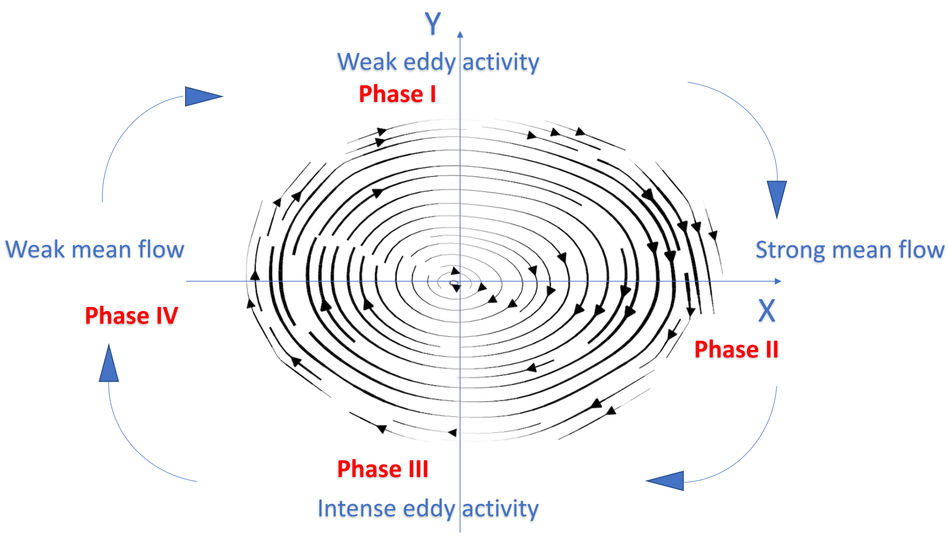

Figure 4.8: Schematic representation of the eddy-mean flow life cycle. See main text for details.

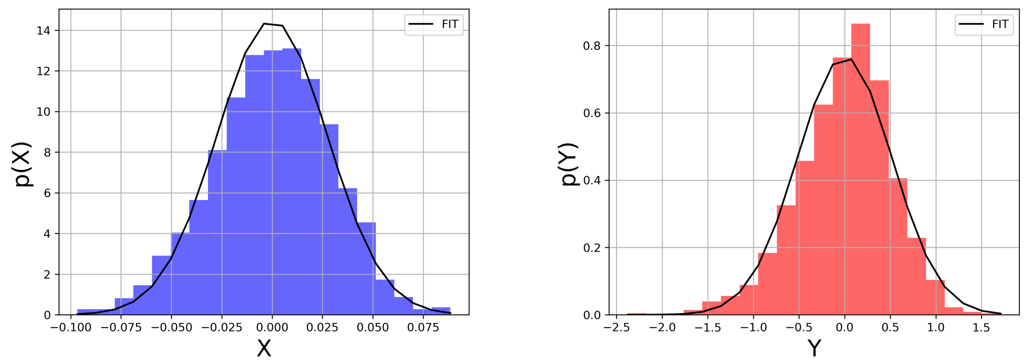

Figure 4.9: Left: sample marginal PDF of X (blue histogram) and prediction obtained with equation (4.58) using the Yule-Walker estimate of the process parameters (continuous black line). Right: sample marginal PDF of Y (red histogram) and prediction obtained with equation (4.59) (continuous black line). Histograms are normalised, i.e. the integral over the bin range is 1.

ible with Gaussian distributions, which is consistent with the theoretical prediction of the simplified model. Further tests of normality (normal probability plots, not shown) reveal that the sample PDF of X is Gaussian to an excellent degree of approximation, whereas the sample PDF of Y displays deviations from normality in the tails. The overall agreement, however, remains satisfactory, and does not present ourselves with strong evidence against the Gaussian hypothesis. We shall test this assumption further after fitting the simplified model to the data.

Next, we turn our attention to the correlation functions. The sample lagged-correlation functions are estimated according to the formula of section 4.2.6 and shown in figure 4.10. The auto-correlation functions of X and Y (top and middle panels respectively) decay non-linearly with a halving time scale of approximately 20 days for X and 10 days for Y. The cross-correlation function $\rho_{XY}(\tau)$ is to a reasonable approximation an odd function of the lag τ, namely:

$$\rho_{XY}(\tau) \approx -\rho_{XY}(-\tau), \tag{4.133}$$

and accordingly displays maximum and minimum at non-zero lags. The peaks are located at $|\tau| \approx 15$ days, and stand out from the 95 % confidence interval associated to the red noise test (see section 4.2.6), the peak at positive lags more markedly

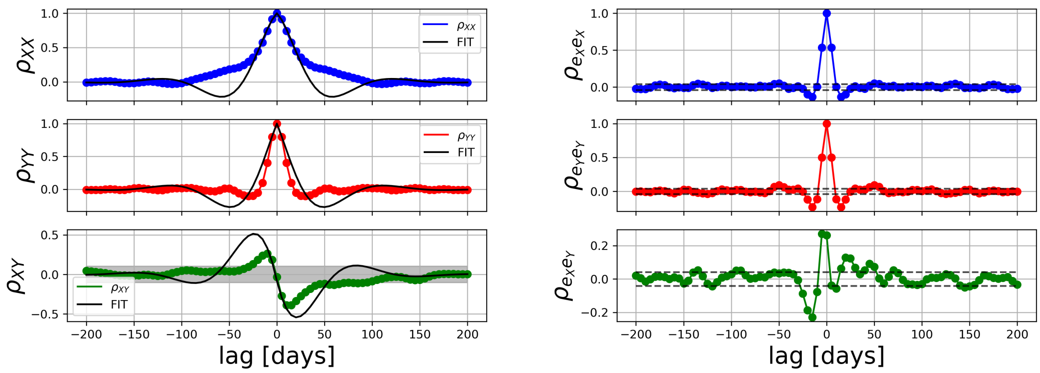

Figure 4.10: Left: lagged auto- and cross-correlation functions of X and Y. Top: $\rho_{XX}(\tau)$. Middle: $\rho_{YY}(\tau)$. Bottom $\rho_{XY}(\tau)$. Continuous, dotted, coloured lines show the sample correlation functions, while continuous black lines show the corresponding predictions obtained with equation (4.130) using the Yule-Walker estimate of the process parameters. The grey shaded interval in the bottom panel marks the 95% confidence interval associated to the red noise test, see section 4.2.6. Right: lagged auto- and cross-correlation functions of e_X and e_Y. Top: $\rho_{e_X e_X}(\tau)$. Middle: $\rho_{e_Y e_Y}(\tau)$. Bottom: $\rho_{e_X e_Y}(\tau)$. The sample correlation functions are shown by continuous, dotted, coloured lines. The dashed black lines mark the 95% confidence interval associated to the white noise test, see section 4.2.6 for details.

so. Thus, based on the sample cross-correlation we can reject at the 95 % confidence level the null hypothesis that X and Y are independent red noise processes. Note that the observed structure of the kernel-averaged trajectories constitutes supporting evidence that X and Y are not independent. At absolute lags larger than $|\tau| \approx 15$ days, the cross-correlation decays to zero. $\rho_{XY}(\tau)$ is negative at positive lags (when X leads Y), which means that a positive fluctuation in X (intensification of mean flow) is on average followed by a negative fluctuation in Y (intensification of eddies). The cross correlation function is approximately an odd function: accordingly, it takes positive values at negative lags. This means that a positive fluctuation in Y (weakening of eddies) is followed, on average, by a positive fluctuation of X (strengthening of mean flow). Thus, the physical picture conveyed by the correlation functions is in agreement with the eddy-mean flow life cycle inferred from the analysis of the phase space diagram, and summarised in figure 4.8.

Table 4.1: Summary of fit results for the idealised channel and the synthetic verification experiment.

Data	γ_{xx} $\left[\mathrm{day}^{-1}\right]$	γ_{yy} $\left[\mathrm{day}^{-1}\right]$	k_{xy} $\left[\frac{\mathrm{Kg}}{J\,\mathrm{day}}\right]$	k_{yx} $\left[\frac{J}{\mathrm{Kg}\,\mathrm{day}}\right]$	R^2	σ_x $\left[\mathrm{day}^{-3/2}\right]$	σ_y $\left[\frac{J}{\mathrm{Kg}\,\mathrm{day}^{3/2}}\right]$
Channel	$-1.50\cdot 10^{-2}$	$-4.30\cdot 10^{-2}$	$2.50\cdot 10^{-3}$	$-8.50\cdot 10^{-1}$	$(0.89, 0.69)$	$4.40\cdot 10^{-3}$	$1.28\cdot 10^{-1}$
Synth	$-1.39\cdot 10^{-3}$	$-4.13\cdot 10^{-2}$	$2.65\cdot 10^{-3}$	$-8.50\cdot 10^{-1}$	$(0.89, 0.71)$	$4.38\cdot 10^{-3}$	$1.27\cdot 10^{-2}$

Fit of the model

The model is fitted to the data with the Yule-Walker method, as explained in section 4.4. The fit consists in the estimate of the simplified model's parameter, the 2×2 matrix of coefficients A and the 2×2 noise amplitude diagonal matrix σ, so that there are in total 6 scalar free parameters. The matrix of coefficients $\mathscr{A}$ relative to the time-continuous SDE (4.50) associated to the discrete process is computed from A according to the formula:

$$\mathscr{A} = \frac{A - \mathbb{I}_2}{\Delta t}. \tag{4.134}$$

The Yule-Walker estimates of A, $\mathscr{A}$, and σ are denoted by $\hat{A}$, $\hat{\mathscr{A}}$, and $\hat{\sigma}$ respectively. The best fit values are reported in table 4.1.

Quantitative predictions and evaluation of the fit

Under the hypothesis that X and Y are governed by the simplified model, and equipped with the estimates $\hat{A}$ and $\hat{\sigma}$, it is possible to make quantitative predictions for the statistical properties of the bivariate process. The predictions for the marginal probability distribution functions of X and Y are computed according to formula (4.58) and (4.59), and shown in figure 4.9 (continuous black lines). The model's predictions are in excellent agreement with the sample PDFs. The agreement lends further evidence to the assumption that the marginal distributions of X and Y can be reasonably approximated by Gaussian distributions.

The predictions for the lagged correlation functions are computed according to equations (4.31), (4.32), and (4.33), and shown in figure 4.10 (continuous black lines). The predictions are only in partial agreement with the sample correlations: the auto-correlation of X is correctly reproduced up to $\tau \approx 25$ days, but the rate of decay is overestimated between $\tau = 30$ days and $\tau = 100$ days approximately. The predicted auto-correlation function of Y decays too slowly, and has more pronounced secondary peaks than observed for the sample correlation. The predicted cross correlation function $\rho_{XY}(\tau)$ peaks at greater lags ($|\tau| \approx 25$ days) than the sample one, and the peaks are larger. Nevertheless, a few observations stand in the simplified model's favour: firstly, the analysis of correlation functions is not commonly attempted in the literature (not, for example, in Ambaum and Novak (2014), Marcheggiani et al. (2022), Novak et al. (2017), Yano et al. (2020). Yano et al. (2020) maintain that this is due to the fact that lagged correlation analysis introduces at least an additional parameter to fit - the time lag), so that even a small achievement in this regard represents a significant improvement with respect to the existing body of research. Secondly, and perhaps more importantly, the model's predictions do succeed in capturing the overall structure of the correlation functions. It is noteworthy, in particular, that the predicted cross-correlation correctly reproduces the symmetry properties of the sample one, with a negative peak at positive lags and a positive peak at negative lags, thus corroborating the physical interpretation of the eddy-mean flow life cycle. The discrepancy between the model's predictions and the sample correlations hints to the fact that equation (4.51) may not be sufficient to describe the full quantitative details of eddy-mean flow interactions in the idealised channel. However, the qualitative agreement strongly supports the idea that the model satisfactorily captures the oscillatory character of the dynamics (plus some of the quantitative details, as for the marginal PDFs above).

We now turn the attention to more conventional metrics of evaluation of the fit. The coefficient of determinations for X and Y are $R_X^2 = 0.89$ and $R_Y^2 = 0.69$ respectively, therefore a large fraction of the variance in the dependent variables $\mathbf{y}$ is explained by variance in the independent variables $\mathbf{x}$. The value of R_Y^2 is small compared to R_X^2, but in the next section we show that this should be attributed to the levels of stochastic noise inherent to the data rather than to model deficiencies. The fit quality is visually demonstrated in figure 4.11. In the left hand side plot, dots represent sample values of X (blue) and Y (red) (i.e., the X and Y components of $\mathbf{y}$), while continuous lines show the corresponding predictions obtained with the Yule-Walker method (i.e., the X and Y components of $\hat{\mathbf{y}}$). We observe that the simplified model predicts the evolution of the sample time series one time step ahead satisfactorily. The scatter plots on the right hand side, or "prediction plots", offer an aggregate perspective on the quality of the fit. In the left hand-side panel, blue dots have coordinates $(\mathbf{y}_{1n}, \hat{\mathbf{y}}_{1n})$, i.e. the horizontal axis represents sample values of X and the vertical axis represents predicted values of X. The right hand side panel shows the same (red dots), but for sample and predicted values of Y. Note that the aspect ratio of the axes is 1. A perfect model with vanishing noise would produce points that fall exactly on the 1:1 line (shown by the dashed black lines). A perfect model with non-zero noise, instead, produces points scattered around the 1:1 line. Deviations from this pattern are indicative of model deficiencies, which need to be assessed on a case by case basis. Bar a few Y outliers, here scatter points are clustered around the 1:1 lines with no obvious patterns of deviation, which provides further confidence in the general quality of the fit. We stress that the scatter plots only demonstrate that the simplified model is not a severely insufficient model. They do not prove, of course, that it is a perfect one.

The final part of the fit evaluation consists in the analysis of the residuals. For a perfect model, the residuals are independent Gaussian random variables, as discussed in section 4.4.3. A common way of testing this hypothesis is to estimate the auto- and cross- correlation functions of the residuals (von Storch and Zwiers, 1999). The sample residual correlation functions, which we name $\rho_{e_X e_X}(\tau)$, $\rho_{e_Y e_Y}(\tau)$, and $\rho_{e_X e_Y}(\tau)$, are shown in figure 4.10 (continuous dotted lines), together with the 95% confidence interval relative to the white noise test (dashed black lines. See section 4.2.6 for the theory). They should be compared with those expected for a bivariate white noise process, when the lag-0 auto-correlation coefficients are equal to 1 and all remaining coefficients are zero. Although none of the three sample

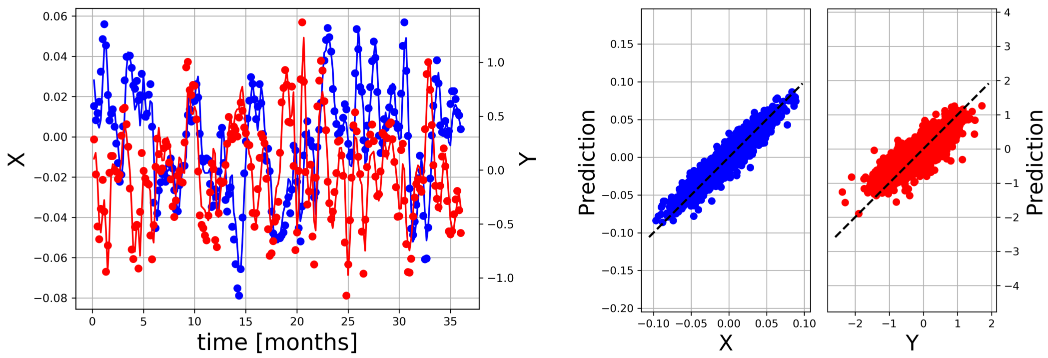

Figure 4.11: Left: three-years excerpts from the sample (dots) and one-step ahead predicted (continuous lines) time series of X (blue) and Y (red). Right: prediction plots for X (left hand side panel) and Y (right hand side panel). The horizontal axes represent sample values of X and Y, while the vertical axes represent one-step ahead predicted values of X and Y. The dashed black lines mark the 1:1 line. The aspect ratio of the axes is 1.

correlation functions satisfies the theoretical prediction perfectly, we observe that: (i) Only the lag-1 and lag-3 coefficients are significantly different from zero at the 95% significance level for $\rho_{e_X e_X}(\tau)$. All other coefficients lie either within or very close to the acceptance interval of the null hypothesis. (ii) The sample correlation $\rho_{e_Y e_Y}(\tau)$ shows more structure, with the first few coefficients clear of the acceptance interval. This is where the most significant deviations from the expected behaviour are found, which once again highlights that some of the fine details pertaining the dynamics of Y may not be captured by the simplified model. The existence of significant persistent correlations in e_Y, in particular, hints that the evolution of Y might be more accurately represented by an auto-regressive model of higher order (the disadvantage of higher order auto-regressive models is that the interpretation of the model coefficients is potentially complicated. For this reason, we only consider the first-order simplified model here.) (iii) A few of the coefficients of $\rho_{e_X e_Y}(\tau)$ stand out clearly of the acceptance interval. Their absolute value, however, is not large (slightly more than 0.2 at most) and the pattern they form is rather irregular, which makes the interpretation of the sample cross-correlation unclear. Overall, this test underscores that the residuals produced by the simplified model are not perfectly

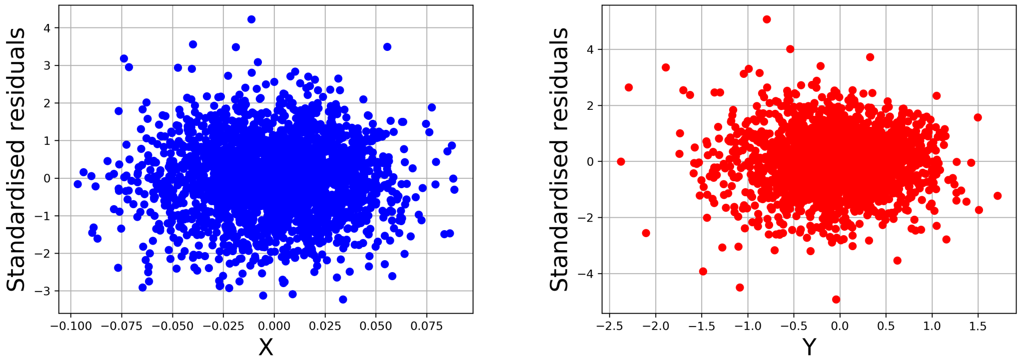

Figure 4.12: Left: scatter plot of $\overline{e}_X$ against X. Right: scatter plot of $\overline{e}_Y$ against Y.

uncorrelated. As it was the case for some of the other indicators considered, it appears that the model's most evident shortcomings are associated to the evolution of the eddy variable Y (another possibility, which we do not explore here, is that the stochastic forcing term is endowed with its own time scales and would be better described by e.g. red noise). Nevertheless, the correlation plots also reveal that, when present, departures from the theoretical expectation are not so severe to put the ability of the model to qualitatively capture the important features of the observed process to question. An alternative is to produce scatter plots of the standardised residuals (i.e., the residuals divided by the estimator of their standard deviation, equation (4.129)) against X and Y (Ross, 2014). Similarly to the prediction plots, the emergence of coherent patterns in the scatter plots signals potential modelling defects. Figure 4.12 shows the scatter plots of $\overline{e}_X$ against X (blue dots, left) and of $\overline{e}_Y$ against Y (red dots, right). We observe that for both variables the vertical coordinate of a large fraction of the dots falls within the interval ± 2, indicating that most of the non-standardised residuals are valued within ± 2 standard deviations. This supports the idea that the residuals are approximately Gaussian, and that their variance is reasonably estimated by $\text{SSE}/N - 3$. Furthermore, the scatter plots do not reveal any obvious patterns such as linear or polynomial dependance of the residuals on X and Y, which places additional confidence in the quality of the model.

Remark: deterministic dynamics

The estimation of the matrix of coefficients $\hat{A}$ (or, alternatively, of its continuous SDE equivalent $\hat{\mathscr{A}}$) allows to deepen our understanding of the phase space diagram. The dashed blue line in figure 4.7 (right) shows the phase space trajectory of the deterministic process:

$$\mathbf{X}_{n+1} = \hat{A}\mathbf{X}_n, \tag{4.135}$$

for arbitrary initial conditions. Simple properties of this process can be investigated by studying the matrix structure of $\hat{\mathscr{A}}$. The components of $\hat{\mathscr{A}}$ satisfy the conditions $\gamma_{xx}, \gamma_{yy} < 0$ and $k_{xy}k_{yx} < 0$ set out in section 4.2.7, therefore the deterministic system above is a damped oscillator. The eigenvalues of $\hat{\mathscr{A}}$ are $\lambda_1 = -0.029 + 0.044i$ and $\lambda_2 = -0.029 - 0.044i$: the imaginary part is non-zero, hence the oscillator is in the sub-critical regime (i.e., deterministic oscillations are not completely suppressed by damping). Accordingly, we observe that the deterministic trajectory shown in figure 4.7 (right) is characteristic of sub-critically damped linear oscillators. What is most remarkable, though, is the stark qualitative difference between the deterministic trajectory (decaying, indicative of dissipative dynamics) and the kernel-averaged trajectories (quasi-periodic, suggestive of conservative dynamics). The important implication is that this analysis shows that quasi-periodic averaged phase space trajectories can be obtained even when the underlying deterministic dynamic is dissipative. We argue that nearly closed kernel-averaged trajectories similar to those observed in Novak et al. (2017) and in this work can arise in one of two ways: (i) The underlying deterministic model is conservative or close to being conservative. In this case, the structure of the kernel-averaged phase space diagram is governed by deterministic effects. This is, for example, the standpoint of Novak et al. (2017), Yano et al. (2020). Yano et al. (2020), in particular, fitted a linear oscillator (see section 4.2.7) to data, obtaining in a few cases eigenvalues with non-zero real part (corresponding to dissipative systems). Since these cases where also those associated with the worst fit scores, though, the authors did not explore the dissipative regime further. (ii) The underlying deterministic dynamics are dissipative, but the kernel-averaged trajectories are shaped by deterministic *and* stochastic effects. This is the perspective presented for the first time in this work. In the next section, we bolster this interpretation by studying a synthetic realisation of process (4.51), so that the kernel-averaged trajectories originate from a dissipative system with certainty, and potential shortcomings of the simplified model in describing the idealised channel data play no role. We stress that the approach presented here is general and can cover both case (i) and case (ii), as one can obtain a conservative system from the fit to the data provided the diagonal elements of $\mathscr{A}$ are approximately zero.

4.5.2 Validation of the fitting procedure

We validate the model fitting procedure by studying a synthetic realisation of the simplified model. We prescribe the problem parameters A and σ, and demonstrate that we can recover their values with good approximation. For illustration purposes, the matrix of coefficients A and the noise amplitude matrix σ are set to the values estimated for the idealised channel, see table 4.1. The synthetic data are generated by integrating equation (4.51) with the Euler-Maruyama method (Higham, 2001), with $\Delta t = 5$ days and up to a final time of $t_f = 30$ years.

Figure 4.13 (left) shows three-year excerpts from the time series of X and Y. As for the case of the idealised channel data, the time series are characterised by intense variability at multiple time scales, and do not lend themselves to straightforward physical interpretation. To filter out part of the noise we apply the phase space kernel averaging technique of section 4.2.8. The kernel-averaged streamlines (black oriented lines: the width of the line is proportional to the local phase space velocity) and the kernel-averaged data density (coloured intervals) are shown in figure 4.13 (right). Remarkably, although phase space trajectories are more regular here, the qualitative structure of the diagram is the same as for the idealised channel, figure 4.7 (right). In particular, the idealised channel and the verification diagrams share the following features:

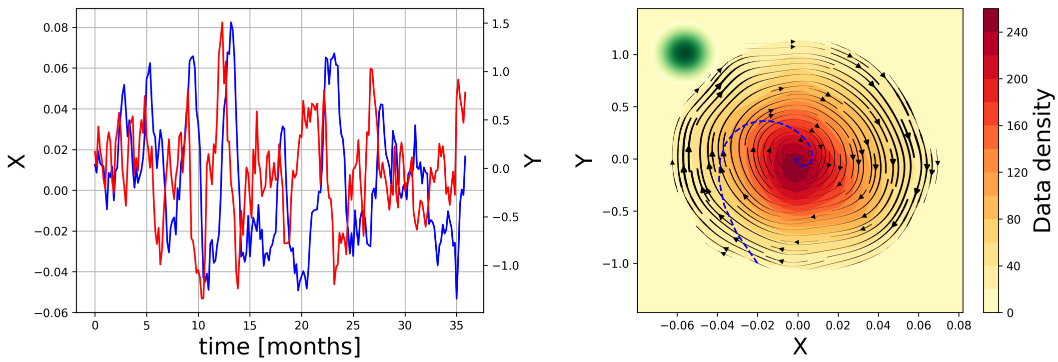

Figure 4.13: Same as figure 4.7, but for the synthetic verification dataset.

1. The data density is largest near $(0,0)$ and decreases outward.

2. The largest values of the phase space velocity are found at the outskirts of the data-populated region.

3. Kernel averaged trajectories are quasi-periodic orbits. Their orientation is consistent with the physical interpretation of the baroclinic life cycle.

4. The deterministic trajectory is typical of dissipative deterministic dynamics and differs qualitatively from the kernel-averaged trajectories.

Since the verification data are generated by numerically integration of the simplified model (a stochastically forced, dissipative oscillator), this analysis confirms that quasi-periodic averaged trajectories can arise even if the underlying deterministic dynamics are dissipative.

The sample marginal PDFs of X and Y are shown in figure 4.14. As for the idealised channel, visual inspection of the plots confirms that $p(X)$ and $p(Y)$ are qualitatively compatible with normal distributions.

The sample lagged-correlation functions are shown in figure 4.15 (left). The auto-correlation functions of X and Y (top and middle panels respectively) are characterised by non linear decay, with a halving time scale of approximately 25 and 15 days respectively. The cross-correlation function $\rho_{XY}(\tau)$ is to a good approximation an odd function of τ. The peaks are located at $|\tau| \approx 25$ days, and stand out clearly

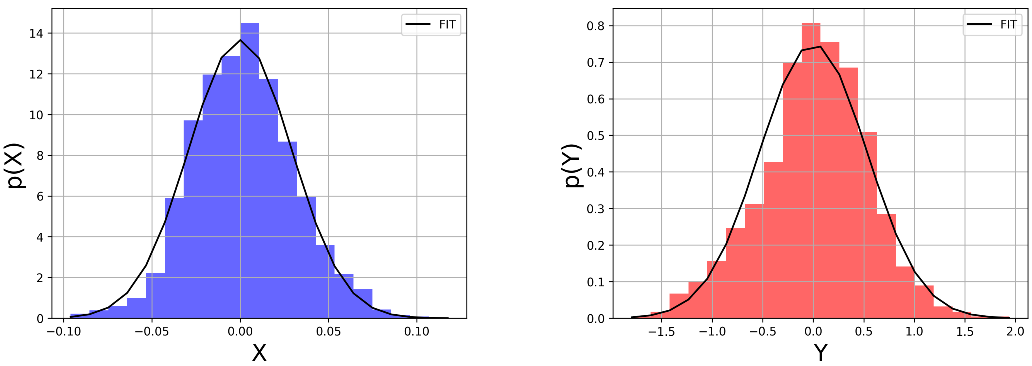

Figure 4.14: Same as figure 4.9, but for the synthetic verification dataset.

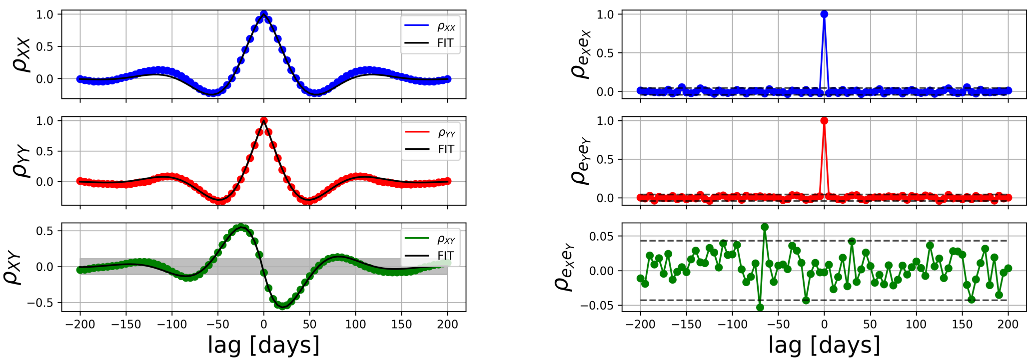

Figure 4.15: Same as figure 4.10, but for the synthetic verification dataset.

from the 95 % confidence interval associated to the red noise test. At absolute lags larger than $|\tau| \approx 25$ days, the cross-correlation decays to zero. Compared to the case of the idealised channel, all correlation functions show more regular dependence on the time lag τ. Moreover, the peaks of the cross-correlation are larger and are located at greater lags. However, the overall structure is qualitatively comparable with that of figure 4.10.

We estimate the matrix of coefficients A and the noise amplitude matrix σ as in the case of the idealised channel. The estimated values are shown in table 4.1. We observe that the prescribed values of the process parameters are recovered accurately: the worst estimate, obtained for the coefficient k_{xy}, departs from the corresponding "true value" by a 7.5% relative error only. For k_{yx}, σ_x, and σ_y, on the other hand, the associated relative error is below 1%. The coefficients of determi-

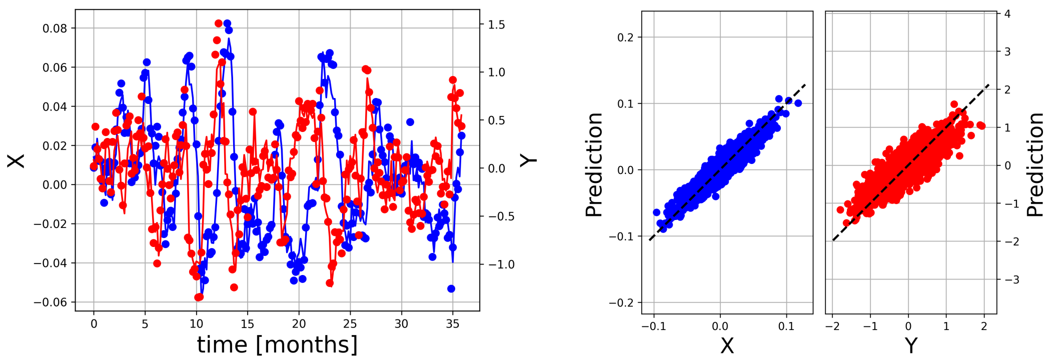

Figure 4.16: Same as figure 4.11, but for the synthetic verification dataset.

nation are $R_X^2 = 0.89$ and $R_Y^2 = 0.71$, therefore in this case too a significant fraction of the dependent variables' variability is explained by the variability of the independent variables. Importantly, the coefficients of determination are quantitatively similar to those obtained for the idealised channel data. In the verification experiment, the values of the coefficients of determination are entirely determined by the prescribed model parameters (and in particular by the amplitude of the stochastic terms), and are not associated to insufficiencies of the model: this suggests that the comparatively low score R_Y^2 obtained for the idealised channel is driven by the levels of noise inherent to the data, and is not indicative of modelling defects. The estimated values of A and σ are used to compute quantitative predictions of the marginal PDFs of X and Y (figure 4.14, continuous black lines) and of their auto- and cross-correlation functions (figure 4.15 (left), continuous black lines). In all cases the predicted curves are in excellent agreement with the sample ones, convincingly demonstrating the validity of the fitting procedure.

The quality of the fit is visually demonstrated in figure 4.16. In the left hand side plot, dots represent sample values of X (blue) and Y (red), while continuous lines show the corresponding predictions obtained with the Yule-Walker method. As expected, we observe that the simplified model reproduces the evolution of the sample time series one time-step ahead satisfactorily. Prediction plots are shown in figure 4.16 (right). The plots confirm that for a perfect model with non-vanishing noise, scatter points are clustered around the 1:1 lines.

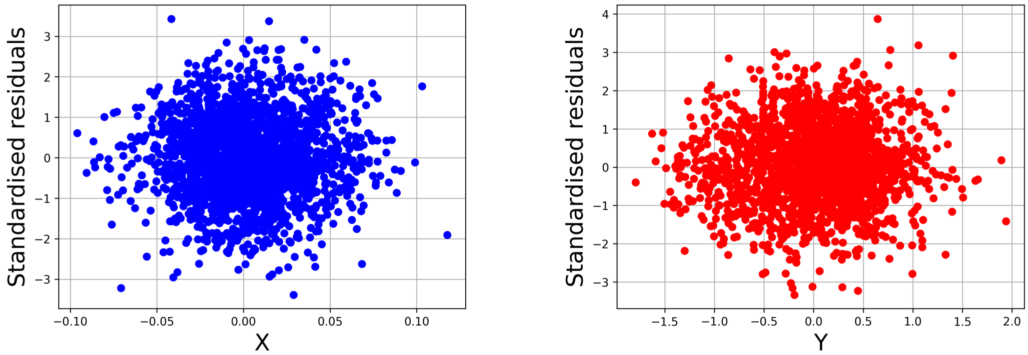

Figure 4.17: Same as figure 4.12, but for the synthetic verification dataset.

The analysis of the residuals is conducted as in the preceding section. Figure 4.15 (right), shows the sample correlation functions $\rho_{e_X e_X}(\tau)$, $\rho_{e_Y e_Y}(\tau)$, and $\rho_{e_X e_Y}(\tau)$ (continuous dotted lines), together with the 95% confidence interval relative to the white noise test (dashed black lines). We observe that the sample residual correlations are entirely consistent with the theoretical expectation for two independent white noise processes, as the only correlation coefficients that stand out clearly from the acceptance intervals are the lag-0 auto-correlations. Otherwise, the coefficients are either within the acceptance interval or in its proximity, in the latter case due to the multiplicity effect discussed in section 4.2.6. Finally, the scatter plots of the standardised residuals against X and Y are shown in figure 4.17. As expected, the residuals are distributed within ± 2 standard deviations from zero, and the plots are free of structural patterns which would point to the existence of a functional relationship between the residuals and the state variables.

4.5.3 Time scales and physical interpretation of the coefficients

The simplified model was introduced based on heuristic arguments, therefore it is not straightforward to relate its parameters, namely the coefficients k's, γ's, and σ's, to those characterising the idealised channel, such as wind stress and buoyancy forcing at the surface. The most direct way to address the issue would be to perform a suite of sensitivity experiments, i.e., to repeat the numerical simulations and subsequent analysis using different values of e.g. wind stress and bottom drag, or other relevant quantities. This method, though, has two main drawbacks: (i) it is computationally expensive, and (ii) the numerical model's changes may in principle project onto any of the simplified model's parameters (with no prior knowledge on which one), which warrants a large number of experiments being executed in order to establish the scaling laws with accuracy. Note that, as shown in section 4.3.2, the theoretical model of Marshall et al. (2017) allows to derive scaling laws for the average values of X and Y: the simplified model, however, describes the variability of deviations of X and Y around their averages, so that equations (4.104) and (4.107) do not easily translate into additional information on the scaling of the coefficients k's, γ's, and σ's. In view of this, and while we acknowledge its potential merits, we do not pursue the sensitivity experiment approach further in this manuscript. Nevertheless, we can interpreter the dynamical coefficients k's and γ's and associate them with typical time scales.

Consider the damping coefficients γ's first: for the sake of clarity, we focus on the deterministic part of the dynamics only, and momentarily assume that the coupling coefficients k's are set to zero. The governing equations are then:

$$\frac{\mathrm{d}}{\mathrm{d}t}X = \gamma_{xx}X \tag{4.136}$$

$$\frac{\mathrm{d}}{\mathrm{d}t}Y = \gamma_{yy}Y\,. \tag{4.137}$$

These equations describe exponential decay of the fluctuations in the mean flow (X) and the eddy activity (Y) variables respectively. The coefficients γ's control the rate of decay, and their inverse represent therefore relaxation times. For each variable, the halving time is computed by multiplying the relaxation time by $\log(2)$, yielding $\tau_x \approx 47$ days for the mean flow and $\tau_y \approx 16$ days for the eddies. The time needed to halve the initial amplitude of a perturbation, hence, is almost three times as long for X than for Y, which also reflects in the fact the the auto-correlation function of X decays more slowly than that associated to Y, figure 4.10 (left). Thus, our model captures the fact that the mean flow is characterised by greater temporal persistence than the eddy field on short time scales.

In order to interpreter the coupling coefficients k's, we go back to the full deterministic system:

$$\frac{\mathrm{d}}{\mathrm{d}t}X = \gamma_{xx}X + k_{xy}Y \tag{4.138}$$

$$\frac{\mathrm{d}}{\mathrm{d}t}Y = \gamma_{yy}Y + k_{yx}X\,. \tag{4.139}$$

It being deterministic, the time evolution of the solution to this system is determined by the initial conditions. To study the time scales associated with the coefficients k's, we consider two different choices for the initial values of X and Y. The first choice is $X(t=0) = 2s_x$ and $Y(t=0) = 0$, where s_x is the standard deviation of X (note that it is different from σ_x, the amplitude of the stochastic forcing on the X equation). The situation described is that of eddy growth following an intense steepening of the isopycnals. In this case, the exact equations (4.138) and (4.139) simplify to:

$$\frac{\mathrm{d}}{\mathrm{d}t}X = \gamma_{xx}X \tag{4.140}$$

$$\frac{\mathrm{d}}{\mathrm{d}t}Y = k_{yx}2s_x\,, \tag{4.141}$$

at $t = 0$. This is a crude approximation for $t > 0$, which can only be valid during the initial stages of the evolution. These equations ignore the feedback of Y on X, the exponential decay of Y, and the effect that changes of X have on the evolution of Y. Nevertheless, they encapsulate the dynamics of the initial stages of the evolution, and allow ourselves to make order of magnitude estimates for the time scales associated with the coupling coefficients. Figure 4.18 (left) shows the solution to the full deterministic system (equations (4.138) and (4.139), continuous lines) and its approximation (equations (4.140) and (4.141), dashed lines) for the considered initial conditions. The dashed blue line shows the exponential decay of X from an initial displacement of two standard deviations, as described by equation (4.140): the halving time $\tau_x \approx 47$ days corresponds to the time at which the curve decreases to 1 s_x (the 1 s_x level is marked by the dashed-dotted blue line): since the initial condition is $X(t=0) = 2s_x$, the halving time coincides with the time needed for a one standard deviation decrease. Note that the exact solution (continuous blue line) decays faster than the approximated one, as the dampening effect of the growing eddies on the mean flow is ignored in the approximated equations. The continuous red line shows instead the exact solution for the eddy activity. We observe that its evolution consists of two phases: firstly, eddy activity grows (i.e. the eddy buoyancy flux Y becomes more negative), as eddies feed on the increased mean flow. This initial phase is qualitatively captured by the solution of the approximated equations (4.140) and (4.141) (dashed red line). Next, eddies start to decay due to the effect of the damping term γ_{yy}, i.e., Y increases. Exponential decay is not included in equation (4.141), therefore this phase is not reproduced by the approximated solution. According to equation (4.141), eddies grow linearly in time. The growth is driven by the coupling coefficient k_{yx}, which represent the conversion of (constant) excess mean flow into eddying motion. A representative time scale for k_{yx} is thus the time needed for the eddy buoyancy flux to increase in absolute value by one standard deviation, which visually corresponds to the intersection between the dashed red line (the approximated solution of Y) and the dashed-dotted red line (the - 1 s_y level, where s_y is the standard deviation of Y) in figure 4.18 (left). We obtain a value of about 10 days (the time scale from the exact equations is 20 days), which is a factor 5 smaller than the halving time $\tau_x \approx 47$ associated with X, suggesting that eddies are particularly effective at extracting energy from the mean flow following positive fluctuations.

The second case we consider is the mirror image of the one above, in that we assume $X(t=0)=0$ and $Y(t=0)=-2s_y$. The situation described is thus that of a decaying eddy field weakening the mean flow, and the relevant approximated equations are:

$$\frac{\mathrm{d}}{\mathrm{d}t}X = -k_{xy}2s_y \tag{4.142}$$

$$\frac{\mathrm{d}}{\mathrm{d}t}Y = \gamma_{yy}Y\,. \tag{4.143}$$

The time evolution of the solution to the exact (continuous lines) and approximated (dashed lines) equations is shown in figure 4.18 (right). According to equation (4.143), Y decays exponentially (dashed red line), and the $\tau_y \approx 16$ days time scale corresponds to the intersection with the - 1 s_y level (dashed-dotted red line). The exact trajectory of Y (continuous red line) decays faster, because the effect of negative deviations of X is ignored in the approximated solution. Similarly to the previ-

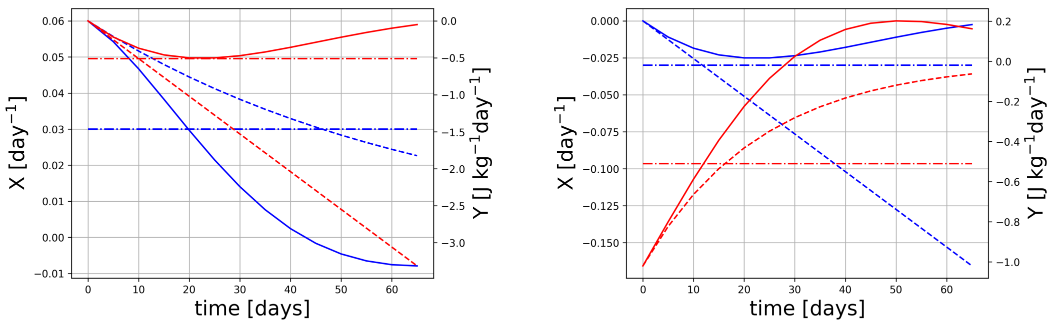

Figure 4.18: Numerical solution of the deterministic equations for $X(t=0) = 2s_x$ and $Y(t=0) = 0$ (left), and $X(t=0) = 0$ and $Y(t=0) = -2s_y$ (right). Continuous lines show the solutions to the exact deterministic equations (4.138) and (4.139), while dashed lines show the solutions to the approximated equations (4.140) and (4.141) (left), and (4.142) and (4.143) (right). Dashed dotted lines mark the $1s_x$ (blue) and $1s_y$ (red) levels. In both panels, X is shown in blue (ticks on the left hand side axis) and Y is shown in red (ticks on the right hand side axis).

ous case, the exact solution for X (continuous blue line) is endowed with two time scales, an initial flattening of the isopycnals driven by the coupling coefficient k_{xy} and a subsequent recovery driven by the damping term γ_{xx} (which includes the effect of the wind acting to restore isopycnals). The approximated solution (dashed blue line) is a straight line, and only captures the initial decay phase. A relevant time scale for k_{xy} is then the time needed for the mean flow to decrease by 1 standard deviation following an intense eddy event, which visually corresponds to the intersection between the dashed blue line and the $-1s_x$ level (dashed-dotted blue line). We obtain a typical time scale of about 12 days, which is comparable with that associated with k_{yx}, but not significantly shorter than the restoring time scale for the eddy buoyancy flux: the implication is that a few eddy events may be needed to induce significant fluctuations in the mean flow.

In summary we have shown that, although it is not possible to relate the simplified model's dynamical parameters to the physical parameters of the idealised channel in a simple way, we can associate them with the typical time scales of the processes they superintend. The time scales thus obtained are all of plausible magnitude, ranging from approximately one to a few weeks, and are consistent with the physical picture of the predator-prey relationship discussed throughout the chapter. Identifying the important time scales associated to each of the simplified model's parameters is a key step towards a comprehensive physical understanding of the dynamics of interaction between eddies and mean flow in the Southern Ocean. Further investigation is required, though, to fill the gap between the simplified model and the physical variables that characterise the Southern Ocean completely, and in the next section we discuss some of the alternatives that may contribute towards this goal.

4.6 Sensitivity of the results to the choice of the averaging domain

In this section, we demonstrate that our results do not depend strongly on the precise choice of the averaging domain. To this end, we show the results of the fit for (i) a domain co-located with that of section 4.5.1, but twice as wide (figure 4.19). (ii) A domain of the same size as that of section 4.5.1, but placed in the northern half of the channel (figure 4.20). (iii) A domain of the same size as that of section 4.5.1, but placed in the southern half of the channel (figure 4.21). The averaging domains are shown superimposed to time-mean, zonal-mean Eady growth rate (left) and eddy buoyancy flux (right) in the top panels of each figure.

In all cases, the qualitative results of sections 4.5.1 are reproduced satisfactorily. Notably, we observe that, while the details change (for example, the magnitude and location of the peaks of the cross-correlation functions), the overall structure of the phase space diagram and correlation functions (middle panels) is preserved throughout the experiments. This highlights that the predator-prey relationship found between eddies and mean flow is a robust feature of the idealised channel, and its detection does not depend critically on how the data are spatially averaged. The quality of the fit is also of consistently high standards for all three test domains, as visually demonstrated by the prediction plots (no structural patterns) and residual correlations (no large deviations from the white noise case) shown in the bottom panels of each figure.

4.7 Summary and conclusions

In this chapter, I have shown that the interaction between eddies and the mean flow in an idealised numerical simulation of the Southern Ocean can be modelled with a two-dimensional stochastic linear oscillator. The work was motivated by the well-documented dynamical analogy between the ACC and the atmospheric storm track, and inspired by the simplified model of atmospheric variability proposed by AN14. Specifically, AN14 and subsequent studies showed that the dynamics of eddy-mean flow interaction in atmospheric jets are akin to those expressed by predator-prey models of population growth. In this picture, eddies behave as a predator feeding on the reserves stored in the mean flow, and feedbacks between the two variables can induce oscillations. Here, I have addressed the following question: can a simple, two-dimensional dynamical system describe the interplay between eddies and the mean flow in the oceanic case too? Note that, while the original model of AN14 is entirely deterministic, here I relaxed this constraint and included a stochastic term in the dynamical equations, which allows to explicitly account for noisy fluctuations of the data.

I have analysed data from a high-resolution configuration of the MITgcm (the idealised channel), described in detail in chapter 3. The model grid resolves the first Rossby deformation radius throughout the domain therefore, importantly, I need not resort to an eddy parametrisation scheme. Similarly to what AN14 did for the atmosphere, I have defined a pair of spatially averaged variables X and Y measuring mean flow and eddy activity respectively for the oceanic case. My choice, supported by previous research, was to compute the spatial averages of the Eady growth rate and of the eddy buoyancy flux over a rectangular region on the meridional plane,

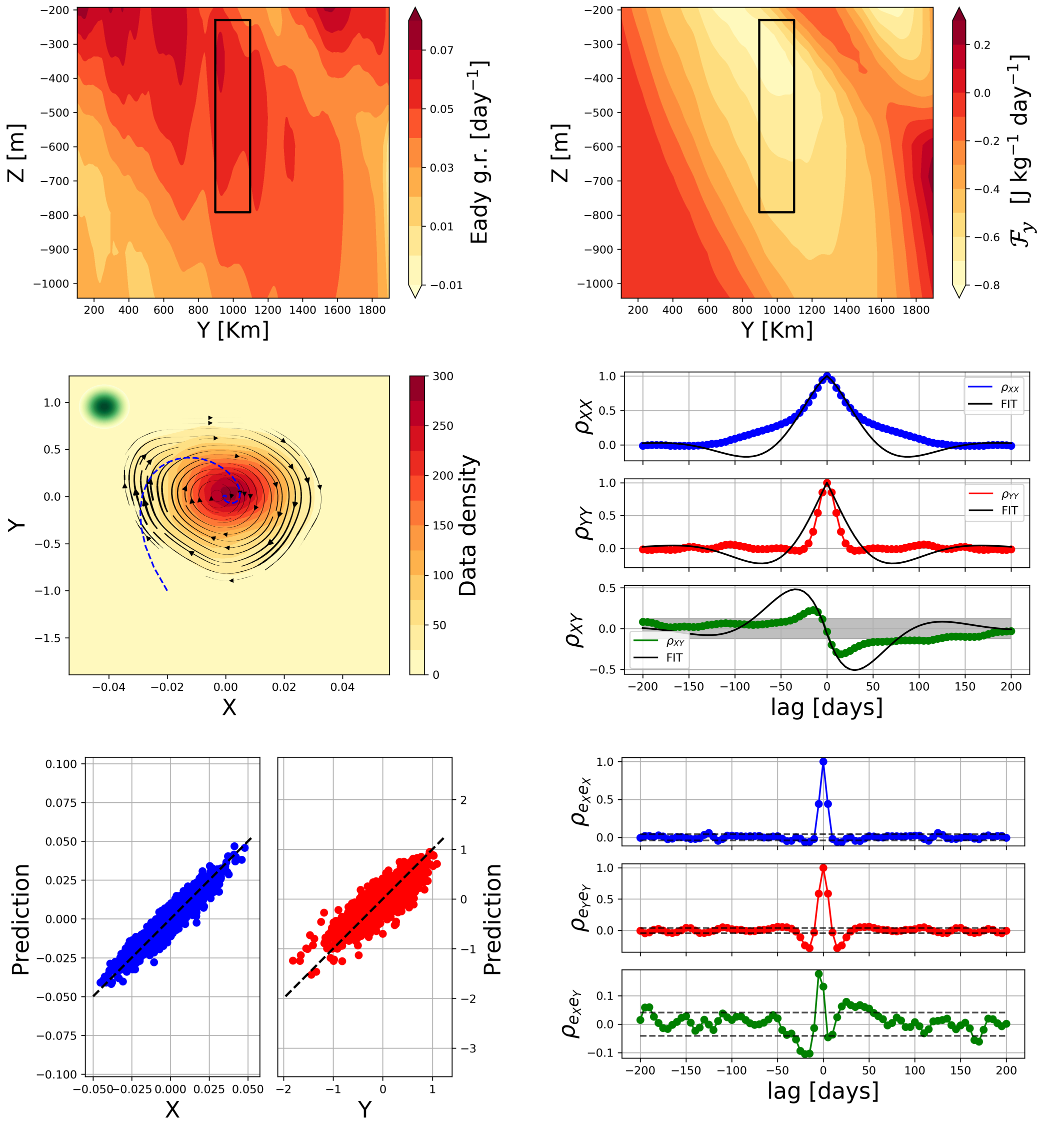

Figure 4.19: Top: the averaging domain (i) (continuous black lines) superimposed to time-mean, zonal-mean Eady growth rate (left) and eddy buoyancy flux (right). Middle: phase space diagram (left) and correlation functions (right). Bottom: prediction plot (left) and residual correlation functions (right). For the middle and bottom panels, all plots are as in section 4.5.1.

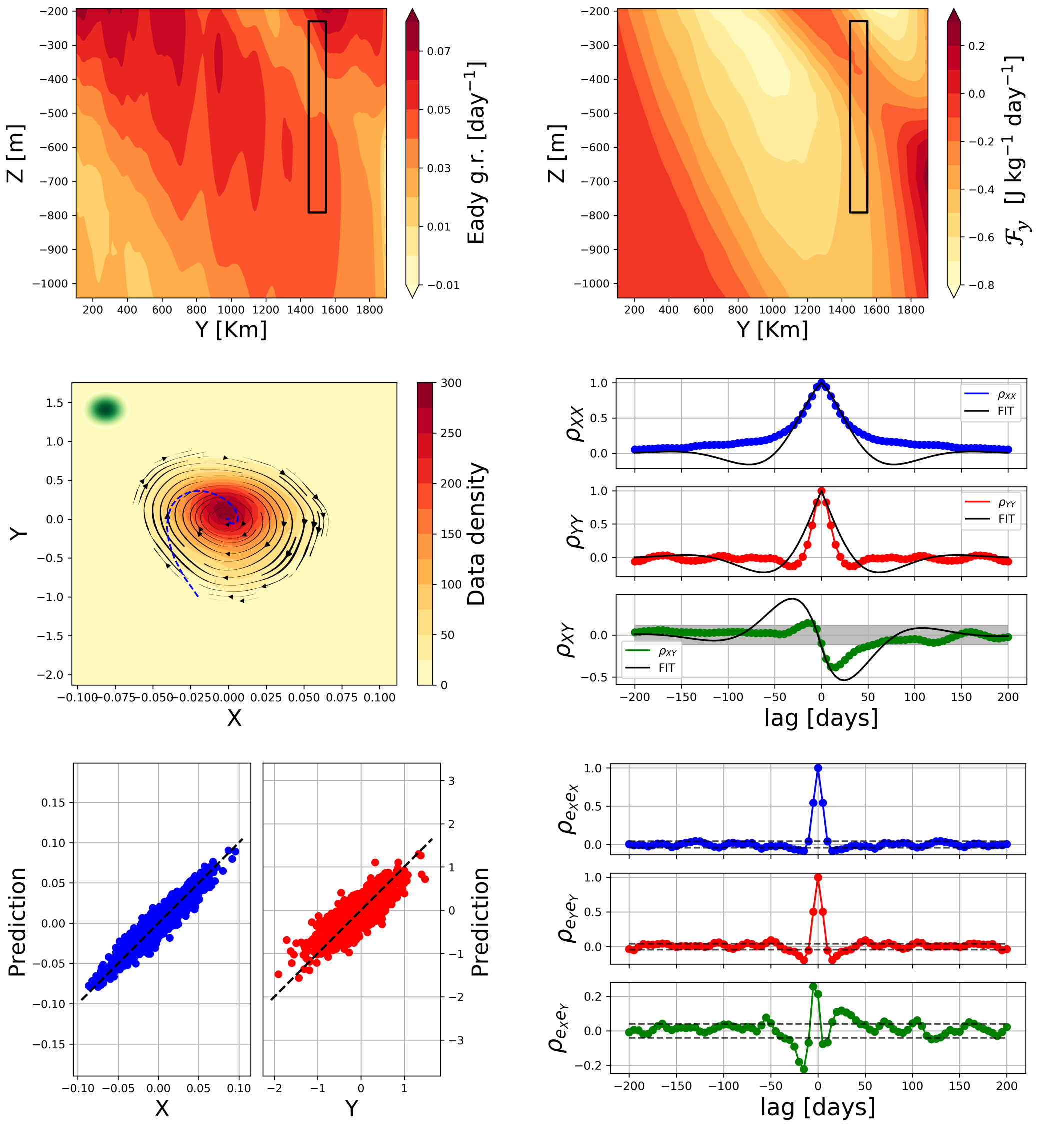

Figure 4.20: Top: the averaging domain (ii) (continuous black lines) superimposed to time-mean, zonal-mean Eady growth rate (left) and eddy buoyancy flux (right). Middle: phase space diagram (left) and correlation functions (right). Bottom: prediction plot (left) and residual correlation functions (right). For the middle and bottom panels, all plots are as in section 4.5.1.

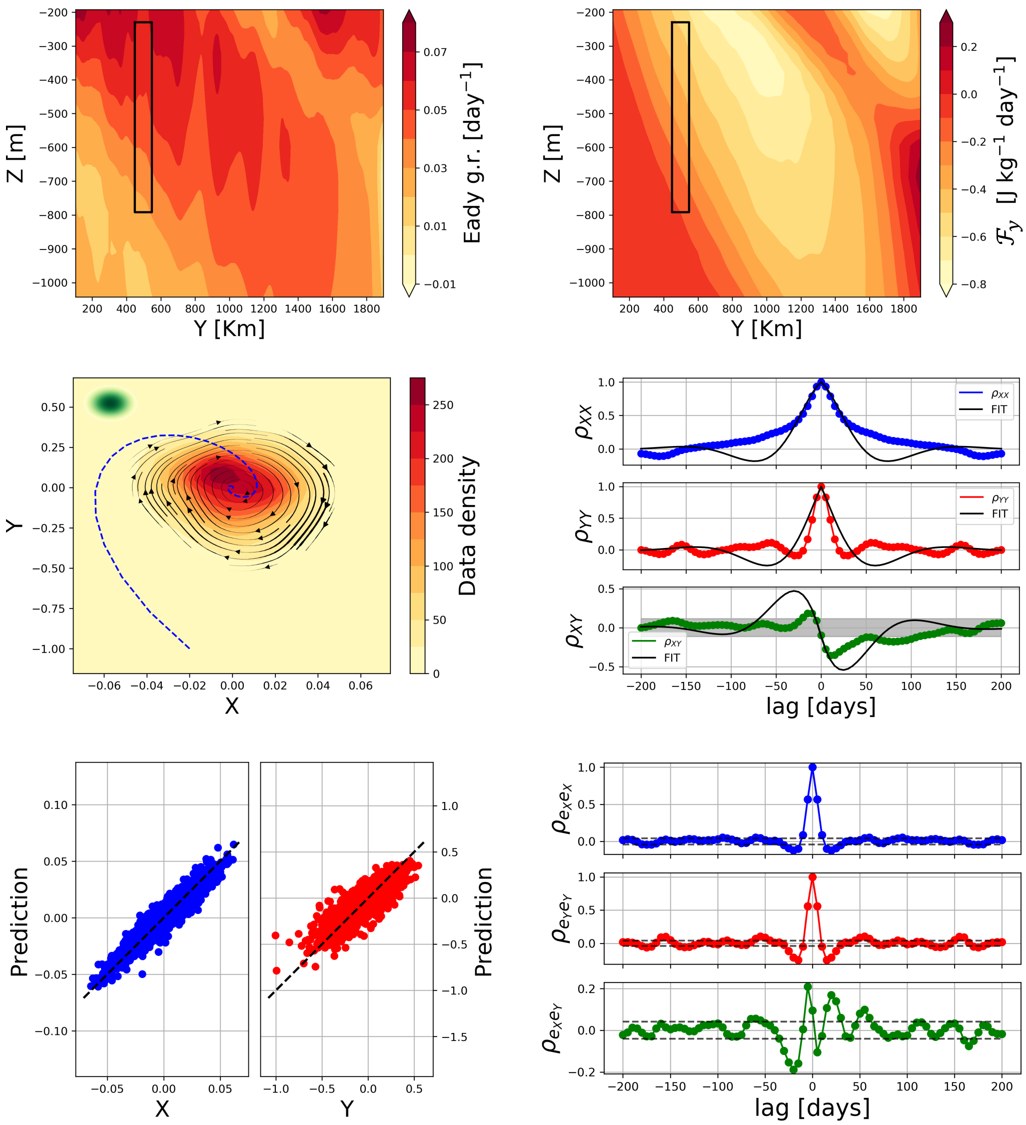

Figure 4.21: Top: the averaging domain (iii) (continuous black lines) superimposed to time-mean, zonal-mean Eady growth rate (left) and eddy buoyancy flux (right). Middle: phase space diagram (left) and correlation functions (right). Bottom: prediction plot (left) and residual correlation functions (right). For the middle and bottom panels, all plots are as in section 4.5.1.

located at intermediate latitudes and in the upper interior of the domain. I have analysed the time series of X and Y both qualitatively and quantitatively.

From a qualitative standpoint, I sought signatures of an oscillatory relationship between the two variables by studying the structure of the kernel-averaged phase space trajectories and of the cross correlation function between X and Y. The data are characterised by intense variability, therefore I expected the oscillations to emerge only on average, and not in the traditional sense of deterministic systems. Quantitatively, I have fitted the simplified model to the data by means of the Yule-Walker method, thereby obtaining an estimate for the simplified model's parameters. The fitting technique was bolstered by performing a verification experiment, where I numerically generated a synthetic realisation of the simplified model and shown that I am able to successfully reconstruct the prescribed parameters.

My results are that:

1. The kernel-averaged phase space trajectories are quasi-periodic orbits, and their orientation is consistent with the predator-prey model of AN14. This strongly supports the idea that eddy-mean flow interactions are endowed with an oscillatory character.

2. The cross-correlation function between X and Y is approximately an odd function of the time lag. Its peaks are statistically significant and located at $|\tau| \approx 15$ days, which can be thought of as a characteristic time for the oscillation. The form of the cross-correlation function is consistent with the physical interpretation of the eddy life-cycle, corroborating the conclusions drawn from the analysis of the phase space diagram.

3. The fit of the simplified model is satisfactory. In particular, the coefficients of determination obtained in the case of the idealised channel are comparable with those obtained for the synthetic validation experiment. This means that the simplified model explains roughly the same amount of data variance in the two cases. The analysis of the residuals revealed no severe modelling defects,

although it also showed that higher order auto-regressive models may slightly enhance fit performances.

4. The estimated parameters can be associated with typical time scales by means of simple physical arguments. The time scales I obtained are plausible and range from one to a few weeks, consistently with the predictions of traditional linear perturbation theory.

5. The estimated parameters can be used to predict a number of properties of the diagnosed time series of X and Y, including their correlation functions and marginal probability distribution functions (PDFs). The simplified model can reproduce the PDFs to a very good approximation. The prediction is less accurate for the correlation functions but, importantly, I have also shown that the qualitative structure of the cross-correlation function is correctly captured.

6. The estimated parameters can also be used to compute deterministic phase space trajectories. Here, I have shown that such trajectories dissipate energy and are therefore qualitatively different from the kernel-averaged orbits, demonstrating for the first time that phase space oscillations can be found even when the deterministic dynamics are dissipative. The implication is that a phase space diagram with quasi-periodic orbits does not constitute a sufficient condition for the dynamics to be conservative.

7. My results do not depend strongly on the choice of the averaging domain, as long as this is not wide enough to contain multiple ACC jets or as narrow as to include only a few grid points.

The results lend support to the idea that reduced-order dynamical models can be effectively employed to study the mechanisms and time scales of mesoscale eddy variability in the Southern Ocean, but a number of important questions remain open. Firstly, I have shown that a predator-prey relationship between eddies and the mean flow can be found in the idealised channel configuration of the MITgcm, but are the implied dynamics actually at play in the real Southern Ocean? Given the many simplifying assumptions made to configure the idealised channel (including constant, zonally symmetric atmospheric forcing, no sea ice, no salinity, and a flat bottom topography), the answer is not obvious. I take a first step in this direction in the next chapter, where I analyse data from a realistic Southern Ocean state estimate product, the SOSE. Secondly, is it possible to exploit the simplified model to make predictions about the time scales of forced response to wind stress changes? Ideally, one could envision numerical step-change experiments where an instantaneous perturbation is applied to one or more of the simplified model's parameters to represent an abrupt increase in wind stress. Unfortunately, this would require knowledge of how the simplified model's parameters scale with the wind strength, which is not known because, following AN14, the simplified model was introduced based on heuristic arguments. Recently, Kobras et al. (2022) made inroads towards deriving a reduced-order model of atmospheric variability from successive approximations of the equations of motion, starting from the two-level QG equations on a beta plane. Although their final model comprises six equations, and is thus inherently more complicated than the one considered here, their study may constitute a promising way forward for the oceanic case too. Alternatively, the dependence of the simplified model's parameters on surface wind stress could be estimated by running the idealised channel with different values of the wind stress parameter τ_0, and fitting the simplified model to the data from each simulation. This "sensitivity experiment" approach has significant provenance, but is computationally demanding as it requires to (i) spin up and (ii) diagnose large amounts of data from at least a few independent configurations of the idealised channel: in view of this, the idea is not pursued further here. Rather, in chapter 6 I study the transient response of the Southern Ocean to wind stress changes directly by generating an ensemble of wind step-change simulations with the idealised channel.

Chapter 5

Time-scales of natural variability in the Southern Ocean: part 2

5.1 Introduction

In chapter 4, I have shown that the interaction between eddies and the mean flow in an idealised channel configuration of the MITgcm is well described by a linear stochastic damped oscillator, named the simplified model. Specifically, the simplified model was fitted to the time series obtained by taking spatial averages of the Eady growth rate and of the eddy buoyancy flux, representing mean flow and eddy activity respectively: the accuracy of the fit was successfully tested against conventional metrics of evaluation, and the simplified model proved able to reproduce the statistical properties of the two time series satisfactorily. Scrutiny of the phase space diagram and of the correlation functions further revealed that the dynamics expressed by the simplified model are compatible with the predator-prey evolution model introduced by AN14 to study the variability of the atmospheric storm track, a result that strengthens the evidence pointing to dynamical analogies between the atmospheric jet stream and the ACC. Given the many idealising assumptions involved in the MITgcm configuration used for the study, however, it remains unclear to what extent these results apply to the real Southern Ocean: here, I seek to bolster the conclusions drawn in the previous chapter by studying the interaction between eddies and the mean flow with a realistic state-estimate of the Southern Ocean, the SOSE.

The chapter is set out as follows. The dataset considered for the analysis is presented in section 5.2: this includes an introduction to SOSE (section 5.2.1), a brief illustration of SOSE's climatological state (section 5.2.2), and a discussion of how the dynamical variables are defined and spatially averaged (section 5.2.3). In section 5.3 the fit method is described, which is more flexibly based on linear regression rather than on the Yule-Walker equations as in the previous chapter. Results are presented in section 5.4. I conclude with a summary and a discussion of perspectives in section 5.5.

5.2 Data

The purpose of this section is to offer a brief description of SOSE's model configuration and physics, and to detail how the dynamical variables used to represent eddies and the mean flow are computed.

5.2.1 SOSE: an overview

The acronym SOSE stands for Southern Ocean State Estimate. In climate science, the expression "state estimate" indicates that the output of a GCM is coupled to actual measurements to produce an accurate representation of a geophysical system. The SOSE is a time-evolving state estimate of the Southern Ocean (south of 25°S) that combines the solution of an eddy permitting configuration of the MITgcm with observations from a variety of oceanic and meteorological datasets (Mazloff et al., 2010). Model data and observations are coupled according to the data assimilation technique developed by the ECCO consortium.

Several versions of SOSE are made publicly available at `http://sose.ucsd.edu/`. In this work, we employ SOSE iteration 100[1], which covers six years of data from 1 January 2005 to 31 December 2010. The model is a realistic configuration of the MITgcm at $1/6°$ horizontal resolution and with 42 vertical levels of varying thickness. Importantly, the model is equipped with a non-linear, two-component equation of state, realistic bottom topography, a sea-ice model (Hibler, 1980), the KPP scheme for mixed layer parametrisation, and a boundary layer scheme for the parametrisation of heat, salinity, and momentum air-sea fluxes (Large and Yeager, 2004). In comparison, the idealised channel features a linear and single-component equation of state (i.e. no salinity), no sea ice, flat bottom topography, and prescribed air-sea fluxes. SOSE's observational constraints come from a variety of products, including Argo floats. Detailed information can be found in Mazloff et al. (2010), or on the Climate Data Guide web page `https://climatedataguide.ucar.edu/climate-data/southern-ocean-state-estimate-sose`. The data assimilation procedure consists in minimising a cost function representing the weighted discrepancy between the model's solution and the observations, where weights are assigned depending on each measurement's uncertainty. In the case of SOSE, the cost function is minimised by iteratively modifying a control vector which contains information about the model's initial conditions and boundary conditions at the surface. The state estimate is then obtained by marching the model forward in time with the optimised initial and boundary conditions. This method should be compared with the conventional data assimilation approach used, for example, in operational weather forecast, where it is the solution to the numerical model itself that is adjusted to better represent observations. In contrast, SOSE remains an exact solution of the MITgcm, and is therefore physically self-consistent.

SOSE's horizontal grid resolution does not fully resolve the first Rossby radius of deformation (which ranges from approximately 10 to 30 km in the Southern Ocean, Chelton et al. (1998)), but the model naturally develops a rich mesoscale eddy field Cerovečki et al. (2019). SOSE has been extensively validated against

[1]Computational resources for the SOSE were provided by NSF XSEDE resource grant OCE130007

observations and estimate products for both the 2005-2006 period (Mazloff et al., 2010) and the full 2005-2010 period (Cerovečki et al., 2019). Although it retains biases with respect to observations (e.g. SST in the subtropics and in the subpolar gyres (Mazloff et al., 2010); isopycnal range of low potential vorticity water, climatological location of the ACC fronts, depth of the mixed layer (Cerovečki et al., 2019)), it provides a reliable representation of the state of the real Southern Ocean.

5.2.2 SOSE: climatology

In this section, selected aspects of SOSE's climatology are explored. Here, by climatology we mean that time-averaged fields are the object of investigation, with time averages computed over SOSE's complete 6-year time span from January 2005 to December 2010. Our limited purpose is to provide the reader with a general feel for the model's physics, while we refer to the cited literature for a comprehensive illustration. In the interest of brevity, we focus on surface properties and on the thermal wind relation.

The surface: Sea Surface Temperature, Sea Surface Height, and fronts

Time-mean Sea Surface Temperature (SST) is shown in figure 5.1 (left). Isothermal lines are approximately zonal, with the coldest temperatures (around zero degrees) found south of the ACC region and near the Antarctic continent, and the warmest (≈ 27 degrees) at the lowest latitudes. Deviations from zonal symmetry are nevertheless apparent throughout the domain, and more markedly so in the vicinity of topographical features, e.g. downstream of Drake passage (at about 60°W) and New Zealand (at about 170°E). The meandering, non-zonal character of the ACC is also manifest in the time-mean surface zonal velocity, figure 5.1 (right).

In spite of the filamented nature of the flow (Marshall and Speer, 2012, Rintoul et al., 2001), regions of strong horizontal gradients of surface properties (or fronts,

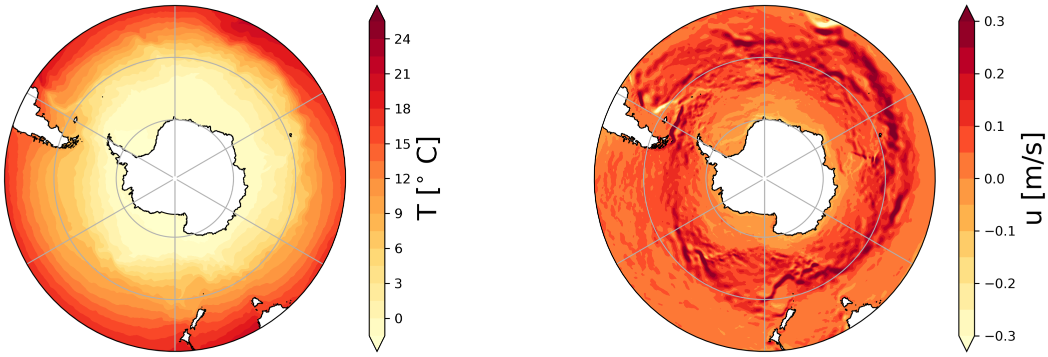

Figure 5.1: Left: time-averaged Sea Surface Temperature. Right: time-averaged surface zonal velocity. Time averages are computed over the full $2005-2010$ period.

following Orsi et al. (1995)) can be exploited to identify the climatological positions of the ACC jets (Olbers et al., 2004). One possibility (Mazloff et al., 2010) is to compute the time-mean, vertically integrated horizontal streamfunction ψ_h:

$$\psi_h(x,y) = \int_{y_S}^{y} \mathrm{d}y \int_{-H(x,y)}^{0} \mathrm{d}z\,\overline{u}(x,y,z), \tag{5.1}$$

which is shown in figure 5.2 (left). By definition, the vertically integrated transport between two streamlines is equal to the difference between the corresponding values of ψ_h. Cerovečki et al. (2019) observe that the 10, 50, and 100 Sv contours of ψ_h (green contours) tend to align with the climatological positions of the SACCF (Southern ACC Front), PF, and SAF respectively in the Southeast Pacific region (the alignment was not tested for other regions). Although characterised by a large-scale zonal symmetry, the fronts show significant local meridional excursions too: as noted by Olbers et al. (2004), all fronts pass through Drake passage, but take otherwise independent circumpolar paths that do not necessarily follow latitude circles. Poleward of the main ACC flow, the Weddell (located around 10°E) and Ross (200°E) cyclonic polar gyres contribute negatively to the transport. Figure 5.2 (right) shows time-mean Sea Surface Height (SSH) anomaly profiles: for illustration purposes, the -0.85, -0.6, and -0.1 m SSH contours are evidenced by black contours, demonstrating an approximate correspondence between SSH and vertically integrated horizontal streamfunction profiles.

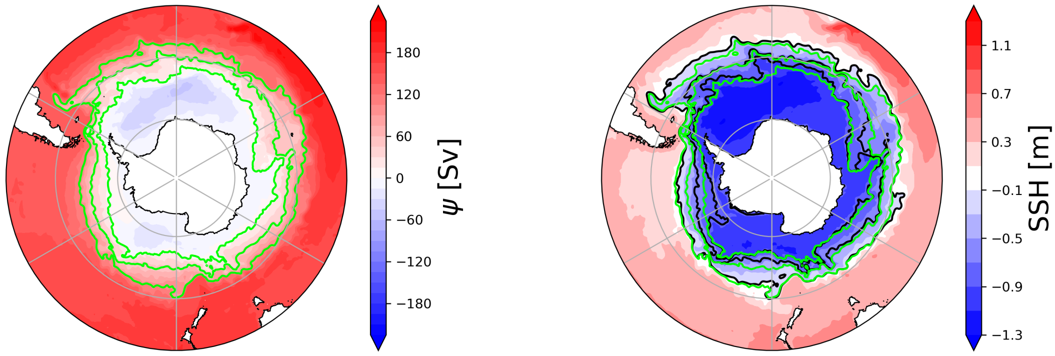

Figure 5.2: Left: time-averaged, vertically-integrated zonal streamfunction. Green contours mark the climatological position of the 10, 50, and 100 Sv levels. Right: time-averaged Sea Surface Height anomaly. Black contours mark the position of the -0.85, -0.6, and -0.1 m levels. Green contours are as in the left panel.

Meridional structure

The most remarkable feature of meridional sections of fluid properties in the ACC region is that isolines are negatively tilted (Rintoul et al., 2001). Dynamically, the sloping profiles are associated with the thermal wind balance, as explained below (Marshall and Speer (2012); see also chapter 3). Hydrographical sections are further characterised by the presence of a series of steps (i.e. narrow regions of strong meridional gradients), which tend to be associated with the location of the major ACC fronts (Rintoul et al., 2001). The time-mean buoyancy contours for two meridional sections at approximately 350°E (section taken in the Atlantic sector) and 240°E (Pacific sector) are shown in figure 5.3. In both cases, the largest values of buoyancy (associated to warm and relatively salty water, with thermal effects dominating over haline contributions) are found near the surface and at the southern boundary of the domain. In the Atlantic section, two regions of steep meridional gradient can be observed at approximately 45°S and 57°S, corresponding to the climatological position of the SAF and PF respectively (compare also with figure 5.2 (left)). The same fronts can be detected at different latitudes in the Pacific section (55°S and 65°S), again consistent with the location of the vertically integrated streamfunction profiles shown in figure 5.2 (left). For both sections, the SAF marks

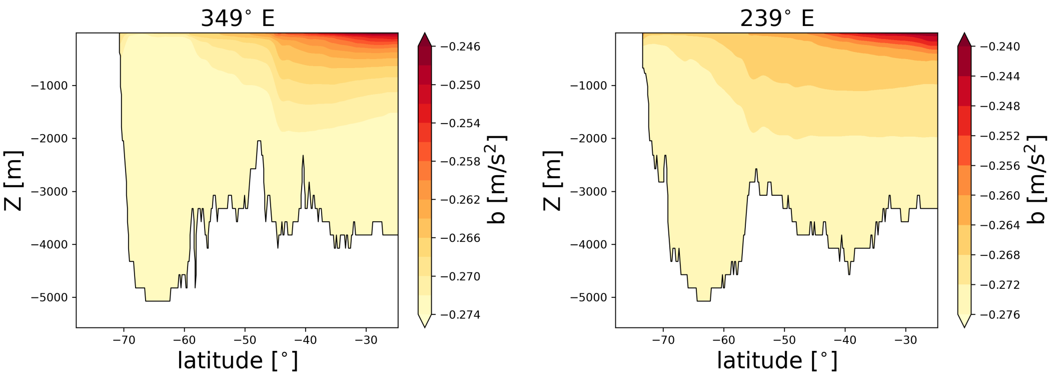

Figure 5.3: Meridional sections of time-mean buoyancy taken at approximately 350°E (Left) and 240°E (Right).

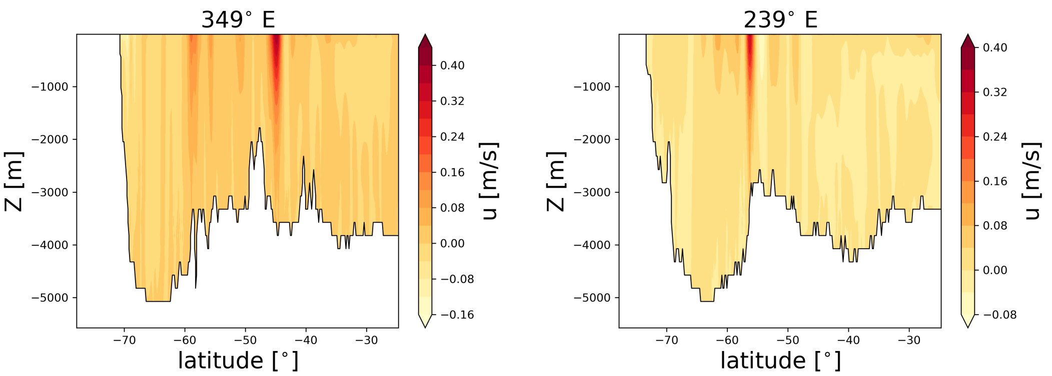

Figure 5.4: Meridional sections of time-mean zonal velocity taken at approximately 350°E (Left) and 240°E (Right).

the southern boundary of the region interested by the largest upward and southward sloping of the isopycnals. The meridional sections of time-mean zonal velocity taken at the same longitudes are shown in figure 5.4. Intense and narrow eastward jets are found at approximately the same latitudes as the meridional steps in time-mean buoyancy, lending support to the dynamical association between fronts and zonal flow. The meridional gradient of buoyancy is related to the vertical shear of zonal velocity via the thermal wind relation, equation (3.23). Figure 5.5 shows the left hand side (vertical shear of time-mean zonal velocity, colours) and the right hand side (minus the meridional gradient of time-mean buoyancy over the Coriolis parameter, contours) of equation (3.23) for the same meridional sections as before and for the upper thermocline. Note that the fields are not zonally averaged. The

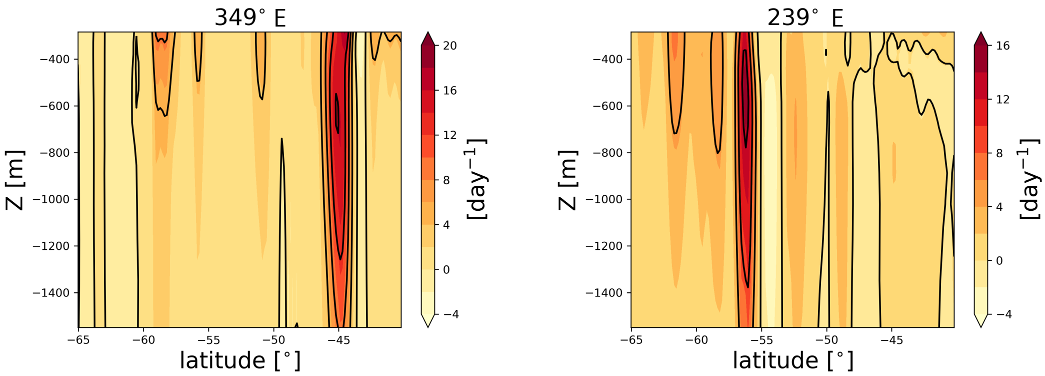

Figure 5.5: Thermal wind balance. Meridional sections of the vertical shear of time-mean zonal velocity (colours) and of $-\partial_y \overline{b}/f$ taken at approximately 350°E (Left) and 240°E (Right).

plots qualitatively demonstrate that thermal wind balance is satisfied to a good approximation. Quantitatively, the thermal wind relation holds within a 20% accuracy at worse, attained at the intense Atlantic jet. The relative error decreases to order 1% away from the jet.

5.2.3 Definition of the dynamical variables

We introduce the dynamical variables X and Y describing mean flow and eddy activity respectively. In chapter 4, we used the Eady growth rate to measure mean flow, equation (4.93), and the eddy buoyancy flux to measure eddy activity, equation (4.96). Both definitions involve application of the zonal average operator, a choice justified by zonal symmetry in the case of the idealised channel. SOSE's time-averaged flow, in contrast, is steered by topography, and the presence of standing meanders breaks zonal symmetry (see for example figure 5.2). Therefore, the ACC flow in SOSE is best characterised by taking averages "following the stream" rather than along a latitude circle. In view of this, we adjust the definitions of section 4.3 by (i) decomposing the horizontal velocity into its along- and across-stream rather than zonal and meridional components, and (ii) taking averages along time-mean SSH profiles rather than along latitude circles, as explained below.

Along- and across-stream components of velocity

Following Karsten and Marshall (2002) and Olbers et al. (2004), we use time-mean Sea Surface Height (SSH) profiles (see figure 5.2 (right)) to identify time-mean streamlines. This choice is motivated by the following considerations: (i) the method is conceptually simple, (ii) it has provenance in the literature, and (iii) time-mean SSH profiles are readily available from model diagnostics. The streamlines thus defined constitute a working approximation of the time-mean circumpolar flow of water masses in the Southern Ocean which, at the cost of an added layer of complication, may be defined based on the baroclinic shear of surface properties (Orsi et al., 1995). There are, of course, a number of alternatives, which we now briefly discuss.

Treguier et al. (2007) defined time-mean streamlines based on the mean contours of the barotropic streamfunction ψ_h . According to the authors, one of the advantages of this choice is that ψ_h is a vertically integrated quantity, therefore the streamlines do not depend on depth by construction. Importantly, though, they also note that the results do not change significantly if SSH is used instead. In view of this, and given the approximate equivalence between mean SSH and ψ_h contours in SOSE (figure 5.2 (right)), we expect that the main thrust of our argument would not vary substantially were time-mean SSH profiles to be replaced with ψ_h ones. In similar spirit, one may use f/H contours to define the streamlines. The rationale is that the ACC is steered by bottom topography and, at leading order, its flow is directed along f/H contours due to conservation of potential vorticity (Patmore et al., 2019). However, this method still constitutes a depth-independent approximation of the streamlines based on the dynamical properties of the ACC, and there is no obvious argument that suggests it would afford a more accurate representation than that obtained by tracking the flow directly with SSH or ψ_h contours. Viebahn and Eden (2012) defined the streamlines based on the time-mean horizontal velocity contours. This choice is rigorous, but is mathematically complicated as the along-stream and across-stream directions depend on depth, which yields a non-orthogonal system of reference. Besides, the structure of the flow being strongly barotropic in the Southern Ocean (Hughes and Ash, 2001), we expect SSH to be a reasonable indicator of the flow in the interior too. Other choices, such as the Bernoulli potential contours used by Polton and Marshall (2007), equally involve a degree of mathematical sophistication.

In summary, using time-mean SSH contours to define the streamlines offers a reasonable compromise between rigour and practicality. A possible improvement would be to define the along- and across- stream directions based on a time-varying SSH field. This may account for slow-time scale deviations of the flow from its time-averaged path: as the ACC varies significantly in time, we expect that the change would reflect in a stronger along-stream mean flow. A likely complication, though, is that the definition of mean flow might then absorb part of the transient eddy cross-stream fluxes, as in the limit case when the streamlines coincide with the instantaneous SSH field (and the across-stream geostrophic velocity is thus identically zero). This warrants a careful choice of the definition of the time-varying streamlines (and thus of what we consider as eddying motion), which should be tested e.g. by applying time filters of varying width to the SSH field. For the sake of simplicity, in this chapter we limit ourselves to the simpler choice of using time-mean SSH contours. Let $\overline{\eta} = \overline{\eta}(x,y)$ be the time-averaged SSH, and let (u,v) be the zonal and meridional components of the horizontal velocity. At each grid-point, we compute the along- and across- stream components $(v_{\parallel}, v_{\perp})$ according to the formula:

$$v_{\parallel} = \frac{\partial_y \bar{\eta}}{||\nabla \bar{\eta}||} u - \frac{\partial_x \bar{\eta}}{||\nabla \bar{\eta}||} v \tag{5.2}$$

$$v_{\perp} = \frac{\partial_x \bar{\eta}}{||\nabla \bar{\eta}||} u + \frac{\partial_y \bar{\eta}}{||\nabla \bar{\eta}||} v, \tag{5.3}$$

where $\nabla \bar{\eta} = (\partial_x \bar{\eta}, \partial_y \bar{\eta})$ is the horizontal SSH gradient, and $||\cdot||$ denotes the Euclidean norm:

$$||\nabla \bar{\eta}|| = \sqrt{(\partial_x \bar{\eta})^2 + (\partial_y \bar{\eta})^2}. \tag{5.4}$$

More compactly, the equations for $(v_{\parallel}, v_{\perp})$ describe a change of basis transformation, from Cartesian to stream-aligned coordinates:

$$\begin{bmatrix} v_{\|} \\ v_{\perp} \end{bmatrix} = S \begin{bmatrix} u \\ v \end{bmatrix}, \tag{5.5}$$

where the change of basis matrix S is:

$$S = \frac{1}{||\nabla \bar{\eta}||} \begin{bmatrix} \partial_y \bar{\eta} & -\partial_x \bar{\eta} \\ \partial_x \bar{\eta} & \partial_y \bar{\eta} \end{bmatrix}. \tag{5.6}$$

Note that the matrix S depends on the horizontal coordinates via $\overline{\eta}(x,y)$, but not on depth or time.

Stream-wise average

Let $\gamma_1 = \gamma_1(x)$ and $\gamma_2 = \gamma_2(x)$ be the circumpolar paths associated to two arbitrary time-mean SSH levels $\eta_2 > \eta_1$ (here, x is the longitude and y the latitude). By definition, the paths γ_1 and γ_2 are such that the value of $\overline{\eta}$ is constant along the path, i.e. $\overline{\eta}(x,\gamma_1(x)) = \eta_1$ and $\overline{\eta}(x,\gamma_2(x)) = \eta_2$ $\forall x$. For simplicity, we assume that $\eta_2 > \eta_1$ implies $\gamma_2(x) > \gamma_1(x)$ $\forall x$, and that $\overline{\eta}(x,\gamma(x)) = \eta$ defines a unique connected path, a condition that is satisfied over most of the ACC region (see figure 5.2 (right)). The stream-wise average $\langle q \rangle$ of a scalar variable $q = q(x,y,z)$ between the contour levels η_1 and η_2 is defined by:

$$\langle q \rangle(z) = \frac{1}{A} \oint \mathrm{d}x \int_{\gamma_1(x)}^{\gamma_2(x)} \mathrm{d}y q(x,y,z), \tag{5.7}$$

where:

$$A = \oint \mathrm{d}x \int_{\gamma_1(x)}^{\gamma_2(x)} \mathrm{d}y \tag{5.8}$$

is the area of the horizontal surface bounded by the circumpolar paths γ_1 and γ_2. Figure 5.6 offers a comparison between zonal and stream-wise averaging: figure 5.6 (left) shows the conventional time-mean, zonal-mean profiles of zonal velocity. Figure 5.6 (right) in contrast, shows the time mean, stream-averaged profiles of along-stream velocity. Here, the stream average is computed between uniformly spaced circumpolar paths separated by an interval $\Delta\eta = 0.052$ m, and parsing SSH

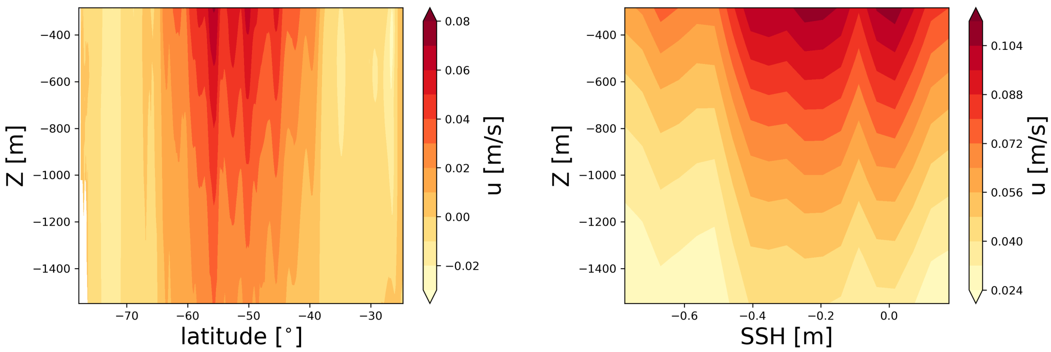

Figure 5.6: Left: time-mean, zonal-mean profiles of zonal velocity. Right: time-mean, stream-wise averaged profiles of along stream velocity.

values from -0.8 m to 0.2 m (namely, $\eta_1 = -0.8$ m, $\eta_j = \eta_1 + (j-1)\Delta\eta$ for $j = 2, \ldots, N+1$ with N such that $\eta_{N+1} = 0.2$ m). Similarly to Olbers et al. (2004), we observe that: (i) the zonal average profiles are characterised by the co-existence of multiple jets of comparable magnitude, while only a few, larger and more intense jets are found in the stream-averaged section. Interestingly, most of the zonal flow is concentrated equatorward of $\overline{\eta} = -0.6$, i.e. north of the PF. Decreasing the parameter $\Delta\eta$ (which amounts to increasing the resolution in SSH space) adds fine scale detail but does not change the overall qualitative structure of the meridional profiles. (ii) The stream-wise averaged velocity is positive (i.e. eastward) everywhere in the depth range considered, whereas negative values likely associated to the presence of standing meanders can be observed for the zonally averaged section. (iii) The maximum of the stream-wise averaged velocity is almost a factor 2 larger than that of the zonally-averaged velocity, expressing the fact that the stream-wise average prevents spurious cancellations due to the non-zonal meanders. Collectively, these observations strengthen the idea that the stream-wise average is better suited to accurately capture the dynamics of the ACC flow in SOSE than the simpler zonal average.

The paths γ_1 and γ_2 in equation (5.7) need not describe closed circuits around the pole, but may only span a finite longitudinal extent. In this case, the formula for the stream-wise average reads:

$$\langle q\rangle(z)=\frac{1}{A}\int_{x_1}^{x_2}\mathrm{d}x\int_{\gamma_1(x)}^{\gamma_2(x)}\mathrm{d}y\,q(x,y,z), \tag{5.9}$$

with:

$$A=\int_{x_1}^{x_2}\mathrm{d}x\int_{\gamma_1(x)}^{\gamma_2(x)}\mathrm{d}y\,q(x,y,z), \tag{5.10}$$

and where x_1 and x_2 mark the longitudinal boundaries of the domain under consideration. We exploit this formula to define the dynamical variables with an emphasis on local rather than hemispheric properties.

Spatial averaging

Equipped with the notion of stream-wise average, we introduce the dynamical variables X and Y. As in chapter 4, the definitions hinge on the Eady growth rate and on the eddy buoyancy flux. Given a longitude interval $[x_1,x_2]$ and two paths $\gamma_1(x),\gamma_2(x)$ associated to the SSH levels η_1,η_2, the Eady growth rate is computed as:

$$\omega=-0.31\langle f\rangle\frac{\langle\partial_z v_{\parallel}\rangle}{\sqrt{\langle\partial_z b\rangle}}\approx-0.31\frac{\langle f\partial_z v_{\parallel}\rangle}{\sqrt{\langle\partial_z b\rangle}}. \tag{5.11}$$

This formula is inherited from equation (4.93), with the difference that the zonal velocity u is replaced by the along-stream velocity $v_{\parallel}$ and that zonal mean is replaced by the stream-wise average (denoted by angular brackets). Here, the stream-wise average is taken over the domain specified by $[x_1,x_2]$ and $\gamma_1(x),\gamma_2(x)$ according to definition (5.9). Experiments with time-averaged fields suggest that multiplication by the Coriolis parameter approximately commutes with the stream-wise average operation, hence the second equality. The eddy buoyancy flux is computed according to:

$$\mathscr{F}_y=\langle v_{\perp}^{+}b^{+}\rangle, \tag{5.12}$$

which is the same as equation (4.96), but with the meridional velocity replaced by the across-stream velocity and with zonal average replaced by the stream-wise average (similarly to section 4.3, $^{+}$ denotes deviations from time mean). As an example, the time mean, stream-wise averaged Eady growth rate and eddy buoyancy

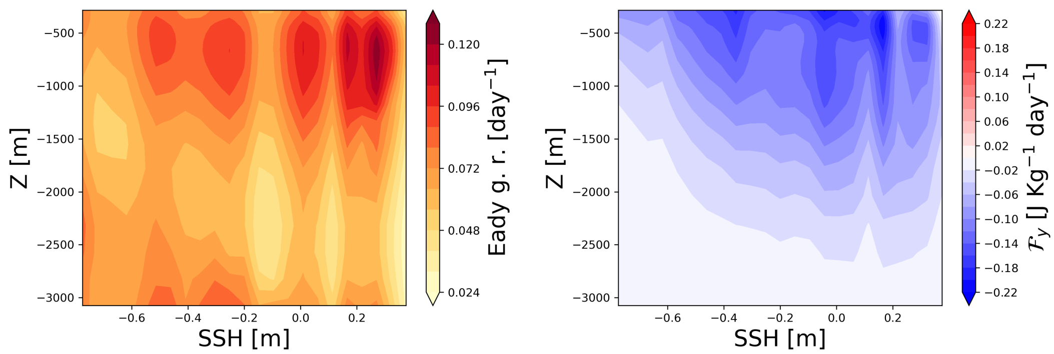

Figure 5.7: Left: time mean, stream-wise average of the Eady growth rate. Right: time-mean, stream-wise average of the eddy buoyancy flux. The stream-wise average is taken between 90°W and 175°W (Indian sector). SSH levels are distanced by an interval $\Delta\eta = 0.052$ m.

flux are shown in figure 5.7, where the stream-wise average is computed over the Indian sector (i.e., between 90°W and 175°W). Note that the Eady growth rate ω and the eddy buoyancy flux $\mathscr{F}_y$, as defined by equations (5.11) and (5.12), are still spatially-structured fields depending on depth z. In order to extract time series from ω and $\mathscr{F}_y$, we take their vertical average over a depth interval $[z_1, z_2]$:

$$\tilde{\omega}(t) = \frac{1}{z_2 - z_1}\int_{z_1}^{z_2} \mathrm{d}z\, \omega(z,t) \tag{5.13}$$

$$\tilde{\mathscr{F}}_y(t) = \frac{1}{z_2 - z_1}\int_{z_1}^{z_2} \mathrm{d}z\, \mathscr{F}(z,t)\,. \tag{5.14}$$

The time series of X and Y are computed from $\tilde{\omega}$ and $\tilde{\mathscr{F}}_y$ by subtracting the corresponding time mean (as in chapter 4) and by removing linear trends (to account for possible time mean drifts in the dataset).

In the case of the idealised channel, the time series of X and Y where computed relative to a single domain located in the middle of the channel. This choice was motivated primarily by the symmetric, simplified geometry of the model configuration. Here, instead, we account for potential local differences in the character of eddy-mean flow interaction by partitioning the ACC region in a greater number of domains Ω_i, where each domain is specified by a choice of x_1, x_2 (the longitudinal extent), z_1, z_2 (the depth range), and η_1, η_2 (the stream paths). In all cases, the depth range is from $z = -1173.5$ m to $z = -550$, so that X and Y are vertically averaged in the thermocline and below the mixed layer, where the seasonal cycle can be neglected. The longitudinal extent of the domains is of approximately 20°, and the SSH paths are taken $\Delta\eta = 0.1$ m apart, from $\overline{\eta} = -0.6$ m to $\overline{\eta} = 0.2$ (corresponding to the region in SSH space where most of the flow is concentrated, see figure 5.6). In total, there are 136 domains that tile the ACC region. We observed that a small number of domains are associated with time series valued far outside the average range of variability (the reasons for this behaviour are unclear, and the domains are not spatially clustered following any obvious patterns): they were flagged as outliers and excluded from the analysis below, leaving a total of 124 domains covering most of the ACC region, and shown in figure 5.8

5.3 Methods

We consider the same problem of section 4.4, which we restate here for convenience: we have extracted the time series X_n and Y_n of the two dynamical variables X and Y from the data, representing Eady growth rate and eddy buoyancy flux respectively. The time interval between two consecutive observations is $\Delta t = 5$ days, $n = 1, \ldots, N$, and we hypothesise that the time series are realisations of the simplified model (4.51):

$$\mathbf{X}_{n+1} = A\mathbf{X}_n + \sigma \mathrm{d}\mathbf{W}_{n+1}, \tag{5.15}$$

The task is to fit the simplified model to the data, compute the best estimate for the unknown matrices A and σ (for a total of 6 free scalar parameters), and evaluate the goodness of the fit. In chapter 4 we tackled the problem with the Yule-Walker equations. Here, instead, we rely on linear regression, as explained below.

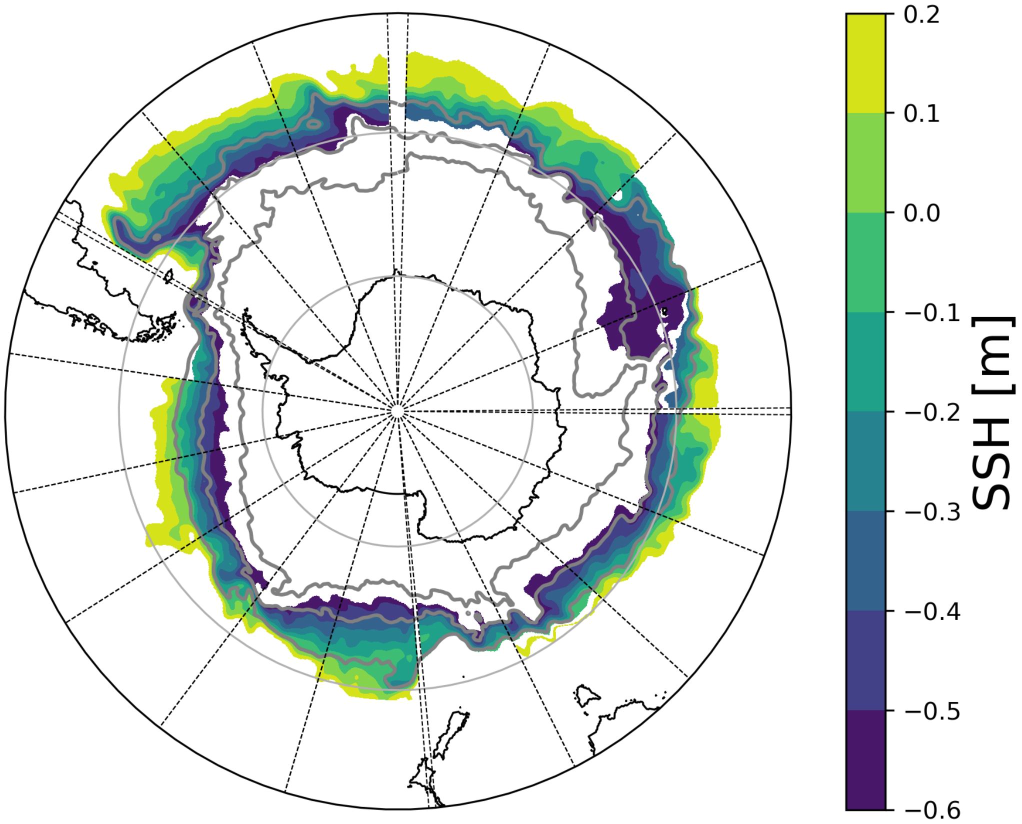

Figure 5.8: Partition of the ACC region into the smaller domains Ω_i. The domains are bounded meridionally by SSH paths uniformly spaced by $\Delta\eta = 0.1$ m (coloured intervals), and are approximately 20° wide in longitude (the longitudinal extent is marked by the dashed black lines). The continuous grey lines show the 10, 50, and 100 Sv levels of the time-averaged, vertically-integrated zonal streamfunction. The small gaps between the regions considered for the analysis were inadvertently introduced when accessing the SOSE dataset, but have no foreseen relevance to the analysis to follow.

5.3.1 Linear regression

Similarly to section 4.4.3, we define:

$$\mathbf{x}_n = \mathbf{X}_n \tag{5.16}$$

$$\mathbf{y}_n = \mathbf{X}_{n+1}\,, \tag{5.17}$$

for $n = 1,...,N-1$. The best estimate $\hat{A}$ of the matrix A is computed by performing an ordinary linear regression (see e.g. Ross (2014) for the theory. In practice, the linear regression is implemented using the Python library scikit-learn, Pedregosa et al. (2011)), with $\mathbf{x}$ and $\mathbf{y}$ being the independent and dependent variables respectively. A similar approach was promoted by Gnanadesikan et al. (2020) to estimate the parameters of a linear model describing oscillatory convection in the Southern Ocean. The values of $\mathbf{y}$ predicted by the linear model are:

$$\hat{\mathbf{y}}_n = \hat{A}\mathbf{x}_n, \tag{5.18}$$

for $n = 1,...,N-1$. The best estimate $\hat{\sigma}$ of the matrix σ is computed from the residuals $\mathbf{e} = \mathbf{y} - \hat{\mathbf{y}}$ as explained in section 4.4.3

5.4 Results

5.4.1 Summary

We test the working hypothesis that the interaction between eddies and mean flow in SOSE is described by a stochastically forced, damped oscillator, the simplified model. The data we analyse consist of 124 bivariate time series (X_n, Y_n), where for each time series X_n is the spatial average of the Eady growth rate (measuring mean flow) and Y_n is the spatial average of the eddy buoyancy flux (measuring eddy activity). The spatial averages exploit the stream-wise averaging technique and are taken at different locations so that, overall, the 124 domains tile most of the ACC region (see figure 5.8). Each time series is 6 years long, and the sampling interval is

5 days. The simplified model is presented in section 4.2.7, together with some of its fundamental mathematical properties. Contrary to chapter 4, the fit method is based on linear regression (section 5.3.1) and not on the Yule-Walker equations but, for a given time series, the fit procedure is otherwise unchanged: section 4.5 contains a summary of the main steps.

5.4.2 Analysis of data from individual domains: local dynamics

From an analytical perspective, there are two important differences between the present case and that of chapter 4: (i) the data is spatially structured, and comprises multiple time series associated with different physical locations, and (ii) the time series cover a shorter period of time (6 rather than 30 years). As a first step, we have individually fitted the simplified model to the bivariate time series from each of the averaging regions. This procedure reveals that the spatial domains can exhibit different behaviours, with some showing clear qualitative evidence of predator-prey oscillations, as detected in the structure of the phase space diagram and of the cross correlation function, and others for which supporting evidence of AN14 dynamics cannot be easily found (a more formal criterion to distinguish between the two cases will be discussed below). As an example, figure 5.9 shows the results of the fit for a domain falling into the first case (with evidence of oscillatory behaviour), while figure 5.10 shows those for a domain falling into the second (without evidence of oscillatory behaviour). The two domains are purposely selected to best demonstrate the argument, but the quantitative results of this section do not depend on their precise choice.

We begin with the scrutiny of the fit outcome for the domain showing evidence of predator-prey dynamics. Figures 5.9 (a) and (d) illustrate the goodness of the fit qualitatively. Figure 5.9 (a) shows the time series of X and Y (blue and red dots respectively) and the corresponding predictions obtained with linear regression one step ahead (continuous blue and red lines): the simplified model captures the time evolution of both variables satisfactorily. Figure 5.9 (d) shows the prediction plots for X and Y. We observe that the scatter points are clustered around the 1:1 line, supporting the view that the simplified model adequately describes the data for this domain. Figures 5.9 (b) and (c) allow ourselves to investigate whether or not the interaction between the dynamical variables is mediated by damped oscillations. Figure 5.9 (b) shows the phase space diagram reconstructed with the Gaussian kernel method of section 4.2.8. Phase space trajectories are characterised by a high degree of coherence, and are consistent with the signature of the predator-prey life cycle studied in chapter 4. The phase space pattern is noisier than in the case of the idealised channel, and the shape of the trajectories is only approximately elliptical (compare for example with figure 4.7 (right)). These effects, however, are most likely an artefact of SOSE's comparatively short time series, see section 5.4.8 for details. Further evidence comes from the inspection of the cross-correlation function, shown in the bottom panel of figure 5.9 (c). Albeit noisy, the sample cross correlation (dotted green line) is approximately an odd function of the time lag with a negative peak at positive lags (eddies intensify after an increase in mean flow) and a positive peak at negative lags (mean flow becomes stronger after a decrease in eddy activity). The best fit prediction for the cross-correlation (continuous black line), in addition, captures this behaviour well. Note that the peaks of ρ_{XY} barely stand out of the acceptance interval of the null hypothesis (in other words, the peaks of the correlation functions are just within the significance region), but this effect too may be partially attributed to the small size of the time series (because the width of the acceptance interval decreases with the square root of the number of observations, section 4.2.6). Overall, the analysis of the phase space diagram and of the cross correlation function demonstrates that eddies and mean flow interact according to the predator-prey oscillatory dynamics of chapter 4 in this domain.

Figure 5.10 shows the outcome of the fit for the domain without evidence of predator-prey dynamics. As before, figures 5.10 (a) and (d) demonstrate that the simplified model is an excellent fit to the data and, in particular, that no obvious patterns of deviation from the 1:1 line appear in the prediction plots, figure 5.10 (d). The phase space diagram and the cross correlation function, however, convey

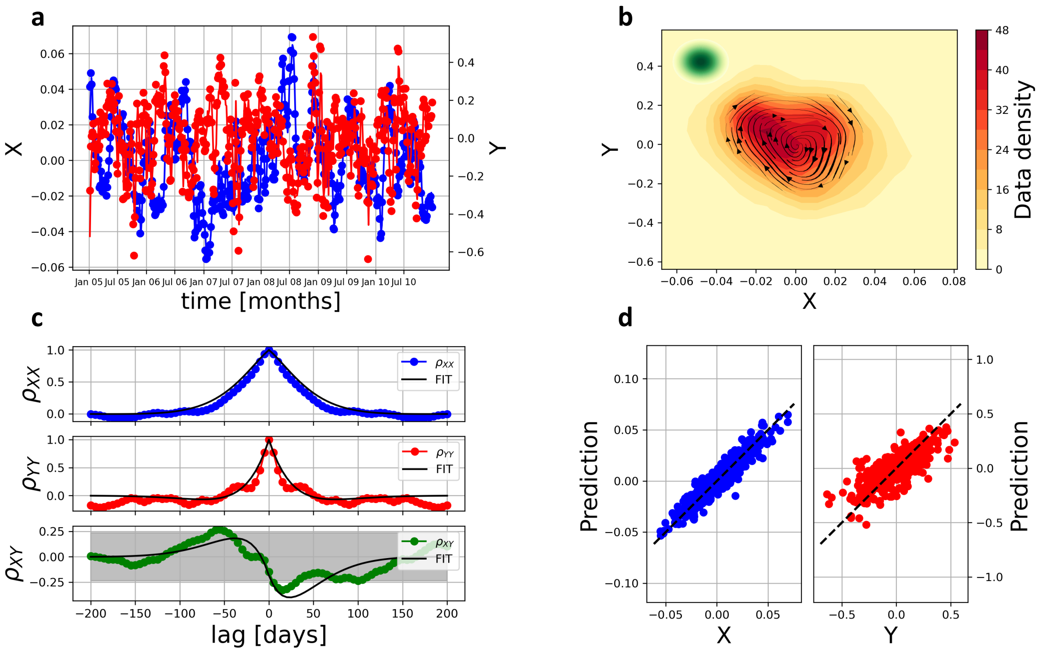

Figure 5.9: Fit of the simplified model to data from an individual domain with phase space oscillations. (a) Sample (dots) and predicted (continuous lines) time series of X (blue) and Y red. (b): Kernel averaged phase space diagram, as in figure 4.7, right. (c) Sample (coloured dotted lines) and predicted (continuous black lines) correlation functions of X and Y. As in figure 4.10, left. (d) Prediction plots for X (blue) and Y (red). As in figure 4.11, right.

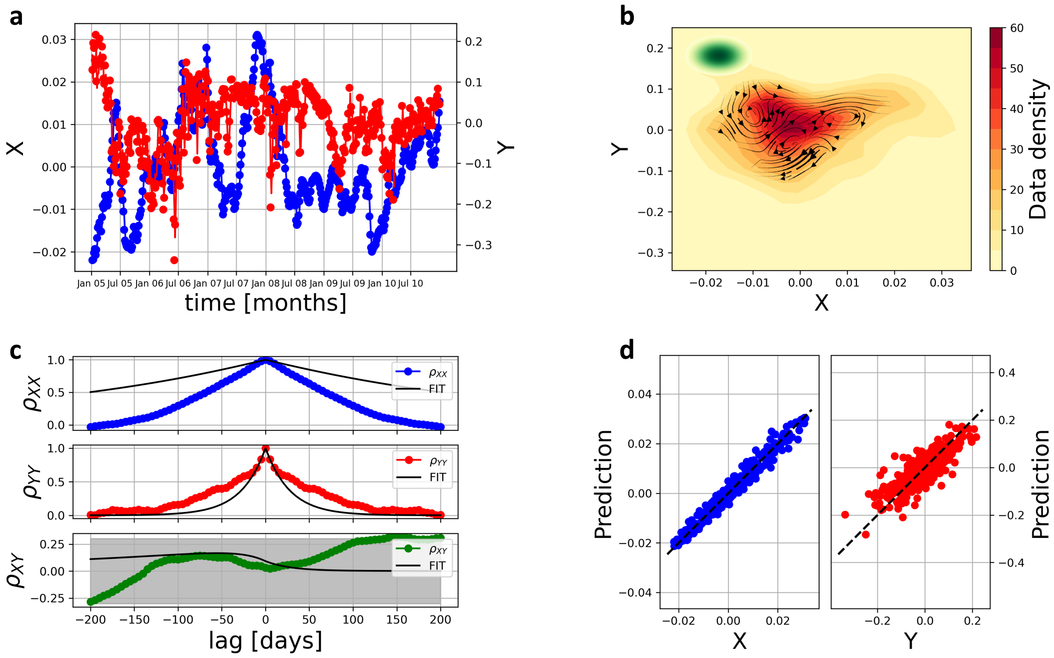

Figure 5.10: As in figure 5.9, but for a domain without phase space oscillations.

a picture markedly different from the previous case. The phase space diagram is shown in figure 5.10 (b). Here, the patterns formed by phase space trajectories are highly irregular, and characterised by the presence of a number of small scale features that cannot be associated with the predator-prey dynamics in a simple way. This could be due to a number of reasons: (i) there is no dynamical interaction between eddies and mean flow in this domain, (ii) the interaction follows AN14 but is too weak to emerge due to the small size of the time series, and (iii) the interaction does not follow AN14 (see section 5.4.8 for further details). The analysis of the cross correlation function, shown in the bottom panel of figure 5.10 (d), confirms the conclusions drawn from the study of the phase space diagram: in particular, no peaks can be detected in the sample cross correlation (dotted green line) that stand out clearly from the background. Note that ρ_{XY} takes comparatively large absolute values at large lags, but these are more likely a manifestation of the noisiness of the data rather than a signal generated by underlying dynamics. The structure of the best fit cross-correlation function (continuous black line) reflects the lack of clear cross correlation between the two variables.

5.4.3 Classification of the domains

The analysis of the two examples above allowed ourselves to put into better focus the qualitative factors differentiating the time series. Next, we seek a formal criterion to separate the dynamics of the various domains (i.e., strong versus weak or absent interaction). The key idea is that the eigenvalues of $\hat{\mathscr{A}}$ (where $\hat{\mathscr{A}}$ is the dynamical matrix estimated from linear regression) are tell-tale signs of the deterministic dynamics expressed by the model: when the eigenvalues are real, deterministic oscillations are completely suppressed by damping. When the eigenvalues have non-zero imaginary part, on the other hand, the deterministic system is in the subcritical regime and decaying oscillations will be seen. The eigenvalues are given by the formula:

$$\lambda = \frac{\gamma_{xx} + \gamma_{yy}}{2} \pm \left[\left(\frac{\gamma_{xx} - \gamma_{yy}}{2} \right)^2 + k_{xy} k_{yx} \right]^{\frac{1}{2}}, \tag{5.19}$$

where the k's are the interaction (or coupling) terms and the γ's are the damping terms. Assuming that k_{xy} and k_{yx} have opposite sign (which is the mathematical requirement to obtain oscillations), the imaginary part of λ becomes larger for larger $k_{xy}k_{yx}$. The mathematical condition that the eigenvalues have non-zero imaginary part thus reflects the physical condition that the coupling terms are strong compared to the dissipation ones.

Importantly, the eigenvalues estimated from linear regression have non-zero imaginary part in the case of figure 5.9, while they are real for figure 5.10 (the fit being successful in both cases). Motivated by this observation, we hypothesise that complex eigenvalues are associated not only with deterministic oscillations but also with the emergence of phase space quasi-periodic orbits. This allows to divide the spatial domains into two groups, where we assign each domain to the first group if the eigenvalues of $\hat{\mathscr{A}}$ have non-zero imaginary component, and to the second group otherwise. We find that, overall, this criterion captures accurately the qualitative character of the time series (see section 5.4.4 for details). The disadvantage is that a small number of domains, associated with real eigenvalues but showing qualitative evidence of phase space oscillations, are assigned to the second group. On the plus side, this choice allows to differentiate between domains with strong *versus* weak or no interaction in a mathematically simple, objective, and unambiguous way. The benefits outweigh the cons and, in the following, we focus primarily on the analysis of those regions that are associated with complex eigenvalues of the dynamical matrix.

Figure 5.11 shows the spatial location of the averaging domains associated with complex eigenvalues (and hence with phase space oscillations). The domains tend to cluster north of the -0.4 SSH level and around the SAF. This is consistent with the observation that, in SSH coordinates, the ACC flow is concentrated between the -0.4 m and the 0.2 m levels approximately (see figures 5.6 and 5.7), as we expect that the interaction between eddies and the jets is strongest where the flow is most intense. The circumpolar path is not continuously tiled by the averaging domains, and a number of gaps are apparent throughout the ACC region, particularly in the Eastern Pacific sector. However, the area around the SAF is more densely populated by the averaging regions if one considers, alongside those with complex eigenvalues, the few domains which show the signature of phase space oscillations despite being associated with real eigenvalues (figure 5.12). Thus, regardless of the fact that one uses a quantitative (the eigenvalues) or qualitative (how the phase space diagram and cross-correlation function look like) criterion to label the domains, those associated with oscillations tend to arrange coherently along the SAF of the ACC. This underscores that the local emergence of predator-prey interaction between eddies and mean flow is dynamically linked to, and plausibly driven by, the strength of the ACC flow.

5.4.4 Analysis of averaged data: domain-scale dynamics

Figures 5.9 and 5.10 offer an anecdotical comparison between time series with real and complex eigenvalues. Can we see the same differences if we focus instead on the averaged behaviour of the two groups? Consider the case of the domains with complex eigenvalues first: we assume that each individual time series is an inde-

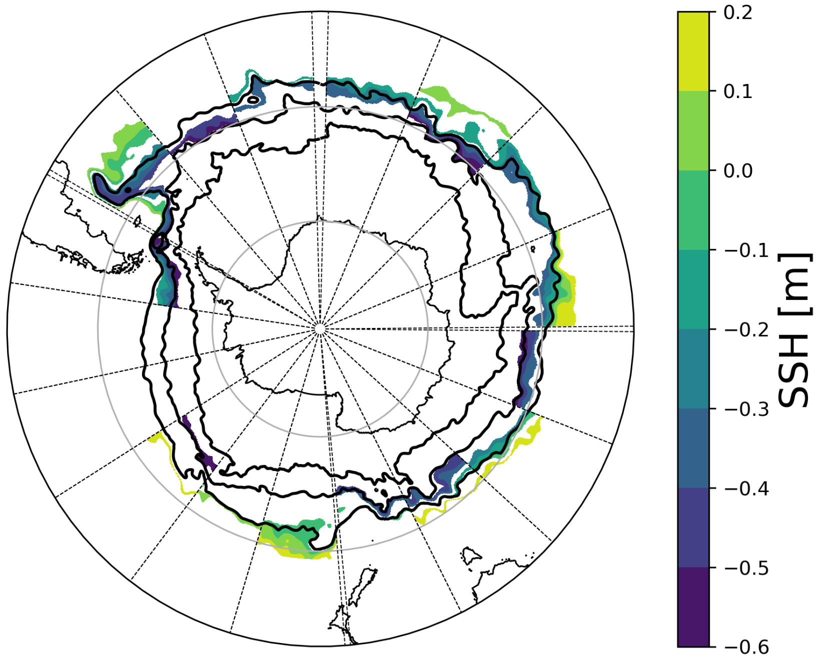

Figure 5.11: As in figure 5.8, but only the domains associated with complex eigenvalues are shown.

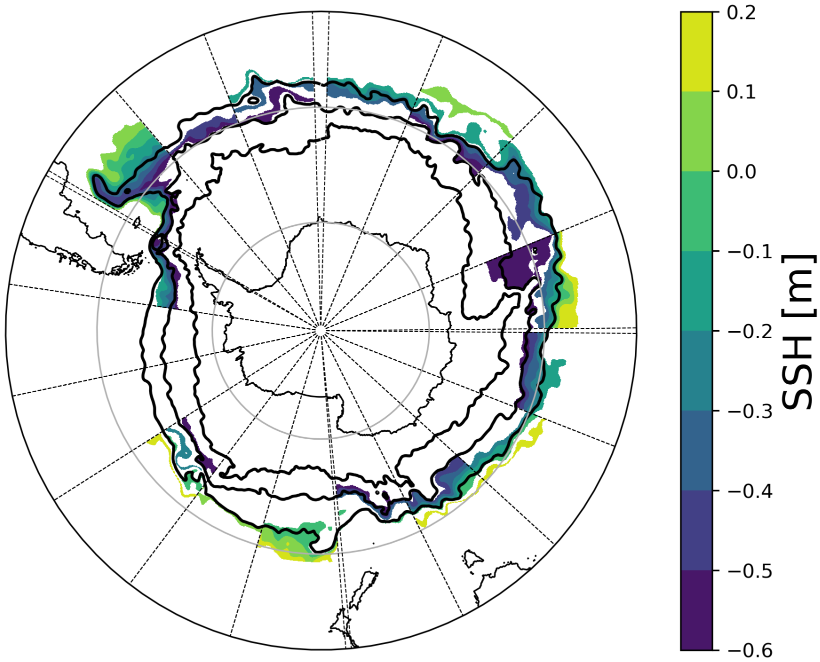

Figure 5.12: As in figure 5.8, but only the domains with complex eigenvalues or with real eigenvalues and qualitative evidence of phase space oscillations are shown.

pendent realisation of the same stochastic process, which represents the averaged dynamics over the ACC belt. One possibility to study this process is to imagine that the short, small-scale time series have been sampled from a fictitious longer, large-scale one, which we reconstruct by concatenating the small-scale time series, namely, by stacking them one at the end of the other. Since from figure 5.9 (c) we know that the typical decorrelation time of the dynamical variables is much shorter than six years (the length of the individual time series), we argue that the boundary points between any pair of contiguous short time-series weigh only marginally on the overall estimation of the statistical properties of the large-scale process. For the same reason, we expect that the specific order of the short time series within the longer one does affect our results substantially. The final step of the analysis is to fit the simplified model to the concatenated time series as detailed in section 5.3. The procedure is then repeated for the case of the domains with real eigenvalues.

Figures 5.13 and 5.14 show the correlation functions (left) and the phase space diagram (right) for the domains with complex (5.13) and real (5.14) eigenvalues. Figure 5.13 demonstrates that phase space oscillations emerge on average, and not only locally, for the domains with complex eigenvalues. The sample cross correlation function between mean flow and eddies (figure 5.13 (left); bottom panel; dotted green line) display the by now familiar structure associated with the predator-prey dynamics: an odd function with a negative peak at positive lags and a positive peak at negative lags. The peaks are at a lag of approximately 20 days, which is comparable with what observed for the idealised channel. The value of ρ_{XY} at the peaks is modest (less than 0.2), but it stands clearly out of the acceptance interval of the null hypothesis (which is computed using the total number of points in the concatenated time series here). Thus, the correlation between the dynamical variables is weak, but statistically significant. As in the case of the idealised channel, the best fit prediction for the cross-correlation captures the functional form of ρ_{XY} well, but with larger peaks at larger absolute lags. Note also the good agreement between the sample and predicted auto-correlation functions for X (top panel) and Y (middle panel). Figure 5.13 (right) shows the phase space diagram for the domains with

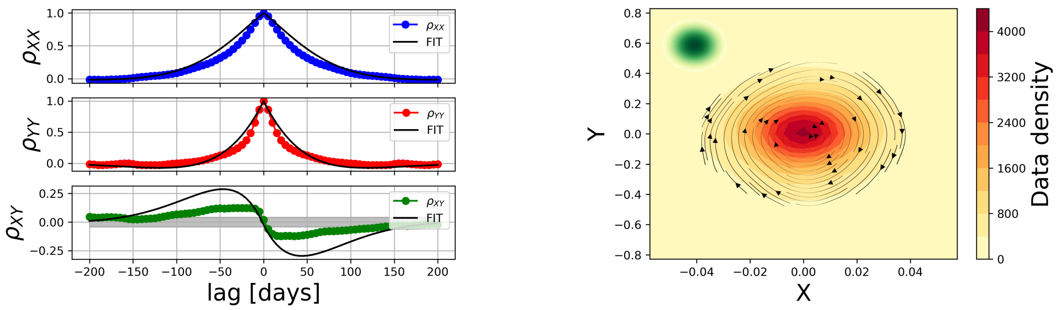

Figure 5.13: Correlation functions and phase space diagram for the time series obtained by concatenating data from all domains associated with complex eigenvalues. Plots are as in figure 5.9.

complex eigenvalues. The structure of the phase space trajectories is characterised by a high degree of regularity and, similarly to that of the idealised channel, is consistent with the physical interpretation of the eddy life cycle as a predator-prey interaction between eddies and mean flow. It also compares well with the phase space diagram obtained from a 30-years long synthetic integration of the simplified model (see section 5.4.8). Thus considering the "global" average rather than a local specimen entails that the dynamical indicators are smoother (which is expected whenever some form of averaging is applied), but does not change the physical picture significantly for the domains with complex eigenvalues.

The case of the domains with real eigenvalues is different. The individual analysis of figure 5.10 does not allow ourselves to judge whether interactions between eddies and mean flow are absent, or too weak to assert themselves due to the short span of the time series. Consideration of the data obtained by concatenating all the time series associated with real eigenvalues, however, tips the balance in favour of the second possibility. The correlation functions are shown in figure 5.14 (left). The sample cross-correlation (bottom panel; dotted green line) is very weak (always less than 0.1) and barely exceeds the width of the acceptance region of the null hypothesis that the variables are uncorrelated. Nevertheless, the familiar structure associated with predator-prey dynamics (approximately odd function; negative peak at positive τ and positive peak at negative τ) is recognisable, and correctly captured by the best fit prediction (continuous black line). The best fit curves are

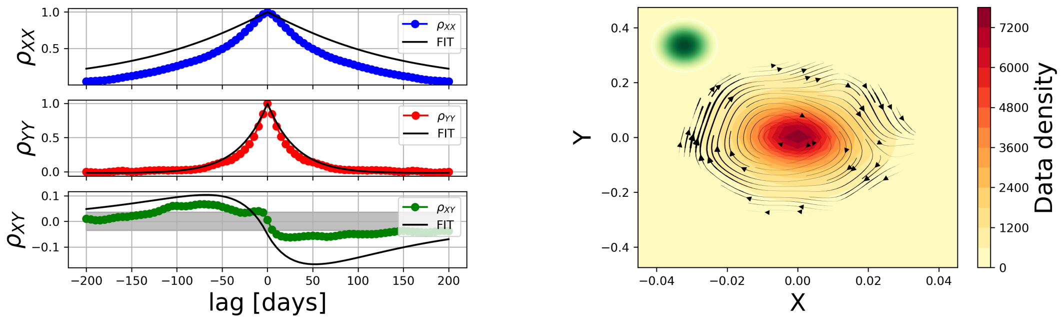

Figure 5.14: Correlation functions and phase space diagram for the time series obtained by concatenating data from all domains associated with real eigenvalues. Plots are as in figure 5.9.

in reasonable agreement with the sample auto-correlations of X and Y too (top and middle panels respectively). The kernel-averaged phase space diagram is shown in figure 5.14 (right). The single most striking feature of this plot is its regularity. Indeed, were it not for the larger pool of data (which shows itself in the values of the data density, coloured intervals) and by the small ripples at the periphery that distort the elliptical orbits, it would hardly be distinguishable from that of figure 5.13. Note that this diagram compares well with that computed from a 30-years long synthetic numerical integration of the simplified model with weak coupling parameters, but not with that from an equally long synthetic integration of two uncorrelated red noise time series (see section 5.4.8). As anticipated the upshot is that, on average, a form of predator-prey dynamics is at play even for the domains associated with real eigenvalues, but that the interaction is too weak to manifest itself locally due to the paucity of data. A second important conclusion we draw is that the phase space diagram is a powerful diagnostic tool which, for a sufficiently large sample, is effective at exposing the hidden dynamical relationship between the variables. It does not contain information, though, about the strength of their interaction, and needs therefore to be complemented by some other diagnostics: here, the correlation functions. Another option may be to use the significance metrics for the phase space diagram introduced by Yano et al. (2020).

As a last step before taking a closer look at the quantitative outcome of the fits, we investigate the Probability Distribution Functions (PDFs) of X and Y (in the interest of brevity, for the domains associated with complex eigenvalues only). The analysis of the PDFs exposes a disadvantage of studying the averaged dynamics in the ACC region by concatenating the time series from multiple local domains, namely, the PDFs relative to the concatenated time series depart from a normal distribution if the individual time series are not identically distributed. In our case, the individual time series are all approximately normally distributed with zero mean, but their standard deviations vary over a comparatively broad range of values. As a consequence, the PDFs associated with the concatenated time series have much heavier tails than what expected from a normal distribution (not shown). A possible remedy is to take the arithmetical average of, rather than concatenating, the time series from different locations in order to study the PDFs. The mathematical basis is that the sum of two independent normal random variables is still a normal random variable even if the two addends are not identically distributed (Ross, 2014). In formula, if $X_1 \sim N(0, \sigma_1^2)$ and $X_2 \sim N(0, \sigma_2^2)$, then $Z = X_1 + X_2 \sim N(0, \sigma_1^2 + \sigma_2^2)$. It follows that, for n independent normal distributions:

$$\begin{aligned} \frac{1}{n}(X_1 + X_2 + ... + X_n) &\sim \frac{1}{n} N(0, \sigma_1^2 + \sigma_2^2 + ... + \sigma_n^2) \\ &\sim N(0, \frac{1}{n^2}(\sigma_1^2 + \sigma_2^2 + ... + \sigma_n^2)). \end{aligned} \tag{5.20}$$

Thus, we can fit the distribution of the arithmetical average of the individual time series with a zero-mean normal distribution, with variance given by the arithmetical average of the individual variances divided by the number of time series. The variance of an individual time series is obtained from the best fit of the lag-0 covariance matrix, as explained in section 4.2.7. The result is shown in figure 5.15: we observe that, while the predicted distributions of X and Y (continuous black lines) have the correct scaling, they are significantly narrower than the sampled ones (histograms) for both variables. Interestingly, the same effect cannot be consistently detected when comparing the predicted versus sampled distributions of individual time series (not shown). While the reason for this behaviour remains unclear, one possible explanation is that small deviation from Gaussianity of individual time series compound to produce larger errors when the average is computed. The PDFs are the only non-additive mathematical objects that we consider in our analysis, and in the

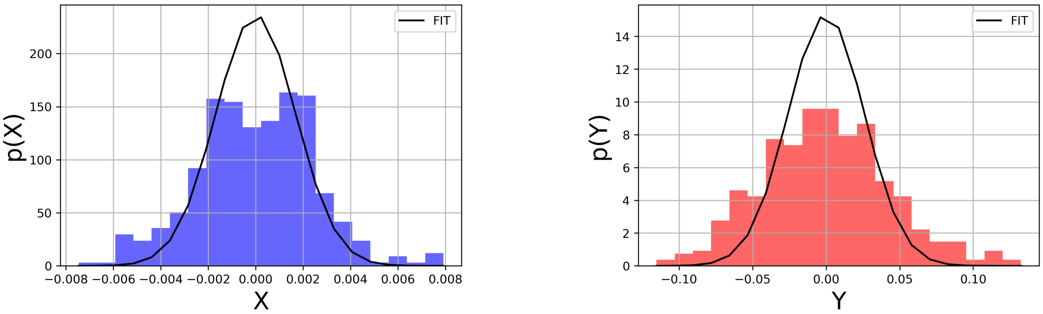

Figure 5.15: Probability Distribution Function relative to the arithmetical average of the time series associated with complex eigenvalues (histograms) and theoretical prediction from the best fit of the data (continuous black lines) for X (left, blue) and Y (right, red).

rest of the section we can safely focus on the properties of the concatenated time series.

5.4.5 Fit of the model

The model is fitted to the data obtained by concatenating all the time series associated with (i) complex and (ii) real eigenvalues as described in section 5.3.1. The best fit parameters from the linear fit for the two cases are shown in table 5.1. Note that the parameters are comparable with those estimated for the case of the idealised channel, table 5.1 (see section 5.4.7 for further comments). Also, the product $|k_{xy}k_{yx}|$ is almost an order of magnitude larger for the data with complex eigenvalues than for those with real eigenvalues, while the sum $\gamma_{xx} + \gamma_{yy}$ is nearly unchanged: thus, the outcome of the fit reflects the mathematical proposition that complex eigenvalues (and hence phase space oscillations) are associated with larger values of the coupling coefficients with respect to the damping terms. In the rest of this section, we concentrate on the fit of the simplified model to the data with the strongest evidence of predator-prey dynamics, i.e. those obtained by concatenating the time series associated with complex eigenvalues.

Table 5.1: Summary of fit results for the data obtained by concatenating the time series of all domains associated with complex (top line) and real (bottom line) eigenvalues

Data	γ_{xx} $\left[\mathrm{day}^{-1}\right]$	γ_{yy} $\left[\mathrm{day}^{-1}\right]$	k_{xy} $\left[\frac{\mathrm{Kg}}{J\,\mathrm{day}}\right]$	k_{yx} $\left[\frac{J}{\mathrm{Kg\,day}}\right]$	R^2	σ_x $\left[\mathrm{day}^{-3/2}\right]$	σ_y $\left[\frac{J}{\mathrm{Kg\,day}^{3/2}}\right]$
Complex eigs	$-1.0 \cdot 10^{-2}$	$-2.7 \cdot 10^{-2}$	$1.0 \cdot 10^{-3}$	$-2.3 \cdot 10^{-1}$	$(0.91, 0.75)$	$2.7 \cdot 10^{-3}$	$6.4 \cdot 10^{-2}$
Real eigs	$-5.7 \cdot 10^{-3}$	$-3.1 \cdot 10^{-2}$	$5.5 \cdot 10^{-4}$	$-8.1 \cdot 10^{-2}$	$(0.94, 0.72)$	$1.6 \cdot 10^{-3}$	$3.9 \cdot 10^{-2}$

5.4.6 Evaluation of the fit

Is the fit of good quality? The coefficients of determination for X and Y are $R_X^2 = 0.91$ and $R_Y^2 = 0.75$ respectively, remarkably close to the values obtained for the idealised channel. Thus, the simplified model explains a large fraction of the variability of the concatenated data. Figure 5.16 and 5.17, in addition, show the conventional visual metrics for the evaluation of the fit. Figure 5.16 (left) shows the prediction plots for X (top panel, blue) and Y (bottom panel, red). The data are clustered around the $1:1$ line, and no clear patterns of deviation are apparent. The correlation functions of the residuals are shown in figure 5.16 (right). Remember that for a perfect model the residuals are independent, identically distributed, normal random variables (i.e. white noise), and only the lag-0 auto-correlation coefficients are expected to be significantly different from zero. Figure 5.16 (right) demonstrates that departures from the ideal case are minimal, as the lag-1 coefficients only stand clearly out of the acceptance interval of the null hypothesis for all three correlations. Importantly this demonstrates that, although more complex, higher-order theoretical models may provide a quantitative better fit to the data, it seems plausible to assume that they would not significantly change the gist of the qualitative conclusions drawn with the simplified model. Lastly, figure 5.17 shows the scatter plots of the standardised residuals for X (left) and Y (right). Most of the scatter points are located within ± 2 standard deviations from zero, and their arrangement is not indicative of linear or polynomial dependence on the dynamical variables. Overall, no severe weaknesses of the model emerge from the analysis of the residuals.

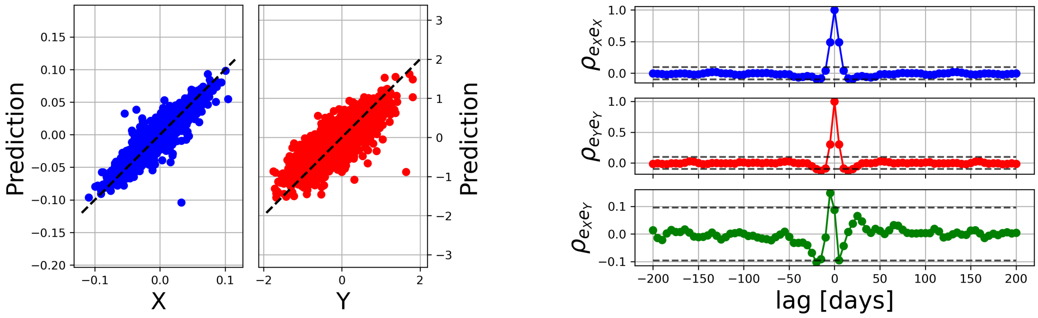

Figure 5.16: Left: prediction plots for X (blue, left panel) and Y (red, right panel). See figure 4.11 for more details. Right: correlation functions for the residuals of the fit. See figure 4.10 for details.

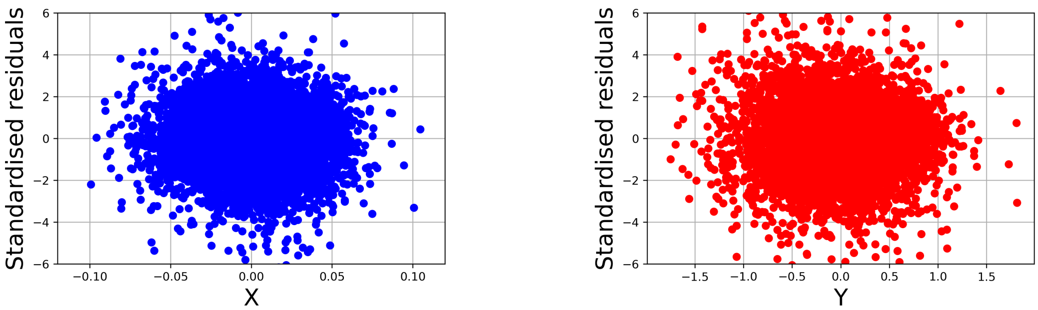

Figure 5.17: Left: scatter plot of $\overline{e}_X$ against X. Right: scatter plot of $\overline{e}_Y$ against Y.

5.4.7 Time scales and physical interpretation of the coefficients

Similarly to chapter 4, we physically interpret the dynamical coefficients of the simplified model by associating them with typical time scales. Here, we focus on the case of the domains with complex eigenvalues.

We start with the damping coefficients γ's: as in section 4.5.3, we consider the deterministic part of the dynamics only and momentarily suppose that the coupling coefficients k's are set to zero, so that the governing equations are given by (4.136)) and (4.137). These equations describe exponential decay, and a relevant time scale is the halving time, $\log 2/\gamma$, yielding $\tau_x \approx 69$ days for the mean flow and $\tau_y \approx 26$ days for the eddies. The time scales are slightly longer than in the case of the idealised channel (where we obtained $\tau_x \approx 47$ days and $\tau_y \approx 16$ days) but, notably, τ_x is approximately a factor 2.7 larger than τ_y, which suggests that in SOSE too the mean flow is characterised by greater temporal persistence than the eddies.

In order to interpret the coupling coefficients k, we consider two different sets of initial conditions for the full deterministic problem, equations (4.138) and (4.139), and approximate the governing equations accordingly (see section 4.5.3 for the details). The first case corresponds to the initial conditions $X(t=0) = 2s_x$ and $Y(t=0) = 0$, where s_x is the standard deviation of X. The situation is that of eddy growth following an intense steepening of the isopycnals at $t = 0$, and the approximated equations are (4.140) and (4.141). In this approximation, the Eady growth rate X decays exponentially, while the eddy buoyancy flux Y evolves by extracting energy from the mean flow via the coupling coefficient k_{yx}. We stress that this a crude approximation, only valid for the initial stages of the evolution of the two variables. The numerical solutions of the exact (continuous lines) and approximated (dashed lines) equations are shown in figure 5.18 (left). The halving time of X visually corresponds to the intersection between the dashed blue line and the horizontal dashed-dotted blue line, marking the $1s_x$ level (note that since the initial condition for X is $X(t=0) = 2s_x$, the halving time coincides with the time needed for a one standard deviation decrease). As in the case of the idealised channel, the exact solution (continuous blue line) decays faster than the approximated one. The dashed red line shows the linear approximation of the exact solution for Y (continuous red line): note that only the initial phase of the evolution, corresponding to eddy growth, is captured by the approximated line. Its intersection with the $1s_y$ level (horizontal dashed-dotted red line) represents the time needed for eddies to grow by one standard deviation in absolute value, and thus the time scale associated with the coupling coefficient k_{yx}. We obtain a value of about 31 days, which is only a factor 2 smaller than the relaxation time scale for the mean flow (compared to the factor 5 found for the idealised channel): this suggests that eddies are less efficient at extracting energy from the mean flow in SOSE than in the idealised MITgcm configuration. The conclusions is supported by the fact that the peaks of the cross-correlation function are larger for the idealised channel than in the case of SOSE.

A time scale for the coupling coefficient k_{yx} is obtained by considering the symmetrical initial condition $X(t=0)=0$ and $Y(t=0)=2s_y$, where s_y is the standard deviation of Y. In this case, the relevant approximated equations are (4.142)) and (4.143), and the situation is depicted in figure 5.18 (right). Physically, it corresponds to the flattening of the isopycnals following a peak in eddy activity. The approximated solution for Y decays exponentially, and is shown by the dashed red line. Its intersection with the dashed-dotted red line corresponds to the halving time τ_y (which also coincides with the time needed for a one standard deviation absolute decrease). The time scale associated with k_{yx} is marked by the intersection between the approximated solution for X (dashed blue line) and the $-1s_x$ level (horizontal dashed-dotted blue line): we obtain a value of about 34 days, which is comparable with that associated with k_{yx} but not shorter (in fact, slightly larger) than the relaxation time scale for Y. As for the case of the idealised channel, this implies that a few eddy events may be needed to induce significant deviations in the mean flow. Overall, we found that the time scales associated with the dynamical coefficients from the best estimate of SOSE's complex domains are longer than, but still comparable with, those of the idealised channel. Importantly though, their relative

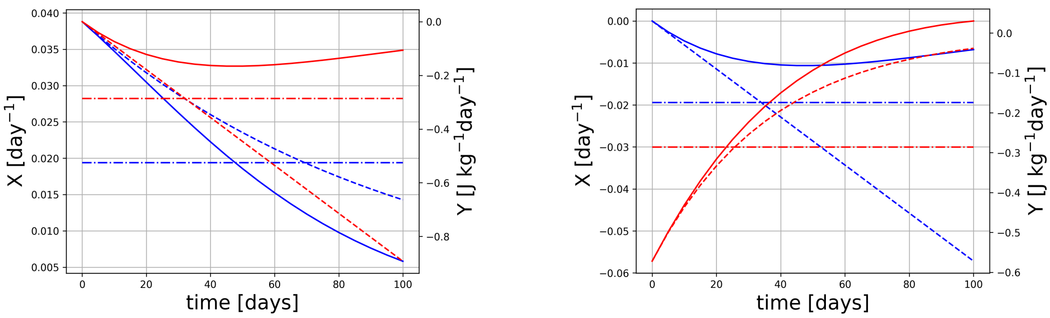

Figure 5.18: Numerical solution of the deterministic equations for $X(t=0) = 2s_x$ and $Y(t=0) = 0$ (left), and $X(t=0) = 0$ and $Y(t=0) = -2s_y$ (right). Continuous lines show the solutions to the exact deterministic equations (4.138) and (4.139), while dashed lines show the solutions to the approximate equations (4.140) and (4.141) (left), and (4.142) and (4.143) (right). Dashed dotted lines mark the $1s_x$ (blue) and $1s_y$ (red) levels. In both panels, X is shown in blue (ticks on the left hand side axis) and Y is shown in red (ticks on the right hand side axis).

magnitude is similar, which suggests that the dynamics of eddy-mean flow interaction expressed by the best fit of the simplified model are qualitatively similar in the two cases.

5.4.8 Comparison with synthetic time series

We corroborate the conclusions draw in the previous sections by generating synthetic realisations of a number of stochastic processes, and comparing the associated phase space diagrams with those constructed from SOSE data.

Figure 5.19 shows the phase space diagrams computed from a 6-years (left) and 30-years (right) long realisations of the simplified model, where the model's parameters are set to the best estimate obtained for SOSE's concatenated time series associated with complex eigenvalues. In both cases, phase space trajectories are quasi-closed orbits whose orientation agrees well with the predator-prey picture of AN14. However, the diagram constructed from the 30-years long time series (right) is more regular than that obtained from the 6-years one (left). We compare these

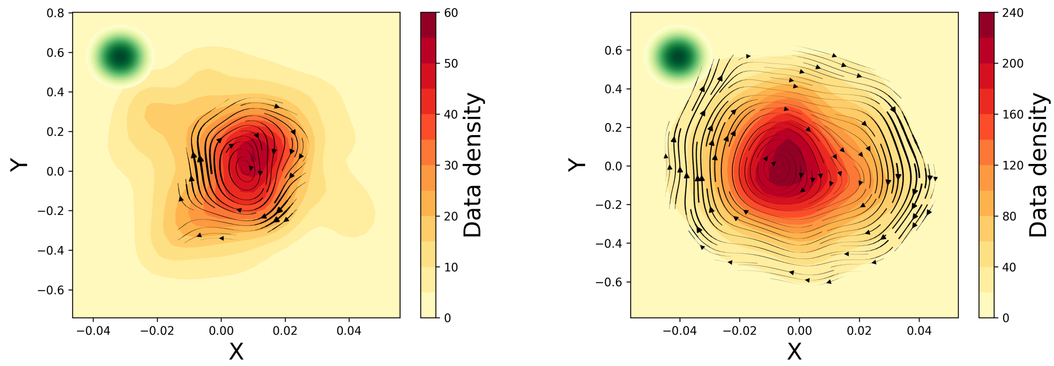

Figure 5.19: Phase space diagrams from a 6-years (left) and 30-years (right) long synthetic time series. The time series are obtained by numerically integrating the simplified model, equation (4.2.7), with the Euler-Maruyama method. The parameters of the simplified model are set to those obtained from the best fit of the concatenated time series associated with complex eigenvalues.

synthetic diagrams with those obtained from SOSE's data: the 30-years synthetic phase space diagram of figure 5.19 (right) exhibits a similar structure to the one constructed from SOSE's concatenated complex time series, figure 5.13 (right). This places additional confidence in the simplified model's ability to capture the average dynamics of the domains with complex eigenvalues. Furthermore, the 6-years synthetic diagram of figure 5.19 (left) is qualitatively compatible with that of figure 5.9 (right) which is obtained from the 6-years time series associated to an individual domain with complex eigenvalues. This demonstrates that distorted phase space orbits relative to an individual domain can be explained by the short span of the time series (however, the comparison does not prove that the length of the time series is necessarily the only factor at play).

Figure 5.20 shows the phase space diagrams constructed from two different 6-years (left) and 30-years (right) long realisations of the simplified model. The model parameters are set to the best estimate for SOSE's concatenated complex time series, but with $k_{xy} = k_{yx} = 0$, for panels (a) and (b), and to the best estimate for SOSE's concatenated real time series for panels (c) and (d).

Panels (a) and (b) thus correspond to a 6-years long (panel (a)) and a 30-years long (panel (b)) bivariate red noise process, which can be thought of as a degenerate instance of the simplified model, with no dynamical coupling between the variables. Regardless of the length of the time series, no evidence of predator-prey dynamics is discernible in the phase space diagrams: hence, we reject the null hypothesis that quasi-periodic orbits can be generated even in the absence of dynamical coupling between the two variables. Both the 6-years long synthetic phase space diagrams, figures 5.20 (a) and (c), are qualitatively comparable with the diagram obtained from SOSE data for an individual domain with real eigenvalues, figure 5.10 (b), confirming that a six-years long time series is not sufficient to differentiate between the cases of uncoupled (panel (a)) and weakly-coupled (panel (c)) dynamics. The difference, however, emerges when longer time series are considered, as only the synthetic 30-years diagram corresponding to weakly coupled dynamics (figure 5.20 (d)) is comparable with that constructed from SOSE's concatenated real time series, figure 5.14 (right). As noted in the previous section, though, the phase space diagram does not provide information about the strength of the dynamical coupling, therefore its analysis must be complemented by that of alternative indicators, such as the correlation functions considered in this study.

5.5 Summary and conclusions

In this chapter, I have analysed data from a realistic state estimate of the Southern Ocean, the SOSE, and shown that the interaction between eddies and the mean flow in localised regions of the ACC is characterised by predator-prey dynamics, and can be modelled mathematically by a two-dimensional, stochastic oscillator. This work was motivated by the results presented in chapter 4, where I have shown that predator-prey dynamics similar to those studied by AN14 for the atmospheric storm track are at play in an idealised channel configuration of the MITgcm. The idealised channel configuration, however, involves a number of restrictive assumptions, including a zonally symmetric domain, flat bottom topography, no salinity, no sea ice, and constant forcing at the surface, which begs the questions of whether the

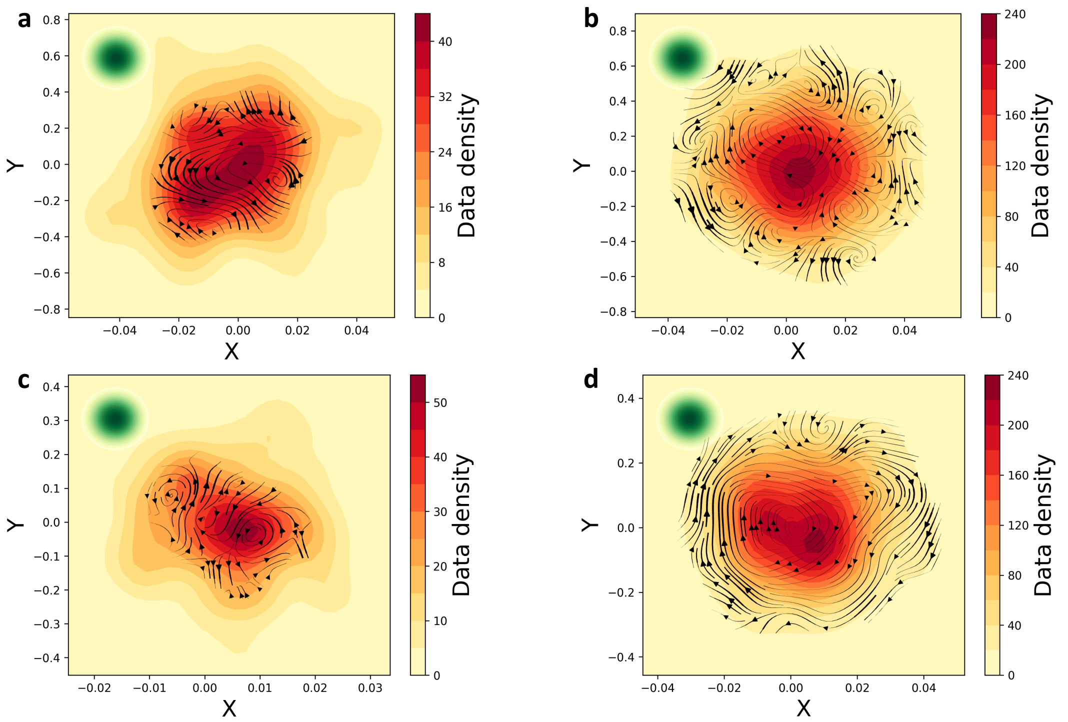

Figure 5.20: Phase space diagrams from a 6-years (left) and 30-years (right) long synthetic time series. The time series are obtained by numerically integrating the simplified model, equation (4.2.7), with the Euler-Maruyama method. The parameters of the simplified model are set to those obtained from the best fit of Top: the concatenated time series associated with complex eigenvalues, but with the coupling parameters k set to zero and Bottom: the concatenated time series associated with real eigenvalues.

AN14 dynamics are relevant to the real Southern Ocean. SOSE, in contrast, is an exact solution of a realistic configuration of the MITgcm, coupled to observations from a variety of sources through a data assimilation technique. Here, I have used SOSE iteration 100, which covers 6 years of time from 2005 to 2010. Previous research demonstrates that this product, while being dynamically self-consistent, provides a reliable representation of the state of the Southern Ocean.

Similarly to what done in chapter 4, I have identified a pair of variables to represent the strength of the mean flow and of the eddies, respectively, the Eady growth rate and the eddy buoyancy flux. Stream-wise averages are more accurate than zonal averages when the geometry of the domain is not zonally symmetric (Olbers et al., 2004), therefore the definitions of chapter 4 were adjusted so that the averages are taken along the time-mean flow of the ACC rather than along latitude circles. Importantly, the dynamical variables are defined locally, over a number of comparatively small averaging domains, rather than hemispherically over the entire ACC region, which allows to capture regional differences in the dynamics of eddy-mean flow interaction. A total of 124 averaging domains tiling the ACC was considered for the analysis. The domains are located in the upper interior and are approximately 20° wide: each domain is associated with a 6-years bivariate time series, representing the time evolution of the spatially averaged Eady growth rate and eddy buoyancy flux. The two-dimensional stochastic oscillator of chapter 4, named the simplified model, was initially fitted to the individual time-series from each of the 124 domains (the fit method is a conventional linear regression). In addition, I fitted the simplified model to the time series obtained by combining the data associated with selected regions of the ACC, as explained below. I have analysed the outcome of the fits, looking in particular for signatures of predator-prey dynamics between the two variables. My main results are:

1. The fit of the simplified model is successful for all the domains considered in the analysis. However, only some of them show qualitative evidence of predator-prey dynamics.

2. When present, the evidence for the predator-prey dynamics is supplied by the phase space diagrams and by the cross-correlation functions. In this case, phase space trajectories are nearly closed orbits. The cross-correlation is approximately an odd function of the time lag. The orientation of the phase space trajectories and the sign of the cross-correlation function are consistent with AN14 dynamics.

3. The shape of the phase space orbits is more irregular than for the idealised channel. However, comparison with synthetic realisations of the simplified model reveals that this effect can be attributed to the comparatively short period of time covered by SOSE.

4. The interaction between eddies and the mean flow could be weak, absent, or depart from the AN14 model in individual domains that do not show quasi-periodic orbits. 6-years synthetic experiments with the simplified model reveal that the phase space diagrams associated with both a weakly-coupled oscillator and with a bivariate uncorrelated red noise process are compatible with those constructed from SOSE data. This suggests that the individual time-series are too short to differentiate between the possible cases.

5. The domains with phase space oscillations tend to cluster around the SAF, where the zonal flow of the ACC is most intense. This underscores that the predator-prey dynamics are not widespread in the ACC region, but physically linked to the presence of strong jets.

6. Mathematically, the domains can be divided into two groups depending on whether the eigenvalues of the dynamical matrix have nonzero imaginary part (pointing to phase space oscillations) or not (no phase space oscillations).

7. The time series of the domains belonging to each group can be combined in order to study the averaged dynamics of the two regions.

8. For the domains with complex eigenvalues, the phase space diagram and the correlation functions are indicative of predator-prey dynamics. The phase space diagram is smoother than for individual domains, and comparable with

that of a 30-years long synthetic realisation of the simplified model. The peaks of the cross-correlation function are weak, but statistically significant.

9. Evidence of AN14 dynamics emerges even when the time series associated with domains with real eigenvalues are combined. In this case, though, the peaks of the cross-correlation function stand barely clear of the null hypothesis acceptance region, suggesting that, albeit present, the dynamical coupling is weak for these domains.

10. The simplified model does not capture the details of the correlation functions and marginal probability distribution functions accurately for the combined domains with complex eigenvalues. Importantly, though, the overall qualitative patterns are well reproduced. The analysis of the residuals does not reveal any obvious shortcoming of the simplified model.

11. The fitted dynamical parameters of the simplified model can be associated with typical time scales by considering suitable approximations of the deterministic equations. The time scales for the combined complex domains are slightly longer than, but overall comparable with, those found in the case of the idealised MITgcm configuration, and range from a few weeks to a couple of months.

The results presented here corroborate the view that some of the important aspects of eddy-mean flow interaction in a realistic representation of the Southern Ocean are captured by a simple, two dimensional stochastic oscillator. Notably, they demonstrate that the dynamics in the oceanic case are similar to those of the AN14 model for the atmospheric storm track: eddies feed on the available potential energy stored in the mean flow, and the dynamical coupling between the two variables generates predator-prey cycles with typical time scales ranging from a few weeks to a couple of months. These time scales are comparable to those found in the case of the idealised channel: the two models's configurations are at opposite ends of the complexity spectrum, suggesting that the simplified model contains the fundamental ingredients necessary to describe the eddy-mean flow interaction. The results also highlight that the predator-prey dynamics are not widespread in the Southern Ocean domain, but tend to cluster around the strongest jets, emphasising the importance of regional dynamics in the ACC (Frenger et al., 2015, Rintoul, 2018). Overall, this work shows that the dynamical analogy between the ACC and the tropospheric jets (Thompson, 2008, Williams et al., 2007) can be exploited to enhance our understanding of the dynamics of mesoscale eddies, which, as explained in chapter 1, is a key ingredient to confidently predict the future evolution of the Southern Ocean under climate change. Moreover, here and in chapter 4 I observed that phase space diagrams with quasi-periodic orbits can be expressed by dissipative deterministic dynamics, whereas the preferential interpretation is that they are associated with conservative dynamics. Thus, these results may contribute to the discussion within the atmospheric community too, where the analysis of phase space diagram is emerging as a valuable diagnostic tool (Yano et al., 2020).

A number of caveats apply: firstly, the simplified model captures the salient qualitative traits of the dynamics well, but not so the quantitative details (for example, the fit does not reproduce the structure of the correlation functions or of the marginal probability distribution functions accurately). Partly, this may be due to the short period of time covered by SOSE, as it is clear from section 5.4 that the statistical indicators constructed from the 6-years time series are characterised by high levels of noise. However, a similar quantitative mismatch between predicted and diagnosed statistics was observed for the 30-years long time series of the idealised channel, suggesting that more complex models such as, e.g. second order auto-regressive processes, may be needed to achieve higher fit performances. The drawback is that increasing the complexity of an empirical models further hinders the physical interpretation of its parameters. Secondly, it is unclear to what extent the transient response of the eddy field to wind stress changes projects onto these modes of interannual variability. The results presented here suggest that eddies strengthen a few weeks after an intensification of the mean flow, which is not incompatible with recent modelling results (Wilson et al., 2015), where no significant non-zero lag for the EKE response to wind stress changes at yearly scale.

However, knowledge of how the parameters scale with the wind forcing is a precondition to investigate how the time scales of natural variability relate to those of the forced response by means of reduced-order mathematical models. As highlighted in chapter 4, a promising way forward is to derive a mathematical model of higher complexity by successive approximations of the equations of motion (Kobras et al., 2022), which naturally endows the model with scaling laws for its parameters. Here, by fitting the simplified model to data from the MITgcm's idealised channel configuration (chapter 4) and from SOSE, the first steps were taken to show that low-dimensional mathematical models based on a dynamical system approach can help us deepen our understanding of the complex nature of eddy dynamics in the Southern Ocean.

Chapter 6

Time-scales of forced variability in the Southern Ocean

6.1 Introduction

The surface of the Southern Ocean has cooled (or warmed weakly) over the last few decades (Fan et al., 2014). Concomitantly, the seasonal sea ice has expanded (Parkinson, 2019). A number of hypotheses have been put forward to explain the observed trends, including enhanced freshwater fluxes (Haumann et al., 2020), Antarctic glacial melt (Rye et al., 2020), and natural variability (Polvani et al., 2021).

A further possibility is that surface wind stress modulations may be responsible for the observed changes. It is widely accepted that ozone depletion over Antarctica has induced a strengthening of the Southern Hemisphere jet stream over recent decades (Polvani et al., 2011). Furthermore, the trend may persist during the 21st century due to increased greenhouse gases concentrations (Thompson et al., 2011). Wind stress is one of the primary drivers of the ACC, therefore its intensification could have a large impact on the Southern ocean circulation, temperature, salinity, and heat and carbon uptake rates, with consequences for global climate.

Considerable efforts have been devoted to investigate the equilibrium response of the Southern Ocean to wind stress perturbations in the last 20 years, particularly with the advent of eddy-resolving general circulation models. There is now general consensus about the fact that the ACC circumpolar transport is only weakly sensitive to wind changes (Munday et al., 2013): the additional energy supplied by the stronger winds powers eddy motion rather than the zonal flow, a phenomenon known as eddy saturation. Many studies also suggest that baroclinic eddies partially compensate for wind-induced changes in the meridional overturning circulation (Viebahn and Eden, 2010), although the exact amount of so-called eddy compensation is not uniform across models.

The transient response of the Southern Ocean to wind stress perturbations is likely endowed with time scales ranging from years to decades (Kostov et al., 2017), and is thus relevant to the future evolution of the Southern Ocean and global climate: however, it is less well understood than the equilibrium response. Notably, a study by Ferreira et al. (2015) proposed that the transient adjustment comprises two time scales, separately driven by different physical processes. The fast time scale is dictated by anomalous northward Ekman transport of cold water: in this phase, the surface of the Southern Ocean cools. The slow time scale, instead, is controlled by anomalous upwelling of warm water from below the seasonal sea ice, which leads to a surface warming. This explanation reconciles an apparent paradox: observations reveal that SST decreases following an anomalous intensification of the winds on interannual timescales, largely due to enhanced northward Ekman transport. On the other hand, models tend to predict long term warming of the surface.

The study of Ferreira et al. (2015) hinges on the climate response function (CRF) formalism of Marshall et al. (2014), where the response of a system to arbitrary time modulations of a forcing field is computed from its response to a step change perturbation. Subsequent studies investigating the transient adjustment of the Southern Ocean with the same technique, though, exposed a number of problems relative to the two-time scales mechanism. First, the time scales themselves are only loosely constrained, and vary widely across models (Kostov et al., 2017). Second, historical simulations are unable to reconstruct the surface temperatures observed in past decades accurately (Seviour et al., 2019).

Critically, though, the GCMs employed in these studies are eddy-parametrising (most, but not all, belonging to the CMIP5 suite). It is well known that the equilibrium response of eddy-parametrising models to wind stress changes is qualitatively different from that of eddy-resolving models (e.g. Hallberg and Gnanadesikan (2006)), which prompts the questions of whether this is the case for the transient response too. Only a handful of papers has addressed the issue so far (Doddridge et al., 2019, Haumann et al., 2020): Doddridge et al. (2019), in particular, found that eddy compensation prevents sustained upwelling of warm water into the mixed layer in their model: as a result, the surface does not warm on the long term, which is the kind of response expected for the real ocean based on observations (Seviour et al., 2019).

In this chapter, I seek to further the current understanding of which processes drive transient surface and interior temperature changes in an eddy-resolving general circulation model following an abrupt wind stress perturbation. A second, complementary goal is to characterise the important time scales of the adjustment. In particular, I will investigate how the spin-up of baroclinic eddies affects the meridional overturning circulation of the Southern Ocean, and thereby influences the surface and subsurface temperature response. To this aim, I will diagnose and analyse the components of the temperature budget: also, I will show that under appropriate conditions the budget equations can be formulated so as to explicitly account for the residual advection terms, which helps separate the role of eddy and mean flow contributions.

The model I employ is an idealised channel configuration of the MITgcm at a high horizontal resolution, whose reference state was discussed in chapter 3. The transient response to the perturbation is studied by realising an ensemble of step change simulations independently branched from the reference state. In order to gain insight on the final state of equilibrium, I also analyse a climatology of the equilibrated channel subject to the wind perturbation.

The chapter is organised as follows: the numerical experiments with the idealised channel configuration of the MITgcm are described in section 6.2. In section 6.3, I present and discuss the temperature budget equations. The equilibrated state of the perturbed channel is studied in section 6.4.1. In section 6.4.2, I test the temperature budget formalism for the reference state. The transient perturbation experiments are investigated in section 6.4.3. I offer a summary and conclusions in section 6.5.

6.2 Data

The response of the Southern Ocean to wind stress changes is investigated by performing a suite of numerical experiments with the idealised channel configuration of the MITgcm. Specifically, we consider three different states of the channel: (i) a statistically equilibrated, unperturbed reference state, (ii) a statistically equilibrated perturbed state, describing the final equilibrium attained after the wind stress perturbation is applied, and (iii) an out-of-equilibrium state, which captures the transient adjustment following the abrupt wind stress perturbation. The reference state, or ”control run”, is the configuration described in chapter 3, while the architecture of the two perturbed experiments is briefly outlined below.

Equilibrated perturbation experiment

The equilibrated perturbation experiment consists in modifying the surface wind stress parameter from $\tau_0 = 0.1$ N m^{-2} to $\tau_0 = 0.3$ N m^{-2}, and the vertical eddy viscosity coefficient from $A_v = 3.0$ m^2 sec^{-1} to $A_v = 3.3$ m^2 sec^{-1}. The change in the wind stress parameter is an idealised and mathematically convenient representation of the observed recent wind trends over the Southern Ocean, as discussed in the previous section and in chapter 1. The modulation in the vertical eddy viscosity coefficient does not alter the physics of the channel significantly, and only serves the purpose of stabilising the model's numerics (a long-integration experiment with the perturbed value of τ_0 lead to numerical divergence of the model, likely due to an intense and rare eddy event that violated the CFL condition. By modifying A_v, we changed the channel's initial conditions so as to avoid the adverse eddy event: this allows to carry out a sufficiently long integration and obtain an appropriate climatology). The channel is spun up until a statistically equilibrated state is reached, as diagnosed from mean kinetic energy. Due to the chaotic nature of the flow, the precise choice of the initial condition is unimportant once statistical equilibrium is attained (i.e. , the system loses memory of its initial state), as long as numerical stability is guaranteed. We found that 15 years of integration at equilibrium provide a sufficiently robust climatology.

Transient perturbation experiment

Similarly to the equilibrated perturbation experiment, the transient perturbation experiment consists in modifying the surface wind stress and eddy viscosity coefficients instantaneously from $\tau_0 = 0.1$ N m^{-2} to $\tau_0 = 0.3$ N m^{-2}, and from $A_v = 3.0$ m^2 sec^{-1} to $A_v = 3.3$ m^2 sec^{-1}. Here, we focus on the transient response of the system to the perturbation: in order to eliminate the dependency of individual trajectories on the initial conditions, we perform an ensemble of simulations and take ensemble averages, which approximates computing the expectation value on the time dependent measure of the system (Lucarini, 2018). The ensemble members are initialised at instants of time spaced by at least three years, so as to guarantee that their trajectories are independent. We realised a total of 15 members for the first three years following the wind stress perturbation, 6 of which include the diagnostics needed to close the temperature budget (see section 6.3 below. We found that 6 members suffice to robustly capture the temperature evolution). 6 ensemble members are run for an additional three years, thus covering a total of 6 years following the perturbation.

6.3 Methods

6.3.1 Closure of the temperature budget

One of the main goals of this chapter is to determine what drives surface and interior temperature changes following the instantaneous wind stress perturbation. The temperature budget is a conservation equation that relates the time derivative of temperature to various physical processes including advection, mixing, and buoyancy fluxes. By closing the temperature budget, we are able to quantify exactly to what extent each process contributes to the temperature tendency at a given time, and at a given spatial location. Below, we briefly outline the equations that govern the temperature budget, and detail how the budget components are diagnosed in the idealised channel configuration.

Fundamental equations

We start from the temperature equation:

$$\partial_t T + \nabla \cdot (\mathbf{v}T) = \mathscr{S}, \tag{6.1}$$

where $\mathscr{S}$ represents source terms. Note that it is possible to obtain a budget equation with energy density units simply by rescaling by $C_p\rho_0$, where C_p is the specific heat capacity of water. Given that we consider a single-component fluid (no salinity) with a linear equation of state, equation (6.1) is the buoyancy equation (2.7) scaled by a factor $g\alpha/\rho_0$. We are interested in the zonally-averaged temperature response to wind changes, therefore we take zonal average of equation (6.1):

$$\partial_t \overline{T} + \nabla \cdot \overline{\mathbf{v}T} = \overline{\mathscr{S}}, \tag{6.2}$$

This is the equation we use to study the temperature budget. With a slight rearrangement, we can write:

$$\partial_t \overline{T} = -\partial_y \overline{vT} - \partial_z \overline{wT} + \overline{\mathscr{S}}, \tag{6.3}$$

which highlights that the temperature equation is a conservation equation. Our task is to diagnose the various terms appearing in equation (6.3) and make sure that the equality is actually satisfied in the idealised channel.

Budget diagnostics

The MITgcm employs complicated tracer advection and time-stepping numerical schemes (see section 3.2.4), hence applying a finite difference scheme to the temperature and velocity model output is not the simplest way to analyse the temperature budget. Instead, we take advantage of the fact that the MITgcm allows for the budget terms, as appearing in equation (6.1), to be diagnosed individually. Below, we will give the essential information relevant to our model configuration, and refer the interested reader to Doddridge et al. (2019) for a more general discussion of the methodology. The model diagnostics needed to close the temperature budget are:

TOTTTEND, ADVy_TH, ADVr_TH, TFLUX, DFrI_TH, KPP_gTH.

A few grid parameters are also necessary:

rA, drF, dyG

It is straightforward to relate the model diagnostics to the physical terms of equation (6.1): the time derivative of temperature $\partial_t T$, at a given time and at a given point (x,y,z) on the tracer grid, is represented by the model as TOTTTEND(ix,iy,iz). We have omitted the here unimportant time index for clarity of notation. It follows that the time derivative of zonal average temperature $\partial_t \overline{T}$ is represented by $\overline{\text{TOTTTEND}}(iy,iz)$, where the bar denotes average over the zonal index ix. The advection terms are related to model diagnostics ADVy_TH, ADVr_TH by the following equations:

$$\partial_y(vT) \approx (\mathrm{ADVy\_TH}(ix, iy+1, iz) - \mathrm{ADVy\_TH}(ix, iy, iz))/V_c(ix, iy, iz), \quad (6.4)$$

and

$$\partial_z(wT) \approx (\mathrm{ADVr\_TH}(ix, iy, iz) - \mathrm{ADVr\_TH}(ix, iy, iz+1))/V_c(ix, iy, iz), \quad (6.5)$$

where V_c is the cell volume:

$$V_c(ix, iy, iz) = \mathrm{rA}(ix, iy) \cdot \mathrm{drF}(iz). \quad (6.6)$$

Here, rA is the cell's horizontal area, and drF the cell's thickness. By applying zonal average to the right hand side of equations (6.4) and (6.5) we recover the advection tendency of zonal-average temperature. The source terms represented by $\mathscr{S}$ in equation (6.1) can be divided in temperature fluxes at the surface and mixing. The surface flux at point (x, y, z) is given by $\mathrm{TFLUX}(ix, iy, iz)/C_p\rho_0\mathrm{drF}(iz)$, where TFLUX is non-zero at the surface only. The mixing term is controlled by the KPP scheme and divides into the implicit vertical diffusive flux:

$$\mathrm{DFrI\_TH}(ix, iy, iz)/V_c(ix, iy, iz),$$

and the non-local flux:

$$\mathrm{KPP\_gTH}(ix, iy, iz)/V_c(ix, iy, iz).$$

Although this is not an exhaustive list of all model diagnostics that are required to close the temperature budget in the most general MITgcm configuration, it suffices for our simplified setup.

Discrete equations

The model's discrete representation of the zonal average temperature equation (6.3) is thus:

$$\mathrm{TTEND} = \mathrm{ADV\_h} + \mathrm{ADV\_v} + \mathrm{TFLUX} + \mathrm{MIX}, \quad (6.7)$$

where, in order to keep the notation simple, we omit the bar denoting zonal average and assume that all budget terms are zonally averaged in the following. Here, TTEND is short for the zonally averaged temperature tendency TOTTTEND, ADV_h and ADV_v are the zonally averaged horizontal and vertical advection terms respectively (equations (6.4) and (6.5)), and MIX is the mixing term:

$$\mathrm{MIX} = \overline{\mathrm{DFrI\_TH}/V_c} + \overline{\mathrm{KPP\_gTH}/V_c}. \tag{6.8}$$

Equation (6.7) is one of the important building blocks for the analysis below.

6.3.2 Residual budget

Residual mean equations

The key advantage of the discrete temperature budget equation (6.7) is that it is an exact equation (up to errors introduced by the model numerics). Its main drawback is that, since it is the residual circulation that advects tracers in the meridional plane, it would be desirable to know how the meridional and vertical *residual* advection terms contribute to the temperature budget. Unfortunately, it is not possible to associate the advection terms in equation (6.7) to the residual or Eulerian contributions in a simple way. For example:

$$\mathrm{ADV\_h} \approx \partial_y \overline{vT} = \partial_y(\overline{v}\overline{T}) + \partial_y \overline{v'T'}, \tag{6.9}$$

which shows that the horizontal advection includes contributions from both the mean flow and the eddy flux. Rather than diagnosing the terms on the right hand side of equation (6.9), and approximate it with a finite difference scheme, we formulate the temperature budget equation directly in the residual framework. We return to the zonal-average temperature equation (6.3), and note that its residual mean equivalent is (see section 2.4.3):

$$\partial_t \overline{T} = -\partial_y(v_{\mathrm{res}}\overline{T}) - \partial_z(w_{\mathrm{res}}\overline{T}) - \nabla \cdot \mathscr{F}_{\mathrm{res}}\{T\} + \overline{\mathscr{S}}. \tag{6.10}$$

TEM theory is formulated in such a way that, away from horizontal boundaries, the divergence of the residual flux vanishes. Let us assume for the moment that we are interested in closing the temperature budget in the interior only. Then, the residual temperature equation reads:

$$\partial_t \overline{T} = -\partial_y(v_{\text{res}}\overline{T}) - \partial_z(w_{\text{res}}\overline{T}) + \overline{\mathscr{S}}. \tag{6.11}$$

The discrete residual budget

Equation (6.11) is the TEM equivalent of equation (6.3), so that the advection terms are now expressed in terms of the residual velocity, which provides a clean framework to relate temperature changes to residual concepts like, for example, eddy compensation. However, the terms on the right hand side need to be computed explicitly. To this aim, we start by computing the residual streamfunction ψ_{res} as detailed in chapter 3, and obtain v_{res} and w_{res} through formula (2.71). Next, we interpolate the zonal average temperature twice to the y- and z- interfaces. The final part is to represent the gradients by finite differencing:

$$\begin{aligned}\text{ADV\_vres} &= -\partial_y(v_{\text{res}}\overline{T}) \approx \\ &\approx -(v_{\text{res}}(iy+1,iz)\overline{T}_v(iy+1,iz) - v_{\text{res}}(iy,iz)\overline{T}_v(iy,iz))/\text{dyG}(iy,iz),\end{aligned} \tag{6.12}$$

and:

$$\begin{aligned}\text{ADV\_wres} &= -\partial_z(w_{\text{res}}\overline{T}) \approx \\ &\approx -(w_{\text{res}}(iy,iz)\overline{T}_w(iy,iz) - w_{\text{res}}(iy,iz+1)\overline{T}_w(iy,iz+1))/\text{drF}(iy,iz),\end{aligned} \tag{6.13}$$

where $\overline{T}_v$ and $\overline{T}_w$ represent the temperature field on y- and z- interfaces respectively, and dyG is the spacing in the meridional direction between points on y-interfaces.

Here, we have approximated the meridional and vertical gradients with a simple, first-order finite difference formula, whereas the idealised channel employs the more complicated Second Order Momentum scheme of Prather. This implies that, even when the divergence of the residual flux vanishes, ADV_vres and ADV_wres do not correspond perfectly to the advection terms actually used by the model to march its state forward. Thus, the discrete representation of equation (6.11):

$$\text{TTEND} = \text{ADV\_vres} + \text{ADV\_wres} + \text{TFLUX} + \text{MIX}, \tag{6.14}$$

will not be an exact equation in general. Note also that equation (6.14) is only valid away from horizontal boundaries: near the surface and the bottom large errors are expected due to the fact that the residual flux is significantly different from zero, which is an expression of the difficulties introduced by TEM theory in the treatment of the boundary layers. Nevertheless, the residual temperature budget can be used effectively in the interior, as illustrated in section 6.4.

6.3.3 The time integrated budget

Suppose that the model is in a statistically equilibrated state. Then, the time average of zonal-mean temperature tendency is zero:

$$\overline{\partial_t T} = \frac{1}{\tau}\int_0^\tau \mathrm{d}t\,\partial_t T(t) = \frac{1}{\tau}(T(\tau) - T(0)) = 0, \tag{6.15}$$

for $\tau >> 1$. The idea is simply that temperature variance must be bounded if the state is statistically equilibrated, and bounded random temperature fluctuations cancel out in the time average. Note that the bar denotes time average here, and the convention is that all budget variables are zonally averaged. In terms of the discrete budget variables, this reads:

$$\overline{\text{TTEND}} = \overline{\text{ADV\_h}} + \overline{\text{ADV\_v}} + \overline{\text{TFLUX}} + \overline{\text{MIX}} = 0\,. \tag{6.16}$$

The second equality holds strictly when the time interval upon which the time average is taken tends to infinity. In the control run, we take time average over a large but finite time interval (18 years), therefore we rewrite the above as:

$$\overline{\text{TTEND}}^c = \overline{\text{ADV\_h}}^c + \overline{\text{ADV\_v}}^c + \overline{\text{TFLUX}}^c + \overline{\text{MIX}}^c = \varepsilon, \tag{6.17}$$

where $\varepsilon \approx 0$ is a small residual due to the finite size of the sample and $\overline{\cdot}^c$ denotes a long time-average in the control run. Equivalently, using the residual framework we can write:

$$\overline{\text{TTEND}}^c = \overline{\text{ADV\_vres}}^c + \overline{\text{ADV\_wres}}^c + \overline{\text{TFLUX}}^c + \overline{\text{MIX}}^c = \varepsilon, \tag{6.18}$$

although in practice the ε term will differ in the two equations due to the fact that the residual budget does not close exactly.

This picture is satisfactory for a statistically equilibrated state, and we will see hands-on how it unfolds for the control run in section 6.4.2. If the model is not in a statistically equilibrated state, however, time averages are not especially meaningful and a formulation based on the equations above becomes of little relevance. Neither it is particularly illuminating to look at the time evolution of the temperature tendency budget components, as one usually reasons in terms of temperature changes and not in terms of changes of its derivative. Then, the next logical step is to take the time integral of the temperature tendency equation, from time 0 to t:

$$T(t) = T(0) + \int_0^t \mathrm{d}s \, \mathrm{TTEND}(s) \, . \tag{6.19}$$

Here, we identify $t = 0$ with the time at which the wind step change is applied. If we expand TTEND by means of equation (6.7), we obtain:

$$T(t) = T(0) + \int_0^t \mathrm{d}s \, [\mathrm{ADV\_h}(s) + \mathrm{ADV\_v}(s) + \mathrm{TFLUX}(s) + \mathrm{MIX}(s)] \, , \tag{6.20}$$

or, more compactly:

$$T(t) = T(0) + \sum_i \int_0^t \mathrm{d}s \, \mathrm{TBDG}_i(s) \, , \tag{6.21}$$

where the summation index i runs on the right hand side terms of equation (6.7). A similar equation holds for the residual budget, equation (6.14).

Rather than temperature itself, the object of study in perturbation experiments is often the temperature anomaly with respect to the averaged control run state (Doddridge et al., 2019, Ferreira et al., 2015, Kostov et al., 2017, Seviour et al., 2017):

$$T'(t) = T(t) - \overline{T}^c. \tag{6.22}$$

Expressing the temperature budget equation (6.7) (or equation (6.14)) in terms of the temperature anomaly gives:

$$T'(t) = T(0) - \overline{T}^c + \sum_i \int_0^t \mathrm{d}s \, \mathrm{TBDG}_i(s). \tag{6.23}$$

The final step is to re-arrange this equation so that the anomalies of the budget components appear on the right hand side. To this aim, we exploit:

$$\sum_i \mathrm{TBDG}_i(s) = \sum_i \mathrm{TBDG}_i'(s) + \sum_i \overline{\mathrm{TBDG}_i}^c = \sum_i \mathrm{TBDG}_i'(s) + \varepsilon\,, \tag{6.24}$$

yielding:

$$T'(t) = T(0) - \overline{T}^c + \sum_i \int_0^t \mathrm{d}s\, \mathrm{TBDG}_i'(s) + \varepsilon t. \tag{6.25}$$

where $\varepsilon \approx 0$ is the small residual of the control run time-averaged temperature budget. When ε is so small that can be neglected (for very long time averages in the control run) the equation further simplifies to:

$$T'(t) = T(0) - \overline{T}^c + \sum_i \int_0^t \mathrm{d}s\, \mathrm{TBDG}_i'(s). \tag{6.26}$$

Equation (6.25) is yet another representation of the temperature budget. We will see in section 6.4.3 that it is especially convenient to study how advective and mixing flux anomalies drive anomalous temperature changes. In particular, we can single out the effect of a particular physical process on temperature anomaly by defining:

$$T_i'(t) = T(0) - \overline{T}^c + \int_0^t \mathrm{d}s\, \mathrm{TBDG}_i'(s), \tag{6.27}$$

where T_i' is the the temperature anomaly that would be observed assuming that all the physical processes appearing in the temperature budget were unchanged with respect to the control run averaged state except for TBDG_i.

6.3.4 Summary of methods

The temperature budget is a powerful diagnostic tool which allows to study the physical processes driving temperature changes at a given place and at a given time. The temperature budget can be expressed in many different forms, but three equations will suffice to our goals: equations (6.7) and (6.14) tell us that the temperature tendency (i.e., the time derivative of temperature) is the sum of several contributions representing the individual effects of different physical processes, including advective fluxes, diffusive fluxes, and boundary effects. The difference between equation

(6.7) and equation (6.14) is that in equation (6.7) the advective fluxes are obtained from the usual zonal-average temperature equation (6.3), and therefore do not differentiate between eddy and mean flow fluxes. Equation (6.14), on the other hand, is based on the residual mean temperature equation (6.10), and (at the cost of a few approximations that are not very accurate near the surface) expresses the advective fluxes in terms of the residual circulation. Equations (6.7) and (6.14) are meaningful in both a statistically equilibrated state and a transient, out-of-equilibrium state. In the latter case, however, they can be complemented by equation (6.25), which describes how the aforementioned processes govern the time evolution of temperature (or of temperature anomaly) rather than that of its time derivative.

6.4 Results

The presentation of our results is organised as follows: firstly, in section 6.4.1 we will briefly survey the equilibrated state which the idealised channel attains after the wind-stress perturbation is applied. Although the main focus of this work is the transient response, the analysis of the perturbed channel at equilibrium provides ourselves with a general sense of how the final state of the channel looks like. Secondly, we will study the closure of the temperature budget in the equilibrated control run, section 6.4.2. This preliminary step is necessary because (i) the control run-averaged budget components appear in the anomalous temperature budget for the perturbation experiments, equation (6.25), and (ii) it is desirable to test the temperature budget technique in the simpler case of the control run before considering the more complicated perturbation experiment. Thirdly, patterns of transient temperature and circulation changes are investigated in section 6.4.3. We conclude in section 6.5.

6.4.1 The equilibrated response

In this section, we briefly describe the physics of the statistically equilibrated idealised channel subject to the wind stress perturbation. We follow the discussion of section 3.3 (to which we refer for details), and we place a special emphasis in highlighting the differences between the equilibrated channel and the control run. As usual, we begin with zonal-mean temperature.

Zonal-mean temperature

Figure 6.1 (left) shows time-mean, zonal-mean temperature for the perturbed channel at equilibrium (colours). The temperature range is unchanged with respect to the control run. This is expected, as both the upper and the lower bounds are dictated by the restoring condition at the northern boundary, which is the same in the two experiments. The mean stratification in the top 1000 meters also shows little sensitivity, with the associated period $2\pi/N$ changing from 46 min in the control run to 49 min in the perturbed run. The mean depth of the mixed layer (dashed black line) is slightly larger in the perturbed run than in the control run in the intermediate and northern regions, while it is drastically deeper close to the southern boundary, where the stratification is weak. The deepening of the mixed layer is likely driven by increased wind stirring through the KPP scheme and is associated to the steepening of the mean isopycnals, which is apparent from the comparison with the control run mean temperature isolines (black contours). The change in slope is particularly marked in the central and southern parts of the domain, but becomes less significant nearer the northern boundary, where temperature is relaxed to a prescribed stratification profile. At intermediate latitudes and near the surface, isopycnal steepening is associated with the disappearance of a region of weak surface meridional gradient found in the control run. Due to the fact that temperature decreases southward and downward, the regions interested by the steepening of the isopycnals display anomalous cooling, as illustrated in figure 6.1 (right). The cool anomaly is especially pronounced near the surface at around $y = 750$ km, where temperature

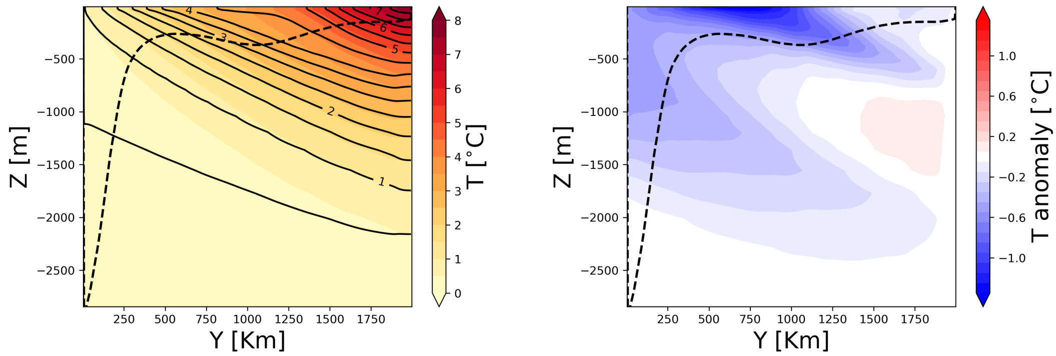

Figure 6.1: Left: time-mean, zonal-mean temperature in the equilibrated channel. The dashed black line marks the depth of the time-mean, zonal-mean mixed layer, and black contours represent time-mean, zonal-mean temperature in the control run. Right: time-mean, zonal-mean temperature anomaly (colours) and time-mean, zonal-mean depth of the mixed layer (dashed black line).

decreases by as much as 1.44° C, and, interestingly, along a cool tongue that propagates the surface anomaly equatorward and towards the interior. A second cold anomaly tongue is found at greater depths, separated from the first one by a region of weaker warm anomaly. We will see below that these patterns of temperature anomaly are associated to modulations of the meridional circulation.

Zonal circulation

The time-mean, zonal-mean zonal component of velocity is shown in figure 6.2 (left). The zonal flow has a meridional structure comparable to that of the control run but, predictably, is characterised by larger average values of the velocity. For a more quantitative comparison, we have computed the barotropic flow, $U_{bt} = 1136$ Sv, and the baroclinic flow, $U_{bc} = 102$ Sv: these values should be contrasted with those for the control run, $U_{bt}^{(c)} = 433$ Sv and $U_{bc}^{(c)} = 88$ Sv respectively (section 3.3.3).

With a factor 2.6 increase, the barotropic transport nearly triples from the control run to the equilibrated perturbed run. The barotropic transport is dominated

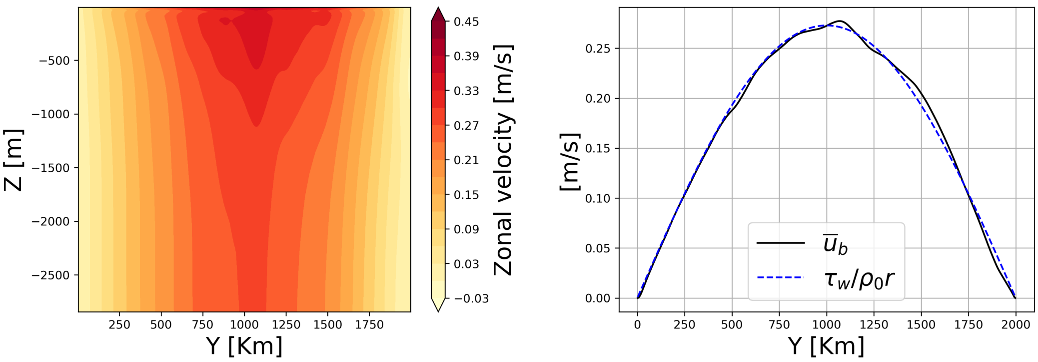

Figure 6.2: Left: time-mean, zonal-mean zonal velocity. Right: zonal momentum balance. The continuous black line shows the meridional profile of the time-mean, zonal-mean zonal bottom velocity, while the dashed blue line shows the theoretical prediction $\tau_w(y)/\rho_0 r_b$.

by the large flow at the bottom which, according to theory, scales linearly with the wind stress, as demonstrated by figure 6.2 (right). Therefore, the large increase in the barotropic flow reflects the threefold increase in the wind forcing. By comparison, the baroclinic transport increases by a modest 15%. This is because U_{bc} is controlled, through the thermal wind relation, by the slope of the isopycnals. The large scale structure of the isopycnals, however, is in turn constrained by the restoring conditions at the northern boundary (Abernathey et al., 2011), implying that the channel is close to a state of eddy saturation (i.e., that baroclinic transport depends weakly on surface wind stress).

Thermal wind

The equilibrated channel is in the low Rossby number regime (formula (2.10) with $U \approx 0.3$ gives $Ro \approx 3 \cdot 10^{-4} - 10^{-3}$) and, similarly to the control run, geostrophic scaling is expected to hold. The thermal wind relation combines the geostrophic and hydrostatic approximations, and stipulates that the vertical shear of the zonal velocity is controlled by the meridional gradient of buoyancy. As in section 3.3.6, we have tested the thermal wind relation by comparing both sides of equation (3.23), figure 6.3 (left), and of equation (3.24), figure 6.3 (right). The qualitative agreement

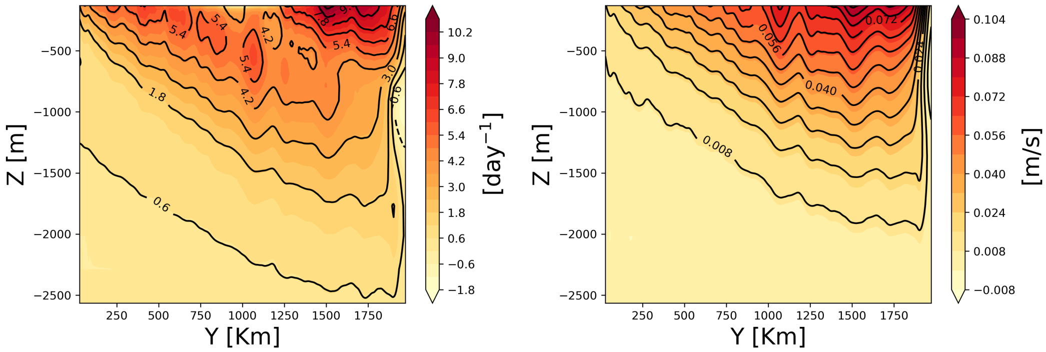

Figure 6.3: Test of the thermal wind relation. Left: time-mean, zonal-mean vertical shear of zonal velocity (colours), and the corresponding prediction from thermal wind given by the right hand side of equation (3.23) (black contours). Right: time-mean, zonal-mean baroclinic zonal velocity (colours), and the corresponding prediction from thermal wind given by the right hand side of equation (3.24) (black contours).

is satisfactory, and we conclude that geostrophic scaling applies to the perturbed channel at equilibrium.

Meridional circulation

Figure 6.4 (left) shows the time-mean Eulerian streamfunction for the perturbed channel at equilibrium. The structure of the Eulerian circulation in the meridional plane is identical to that of the control run, see also the anomaly field in figure 6.4 (right), and in good agreement with the theoretical prediction of equation (2.136) (black contours. The theoretical streamfunction is parametrised so that it scales linearly to zero within the surface and bottom frictional layers). The magnitude of the circulation is different in the two experiments though, with $\overline{\psi}$ at $y = 200$ km (depth is unimportant as long as the value is taken in the interior) ranging from 1.12 Sv in the control run to 3.34 Sv in the equilibrated channel. This three-fold increase neatly reflects the linear dependency of $\overline{\psi}$ on surface wind stress, equation (2.136).

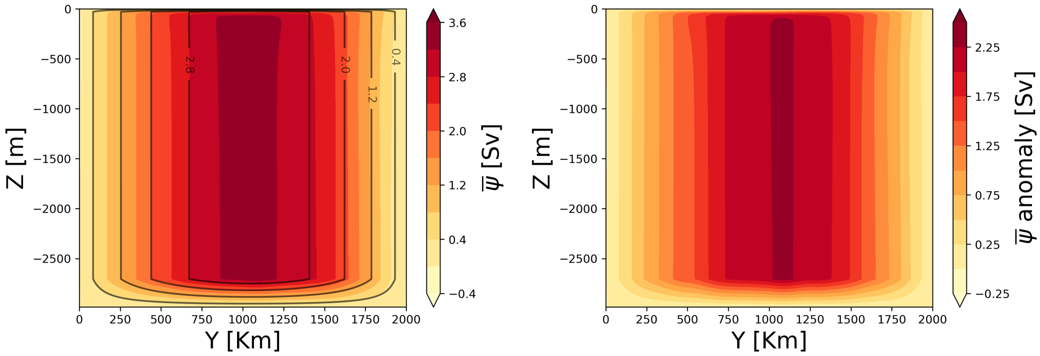

Figure 6.4: Left: time-mean Eulerian streamfunction (colours). Black contours represent the theoretical prediction from equation (2.136) (we assume that $\overline{\psi}$ goes linearly to zero in the frictional layers). Right: time-mean Eulerian streamfunction anomaly.

The modulations of the residual streamfunctions are more nuanced. Figure 6.5 shows the residual streamfunction in isopycnal coordinates: similarly to the control run (panel (a)), in the perturbation experiment (panel (b)) we observe that contours are largely horizontal away from the surface and northern diabatic layers (the dashed black line marks the time-mean, zonal-mean sea surface temperature and thus represents an estimate of the mean location of the surface diabatic layer), which is an expression of the fact that the residual circulation advects tracers along isopycnals in the interior. Diabatic exchanges at the fluid's boundaries support a cross-isopycnal flow and allow for closure of the circulation.

The residual circulation is again partitioned in three separate units (negative lower cell, intermediate positive cell, and negative upper cell), with positive values of the streamfunction associated with clock-wise circulation, and *vice versa*. Although it is not immediately straightforward to interpreter changes in the location, extent, and magnitude of the individual cells by analysing the residual circulation in isopycnal coordinates, we note that the positive anomaly in the temperature range $1-3$ °C, figure 6.5 (c), is associated with the intermediate cell crossing the diabatic layer (Abernathey et al., 2011). In other words, while the positive intermediate cell attains its maximum value within the surface diabatic layer in the control run, the same cell resides entirely in the interior in the perturbed channel. Structural changes

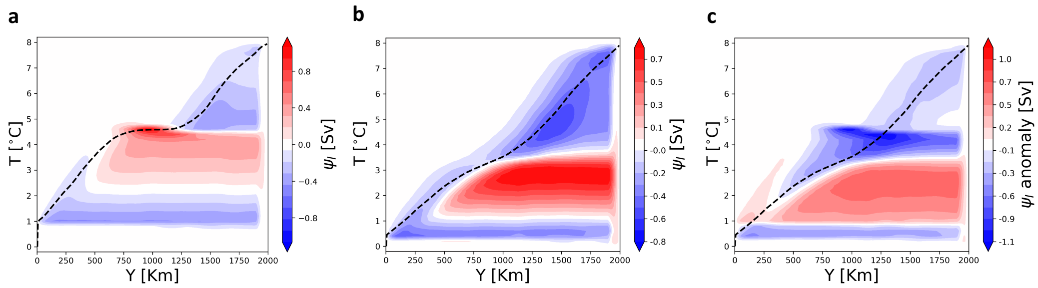

Figure 6.5: Time-mean residual streamfunction in isopycnal coordinates (colours) for the control run (a) and the equilibrated perturbation experiment (b). The dashed black line represents the control and equilibrated time-mean, zonal-mean sea surface temperature, respectively. (c): Equilibrated residual streamfunction anomaly. The dashed black line is as in (b).

in the residual circulation are more transparent when the streamfunction is mapped to height coordinates, figure 6.6.

Changes in the intermediate cell are accompanied by a deepening of the negative lower cell (which corresponds to the upwelling branch of the meridional circulation moving at greater depths) and of the negative upper cell, see also the anomaly field in figure 6.6 (c). The response of the meridional circulation's magnitude is complex: following Abernathey et al. (2011), we quantify the magnitude of the intermediate and lower cells by taking the maximum and minimum of the residual streamfunction at $y = 1800$ km and below $z = -500$ m depth respectively (the precise meridional location where the extrema are taken is not important because ψ_{res} is nearly constant along isopycnals below the surface layer). We find that the intermediate cell is weakly sensitive to the wind stress perturbation, with the maximum of ψ_{res} increasing from approximately 0.4 Sv in the control run to approximately 0.8 Sv in the perturbed channel, corresponding to a factor 2 increase. This should be contrasted with the three-fold increase in the magnitude of the Eulerian circulation, discussed above. Note, however, that the intermediate cell attains its maximum within the boundary layer in the control run, with a value of approximately 0.8 Sv (see figure 3.8 (right)): the transport increase associated with the intermediate cell vanishes if computed against this value. The lower cell is practically insensitive to the pertur-

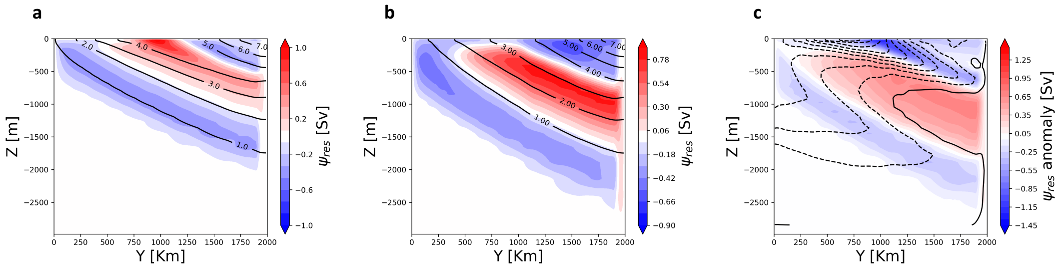

Figure 6.6: Time-mean residual streamfunction in height coordinates (colours) for the control run (a) and the equilibrated perturbation experiment (b). Black contours represent the control and equilibrated time-mean, zonal-mean temperature profiles, respectively. (c): Equilibrated time-mean residual streamfunction anomaly in height coordinates (colours). Black contours represent the equilibrated time-mean, zonal-mean temperature anomaly profiles.

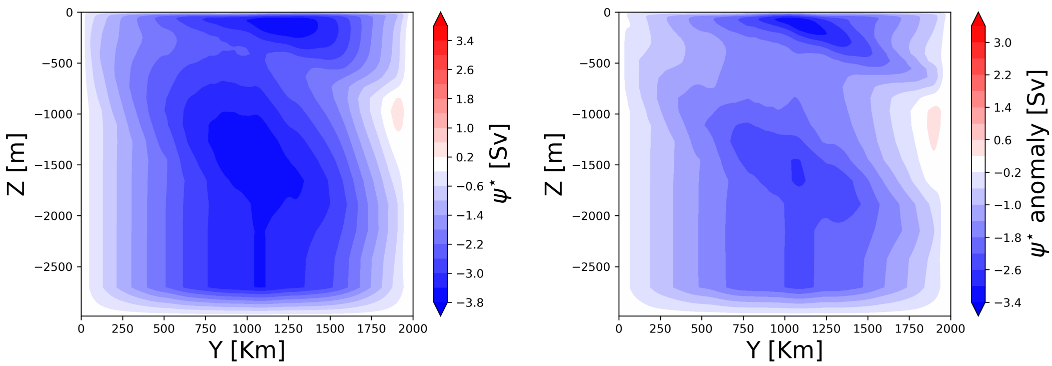

Figure 6.7: Left: time-mean eddy-induced streamfunction. Right: time-mean eddy-induced streamfunction anomaly.

bation too, with the minimum of ψ_{res} changing from -0.4 Sv in the control run to -0.3 Sv in the equilibrated channel. These results are in broad agreement with the findings of Abernathey et al. (2011). Overall, we find that the model is close to a state of eddy compensation: the magnitude of the residual overturning cells is less sensitive to the wind stress change than the Eulerian circulation because baroclinic eddies spin up to counter-balance the anomalous wind-induced circulation. This is reflected in the absolute increase of the eddy-induced circulation, illustrated in figure 6.7.

It is remarkable that although changes in the intensity of individual cells are moderate, local streamfunction anomalies can take on much larger values (figure 6.6 (c)), which tend to be associated with the spatial repositioning of the cells rather than with the modulation of their strength, as described above. These changes are important to the equilibrated structure of zonal-mean temperature, as is clear from comparison of the time-mean residual streamfunction anomaly (figure 6.6 (c); colours) with the time-mean, zonal-mean temperature anomaly (contours). In particular, the upper cold anomaly tongue propagating equatorward and towards the interior from the surface appears related to the negative near-surface anomaly of the residual streamfunction, while the lower cold anomaly tongue seems associated to the deepening of the lower cell of the MOC. We conclude that studying changes in the magnitude of the residual streamfunction only is not sufficient to fully understand temperature changes in the channel.

We will explore the relation between (transient) streamfunction and temperature anomalies further in section 6.4.3 by taking advantage of the temperature budget framework. Specifically, we will average the temperature tendency and the temperature budget components over four rectangular domains on the meridional plane, and, for each domain, we will determine which processes drive the observed temperature changes. The averaging domains are shown in figure 6.8 (continuous black lines), together with the perturbed channel's equilibrium temperature anomaly (colours). The domains are chosen so as study the physical mechanisms driving, respectively:

1. The cold interior anomaly, poleward of the main ACC flow
2. The warm interior anomaly, equatorward of the main ACC flow
3. The cold surface anomaly at intermediate latitudes
4. The cold subsurface anomaly at intermediate latitudes

Note that domains 1 and 2 are co-located with the regions of anomalous wind-induced upwelling and downwelling respectively, see figure 6.4 (right). The surface domains 3 and 4 are positioned where the surface response is largest.

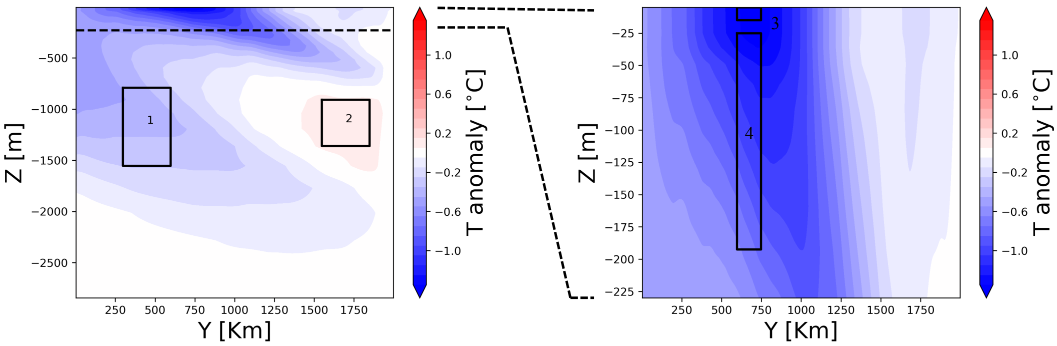

Figure 6.8: Left: Equilibrated time-mean, zonal-mean temperature anomaly (colours), and the rectangular averaging domains (continuous black lines). Right: a close-up of the surface layer.

6.4.2 Temperature budget in the control run

Before turning our attention to the transient adjustment of the idealised channel to the wind stress perturbation, we demonstrate the closure of the temperature budget through equations (6.7) and (6.14) in the simpler case of the control run. In particular, we average the temperature tendency and the temperature budget components spatially over the four domains shown in figure 6.8 and in time over 18 years of simulation, and assess whether the sum of the individual components equals the average temperature tendency.

The time-mean, zonal-mean temperature budget for domain 1, computed through equation (6.7), is shown in figure 6.9 (a). Only the advection terms contribute to the budget, which is not surprising since the domain is located in the interior, where KPP mixing and the surface temperature flux vanish. The two advection terms have opposite sign, with horizontal advection warming the domain on average, and vertical advection cooling it. Their sum is nearly zero, matching the diagnosed time averaged temperature exactly (we expect zero or small time averaged temperature tendency because the control run is a statistically equilibrated state of the idealised channel). The temperature budget for the same domain, but computed according to the residual equation (6.14), is shown in figure 6.9 (b). Once again, it is only the advection terms that contribute to the budget non negligibly: their sum is nearly zero, and matches the diagnosed temperature tendency. Interestingly, though, their sign is reversed compared to the previous case, with the residual horizontal advection term acting to cool the domain and the vertical term acting to warm it. Domain 1 is partially co-located with the downwelling limb of the intermediate cell, see figure 3.9 (left): therefore, the horizontal residual velocity advects on average cold water equatorward, and the vertical residual velocity advects warm surface water downward into the domain. The implied tendencies are in agreement with figure 6.9 (b). The advection terms in equation (6.7), on the other hand, include both eddy and mean flow contributions, so that their physical interpretation through figure 6.9 (a) is less straightforward than in the case of the residual budget.

The ordinary and residual time-mean temperature budgets for domain 2 are shown in figures 6.9 (c) and 6.9 (d) respectively. Domain 2 is in the interior and partially co-located with the upwelling branch of the MOC, see figure 3.9 (left): as a consequence, the outcome of the temperature budgets is symmetrical to that of domain 1, which is located in a region of downwelling. Specifically, only the advection terms contribute to the temperature tendency, with residual vertical transport acting to cool the domain and horizontal transport acting to warm it. The sign of the advection terms is reversed with respect to the ordinary budget equation. For both figure 6.9 (c) and figure 6.9 (d), the sum of the advection terms is nearly zero (note the different scale on the vertical axis), and matches the diagnosed tendency well (although a small error is noticeable for the residual budget: this is a manifestation of the fact that the residual budget is not an exact equation, as discussed above).

The ordinary and residual temperature budgets for the surface domain 3 are shown in figures 6.10 (a) and (b). In this case, all terms in the budget contribute to the temperature tendency. Domain 3 is located in a region where the buoyancy flux at the surface is into the ocean (see for example figure 3.2), therefore the temperature flux term acts to warm the domain. This tendency is nearly compensated by the mixing term, which acts to transfer heat downward and cool the surface. The temperature flux and mixing terms do not change between the ordinary and residual

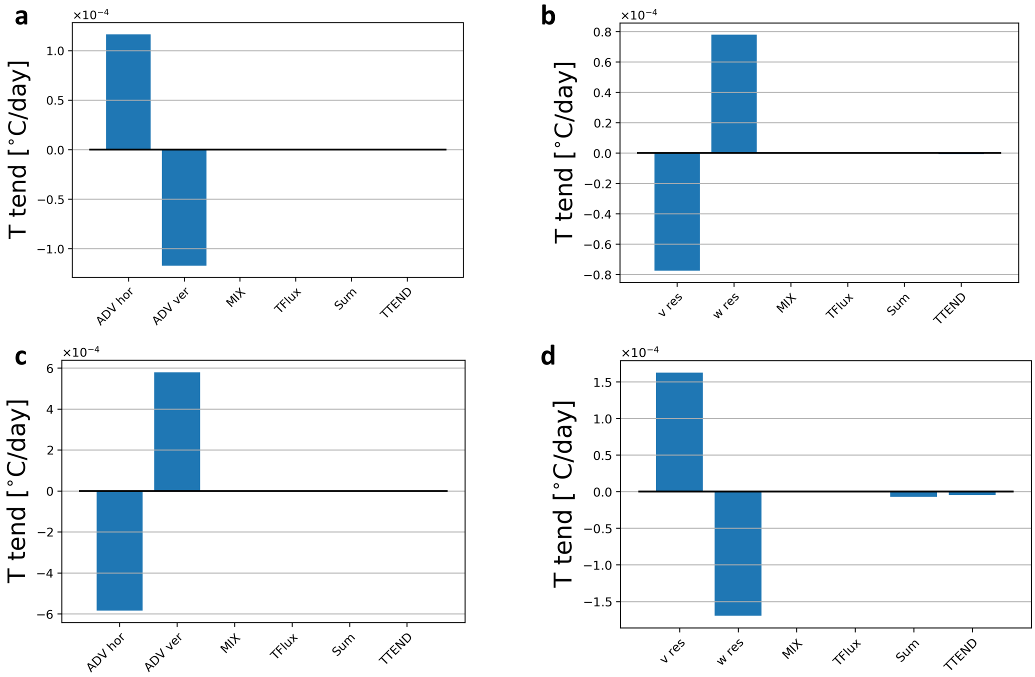

Figure 6.9: Time-mean, zonal-mean temperature budget for domains 1 (top) and 2 (bottom) in the control run. The budget components are computed according to equations (6.7) (left) and (6.14) (right). The bar plots also show the sum of the budget components (Sum) and the diagnosed temperature tendency (TTEND): the two terms are equal if the budget closes exactly. The time mean is computed over a period of 18 years.

formulations of the temperature budget (note the different scale on the vertical axis). The horizontal and vertical advection terms are in near balance with each other and, in the case of the ordinary budget of figure 6.10 (a), cancel out exactly with the mixing and temperature flux terms. Therefore, the total sum of the budget components is zero, in perfect agreement with the diagnosed temperature tendency. Note that the temperature tendency contributed by the individual budget components is two orders of magnitude larger than that for the interior domains of figure 6.9. In the case of the residual budget of figure 6.10 (b), instead, the advection terms do not cancel with the mixing and temperature flux terms: the sum of the budget components yields an overall tendency of about $-2 \cdot 10^{-3}$ °C/day, whereas the diagnosed tendency is zero. The discrepancy between the predicted and diagnosed tendencies may appear modest in comparison with the magnitude of the individual budget terms, but the error is large as it corresponds to a spurious tendency of about 0.7 °C/year. This underscores that the residual budget, as formulated in equation (6.14), cannot be employed effectively near the surface layer. As discussed in section 6.3, the limitation arises from the fact that the residual flux is not zero near the horizontal boundaries. Parametrising the quasi-Stokes streamfunction within the boundary layer may provide a pathway to circumvent the problem (see the discussion in section 2.4.3), but to test the idea is beyond the scope of this work.

Finally, the ordinary and residual temperature budgets for the sub-surface domain 4 are shown in figures 6.10 (c) and (d) respectively. The temperature flux term is non-zero only at the surface, and thus does not contribute here. The mixing term is positive which, consistently with the discussion above, implies that mixing acts to transfer heat from the surface (where it is injected in the system by the fixed buoyancy flux) to the base of the mixed layer at the considered latitudes. At leading order, the warming tendency driven by mixing is balanced by the horizontal advection term in the ordinary budget, with the vertical advection also contributing negatively to the total tendency: the sum of the components is approximately zero, and captures the diagnosed tendency exactly. Similarly to the case of domain 3, instead, the residual temperature budget is not closed: the components sum to

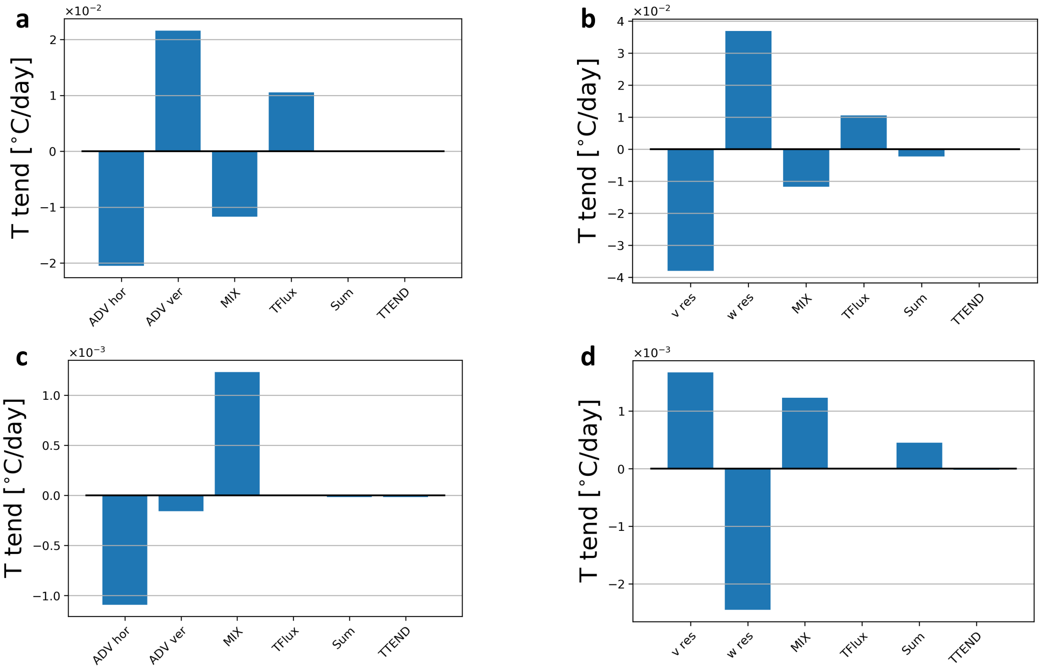

Figure 6.10: Time-mean, zonal-mean temperature budget for domains 3 (top) and 4 (bottom) in the control run. The budget components are computed according to equations (6.7) (left) and (6.14) (right). The bar plots also show the sum of the budget components (Sum) and the diagnosed temperature tendency (TTEND): the two terms are equal if the budget closes exactly. The time mean is computed over a period of 18 years.

approximately $5 \cdot 10^{-4}$ °C/day, whereas the diagnosed tendency is at least on order of magnitude smaller. The error is smaller than in the case of the surface layer but still sizeable (about 0.18 °C/year), implying that the residual budget is not suited to study temperature changes in the sub-surface domain.

Overall, we have shown that the ordinary temperature budget based on equation (6.7) is satisfied exactly for the control run within all of the four averaging domains considered. The residual budget of equation (6.14) is easier to interpreter physically, but is only closed to a satisfactory degree of approximation for the interior domains 1 and 2. For this reason, in the next section we will analyse temperature changes in the surface layers based on the ordinary formulation of the temperature budget only.

6.4.3 The transient response

Temperature response

Figure 6.11 shows the ensemble-mean, zonal-mean temperature anomaly at different instants of time after the perturbation. The ensemble average is computed over 15 members up until month 36, and over 6 members after that time. A rolling average with a window of six months was applied to all data for smoothing purposes. The initial temperature response ($t = 1$ month) is weak, negative, and largely confined to the surface layers. This surface negative anomaly is a robust feature of Southern Ocean models (Abernathey et al., 2011, Doddridge et al., 2019, Ferreira et al., 2015), and is usually associated to anomalous meridional Ekman transport (i.e., to the strong, wind-induced meridional velocity perturbation acting on the background gradient of surface temperature, see below). One year after the perturbation ($t = 12$ months), the negative surface anomaly has considerably intensified and propagated downward, beneath the surface layer. A meridional dipole has also developed, with weak cold anomalies appearing in the southern regions of the domain, and weak warm anomalies in the northern ones. The meridional dipole is associated to anomalous patterns of upwelling and downwelling, as discussed below. The meridional structure of temperature anomaly at subsequent times ($t = 24 - 72$ months) is similar, with the cold surface anomaly continuing to propagate downward along anomalous mean residual streamlines (shown in black contours). Six years after the perturbation ($t = 72$ months), temperature appears approximately equilibrated in the top layers (compare with the equilibrated temperature response of figure 6.1). At greater depths, the deep cold anomaly tongue descried in section 6.4.1 has not fully developed, and the interior layers are thus still in a state of slow transient evolution.

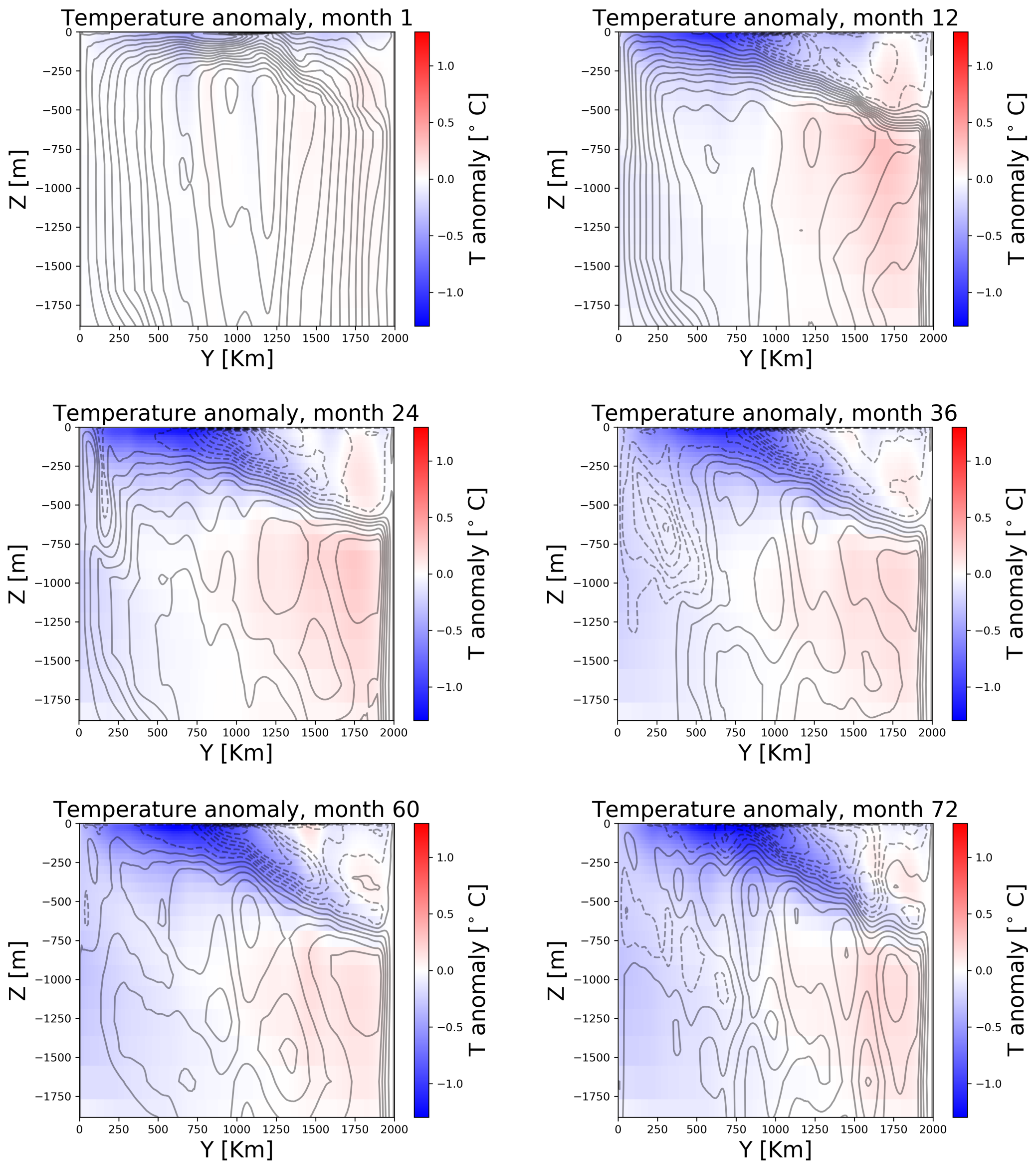

Figure 6.11: Ensemble-mean, zonal-mean temperature anomaly at various times after the wind stress perturbation is applied. Black contours represent the corresponding ensemble-mean residual streamfunction anomaly. A six-months rolling average is applied to all data for smoothing. The ensemble average is computed over 15 members up until month 36 after the perturbation. After that time, 6 members only are considered.

Circulation response

Temperature changes are related to circulation changes. The ensemble-mean residual streamfunction anomaly at various instants of time after the perturbation is shown in figure 6.12 (see also the black contours in figure 6.11). A six-months rolling average was applied to all data for smoothing purposes. The initial response of the residual streamfunction ($t = 1$ month) is conspicuous and resembles the Eulerian circulation anomaly, shown in figure 6.4 (right): streamlines are vertical in the interior, while Ekman stresses in the surface and bottom Ekman layers allow for closure of the meridional circulation. Within the surface layer, the anomalous Eulerian circulation is northward. The anomalous velocity acts on a positive meridional gradient of surface temperature, therefore the initial velocity anomaly advects cold water equatorward (Ferreira et al., 2015). Deep water is upwelled to the surface in the southern flank of the ACC, and surface water is downwelled in the northern flank. The anomalous Eulerian vertical velocity acts on a positive vertical gradient of temperature, therefore this mechanism produces cooling to the south and warming to the north (figure 6.11). In the presence of a temperature inversion, the same mechanism can result in subsurface warming in the southern flank of the ACC, which corresponds to the slow time-scale mechanism proposed by Ferreira et al. (2015). A southward return flow located in the bottom layer closes the meridional circulation.

Overall, the initial response of the residual streamfunction is dominated by the so-called Deacon cell pattern, discussed at some length in sections 3.3.8 and 6.4.1. By year 1 after the perturbation, however, the Deacon pattern is distorted by the appearance of a near-surface negative anomaly which, as illustrated in section 6.4.1, is associated to the downward propagation of the intermediate positive cell. Due to $\psi_{\text{res}} = \overline{\psi} + \psi^*$, equation (2.73), the departure from the Eulerian anomaly is also associated with the invigoration of baroclinic eddies. As noted in the previous section, negative surface temperature anomalies appear to propagate along the corresponding anomalous residual streamlines, figure 6.11.

Similarly to the temperature response, the residual circulation in the top layers equilibrates rather quickly (compare with the equilibrated residual streamfunction anomaly, shown in figure 6.6 and in black contours in figure 6.12), while the transient adjustment takes significantly longer in the interior: 6 years after the perturbation, the deep circulation is not fully equilibrated. We can estimate the adjustment time scales of the large scale circulation by averaging the residual, Eulerian, and eddy-induced streamfunctions over a large domain located at the centre of the channel (similarly to Doddridge et al. (2019)), specifically between $y = 250$ km and $y = 1750$ km, and between 500 m and 2000 m depth. The evolution of the corresponding ensemble average time series is shown in figure 6.13. The response of the Eulerian circulation (continuous green line) is fast, and in approximately 6 months $\overline{\psi}$ attains its equilibrium value (dashed green line). On the contrary, the eddy-induced circulation (continuous red line) does not change significantly in the first few months after the perturbation is applied, and ψ^* only nears its equilibrium value (dashed red line) after approximately 3 years. This eddy relaxation time-scale is consistent with the findings of Doddridge et al. (2019). The response of ψ_{res} mirrors that of $\overline{\psi}$ and ψ^*: thus, the initial adjustment is dominated by the Eulerian response, and it is only at a later stage that baroclinic eddies activate to counteract the wind induced circulation, and the residual streamfunction equilibrates. A remarkable difference between the evolution of the large-scale circulation metrics in our model and that in Doddridge et al. (2019) is that here the initial residual streamfunction anomaly does not exceed the Eulerian anomaly, which is instead the case in Doddridge et al. (2019). Our findings are more aligned with the physical understanding of baroclinic instability as a process that competes against the wind-induced circulation by flattening isopycnals. On the contrary, $\psi_{\text{res}} > \overline{\psi}$ seems to imply that eddies act to reinforce the Eulerian circulation at least initially. We note though that in our experiments too the residual anomaly can exceed the Eulerian anomaly locally (not shown), an interesting but unclear feature which we do not investigate further.

Large-scale measures of circulation such as that of figure 6.13 provide a general sense for the typical time scales of adjustment but, importantly, they do not capture the variations of the equilibration rate with depth illustrated above. Moreover, we have seen that it is local circulation anomalies (possibly corresponding to relocation of the meridional circulation cells) that are important to temperature changes, and not modulations in the overall magnitude of the large-scale circulation. For this reason, in the next section we study the relationship between the local temperature and circulation response by means of the spatially averaged temperature budget.

Temperature budget

We can establish a physical link between temperature and circulation changes by means of the temperature budget technique of section 6.3. Specifically, we consider the four rectangular domains shown in figure 6.8, and for each domain we compute the ensemble average temperature tendency anomaly and the temperature anomaly evolution. The domains are located in the regions of cold (domain 1) and warm (domain 2) equilibrium interior anomaly, and in the region of largest surface (domain 3) and sub-surface (domain 4) cold equilibrium anomaly: we seek to determine which physical processes drive the observed temperature response within the domains, and to characterise the important time scales of the adjustment. For the interior domains 1 and 2, we consider both the ordinary (equation (6.7)) and the residual (equation (6.14)) budget closure formula. The ordinary formula only, however, is considered for the surface and sub-surface domains 3 and 4, as discussed in section 6.4.2. Ensemble averages are computed over six ensemble members, and anomalies of the budget components are with respect to the control run time-averaged values, see section 6.4.2. Finally, the time evolution of temperature anomaly and of its components are computed according to formula (6.25) and (6.27) respectively.

Figure 6.14 shows the temperature budget averaged over domain 1 for the first three years following the wind stress perturbation. Figures 6.14 (a) and (c) show the ensemble-mean diagnosed temperature tendency anomaly (red dots) and the ensemble-mean tendency components anomaly (see legend). The sum of the bud-

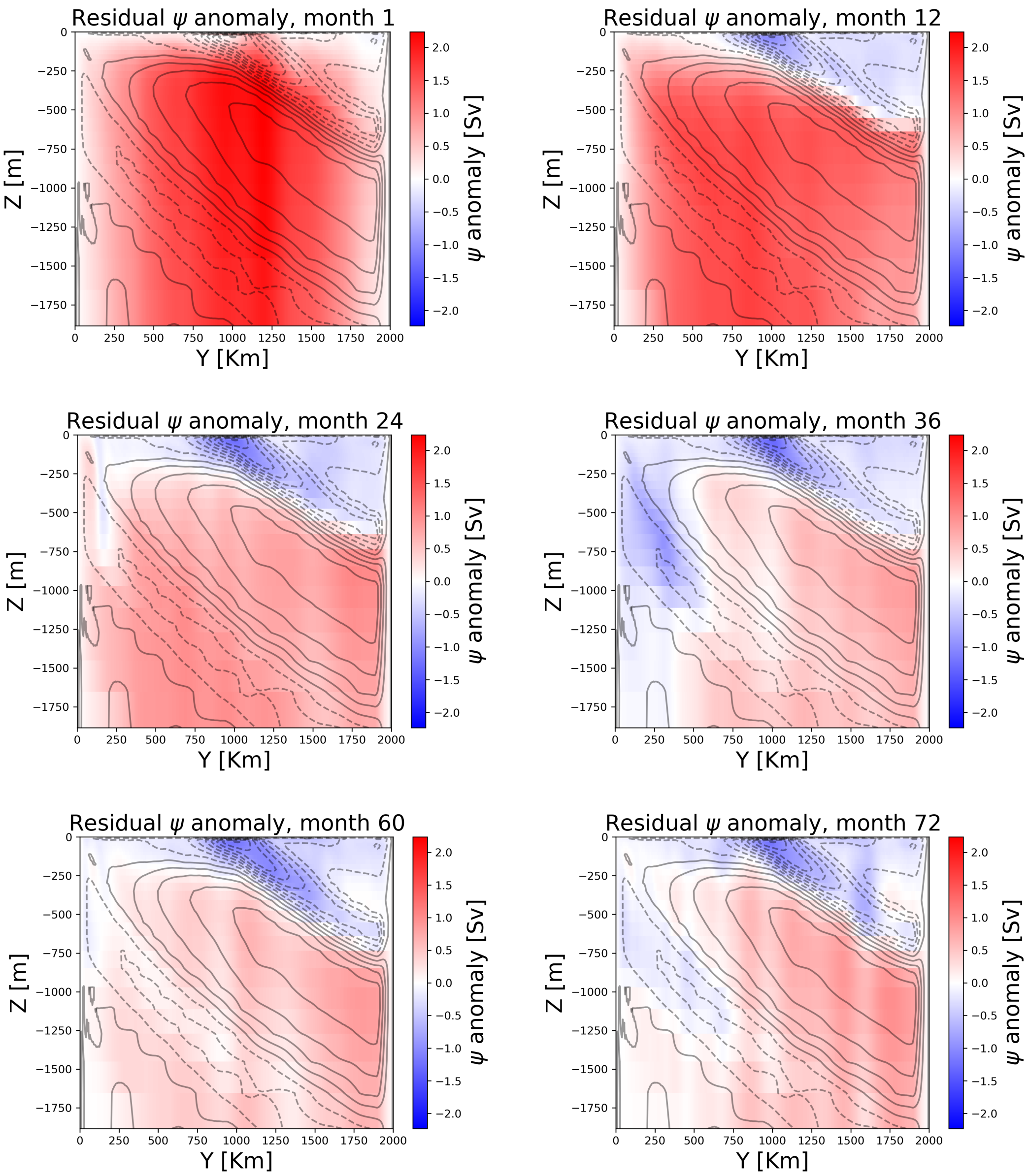

Figure 6.12: Ensemble-mean residual streamfunction anomaly at various times after the wind stress perturbation is applied. Black contours represent the equilibrated residual streamfunction anomaly. A six-months rolling average is applied to all perturbation experiment data for smoothing. The ensemble average is computed over 15 members up until month 36 after the perturbation. After that time, 6 members only are considered.

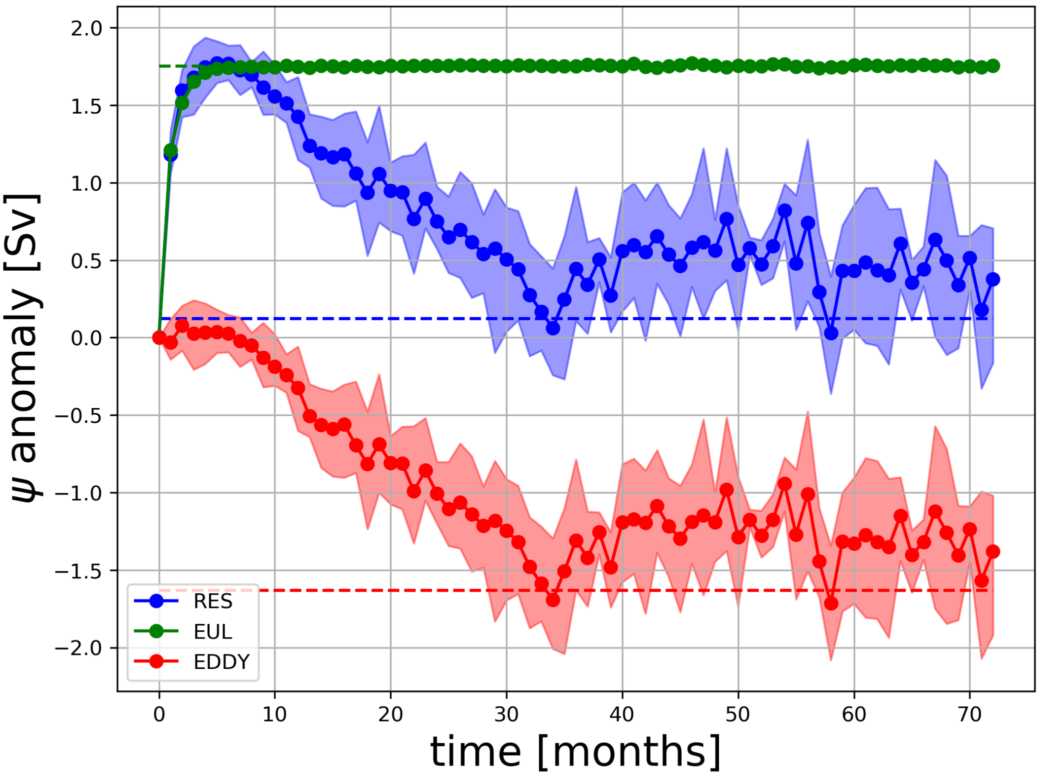

Figure 6.13: Time evolution of the Eulerian (dotted green line), residual (dotted blue line), and eddy (dotted red line) streamfunction anomalies following the wind-stress perturbation averaged between $y = [250, 1750]$ km and $z = [-500, -2000]$ m. The dashed lines mark the corresponding values for the equilibrated run, and the shaded intervals mark the ensemble spread as quantified by the 84.1 and 15.9 percentiles. The ensemble average is computed over 15 members up until month 36 after the perturbation. After that time, 6 members only are considered.

get components, which in the case of a perfectly closed budget coincides with the diagnosed tendency, is shown by the starred blue line. The tendency components are computed according to the ordinary formula (6.7) in panel (a), and to the residual formula (6.14) in panel (c). A six-months rolling average is applied to all data for smoothing purposes. Figures 6.14 (b) and (d) show instead the ensemble-mean, time-integrated temperature tendency anomaly (i.e., the temperature anomaly time evolution, continuous red line), and the ensemble-mean, time-integrated anomalous budget components (see legend), computed according to formula (6.25) for, respectively, the ordinary and residual budget equations. The sum of the components is shown by blue dots and the continuous blue line. Integral quantities are smoother by construction, so no rolling average is applied in this case.

The anomalous temperature tendency is negative at all times and quasi-steady in the period considered, see figures 6.14 (a) and (c). Accordingly, the anomalous temperature evolution is approximately a straight line, panels (b) and (d), implying that domain 1 cools linearly with time. Three years after the perturbation is applied, the temperature anomaly has not achieved its equilibrium value yet, see figure 6.15 (a), implying that the time scale of adjustment for domain 1 is larger than three years, and cannot be estimated directly with our experiment. However, by assuming that temperature decreases linearly with constant cooling rate until the equilibrium value is attained, and estimating an average tendency of -0.025 °C/month (see panel (c)), we obtain a relaxation time scale of about 11 years. Note however that this may be an underestimate if the linear approximation is only valid initially, and the full adjustment follows an exponential decay (as, anticipating on our results, is the case for domains 3 and 4, see figures 6.17 and 6.18 respectively).

The diagnosed temperature tendency is captured by the ordinary temperature budget perfectly, panel (a), confirming that equation (6.7) is exact down to numerical precision. The residual budget is also in excellent agreement with the diagnosed temperature tendency, see panel (c), which implies that TEM theory assumptions are well satisfied in the interior of the idealised channel. In both the standard and the residual picture, mixing and temperature flux anomalies play no role, and vertical advection contributes negatively to the temperature budget (which is expected as the domain is located in a region of anomalous upwelling of cold, deep water). Note also that, in both cases, vertical advection is the leading term of the budget: the physical interpretation of the balance, however, is simpler in the residual framework of figures 6.14 (c) and (d).

To see why, consider the ordinary temperature balance of figures 6.14 (a) and (b) first. Domain 1 is located in the region of interior upwelling: since the initial circulation adjustment is dominated by the Deacon cell anomaly, we would expect vertical advection driven by Ekman suction to be the only non-zero term of the budget. Moreover, as the streamlines of the equilibrium eddy-induced streamfunction are approximately vertical in the region considered (see figure 6.7), we may also imagine that the spin up of baroclinic eddies should manifest itself mainly as a modulation of the vertical advection term, with horizontal advection playing a marginal role. Figure 6.14 (a), however, shows that ADV_v is not a very good approximation for the total temperature tendency after the first few months. This is because, somewhat counter-intuitively, the horizontal advection term is non-negligible for most of the simulation. The issue is apparent in figure 6.14 (b) too, where the temperature evolution is not well approximated by the time-integrated vertical advection tendency (note the different scales on the vertical axis). Furthermore, it is not immediately obvious how to interpreter the term corresponding to horizontal advection physically.

Thus, while the ordinary temperature budget is a legitimate expression of the temperature equation, it is not a particularly effective one when it comes down to decoupling the contributions from mean flow and eddy fluxes. The problem, though, may be relaxed by considering the residual budget of figures 6.14 (c) and (d) instead. Here, vertical transport is dominant for the first year of simulation. Accordingly, the residual vertical advection tendency (dashed black line) approximately coincides with the total tendency (red circles). We may give an order of magnitude estimate for ADV_wres based on the residual formula $w_{\text{res}} = w_{ek} + w^*$. Assuming that the response is initially dominated by Ekman transport in the interior of the ocean, the anomalous temperature tendency equation for the initial stages of the evolution simplifies to:

$$\partial_t T'(t) = -w'_{ek}\partial_z \overline{T}^c, \tag{6.28}$$

where T' is the zonal-mean temperature anomaly, $\overline{T}^c$ is the control run time-mean, zonal-mean temperature, and w'_{ek} is the vertical Ekman velocity anomaly. Using the diagnosed values $w'_{ek} \approx 2.4 \cdot 10^{-6}$ m/s and $\partial_z \overline{T}^c \approx 0.62$ °C/km we obtain $-w'_{ek}\partial_z \overline{T}^c \approx -1.3 \cdot 10^{-4}$ °C/day, in broad agreement with the initial tendency shown in figure 6.14 (c). The subsequent modulations of the residual vertical term (and, two years after the wind step-change, of the residual horizontal term) are a manifestation of the spin up of baroclinic eddies, which act to disrupt the Deacon cell anomaly. Their presence does not strongly affect the nature of the balance, though, and we conclude that upwelling of cold water is the main driver of the temperature response in domain 1.

Figure 6.16 shows the same variables of figure 6.14, but for domain 2 in figure 6.8. The anomalous temperature tendency is positive during the first 16 months of simulation (as expected from the fact that domain 2 is located in a region of downward Ekman pumping), and weakly negative afterwards, panels (a) and (c): contrary to the previous case, thus, the temperature tendency is not in quasi-steady conditions. Consequently, the temperature anomaly evolution is non monotonic, panels (b) and (d): in fact, we observe that the anomaly overshoots its equilibrium value in about three and a half months, figure 6.15 (b), and it is only after the 16 months turnover time that temperature starts to fall back towards the equilibrium. The ordinary and residual frameworks convey markedly different physical pictures as to what processes drive the diagnosed tendency: in the ordinary framework, the tendency contribution from vertical advection (panel (a), dashed black line) is nearly constant and positive. Accordingly, the associated temperature evolution in panel (c) is a linear warming. Panel (a) shows that this warming tendency is partially compensated by horizontal advection, most likely dominated by eddy fluxes: therefore, the cross-over time of 16 months corresponds to the spin-up time of the horizontal advection term in the ordinary framework.

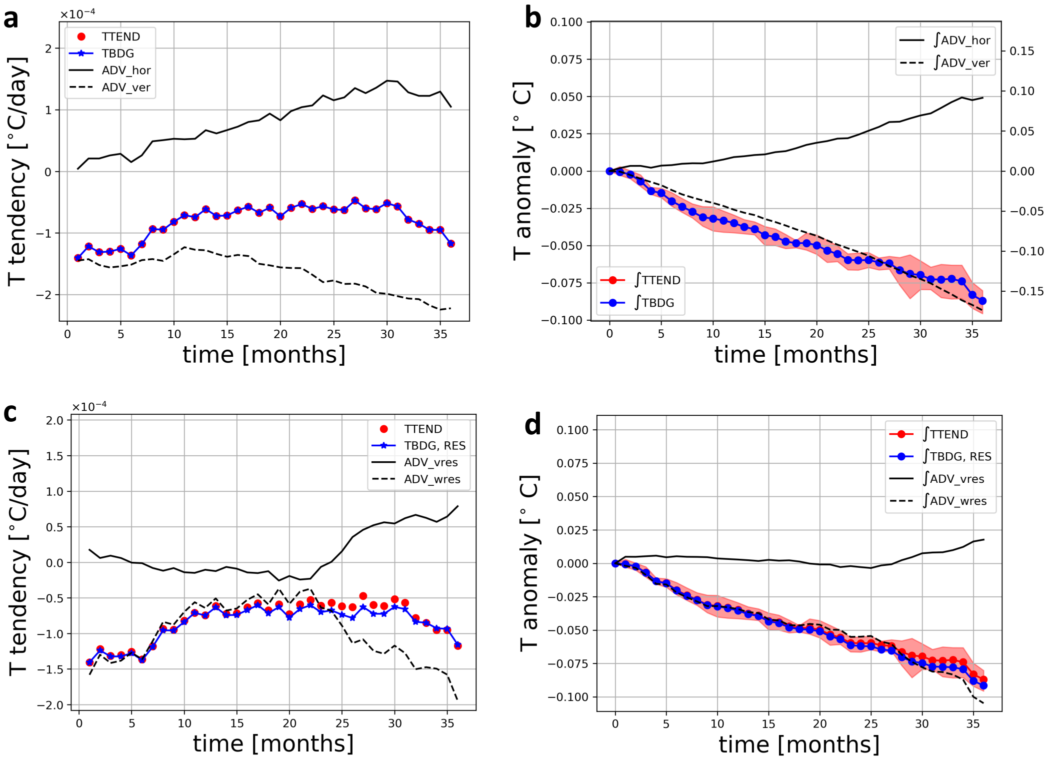

Figure 6.14: Spatially averaged temperature budget for domain 1 in figure 6.8. (a) and (c): temperature tendency (red dots) and anomalous budget components obtained with the ordinary and residual budget equations respectively. (b) and (d): time-integrated anomalous temperature tendency (continuous red line) and time-integrated anomalous budget components obtained with the ordinary and residual budget equations respectively. The red-shaded intervals mark the ensemble spread of the temperature anomaly, as quantified by the 84.1 and 15.9 percentiles. Note the different scale between the time-integrated total temperature (left hand side axis) and its components (right hand side axis) in panel (b). All quantities are ensemble mean, and all tendency fields are smoothed with a six-months rolling average.

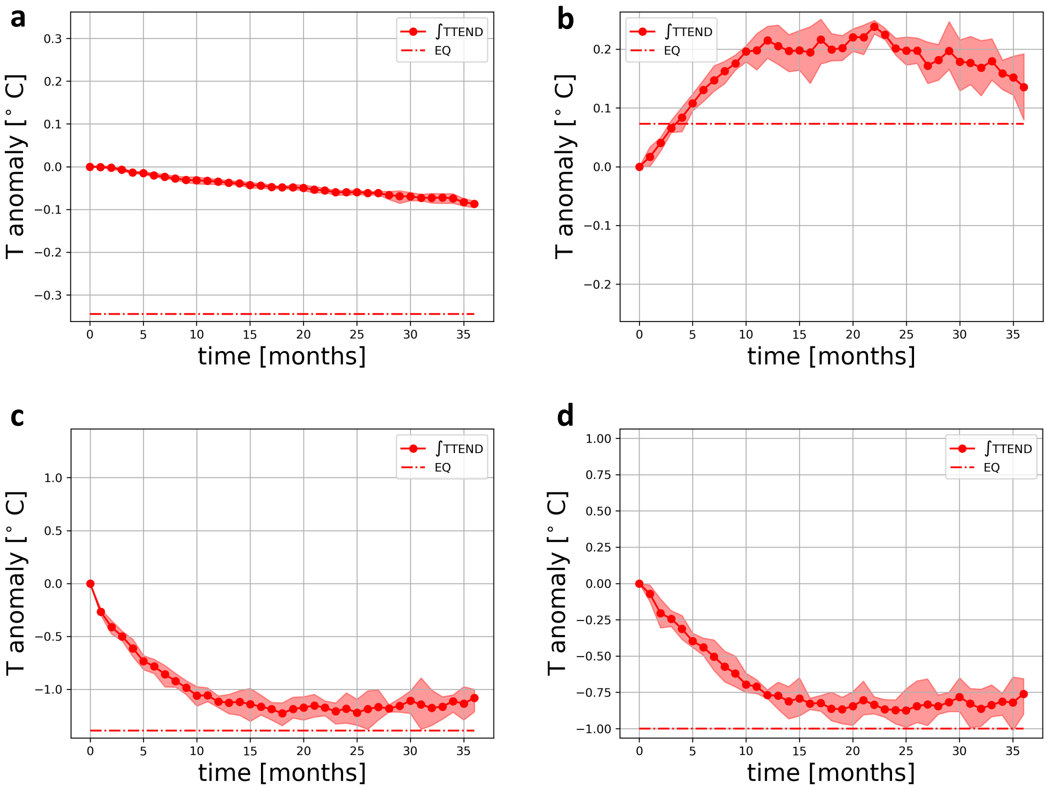

Figure 6.15: Time evolution of temperature anomaly (red dots and continuous red line) and equilibrium value of the temperature anomaly (dashed-dotted red line) for (a): domain 1, (b): domain 2, (c): domain 3, and (d): domain 4 in figure 6.8. The equilibrium temperature anomaly is averaged over the 15-years long climatology of the equilibrated perturbed channel.

As in the previous case, the physical interpretation of the temperature anomaly evolution is simpler in the residual framework, panels (b) and (d). Here, horizontal residual advection plays a minor role, contributing a weak positive tendency in the early stages of the adjustment only. The dominant contribution comes from vertical residual advection, and the cross-over time of about 16 months coincides with the time it takes for ADV_wres to change sign. The initial positive tendency, and the corresponding warming, are associated to residual downwelling acting upon a positive vertical gradient of temperature: assuming again that the initial temperature tendency is given by $\partial_t T'(t) = -w'_{ek}\partial_z \overline{T}^c$, and estimating $w'_{ek} \approx -3.5 \cdot 10^{-6}$ m/s and $\partial_z \overline{T}^c \approx 2.8$ °C/km, we obtain $-w'_{ek}\partial_z \overline{T}^c \approx 8.5 \cdot 10^{-4}$ °C/day, in broad agreement with figure 6.16 (c). The Ekman tendency is larger than for domain 1 because the ocean is more stratified close to the northern boundary.

It is more complicated to interpreter the subsequent migration to negative tendency values: assuming that the vertical temperature gradient is still dominated by the background temperature (i.e., we neglect non-linear effects in the interior), then the relative cooling observed after month 16 must be associated to residual upwelling of cold water. Inspection of figure 6.11 ($t = 24, 36$, black contours) confirms that the initial Deacon cell pattern is distorted in such a way that the dominant residual vertical flow is directed upward in domain 2, although it is not clear whether this effect would be observed with more ensemble members. It is remarkable that, for the time scales considered here, when eddies eventually intensify after the wind-stress perturbation their presence does not manifest itself as a simple damping effect on the initial Eulerian-like tendency, but in a more complicated form depending on the detailed structure of the meridional circulation's cells (rather than simply on their overall strength).

The analysis of the temperature budget for domain 3 is shown in figure 6.17: as explained in section 6.4.2, we only study the ordinary budget for the surface domains. Figure 6.17 (left) shows the anomalous tendency components: the initial diagnosed temperature tendency is negative and large (note the difference in scale between figure 6.17 (left) and figure 6.14 (a)). The tendency anomaly fades to ap-

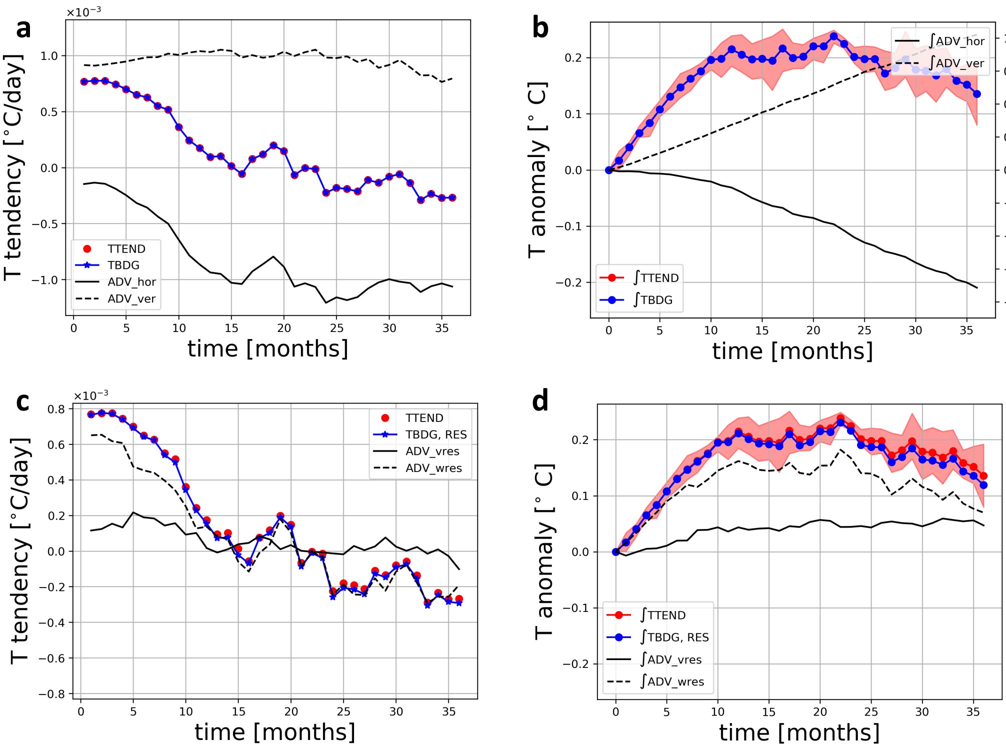

Figure 6.16: Same as in figure 6.14, but for domain 2 in figure 6.8.

proximately zero in a comparatively short time (about 20 months), suggesting that the surface tends to equilibrate faster than the interior of the channel. The time evolution of the temperature anomaly reflects that of the anomalous tendency: accordingly, domain 2 cools during the first 20 months following the perturbation, figure 6.17 (right), while its average temperature is rather stable afterwards. The process of adjustment approximately follows an exponential profile: however, after 36 months of simulation the temperature anomaly has not quite attained its equilibrium value (see figure 6.15 (c)), implying that the final part of the adjustment is not captured by our perturbation experiment.

In spite of the fact that the initial diagnosed temperature tendency is negative, the largest individual contribution to the temperature budget is positive, and comes from anomalous vertical advection, figure 6.17 (left). The negative contributions, which initially dominate over the vertical advection term, are supplied by anomalous horizontal advection and vertical mixing. There is no contribution from anomalous air-sea fluxes because the buoyancy flux at the surface is fixed, namely, it does not change from the control run to the perturbation experiment. Note that the negative mixing tendency, associated with cooling of domain 3, is larger in absolute value than the diagnosed tendency at all times: the implication is that the net effect of the advection terms is to warm the domain. The anomalous northward Ekman transport, however, acts on a positive meridional gradient of background temperature, and is therefore associated with a cooling tendency too. This suggests that the warming tendency necessary to close the budget must be supplied by eddies: we stress though that the horizontal and vertical advection terms include contributions from both the mean flow and the eddy fluxes in the ordinary framework, and is thus not straightforward to interpreter them physically in isolation as we did for the domains in the interior. The negative tendency from mixing and from horizontal advection are quasi-steady in the period of time considered, see figure 6.17 (left): therefore, the associated anomalous temperature evolution corresponds in both cases to a linear cooling, figure 6.17 (right). Interestingly, this implies that the temperature tendency anomaly equilibrates through increased warming via anomalous vertical advection, which is likely associated to the local spin-up of eddies counteracting the wind-induced circulation. Overall, the situation is rather different from that depicted in Ferreira et al. (2015), where the surface budget was characterised by a two-way balance between anomalous fluxes and horizontal advection, see for example their figure 8. Doddridge et al. (2019) primarily attribute the surface cold anomaly to anomalous horizontal advection too. The fixed *versus* interactive boundary condition is likely an important separating factor between our results and those of the cited studies.

Finally, figure 6.18 shows the temperature budget for domain 4 in figure 6.8. The initial diagnosed temperature tendency anomaly is negative, figure 6.18 (left), and, similarly to the surface case, it stabilises in approximately 20 months. Accordingly, the temperature anomaly evolution follows an approximately exponential decay, with typical time scale of about 20 months, see figure 6.18 (right). Three years into the simulation, however, the anomaly has not attained the equilibrium value (figure 6.15 (d)), therefore in the subsurface layer too three years of simulation are not

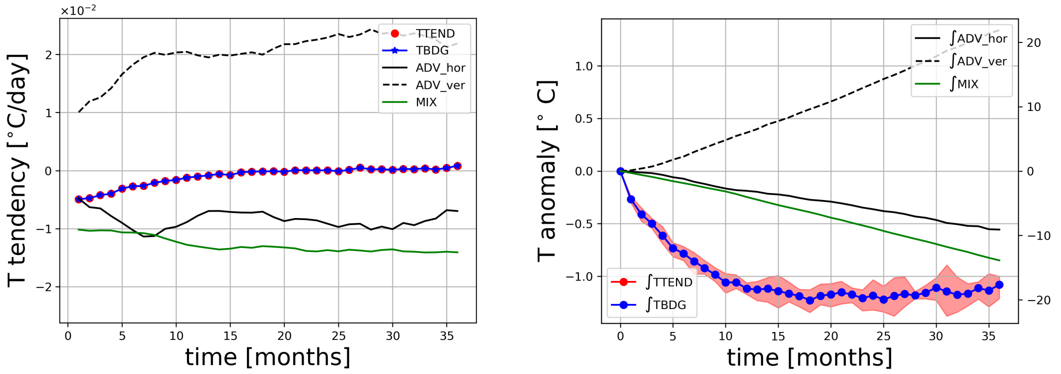

Figure 6.17: Same as in figure 6.14, but for domain 3 in figure 6.8. The ordinary temperature budget only is shown.

sufficient to capture the full temperature response. The main difference between this case and that of domain 3 is that the mixing term is now associated with a warming tendency, i.e., heat is extracted from the surface and communicated to the sub-surface layers via enhanced stirring. The result is in agreement with Doddridge et al. (2019), although in our case anomalous subsurface mixing does not actually induce subsurface warming. This is because horizontal advection acts to cool the domain, is the largest term in the balance, and dominates over the warming tendency driven by mixing and vertical advection. Similarly to the previous case, horizontal advection includes the contribution from northward Ekman transport of cold water, acting to cool domain 3. The adjustment of the horizontal and vertical advection terms, which eventually leads to an approximate stabilisation of the temperature tendency in about twenty months, is associated with the spin-up of eddies, which induces a re-organisation of the flow and counteracts the initial Eulerian response.

6.5 Summary and conclusions

In this chapter, I have studied the transient response of the Southern Ocean to an abrupt wind-stress perturbation with an eddy-resolving, idealised channel configuration of the MITgcm.

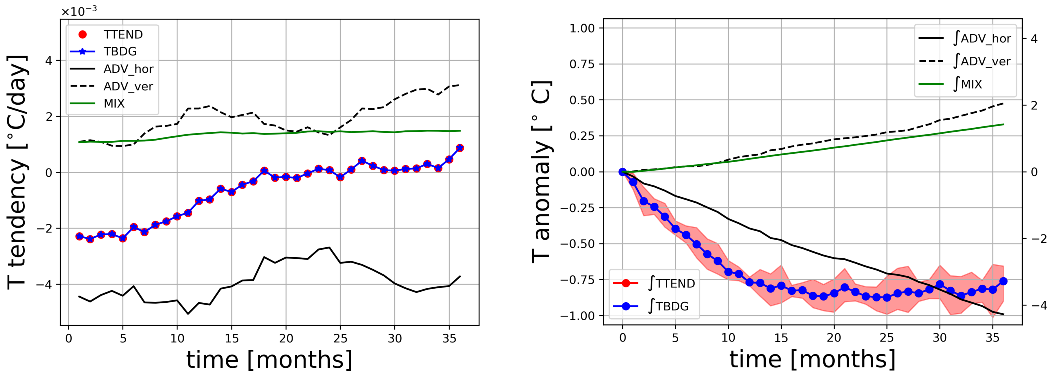

Figure 6.18: Same as in figure 6.14, but for domain 4 in figure 6.8. The ordinary temperature budget only is shown.

Ozone depletion over Antarctica acted to strengthen the winds driving the ACC over the last few decades. The transient response of the Southern Ocean to wind stress modulations may extend for up to tens of years (Kostov et al., 2017), and is thus key to understand how the climate of the Southern Ocean will evolve in the near future. A hypothesis under scrutiny is that trends in the Southern Hemisphere jet stream may explain the recently observed surface cooling in the region poleward of the ACC. However, most of the studies addressing the issue so far relied on numerical simulations performed with eddy-parametrising general circulation models, leaving the response of eddy-resolving models largely unexplored.

Here, I have investigated the time scales and the physical mechanisms driving the temperature and circulation response of the eddy-resolving, idealised channel configuration of the MITgcm. Specifically, I have run and analysed an ensemble of wind step-change experiments, where the forcing modulation corresponds to an instantaneous threefold increase of the wind stress parameter. The transient ensemble experiment was complemented by a climatology of the perturbed channel in the statistically equilibrated state attained following the wind stress perturbation. In order to determine exactly which processes drive temperature changes in the channel, I have diagnosed and studied the terms of the temperature budget equation. Also, I have presented an alternative formulation of the temperature budget which, in the interior of the ocean, allows to isolate the residual circulation's contributions to the temperature tendency equation. My results are:

1. The idealised channel is close to a state of complete eddy saturation, i.e., the equilibrium ACC circumpolar transport increases only modestly despite the large increase in wind stress. This result largely depends on the choice of the boundary conditions, and aligns well with previous research (Abernathey et al., 2011).

2. The channel is also close to a state of eddy compensation, as the equilibrium intensity of the MOC cells is weakly sensitive to the wind forcing. However, the structure of the residual cells on the meridional plane shows significant changes: notably, the intermediate cell re-positions beneath the diabatic layer.

3. The surface of the perturbed channel at equilibrium is cooler than in the reference state at all latitudes, with a difference locally larger than 1 °C. In the interior, a cold anomaly is found south of the main flow of the ACC, whereas a warm anomaly tongue can be found north of it.

4. The patterns of equilibrium temperature anomaly closely resemble those of residual circulation anomaly, suggesting that the two are related.

5. The residual budget technique is tested for the reference state of the channel within four averaging domains located in the regions of largest interior, surface and near-surface equilibrium temperature anomaly (there are two domains in the interior, capturing both the cold and the warm anomaly). The method allows to close the budget to a satisfactory accuracy in the interior, but fails near the surface. This is expected because the residual flux is non-zero in the vicinity of horizontal boundaries.

6. The transient adjustment of the residual circulation is initially dominated by the wind-induced anomaly: this pattern of anomalous circulation is reminiscent of the Deacon cell, with upwelling to the south and downwelling to the north. The circulation is closed by anomalous northward and southward Ekman flow in the surface and bottom frictional layers, respectively.

7. The typical time scales of circulation response are estimated by averaging the Eulerian, residual, and eddy-induced streamfunctions over a large domain.

The Eulerian circulation reacts very quickly to the wind stress perturbation, and is nearly equilibrated after 6 months. The adjustment of the eddy-induced circulation is considerably slower, and takes approximately 3 years. This result is in broad agreement with previous research (Meredith and Hogg, 2006) and, through $\psi_{\text{res}} = \overline{\psi} + \psi^*$, explains why the initial response of the residual streamfunction is dominated by the Eulerian anomaly.

8. Locally, though, the typical time scales of adjustment depart from the large scale average significantly. Both the temperature and circulation anomalies equilibrate faster close to the surface than in the interior. Large-scale measures of circulation response fail to capture this difference, and are unable to account for structural changes of the MOC cells that involve variations of their spatial location, but not of their intensity.

9. The cold and warm interior anomalies are primarily driven by, respectively, anomalous Ekman upwelling and downwelling acting on a positive vertical gradient of background temperature. However, the temperature evolution in the warm anomaly region north of the ACC is non-monotonic in the period of time analysed. This is attributed to a faster spin-up of eddies, which distort the initial Deacon cell pattern and prevent sustained Ekman pumping of warm surface water into the domain. Three years into the simulation, the temperature anomaly in the interior domains is far from the final equilibrium values.

10. The adjustment of the surface domain is fast, and the temperature anomaly approximately stabilises to its equilibrium value in about 20 months. Surprisingly, the cooling trend is supported primarily by enhanced vertical mixing, although horizontal advection also contributes negatively to the budget. Moreover, the cooling tendency due to vertical mixing is larger in absolute value than the total temperature tendency. Thus, despite the fact that northward Ekman transport acts to cool the domain, the combined effect of the advection terms is to warm the domain. This suggests that eddies spin-up very quickly to counteract the wind-induced anomaly within the surface domain.

11. The adjustment of the sub-surface domain is similar to the surface case, but the tendency components contribute to the balance in a different way. Here, the dominant term is horizontal advection, which acts to cool the domain. The mixing term is positive, implying that heat is transferred from the surface to the sub-surface layers through enhanced wind stirring.

Overall, the results presented in this chapter suggest that the spatial rearrangement of the MOC cells can have a large impact on the transient temperature evolution even if the channel is nearly eddy-compensated. In addition, the analysis above confirms that anomalous northward Ekman transport is critical to the adjustment of the surface layers: however, it also highlights that vertical mixing plays at least an equally important role. Doddridge et al. (2019) found that vertical mixing is the dominant term of the balance immediately below the mixed layer. Here, I have demonstrated that it is key to the surface response as well. Finally, the surface budget indicates that close to the surface the spin-up of eddies may be so fast as to be influential in the very initial stages of the adjustment. Indeed, given that the combined effect of the advection terms is to warm the domain, and that a purely Eulerian circulation would instead act to cool it, eddies must supply the warming tendency necessary to close the budget. Thus, the results presented here suggest that baroclinic eddies can not only alter the slow time-scale mechanism of adjustment, as found by Doddridge et al. (2019), but may even play an important role during the fast phase of the response.

A number of caveats apply. The idealised channel is configured based on several simplifying assumptions that do not apply to the real Southern Ocean. These include: flat bottom topography (and thus no standing meanders), no salinity, a linear equation of state, idealised mechanical and thermodynamical forcing. For this reason, the results presented here are not intended to provide a quantitative estimate of the real ocean. Nevertheless, the climatology of the idealised channel reproduces many aspects of the observed Southern Ocean accurately (chapter 3), which supports the view that this analysis is relevant to more comprehensive model configurations and to the real ocean. Importantly, the idealised channel does not include sea ice, and thus does not capture the region of temperature inversion below the seasonal sea ice. Therefore, I could not test whether upwelling of warm water from the temperature inversion induces surface warming within the transient perturbation experiments. The residual formulation of the temperature budget facilitates the physical interpretation of temperature changes. However, it does not allow to separate between the mean flow and eddy parts of the residual circulation explicitly and, critically, it is not valid near the surface because TEM theory breaks down at horizontal boundaries. A target for future work is thus to improve on the methods introduced in this manuscript by diagnosing the right hand side of equation (6.9), i.e. eddy and Ekman fluxes, directly. Finally, the fixed flow boundary conditions and the restoring sponge layer at the northern boundary impose severe constraints on the response of the channel to the wind stress perturbation. For example, by controlling the large scale slope of the isopycnals they also limit the circumpolar transport sensibility (via the thermal wind relation). This work informs the interpretation of higher-complexity general circulation model configurations for the Southern Ocean. A natural next step would be to evaluate the relative importance of vertical mixing, eddy fluxes, and mean flow contributions to the surface response of eddy-resolving models endowed with a region of temperature inversion. Determining how interactive *versus* fixed flow thermodynamical boundary conditions influence the process of transient adjustment is a further goal going forward.

Chapter 7

Conclusions

In this Thesis, I have investigated the role of mesoscale eddies in setting the time scales of natural and forced variability in the Southern Ocean.

Mesoscale eddies are generated primarily via baroclinic instability, and contribute to the dynamical balance of the Southern Ocean at leading order. Importantly, they communicate surface zonal momentum downwards, transfer heat polewards and, by countering the wind-induced steepening of the isopycnals, shape the residual circulation on the meridional plane: it is the residual MOC that advects tracers in the Southern Ocean, and provides a quasi-adiabatic upwelling pathway for deep water to reach the surface. Mesoscale eddies are also key to understand the response of the Southern Ocean to modulations of surface radiative (e.g., ozone thickness), thermodynamical (freshwater fluxes), or mechanical (wind stress) forcing, with numerous studies showing that coarse-resolution, eddy-parametrising GCMs disagree significantly with eddy-resolving models about the amplitude and the structure of the circulation response.

It is an open problem whether the observed decadal trends in the strength of the Southern Hemisphere jet stream can explain the observed weak surface warming (compared e.g. to the Arctic region. Poleward of the ACC, the surface of the Southern Ocean has actually cooled in recent decades) and induce sustained temperature, sea-ice, and heat and carbon absorption anomalies in the Southern Ocean. The question is compelling, as the Southern Ocean is an important regulator of global climate, and a prominent sink of anthropogenic heat and carbon. Nevertheless, evidence suggesting that changes in the ocean circulation can drive the observed trends continues to accrue (Doddridge et al., 2019, Gruber et al., 2019).

The current understanding of the equilibrated circulation response to wind stress changes hinges on the concepts of eddy saturation and eddy compensation. Eddy saturation refers to the weak (but not necessarily zero) sensitivity of the ACC circumpolar transport to wind changes. The definition of eddy compensation is less unambiguous, but the term is generally used to indicate that the MOC of the Southern Ocean scales sub-linearly with wind stress. While there is general consensus about the real Southern Ocean being close to a state of eddy saturation, the response of the residual overturning circulation varies significantly across models.

The transient adjustment of the Southern Ocean to wind stress changes is endowed with time scales ranging from years to decades (according to, for example, GCMs included in the CMIP5 suite), and is thus as important as the equilibrated response to understand current and future changes in the Antarctic region. Notably, Ferreira et al. (2015) proposed that the recently observed surface cooling could be driven by a fast phase of response similar to the dominant mode of SAM-SST interannual variability, with anomalous Ekman transport advecting cold water northwards and cooling the ACC region. However, sustained upwelling of warm water from below the seasonal sea ice could drive a slow phase of response, leading to long-term surface warming. This mechanism was tested by a number of studies, with contrasting results. Models, in particular, disagree about (i) whether or not the response of the Southern Ocean is actually endowed with a long-term warming phase, (ii) the time scales of the response, and (iii) whether or not the proposed mechanism can explain the observed trends. Crucially, though, most of the GCMs employed in these works parametrise eddies, with only a handful (Doddridge et al., 2019, Haumann et al., 2020) having enough resolution to explicitly resolve them. Doddridge et al. (2019), in particular, found that the residual overturning circulation compensates almost perfectly in approximately three years in their eddy-resolving model, potentially quenching the long-time scale warming phase of the response.

Thus, our ability to understand the present state of the Southern Ocean and to accurately predict its future evolution depends critically on the full comprehension of the physics of baroclinic eddies. Yet, the mechanisms that govern their interaction with the mean flow, and set the time scales and structure of natural and forced variability, are not fully understood. In this Thesis, I sought to put these mechanisms into better focus primarily by running and analysing numerical simulations of the Southern Ocean with an idealised, eddy-resolving channel configuration of the MITgcm.

The approach I pursued is twofold and, accordingly, the original material presented in this Thesis divides in two parts. In the first part (chapters 4 and 5), I concentrated on developing a simple mathematical model of eddy-mean flow interaction, which I tested against data from an idealised MITgcm configuration and a realistic state estimate of the Southern Ocean, the SOSE. In the second part (chapter 6), I studied the transient response of the idealised channel to an abrupt increase in wind stress, with an emphasis on the relationship between circulation and temperature changes. Since most of the analysis presented here is based on numerical simulations run with the idealised channel configuration of the MITgcm, the results of this Thesis are not intended to provide a quantitative description of the real Southern Ocean (although, whenever possible, I checked that the physics of the idealised channel agrees with observational estimates). The upside is that the idealised configuration allows to investigate the fundamental mechanisms governing the dynamics and time scales of mesoscale eddies with a minimum of complicating factors. Moreover, the model can be run at a high horizontal resolution, at a comparatively moderate computational cost: the first Rossby radius of deformation is resolved throughout the domain, and the idealised channel develops a vigorous eddy field with no need for an eddy parametrisation scheme. Therefore, although the simplified assumptions made to compile the idealised channel configuration hinder direct comparison with observations, the dynamics discussed here can be expected to be qualitatively relevant for higher-complexity GCM configurations and for the real ocean.

The mathematical model of unforced eddy-mean flow variability discussed in chapters 4 and 5 is inspired by the work of Ambaum and Novak (2014) (AN14) on atmospheric storm track variability, and motivated by the dynamical analogy between the ACC and the tropospheric jet stream (Thompson, 2008, Williams et al., 2007). AN14 and subsequent studies showed that the interaction between eddies and the mean flow in the atmosphere is characterised by predator-prey oscillatory dynamics, similar to those typical of population growth problems. In essence, eddies (the predator) feed on the available potential energy stored in the mean flow (the prey), and tend to intensify following an anomalous steepening of the isopycnals. Note that the slope of the isopycnal is a common measure of mean flow: when the isopycnals are steeper, more eddies are released by baroclinic instability. Conversely, when eddies are anomalously weak baroclinicity is replenished by diabatic forcing (or wind stress at the surface in the oceanic case), and the isopycnal slope increases. Here, I demonstrated that a similar approach can be applied to the natural oceanic equivalent of the atmospheric jet stream, the ACC.

Firstly, I identified a pair of spatially-averaged dynamical variables in order to represent the intensity of the mean flow and of the eddies: following previous studies (Novak et al., 2017), I employed the Eady growth rate and the eddy buoyancy flux, respectively. The two variables were averaged in a domain of meridional width comparable with that of an individual ACC jet, located in the upper interior and at intermediate latitudes, in the case of the idealised channel (chapter 4). The results, however, are not critically dependant on the precise choice of the averaging domain. In the case of SOSE (chapter 5), I considered a higher number of domains tiling the ACC region in order to capture regional differences in the dynamics. Also, streamwise averages were preferred to zonal averages to better capture the meandering nature of the ACC flow in SOSE. The bivariate time series thus obtained were characterised by means of a small number of statistical indicators: specifically, the correlation functions, kernel-averaged phase space diagrams, and marginal probability distribution functions were computed. Construction of the kernel-averaged phase space trajectories, in particular, helps mitigate the effect of random fluctuations on the evolution of the state variables, and thus allows one to probe the dynamical coupling between the two variables even in the presence of noise (Novak et al., 2017). Despite the fact that the raw time series fluctuate randomly (and thus do not define oscillators in the common deterministic sense), the phase space trajectories associated with both the idealised channel and, locally, SOSE, are quasi periodic orbits. The sense of circulation of the orbits is such that a complete circuit can be interpreted as a full predator-prey life cycle, suggesting that the mean flow and the eddies mutually interact according to the oscillatory picture of AN14. This conclusion was corroborated by examination of the cross-correlation functions which, in both cases, are approximately odd functions of the time lag with statistically significant peaks. The structure of the cross-correlation agrees with the physical interpretation of the dynamics. In the case of SOSE, not all the spatial domains I considered display evidence of predator-prey oscillations. Those which do, however, tend to cluster around the Sub Antarctic Front of the ACC, where the flow is strongest, lending further support to the idea that the oscillatory dynamics are physically linked to the presence of intense zonal jets. Since SOSE's time series are only 6-years long, I combined the data from these regions to reduce the level of noise in the time series and facilitate their analysis.

The hypothesis was tested that the bivariate time series can be described by a two-dimensional dynamical system with predator-prey dynamics. To this aim, I adapted the original model of AN14 by linearising it (so as to retain analytical tractability) and by including a stochastic forcing term, which allows to account for the data's fluctuations explicitly. The model so obtained (named simplified model) is a stochastic linear damped oscillator or, in the discrete, a bivariate auto-regressive process of order one. The simplified model was fitted to the data with the Yule-Walker method (chapter 4) and with ordinary linear regression (chapter 5). The fit was successful in both cases, and the simplified model could accurately predict the overall structure of the statistical indicators considered (although not all the quantitative details, indicating that more complex models may achieve better results). The best fit parameters were interpreted in terms of typical time scales, which were found to range from a few weeks to a couple of months. In particular, the analysis revealed that eddies intensify following a steepening of the isopycnals with a lag of $15-20$ days. This time scale describes the *natural* variability of eddies and mean flow, and does not necessarily coincide with the time scales of forced response to wind stress changes (see comments below). Importantly, the time scales associated to the best fit parameters were comparable in the case of the idealised channel and of SOSE. Given the stark difference between the two models' configurations, there is reason to conclude that the simplified model robustly captures a fundamental mechanism of natural eddy-mean flow variability. Thus, although both GCMs represent complex, three-dimensional geophysical system and comprise a high number of degrees of freedom, the first part of this thesis demonstrates that as few as two stochastic oscillatory equations are needed to effectively capture the important dynamics of eddy-mean flow interaction. Finally, the conclusions drawn in chapters 4 and 5 were supported by testing the fit procedure against synthetic realisations of the simplified model. The numerical experiments confirmed that quasi-periodic phase space trajectories can emerge even when the underlying dynamics are dissipative, due to the combined effect of stochastic and deterministic contributions. This result is relevant to the atmospheric case as well, where the phase space diagram is emerging as a powerful diagnostic tool (Yano et al., 2020).

The mechanisms and time scales of forced response to wind stress changes are investigated in the second part of this Thesis' original material, chapter 6. Specifically, the chapter is dedicated to the analysis of wind step-change experiments performed with the idealised channel configuration of the MITgcm. Due to its relevance to the Antarctic and global climate in future decades, the transient response of the Southern Ocean to wind stress changes has been extensively investigated. Abrupt perturbation experiments, in particular, have attracted considerable attention because they are conceptually simple, and linear theory allows to extend the results to the case of arbitrary time modulations of the forcing. Critically, though, only a handful of the studies addressing the problem so far employed eddy-resolving models, which leaves the effect of baroclinic eddies on the mechanisms of adjustments largely unexplored. Here, I sought to further the current understanding of the subject by realising an ensemble of wind step-change experiments with the eddy-resolving idealised channel, where the perturbation consists in an instantaneous threefold increase of the wind stress parameter τ_0. The ensemble members are branched from the unperturbed, equilibrated state of the channel (named control run), and the initial conditions were chosen so as to reasonably guarantee that members evolve independently. Although the time scales of equilibration of idealised models are generally faster than those of comprehensive GCMs (compare for example Seviour et al. (2016) and Doddridge et al. (2019)), six years are not sufficient for the idealised channel to reach a statistically equilibrated state, particularly in the interior. For this reason, I complemented the ensemble perturbation experiments with a climatology obtained by running a single realisation of the idealised channel subject to the wind perturbation to equilibrium.

The first step of the analysis consisted in the study of the equilibrated state of the perturbed channel. I found that the baroclinic circumpolar transport and the strength of the MOC are weakly dependant on wind stress, i.e. that the channel is close to a state of both eddy saturation and eddy compensation: this result agrees well with what found by previous studies adopting a similar GCM configuration (Abernathey et al., 2011). Importantly though, I also observed that, while the volume transport associated with the individual MOC cells does not change significantly, their structure and spatial arrangement varies greatly with respect to the control run. This means that local residual streamfunction anomalies are large even though the channel is in a nearly compensated state. A notable example is the intermediate, clockwise rotating cell, which is essentially confined to the surface diabatic layer in the control run, but reaches well below it in the perturbed state. The implication is that the overall strength of the MOC does not characterise the perturbed state completely, and structural changes should also be considered. The conclusion is corroborated upon scrutiny of temperature anomalies on the meridional plane, which organise in a way that closely resembles the patterns of residual circulation changes. In particular, I found that: (i) the surface of the ocean cools at all latitudes, (ii) the interior cools in the southern part of the domain, and (iii) a warm tongue can be found in the interior northern region, approximately co-located with the downward-shifted intermediated cell of the MOC. Note that the idealised channel does not include sea ice, therefore upwelling of warm water from below the seasonal ice plays no role here.

To study the relationship between circulation and temperature changes in the ensemble experiment, I have diagnosed the terms of the temperature budget, which allows to determine exactly which processes drive the temperature tendency at a specified location. In addition, I have also shown how an approximate temperature budget can be formulated which expresses the advection tendency in terms of the residual circulation, thereby greatly facilitating the physical interpretation of the results. This method was found to work remarkably well in the ocean interior, but not close to the surface (as expected from limitations inherent to TEM theory). Analysis of the temperature budget revealed that temperature changes in the interior are largely driven by anomalous vertical flows, specifically, upwelling in the southern part of the domain (associated with cooling) and downwelling in the northern part (warming). This initial pattern of anomalous circulation resembles the Deacon cell and is consistent with the results of previous modelling studies (Ferreira et al., 2015). Interestingly, the cooling rate remains more or less constant in the southern half of the domain three years after the perturbation, whereas an inversion of tendency is observed in the northern region one year and a half into the simulation. The difference is attributed to regional variations in the time scales of circulation response, i.e. the residual streamfunction adjusts faster close to the northern boundary of the domain. While this is probably at least partly due to the specifics of the boundary conditions, the result underscores the importance of local modulations of the circulation response: by way of comparison, a domain-scale average of the residual streamfunction equilibrates in approximately three year, which does not reflect neither of the time scales of temperature adjustments in the two interior regions considered. Here, the three-years equilibration time scale corresponds to the spin-up of the eddy field following the wind stress perturbation. This estimate is in overall agreement with results from previous research (Doddridge et al., 2019, Meredith and Hogg, 2006). The wind-induced circulation, instead, reacts on much faster scales to the perturbation (six months approximately), so that the initial circulation response is dominated by the Deacon cell pattern, as noted above. Finally, the cold surface anomaly was found to be primarily driven by enhanced vertical mixing, with horizontal advection (associated with anomalous Ekman transport) supplementing the cooling tendency. It equilibrates in about 20 months due to a re-organisation of the circulation linked with the kick-off of baroclinic eddies. This initial surface cooling response to the wind perturbation is a robust feature of GCMs (Ferreira et al., 2015, Kostov et al., 2017), but is normally attributed to anomalous northward Ekman transport only. The results presented in this Thesis, instead, suggest that vertical mixing plays a greater role than previously thought, which could inform experiments run with more realistic model configurations.

Overall, the two parts of this Thesis's results complement each other and, collectively, provide a characterisation of the time scales of both the natural and forced variability of mesoscale eddies in the Southern Ocean. The results, naturally, come with a number of caveats, including:

1. The idealised channel configuration is based on simplifying assumptions that are not valid in the real ocean, and many important processes are not represented. Therefore, the time scales I identified in this manuscript cannot be expected to closely reflect those of the actual Southern Ocean.
2. The fixed-flow buoyancy boundary conditions at the surface and the restoring sponge layer at the northern boundary, in particular, impose severe constraints on the response of the channel to wind stress changes (for example, they enforce a state of near eddy saturation by fixing the large-scale isopycnal slope).
3. The idealised channel does not include sea ice, and thus the sub-surface region

of temperature inversion. Therefore, it is not possible to test whether or not upwelling of warm subsurface water into the mixed layer induces long-term warming

4. The simplified model captures the qualitative statistical features of the data well, but not all the quantitative details. Furthermore, the six-years time series of SOSE are too noisy to be fitted regionally, which means that part of the information regarding regional variations of the dynamics is lost when the data were aggregated.

5. The dependency of the simplified model's coefficients on the physical parameters is not known, which limits one's ability to use the simplified model to predict the time scales of the forced response to wind changes.

This work, I believe, leaves a number of interesting options open, which in some cases connect directly with the limitations listed above. Firstly, the simplified model expands the framework introduced by AN14 to study eddy-mean flow variability in the atmosphere. A natural follow-up of this study would thus be to fit the simplified model to atmospheric storm-track data, and compare the results with the oceanic case. A major (dynamical) difference between the ACC system and the tropospheric jet stream is their scale separation (i.e. the ratio between the typical spatial scale of the zonal flow and the Rossby radius of deformation is larger in the ocean, Williams et al. (2007)) and it is intriguing to ask whether the simplified model has enough structure to capture it. A second, attractive research avenue is that it may be possible to explain the forced time scales of the transient response to wind changes based on those of natural variability. This idea has provenance in climate science (Gritsun and Branstator, 2007), and is theoretically rooted in the fluctuation-dissipation theorem (Breul et al., 2022). In the context of this Thesis, this would mean attempting to explain the time scales of eddy response observed in the ensemble perturbation experiments based on the dynamics of the simplified model: for example, by running a simplified version of the step change experiments with the simplified model itself. The major obstacle in this sense is that the dependency of the simplified model's coefficients on physical parameters such as wind stress is not known (point 5 above), which makes it difficult to represent the applied perturbation within the simplified model in a consistent way. There are at least two options to circumvent the problem which can be envisioned at this stage. The first is to infer the simplified model's scaling by running sensitivity experiments with the idealised channel, i.e., by fitting the simplified model to equilibrated states of the channel with different values of the wind stress parameter. This approach has the merit of simplicity, but is computationally expensive and may lead to overlook the dependency of the model's coefficients on other physical parameters (e.g. bottom drag, buoyancy forcing at the surface, etc.). The second, more radical, option is to formulate a mathematical model of eddy-mean flow variability from successive approximations of the equations of motion rather than based on empirical arguments. The advantage of this approach is that the scaling of the model's coefficients with the physical parameters would be obtained naturally from the derivation of the model itself. On the other hand, analytical tractability may impose severe restrictions on the realism of the theoretical setup, so that it is not clear to what extent such a model would afford comparison with a GCM such as the idealised channel. Nevertheless, inroads have recently been made towards formulating a model that describes the interplay between eddies and mean flow from first principles in the atmosphere (Kobras et al., 2022), so that it would certainly be of interest to explore the feasibility of this strategy for the oceanic case as well. By demonstrating the value of simplified mathematical models of eddy variability, this work may pave the way to further investigation aimed at bridging the gap between the time scales of natural and forced variability.

The Southern Ocean is a key regulator of global climate, and enhancing the comprehension of the physical processes that drive its evolution under climate change is a research priority and a fascinating challenge. This Thesis is dedicated to the dynamics of mesoscale eddies: by modelling their time scales of natural variability, and by bringing their relationship with the transient circulation and temperature response to wind stress modulations into clearer focus, it contributes to the research efforts devoted to this compelling scientific task.

Bibliography

Abernathey, R. and Ferreira, D. (2015). Southern Ocean isopycnal mixing and ventilation changes driven by winds. *Geophys. Res. Lett.*, 42:10,357–10,365.

Abernathey, R., Marshall, J., and Ferreira, D. (2011). The dependence of Southern Ocean meridional overturning on wind stress. *J. Phys. Oceanogr.*, 41:2261–2278.

Adcroft, A., Campin, J.-M., Doddridge, E., Dutkiewicz, S., Evangelinos, C., Ferreira, D., Follows, M., Forget, G., Fox-Kemper, B., Heimbach, P., Hill, C., Hill, E., Hill, H., Jahn, O., Klymak, J., Losch, M., Marshall, J., Maze, G., Mazloff, M., Menemenlis, D., Molod, A., and Scott, J. (2022). Mitgcm's user manual.

Ambaum, M. and Novak, L. (2014). A nonlinear oscillator describing storm track variability. *Q. J. R. Metereol. Soc.*, 140:2680–2684.

Bishop, S., Gent, P., Bryan, F., Thompson, A., Long, M., and Abernathey, R. (2016). Southern Ocean Overturning Compensation in an Eddy-Resolving Climate Simulation. *J. Phys. Oceanogr.*, 46:1575–1592.

Bitz, C. M. and Polvani, L. M. (2012). Antarctic climate response to stratospheric ozone depletion in a fine resolution ocean climate model. *Geophys. Res. Lett.*, 39:L20705.

Böning, C. W., Dispert, A., Visbeck, M., Rintoul, S. R., and Schwarzkopf, F. U. (2008). The response of the Antarctic Circumpolar Current to recent climate change. *Nature Geoscience*, 1:864–869.

Bretherton, C., Widmann, M., Dymnikov, V. P., Wallace, J., and Blade, I. (1999).

The effective number of spatial degrees of freedom of a time-varying field. *J. Climate*, 12:1990–2009.

Breul, P., Ceppi, P., and Shepherd, T. G. (2022). Relationship between southern hemispheric jet variability and forced response: the role of the stratosphere. *Weather Clim. Dynam.*, 3:645–658.

Cerovečki, I., Meijers, A. J. S., Mazloff, M. R., Gille, S. T., Tamsitt, V. M., and Holland, P. R. (2019). The effects of enhanced sea ice export from the Ross Sea on recent cooling and freshening of the southeast Pacific. *Journal of Climate*, 32:2013–2035.

Cessi, P., Young, W. R., and Polton, J. A. (2006). Control of large-scale heat transport by small-scale mixing. *J. Phys. Oceanogr.*, 36:1877–1894.

Charney, J. G. (1947). The dynamics of long waves in a baroclinic westerly current. *J. Meteor.*, 4:135–162.

Chelton, D. B., deSzoeke, R. A., Schlax, M. G., Naggar, K. E. I., and Siwertz, N. (1998). Geographical variability of the first baroclinic Rossby radius of deformation. *J. Phys. Oceanogr.*, 28:433–460.

Ciasto, L. M. and Thompson, D. W. (2008). Observations of large-scale ocean-atmosphere interaction in the Southern Hemisphere. *J. Climate*, 21:1244–1259.

Damerell, G., Heywood, K. J., Calvert, D., Grant, A. L. M., Bell, M. J., and Belcher, S. E. (2020). A comparison of five surface mixed layer models with a year of observations in the North Atlantic. *Progress in Oceanography*, 187.

Doddridge, E. W., Marshall, J., Song, H., Campin, J.-M., Kelley, M., and Nazarenko, L. S. (2019). Eddy compensation dampens Southern Ocean sea surface temperature response to westerly wind trends. *Geophysical Research Letters*, 46:4365–4377.

Doddridge, E. W. and Marshall, J. C. (2017). Modulation of the seasonal cycle of Antarctic sea ice extent related to the Southern Annular Mode. *Geophys. Res. Lett.*, 44:9761–9768.

Doos, K. and Webb, D. (1994). The Deacon cell and other meridional cells of the Southern Ocean. *J. Phys. Oceanogr.*, 24:429–442.

Durack, P. J. and Wijffels, S. E. (2010). Fifty-year trends in global ocean salinities and their relationship to broad-scale warming. *J. Clim.*, 23:4342–4362.

Eady, E. (1949). Long waves and cyclone waves. *Tellus*, 1:33–52.

Eden, C. and Greatbatch, R. J. (2008). Towards a mesoscale eddy closure. *Ocean Modelling*, 20:223–239.

Fan, T., Deser, C., and Schneider, D. P. (2014). Recent Antarctic sea ice trends in the context of Southern Ocean surface climate variations since 1950. *Geophys. Res. Lett.*, 41:1206–1226.

Farneti, R., Delworth, T. L., Rosati, A. J., Griffies, S. M., and Zeng, F. (2010). The role of mesoscale eddies in the rectification of the Southern Ocean response to climate change. *J. Phys. Oceanogr.*, 40:1539–1557.

Ferrari, R. and Wunsch, C. (2009). Ocean circulation kinetic energy: Reservoirs, sources, and sinks. *Annu. Rev. Fluid Mech.*, 41:253–282.

Ferreira, D., Marshall, J., Bitz, C. M., Solomon, S., and Plumb, A. (2015). Antarctic ocean and sea ice response to ozone depletion: A two-time-scale problem. *J. Clim.*, 28:1206–1226.

Ferreira, D., Marshall, J., and Heimbach, P. (2005). Estimating eddy stresses by fitting dynamics to observations using a residual-mean ocean circulation model and its adjoint. *J. Phys. Oceanogr.*, 35:1891–1910.

Frankignoul, C. and Hasselmann, K. (1977). Stochastic climate models, Part II Application to sea-surface temperature anomalies and thermocline variability. *Tellus*, 29:289–305.

Frenger, I. M., Münnich, M., Gruber, N., and Knutti, R. (2015). Southern Ocean eddy phenomenology. *J. Geophys. Res. Oceans*, 120:7413–7449.

Frölicher, T. L., Sarmiento, J. L., Paynter, D. J., Dunne, J. P., Krasting, J. P., and Winton, M. (2015). Dominance of the Southern Ocean in anthropogenic carbon and heat uptake in CMIP5 models. *Journal of Climate*, 28:862–886.

Fyfe, J. C. and Saenko, O. A. (2006). Simulated changes in the extratropical Southern Hemisphere winds and currents. *Geophys. Res. Lett.*, 33:L06701.

Gille, S. (2008). Decadal-scale temperature trends in the Southern Hemisphere ocean. *J. Clim.*, 21:2749–2765.

Gnanadesikan, A., Speller, C. M., G., R., San Soucie, J., Thomas, J., and Pradal, M.-A. (2020). Feedbacks driving interdecadal variability in Southern Ocean convection in climate models: A coupled oscillator mechanism. *J. Phys. Oceanogr.*

Gritsun, A. and Branstator, G. (2007). Climate response using a three-dimensional operator based on the fluctuation-dissipation theorem. *J. Atmos. Sci.*, 64:2558–2575.

Gruber, N., Landschutzer, P., and Lovenduski, N. P. (2019). The variable Southern Ocean carbon sink. *Annu. Rev. Mar. Sci.*, 11:159–186.

Hallberg, R. and Gnanadesikan, A. (2006). The role of eddies in determining the structure and response of the wind-driven Southern Hemisphere overturning: Results from the modeling eddies in the Southern Ocean (MESO) project. *J. Phys. Oceanogr.*, 36:2232–2252.

Hasselmann, K., Sausen, R., Maier-Reimer, E., and Voss, R. (1993). On the cold start problem in transient simulations with coupled atmosphere-ocean models. *Clim. Dyn.*, 9:53–61.

Haumann, F. A., Gruber, N., and Münnich, M. (2020). Sea-ice induced Southern Ocean subsurface warming and surface cooling in a warming climate. *AGU Advances*, 1:e2019AV000132.

Haumann, F. A., Gruber, N., Münnich, M., Frenger, I., and Kern, S. (2016). Sea-ice transport driving Southern Ocean salinity and its recent trends. *Nature*, 537 (7618):89–92.

Hibler, W. D. I. (1980). Modeling a variable thickness sea ice cover. *Mon. Wea. Rev.*, 108:1943–1973.

Higham, D. J. (2001). An algorithmic introduction to numerical simulation of stochastic differential equations. *SIAM Review*, 43(3):525–546.

Hill, C., Ferreira, D., Campin, J., Marshall, J., Abernathey, R., and Barrier, N. (2012). Controlling spurious diapycnal mixing in eddy-resolving height-coordinate ocean models: Insights from virtual deliberate tracer release experiments. *Ocean Modelling*, 45-46:14–26.

Hogg, A. M. C. and Blundell, J. R. (2006). Interdecadal variability of the Southern Ocean. *J. Phys. Oceanogr.*, 36(8):1626–1645.

Hogg, A. M. C., Meredith, M. P., Blundell, J. R., and Wilson, C. (2008). Eddy heat flux in the Southern Ocean: Response to variable wind forcing. *J. Climate*, 21:608–620.

Hogg, A. M. C., Meredith, M. P., Chambers, D. P., Abrahamsen, E. P., Hughes, C. W., and Morrison, A. K. (2015). Recent trends in the Southern Ocean eddy field. *J. Geophys. Res. Oceans*, 120:257–267.

Hughes, C. W. and Ash, E. R. (2001). Eddy forcing of the mean flow in the Southern Ocean. *Journal of Geophysical Research*, 106 (C2):2713–2722.

Jansen, M. and Ferrari, R. (2012). Macroturbulent equilibration in a thermally forced primitive equation system. *Journal of the Atmospheric Sciences*, 69:695–713.

Karsten, R. and Marshall, J. (2002). Constructing the residual circulation of the ACC from observations. *J. Phys. Oceanogr.*, 32:3315–3327.

Kobras, M., Ambaum, M. H. P., and Lucarini, V. (2022). Eddy saturation in a reduced two-level model of the atmosphere. *Geophysical and Astrophysical Fluid Dynamics*, 116:1:38–55.

Kostov, Y., Ferreira, D., Armour, K. C., and Marshall, J. (2018). Contributions of greenhouse gas forcing and the southern annular mode to historical Southern Ocean surface temperature trends. *Geophys. Res. Lett.*, 45:1086–1097.

Kostov, Y., Marshall, J., Hausmann, U., Armour, K. C., Ferreira, D., and Holland, M. (2017). Fast and slow responses of Southern Ocean sea surface temperature to SAM in coupled climate models. *Climate dynamics*, 48:1595–1609.

Large, W. G., Mcwilliams, J. C., and Doney, S. C. (1994). Oceanic vertical mixing - a review and a model with a nonlocal boundary-layer parameterization. *Rev. Geophys*, 32:363–403.

Large, W. G. and Yeager, S. G. (2004). Diurnal to decadal global forcing for ocean and sea-ice models: The data sets and flux climatologies. *NCAR Tech. Rep.*, TN-4601STR:112 pp.

Le Bars, D., Viebahn, J. P., and Dijkstra, H. P. (2016). A Southern Ocean mode of multidecadal variability. *Geophys. Res. Lett.*, 43:2102–2110,.

Ledwell, J. R., Watson, A. J., and Law, C. S. (1998). Mixing of a tracer in the pycnocline. *J. Geophys. Res.*, 103:21499–21529.

Lembo, V., Lucarini, V., and Ragone, F. (2020). Beyond forcing scenarios: Predicting climate change through response operators in a coupled general circulation model. *Sci Rep*, 10:8868.

Leveque, R. J. (2002). *Finite volume methods for hyperbolic problems*. Cambridge university press.

Lorenz, E. (1963). Deterministic nonperiodic flow. *J. Atmos. Sci.*, 20:130–141.

Lucarini, V. (2018). Revising and extending the linear response theory for statistical mechanical systems: Evaluating observables as predictors and predictands. *Journal of Statistical Physics*, 173 (6):1698–1721.

Marcheggiani, A., Ambaum, M. H. P., and Messori, G. (2022). The life cycle of meridional heat flux peaks. *Q. J. R. Metereol. Soc.*

Marshall, D. P., Ambaum, M. H. P., Maddison, J. R., Munday, D. R., and Novak, L. (2017). Eddy saturation and frictional control of the Antarctic Circumpolar Current. *Geophys. Res. Lett.*, 44:286–292.

Marshall, J., Adcroft, A., Hill, C., Perelman, L., and Heisey, C. (1997a). A finite-volume, incompressible Navier Stokes model for studies of the ocean on parallel computers. *J. Geophys. Res.*, 102:5753–5766.

Marshall, J., Armour, K. C., Scott, J. R., Kostov, Y., Hausmann, U., Ferreira, D., Shepherd, T. G., and Bitz, C. M. (2014). The ocean's role in polar climate change: Asymmetric Arctic and Antarctic responses to greenhouse gas and ozone forcing. *Philos. Trans. Roy. Soc. London*, A372.

Marshall, J., Hill, C., Perelman, L., and Adcroft, A. (1997b). Hydrostatic, quasi-hysdrostatic, and non-hydrostatic ocean modeling. *J. Geophys. Res.*, 102:5733–5752.

Marshall, J. and Plumb, R. A. (2008). *Atmosphere, Ocean and Climate Dynamics*. Elsevier.

Marshall, J. and Radko, T. (2003). Residual-mean solutions for the Antarctic Circumpolar Current and its associated overturning circulation. *J. Phys. Oceanogr.*, 33:2341–2354.

Marshall, J. and Radko, T. (2006). A model of the upper branch of the meridional overturning of the Southern Ocean. *Progr. Oceanogr.*, 70:331–345.

Marshall, J. and Speer, K. (2012). Closure of the meridional overturning circulation through Southern Ocean upwelling. *Nature Geoscience*, 5:171–180.

Mazloff, M., Heimbach, P., and Wunsch, C. (2010). An Eddy-Permitting Southern Ocean State Estimate. *J. Phys. Oceanogr.*, 40:880–899.

McIntosh, P. C. and McDougall, T. J. (1996). Isopycnal averaging and the residual mean circulation. *J. Phys. Oceanogr.*, 26:1655–1660.

Meehl, G. A., Arblaster, J. M., Chung, C. T. Y., Holland, M. M., DuVivier, A., Thompson, L., Yang, D., and Bitz, C. (2019). Sustained ocean changes contributed to sudden Antarctic sea ice retreat in late 2016. *Nature Communications*, 10 (14).

Meredith, M. P. and Hogg, A. M. (2006). Circumpolar response of Southern Ocean eddy activity to a change in the Southern Annular Mode. *Geophys. Res. Lett.*, 33:L16608.

Meredith, M. P., Woodworth, P., L., Hughes, C. W., and Stepanov, V. (2004). Changes in the ocean transport through Drake Passage during the 1980s and 1990s, forced by changes in the Southern Annular Mode. *Geophys. Res. Lett.*, 31:L21305.

Moreton, S., Ferreira, D., Hewitt, H., and Roberts, M. (2020). Evaluating surface eddy properties in climate simulations with 'eddy-present' and 'eddy-rich' ocean resolution. *Ocean modelling*, 147:101567.

Morrison, A. K. and Hogg, A. M. (2013). On the relationship between Southern Ocean overturning and ACC transport. *J. Phys. Oceanogr*, 43:140–148.

Munday, D. R., Johnson, H. L., and Marshall, D. P. (2013). Eddy saturation of equilibrated circumpolar currents. *J. Phys. Oceanogr*, 43:507–532.

Novak, L., Ambaum, M. H. P., and Tailleux, R. (2017). Marginal stability and predator-prey behaviour within storm tracks. *Q. J. R. Metereol. Soc.*, 143:1421–1433.

Olbers, D., Borowski, D., Völker, C., and Wölff, J.-O. (2004). The dynamical balance, transport and circulation of the Antarctic Circumpolar Current. *Antarctic Science*, 16:439–470.

Orsi, A. H., Whitworth III, T., and Nowlin Jr., W. D. (1995). On the meridional extent and fronts of the Antarctic Circumpolar Current. *Deep-Sea Res.*, 42:641–673.

Parkinson, C. L. (2019). A 40-y record reveals gradual Antarctic sea ice increases followed by decreases at rates far exceeding the rates seen in the arctic. *Proceedings on the National Academy of Sciences*, 116 (29):14414–14423.

Patara, L., Böning, C. W., and Biastoch, A. (2016). Variability and trends in Southern Ocean eddy activity in 1/12 ocean model simulations. *Geophys. Res. Lett.*, 43:4517–4523.

Patmore, R. D., Holland, P. R., Munday, D. R., Naveira Garabato, A. C., Stevens, D. P., and Meredith, M. P. (2019). Topographic Control of Southern Ocean Gyres and the Antarctic Circumpolar Current: A Barotropic Perspective. *J. Phys. Oceanogr.*, 49:3221–3244.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., and Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830.

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

Plumb, R. A. and Ferrari, R. (2005). Transformed eulerian-mean theory. part I: Nonquasigeostrophic theory for eddies on a zonal-mean flow. *J. Phys. Oceanogr.*, 35:165–174.

Polton, J. A. and Marshall, D. P. (2007). Overturning cells in the Southern Ocean and subtropical gyres. *Ocean Sci.*, 3:17–30.

Polvani, L. M., Banerjee, A., Chemke, R., Doddridge, E. W., Ferreira, D., Gnanadesikan, A., Holland, M. A., Kostov, Y., Marshall, J., Seviour, W. J. M., Solomon, S., and Waugh, D. (2021). Interannual SAM modulation of Antarctic sea ice extent does not account for its long-term trends, pointing to a limited role for ozone depletion. *Geophys. Res. Lett.*, 48 (21):e2021GL094871.

Polvani, L. M. and Smith, K. (2013). Can natural variability explain observed Antarctic sea ice trends? New modeling evidence from CMIP5. *Geophys. Res. Lett.*, 40:3195–3199.

Polvani, L. M., Waugh, D. W., Correa, G. J. P., and Son, S.-W. (2011). Stratospheric ozone depletion: The main driver of twentieth-century atmospheric circulation changes in the Southern Hemisphere. *J. Climate*, 24:795–812.

Poulsen, M. B. (2018). *Parameterizing Southern Ocean eddy-induced circulation in coarse resolution ocean models*. PhD Thesis.

Prather, M. (1986). Numerical advection by conservation of second-order moments. *J. Geophys. Res.*, 91:6671–6681.

Rintoul, S. C. (2018). The global influence of localized dynamics in the Southern Ocean. *Nature*, 558:209–218.

Rintoul, S. C., Hughes, C., and Olbers, D. (2001). The Antarctic Circumpolar Current system. *Ocean Circulation and Climate*, pages 171–302.

Roncoroni, S. (2018). *Nonlinear transient adjustment of the Southern Ocean to wind stress changes*. University of Reading internal report.

Ross, S. M. (2014). *Probabilitá e statistica*. Maggioli Editore.

Rye, C. D., Marshall, J., Kelley, M., Russell, G., Nazarenko, L. S., Kostov, Y., Schmidt, G. A., and Hansen, J. (2020). Antarctic glacial melt as a driver of recent Southern Ocean climate trends. *Geophys. Res. Lett.*, 47(11):e2019GL086892.

Sallée, J.-B. (2018). Southern Ocean warming. *Oceanography*, 31(2):52–62.

Screen, J. A., Gillett, N. P., Stevens, D. P., Marshall, G. J., and Roscoe, H. K. (2009). The role of eddies in the Southern Ocean temperature response to the Southern Annular Mode. *J. Climate*, 22:806–818.

Seviour, W., Gnanadesikan, A., and Waugh, D. (2016). The transient response of the Southern Ocean to stratospheric ozone depletion. *J. Climate*, 29:7383–7396.

Seviour, W. J. M., Codron, F., Doddridge, E. W., Ferreira, D., Gnanadesikan, A., Kelley, M., Kostov, Y., Marshall, J., Polvani, L. M., Thomas, J. L., and Waugh, D. W. (2019). The Southern Ocean sea surface temperature response to ozone depletion: A multimodel comparison. *J. Climate*, 32 (16):5107–5121.

Seviour, W. J. M., Gnanadesikan, A., Waugh, D. W., and Pradal, M.-A. (2017). Transient response of the Southern Ocean to changing ozone: Regional responses and physical mechanisms. *J. Climate*, 30:2463–2480.

Singh, H., Polvani, L. M., and Rasch, P. J. (2019). Antarctic sea ice expansion, driven by internal variability, in the presence of increasing atmospheric CO2. *Geophys. Res. Lett.*, 46 (24):14762–14771.

Sinha, A. and Abernathey, R. P. (2016). Time scales of Southern Ocean eddy equilibration. *J. Phys. Oceanogr.*, 46:2785–2805.

Straub, D. N. (1993). On the transport and angular momentum balance of channel models of the Antarctic Circumpolar Current. *J. Phys. Oceanogr.*, 23:776?782.

Swart, N. C. and Fyfe, J. C. (2012). Observed and simulated changes in the Southern Hemisphere surface westerly wind-stress. *Geophys. Res. Lett.*, 39:L16711.

Swart, N. C., Fyfe, J. C., Gillet, N., and Marshall, G. J. (2015). Comparing trends in the Southern Annular Mode and surface westerly jet. *Journal of Climate*, 28 (22):8840–8859.

Tamsitt, V., Drake, H. F., Morrison, A. K., Talley, L. D., Dufour, C. O., Gray, A. R., Griffies, S. M., Mazloff, M. R., Sarmiento, J. L., Wang, J., and Weijer, W. (2017). Spiraling pathways of global deep waters to the surface of the Southern Ocean. *Nat. Commun.*, 8.

Thompson, A. F. (2008). The atmospheric ocean: eddies and jets in the Antarctic Circumpolar Current. *Phil. Trans. R. Soc. A*, 366:4529–4541.

Thompson, D. W. J., Solomon, S., Kushner, P. J., England, M. H., Grise, K. M., and Karoly, D. J. (2011). Signatures of the Antarctic ozone hole in Southern Hemisphere surface climate change. *Nat. Geosci.*, 4:741–749.

Trefethen, L. N., Birkisson, A., and Driscoll, T. A. (2018). *Exploring ODEs*. SIAM.

Treguier, A. M., England, M. H., Rintoul, S. R., Madec, G., Le Sommer, J., and Molines, J.-M. (2007). Southern Ocean overturning across streamlines in an eddying simulation of the Antarctic Circumpolar Current. *Ocean Sci.*, 3:491–507.

Treguier, A. M., Held, I. M., and Larichev, V. D. (1997). Parameterization of quasigeostrophic eddies in primitive equation ocean models. *J. Phys. Oceanogr.*, 27:567–580.

Tsay, R. S. (2013). *Multivariate Time Series Analysis : With R and Financial Applications*. John Wiley and Sons, Incorporated.

Vallis, G. K. (2017). *Atmospheric and Oceanic Fluid Dynamics*. Cambridge university press.

Vatiwutipong, P. and Phewchean, N. (2019). Alternative way to derive the distribution of the multivariate Ornstein-Uhlenbeck process. *Advances in Difference Equations*, 276.

Viebahn, J. and Eden, C. (2010). Towards the impact of eddies on the response of the Southern Ocean to climate change. *Ocean Modelling*, 34:150–165.

Viebahn, J. and Eden, C. (2012). Standing eddies in the meridional overturning circulation. *J. Phys. Oceanogr.*, 42:1486–1508.

von Storch, H. and Zwiers, F. (1999). *Statistical Analysis in Climate Research*. Cambridge university press.

Ward, M. L. and Hogg, A. M. (2011). Establishment of momentum balance by form stress in a wind-driven channel. *Ocean Modell.*, 40:133–146.

Williams, R. G., Wilson, C., and Hughes, C. W. (2007). Ocean and atmosphere storm tracks: The role of eddy vorticity forcing. *J. Phys. Oceanogr.*, 37:2267–2289.

Wilson, C., Hughes, C. W., and Blundell, J. R. (2015). Forced and intrinsic variability in the response to increased wind stress of an idealized Southern Ocean. *J. Geophys. Res. Oceans,*, 120:113–130.

Wolfe, C. L. and Cessi, P. (2009). Overturning in an eddy-resolving model: The effect of the pole-to-pole temperature gradient. *J. Phys. Oceanogr.*, 39:125–142.

Wolfe, C. L. and Cessi, P. (2014). Salt feedback in the adiabatic overturning circulation. *J. Phys. Oceanogr.*, 44:1175–1194.

Wolfe, C. L. and Cessi, P. (2015). Multiple regimes and low-frequency variability in the quasi-adiabatic overturning circulation. *J. Phys. Oceanogr.*, 45:1690–1708.

Yano, J.-I., Ambaum, M. H. P., Dacre, H. F., and Manzato, A. (2020). A dynamical-system description of precipitation over the tropics and the midlatitudes. *Tellus A: Dynamic Meteorology and Oceanography*, 72:1:1–17.