ON THE EMPIRICAL ORTHOGONAL FUNCTIONS REPRESENTATION OF THE OCEAN CIRCULATION

Dissertation

Zur Erlangung des Doktorgrades der Naturwissenschaften

an der Fakultät für Mathematik, Informatik und Naturwissenschaften

Fachbereich Geowissenschaften

der Universität Hamburg

vorgelegt von

Jairo Alonso Segura Bermúdez

aus Lorica, Kolumbien

Hamburg, 2020

Tag der Disputation: 22.06.2020
Folgende Gutachter empfehlen die Annahme der Dissertation:
Dr. Christian Franzke
Dr. Thomas Frisius

„… That is the plan: just try. If we can do it, wonderful. If not, then fail in the most beautiful way“

J. Klopp

Contents

Zusammenfassung

Die Modellierung der Ozeandynamik ist ein wichtiger Bestandteil von Erdsystemmodellen. Das Verständnis der Wechselwirkungen zwischen Ozean und Atmosphäre erfolgt durch numerische Modelle wichtiger Variablen, wie es der Wind ist. Die windgetriebene Ozeanzirkulation mittlerer Breite umfasst zum Beispiel die subtropischen Gyres, die warmes Wasser aus äquatorialen Regionen polwärts transportieren, und die subpolaren Gyres, die kaltes Polarwasser zum Äquator transportieren, bis zu dem Punkt, an dem beispielsweise Europa erwärmt wird durch die Erweiterung des Golfstroms.

Eine Möglichkeit zur Reduzierung der Komplexität eines Modells könnte mithilfe der Reduktionsmethode „Empirical Orthogonal Functions (EOF)“ erreicht werden. Die EOF-Reduktionsmethode wurde zuvor in der Atmosphäre mit erfolgreichen Ergebnissen (*Achatz and Schmitz*, 1997) verwendet, was die Möglichkeit eröffnet, die Methode für Ozeanmodelle zu verwenden, was bisher selten durchgeführt wurde. Diese Prämisse wird getestet, indem ein Prototyp eines windgetriebenen barotropischen Ozeanmodells verwendet und ein zusätzlicher externer Antrieb angewendet wird, um einen realistischeren Aufbau zu simulieren.

Die Ergebnisse der Rekonstruktion aus der EOF-Reduktionsmethode werden unter anderem anhand der optimalen Zuordnung und der Wasserstein-Abstände ausgewertet. Konzepte, die hauptsächlich in der Wirtschaft verwendet werden und eine neue Alternative als Methode zum Vergleichen von Attraktoren bieten.

Das getestete Modell mit mittlerer Komplexität (*Böning*, 1986) erzeugt Attraktoren mit einer Phasendifferenz zur Referenz, die mit herkömmlichen Fehlerberechnungsmethoden nicht einfach zu vergleichen sind. Bereits eine kleine Anzahl von EOFs (3-7) reicht aus, um die wesentliche Phasenraumdynamik und Stromfunktion zu reproduzieren. Bei den Hauptkomponenten verringern sich die Unterschiede zum Referenzmodell, wenn die Anzahl der EOFs zunimmt.

Diese Arbeit zeigt, dass die Methode der EOF-Reduktion und die Rekonstruktion vielversprechend ist und zufriedenstellende Reproduktionen der kinetischen Energie, der Stromfunktion und der Phasenraumdynamik auch für eine modifizierte Version des *Böning* (1986) -Modells erhalten werden. Dies erfolgt durch Einbeziehen eines zusätzlichen externen Antriebs mit kohärenter räumlicher Struktur und stochastischer zeitabhängiger

Amplitude. Die neue Kraft bricht die im unveränderten Modell erzeugte symmetrische Struktur. Dies stellt eine Herausforderung dar, mit dem rekonstruierten Modell zu reproduzieren, aber für kleine Variationen wird eine zufriedenstellende Reproduktion erreicht.

Schließlich gibt es viele mögliche Verbesserungen und empirische Korrekturen, die implementiert werden könnten, um die Methode effizienter zu gestalten. Es gibt auch mehr potenzielle Prototypen von Ozeanmodellen, die einen Vorteil für zukünftige Arbeiten darstellen könnten.

Abstract

Modelling the ocean dynamics is an important part in Earth system models. The understanding of the interplays between ocean and atmosphere is performed through numerical models of important variables, as it is the wind. The wind-driven mid-latitude ocean circulation, for example, includes the subtropical gyres which transport warm water poleward from equatorial regions, and the subpolar gyres that carry cold polar water towards the equator. To the point that Europe, for instance, is warmed by the extension of the Gulf Stream.

A way to reduce a model complexity could be achieved by using the Empirical Orthogonal Functions (EOF) reduction method. The EOF reduction method has been used in the atmosphere before with successful results (*Achatz and Schmitz*, 1997), which opens the possibility of using the method for ocean models, something that has been rarely done. This premise is tested by using a prototype wind-driven barotropic ocean model and applying an additional external forcing to simulate a more realistic setup.

The results of the reconstruction from the EOF reduction method are evaluated by using the optimal assignment and Wasserstein distances among others. Concepts that are mostly used in economy and that offer a new alternative as a method to compare attractors.

The intermediate complexity model tested (*Böning*, 1986), produces attractors that have a phase difference to the reference that are not easy to compare with traditional error calculations methods. Already a small number of EOFs (3-7) suffice to reproduce the essential phase space dynamics and streamfunction. For the principal components, the differences with the reference model decrease when the number of EOFs increases.

This thesis shows that the method of EOF reduction and reconstruction is promising, satisfactory reproductions of the kinetic energy, the streamfunction and the phase space dynamics are also obtained for a modified version of the *Böning* (1986) model. This is done by including an additional external forcing with coherent spatial structure and stochastic time dependent amplitude. The new force breaks the symmetric structure produced in the unmodified model. This represents a challenge to reproduce with the reconstructed model but for small variations, a satisfactory reproduction is achieved.

Finally, there are many potential improvements and empirical corrections that could be implemented to make the method more efficient. Also, there are more potential prototype ocean models that could represent an advantage for future works.

1. Introduction

The climate is a complex system and oceans are an essential part of it. The hydrosphere consists of mostly oceans because they cover more than two thirds of the surface of the Earth, which plays a key role in the natural variability of the weather. Ocean modeling by extension is, therefore, important and advantageous in order to understand our current climate and the effects of climate change (*Miller*, 2007).

Oceans have a profound influence on our climate, our weather and the ecosystems of the Earth. The great heat capacity of the oceans (water alone has a heat capacity 4 times greater than air and a volumetric density a thousand times higher) exerts, in some cases, a control effect on the climate of the Earth, stores and transports heat, CO_2, nutrients and, of course, water, which in the long term became a great pillar for life on our planet. The oceans also convert short-term climatic fluctuations into long-term climatic variations (*Hasselmann*, 1982).

The waters of the earth are gathered in shallow, irregular and interconnected basins. The general circulation of the ocean is the large scale pattern of the flow in these basins. Alongside the pattern of motion, the related fields of pressure, temperature, salinity and density are as well important parts of a complete theory for the ocean circulation, since they are dynamically connected to the motion of the oceans (*Pedlosky*, 1996).

The wind-driven mid-latitude ocean circulation forms an important part of the global climate system. The western boundary currents of subtropical gyres transport warm water poleward from equatorial regions, whilst those of subpolar gyres carry cold polar water towards the equator. The resulting sea surface temperature anomalies and the effects they have on evaporation have a profound influence on the climate of contiguous continental regions. For instance, Europe is warmed by the extension of the Gulf Stream, whilst Newfoundland is cooled by the Labrador current (*Kiss*, 1998). Studies have also shown that in the last decade, the upper ocean has warmed up much more in comparison with deeper layers therefore contributing in a higher rate of sea level rise due to thermal expansion (*Llovel et al.*, 2014).

When a model improves in replicating reality inevitably it becomes more complex due to the inclusion of new processes and scales and their interactions, the most complex ocean circulation models include as much physics as is computationally feasible and have an

output almost as complex as the real system. Models and their discrepancies with the real climate have been studied before (*Lambert and Boer*, 2001) and it has been shown that these discrepancies are more visible regarding ocean variables. Everything plays an important role in the real ocean and this accentuates the difficulty in discarding variables when studying their influence over the whole system.

A stochastic model of climate variability was considered by *Hasselmann* (1976) in which slow changes of climate were explained as the integral response to continuous random excitation by short-period weather perturbations. The coupled system ocean-atmosphere-cryosphere-land, which can be divided into a fast responding system (the atmosphere) and a slow one (the ocean), produces a climate response to the random atmospheric forcing that could be described as a random walk or diffusion process (first-order Markov process). The nature of the study allows that some of the general properties of stochastic climate models could be contrasted against observed climatic variability.

Acknowledging that current comprehensive Earth system models for weather and climate cannot resolve all processes and scales, *Franzke et al.* (2015) proposed the use of stochastic parameterization schemes for existing numerical weather and climate prediction models, and showed that for data on the available timescales, the method provides more skillful weather forecasts than traditional ensemble prediction methods. Their effect was even demonstrated in laboratory experiments (*Williams et al.*, 2005, 2008), and their corresponding uncertainty estimations were done within the parameterizations that models accurately simulate the large scales, while capturing the statistical properties of the small and unpredictable scales.

Stochastic forcing was also used to explain the decadal variability in certain regions of the ocean (*Frankignoul et al.*, 1997), in this particular case the extratropical ocean, and the model was a simple linear stochastic forcing model where the wind stress forcing was modeled as stochastic with a white frequency spectrum, so its influence in the interior ocean would be representative of short-timescale weather fluctuations. The model was able to reproduce a red baroclinic spectrum appropriately and the baroclinic predictions for a white-noise wind stress curl spectrum was consistent with real ocean decadal changes. The model was coarse and could not reproduce details but it was suited for improvements. The stochastic forcing approach was also used to study El Niño-like phenomena in the North Atlantic using a coupled ocean-atmosphere system in a global circulation model, aiming to improve the interannual to decadal predictability of these fluctuations (*Grötzner et al.*, 1999).

Sura et al. (2001) from their side, analyzed the influence of adding a stochastic wind stress component in a double-gyre, gravity reduced wind driven ocean model. The stochastic variability represented atmospheric transient eddies. This addition caused the system to

undergo through a bimodal behaviour, the spatially inhomogeneous stochastic wind field pushed the system into a non symmetric flow pattern in a phenomenon known as „Noise-induced transition". Later, deepening on their previous study with the same model, (*Sura and Penland*, 2002) analyzed how relatively small variations of the stochastic forcing influence the system, finding that physical processes with widely spread timescales interacted, that the distribution of the forcing was important to the dynamical response of the system. In this case, the stochastic forcing represented the macroscopic manifestation of unresolved nonlinear interactions. The prediction of extreme events on seasonal to interannual timescales depend heavily on those unresolved aforementioned processes.

1.1. Variability in the ocean

The ocean distinctive dynamics translates into variabilities on different time scales and explaining them has been the subject of oceanographic research for years, ranging from a work by *Veronis* (1963) motivated by *Lorenz* (1963) to nowadays. *Zacharuk et al.* (2018) for example, analyzed a simple ocean model using stochastic parameterization of the subgrid. The causes of these variabilities are manifested in their interaction with the atmospheric system (*Bjerknes*, 1964), the thermohaline system (*Winton and Sarachik*, 1993; *Greatbatch and Zhang*, 1995) and the wind-driven circulation (*Meacham and Berloff* , 1997). Studies have shown that the latter alone has enough potential to produce a low-frequency time dependency (*Meacham*, 2000). Thus it is not far fetched to think that understanding the natural part of the variability would help to better understand and predict the impact of human activity on the climate (*Dellnitz*, 2000).

Those potential changes on the ocean circulation are the focus of attention for decadal climate predictability. A study by *Nilsen et al.* (2003) suggests that an atmospheric pattern resembling the North Atlantic Oscillation (NAO) is the main driving force for the variations in the water volume exchanges in the North Atlantic-Nordic Seas. This Implies a strong link between the Meridional Overturning Circulation (MOC) and the NAO and means that considering those phenomena separately may result in important details being missed (see *Marshall et al.*, 2001a).

1.2. Modes of variability

There are plenty known variabilities in the ocean that are actively studied and, although some of them are mentioned through this document, the emphasis of this study is the influence of wind driven circulation at midlatitudes.

1.2.1. The North Atlantic oscillation (NAO)

The NAO is one of the most prominent and recurrent patterns of atmospheric circulation variability, hence it influences significantly the properties and circulation of the ocean. It dictates climate variability from the eastern United States to Siberia and from the Arctic to the subtropical Atlantic and even north Africa, understanding the dynamics of the interaction between the NAO and the ocean is an important task to produce more accurate coupled ocean-atmosphere models and in a way attacking important issues like climate change (*Visbeck et al.*, 2003).

Since the NAO exhibits variations at interannual to multidecadal time scales its importance is undeniable. The connection between NAO and the ocean has been an object of study for decades. *Mehta et al.* (2000), e.g., revised more than 40 years of data of SST and contrasted those values to several experiments of atmospheric general circulation models, and it showed that the relation between NAO and SST is not so simple; one influences the other. A sensitivity study done by *Eden and Willebrand* (2001) on a model of the Atlantic Ocean that was forced with decadal-scale time series of surface fluxes, showed a fast barotropic response and a delayed baroclinic oceanic response to the NAO. a fast response of a barotropic anticyclonic circulation anomaly to a high NAO and an enhanced meridional overturning due to a spinup of the subpolar gyre.

Hurrell and Deser (2010) stated that the NAO also affects the ocean through changes in heat content, gyre circulations, mixed layer depth, salinity, high latitude deep water formation and sea ice cover. It was also found that there is no preferred time scale of variability for the NAO: large changes occur from one winter to the next and from one decade to the next. There was also a large amount of within-season variability in the patterns of atmospheric circulation of the North Atlantic, so that most winters cannot be characterized solely by a typical NAO structure.

There is evidence of links between the ocean circulation and the large-scale atmospheric flow. *Czaja et al.* (2003) discussed the feedback between the atmospheric circulation, in terms of the NAO, and the ocean circulation, both horizontal and overturning. At mid-latitudes the large-scale mean horizontal circulation of the upper $1 - 2$ km of the global ocean is dominated by subtropical and subpolar gyres driven by surface wind stress. These gyres are recirculations spanning ocean basins, with slow meridional flow in most of the basin returned by narrow, rapid boundary currents at the western side. These currents separate from the coast at some point and meander across the interior between the gyres. Western-intensified circulations of this type appear at mid-latitudes in all ocean basins, examples of the poleward western boundary current being the Gulf Stream in the North Atlantic, the Kuroshio in the North Pacific and the East Australia Current in the South Pacific. Subtropical and subpolar gyres have a similar western-intensified form, but

subtropical gyres have an anticyclonic circulation, whilst subpolar gyres are cyclonic.

1.3. Wind driven circulation and ocean dynamics in midlatitudes

The study of wind driven ocean circulation was a focus of interest from different points of views, from mathematics, fluid dynamics, meteorology, geophysical among others. The ocean acquires momentum (and kinetic energy) due to the force applied from the wind against the ocean upper boundary. The ocean circulation is dominated (in the first hundreds meters) by the wind stress (*Wunsch*, 2002) so it is important then to study in depth the influence the wind has over the overall ocean behaviour.

Simple models with a manageable number of degrees of freedom can be used to model and understand certain characteristics of the flow. It serves to lay the foundation of a model hierarchy consisting of different simple models in which individual processes can be viewed in isolation. The understanding of these separate processes, their interactions, as well as the inherent mathematical-numerical structures will help in the future when it comes to application of detailed numerical models of the general oceanic and atmospheric circulation (*Veronis*, 1966; *Primeau*, 1998). The linear theories of *Stommel* (1948) and *Munk* (1950) and the theory of nonlinear models according to *Fofonoff* (1954) and *Charney* (1955) represent the extremes of a continuum of models (*Meacham and Berloff* , 1997).

Böning (1986) studied a wind-driven ocean circulation in an idealized, rectangular ocean model, which was forced by steady zonal winds and damped by lateral and bottom friction. He found that for diffusivity sufficiently large and using free-slip boundary conditions, steady solutions were found for all chosen values of Rossby number (R_O). If lateral diffusion dominated the bottom friction, a strong recirculating subgyre emerged in the northwestern corner of the basin. In the *Veronis* case, inertial recirculation only took place with the values of R_O so large that the eastward jet reached the eastern boundary. This study established the foundation of many following studies that included a more complete set of physical parameters.

Western boundary currents vary seasonally and have attracted significant attention when studying the ocean circulation. *Jiang et al.* (1995) used a reduced-gravity shallow-water wind driven double gyre model and analyzed the behaviour of western boundary currents and found that when the meridionally symmetric wind stress is sufficiently strong, two steady solutions are achieved with the model: with the oceanic response from the imposed wind stress, a periodic oscillation arises, and increasing the wind forcing and or decreasing the viscosity would translate into aperiodic solutions.

Something that had been suffering from a lack of attention in the recent decade are the spectral models for global ocean circulation, although they are widely used in atmospheric general circulation models. *Frisius et al.* (2009) introduced SOM (Spectral Ocean Model) a global spectral barotropic ocean model for general circulation and weather prediction, and it was shown that the coupling of this spectral ocean circulation model with the correspondent spectral atmospheric model was notably smoother due to their grid structure nature and that the singularities at the poles caused no issue. On the downside, the model became inefficient at high spatial resolutions, which makes SOM good for studying the atmosphere-ocean interactions in idealized contexts but not for simulating the large-to-small scale interactions in a realistic ocean.

The role of the North Atlantic ocean on the NAO has been discussed and it is said that it modulates the NAO. *Dewar* (2001) used a quasigeostrophic ocean model with a highly idealized atmospheric input to study the aforementioned influence, and it was found that the intrinsic variability of the wind-driven ocean circulation damped the SST anomalies at long timescales. Low-frequency modes of instability were observed and these instabilities are at a maximum near the western boundary current. In addition, he highlighted the importance of nonlinear processes associated with mesoscale eddies in the ocean for driving variability in the SST and the overlying atmosphere. Variability of this type has the potential to alter heat transport in the ocean, and is therefore important to climate variability.

In the present work, the main focus is on a wind-driven barotropic model for a rectangular ocean basin. This model is based on Fourier functions and can be formulated as a dynamical system including quadratic nonlinearities, so it is proper to review other similar studies.

When studying the motion of fluids it will always be about solving the Navier-Stokes equations on any of its approximations, but solving the Navier-Stokes equations is not a trivial matter. Numerical solvers must balance accuracy, stability and consistency, which poses constraints on the time step and grid resolution that one can employ. One way is to translate these equations to computational mathematical language using analytical mathematical techniques that are available. So the focus can be shifted to scales that are of interest for the researcher by using specific modeling techniques. The biggest issue is that exact solutions to this set of equations are yet not known. So it falls into constructing idealized models that are valid enough approximations to the full system. Some mathematicians have also studied the derivation of some basic and classical equations of meteorology and oceanography from the Navier-Stokes equations in a rotating frame with viscous dissipation (see *Desjardins and Grenier*, 1999).

Idealized models are one way of addressing this difficulty; thus, it has been the resource for many researchers. For instance, the wind-driven ocean circulation, which is useful for

duplicating the real dynamics of the ocean, has been studied since the late 40s (*Stommel*, 1948; *Munk*, 1950). It has been possible to obtain considerable good results from models on how the ocean circulation reaches an equilibrium state in a simple, closed and rectangular basin in response to a steady wind stress pattern. The simple, linear one-layer models have shown the central role of the planetary vorticity gradient in establishing the westward intensification of the midlatitude gyres.

The understanding of important processes can be pinpointed clearer in models with less input variables than in complicated complex models, also prototype models imply faster simulations and a clearer opportunity to analyze the influence of large scale phenomena on the general circulation patterns. Highly idealized models have been used to reduce the problem to the most relevant physics processes, and it has been found that it is still reasonably amenable to simple explanation (*Scott*, 1998). An additional approach to reduce the complexity of the system is by applying reduction methods to those simplified ocean models, therefore analyzing how variables impact on the general ocean circulation comes with ease.

In addition to the barotropic ocean models already mentioned, e.g. *Böning* (1986); *Bryan* (1963); *Munk* (1950), other interesting approaches have been taken. *Barcilon* (1998) carried out a very detailed analysis of the classical barotropic ocean model with bottom friction, varying the Rossby number (R_O) which represents the effects of the Earths rotation, and the vertical Ekman number or bottom friction parameter (λ_B) with the space divided into four sectors associated with four different dynamical regimes discriminated by contrasting R_o with λ_B. A large-scale clockwise circulation in the subtropical North Atlantic was found and similarly patterns of circulation have been found in each of the oceans in both hemispheres. Such large-scale patterns of flow conform to the general circulation of the ocean.

One of the commonly used approaches to make the system closer to the real ocean, is by modifying the external forcing. This could be by simulating the effect of the NAO with an anomaly in the form of a cosine windstress the midlatitude domain (*Badin et al.*, 2003). This is something that proved to give satisfactory outcomes and the resulting wind anomalies resembled in acceptable means the real climate anomalies.

Using this kind of simple models for a wind-driven barotropic ocean in a closed basin of constant depth, *Verron and Jo* (1994) analyzed the circulation patterns of the transitions from a δ_i (non-linearity, inertia) system to one dominated by δ_l (lateral viscous dissipation), and from a δ_i one to a δ_b (bottom friction dissipation) one. This analysis gives insight into how the system behaves with variations of these parameters and also when increasing the whole system complexity, from a simple inertia model (*Fofonoff* , 1954) to a bottom friction one (*Stommel*, 1948) to a Laplacian lateral friction (*Munk*, 1950) to finally a biharmonic lateral friction model (*Holland*, 1978) and extending them by including non-linear

phenomena. A general scheme on the parameters space was derived, so it was easier to see that the relative importance of δ_i, δ_b and δ_l determines which process constitutes the main controlling factor with respect to the dynamics of the western boundary current. The models used are conceptually simple, but the number of possible configurations is high and thus it is possible to increase the complexity. The transition from the linear systems to those strongly influenced by non-linearities could help with understanding of the behaviour of the real ocean.

In the same direction, using a homogeneous model of wind-driven circulation, *Badin and Crisciani* (2009) observed the transition from non-linear to linear regimes, transitioning continuously the values of δ_i from non-linear regimes governed by the steady form of *Bryan* (1963) model to linear regimes governed by the *Munk* (1950) model, exposing that the linear regime should be considered autonomous and separate from the non-linear regimes due to the loss of the Sverdrup balance in some intermediate phases.

An interesting numerical phenomena is observed when eddy viscosity coefficients are low enough; then, the *Munk* model presents a couple of zonally lined up gyres, phenomena not present in the real ocean, which is something that caught the attention of *Badin et al.* (2009). This proves that this happens because the flow is not in Sverdrup balance and the aforementioned aligned gyres cannot take place in the real ocean, and although in other regimes the model was consistent in reproducing the ocean dynamics, it was proven that the *Munk* model has its limitations and can be improved by the use of nonlinear circulation models.

To test the method, simplified ocean models have been used: initially a wind-driven barotropic model of a square basin with a rigid lid discarding vertical variations (*Dellnitz*, 2000) based on Fourier functions formulated as a dynamical system. In the North Atlantic, for example, there are some variabilities that can be understood by studying these kinds of simplified models.

Empirical Orthogonal Function (EOF) analysis had an early introduction to meteorology by *Lorenz* (1956) and has become a statistical tool of fundamental importance in atmospheric, oceanic, and climate science for exploratory data analysis and dynamical mode reduction (*Selten*, 1995). The EOFs give us information about the modes of variability which explain the most variance in the system and are not only used on multivariate statistics or atmospheric sciences but also to analyze stochastic fields and interaction terms. The EOF reduction method has been used in the atmosphere before (*Achatz and Schmitz*, 1997; *Franzke et al.*, 2005; *Franzke and Majda*, 2006) with successful results, which open the possibility of using the method for ocean models.

1.4. Research questions and outline

On the whole, the importance of the wind input on accurate Earth system models, and being aware of its influence in the global ocean circulation. The research on the influence of the wind forcing over prototype ocean models is clearly relevant. All of the above leads to specific questions that motivates the main focus of this document and will be addressed in the course of it. The main question being

Is it possible to reproduce the oceanic circulation of a prototype wind driven ocean model with a EOF reduced model?

and to answer this, it is necessary to tackle more specific questions such as

» *Would the aforementioned reduced model be consistent with changes on the nature of windstress input and a more realistic setting?*

» *Can the same reduced model reproduce slightly different atmospheric states and potentially be used for predictions?*

» *How would the model solutions react to changes of Reynolds number?*

This study is organized as follows: in Chapter 2, a review of theoretical concepts and relevant approximations for wind driven ocean models are discussed and the concept of EOFs reduction method is applied to the ocean models. Chapter 3 covers the establishment of the dynamical system and its relevant parameters, the structure of the wind forcing that is used and its variations.

The analysis of the EOF reduction model is discussed in chapter 4. Where three main ocean model are analyzed, starting with a low spectral order model developed by *Veronis* (1963), followed by a more complex model (*Böning*, 1986) and finally a variation of the latter to produce an ocean model with a realistic windstress input. And in Chapter 4, an overall summary, closing remarks, and a brief outlook of the study is provided.

2. Spectral ocean models and EOFs

Solving the Navier-Stokes equations is not a trivial matter. Numerical solvers must face concerns regarding accuracy, stability and consistency, which pose constraints on the time step and grid resolution that one can employ. The usual approach is translating the Navier-Stokes equations to computational mathematical language using analytical mathematical techniques available to us, that it is possible to focus on scales that we are interested in, using specific model techniques, like the Reynolds averaged Navier-Stokes equations.

Wind driven oceans can be considered a non divergent fluid ($\nabla \cdot \vec{v} = 0$) and non compressible (in overturning circulation models variations in density have to be considered), which means that the Boussinesq approximation can be used in the Navier-Stokes equations, and it is reflected in the continuity equation. In a prototype ocean, with a constant depth H and a rectangular basin, the divergence of the vertically averaged horizontal flow.

This yields an ocean model driven only by the external forcing where forcing corresponds to the wind forcing. It is known that on a β-plane, or on the surface of a rotating sphere, an eastward mean zonal flow can be maintained by a vorticity stirring that imparts no net momentum to the fluid. That momentum converges in a rapidly rotating flow mixed in a meridionally localized region. Scaling analysis suggests that over most of the mid-latitude ocean basins, inertial and frictional terms in the vorticity equation can be neglected in favor of a balance between the advection of planetary vorticty and the input of vorticity by the curl of the wind-stress.

$$\int \beta v dz = \frac{\partial \tau_0^y}{\partial x} - \frac{\partial \tau_0^x}{\partial y} \tag{2.1}$$

This is the Sverdrup relation: at any location in the ocean, *the vertically integrated meridional velocity is given by the curl of the wind stress at the surface* implying that a non stratified wind driven ocean model is an amenable assumption. This relation is one of the pillars of physical oceanography and a more complete discussion of it can be found in *Vallis* (2019).

Since the early papers of *Stommel* (1948) and *Munk* (1950), a particular class of models has been studied, one that deals with the question of how the ocean circulation equilibrates in a simple, closed and often rectangular basin in response to a simple, steady windstress

pattern. The imminent issue associated with the mentioned models is that the solutions obtained are simply approximations to the full set of equations with a yet unknown analytical solution. The system can be further simplified by making approximations and idealizations.

A crucial disadvantage of the spectral method used to be the enormous amount of computation that required to evaluate the convolution sums of non-linear terms in the spectral space. For this reason, the application of the spectral method was initially restricted to systems with few degrees of freedom. It is the case for the oceanic circulation model of *Veronis* (1963, 1966) widely considered as a „low-order“ models or in the case of the atmosphere, the *Lorenz* (1963) atmospheric model (*Böning*, 1985).

2.1. The barotropic model: a homogeneous model of the ocean

The barotropic quasigeostrophic vorticity equation describes the horizontal motion of a homogeneous fluid on the $\beta-$plane. This fluid can be subdivided into three layers in the vertical direction. In the middle, there is a broad homogeneous geostrophic layer that lies between two thin Ekman layers. In the upper Ekman layer, the wind shear prompts Ekman transport, creating a vertical velocity at the top of the inner layer. The lower Ekman layer dampens the flow caused by the bottom friction. The absolute vorticity of this quasigeostrophic model can be altered by the horizontal diffusion (*Pedlosky*, 1987).

One of the first efforts to study the ocean through models was made by *Stommel* (1948), trying to explain the western boundary current phenomena. For that, a wind-driven ocean circulation was devised including the beta-effect of the Coriolis force. This and the well established *Munk* (1950) approach are simple linear stationary wave theories introducing bottom and lateral friction respectively. In some regimes these models are consistent in reproducing the ocean dynamics, but it was clear that they have their limitations, some of which can be improved by the use of non linear circulation models as is the case for the non-stationary, nonlinear *Veronis* (1963) model, defined by

$$\frac{\partial}{\partial t}\nabla^2\psi + \mathbf{J}(\psi, \nabla^2\psi) + \beta\frac{\partial\psi}{\partial x} = -K\nabla^2\psi + \frac{\nabla\times\tau}{H} \tag{2.2}$$

At its time, the numerical solution of this model represented a computational challenge and was mostly limited by it. Eventually, this issue was overcome and models with higher complexity but still simplistic and idealized were created, as was the case for the model of *Böning* (1986).

Having a simple system facilitates pinpointing variations and sensitivities in the system. That is why the main focus of the work of *Böning* (1986) is on analyzing the horizontal transport in a closed square-ocean basin of width L and constant depth H on a mid-latitude β-plane. If lateral friction is included by a simple eddy viscosity hypothesis (*Munk*, 1950) and vertical friction by means of Ekman layers at both the top and the bottom, the model is stated in terms of the barotropic vorticity equation (BVE, eq.2.4), with the stream function ψ for the vertically averaged velocity components u (east) and v (north).

$$u = -\frac{\partial \psi}{\partial y} \quad , \quad v = \frac{\partial \psi}{\partial x} \tag{2.3}$$

Then barotropic vorticity equation has the structure

$$\frac{\partial}{\partial t}\nabla^2\psi + \mathbf{J}(\psi, \nabla^2\psi) + \beta\psi_x = \frac{1}{H}\hat{k} \cdot (\nabla \times \tau) + A\nabla^4\psi - K\nabla^2\psi \tag{2.4}$$

where J is the Jacobian operator, β is the gradient of planetary vorticity, $f = f_0 + \beta y$, H is the constant model depth, $\hat{k}$ is the unit vector in the vertical (z up), τ is the wind shear stress, A and K are the exchange coefficients for the lateral and bottom friction on the replacement concept, respectively (*Munk*, 1950). For comparison, some calculations were performed with friction approach $-k\nabla^2\psi$ by *Stommel* (1948).

Equation eq. 2.4 describes the local-temporal change of relative vorticity $\zeta = \nabla^2\psi$ by the rotation of the wind shear stress, by dissipation due to the lateral diffusion (or the linear bottom friction), and by the transport of vorticity due to the advective terms and the advection of planetary vorticity.

To describe the strategies employed for emulating the simple model, it is useful to consider the symbolic representation of a dynamical model, which can be split into three main types summarized in terms of a schematic form of the governing equations (either primitive equations or a balance set)

$$\dot{\mathbf{a}} = L\mathbf{a} + N\mathbf{a} + F \tag{2.5}$$

where $\mathbf{a}$ is a state of the model vector corresponding to the prognostic variables, $\dot{\mathbf{a}}$ is the corresponding tendency, L is a linear operator, corresponding to linearization on some fixed flow, $N(\mathbf{a})$ is a complex nonlinear operator dependent on $\mathbf{a}$ and F represents the forcing terms.

In principle, construction of eq.2.5 can be done using purely empirical procedures once a way of discretizing $\mathbf{a}$ has been decided. The need to define only the functional form of N arises and so the coefficients of those functions as well as the elements of L and F can be found using an optimization approach. General circulation studies typically involve

altering the forcing function, F, and integrating eq.2.5 until a statistically stable state has been obtained (*Valdes and Hoskins*, 1989).

A wind-driven barotropic model for a rectangular ocean basin, based on Fourier functions, can be formulated as a dynamical system including non linearities which fits the description of systems suitable for an EOF representation, for it can be reshaped to take the form of eq.2.5 as follows

$$\begin{aligned} \frac{\mathrm{d}X_i}{\mathrm{d}t} &= \sum_{j=1}^{n}\sum_{k=1}^{n} A_{ijk}X_jX_k + \sum_{j=1}^{n} B_{ij}X_j + F_i(t) \\ &\text{with} \quad i = 1, \cdots, n \quad \text{and} \quad n = n_x \times n_y \end{aligned} \tag{2.6}$$

Where n is the total number of spectral coefficients, n_x and n_y are the zonal and meridional spectral numbers of the approximation.

Except in the case when fast surface waves are of interest, we can exploit the fact that large-scale motions in the ocean are relatively slow and introduce the rigid-lid approximation. Large-scale movements with small Rossby numbers are close to geostrophic equilibrium, so it would be safe to assume that the displacements of the surface are negligible compared with interface displacements (*Cushman-Roisin and Beckers*, 2011).

2.2. Empirical orthogonal functions (EOFs)

Empirical orthogonal function analysis (EOF) is a statistical orthogonal decomposition method for studying the variability of a single field. The method finds spatial patterns of variability, their time variation, and gives a measure of the „magnitude“ of each pattern (*Björnsson and Venegas*, 1997). The essential idea is to generate an optimal basis for the representation of data set of measurements, or of several simulations of a dynamic system. Some important properties of the EOFs to highlight are:

- Provides a set of orthogonal functions that effectively compresses data, being able to explain in the first terms the biggest variability of the data.
- It allows the reduction of the number of variables and provides a method of eliminating noise or the less predictable part of the data.
- As an orthogonal set of functions with adaptable structure of eigenfunctions defined by the structure of the data in itself, it provides broader range in contrast with the Fourier Analysis which assumes a sinusoidal variation of the data set (*Muñoz Pérez et al.*, 2001).

Simply put, the EOFs are eigenvectors of the covariance matrix and they describe typical variability patterns, like propagating waves. In essence, they find a new set of variables that capture most of the observed variance from the data through linear combinations of the original variables. This is done through the decomposition of a continuous space-time field $\boldsymbol{\varphi}(t, \vec{r})$ which *Hannachi et al.* (2007) describes as

$$\boldsymbol{\varphi}(t, \vec{r}) = \sum_{i=1}^{m} \varphi_i(t)\phi_i(\vec{r}) \tag{2.7}$$

where t represents time and $\vec{r}$ the position vector and m is the modes in the field (and in theory $m \to \infty$). These modes are composed of space functions $\phi_i(\vec{r})$ and a set of time dependent terms $\varphi_i(t)$ with $t = 1, \cdots, p$. In the matrix $\boldsymbol{\varphi}$, each row is the spatial measurement, and each column is a time series of observations.

For the decomposition it is important to calculate the anomaly field over a given time, which is the calculation of the time anomaly version of the original data set ($\boldsymbol{\varphi}_O$). This can be done by removing the respective mean state of the time series $\boldsymbol{\varphi}_O$, that represented by $\boldsymbol{\varphi}$. Then the covariance matrix of the scalar field $\boldsymbol{S}$ can be calculated using the relation

$$\boldsymbol{S} = \frac{1}{p-1}\boldsymbol{\varphi}^T\boldsymbol{\varphi} \tag{2.8}$$

The calculation of the EOFs is then reduced to solve the eigenvalue problem:

$$(\boldsymbol{S} - \lambda\mathbf{I})\vec{E} = \vec{0} \tag{2.9}$$

The EOF E_i is defined as the eigenvector of the covariance matrix $\boldsymbol{S}$ and λ_i is its corresponding eigenvalue. Sorting the eigenvectors from the biggest eigenvalue to the lowest one. The 1st EOF is then associated to the biggest eigenvalue λ_1, so that the higher the number of EOF, the lower is the value of its eigenvalue. For a measure of the explained variance of $\boldsymbol{S}$ of a certain eigenvalue λ_i ($\sigma^2_{\lambda_i}$), said eigenvalue λ_i can be divided by the total sum all the eigenvalues, meaning

$$\sigma^2_{\lambda_i} = 100\% \times \frac{\lambda_i}{\sum_{i=1}^{m} \lambda_i} \tag{2.10}$$

The projection of the anomaly field $\boldsymbol{\varphi}$ onto the i^{th} EOF, E_i, provides the evolution in time of the pattern, which is

$$\vec{P}_i = \boldsymbol{\varphi}\vec{E}_i \tag{2.11}$$

When plotting an EOF as a map, the pattern obtained represents an oscillation and its evolution on time is given by their principal component, which describes the magnitude and timing of the oscillation of said pattern (*Björnsson and Venegas*, 1997). Consequently, $\vec{P_i}$ is then the i^{th} principal component (PC) that corresponds to E_i, with elements P_{ti} defined as:

$$P_{ti} = \sum_{j=1}^{m} \varphi_{tj} E_{ji} \qquad \text{where} \qquad t = 1, \cdots, p \tag{2.12}$$

With this, eq.2.12 and eq.2.7 share similarities worthy of highlight: $\varphi_i(t)$ is analogous to φ_{tj} and the space function $\phi_i(\vec{r})$ to E_{ji}, respectively.

The EOFs are orthogonal and the PCs are uncorrelated by definition, and this orthogonality provides a complete orthogonal basis for the data matrix so any field can be expanded in terms of such EOFs. The decomposition is defined as

$$\boldsymbol{\varphi}_t = \sum_{i=1}^{m} \vec{P}_{ti} \boldsymbol{E}_i \tag{2.13}$$

where P_{ti} is the i^{th} principal component P_i at a given time t (*Hannachi et al.*, 2007). This expansion transforms the problem from a set of partial differential equations to a coupled system of ordinary differential equations describing the evolution of the system. Since in theory the original streamfunction has an infinite number of Fourier components, so the expanded number with the new basis is finite, the method represents a reduced form of the barotropic vorticity equation (BVE) eq.2.4 (*Selten*, 1995).The field variable X_i in 2.6 can be expanded in terms of EOFs, taking the form:

$$\mathbf{X}_i(t) = \sum_{j=1}^{n_T} \vec{P}_j(t) E_{ij} + \overline{X^r} \tag{2.14}$$

Here $\mathbf{X}^r$ denotes the reference states and n_T is the number of EOFs. Knowing that n_T represents the dimension of the truncated phase space ($n_T \ll n$), using the first few EOFs (which possess the largest eigenvalues) can be assumed that the first n_T eigenvectors are capturing the large-scale dynamical behavior of the system. With this new formulation, the dynamical system with the EOF reconstruction leads to

$$\frac{\mathrm{d}a_\nu}{\mathrm{d}t} = \sum_{\varepsilon=1}^{n_T} \sum_{\rho=1}^{n_T} N^q_{\nu\varepsilon\rho} a_\varepsilon a_\rho + \sum_{\varepsilon=1}^{n_T} L^q_{\nu\varepsilon} a_\varepsilon + G^q_\nu(t) \tag{2.15}$$

which keeps the formal structure as the original system eq.2.7 but with different coefficients $N^q_{\nu\varepsilon\rho}$, $L^q_{\nu\varepsilon}$ and external forcing $G^q_\nu(t)$ and with fewer equations. Since the relevant dynamics are contained in the first EOFs, it is possible to truncate the system to a

small number of equations that describe the system accurately to a certain extent. Therefore, a EOF reduction would mean a simplified dynamical system (*Niu et al.*, 2015).

2.2.1. Empirical-based correction to the EOFs

As a probable result of the truncation and thus neglecting some physical processes and scales, the fields studied with this method lack necessary properties to accurately resemble the reference field. In this kind of reduction method, these factors have already been studied. When comparing the reconstructed model to the reference model, differences that grow over time emerge.

It has also been reported (*Achatz and Schmitz*, 1997) that using the structure of the reference model 2.6 to add an empirically based correction to the forcing and the linear operator minimizes the error considerably.

One of these solutions was proposed by *Achatz and Branstator* (1999), replacing the operators of the dynamical system such as the forcing G^q and the linear operator L^q with corrected empirically based quantities in the form

$$\vec{G} = \vec{G}^e + \vec{G}^q = -\overline{N^q} \tag{2.16}$$

for the external forcing, where $\vec{G}$ is divided in two parts, $\vec{G}^q$ represents the original uncorrected forcing and $\vec{G}^e$ is the empirically based corrected form of it. In this case the terms of the nonlinear operator $\vec{N}$ are defined as

$$N_\nu^q = \sum_{\varepsilon=1}^{n_T} \sum_{\rho=1}^{n_T} N_{\nu\varepsilon\rho}^q a_\varepsilon^r a_\rho^r \tag{2.17}$$

and similarly for the linear interaction coefficient $\boldsymbol{L}$

$$\vec{L} = \vec{L}^e + \vec{L}^q = -\overline{(\dot{\mathbf{a}}^r - \mathbf{N}^q) \cdot \mathbf{a}^{rT}} \; \overline{\mathbf{a}^r \mathbf{a}^{rT}}^{-1} \tag{2.18}$$

Here a^r represents the principal components of the original reference state. As the influence of the linear operator in the regime studied is greater than the influence of the non-linear one, correcting the linear operator has a more significant and direct impact on the results. Non-linearity affects the physics of the boundary layer (*Pedlosky*, 1996).

3. The model and the experimental setup

The Equation 2.4 was integrated numerically and the system was evolved in time using a fourth-order Runge-Kutta scheme and a rigid lid approximation. Thus, the spatial dependence of the inhomogeneous term and the dependent variable in the BVE (eq. 2.4) is represented by a truncated series of orthogonal functions, and a system of ordinary differential equations for the time-dependent expansion coefficients is constructed by the Galerkin approximation.

A spectral model simulates two-dimensional non-divergent flows in a rectangular model domain. The model fields were expanded in a double Fourier series, so that the model equations form a dynamical system for the Fourier coefficients. The wave numbers at which the expansion is truncated could be chosen freely. Therefore, it sparks interest to compare results of low and high order models for the same physical system.

All the reference models were run using part of the ICSM series (Interaction Coefficient Spectral Model) routine, specifically the ICSM_OCE, written by *Frisius* (1998) and modified by the author for the specific purposes of this project.

To test the reduction method, simplified ocean models have been divided into more idealized ones. The first one is a wind-driven barotropic model of a square basin with a rigid lid discarding vertical variations (*Dellnitz*, 2000) based on Fourier functions and formulated as a dynamical system. Starting from the simplest ocean model and increasing the complexity, the models studied as reference were:

- *A toy ocean model (Veronis, 1963)*
- *Idealized ocean models of higher complexity (Böning, 1986)*
- *Higher complexity with additional anomalous forcing (Badin et al., 2003; Dewar, 2001)*
- *Higher complexity with additional anomalous forcing and changes in atmospheric conditions*

It is important to understand these selection of ocean models because in the North Atlantic, for example, there are oceanic variabilities that can be explained by studying these kinds of simplified models.

3.1. Description of the model

Some considerations have to be made for integrating numerically eq.2.4 by means of a spectral method. Spectral methods generally involve the approximate representation (usually spatial dependency) of the solution function $\psi(r,t)$ in a finite series of orthogonal functions ϕ_n, which is, namely

$$\psi(\vec{r},t) \sim \sum_{\mu=1}^{n} \psi_\mu(t)\phi_\mu(\vec{r}) \tag{3.1}$$

The spectral method known as Galerkin method involves the transformation of the partial differential equation into a system of ordinary differential equations for the (time-dependent) evolution coefficients, the „spatially truncated spectral equations“, which involves constructing a numerical scheme for ψ_n.

The rectangular basin of the model has free-slip boundaries, that is

$$\psi = 0\,, \nabla^2\psi = 0 \quad \begin{cases} x = 0 \quad , \quad x = L \\ y = 0 \quad , \quad y = L \end{cases} \tag{3.2}$$

To satisfy the boundary conditions 3.2 the fields are represented by a double Fourier sine series. Therefore, 3.1 becomes

$$\psi(x,y,t) \sim \sum_{k=1}^{n}\sum_{l=1}^{n} \psi_{k,l}(t)\sin(kx)\sin(ly) \tag{3.3}$$

and

$$\mathrm{curl}_z\tau(x,y) \sim \sum_{k=1}^{n}\sum_{l=1}^{n} T_{k,l}\sin(kx)\sin(ly) \tag{3.4}$$

$$\mathrm{curl}_z\tau(x,y) \sim T_0\sum_{k=1}^{n}\sum_{l=1}^{n} \sin(kx)\sin(ly) \tag{3.5}$$

Thus, the default parameters of the model are L, H, β, T_0, A and K. The number of parameters can be reduced by non-dimensionalizing the equation. The variables are scaled according to

$$\begin{aligned}(x, y) &= L(x', y') \\ t &= \frac{1}{\beta L} t' \\ \psi &= \frac{T_0}{\beta H} \psi'\end{aligned} \tag{3.6}$$

resulting in the dimensionless form of eq.2.4.

$$\nabla^2 \psi_t + R_O \mathbf{J}(\psi, \nabla^2 \psi) + \psi_x = E_L \nabla^4 \psi - E_B \nabla^2 \psi + \nabla \times \tau \tag{3.7}$$

where

$$R_O = \frac{T_0}{H \beta^2 L^3} = \left(\frac{\delta_I}{L}\right)^2, \quad E_L = \frac{A}{\beta L^3} = \left(\frac{\delta_M}{L}\right)^3, \quad E_B = \frac{K}{\beta L} = \left(\frac{\delta_S}{L}\right) \tag{3.8}$$

and in this case, it is assumed that the source of vorticity due to the curl of the wind-stress is given by

$$\text{curl}_z \tau(x, y) = \hat{k} \cdot (\nabla \times \tau) = \omega_E(y, t) \tag{3.9}$$

The nondimensionalized boundary conditions become

$$\psi = 0\,, \nabla^2 \psi = 0 \quad \text{for} \quad x\ ,\ y = 0\,, \pi \tag{3.10}$$

The variables R_O, E_L and E_B are known as the inertial parameter or Rossby number, the horizontal Ekman number or lateral friction coefficient (*Munk*, 1950) and the vertical Ekman number or bottom friction coefficient (*Stommel*, 1948) respectively. The nondimensional parameters anticipate which fluxes are expected to be important in this process; the frictional widths, δ_M and δ_S, can be compared to the inertial width, δ_I, for which a boundary layer Reynolds number is suitable, defined as the ratio of the inertial advection of relative vorticity to the diffusion of vorticity meaning:

$$R_e = \frac{R_O}{E_L} = \frac{L \delta_I^2}{\delta_M^3} \tag{3.11}$$

3.2. Structure of the wind forcing and the anomaly

In addition to the external parameters, the shape of the forcing (wind stress) strongly affects the model dynamics. Inertial currents are those driven by the wind and kept in movement by the Coriolis force and gravity. And the wind driven turbulence happens in the mixed layer, where most of the activity happens. Knowing that the wind-driven currents affect about 20% of the total volume of the ocean, a wind driven ocean is suitable and important, which is the main focus of the work of *Böning* (1986). A typical wind stress for a box model is constant in time and has a coherent spatial structure of a double ocean gyre. For this work, the basic structure of the wind forcing (3.9) is a one-dimensional sinusoidal function with a structure

$$\omega_{E0}(y) = -\sin(2y) \tag{3.12}$$

Counting on additional external forcings that are used for making a more realistic approach by simulating the effect of the NAO with an anomaly, the total wind-stress curl field was defined as a linear combination of the unperturbed wind-stress curl field component (ω_{E0}) and the anomalous components

$$\omega_E(y,t) = \omega_{E0}(y) + \omega_{\text{ANOM}}(y,t) \tag{3.13}$$

The climate anomalies are modelled by the ω_{ANOM} wind field structure which is latitudinally restricted, in a broad sense similar to the oceanographic setting. The spatial distribution of these anomalies is assumed fixed, in accordance to *Badin et al.* (2003) and *Dewar* (2001).

$$\omega_{\text{ANOM}} = \eta(t)\cos(2\pi y), \qquad \frac{\pi}{4} \leq y \leq \frac{3\pi}{4} \tag{3.14}$$

The temporal part of this input is then defined as two independent parts that represent only a stochastic amplitude and an annual cycle that simulates the NAO. The criteria to chose the amplitude of $\eta(t)$ is determined by the model used, this two possible structures of the amplitude have the form:

$$\eta(t) = \begin{cases} \eta_w = -\dfrac{w_t}{M} \\ \eta_T = -w_t \left|\sin\left(\dfrac{\pi t}{t_T}\right)\right| \end{cases} \tag{3.15}$$

The two terms are defined as $\eta_w = {}^{w_t}/_{M}$, which represents the stochastic atmospheric wind stress anomaly associated with the NAO, and is updated as reported in literature (*Dewar*, 2001), and the periodic side, η_T corresponding to $w_t \left|\sin\left({}^{\pi t}/_{t_T}\right)\right|$, could represent the influence of annual cycle on the stochastic wind stress variability.

Additionally, w_t is defined by the Box-Muller transformation (*Thistleton et al.*, 2007) with the form

$$w_t = \mu + \sigma \left\{-2\ln(Z_1)\right\}^{1/2} \cos(2\pi Z_2) \tag{3.16}$$

where Z_1 and Z_2 are pseudo random numbers, the stochastic variable w_t follows a normal distribution $w_t \sim N(\mu, \sigma^2)$ with mean μ and variance σ, and w_t changes its value stochastically every t_w time period. Throughout all the runs with ω_{ANOM} the chosen values for the mean and variance were left fixed, $\mu = 9 \times 10^{-1}$ and $\sigma = 3 \times 10^{-2}$ (*Visbeck et al.*, 1998). The chosen values for the wind stress profile are similar to the ones reported in the literature; in this specific case $t_T = 2$ years (*Hellerman and Rosenstein*, 1983) and the stochastic part term is updated monthly ($t_w = 30$ days).

3.3. Experimental setup

The division of the simplified ocean models into more idealized ones was carried out by first using a barotropic model driven by the wind in a squared basin using the mentioned initial conditions and starting from the simplest ocean model. The complexity was increased for the chosen reference models. Additionally, for all further setups a rigid lid approximation was chosen, so vertical variations can be safely discarded. For the different runs already mentioned, the set of parameters that were fixed depended on the approach and reference setup as follows.

3.3.1. A toy ocean model

The *Veronis* (1963) model serves as an introduction to the reduction method as an example of a simple dynamic system. When solving the reference model of the barotropic vorticity equation eq.3.17 with few Fourier components, it is easy to select the same spectral number as the number of EOFs. The chosen reference model was a bottom friction approach, in order to ensure sufficient dissipation for given values of E_B and R_O (see Table 3.1). The *Veronis* (1963) model found oscillating solutions, which would represent the simplest solution to reproduce.

$$\nabla^2\psi_t + R_O\mathbf{J}(\psi, \nabla^2\psi) + \psi_x = -E_B\nabla^2\psi + \nabla \times \tau \tag{3.17}$$

For this particular case, the wind stress curl ($\nabla \times \tau$) is approximated as

$$\text{curl}_z\tau(x, y) \sim \sin x \sin y \tag{3.18}$$

Table 3.1.: Parameters Veronis

Parameter	Symbol	Quantity	Unit
Horizontal length	L	$2.0\pi \times 10^8$	cm
Beta parameter	β	2.0×10^{-13}	$cm^{-1}s^{-1}$
Rossby number	R_O	0.3	
Vertical Ekman number	E_B	1.0×10^{-2}	

For the reference run, and for the EOF reduced one, a total of 4 EOFs were chosen to reproduce the 2×2 spectral wave number full model.

3.3.2. Idealized ocean models of higher complexity

As mentioned, an idealized, rectangular ocean model can describe well enough many of the features of large-scale circulation. *Böning* (1986) found statistically steady states solutions for specific values of Rossby number (R_O) and horizontal Ekman Number (E_L).

The simplicity of this approach and the fairly realistic results made the study attractive for ocean modelling, establishing itself as the basis of many studies. And to build an identifiable reference model with even more realistic setups, the calculations were made by means of a double-gyre wind forcing (see eq. 3.12).

In that case, it was found that the energy from the wind forcing was not dissipated by the friction acting on the mean state, instead the energy was transferred to the fluctuating part of the solution by Reynolds stress interaction work. The physical parameters that produce this behavior are depicted in Table 3.2

The spectral number chosen for the reference model was in congruence with what was found in the study by *Böning* where 63×63 wave numbers were shown to be necessary to accommodate the solutions for this particular setup and parameters.

3.3.3. Higher complexity with additional anomalous forcing

Without drifting away from the already known Böning setup for the ocean box and in order to bring it closer to a slightly more realistic mid-latitude realm (*Mildner*, 2013), the chosen parameters for the second tier were similar to the previous one (see Table 3.3), with the addition of the anomalous forcing of eq. 3.14 with amplitudes oscillating stochastically

Table 3.2.: Böning Run

Parameter	Symbol	Quantity	Unit
Horizontal length	πL	2.00×10^{8}	cm
Vertical length	H	1.38×10^{5}	cm
Horizontal viscosity	A_h	5.20×10^{6}	$cm^2 s^{-1}$
Beta parameter	β	2.00×10^{-13}	$cm^{-1} s^{-1}$
Amplitude of windstress	T_0	2.00	$cm^2 s^{-2}$
Rossby number	R_O	1.40×10^{-3}	
Horizontal Ekman number	E_L	1.00×10^{-4}	
Vertical Ekman number	E_B	$- - -$	
Reynolds number	R_e	14	
Spectral wave number	n	63×63	

between 70% of T_0 and T_0, and adding bottom friction with values very similar to the ones reported in the literature for the North Atlantic (*Ierley and Sheremet*, 1995).

Table 3.3.: Böning_2 Run with* and without bottom friction

Parameter	Symbol	Quantity	Unit
Horizontal length	L	$2.0\pi \times 10^{3}$	km
Vertical length	H	1.38	km
Horizontal viscosity	A_h	44.58	$km^2 d^{-1}$
Beta parameter	β	1.73×10^{-3}	$km^{-1} d^{-1}$
Amplitude of windstress	T_0	1.49	$km^2 d^{-2}$
Rossby number	R_O	1.40×10^{-3}	
Horizontal Ekman number	E_L	1.00×10^{-4}	
Vertical Ekman number*	E_B	5.00×10^{-3}	
Reynolds number	R_e	14	
Spectral wave number	n	63×63	

3.3.3.1. Altered atmospheric conditions

In order to test the ability of the method to predict slight variations in the initial conditions, the reduced model was used on different reference models with different initial windstress conditions. The base reference model chosen used the parameters in Table 3.3, but the windstress amplitude of the anomaly (eq. 3.15) was altered for every run by

increasing the anomaly forcing by 10% and 5%, and in other runs with 10% and 5% weaker anomaly (η_T) with respect to the base original (T_0).

This would also mean a change in standard deviation and the mean of w_t (see eq.3.16) to values that would allow the higher limit to not exceed the windstress amplitude T_0, ending up with anomaly values comparable to, or to half of the windstress, which is the case for $\mu = 8.2 \times 10^{-1}$ and $\sigma = 3 \times 10^{-2}$.

As said by *Marshall et al.* (2001b), typical values of the windstress curl for the North Atlantic are around the ones given by Table 3.3 (*Ierley and Sheremet*, 1995) while those correlated with the NAO are smaller than, but comparable to, the values already mentioned in Table 3.3 (*Dewar*, 2001).

This is done mainly to test the ability of the method for the potential use as a predictor of different states due to changes in the atmosphere but keeping the simplicity and clarity provided by the reduced form provided by the EOF reduction.

3.3.3.2. Variations in Reynolds number

The Reynolds number (R_e) is a representation of the nonlinearity of the oceanic model. If the Reynolds number is increased (increasing the influence of the non-linearity) then the center of the circulation cell for a single gyre, for example, shifts to the north. A field of Rossby waves develops, leading to a recirculation vortex in the northwest corner of the basin which is what increases the nonlinearity.

A link between the coefficient of the lateral turbulent friction A_L and instabilities has been reported previously. The smaller the coefficient A_L is, the richer the instability of the physical system. Decreasing diffusivity increased the meridional mass transport in Northern Hemisphere (*Danabasoglu and Mc Williams*, 1995), which reduced the horizontal diffusion coefficient and increased north-south temperature gradients, thus enhancing midlatitude zonal currents (*Meehl et al.*, 1982), diffusion increases with the water temperature.

Using Böning_2 values as a reference (see Table 3.3) are used, and values of δ_M which are approximate values for the North Atlantic. In the case for no-slip conditions, corresponding to the typical North Atlantic subtropical gyre wind forcing, increasing values of Reynolds number R_e means decreasing the viscosity of the system (*Sheremet et al.*, 1997), which in this particular case was done as shown in Table 3.4.

Table 3.4.: Variations in Reynolds number using Böning_2 parameters of Table 3.3

R_e	A_h (km^2/day)	E_L
140	4.46	1×10^{-5}
28	22.29	5×10^{-5}
14	44.58	1×10^{-4}
7	89.17	2×10^{-4}
1.4	445.85	1×10^{-3}

Increasing the Reynolds number of the model by the decreasing the horizontal advection could be used as a form of simulating gradients of temperature in the ocean without using the temperature as a variable *Verron and Jo* (1994).

4. EOF of wind driven ocean models

One important part of reducing wind driven ocean models is establishing the references from which the reduction model is built. The reference model used, starts with an idealized model and progressively is turned into a more realistic one by small steps.

After the clear reference model structure is clarified, the EOF reduction and posterior representation is carried out by using the method in section 2.2. The reconstruction is done with the results of the reductions and then compared to the original full reference model.

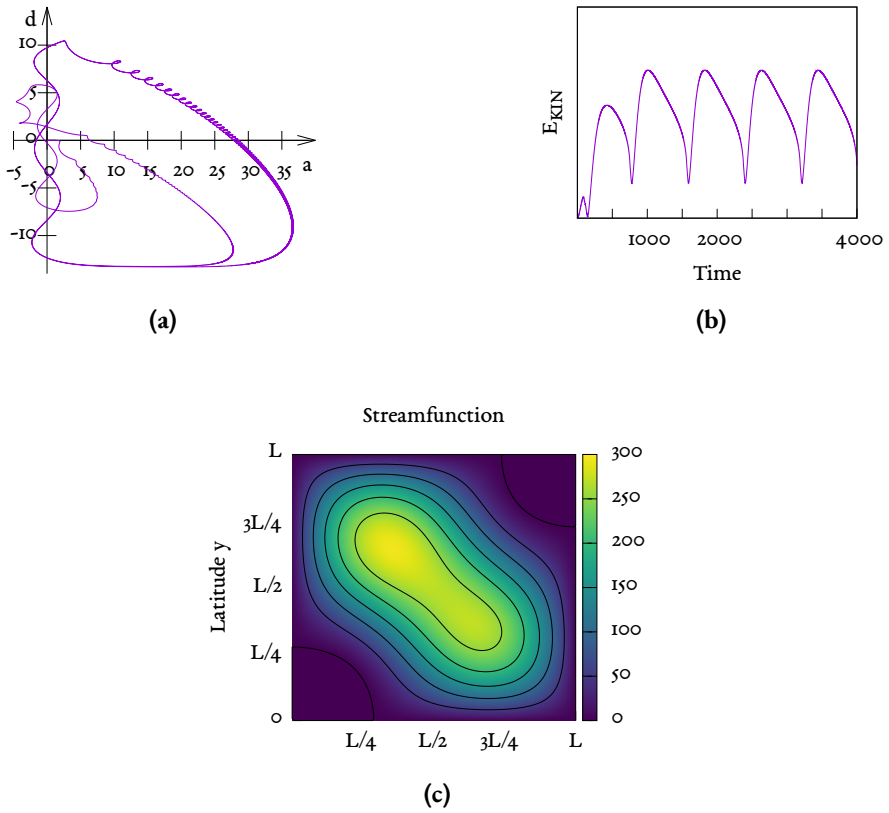
4.1. A toy ocean model

In this section, the work of *Veronis* (1963) is used as an example of a simple dynamic system and as a stepping stone for subsequent calculations and considerations. It is important to be wary of some issues caused by modeling the version of the BVE described by eq.3.17 with only a few Fourier components.

To better determine the reference points, the results from *Veronis* (1963) were taken as starting point. The parameters given in Table 3.1 were proven to provide oscillatory solutions, represented by the limit cycle of the first spectral coefficients (see Figure 4.1a) providing an idea of the structure of the principal components of the system.

The streamfunction is expanded in a double Fourier sine series in the *Veronis* (1963) reference model. For a certain range of the Rossby number (or the strength of the wind stress), the ocean never settles down to a steady state (see Figure 4.1b). After an initial transient phase, a periodic limit cycle appears, which constitutes the necessary variability for the EOF representation.

During the time of the statistically steady state, the system oscillates between two major energetic levels. A mean state is taken as a representative characteristic of the model (discarding the first 800 time steps of the transient) the gridpoint space representation of the streamfunction takes a familiar form (Figure 4.1c) already reported in the literature (*Veronis*, 1963, 1966).

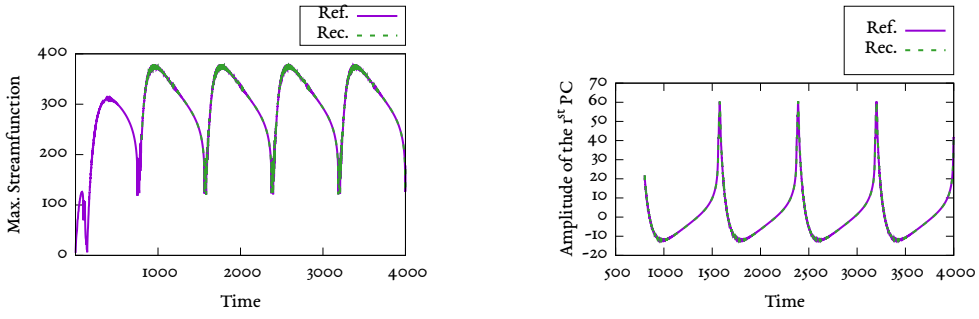


(a) (b) (c)

Figure 4.1.: (a) Phase space contrast of the 1^{st} spectral coefficient vs the 3^{rd}, **(b)** basin integrated kinetic energy vs time, **(c)** mean streamfunction of the *Veronis* (1963) model over 3000 time steps, discarding the initial transient.

4.1.1. EOF representation

Choosing the complete set of available spectral numbers (four), the EOF reduced model in theory should reproduce the entire variability of the system, something that is in congruence with what is obtained. The 1^{st} EOF shows the maximum variability of the reference model, and evolving this map over time by using the principal component described in Figure 4.2b, produces a fair, close to identical, reproduction of the reference model (see Figure 4.2a). This is also visible on the phase space comparison of the first 2 principal components.



(a) (b)

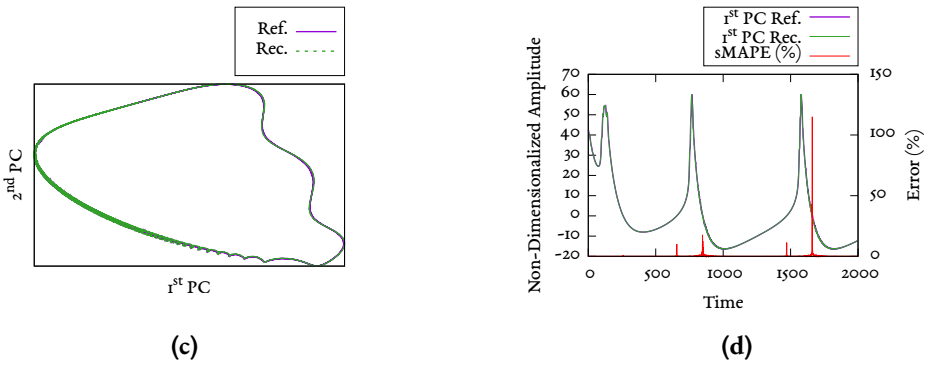


Figure 4.2.: **(a)** Time series of the maximum streamfunction of the *Veronis* (1963) model and the reconstructed data, **(b)** time series comparison of the 1st PC of the reference model (Ref.) and the reduced one (Rec.), **(c)** phase space representation of the 1st and 2nd PC of Ref. and Rec., **(d)** comparison between the 1st PC of the reference and reconstructed models and the symmetric mean absolute percentage error (sMAPE) between them.

It must be clarified that the error calculated when comparing the reference and the reconstructed models is the symmetric mean absolute percentage error, sMAPE (*Morley et al.*, 2018; *Botchkarev*, 2019), defined as

$$\text{sMAPE} = \frac{100\%}{n} \sum_{\alpha=1}^{n} \frac{|R_\alpha - m_\alpha|}{(|R_\alpha| + |m_\alpha|)/2} \tag{4.1}$$

Where R_α is the reference value, m_α the measurement and n the number of measurements. From eq.4.1 the limitation of the sMAPE becomes apparent. If the reference value is 0, the value of the sMAPE spikes to the upper-limit of 200% (*Tofallis*, 2015). This is something that is noticeable in Figure 4.2d.

Using the reconstructed spectral field, it is possible to calculate the kinetic energy of the reduced system and compare it with the original reference system (see Figure 4.3). Just by contrasting them, it is possible to attest to the efficiency of the model. The correspondence between the time series of the kinetic energy of the two fields is almost $1:1$ and yields insight of how promising the method is for stable non chaotic systems.

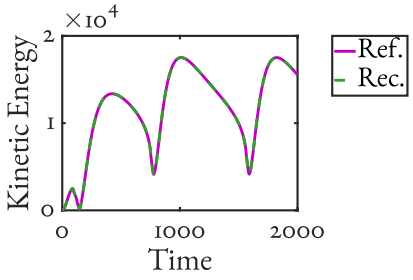


Figure 4.3.: Kinetic energy diagram for the reference model (Ref.) and the reconstructed model (Rec.) from the EOF reduction. Note that both lines are overlapping.

4.2. Idealized ocean models of higher complexity

The following step after the successful reproduction of the *Veronis* (1963) model is setting up the parameters that better resemble a real ocean. The model chosen was the one devised by *Böning* (1986). There, it was found that the needed statistically steady state of the system was only achieved for particular cases and values of the given parameters. The attention of this section is centered on one particular experiment of the ones proposed by *Böning* (1986). In it, the desirable state was reached after an initial spin-up time of around 700 days of fluctuations. After that transitional period, the system settles down into the statistically steady state with regular oscillations (see Figure 4.4a).

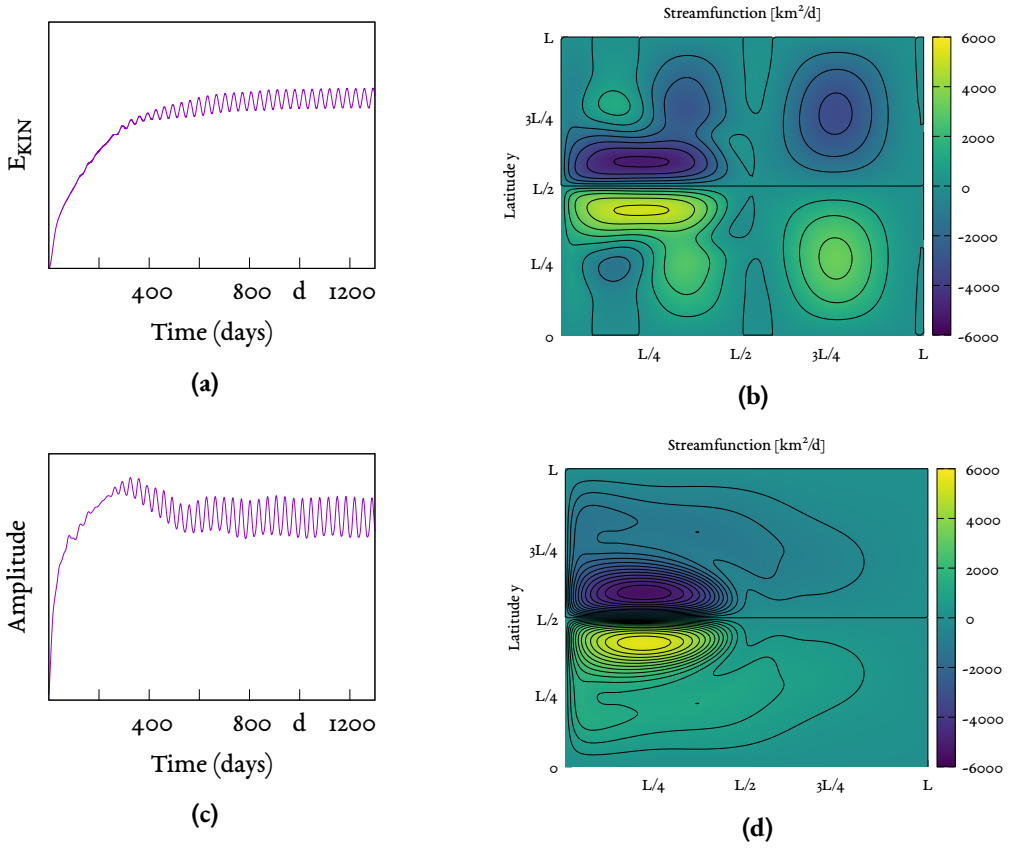


Figure 4.4.: (a) Kinetic Energy in Exp. 14 in *Böning* (1986), **(b)** Stream function at T=4years, **(c)** Normalized Max Stream function over time, **(d)** Mean stream function found as Fig. 17 in *Böning* (1986).

The friction acting on the mean state is not able to dissipate the total wind input of energy, producing the effect present in this particular experiment. It has been implied that the decay times for initial oscillations become longer as the diffusivity becomes smaller (*Böning*, 1986).

4.2.1. EOF analysis

After decomposing the reference model into its principal components and calculating the corresponding eigenvalues, it is possible to generate a variance plot as can be seen in Figure 4.5, which shows that the first 5 PCs contain more than 95% of the variance of the field. This could be interpreted as a clue of the amount of EOFs that could possibly be used to reproduce the majority of the system variability.

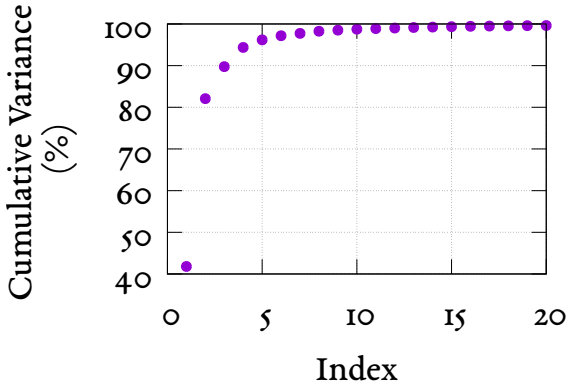
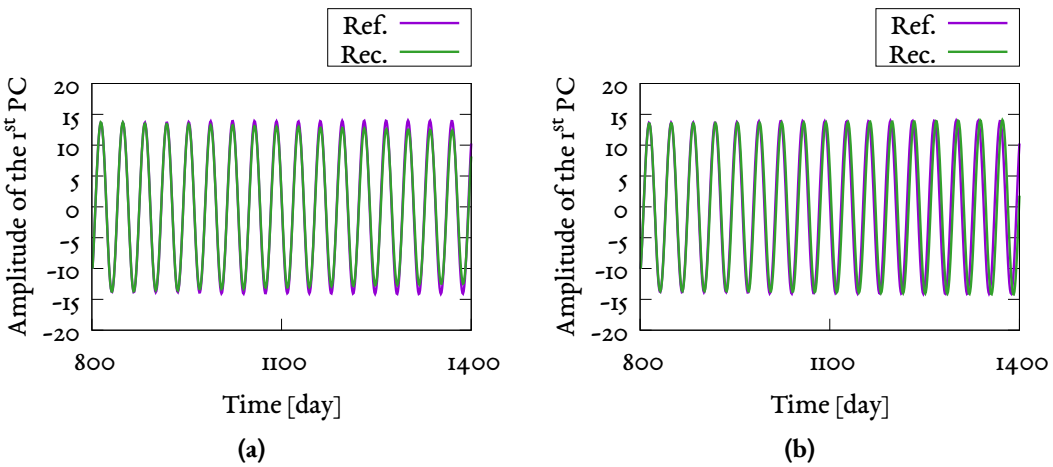
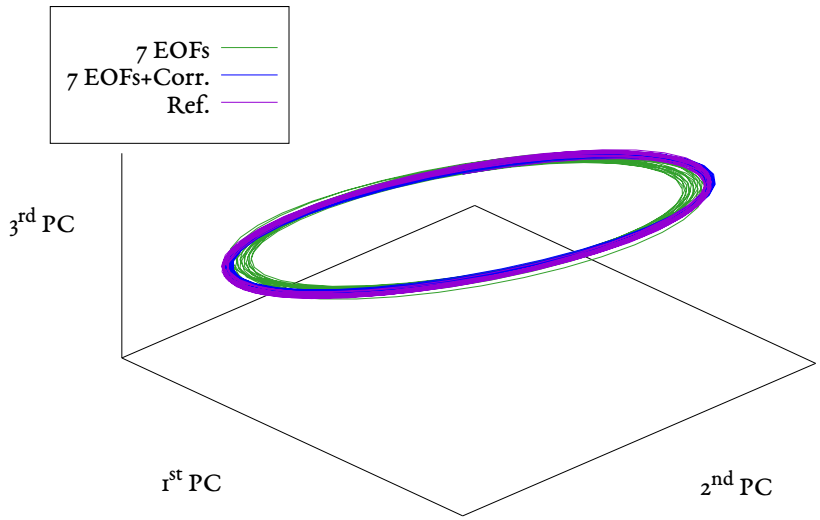


Figure 4.5.: Explained variance of principal components for the *Böning* (1986) model.

When reducing the reference field, it is clear that although the principal components of the reference and reduced models share a similar structure, an increasing discrepancy that grows over time appears (see Figure 4.6a), even when using only 7 EOFs, which could describe around 99% of the variance. This phenomena has been dealt with using various methods. The one chosen is the semi empirical correction proposed by *Achatz and Branstator* (1999), described in Section 2.2.1.





(c)

Figure 4.6.: (a) Comparison between the 1^{st} PC for the *Böning* (1986) reference model (Ref.) and the reconstructed model (Rec.) using 7 EOFs, **(b)** 1^{st} PC of Ref. and 1^{st} PC of Rec. using 7 EOFs with empirical corrections according to *Achatz and Branstator* (1999), **(c)** phase space diagram of first 3 PCs of the Ref., the 7 EOFs model and the 7 EOFs model with corrections.

The correction proposed by *Achatz and Branstator* (1999) significantly improves the behavior of the principal components of the reduced model (see Figure 4.6b). This is also visible in the phase space diagram where the attractors are more similar when the correction is implemented, although initially the correction overcorrects the phase difference. Nevertheless this particularity could be associated to the emerging of a second transient phase that can be avoided by choosing a longer time series.

The correction allows a longer time series for the reduced model and additionally it allows the use of fewer EOFs to reproduce the field as initially thought. Using the information provided by Figure 4.5, it is now possible to choose a smaller number of EOFs (3 in this case) that should suffice to explain at least 90% of the variance.

It is worth of highlighting that the improvement of the reconstruction of the attractor produced in the phase space projection of the first PC of both models is associated with the selection of a longer time series in combination with the empirical correction (Figure 4.7).

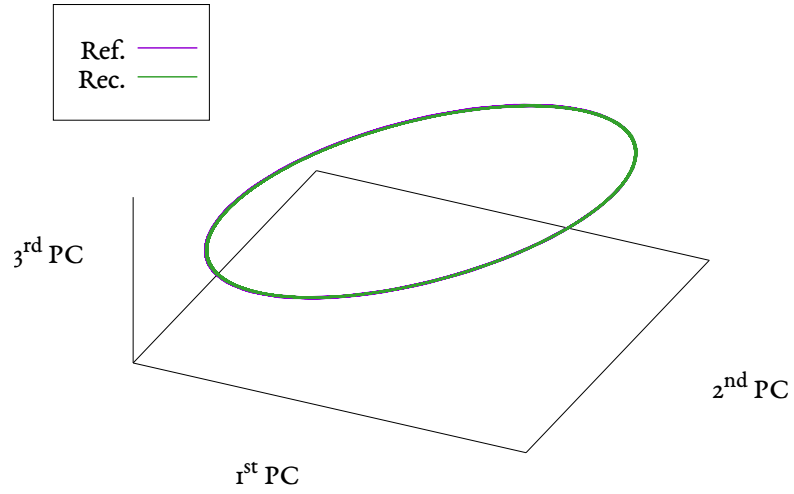


Figure 4.7.: Phase space diagram of the reference and the reconstructed 3 EOFs model.

Redoing the calculations and comparing the new results with Figure 4.6b over a longer time period and using only 3 EOFs, the situation changes (see Figure 4.8). This time the sMAPE error calculation method is shown to be less effective as the PCs oscillate around the zero axis. Note that Figure 4.8 shows only a year of the run so the sMAPE would not overshadow the other principal components.

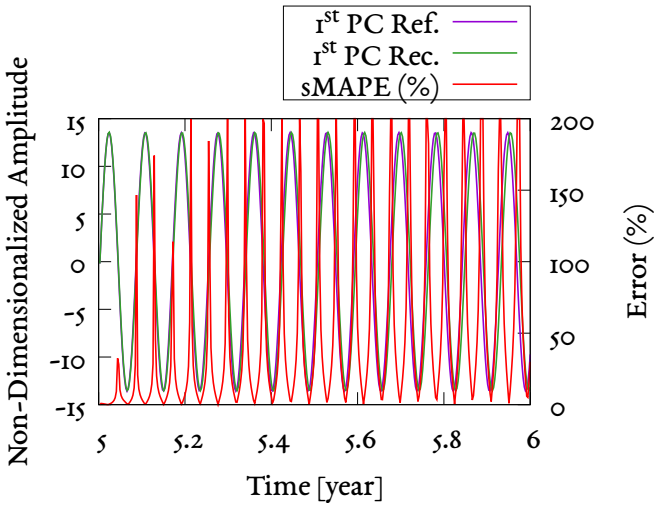


Figure 4.8.: Time series of the 1^{st} PC of: the reference model (Ref.) for *Böning* (1986), the reconstructed model (Rec.) and the error sMAPE between them.

As already mentioned, another valuable characteristic of the models that is useful to compare is the kinetic energy after the reconstruction. An initial comparison is using the power spectrum of the detrended energy time series (see Figure 4.9).

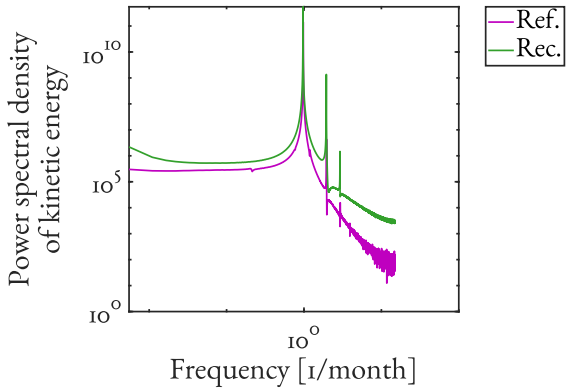


Figure 4.9.: Power spectrum of the basin integrated kinetic energy over time for the reference model (Ref.) and the reconstructed one (Rec.).

Despite the reconstructed model being more energetic than the reference one, it is safe to conclude that the 3 EOF reduced model is able to reproduce many important aspects of the system dynamics. This is further verified by looking at the autocorrelation patterns of their respective kinetic energies (see Figure 4.10a).

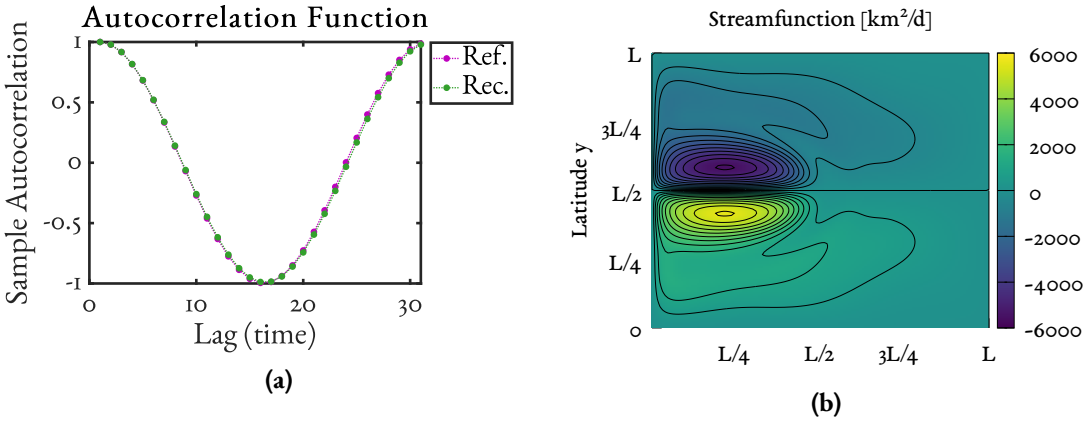


Figure 4.10.: (a) Autocorrelation function of the reference model (Ref.) for *Böning* (1986) and its correspondent reconstructed version (Rec.), **(b)** Mean streamfunction of the reconstructed system using 3 EOFs.

Notice that the autocorrelation functions for both fields behave almost identically for the first lags, hinting at a successful reproduction of the properties of the studied field. Furthermore, the reconstructed model (Figure 4.10a), is able to reproduce the structure of the streamfunction of the reference model (Figure 4.4d).

To further understand the differences between the models, it is useful to look at the statistical parameters of each series (see Table 4.1).

Table 4.1.: Numerical comparison of the maxima (max), mean (μ) and standard deviation (σ) of the kinetic energy for the reference model (*Böning*, 1986) and the reconstructed data and the Bootstrap confidence interval of 90% of the mean (CI).

Parameter	Ref.	Rec.
max	3.87×10^3	3.98×10^3
μ	3.73×10^3	3.70×10^3
σ	100.24	197.10
CI(%) [5 : 95]	[3722.7 : 3726.0]	[3697.4 : 3703.7]

The correspondence between the compared aspects of both systems seem satisfactory, but the sMAPE error calculation falls short when quantifying this congruence. Considering the principal components of the models as two sets of discrete points in the phase space, it is possible to use the optimal transport properties (see Appendix A) to find how much the reconstructed PCs differ from the reference ones (Figure 4.11a).

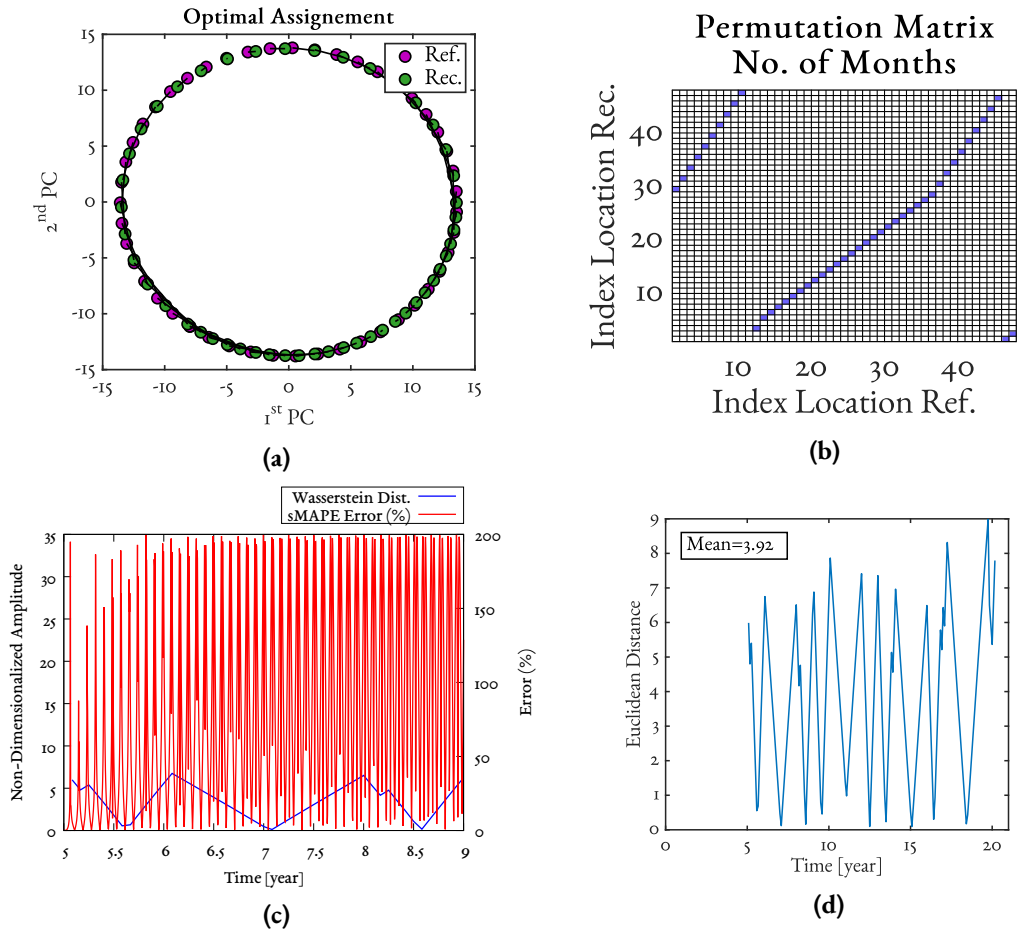


Figure 4.11.: **(a)** Optimal Assignment of the first two PCs of the reconstructed and reference models using the Wasserstein distance algorithm, **(b)** Permutation matrix as a byproduct of the optimal assignment for monthly data of the first 4 years after the transient, **(c)** comparison between the two error methods selected in the time series, **(d)** Monthly calculated Wasserstein distance of the aforementioned phase space diagram segment.

The time series of the principal components were subdivided into segments of 48 months to ease the readability and to better express the optimal assignment. One of the byproducts of the optimal assignment is the permutation matrix (Figure 4.11b) which is the mapping of the reconstructed data point to the correspondent data point in the reference model, and transforming the reconstructed data set into the reference one.

When comparing the performance of the sMAPE with the Wasserstein distance, it is clear that the Wasserstein distance offers stability at points close to zero, which is a significant advantage over the sMAPE. Although the Wasserstein distance seems to oscillate on each subdivision, it does not have a clear periodic pattern. The mean distance of 3.92 of a maximal euclidean distance of 30 corroborates the quality of the reduction model with only 3 EOFs.

4.3. Higher complexity models with additional anomalous forcing

It has been said that the Ekman pumping and Sverdrup transport in the subtropical gyre of the North Atlantic is driven largely by the meridional gradient of the zonal wind stress and it is believed that this is also true even during the last glacial maximum (LGM) times (*Nürnberg et al.*, 2015). This means that when getting the windstress input of the model right, the realistic aspect of the already promising *Böning* (1986) model would improve significantly.

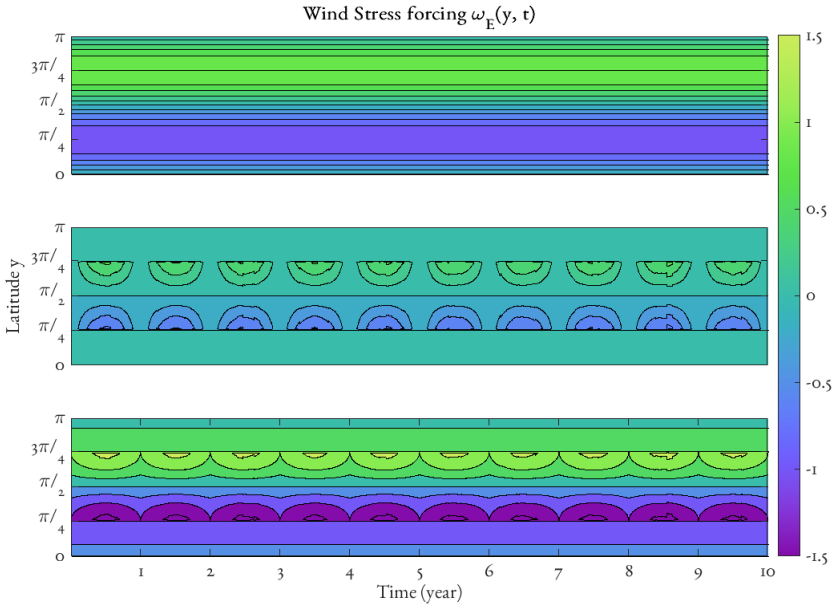


Figure 4.12.: Time series of the windstress forcing for the traditional double gyre (upper panel), the time dependent anomaly that resembles the NAO (middle) and the total forcing when combining those two (lower panel).

A more realistic wind forcing can be obtained by implementing the aforementioned external forcing on the model (*Badin et al.*, 2003). The additional wind forcing is modeled as a white gaussian noise with a structure given by eq.3.14 and updated every day like suggested by *Sura et al.* (2001). The already introduced double-gyre wind profile can be described as the result of a linear combination of an unperturbed component and an anomalous component (see Figure 4.12).

Adding an anomalous windstress forcing with a specific structure to the already established model would resemble a NAO-like phenomena. The traditional windstress proposed is the double gyre described by eq.3.12 (ω_{E0}) as shown in the upper panel of the Figure 4.12, representing the steady windstress over the whole basin. The isolated input of the chosen anomaly over half of the basin, ω_{ANOM} (3.14) is shown in the middle panel and the final structure of the windstress input as a result of the other two in the bottom panel.

This is easily visible in a cross section from north to south through the mean field (see Figure 4.13), where the green line (ω_{ANOM}) changes its amplitude over time.

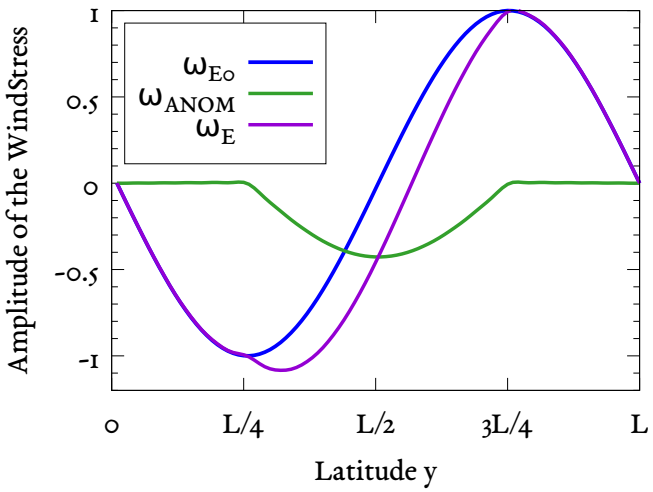


Figure 4.13.: Modelled Ekman pumping of the standard double gyre forcing, the perturbation pumping and the total wind-stress curl field adding the two previous ones.

Thus, the NAO-like structure modelled is more latitudinally restricted than the ocean circulation gyres, similar to other studies of the windstress oceanographic setting, e.g. *Jin* (1997), where the spatial distribution of these anomalies is assumed to be well-structured and fixed but with a temporally stochastic wind forcing.

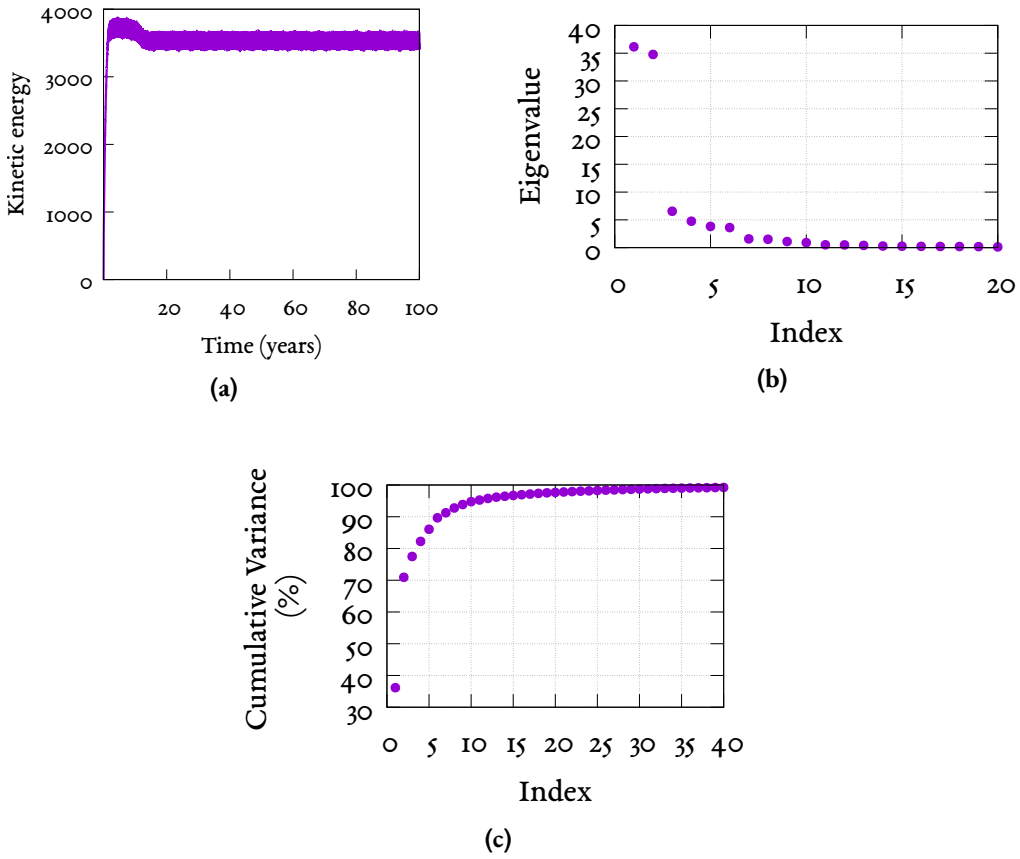


Figure 4.14.: (a) Kinetic energy diagram with the small white noise, **(b)** Eigenvalue spectrum with confidence limits where index is the number of the eigenvalue, ordered from largest to smallest, **(c)** illustration of explained variance as a function of the number of principal components retained out of the total possible.

To verify the efficiency of the method when including a small wind alteration, only the η_w part of the eq.3.15 is used for the anomaly amplitude. The parameters of eq.3.16 chosen for this purpose are: $\mu = 7 \times 10^{-1}$, $\sigma = 7 \times 10^{-2}$, $M = 62$ and updating Z_1 and Z_2 monthly, which translates into a weak anomaly (varying stochastically its amplitude between 0.8% and 1.6% of the reference windstress).

It can be seen that the system does not change much but it is only after 20 years that it stabilizes itself into a statistically steady state, showing the strong influence that the wind input represents in the model (see Figure 4.14a).

Through the study, the criteria selected to evaluate how well the reduced model reproduces the reference one, can be divided into 3 major parts:

- *The kinetic energy of the system*
- *The streamfunction (ψ)*
- *The principal components*

The kinetic energy part consists of directly comparing the time series of the reference model with the reconstructed kinetic energy, its major statistical properties (variance, mean, maximum value), the power spectra resulting from those time series and their corresponding autocorrelation functions. For the streamfunction, the criteria would be a side to side comparison of the gridpoint representation of $\bar{\psi}$ and their maxima.

Finally, evaluating the principal components is done by contrasting the attractors formed by the first three PCs, the time series of the 1^{st} PC against the reconstructed one, and calculating the optimal transport and the respective Wasserstein distance associated to it. If the results of reconstruction do not deviate more than $\sim 20\%$ from the reference in each of the three aspects, then the reduced model is considered satisfactory.

The Wasserstein distance should be interpreted differently for each model. As the cost of finding optimal distances is computationally high and increases exponentially with the amount of data to assign, the times series of the principal components were subdivided into segments of 48 months to facilitate the optimal assignment and the visualization of the data. This is how a temporal quality is associated to the optimal assignment and Wasserstein distances, something that provides an advantage when identifying pattern.

To start the comparison, it is important to know those aspects from the reference model. In the case of the streamfunction, the patterns obtained are shown in Figure 4.15. Comparing Figure 4.15b with Figure 4.4b, it can be seen that the latitudinal symmetry is broken as an effect of the new forcing, but still the system does not devolve into a chaotic state.

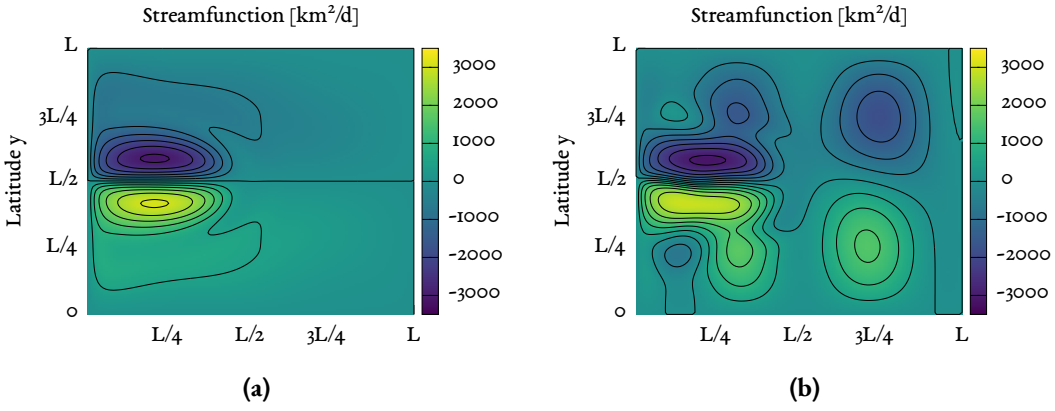


Figure 4.15.: Gridpoint space representation of **(a)** the mean streamfunction of the reference model ($\psi_{max} = 3.81 \times 10^{3} km^2/d$) and **(b)** the streamfunction of the system after a 100 years of evolution.

This, combined with the information provided by Figure 4.14c, a 3 EOF reduced model is assumed to be sufficient to explain the most relevant characteristics of the studied model.

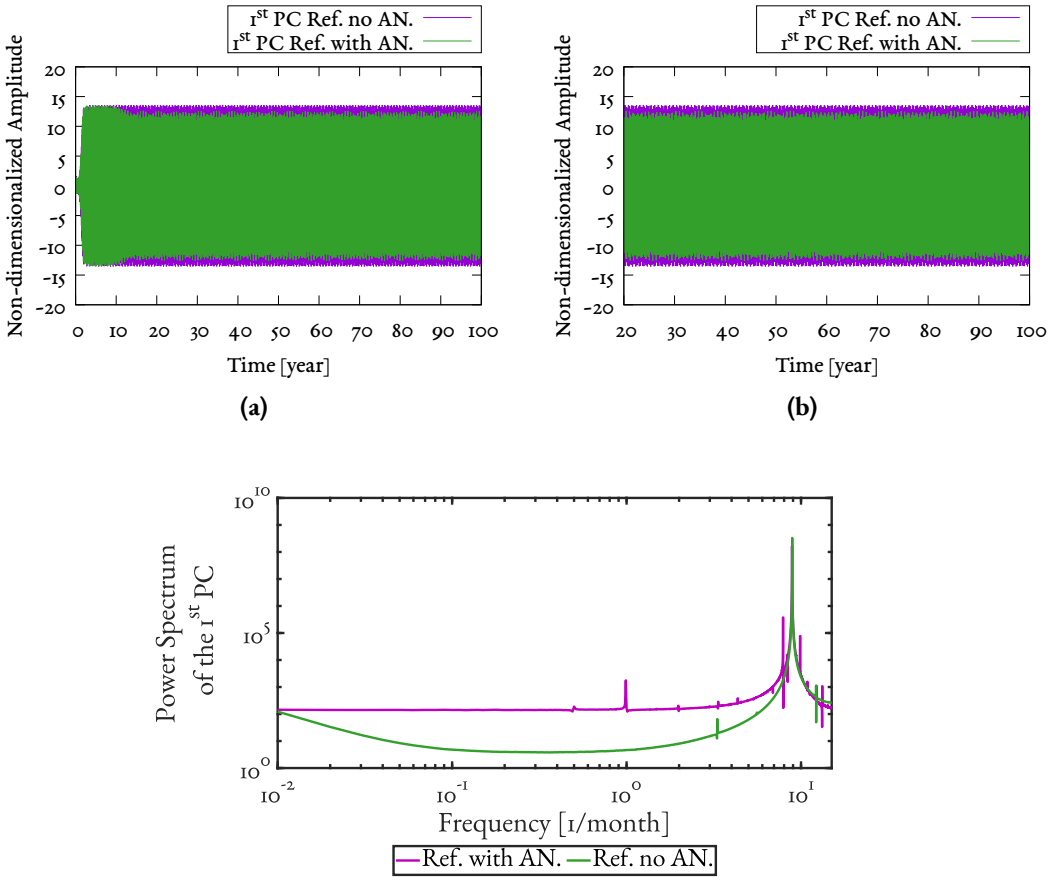


Figure 4.16.: Time series of the 1^{st} PC of the reference model with and without the windstress anomaly η_w for **(a)**, the entire run and **(b)** the chosen time period to analyze, **(c)** power spectra representation of Figure 4.16b.

4.3.1. EOF reconstruction

A total of 3 EOFs were used on the reduced model for the reconstruction. The first 24.1 years shown in Figure 4.14a are considered as a transition period necessary for the system to stabilize and reach the statistical steady state necessary for the EOF reduction. After the reconstruction, it is possible to review all of the criteria one by one.

Starting with the kinetic energy in Figure 4.17a, the reconstructed energy is able to reproduce parts of the fluctuations from the reference system and reaching approximately the same higher energetic values. But for the lower energy states, the reconstructed system goes to values $\sim 5\%$ lower than the minimal energy of the reference.

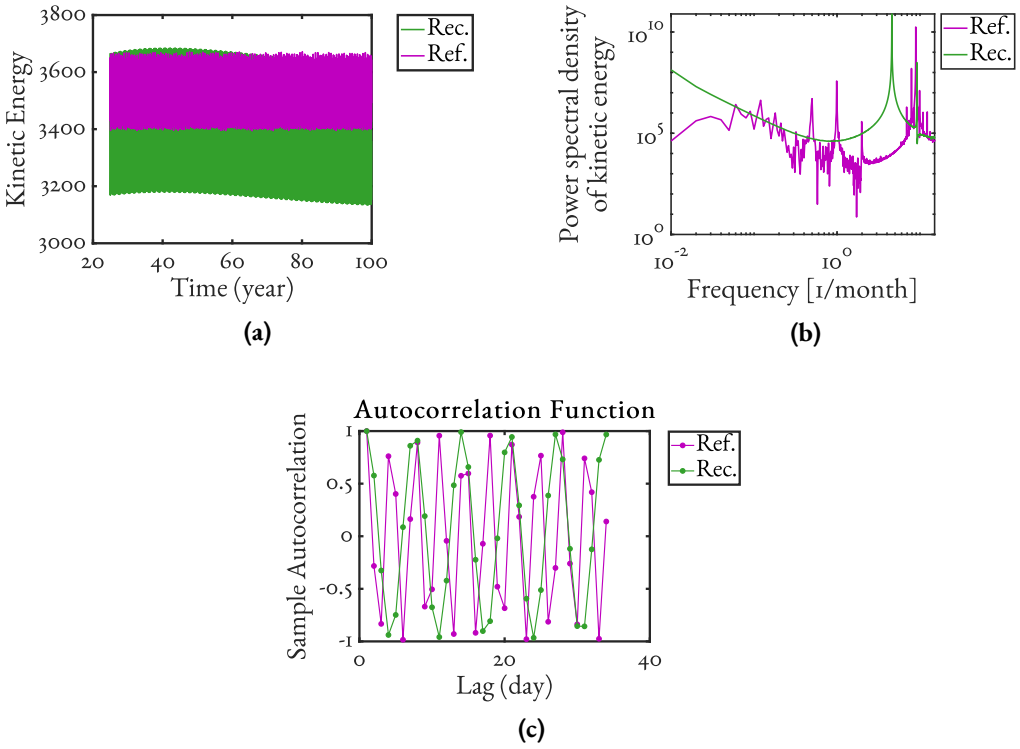


Figure 4.17.: (a) Time series of the nondimensionalized kinetic energy of the reference model and the reconstructed one, **(b)** the power spectra representation of both energies, **(c)** autocorrelation function of the studied kinetic energies.

Figure 4.17b shows that the reconstructed energy is much simpler and smoother than the reference one, which means that higher frequency variabilities are not really represented but it is also seen that they both share some of the same period of variability, and although they do not share the same peak at the same frequency, the overall behavior of the curve is similar.

But by the autocorrelation function it is possible to determine that the reconstructed system seems to skip every second lag, which corroborates the lack of high frequency variability. This is something that could potentially be changed by adding more EOFs to the reduction, depending on what is desired to be reproduced. 3 EOFs seems appropriate to reconstruct the kinetic energy behavior.

Table 4.2.: Numerical comparison of the maxima (max), mean (μ) and standard deviation (σ) of the kinetic energy for the reference model with additional anomalous forcing, and the reconstructed data and the Bootstrap confidence interval of 90% of the mean (CI).

Parameter	Ref.	Rec.
max	3.67×10^3	3.68×10^3
μ	3.53×10^3	3.41×10^3
σ	8.92×10^1	1.77×10^2
CI(%) [5 : 95]	[3528.1 : 3531.2]	[3415.0 : 3421.0]

The maximum value of the reproduced energy deviates only by less than 1% from the reference, and the mean by $\sim$ 4%. The standard deviation on the other hand, is almost double of the reference, which represents the dispersion of the lower energy values.

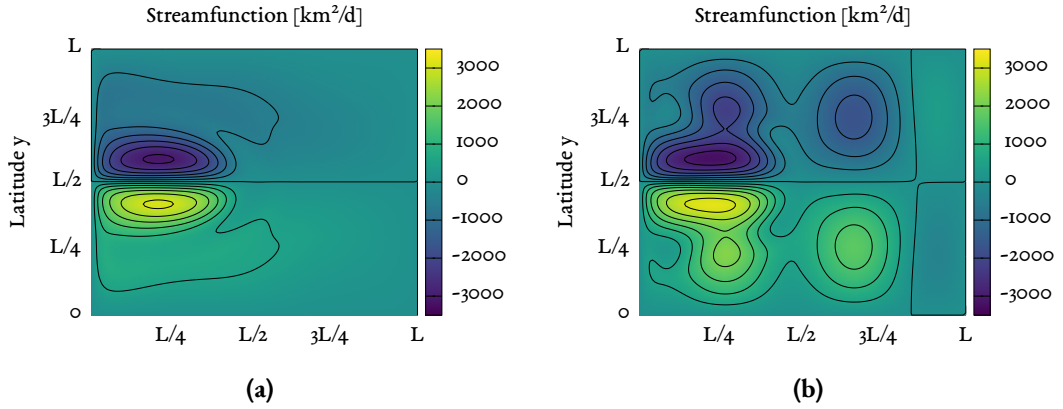


Figure 4.18.: Gridpoint space representation of **(a)** the mean streamfunction of the reconstructed system over the 75.9 years studied ($\psi_{max} = 3.48 \times 10^{3} km^2/d$) and **(b)** the streamfunction of the reconstructed system at the end of the integration.

The streamfunction patterns of the reconstructed version (Figure 4.18a and Figure 4.18b) can be compared to the reference model in Figure 4.15a and Figure 4.15b , respectively. The ψ_{max} of the reconstructed system differs only by $\sim$ 9% from the reference and the structure of both are similar, which means that the main structure of ocean circulation is represented.

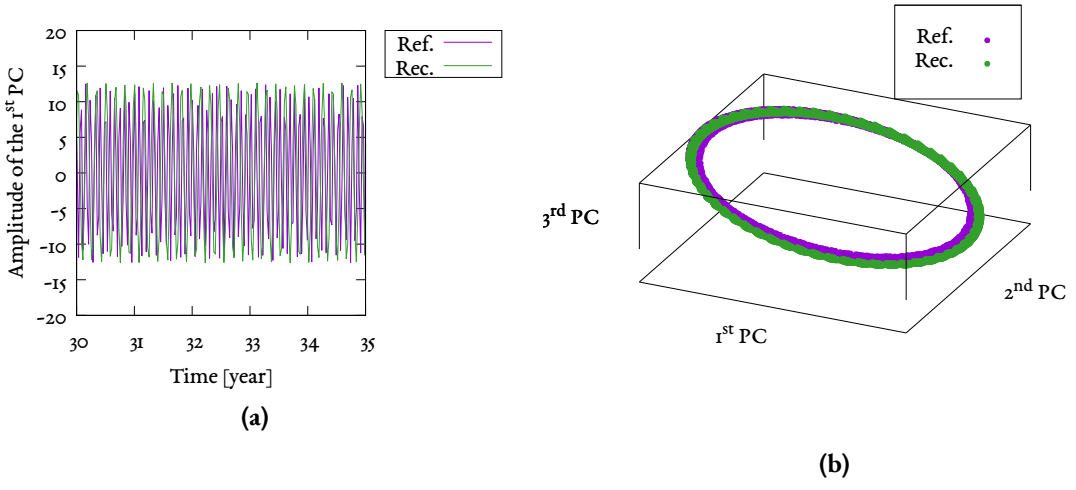


Figure 4.19.: **(a)** Time series of the 1^{st} PC for the reference and reconstructed systems, note that only a period of 5 years out of the 75.9 years available is shown to optimize visibility **(b)** phase space representation of the first 3 PCs.

Figure 4.19a shows the resemblance of the PC is visible and also the lack of representation of

the high frequency variability. But this does not affect the attractor, which reproduces the reference considerably well.

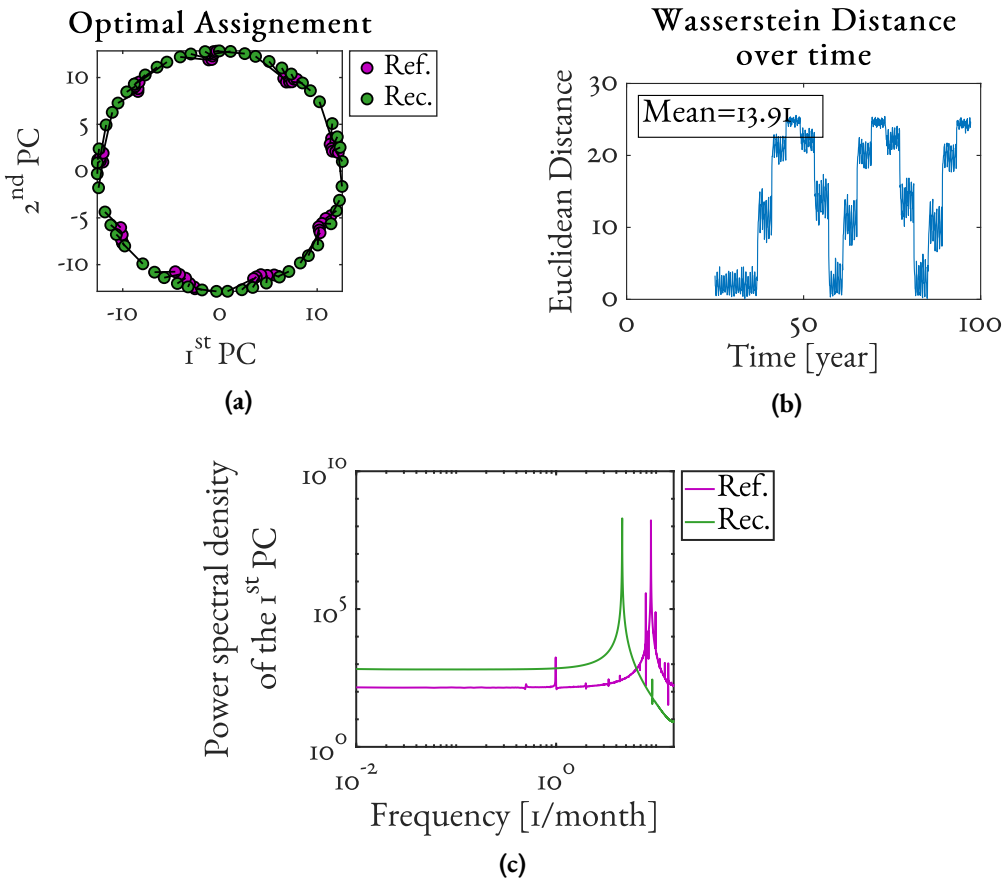


Figure 4.20.: **(a)** Optimal assignment diagram of the first 2 PCs, **(b)** total Wasserstein distance calculated over periods of 4 years, **(c)** power spectrum of the 1^{st} PC.

The optimal assignment diagram (Figure 4.20a) shows that while the PCs of the reference tend to accumulate in certain regions of the phase space , the reconstructed PCs are more evenly distributed. That implies a long term periodicity in the reference model that the reconstruction was not able to reproduce. It can also be seen in the Wasserstein distance (Figure 4.20b) that out of an maximum distance of 30, the mean is about 50% of the maximum, which is deceiving due to the generation of periodicity that resembles a quasi bidecadal oscillation composed of small periods of consistent distance.

This could be interpreted as a sign of a potential weakness of the model in reproducing long term oscillations that might appear from small variations of the wind forcing input. In general, it can be said that the reconstruction is able to successfully reproduce the main aspects of the streamfunction and in general the kinetic energy. For the case of principal components, the system appears to have a fixed phase difference, which affects the other two aspects. Despite that, the model can be considered successful and promising and could be improved further by applying additional corrections to overcome the phase difference.

4.3.2. Different atmospheric conditions

One approach to test the ability of the reduction method to predict different atmospheric conditions is to select a base state for reference, reproduce it, and then use that base state data to reconstruct other systems with slightly different atmospheric forcing input.

It is important to notice that the wind forcing influences to a high degree the overall oceanic circulation and could represent and influence other bigger phenomena. It is shown that large-scale, long-time sea surface temperature (SST) anomalies might be explained naturally as the response of the oceanic surface layers to short-time-scale atmospheric forcing " (*Frankignoul and Hasselmann*, 1977).

Using the parameters in Table 3.3 and η_w (see eq.3.15) with values not higher than the windstress amplitude T_0,the values of μ and σ in eq.3.16 are $\mu = 8.2 \times 10^{-1}$ and $\sigma = 3 \times 10^{-2}$. The EOF reduction is applied at the model with this conditions (from here on referred as Base), and using this data to try to reproduce different atmospheric conditions and compare it with their corresponding reference models

Setting a fixed value of amplitude of the windstress anomaly for the base model (see Section 3.3.3.1), the amplitude of the anomaly for the other atmospheric conditions would oscillate stochastically around that value within $\pm 10\%$ and $\pm 5\%$, see Figure 4.21.

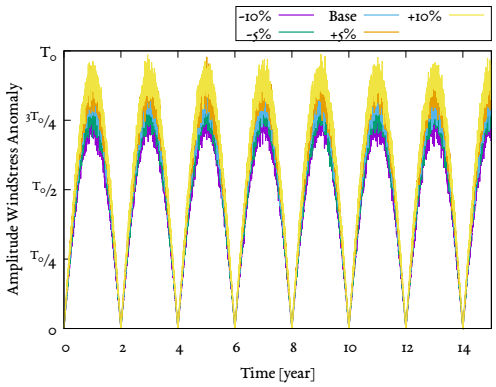


Figure 4.21.: Amplitude of the windstress anomaly η_T (eq. 3.15) for different atmospheric conditions of 90%, 95%, 105% and 110% respect to the base original.

Due to the more chaotic nature of the system, a higher number of EOFs were required for this cases. An idea of it can be drawn from the cumulative valuable plot of the eigenvalues (see Figure 4.22).

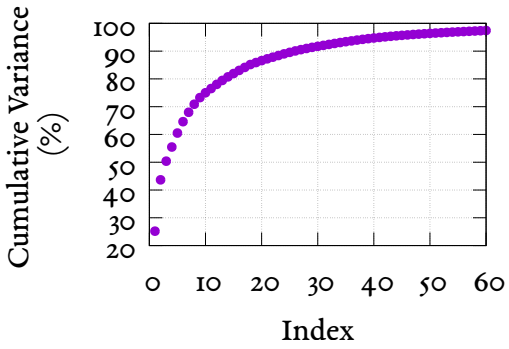


Figure 4.22.: Explained variance as a function of the number of eigenvalues out of the total possible.

The number of EOFs selected for this section was 23. Continuing with the previously discussed criteria and comparing all the kinetic energies reproduced for all the stated atmospheric conditions, Figure 4.23 is obtained.

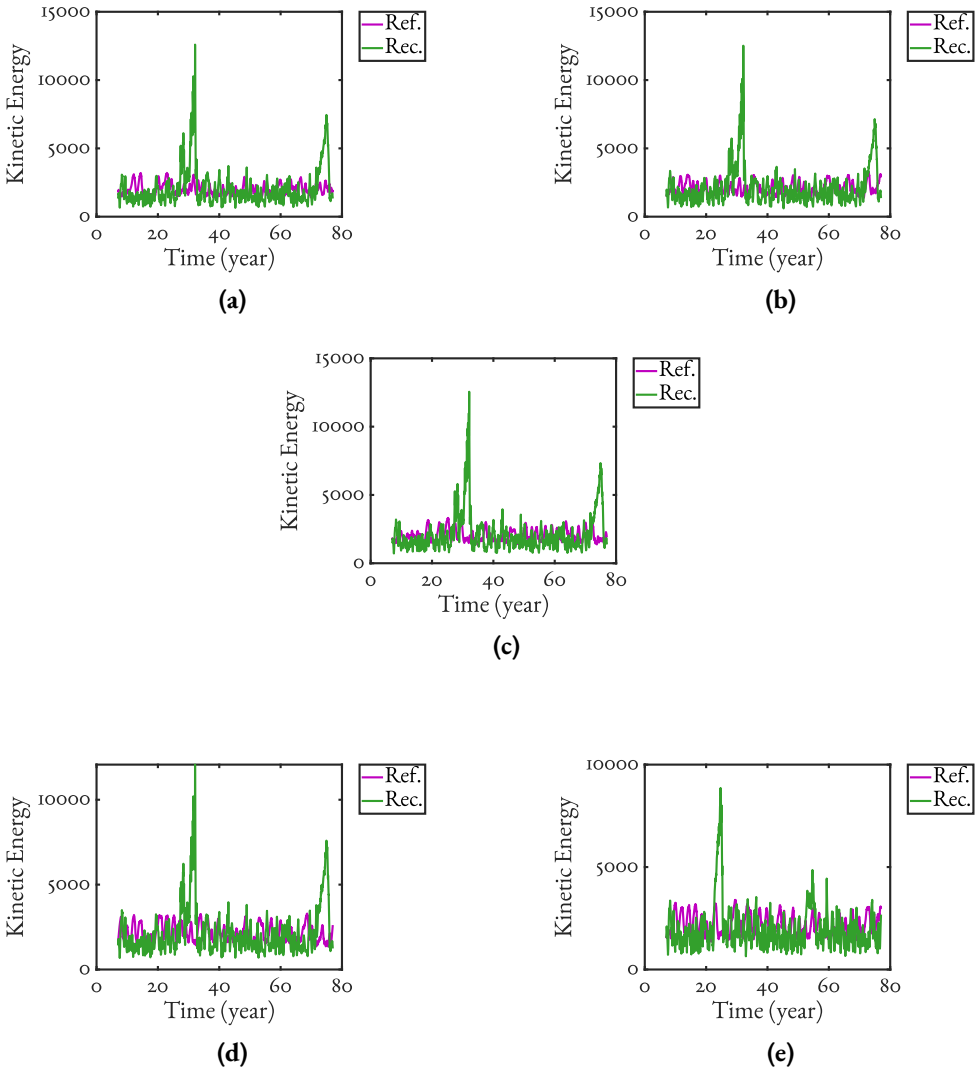


Figure 4.23.: Kinetic energy of the models with changes in the amplitude of windstress anomaly in a magnitude of: **(a)** $+10\%$, **(b)** $+5\%$, **(c)** 0%, **(d)** -5%, **(e)** -10%, with respect to the base model.

All the patterns of the reconstructed kinetic energy have some characteristics in common. Starting with the overestimated mean energy states with respect to the reference, the phase difference in the oscillating part of the energy and the sudden increase of energy of the reduced model between years 20 and 40 and then again at the end of the integration. But, except the spikes, the energy structure is mostly maintained.

Table 4.3.: Numerical comparison of the maxima (max), mean (μ) and standard deviation (σ) of the kinetic energy of 5% and 10% of the amplitude of the windstress anomaly and the Bootstrap confidence interval of 90% of the mean (CI).

	-5%		$+5\%$	
Parameter	Ref.	Rec.	Ref.	Rec.
max	3.30×10^3	1.21×10^4	3.29×10^3	1.25×10^4
μ	2.15×10^3	2.16×10^3	2.08×10^3	2.12×10^3
σ	497.57	1.39×10^3	431.97	1.32×10^3
CI(%) [5 : 95]	[2142.3 : 2158.7]	[2148.0 : 2195.2]	[2069.6 : 2083.9]	[2096.7 : 2139.9]

(a)

	Base	
Parameter	Ref.	Rec.
max	3.32×10^3	1.26×10^4
μ	2.03×10^3	2.16×10^3
σ	431.79	1.36×10^3
CI(%) [5 : 95]	[2023.2 : 2036.9]	[2087.6 : 2131.9]

(b)

	-10%		$+10\%$	
Parameter	Ref.	Rec.	Ref.	Rec.
max	3.40×10^3	8.85×10^3	3.21×10^3	1.26×10^4
μ	2.16×10^3	1.99×10^3	1.98×10^3	2.05×10^3
σ	489.42	1.08×10^3	392.59	1.38×10^3
CI(%) [5 : 95]	[2156.8 : 2173.4]	[1987.0 : 2023.0]	[1970.1 : 1982.4]	[2039.9 : 2084.8]

(c)

Tables 4.4a, 4.4b and 4.4c show that the mean energy states are conserved in the reconstruction, and if it were not for the abnormal spikes in the reconstructed model, the maximum values would be also similar.These spikes are probably due to the reduced model mixing the intrinsic variability between lower and higher EOF modes. It is known that higher EOFs with small eigenvalues can be dynamically relevant and their inclusion or exclusion could cause an interference at the moment of reconstructing the original reference data.

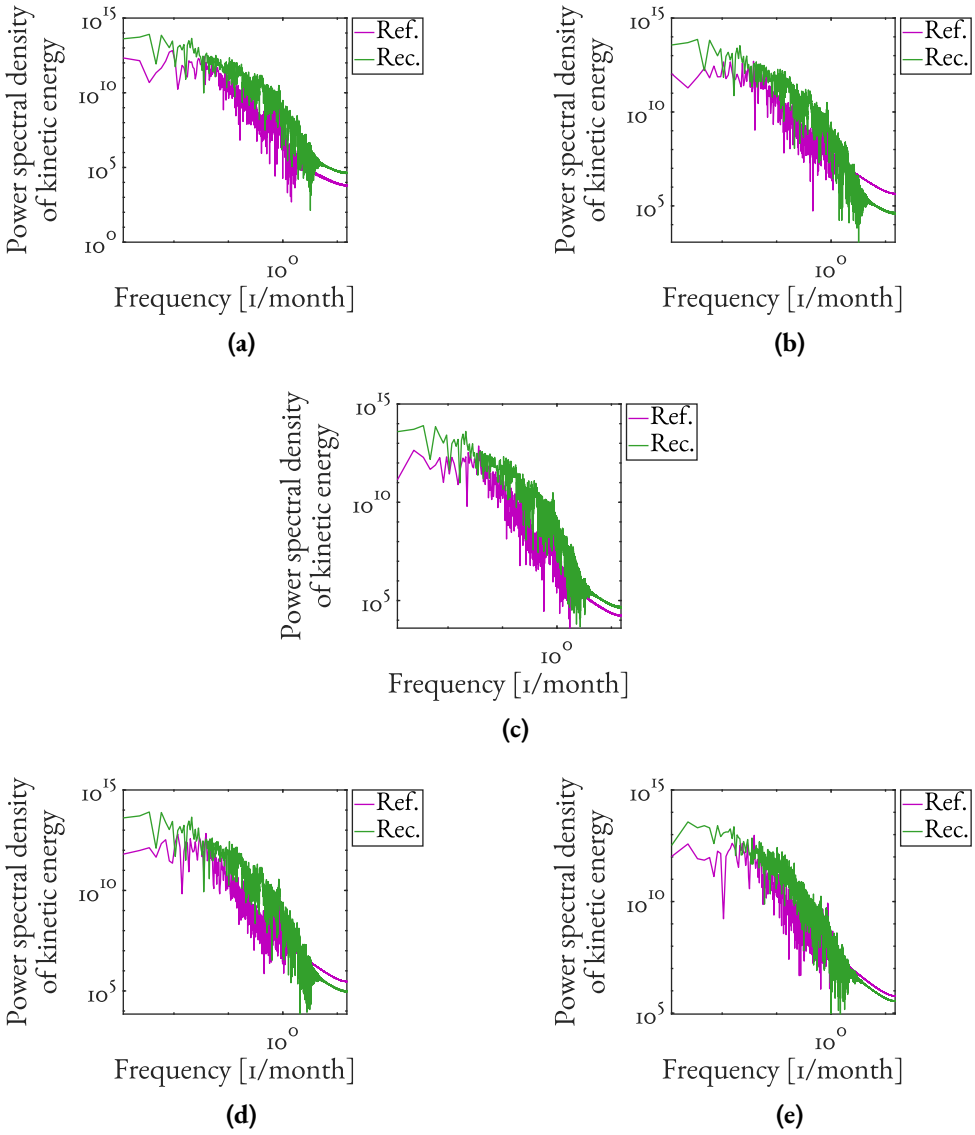


Figure 4.24.: Power spectrum of the models with changes in the amplitude of windstress anomaly in a magnitude of: **(a)** $+10\%$, **(b)** $+5\%$, **(c)** 0%, **(d)** -5%, **(e)** -10%, with respect to the base model.

That can be similarly seen in the power spectra of the corresponding kinetic energies, where the references data and reconstructed data all share the same or similar period of variability only differing in the amount of energy in the system.

The autocorrelation functions associated with the kinetic energies (Figure 4.25) are also consistent and show a similar behavior for the first lags. It is fair to consider the reconstruction of the behavior of kinetic energy as a relative success.

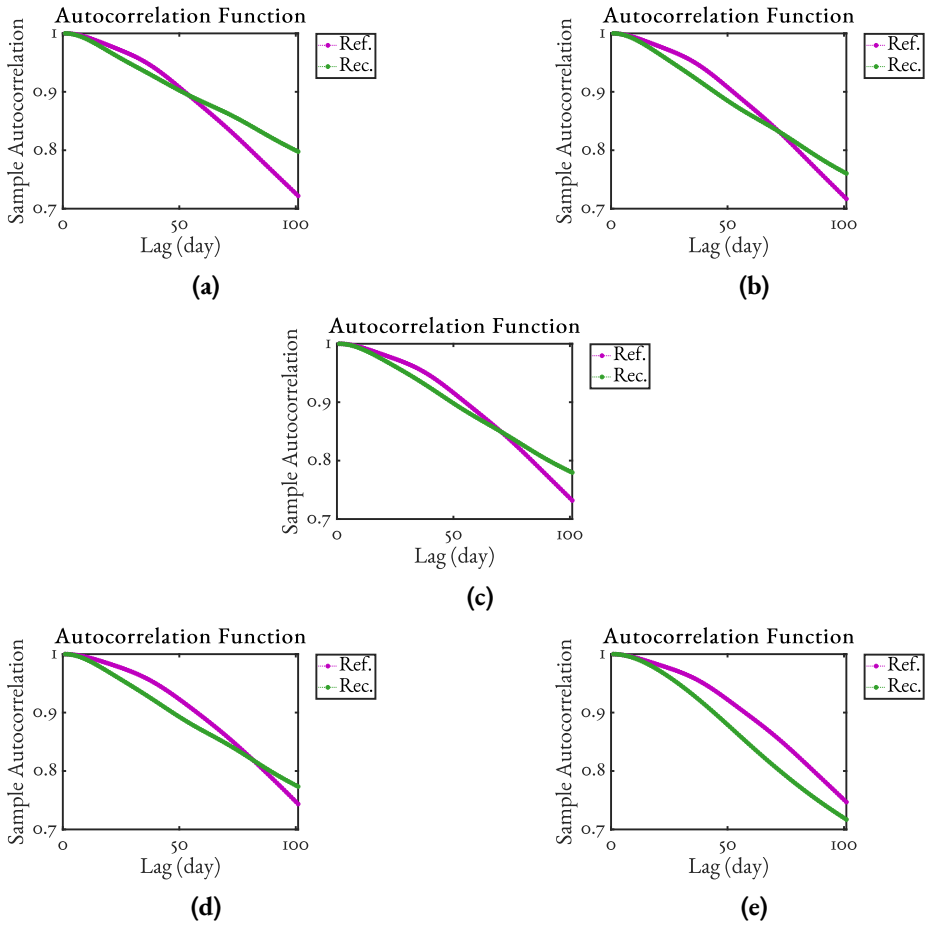


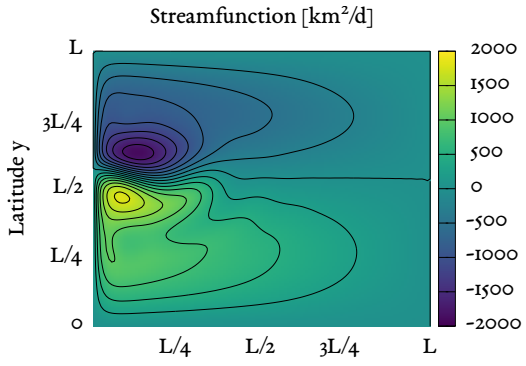
Figure 4.25.: Autocorrelation function of the models with changes in the amplitude of windstress anomaly in a magnitude of: **(a)** $+10\%$, **(b)** $+5\%$, **(c)** 0%, **(d)** -5%, **(e)** -10%, with respect to the base model.

The structure of the streamfunction is for the most part conserved, but the same spike in amplitude is present in the streamfunction. The average values for ψ in the studied period of time are shown in Table 4.4.

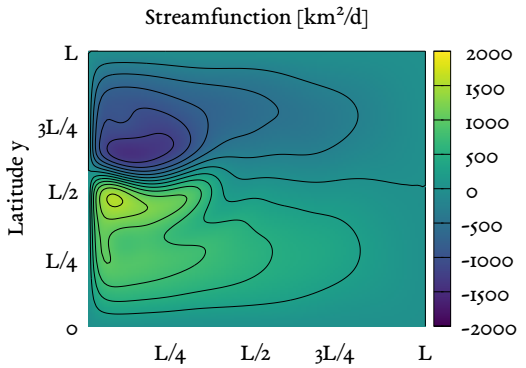
Table 4.4.: Mean streamfunction over 75.9 years of integration for the given atmospheric conditions in km^2/d.

	Ref.	Rec.
$\bar{\psi}_{-10\%}$	0.402×10^2	0.328×10^2
$\bar{\psi}_{-5\%}$	0.424×10^2	0.300×10^2
$\bar{\psi}_{base}$	0.446×10^2	0.316×10^2
$\bar{\psi}_{+5\%}$	0.469×10^2	0.346×10^2
$\bar{\psi}_{+10\%}$	0.492×10^2	0.368×10^2

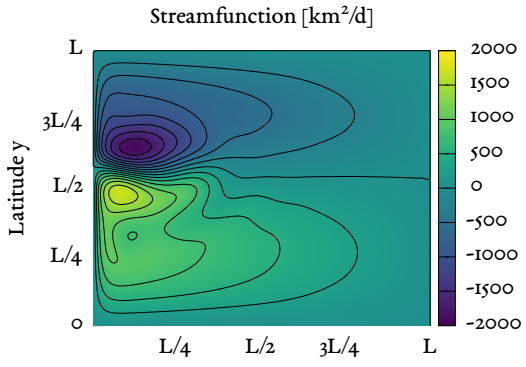
The model seems to better handle the small variations when reconstructing ψ. There are small differences in the circulation patterns and slightly less intense streamfunctions, but in general the structure and the ranges are similar. And the differences increase with the increments of the amplitude of the windstress anomaly (see Figure 4.26). But staying within the same range, corroborating that for ψ, the method is effective.



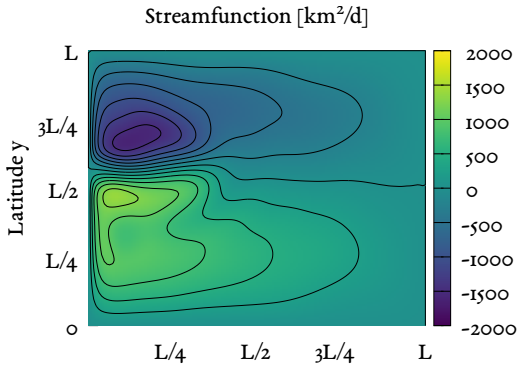
(a) Ref. -10%



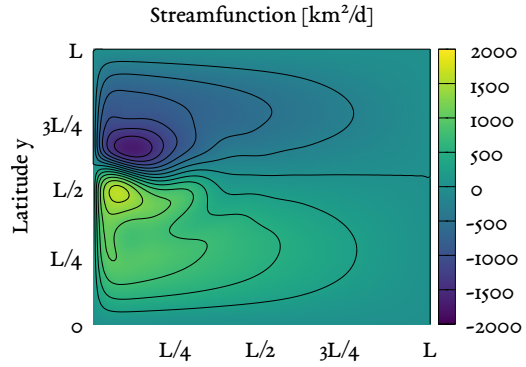
(b) Rec. -10%



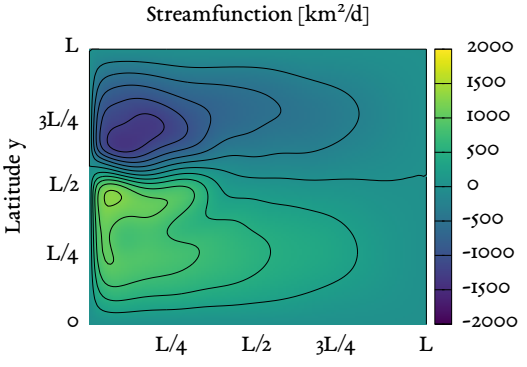
(c) Ref. -5%



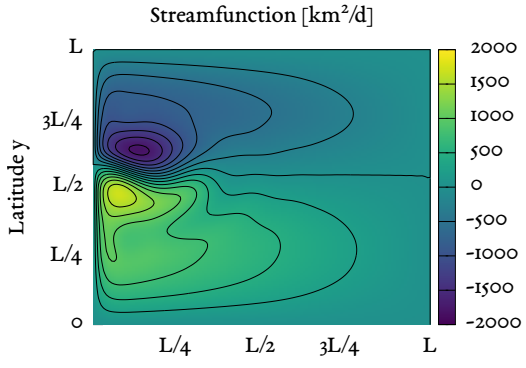
(d) Rec. -5%



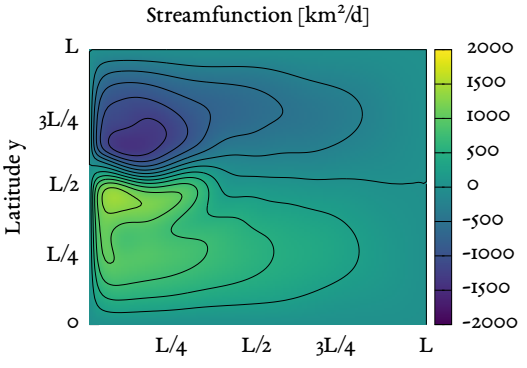
(e) Ref. Base



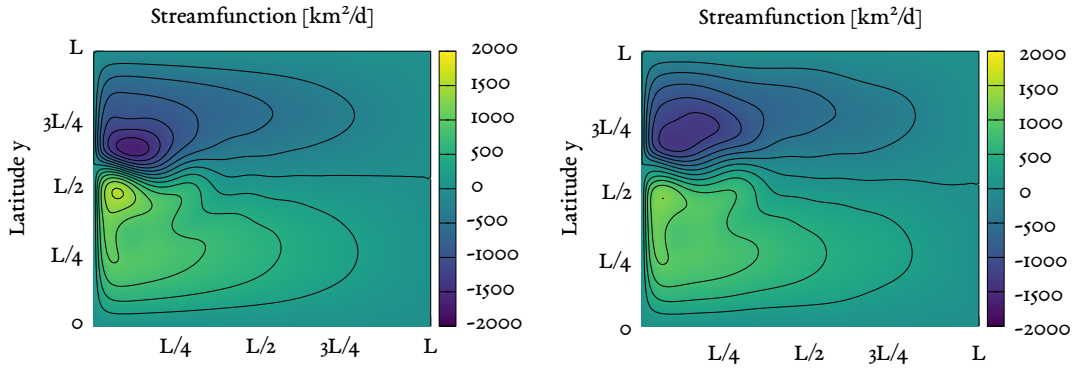
(f) Rec. Base



(g) Ref. $+5\%$



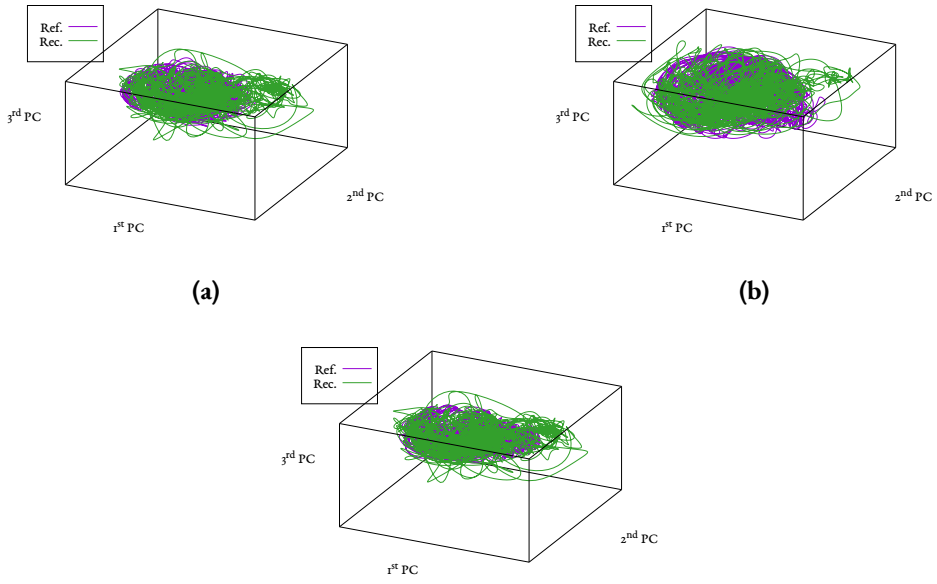
(h) Rec. $+5\%$



(i) Ref. $+10\%$ **(j)** Rec. $+10\%$

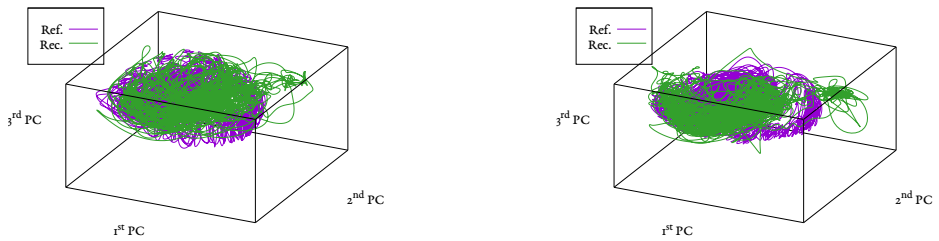
Figure 4.26.: Mean streamfunction over 75.9 years of the reference and reconstructed models for changes in the amplitude of windstress anomaly in magnitudes of: **(a)** and **(b)** $= -10\%$, **(c)** and **(d)** $= -5\%$, **(e)** and **(f)** $= 0\%$, **(g)** and **(h)** $= +5\%$ and **(i)** and **(j)** $= +10\%$, with respect to the base model.

After corroborating a successful reconstruction of the streamfunction, the criteria left to compare is the behavior of the principal components. Comparing the attractors formed by the first three PCs for all the reference models and their respective reconstructed versions, it is possible to see that the system is comparatively more chaotic than the previous analyzed models. This leads to chaotic reconstructions of the PCs with lower agreement (see Figure 4.27).



(a) **(b)**

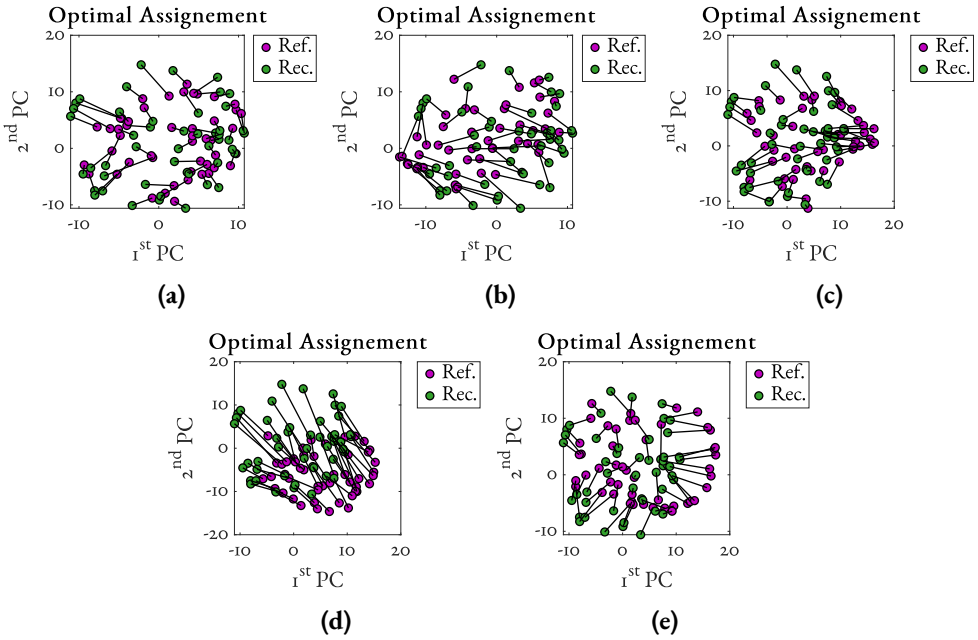
(c)



(d) **(e)**

Figure 4.27.: Phase space representation of the reference and reconstructed models for changes in the amplitude of windstress anomaly in magnitudes of: **(a)** $+10\%$, **(b)** $+5\%$, **(c)** base, **(d)** -5%, **(e)** -10%.

The attractors are not well represented (see Figure 4.27), and it can be further confirmed by the optimal assignment between the first 2 PCs where the principal components tend to cluster in the middle of the phase space. This clustering means that the 2 PCs carry similar information about the variability, which is something that does not happen for the reconstructed model.



(a) **(b)** **(c)**

(d) **(e)**

Figure 4.29.: Optimal assignment formed from the 1^{st} PCs and 2^{nd} PCs in the phase space for the reference and reconstructed models with changes in the amplitude of η in magnitudes of: **(a)** $+10\%$, **(b)** $+5\%$, **(c)** base, **(d)** -5%, **(e)** -10% of T_0.

The behavior can be seen as well in the pattern of the Wasserstein distance (Figure 4.30). which has no clear periodic structure but rather a noise-like one.

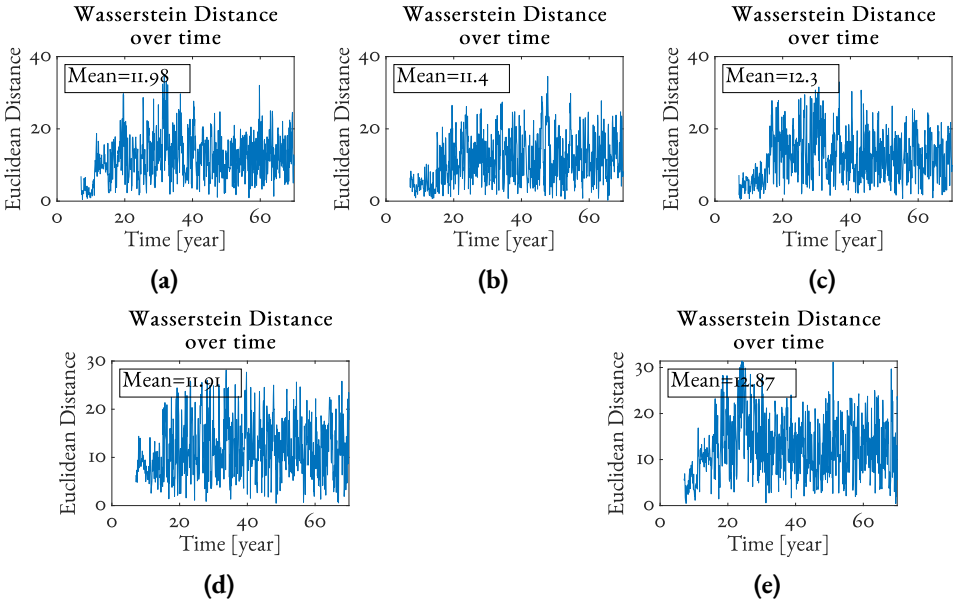


Figure 4.30.: Wasserstein distance between the 1^{st} PCs and 2^{nd} PCs of the reference models and reconstructed models for changes in the amplitude of η in magnitudes of: **(a)** $+10\%$, **(b)** $+5\%$, **(c)** 0%, **(d)** -5%, **(e)** -10%.

The distances seem high (all the mean values are circling 20% of the maximum distance) but for the initial 4 years of calculations, the distances are much smaller (Figure 4.31).

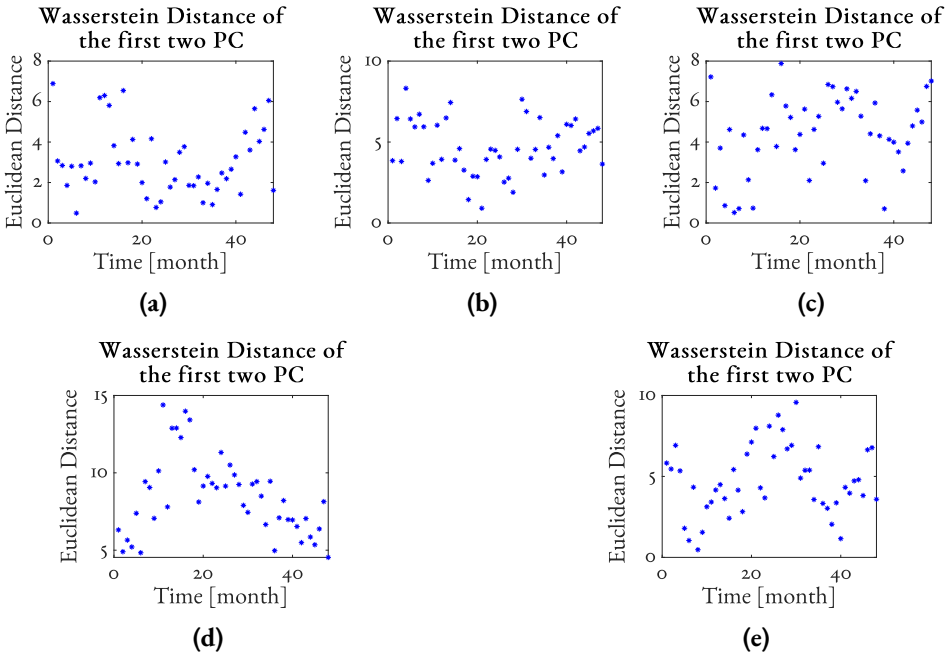


Figure 4.31.: Wasserstein distance of the first months between the 1^{st} PCs and 2^{nd} PCs of the reference models and reconstructed models for changes in the amplitude of η in magnitudes of: **(a)** $+10\%$, **(b)** $+5\%$, **(c)** base, **(d)** -5%, **(e)** -10%.

4.3.3. Reynolds number variation

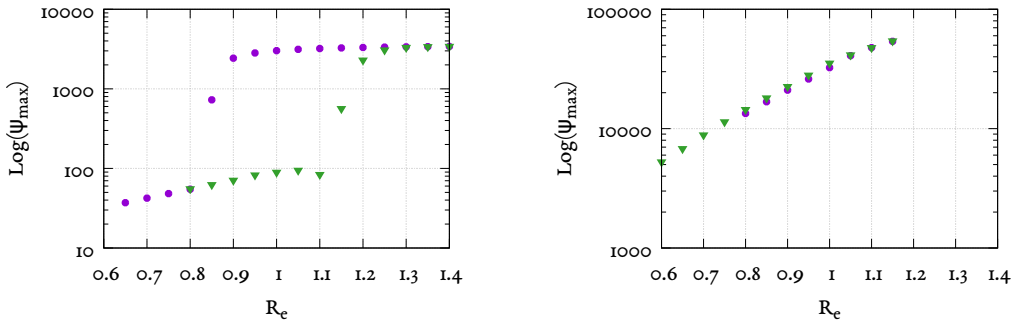
One of the known characteristics of the ocean box models is that the wind-driven double-gyre circulation in a defined rectangular basin goes through several dynamical regimes. One way to produce these is by decreasing the lateral friction in the system. But generally in the box models, the bottom friction is very efficient in dissipating the vorticity produced by the wind (*Pedlosky*, 1996).

Additionally, the regimes are also present when the Reynolds number is increased (*Speich et al.*, 1995). Increasing the Reynolds number of the model by increasing the Rossby number would mean changes in the vertical depth of the model. A single gyre windstress forcing was used (eq. 3.18) together with lateral friction $E_L = 9 \times 10^{-4}$ producing what in the research from *Böning* (1986) is called *experiment 4*, leaving the lateral Ekman number constant and R_O as:

Table 4.5.: Variations in Reynolds number using the Rossby number

R_e	R_O	R_e	R_O
0.60	3.36×10^{-3}	0.90	7.54×10^{-3}
0.65	3.94×10^{-3}	0.95	8.40×10^{-3}
0.70	4.57×10^{-3}	1.00	9.31×10^{-3}
0.75	5.24×10^{-3}	1.05	1.03×10^{-2}
0.80	5.96×10^{-3}	1.10	1.13×10^{-2}
0.85	6.73×10^{-3}	1.15	1.23×10^{-2}

As a starting point, the comparisons are made between a known bifurcation point in this setup and then the anomaly is included (Figure 4.32a).



(a) **(b)**

Figure 4.32.: (a) Bifurcation diagram of a *Böning* (1986) model without anomaly, **(b)** maxima of the streamfunction as a function of a changing Reynolds number for the *Böning* (1986) model with anomaly.

By adding the anomalous external forcing, the final states of the systems are disrupted and the new forcing dissipates the bifurcation that was present; nevertheless, the system does not reach a steady state but a statistical steady state that allows us to use the EOF reduction.

One can see that by the addition of the anomalous forcing, the bifurcation point vanishes and the maximum value of the streamfunction is almost linearly related with the increase in Reynolds number. Selecting some representative points, analogous to the same model without anomaly, it is possible to see that the structure of the streamfunction does not change substantially.

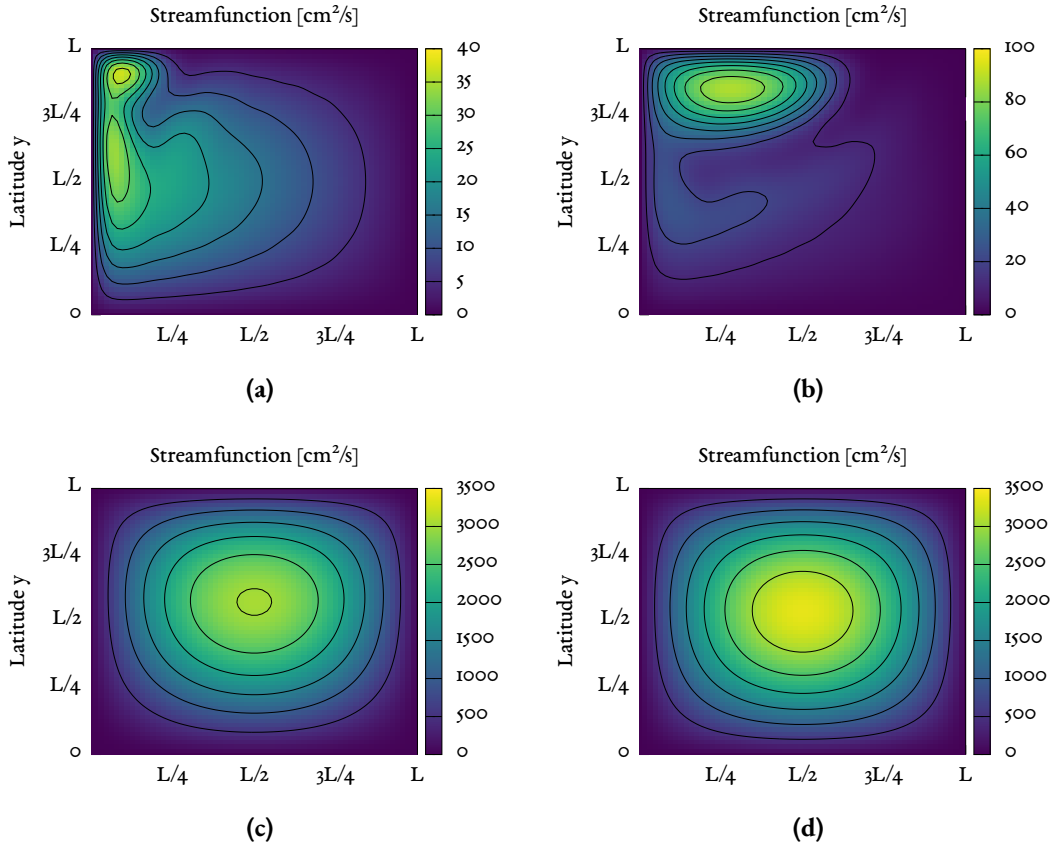
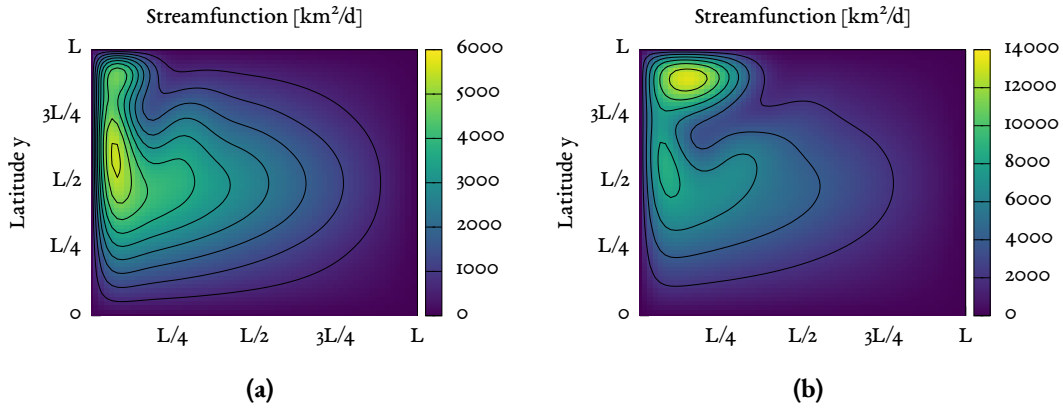
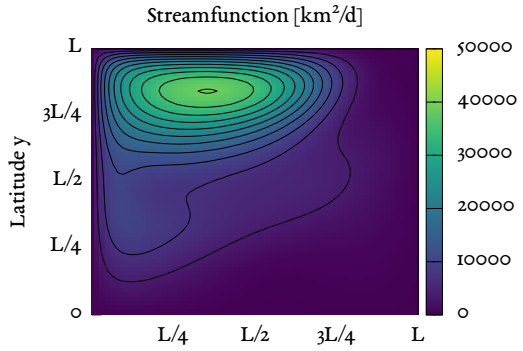


Figure 4.33.: Gridpoint representation of the mean streamfunctions of the Figure 4.32a for values of **(a)** $R_e = 0.65$, **(b)** $R_e = 1.0$ (▼), **(c)** $R_e = 1.0$ (●), **(d)** $R_e = 1.3$.

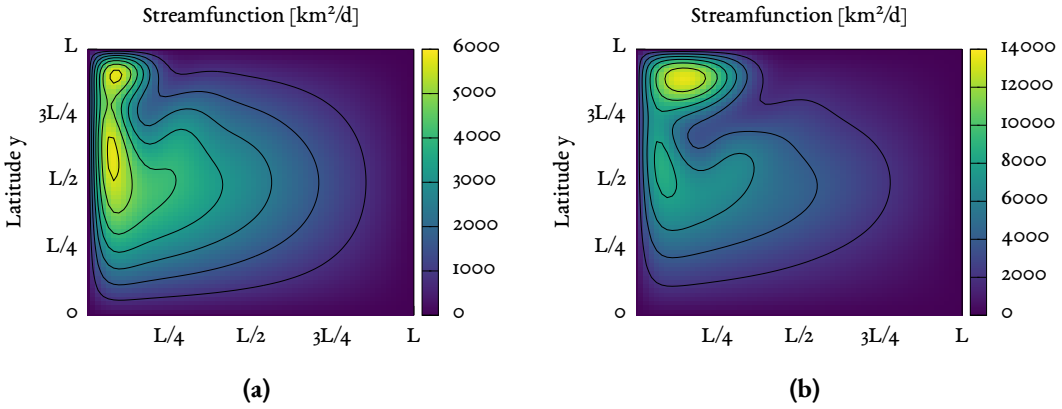




(c)

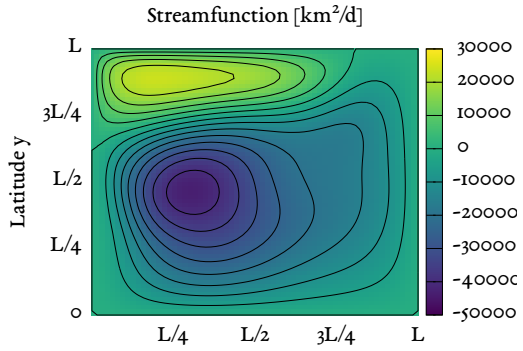
Figure 4.33.: Gridpoint representation of the mean streamfunctions of the Figure 4.32b for values of **(a)** $R_e = 0.65$, **(b)** $R_e = 0.85$, **(c)** $R_e = 1.15$.

Reconstructing those states for the chosen values of R_e, it is possible to see that the reduced model was able to reproduce different parts of the reference model but not all at the same time. In the case of the streamfunction, the resemblance seems uncanny for small Reynolds numbers ($R_e < 1$).



(a)

(b)



(c)

Figure 4.35.: Gridpoint representation of the mean streamfunctions of the reconstructed form of the systems shown in Figure 4.33 for values of **(a)** $R_e = 0.65$, **(b)** $R_e = 0.85$, **(c)** $R_e = 1.15$.

The first two principal components have different behavior depending on the Reynolds number. The phase space representation together with the optimal assignment and Wasserstein distances show that the correspondence is not the best and for systems with a higher chaotic behavior the reduced model loses accuracy.

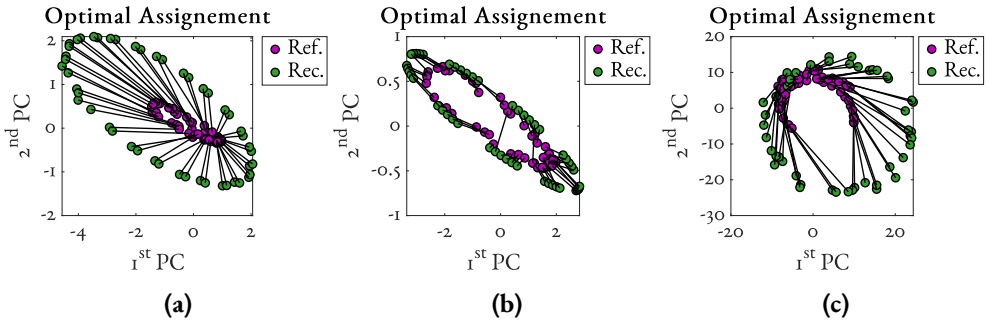


Figure 4.36.: Optimal assignment formed from the 1^{st} PCs and 2^{nd} PCs in the phase space for the reference and reconstructed models with changes of **(a)** $R_e = 0.65$, **(b)** $R_e = 0.85$, **(c)** $R_e = 1.15$.

Having in mind that the maximal Wasserstein distance is given by the span of the axis in the phase space. In the case of $R_e = 0.65$, all the distances over time were above 50% of the maximum available. Assuming that an acceptable distance would be smaller than 20% of the maxima, the optimal distance is not ideal.

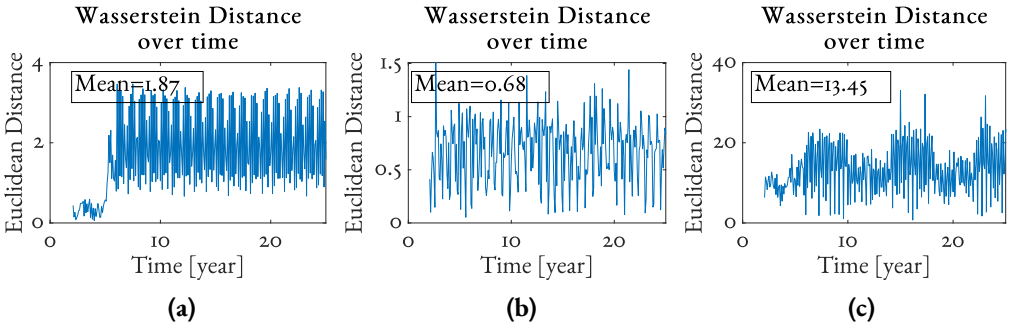


Figure 4.37.: Wasserstein distance between the 1^{st} PCs and 2^{nd} PCs of the reference models and reconstructed models for changes in: **(a)** $R_e = 0.65$, **(b)** $R_e = 0.85$, **(c)** $R_e = 1.15$.

But this does not seem to hold for the kinetic energy, where the reduced model overestimated all of the points and seem to add an additional transient phase.

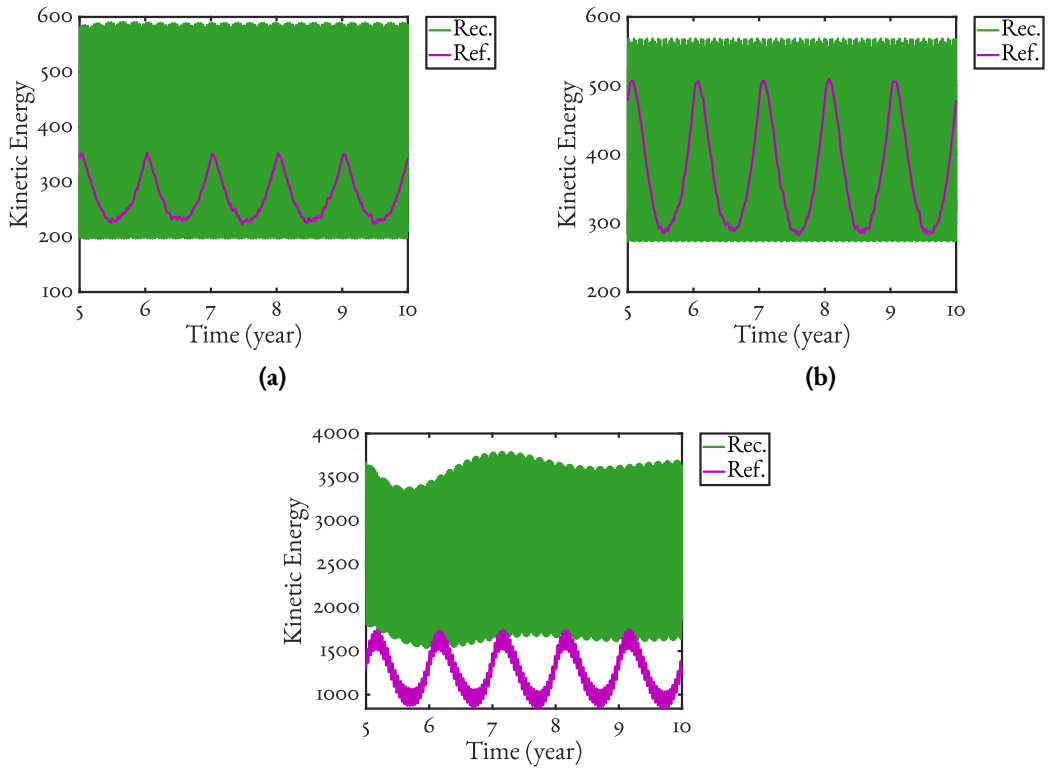


Figure 4.38.: Kinetic energies of the reference models and their respective reconstructed models for values of **(a)** $R_e = 0.65$, **(b)** $R_e = 0.85$, **(c)** $R_e = 1.15$.

Table 4.6.: Numerical comparison of the maxima (max), mean (μ) standard deviation (σ) and the Bootstrap confidence interval of 90% of the mean (CI) of the kinetic energy for the reference model and the reconstructed model for different values of R_e due to changes in R_O.

	$R_e = 0.65$		$R_e = 0.85$	
Parameter	Ref.	Rec.	Ref.	Rec.
max	3.70×10^2	6.21×10^2	5.70×10^2	5.87×10^2
μ	2.72×10^2	3.61×10^2	3.73×10^2	3.92×10^2
σ	3.91×10^1	1.38×10^2	7.57×10^1	1.05×10^2
CI(%) [5 : 95]	[272.1 : 273.3]	[358.1 : 362.7]	[372.1 : 374.5]	[390.1 : 393.4]

(a)

	$R_e = 1.15$	
Parameter	Ref.	Rec.
max	1.76×10^3	4.29×10^3
μ	1.25×10^3	2.60×10^3
σ	2.56×10^2	7.62×10^2
CI(%) [5 : 95]	[1252.0 : 1260.6]	[2581.7 : 2606.0]

(b)

The dominant frictional effect plays an important role in the oceanic circulation, and this transition can be seen when analyzing the power spectra of various runs with and without

bottom friction, with and without anomalous forcing. The effect of the bottom friction is undoubtedly important, making the system go to an early steady state and to produce a lot less energetic systems.

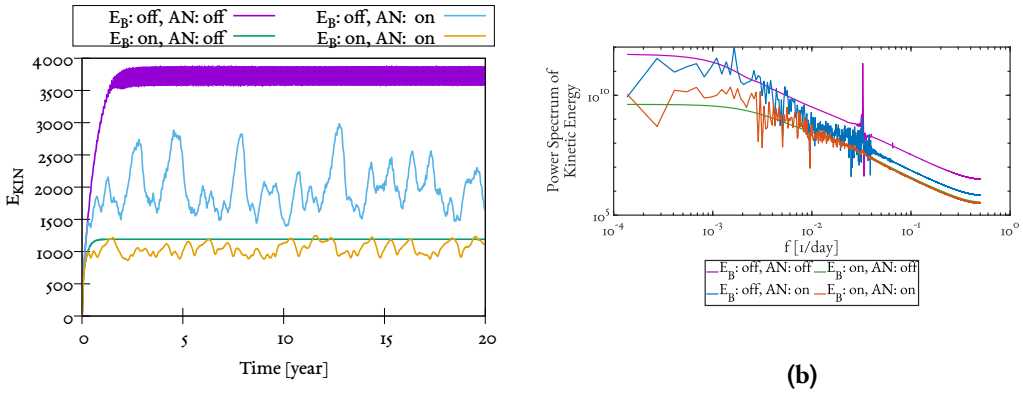


Figure 4.39.: (a) Kinetic energy diagram for different reference models with changes in E_B and a windstress anomaly of the form of η_T with $t_T = 2$ years, **(b)** power spectral density representation of the kinetic energies of Figure 4.39a.

Therefore the reference models taken were the ones with bottom friction ($E_B = 5 \times 10^{-3}$), allowing the system to be a little less energetic and better defined. The EOF analysis was carried on with an initial base state where $R_e = 14$, and then doubling it and increasing the Reynolds number by ten times by means of the E_L. The runs for $R_e = 1.4$ and $R_e = 7$ for this specific setup have a persistent tendency to reach a steady state, limiting the EOF reduction to only the values of $R_e = 14$ and higher (see Figure 4.40).

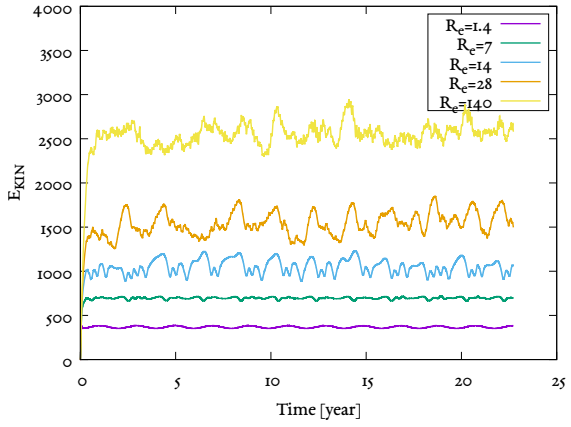


Figure 4.40.: Time series of the integrated kinetic energy for models with different R_e driven by changes in E_L.

Of the three setups studied in this part, only the $R_e = 140$ seems to produce reasonable reproduction of the kinetic energy. The other two overestimated the values by several orders

of magnitude, a problem that seems consistent on every model that includes E_B.

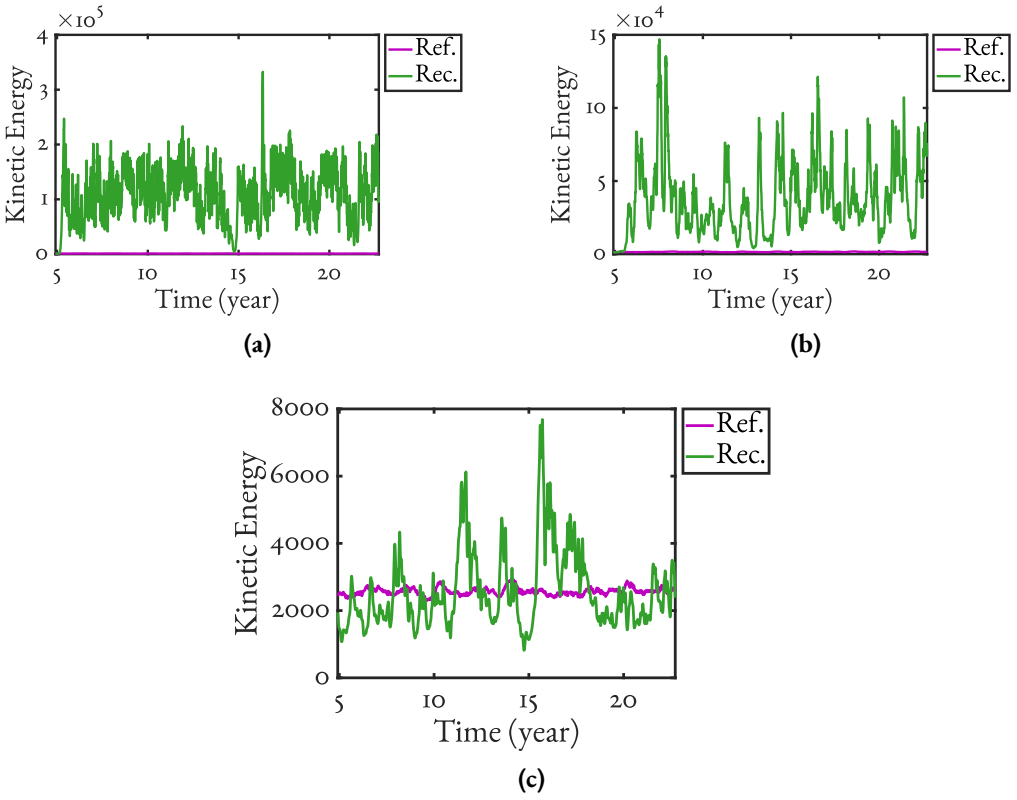


Figure 4.41.: Time series of the kinetic energy of the models with different R_e via changes in E_L in multiples of **(a)** $R_e \times 1$, **(b)** $R_e \times 2$, **(c)** $R_e \times 10$.

Table 4.7.: Numerical comparison of the maxima (max), mean (μ) and standard deviation (σ) and the Bootstrap confidence interval of 90% of the mean (CI) of the kinetic energy for the reference model and the reconstructed model for different values of R_e due to changes in E_L by factors of **(a)** $R_e \times 1$, $R_e \times 2$ and **(b)** $R_e \times 10$

	$R_e \times 1$		$R_e \times 2$	
Parameter	Ref.	Rec.	Ref.	Rec.
max	1.23×10^3	3.33×10^5	1.47×10^5	1.85×10^3
μ	1.05×10^3	1.07×10^5	3.78×10^4	1.55×10^3
σ	8.05×10^1	4.56×10^4	2.24×10^2	2.48×10^4
CI(%) [5 : 95]	[1053.1 : 1055.8]	[2556.8 : 2595.8]	[107501.7 : 108980.4]	[37233.2 : 38088.0]

(a)

	$R_e \times 10$	
Parameter	Ref.	Rec.
max	2.94×10^3	7.69×10^3
μ	2.57×10^3	2.75×10^3
σ	1.09×10^2	1.29×10^6
CI(%) [5 : 95]	[2566.9 : 2570.5]	[2556.8 : 2595.8]

(b)

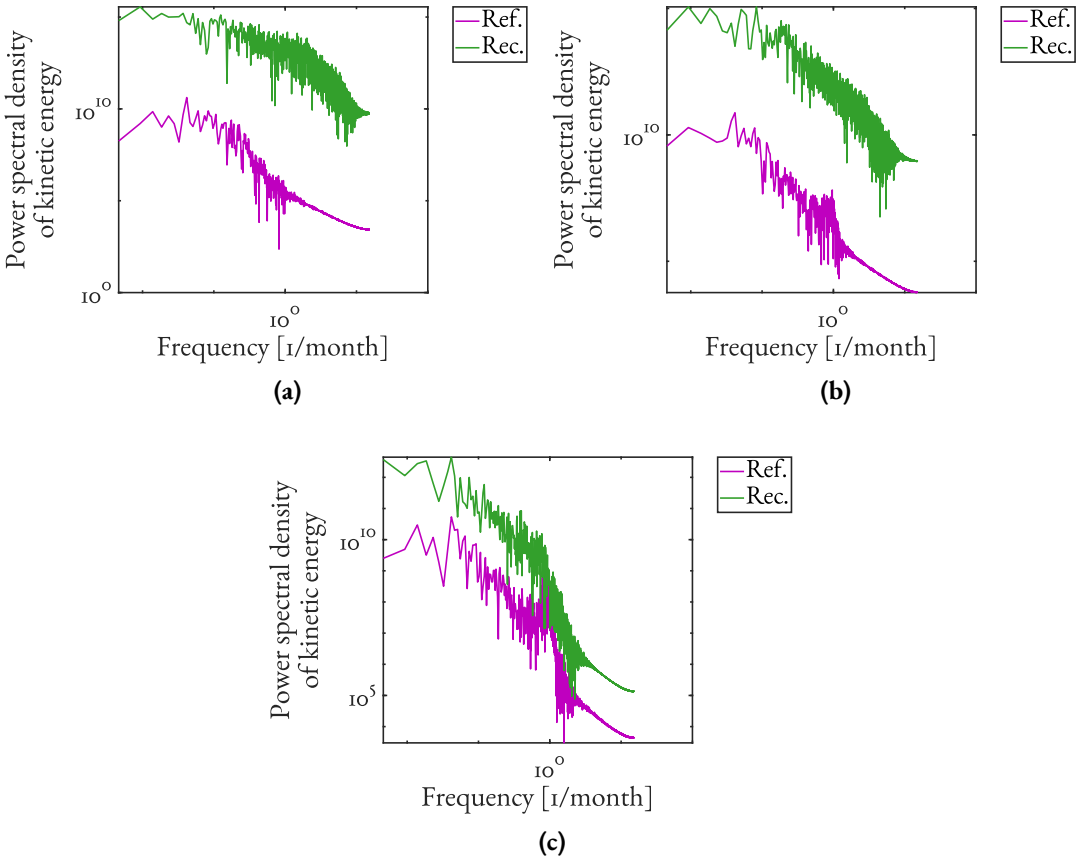
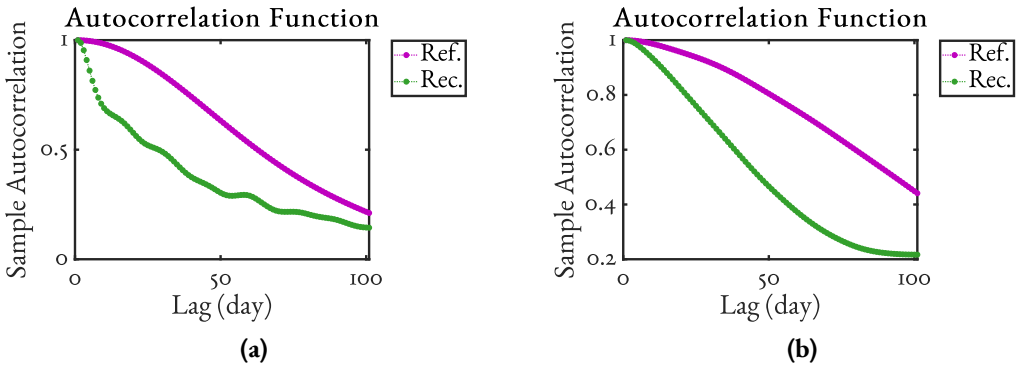


Figure 4.42.: Power spectrum representation of the kinetic energy of the models with different R_e via changes in E_L in multiples of **(a)** $R_e \times 1$, **(b)** $R_e \times 2$, **(c)** $R_e \times 10$.

The power spectrum representation of the kinetic energy shows, in a more clear way, the large differences in the reference and reconstructed data. This is the first indication that the model starts to fall apart for systems of a more chaotic nature and with ample deviations from the starting point.



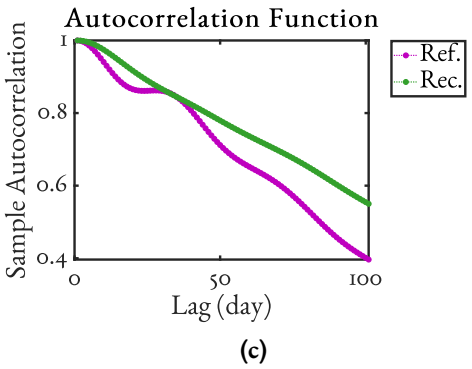


Figure 4.44.: Autocorrelation function associated to the the kinetic energy of the models with different R_e via changes in E_L in multiples of **(a)** $R_e \times 1$, **(b)** $R_e \times 2$, **(c)** $R_e \times 10$.

Only for the autocorrelation that corresponds to the $R_e = 140$ do the reduced models seem to initially follow a similar behavior, but not enough to be associated with one another.

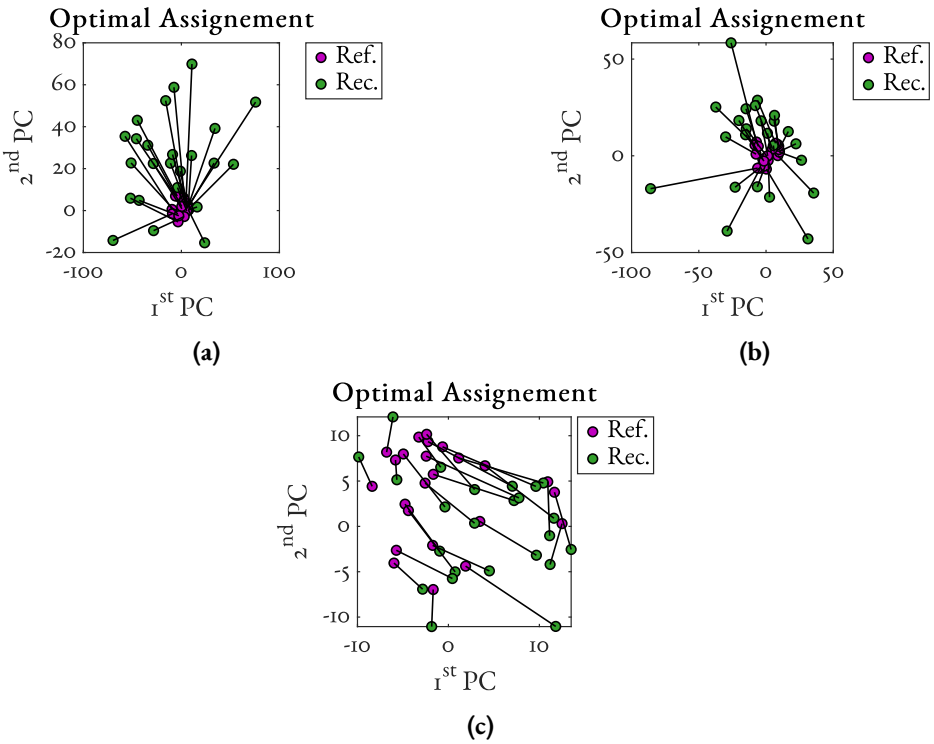


Figure 4.45.: Optimal assignment formed from the 1^{st} PCs and 2^{nd} PCs in the phase space for the reference and reconstructed models with different R_e via changes in E_L in multiples of **(a)** $R_e \times 1$, **(b)** $R_e \times 2$, **(c)** $R_e \times 10$.

Even using the Wasserstein distance and optimal assignment, it is clear that the principal components describe completely unrelated attractors with exorbitant values for the Wasserstein distance.

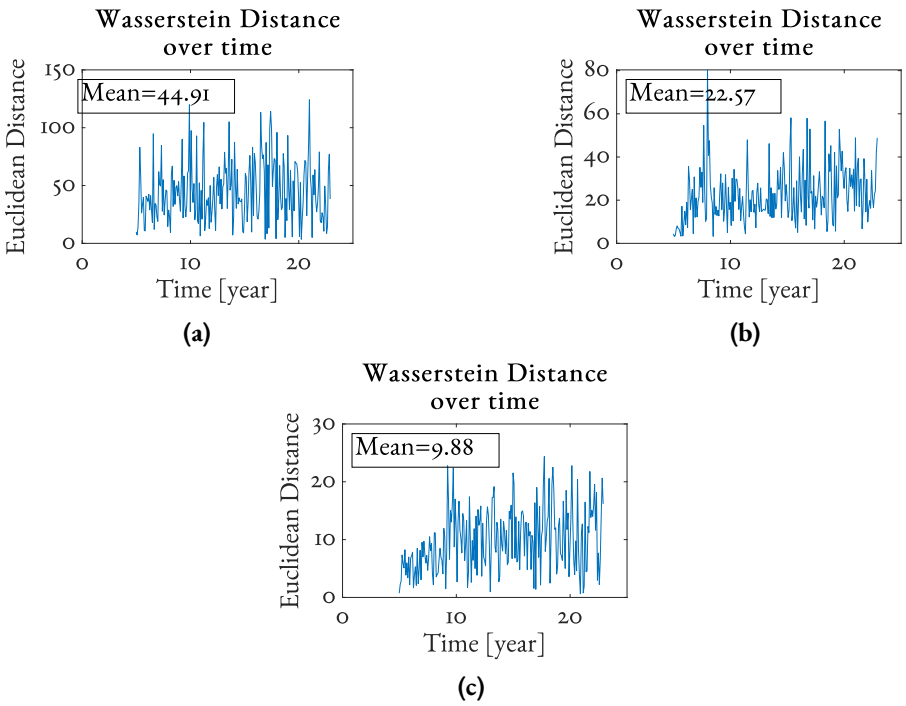


Figure 4.46.: Wasserstein distance associated to the 1^{st} PCs and 2^{nd} PCs in the phase space for the reference and reconstructed models with different R_e via changes in E_L in multiples of **(a)** $R_e \times 1$, **(b)** $R_e \times 2$, **(c)** $R_e \times 10$.

The Wasserstein distance is complemented by the phase space representation of the optimal assignment. Where a real meaning of the distance can be grasp. The more disperse the points of the principal components in the phase space, the smaller the Wasserstein distances. A small Wasserstein distance means that the cost of transforming the PCs of the reconstructed data to the reference data is. In this case (Figure 4.46), the Wasserstein distances are so high because of the concentration of the reference data in the origin of the phase space. This implies that the variabilities associated with these PCs are too low and henceforth now entirely reproducible with the EOF reduction method.

This is a similar result as the results for the high complexity model without the empirical correction. An additional empirical correction thus could be of use to resolve this issue.

5. Conclusions

Before the closing remarks, it is important to review the main results obtained for each tier established. First, the toy ocean model, in this case the reduced model and the reference model (*Veronis*, 1963) shared the same amount of spectral numbers and EOFs, which led to a successful full reproduction of the reference to a high degree.

For the second tier of complexity, the idealized with higher complexity, the reduction started to be inconsistent but because of the empirical corrections that were implemented and maintained along the rest of the study, the reduction was satisfactory to a degree where less EOFs were needed to recreate the kinetic energy, streamfunction and principal components of the reference.

The most realistic complex model studied was subdivided into a stochastic forced one, one submitted under different atmospheric conditions and one that could resemble gradients in temperature. For each of them the reduction worked differently, the stochastic anomaly was handled better and although differences between the reconstructions and the reference arose, the selection of a different number and the appropriate EOFs of the system would reduce those differences to an acceptable value.

Coming back to the research questions initially stated, the reduction method was tested with changes of the nature of the windstress input, and with it more realistic settings. When the windstress curl is perturbed by a stochastic noise of low amplitude, the reduced model was able to reconstruct successfully the streamfunction structure and kinetic energies. But the method reached its limit for stronger additional wind forcings (comparable to the default windstress amplitude T_0) and not all the studied characteristics were reproduced. This nevertheless, could be improved with the use of additional correction methods.

With slightly different atmospheric states the reduction method overestimated the kinetic energy and was unable to reconstruct the attractor and because of it, the principal components affected the streamfunction and energy of the system, hence its potential use for predictions cannot be confirmed.

In the case of different Reynolds number, only regimes where the reference model had no bottom friction influence were reconstructed. There is still a lot to explore in this aspect. And with all these answers, it can be said that the reduced model is able to successfully reconstruct important characteristic of the idealized wind driven ocean models studied between them, the circulation patterns of the streamfunction and kinetic energy.

The solutions obtained by the EOF reduction method are promising and the model works better with a simpler reference. The selected criteria seemed appropriate and the reconstructions, on the most part, were able to reproduce the characteristics compared. Differences in the statistics for the complex models are associated with the inclusion or exclusion of relevant EOFs that are probably not in the first ones.

When several time scales appear in the system the EOF analysis could be applied to low pass filtered model data. Then, the reduced model should only describe the low frequency variability while high frequency perturbations are discarded. Therefore, the EOF reduction method could be also used to design filtered models. Furthermore, one may separate the time scales by applying the EOF analysis to time filtered model output. Then, the first EOFs only describe low frequency variability and the high frequency waves which do not influence the slow dynamics can be filtered out.

Comparing the firsts principal components of the reduced model (EOF method) with the correspondent principal components of the spectral solution, for either idealized or quasi realistic input parameters, it can be noticed that already few EOFs (most cases in this study 3 EOFs) are sufficient to reproduce the essential phase space dynamics. When increasing the complexity of the spectral model by adding external forcing of different nature it was then when a higher number of EOFs and an empirical correction made the reduction method work.

After all the discussion and witnessing the weaknesses that are carried by the reduction, it is worthy to mention that some other approaches could turn the method into a more robust one. It would be interesting to explore the possibility of using the permutation matrix (obtained from the optimal assignment routine) as part of a next level correction. Preliminary tests have been carried out for spectral thermohaline circulation models and the results are also promising and could be integrated into the wind driven one to build a global reduced model.

Although the selection of the number of EOFs seemed straight forward along the study, the lack of a more direct indication prevented the model to be as efficient as it could be. Devising a method or a routine to estimate the number of EOFs to be used would increase the practicability of the method.

Appendices

A. Optimal assignment and Wasserstein metric

In order to calculate in a more accurately form how one of the attractors differs from another, a method for finding the optimal distance through the Wasserstein distance is implemented, which is finding the optimal permutation to convert one set of points into another. This is done by reducing a cost function by solving the **Monge–Kantorovich** problem, also known as the optimal transport problem (*Villani*, 2009; *Vissio*, 2018).

Depending on the measures μ and ν the optimal transport problem can take different forms, but first it is important to understand the concepts surrounding it. For two complete sets of metric spaces that are both completely metrizable and separable $\mathcal{X}$ and $\mathcal{Y}$ (also known as Polish spaces), the interest is put on subsets of $P(\mathcal{X})$ and $P(\mathcal{Y})$ as $\mathcal{P} \subset P(\mathcal{X})$ and $\mathcal{Q} \subset P(\mathcal{Y})$, respectively. Then defining the set $\Pi(\mathcal{P}, \mathcal{Q})$ of all transference plans whose marginals lie in $\mathcal{P}$ and $\mathcal{Q}$, respectively.

Thinking of the optimal transport as a simple transportation problem, quantities of a certain product, μ_i, must be shipped from each of m locations and received in amounts ν_j at each of n destinations. It is assumed that the system is balanced in the sense that the total amount shipped is equal to the total amount received. That is,

$$\sum_{i=1}^{m} \mu_i = \sum_{j=1}^{n} \nu_j \tag{A.1}$$

Associated with the shipping of a unit of product from origin i to destination j is a unit shipping cost c_{ij}. The aim is to determine the amounts π_{ij} to be shipped between each origin–destination pair $i - j$, so as to satisfy the shipping requirements and minimize the total cost of transportation. This requires a solution in the form of a linear programming problem. In other words, it can be reduced to find a set of π_{ij} that minimizes

$$\sum_{i=1}^{m} \sum_{j=1}^{n} c_{ij} \pi_{ij} \tag{A.2}$$

while satisfying

$$\sum_{j=1}^{n} \pi_{ij} = \mu_i \quad \text{for} \quad i = 1, \cdots, m \tag{A.3}$$

$$\sum_{i=1}^{m} \pi_{ij} = \nu_j \quad \text{for} \quad j = 1, \cdots, n \tag{A.4}$$

for $\pi_{ij} \geq 0$, so the constraints eq. A.3 and eq. A.4 remain consistent.

The transportation problem is now clearly seen to be a linear programming problem in mn variables. The equations A.3 and A.4 can be combined and expressed in matrix form resulting in an $(m + n) \times (mn)$ coefficient matrix consisting of zeros and ones only.

The optimal transport cost between the two measures:

$$C(\mu, \nu) = \inf_{\pi \in \Pi(\mu,\nu)} \int c d\pi \tag{A.5}$$

where c stand for the value of the optimal transport cost of transport between μ and ν.

Generally, this value does not, strictly speaking, represent a distance because it does not satisfy the axioms of a distance function. However, when the cost is defined in terms of a distance, it is possible to define a distance from it, the Wasserstein distance *Villani* (2009).

One very important aspect of optimal transport is that it gives rise to the Wasserstein distance. Simply put, the p-dimensional Wasserstein distance between two probability measures μ and ν on a complete, separable metric space $(\mathcal{X}, d)$ for $p \geq 1$ is the p^{th} root of the optimal transport cost above, with respect to the cost function (*Luenberger and Ye*, 2008), meaning

$$W_p(\mu, \nu) = \left(\min \int_{\mathcal{X} \times \mathcal{X}} d^p(x, y) d\pi(x, y) \right)^{\frac{1}{p}} \tag{A.6}$$

W_p defines a metric on the space of probability measures over $\mathcal{X}$ of dimension p, which is, the set

$$P_p(\mathcal{X}) = \left\{ \mu \in P(\mathcal{X}) : \int_{\mathcal{X}} d^p(x_0, x) d\mu(x) < \infty \right\} \tag{A.7}$$

where $P(\mathcal{X})$ is the set of probability measures on $\mathcal{X}$ and $x_0 \in \mathcal{X}$ is an arbitrary element.

In this particular case, $p = 2$ and this study is centered on the Kantorovich formulation of finite, and hence discrete, optimal transport and can be written as a finite linear program. For this, the measures μ and ν are considered as the following finite sums

$$\mu = \sum_{i=1}^{m} \mu_i \delta_{x_i} \qquad \text{and} \qquad \nu = \sum_{j=1}^{n} \nu_j \delta_{y_j} \tag{A.8}$$

The routines used are inspired on the ones coded by Gabriel Peyré, he made them free and openly available at `http://www.numerical-tours.com/matlab/` from the section of Optimal Transport with Linear Programming (*Peyré*, 2011).

Bibliography

Achatz, U., and G. Branstator (1999), A two-layer model with empirical linear corrections and reduced order for studies of internal climate variability, *J. Atmos. Sci.*, *56*(17), p3140–3160, doi:10.1175/1520-0469(1999)056<3140:ATLMWE>2.0.CO;2.

Achatz, U., and G. Schmitz (1997), On the closure problem in the reduction of complex atmospheric models by pips and eofs: A comparison for the case of a two-layer model with zonally symmetric forcing, *J. Atmos. Sci.*, *54*(20), p2452–2474, doi:10.1175/1520-0469(1997)054<2452:OTCPIT>2.0.CO;2.

Badin, G., and F. Crisciani (2009), On the transition from non-linear to linear regimes in the homogeneous models of the wind-driven ocean circulation, *Nuovo Cimento B*, *124*(4), p459–471, doi:10.1393/ncb/i2009-10773-y.

Badin, G., F. Crisciani, and G. Furlan (2003), On the dynamics of quasi-geostrophic intergyre gyres, *Nuovo Cimento C*, *26*(6), p621–632, doi:10.1393/ncc/i2003-10004-x.

Badin, G., F. Cavallini, and F. Crisciani (2009), Zonally aligned gyre solutions of linear models of wind-driven ocean circulation, *Nuovo Cimento B*, *124*(6), p653–669, doi:10.1393/ncb/i2009-10792-8.

Barcilon, V. (1998), On the barotropic ocean with bottom friction, *J. Mar. Res.*, *56*(4), p731–771, doi:10.1357/002224098321667341.

Bjerknes, J. (1964), Atlantic air-sea interaction, in *Adv. Geophys.*, vol. 10, edited by H. Landsberg and J. V. Mieghem, pp. p1–82, Elsevier, doi:10.1016/S0065-2687(08)60005-9.

Björnsson, H., and S. Venegas (1997), A manual for eof and svd analyses of climatic data, *CCGCR Report*, *97*(1), p112–134.

Böning, C. W. (1985), Eine untersuchung der dynamik der windgetriebenen ozeanischen zirkulation mit einem wirbelauflösenden barotropen modell, Phd/ doctoral thesis, Christian-Albrechts-Universität Kiel, doi:doi:10.3289/ifm_ber_137.

Botchkarev, A. (2019), *Interdisciplinary Journal of Information, Knowledge, and Management*, *14*, 045–076, doi:10.28945/4184.

Bryan, K. (1963), A numerical investigation of a nonlinear model of a wind-driven ocean, *J. Atmos. Sci.*, *20*(6), p594–606, doi:10.1175/1520-0469(1963)020<0594:ANIOAN>2.0.CO;2.

Böning, C. W. (1986), On the influence of frictional parameterization in wind-driven ocean circulation models, *Dyn. Atmos. Oceans*, *10*(1), p63–92, doi:10.1016/0377-0265(86)90010-2.

Charney, J. G. (1955), The gulf stream as an inertial boundary layer, *Proc. Nat. Acad. Sci. U.S.A.*, *41*(10), p731–740.

Cushman-Roisin, B., and J.-M. Beckers (2011), Chapter 7 - geostrophic flows and vorticity dynamics, in *Introduction to Geophysical Fluid Dynamics*, *International Geophysics*, vol. 101, edited by B. Cushman-Roisin and J.-M. Beckers, pp. p205–238, Academic Press, doi:https://doi.org/10.1016/B978-0-12-088759-0.00007-9.

Czaja, A., A. W. Robertson, and T. Huck (2003), The role of atlantic ocean–atmosphere coupling in affecting north atlantic oscillation variability, in the north atlantic oscillation: Climatic significance and environmental impacts, in *Geophys. Monogr. Ser.*, vol. 134, pp. p147–172, edited by J. W. Hurrell et al. AGU, Washington, D. C., doi:10.7916/D8N87M17.

Danabasoglu, G., and J. C. Mc Williams (1995), Sensitivity of the global ocean circulation to parameterizations of mesoscale tracer transports, *J. Climate*, *8*(12), p2967–2987, doi:10.1175/1520-0442(1995)008<2967:SOTGOC>2.0.CO;2.

Dellnitz, J. (2000), Niedrigdimensionale spektralmodelle des ozeans: Physik und strukturen., Dissertation, Universität Hamburg, Hamburg, Deutschland.

Desjardins, B., and E. Grenier (1999), On the homogeneous model of wind driven ocean circulation, *SIAM J. Appl. Math.*, *60*(1), 43–60, doi:10.1137/S0036139997324261.

Dewar, W. K. (2001), On ocean dynamics in midlatitude climate, *J. Climate*, *14*(23), p4380–4397, doi:10.1175/1520-0442(2001)014<4380:OODIMC>2.0.CO;2.

Eden, C., and J. Willebrand (2001), Mechanism of interannual to decadal variability of the north atlantic circulation, *J. Climate*, *14*(10), p2266–2280, doi:10.1175/1520-0442(2001)014<2266:MOITDV>2.0.CO;2.

Fofonoff, N. P. (1954), Steady flow in a frictionless homogeneous ocean, *J. Mar. Res.*, *13*, p254–262.

Frankignoul, C., and K. Hasselmann (1977), Stochastic climate models, part ii application to sea-surface temperature anomalies and thermocline variability, *Tellus*, *29*(4), p289–305, doi:10.1111/j.2153-3490.1977.tb00740.x.

Frankignoul, C., P. Müller, and E. Zorita (1997), A simple model of the decadal response of the ocean to stochastic wind forcing, *J. Phys. Oceanogr.*, *27*(8), p1533–1546, doi:10.1175/1520-0485(1997)027<1533:ASMOTD>2.0.CO;2.

Franzke, C., and A. Majda (2006), Low-order stochastic mode reduction for a prototype atmospheric gcm, *J. Atmos. Sci.*, *63*(2), p457–479, doi:10.1175/JAS3633.1.

Franzke, C., A. Majda, and E. Vanden-Eijnden (2005), Low-order stochastic mode reduction for a realistic barotropic model climate, *J. Atmos. Sci.*, *62*(6), p1722–1745, doi:10.1175/JAS3438.1.

Franzke, C. L. E., T. J. O'Kane, J. Berner, P. D. Williams, and V. Lucarini (2015), Stochastic climate theory and modeling, *WIREs Clim. Change*, *6*(1), p63–78, doi:10.1002/wcc.318.

Frisius, T. (1998), A mechanism for the barotropic equilibration of baroclinic waves, *J. Atmos. Sci.*, *55*(18), p2918–2936, doi:10.1175/1520-0469(1998)055<2918:AMFTBE>2.0.CO;2.

Frisius, T., K. Fraedrich, X. Zhu, and W. Wang (2009), A spectral barotropic model of the wind-driven world ocean, *Ocean. Model.*, *30*(4), p310–322, doi:https://doi.org/10.1016/j.ocemod.2009.07.008.

Greatbatch, R. J., and S. Zhang (1995), An interdecadal oscillation in an idealized ocean basin forced by constant heat flux, *J. Climate*, *8*(1), p81–91, doi:10.1175/1520-0442(1995)008<0081:AIOIAI>2.0.CO;2.

Grötzner, A., M. Latif, A. Timmermann, and R. Voss (1999), Interannual to decadal predictability in a coupled ocean–atmosphere general circulation model, *J. Climate*, *12*(8), p2607–2624, doi:10.1175/1520-0442(1999)012<2607:ITDPIA>2.0.CO;2.

Hannachi, A., I. T. Jolliffe, and D. B. Stephenson (2007), Empirical orthogonal functions and related techniques in atmospheric science: A review, *Int. J. Climatol.*, *27*(9), p1119–1152, doi:10.1002/joc.1499.

Hasselmann, K. (1976), Stochastic climate models part i. theory, *Tellus*, *28*(6), p473–485, doi:10.1111/j.2153-3490.1976.tb00696.x.

Hasselmann, K. (1982), An ocean model for climate variability studies, *Prog. Oceanogr.*, *11*(2), p69–92, doi:10.1016/0079-6611(82)90004-0.

Hellerman, S., and M. Rosenstein (1983), Normal monthly wind stress over the world ocean with error estimates, *J. Phys. Oceanogr.*, *13*(7), p1093–1104, doi:10.1175/1520-0485(1983)013<1093:NMWSOT>2.0.CO;2.

Holland, W. R. (1978), The role of mesoscale eddies in the general circulation of the ocean—numerical experiments using a wind-driven quasi-geostrophic model, *J. Phys. Oceanogr.*, *8*(3), p363–392, doi:10.1175/1520-0485(1978)008<0363:TROMEI>2.0.CO;2.

Hurrell, J. W., and C. Deser (2010), North atlantic climate variability: The role of the north atlantic oscillation, *J. Marine Syst.*, *79*(3), p231–244, doi:10.1016/j.jmarsys.2009.11.002.

Ierley, G. R., and V. A. Sheremet (1995), Multiple solutions and advection-dominated flows in the wind-driven circulation. part i: Slip, *J. Mar. Res.*, *53*(5), p703–737, doi:doi:10.1357/0022240953213052.

Jiang, S., F.-f. Jin, and M. Ghil (1995), Multiple equilibria, periodic, and aperiodic solutions in a wind-driven, double-gyre, shallow-water model, *J. Phys. Oceanogr.*, *25*(5), p764–786, doi:10.1175/1520-0485(1995)025<0764:MEPAAS>2.0.CO;2.

Jin, F.-F. (1997), A theory of interdecadal climate variability of the north pacific ocean–atmosphere system, *J. Climate*, *10*(8), p1821–1835, doi:10.1175/1520-0442(1997)010<1821:ATOICV>2.0.CO;2.

Kiss, A. (1998), Chaos in the „sliced cone"model of wind-driven ocean circulation, *Proceedings of the 1998 Woods Hole Geophysical Fluid Dynamics Program*, *WHOI Tech. Rep., Woods Hole, MA*, Available from Woods Hole Oceanographic Institution, Woods Hole, MA 02543.

Lambert, S. J., and G. J. Boer (2001), Cmip1 evaluation and intercomparison of coupled climate models, *Clim. Dynam.*, *17*(2), p83–106, doi:10.1007/PL00013736.

Llovel, W., J. K. Willis, F. W. Landerer, and I. Fukumori (2014), Deep-ocean contribution to sea level and energy budget not detectable over the past decade, *Nat. Clim. Change*, *4*, p1031–1035, doi:10.1038/nclimate2387.

Lorenz, E. N. (1956), Empirical orthogonal functions and statistical weather prediction,, Department of Meteorology, Massachusetts Institute of Technology, Cambridge, MA, 57 pp, statistical Forecasting Scientific Rep. 1.

Lorenz, E. N. (1963), Deterministic nonperiodic flow, *J. Atmos. Sci.*, *20*(2), p130–141, doi:10.1175/1520-0469(1963)020<0130:DNF>2.0.CO;2.

Luenberger, D. G., and Y. Ye (2008), *Transportation and Network Flow Problems*, pp. p145–179, Springer US, New York, NY, doi:10.1007/978-0-387-74503-9_6.

Marshall, J., Y. Kushnir, D. Battisti, P. Chang, A. Czaja, R. Dickson, J. Hurrell, M. McCartney, R. Saravanan, and M. Visbeck (2001a), North atlantic climate variability: phenomena, impacts and mechanisms, *Int. J. Climatol.*, *21*(15), p1863–1898, doi:10.1002/joc.693.

Marshall, J., H. Johnson, and J. Goodman (2001b), A study of the interaction of the north atlantic oscillation with ocean circulation, *J. Climate*, *14*(7), p1399–1421, doi:10.1175/1520-0442(2001)014<1399:ASOTIO>2.0.CO;2.

Meacham, S. P. (2000), Low-frequency variability in the wind-driven circulation, *J. Phys. Oceanogr.*, *30*(2), p269–293, doi:10.1175/1520-0485(2000)030<0269:LFVITW>2.0.CO;2.

Meacham, S. P., and P. S. Berloff (1997), Barotropic, wind-driven circulation in a small basin, *J. Mar. Res.*, *55*(3), p523–563, doi:10.1357/0022240973224364.

Meehl, G. A., W. M. Washington, and A. J. Semtner (1982), Experiments with a global ocean model driven by observed atmospheric forcing, *J. Phys. Oceanogr.*, *12*(4), p301–312, doi:10.1175/1520-0485(1982)012<0301:EWAGOM>2.0.CO;2.

Mehta, V. M., M. J. Suarez, J. V. Manganello, and T. L. Delworth (2000), Oceanic influence on the north atlantic oscillation and associated northern hemisphere climate variations: 1959–1993, *Geophys. Res. Lett.*, *27*(1), p121–124, doi:10.1029/1999GL002381.

Mildner, T. C. (2013), Past and present ocean dynamics in the western subtropical atlantic, Ph.D. thesis, Universität Hamburg, Von-Melle-Park 3, 20146 Hamburg.

Miller, R. N. (2007), *Numerical Modeling of Ocean Circulation*, Cambridge University Press, doi:10.1017/CBO9780511618512.

Morley, S. K., T. V. Brito, and D. T. Welling (2018), Measures of model performance based on the log accuracy ratio, *Space Weather*, *16*(1), p69–88, doi:10.1002/2017SW001669.

Munk, W. H. (1950), On the wind-driven ocean circulation, *J. Meteorol.*, *7*(2), p80–93, doi:10.1175/1520-0469(1950)007<0080:OTWDOC>2.0.CO;2.

Muñoz Pérez, J., L. Tejedor, and R. Medina (2001), Las funciones empíricas ortogonales y los cambios en el perfil de playa a corto, medio y largo plazo., *Física de la Tierra*, *13*, 139, doi:-.

Nilsen, J. E. Ø., Y. Gao, H. Drange, T. Furevik, and M. Bentsen (2003), Simulated north atlantic-nordic seas water mass exchanges in an isopycnic coordinate ogcm, *Geophys. Res. Lett.*, *30*(10), p1536, doi:10.1029/2002GL016597.

Niu, M., S. Sun, J. Wu, and Y. Zhang (2015), Short-term wind speed hybrid forecasting model based on bias correcting study and its application, *Math. Probl. Eng.*, p. 13, doi:https://doi.org/10.1155/2015/351354.

Nürnberg, D., A. Bahr, T. Mildner, and C. Eden (2015), *Loop Current Variability—Its Relation to Meridional Overturning Circulation and the Impact of Mississippi Discharge*, pp. p55–62, Springer International Publishing, Cham, doi:10.1007/978-3-319-00693-2_10.

Pedlosky, J. (1987), *Geophysical Fluid Dynamics*, Springer-Verlag New York, doi:10.1007/978-1-4612-4650-3.

Pedlosky, J. (1996), *Ocean circulation theory*, Springer, Berlin, Heidelberg, doi:10.1007/978-3-662-03204-6.

Peyré, G. (2011), The numerical tours of signal processing - advanced computational signal and image processing, *IEEE Comput. Sci. Eng.*, *13*(4), 94–97, doi:10.1109/MCSE.2011.71.

Primeau, F. (1998), Multiple equilibria and low-frequency variability of the wind-driven ocean circulation, Ph.D. thesis, Massachusetts Institute of Technology and Woods Hole Oceanographic Institution.

Scott, R. B. (1998), Geostrophic energetics and the small viscosity behaviour of an idealized ocean circulation model, Ph.D. thesis, McGill University Libraries.

Selten, F. M. (1995), An efficient description of the dynamics of barotropic flow, *J. Atmos. Sci.*, *52*(7), p915–936, doi:10.1175/1520-0469(1995)052<0915:AEDOTD>2.0.CO;2.

Sheremet, V. A., G. R. Ierley, and V. M. Kamenkovich (1997), Eigenanalysis of the two-dimensional wind-driven ocean circulation problem, *J. Mar. Res.*, *55*(1), p57–92, doi:doi:10.1357/0022240973224463.

Speich, S., H. Dijkstra, and M. Ghil (1995), Successive bifurcations in a shallow-water model applied to the wind-driven ocean circulation, *Nonlin. Proc. Geophys.*, *2*, p241–268, doi:10.5194/npg-2-241-1995.

Stommel, H. (1948), The westward intensification of wind-driven ocean currents, *Eos. Trans. AGU*, *29*(2), p202–206, doi:10.1029/TR029i002p00202.

Sura, P., and C. Penland (2002), Sensitivity of a double-gyre ocean model to details of stochastic forcing, *Ocean. Model.*, *4*(3), p327–345, doi:10.1016/S1463-5003(02)00008-2.

Sura, P., K. Fraedrich, and F. Lunkeit (2001), Regime transitions in a stochastically forced double-gyre model, *J. Phys. Oceanogr.*, *31*(2), 411–426, doi:10.1175/1520-0485(2001)031<0411:RTIASF>2.0.CO;2.

Thistleton, W. J., J. A. Marsh, K. Nelson, and C. Tsallis (2007), Generalized box–mÜller method for generating *q*-gaussian random deviates, *IEEE T. Inform. Theory*, *53*(12), p4805–4810, doi:10.1109/TIT.2007.909173.

Tofallis, C. (2015), A better measure of relative prediction accuracy for model selection and model estimation, *J. Oper. Res. Soc.*, *66*(8), p1352–1362, doi:10.1057/jors.2014.103.

Valdes, P., and B. Hoskins (1989), Linear stationary wave simulations of the time-mean climatological flow, *J. Atmos. Sci.*, *46*(16), p2509–2527, doi:10.1175/1520-0469(1989)046<2509:LSWSOT>2.0.CO;2.

Vallis, G. K. (2019), *Essentials of Atmospheric and Oceanic Dynamics*, Cambridge University Press, doi:10.1017/9781107588431.

Veronis, G. (1963), An analysis of wind-driven ocean circulation with a limited number of fourier components, *J. Atmos. Sci.*, *20*(6), p577–593, doi:10.1175/1520-0469(1963)020<0577:AAOWDO>2.0.CO;2.

Veronis, G. (1966), Wind-driven ocean circulation—part 2. numerical solutions of the non-linear problem, *Deep-Sea Res.*, *13*(1), p31–55, doi:10.1016/0011-7471(66)90004-0.

Verron, J., and J.-H. Jo (1994), On the stability of wind-driven barotropic ocean circulations, *Fluid Dyn. Res.*, *14*(1), p7–27, doi:10.1016/0169-5983(94)90019-1.

Villani, C. (2009), *Optimal Transport: Old and New*, Grundlehren der mathematischen Wissenschaften, Springer Berlin Heidelberg.

Visbeck, M., H. Cullen, G. Krahmann, and N. Naik (1998), An ocean model's response to north atlantic oscillation-like wind forcing, *Geophys. Res. Lett.*, *25*(24), p4521–4524, doi:10.1029/1998GL900162.

Visbeck, M., E. P. Chassignet, R. G. Curry, T. L. Delworth, R. R. Dickson, and G. Krahmann (2003), The ocean's response to north atlantic oscillation variability. in the north atlantic oscillation: Climatic significance and environmental impact, in *Geophys. Monogr. Ser.*, vol. 134, pp. p113–145, eds by J. W. Hurrell et al. AGU, Washington, D. C., doi:10.1029/134GM06.

Vissio, G. (2018), Statistical mechanical methods for parametrization in geophysical fluid dynamics, Ph.D. thesis, MPI für Meteorologie, Hamburg, doi:10.17617/2.3012736.

Williams, P. D., T. W. N. Haine, and P. L. Read (2005), On the generation mechanisms of short-scale unbalanced modes in rotating two-layer flows with vertical shear, *J. Fluid Mech.*, *528*, p1–22, doi:10.1017/S0022112004002873.

Williams, P. D., T. W. N. Haine, and P. L. Read (2008), Inertia–gravity waves emitted from balanced flow: Observations, properties, and consequences, *J. Atmos. Sci.*, *65*(11), 3543–3556, doi:10.1175/2008JAS2480.1.

Winton, M., and E. S. Sarachik (1993), Thermohaline oscillations induced by strong steady salinity forcing of ocean general circulation models, *J. Phys. Oceanogr.*, *23*(7), p1389–1410, doi:10.1175/1520-0485(1993)023<1389:TOIBSS>2.0.CO;2.

Wunsch, C. (2002), What is the thermohaline circulation?, *Science*, *298*(5596), p1179–1181, doi:10.1126/science.1079329.

Zacharuk, M., S. I. Dolaptchiev, U. Achatz, and I. Timofeyev (2018), Stochastic subgrid-scale parametrization for one-dimensional shallow-water dynamics using stochastic mode reduction, *Q J R Meteorol Soc.*, *144*(715), p1975–1990, doi:10.1002/qj.3396.

Acknowledgements

First of all I would like to thank to Thomas Frisius, Gualtiero Badin and Christian Franzke, for your patience, your good advice and being there when I needed it the most, I will be always grateful. Special thanks to Berit and Ingo, without whom this thesis would have never been possible, your support made the difference. Gracias to Frau Kutlu, for always having ready a smile. Thanks to Prof. Dr. Johanna Baehr for being part of my committee. MC, thanks for your proofreading and comments. And thanks to you Janine, for always being part of my team. To my family, my friends and to everybody who made me feel at home in a foreign land. Thanks to you all.

Versicherung an Eides statt

Hiermit versichere ich an Eides statt, dass ich die vorliegende Dissertation mit dem Titel: *„On the Empirical Orthogonal Functions Representation of the Ocean Circulation“* selbstständig verfasst und keine anderen als die angegebenen Hilfsmittel — insbesondere keine im Quellenverzeichnis nicht benannten Internet-Quellen — benutzt habe. Alle Stellen, die wörtlich oder sinngemäß aus Veröffentlichungen entnommen wurden, sind als solche kenntlich gemacht. Ich versichere weiterhin, dass ich die Dissertation oder Teile davon vorher weder im In- noch im Ausland in einem anderen Prüfungsverfahren eingereicht habe und die eingereichte schriftliche Fassung der auf dem elektronischen Speichermedium entspricht.

Hamburg, 06.04.2020 Jairo Segura