Doctoral Thesis

Exploring teleconnections and climate dynamics by means of complex networks

by

Fernando Arizmendi

Supervisors: Dr. Marcelo Barreiro and Dr. Arturo Martí

A thesis submitted in fulfilment of the requirements

for the degree of Doctor in Physics

Atmospheric Science Department, Physics Institute

University of the Republic, Montevideo, Uruguay

March 2018

Acknowledgments

To Marcelo Barreiro and Arturo Martí. They gave me the opportunity to do my doctoral thesis and had the necessary patience to guide me through the process.

To the Institute of Physics of the University of the Republic, Montevideo, Uruguay. Many professors, technical and administrative support, partners and students, that not only help me within the University but with whom I spent so many nice moments eating 'asados', playing 'futbol' and much more.

To the LINC project. Both professors and partners from whom I've learnt a lot. Each travel, each meal, each beer or just any time shared with my LINC friends is with me forever. Hope to see you soon.

To my family (blood, in-laws and from heart). To the ones that left and the ones to come. To my friends that, even tough times goes by and maybe we don't see each other in months or even years, are always with me. To Olimarudo +33, my football team. To Ricardo, my psycologist.

Last but not least... Anita, Rufi and Luca: you mean the world to me. I love you so much.

Contents

List of Figures

List of Tables

Abstract

Understanding Earth's climate system is of paramount importance given its impact on every form of life, including of course human society. In these times of global warming, and considering the nonlinear nature of climate dynamics, it is necessary to continue improving our knowledge of the underlying aspects of climate in order to mitigate the eventual biological, social and economical consequences. Even more, it is precise to understand these consequences to awaken as a society and take action on the matter. In the last 40 years, since the satellite era began, the quality and coverage of the climatic variables has improved significantly. The better understanding of the underlying physics together with the increase of computing power calculation result in very precise databases, called reanalysis data, that integrate the infromation through state-of-the-art general circulation models. However, there are still several aspects about climate yet to be learn. It is necessary to explore further this information by means of both standard statistical techniques and recently developed ones, such as complex networks.

The theory of complex systems has grown a great deal in the last 2 decades. It has been used not only in physics but in several and diverse scientific areas such as biology, sociology, economics and genetics among others. The emergence of coherent collective behavior, out of the interaction of a large number of active elements with relatively simple individual dynamics, is the key signature of a complex system. In climate, this perspective can be explored by means of functional networks considering different geographical regions, such as the nodes of the grid of reanalysis databases, as individual interacting agents. Then, assessing the statistical similarity of the evolution of the agents, a climate network can be built by putting links between every pair of nodes that behave similarly with certain statistical significance.

This thesis is aimed to explore teleconnections and climate variability from a complex

networks perspective. To that end, in Chapter 3, next to the introductory chapters (Chapter 1: Introduction to climate system and atmospheric circulation; Chapter 2: Methods and databases), we build climate networks of the 200 mb eddy geopotential height field. We choose this particular pressure level because it is where the maximum divergence takes place and, therefore, the level at which propagation of Rossby waves occurs. We study networks of different nature using standard statistical techniques and concepts from information theory like the nonlinear mutual information that, together with symbolic analysis, allows focusing on different time scales (intra-seasonal, intra-annual and inter-annual). With this set of different networks, we study the evolution in the connectivity of key teleconnection areas in the tropical Pacific during the 20th century. Moreover, using a set of outputs of an Atmospheric General Circulation Model (AGCM), we analyze the contribution of the tropical forcing and that of the internal atmospheric variability to the network structure.

After this exploratory analysis, in Chapter 4, we focus on the seasonal variability and the impact of the different stages of El Niño Southern Oscillation (ENSO). We further look into the network connectivity differences found on El Niño (EN) and La Niña (LN) during austral spring, particularly on the Southern Hemisphere (SH), assessing whether could it be explain in terms of the dispersion of tropically forced Rossby waves or as due to the interaction of these stationary waves with midlatitude dynamics.

In Chapter 5, in another important topic of this thesis, we investigate the response of the climatic system to the incoming solar radiation (insolation). Particularly, based on a forcing-response scheme, we explore the differences in the evolution of the Sea Air Temperature (SAT) field with respect to insolation. Given the importance of the SAT field into regional climate, we further calculate the unpredictability by means of the Shannon entropy. Surprisingly, the large-scale patterns that emerge in both calculation are quite similar and reveal interest aspects of climate dynamics, especially related to precipitation.

Finally, in Chapter 6, we present a summary of the work done and discuss possibilities for future works.

Resumen

Comprender el sistema climático de la Tierra es de suma importancia dado su impacto en todas las formas de vida, incluida, por supuesto, la sociedad humana. En estos tiempos de calentamiento global, y considerando la naturaleza no lineal de la dinámica del clima, es necesario continuar mejorando nuestro conocimiento de los aspectos subyacentes del clima para mitigar las eventuales consecuencias biológicas, sociales y económicas. Aún más, es preciso comprender estas consecuencias para despertar como sociedad y tomar medidas al respecto. En los últimos 40 años, desde que comenzó la era del satélite, la calidad y la cobertura de las variables climáticas han mejorado significativamente. La mejor comprensión de la física subyacente junto con el aumento del poder de cálculo computacional dan como resultado bases de datos muy precisas, llamadas datos de reanálisis, que integran la información a través de modelos de circulación general de última generación. Sin embargo, todavía hay varios aspectos sobre el clima que aún deben aprenderse. Es necesario explorar más a fondo esta información mediante técnicas estadísticas estándar y desarrolladas recientemente, como redes complejas.

La teoría de los sistemas complejos ha crecido mucho en las últimas 2 décadas. Se ha utilizado no solo en física, sino en varias y diversas áreas científicas como biología, sociología, economía y genética, entre otras. La aparición de un comportamiento colectivo coherente, a partir de la interacción de un gran número de elementos activos con dinámicas individuales relativamente simples, es la base de un sistema complejo. En clima, esta perspectiva puede explorarse a través de redes funcionales que consideran diferentes regiones geográficas, como pueden ser los nodos de la grilla de bases de datos de reanálisis, como agentes interactuantes individuales. Luego, al evaluar la similitud estadística de la evolución de los agentes, se puede construir una red climática poniendo enlaces entre cada par de nodos que se comportan de manera similar con cierta significacia estadística.

Esta tesis tiene como objetivo explorar las teleconexiones y la variabilidad climática desde una perspectiva de redes complejas. Para ello, en el Capítulo 3, después de los capítulos introductorios (Capítulo 1: Introducción al sistema climático y a la circulación atmosférica, Capítulo 2: Métodos y bases de datos), construimos redes climáticas del campo de altura geopotencial de 200 mb. Elegimos este nivel de presión particular porque es donde la divergencia es máxima y, por lo tanto, es el nivel en el que se produce la propagación de las ondas de Rossby. Estudiamos redes de diferente naturaleza utilizando técnicas estadísticas usuales y conceptos de la teoría de la información como la información mutua, medida no lineal que, junto con el análisis simbólico, permite enfocarse en diferentes escalas de tiempo (intraestacional, intraanual e interanual). Con este conjunto de redes diferentes, estudiamos la evolución en la conectividad de las áreas clave de teleconexión en el Pacífico tropical durante el siglo XX. Además, utilizando un conjunto de resultados de un Modelo de Circulación Atmosférica General (AGCM según sus siglas en inglés), analizamos la contribución del forzamiento tropical y de la variabilidad atmosférica interna a la estructura de la red.

Luego de este análisis exploratorio, en el Capítulo 4, nos enfocamos en la variabilidad estacional y el impacto de las diferentes etapas de El Niño Southern Oscillation (ENSO). Examinamos además las diferencias de conectividad de red encontradas en El Niño (EN) y La Niña (LN) durante la primavera austral, particularmente en el hemisferio sur (SH), evaluando si podría explicarse en términos de la dispersión de las ondas de Rossby forzadas tropicalmente o a través de la interacción de estas ondas estacionarias con la dinámica de latitudes medias.

En el Capítulo 5, en otro tema importante de esta tesis, investigamos la respuesta del sistema climático a la radiación solar entrante (insolación). Particularmente, en base a un esquema de forzamiento-respuesta, exploramos las diferencias en la evolución del campo de temperatura superficial del aire (SAT) con respecto a la insolación. Dada la importancia del campo SAT en el clima regional, calculamos aún más la predictabilidad mediante la entropía de Shannon. Sorprendentemente, los patrones a gran escala que surgen en ambos cálculos son bastante similares y revelan aspectos de interés de la dinámica del clima, especialmente relacionados con la precipitación.

Finalmente, en el Capítulo 6, presentamos un resumen del trabajo realizado y discutimos las posibilidades de trabajos futuros.

Chapter 1

Introduction

1.1 The climate system

Etimology says that the word *climate* comes from the greek word $\kappa\lambda\iota\mu\alpha$ that means *inclination*. Usually, etimology gives insights of the original understanding of the referred words. In this case, climate was thought to be related to the inclination of the solar radiation, or, in other words, to the latitude. It can be understood as a first order approximation taking into account the larger incoming radiation in the Equator decreasing towards the poles. From those days to the present, there have been huge advances to improve our understanding of the climate system.

Nowadays, the current understanding is that the climate system is composed by five components: atmosphere, hydrosphere, cryosphere, lithosphere, and biosphere. In Fig. 1-1 there is a scheme of the climate system, its components and their interactions. The complex interaction among these components, averaged in time, determines the climate of any region in the world. The fact that there are processes of different temporal and spatial scales makes quite difficult to understand its dynamics and even more to predict it. In addition, taking into account the global warming of the last decades, the understanding of its dynamics is becoming more and more important. The potential consequences of changes in the global temperature and accompanying changes in rainfall patterns and wind distribution are of supreme importance, life on Earth as we know is at stake.

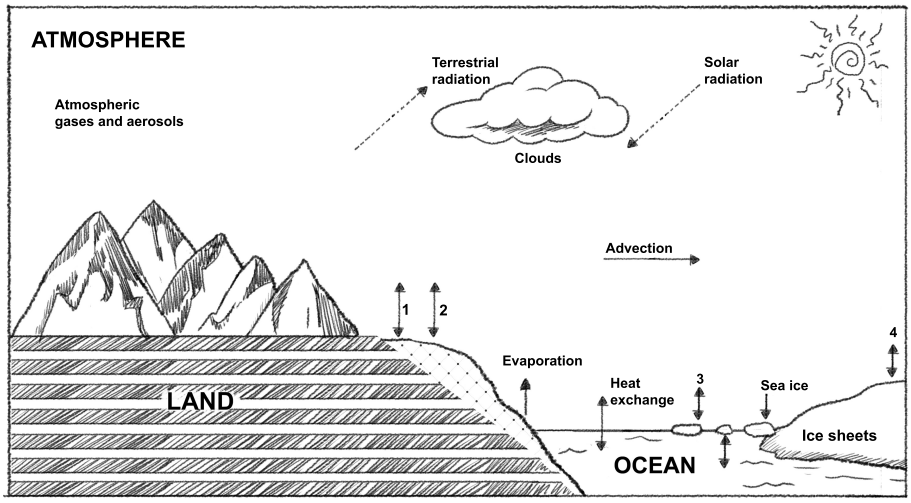

Figure 1-1: The climate system, its components and their interactions. 1: Atmopshere-land coupling, 2: Atmosphere-biosphere coupling, 3: Atmosphere-sea ice coupling, 4: Atmosphere-ice sheet coupling. Ilustration credits: Ana Revello.

In this thesis, we study mainly climate variability on seasonal time scales and also related aspects of the surface temperature response to solar focing. On these time scales the components of the climate system that dominate the variability are the atmosphere and the ocean. The ocean plays a fundamental role as it provides to the system the memory that allows to make extended predictions. For example, it is well known that weather cannot be forecasted with many days in advance. The fact that the dynamics of the atmosphere is chaotic [1] and that it is impossible to know exactly any initial condition, makes it impossible to know if 15 days from today will be hot and sunny or if there will be a storm with lower temperatures. However, in some regions it is possible to estimate with months in advance, if it is likely to rain more or less than usual, or whether it will be hotter or colder than usual given that this is highly dependent on the evolution of the sea surface temperature of the tropical oceans. This is because the surface temperature of the tropical oceans, which dynamics is quite well understood and can be predicted with months in advance, strongly influences the atmospheric circulation. The expected impact on the circulation gives insights of the expected climate anomalies for the following months on a particular region of the world. This interaction of processess on different time scales is one sample of the complexity that the climate system exhibits.

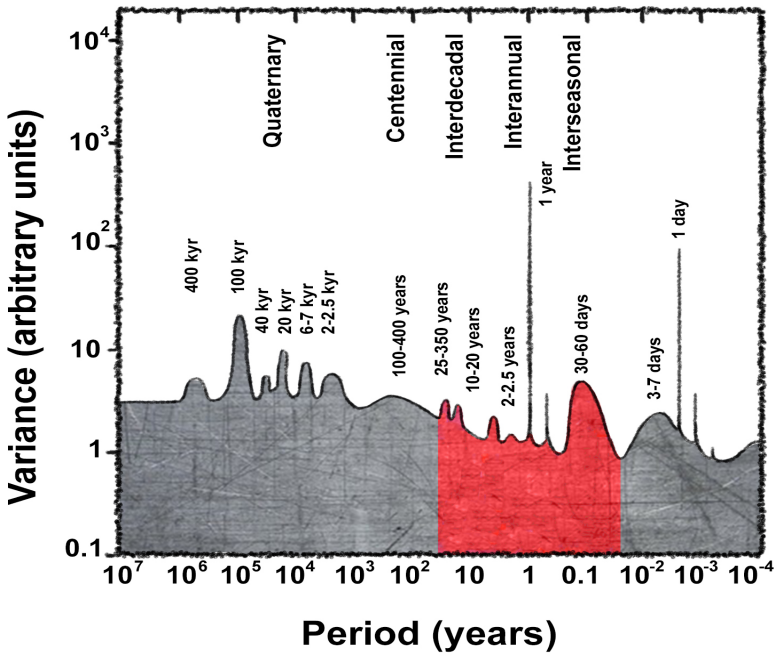

Figure 1-2: Artistic rendering of the power spectrum of climate variability, adapted from [2]. In red, the time scales that we tackle on this thesis. Ilustration credits: Ana Revello.

1.1.1 Climate variability

In Figure 1-2, we show a scheme of the power spectrum of climate variability [2]. The two high and narrow peaks that stand out are due to the sun forcing, specifically the response to the incoming solar radiation (insolation), with periods of a day and a year. In addition, next to the right of each of these peaks, there appear equally narrow peaks with smaller amplitude that represent their higher frequency armonics. It is worth to mention that, starting from the right, the first broad-band process has a period of 3 to 7 days and represents the weather synoptic scale. In the middle latitudes, it is related with the spatial distribution of low- and high-pressure systems, which has a direct impact on daily weather.

Other characteristics of the temporal variability at several time scales can be also appreciated in Fig. 1-2. In particular, the intraseasonal variability *i.e.* periods of around 30-60 days is relevant. One of the most important modes with significant impact in precipitation variability around the world acting in this frequency range is the Madden-Julian Oscillation

(MJO). It presents a quite large interannual variability wihch is, in another example of the interaction of processes of different time scales, related to the phase of El Niño Southern Oscillation (ENSO). Certainly, MJO activity is larger when ENSO is neutral or a weak La Niña is present and almost null if there is a strong El Niño. ENSO, as well as other large-scale variability patterns such as the Pacific North America Oscillation (PNA), the North Atlantic Oscillation (NAO), the Pacific Decadal Oscillation (PDA) or the Atlantic Multidecadal Oscillation (AMO), also appear in broad-bands with interannual and interdecadal variabilities. The large-scale oscillation patterns relevant for this thesis are further described in Section 1.5. Other oscillations with larger characteristic times, which are out of the scope of this work, as the Centennial variability, the Dansgaard-Oeschger Cycles, the Heinrich Events, and the Milankovitch Cycles, can be recognized in the scheme.

1.2 Energy and mass exchange

1.2.1 Thermal radiation

The Stefan-Boltzmann's law describes the power radiated from a black body in terms of its temperature T [3]. Specifically, the relation is as follows:

$$F = \sigma T^4, \tag{1.1}$$

where F is the flux of radiation per square meter, $\sigma = 5.67 \times 10^{-8}\,\mathrm{Wm^{-2}K^{-4}}$ is the Stefan-Boltzmann constant and T the absolute temperature. Then, considering that the temperature of the Sun's surface is $T_S \approx 6000\,\mathrm{K}$ and that of the Earth's surface is $T_E \approx 288\,\mathrm{K}$, we have $T_S > 20T_E$. Therefore, the Sun emits more than 16000 times as much radiation per unit area as the Earth.

On the other hand, Wien's displacement law complements Stefan-Boltzmann and states that the wavelength of maximum radiation intensity is inversely proportional to the absolute temperature [3]. Then, temperature T in K and the wavelength λ in μm can be related as:

$$\lambda_{\max} = \frac{2897}{T}. \tag{1.2}$$

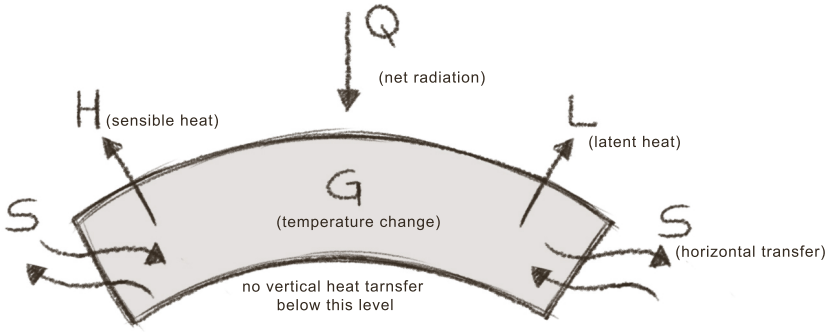

Figure 1-3: Model of energy transfer in the atmospheric system. Adapted from [4]. Ilustration credits: Ana Revello.

In this way, the maximum radiation wavelengths of the Earth and the Sun are, respectively, $10\,\mu$m (infrared) and $0.48\,\mu$m (mostly in the visible portion of the electromagnetic spectrum). It is clear that there is a huge difference between solar and terrestrial radiation and, accordingly, in their impact on the atmosphere and on the Earth's surface.

1.2.2 The heat budget

The particular climate at any place in the world is ultimately related to the net radiation and also depends on certain interacting variables [4]. Firstly, as we state at the beginning of this chapter, insolation varies with latitude. In addition, according to the surface (land, water or ice) and its characteristics (topography, vegetation or land use) energy transformations and, thus, climate can be significantly different in different regions.

In Fig. 1-3 we present a scheme of the heat budget at an arbitrary place. From the variables in the figure, if we assume that there are no changes in the temperature (for example in an annual period) such that $G = 0$, we can get the following relation:

$$Q = L + H \pm S \tag{1.3}$$

where L is the latent heat, H the sensible heat and S the horizontal energy transport done by the atmospheric or oceanic circulation. This equation applies to a mobile column, such as the ocean or the atmosphere. In the case of land, where subsurface flow of heat is negligible,

and therefore $S \approx 0$, we have:

$$Q = L + H. \tag{1.4}$$

The ratio between L and H is defined as the Bowen Ratio B [5] and is closely related with the problem that we tackle in Chapter 5. When the magnitude of B is less (greater) than one, a greater proportion of the available energy at the surface is passed to the atmosphere as latent heat (sensible heat) than as sensible heat (latent heat). Then, in a moist environment most of the available energy is used for evaporation with little available for the sensible heat. On the contrary, if the environment is dry, large amounts of energy are available for sensible heat and the local temperature is more sensitive to insolation. As we mentioned above, the surface and moisture strongly influence the climate at any location. In a moist environment, the energy as the latent heat goes back to the environment when the phase changes reverse. This fact is crucial to the redistribution of the incoming energy all over the globe as it is described in the next section.

1.3 Atmospheric circulation

The general structure of the large-scale atmospheric circulation is the result of the redistribution of the energy coming from the Sun and the fact that the Earth is rotating. In this section we briefly introduce the main governing equations of the atmosphere and the global large-scale circulation patterns.

1.3.1 Primitive equations

The whole set of equations used to describe atmospheric dynamics is called primitive equations. These equations, that were firt written by Bjerknes in 1904 [6], are used to approximate the global atmospheric flow and are used in most atmospheric models. The equations are the following [7]:

The continuity equation

It states the conservation of mass:

$$\frac{d\rho}{dt} + \rho\nabla\cdot\vec{v} = 0 \tag{1.5}$$

The momentum conservation

From Newton's laws of motion, Navier-Stokes equations describes the motion of viscous fluid substances. The atmosphere is one example and the motion equations are as follows:

$$\frac{D\vec{v}}{Dt} = -\frac{1}{\rho}\nabla p - 2\vec{\Omega}\times\vec{v} - \vec{f_r} + \vec{g}, \tag{1.6}$$

where $\frac{D}{Dt} = \frac{\partial}{\partial t} + (\vec{v}\cdot\nabla)$ is the material derivative, p the pressure, ρ the density, $\vec{\Omega}$ the angular velocity, $\vec{f_r}$ the frictional term and $\vec{g}$ the gravity force. The terms of the right-hand side of the equation stand for the pressure gradient force, Coriolis force, frictional force and gravity force, respectively.

The gas law

It was first stated by Èmile Clapeyron in 1834 as a combination of the empirical Boyle's law, Charles's law and Avogadro's law. It represents the equation of state of an ideal gas.

$$p = \rho RT, \tag{1.7}$$

where R is the gas constant ($R = 8.31432 \times 10^3\,\mathrm{N\,m\,kmol^{-1}K^{-1}}$, according to the U.S. Standard Atmosphere).

The thermal energy equation

From the first law of thermodynamics, the thermal energy equation can be deduced [7]:

$$c_v\frac{DT}{Dt} + p\frac{D\alpha}{Dt} = J, \tag{1.8}$$

where J is the rate of heating per unit mass due to radiation, conduction, and latent heat release. c_v the specific heat capacity and $\alpha = RT/p$.

These equations, together with the Stefan-Boltzmann law (Eq. 1.1 presented in Sec.1.2.2) and the Clausius-Clapeyron equation which characterises the relationship between the temperature of the air and its maximum water vapour pressure, are the basic equations that most global climate models use. However, the complexity of the models has never stopped increasing and, nowadays, also consider volcanic activity, sulfates, carbon cycle, aerosols, rivers, ocean circulation and interactive vegetation, among others.

1.3.2 Vorticity equation

Helmholtz decomposition

It is usefull to consider the atmospheric flux in their rotational and divergent components, $\vec{v}_\psi$ and $\vec{v}_\chi$ respectively. This is done through the Helmholtz decomposition (also called fundamental theorem of vector calculus) that states that any vector field that satisfies certain smoothness and decay conditions, that the atmosphere velocity field meets, can be decomposed as the sum of an irrotational vector field and a solenoidal vector field. Therefore, considering that an irrotational vector field has a scalar potential and a solenoidal vector field has a vector potential, we have the following relation:

$$\vec{v} = \vec{v}_\psi + \vec{v}_\chi = \hat{k} \times \nabla\psi + \nabla\chi, \tag{1.9}$$

where the vector potential ψ is the stream function and the scalar potential χ the velocity potential [8, 9].

Conservation of absolute vorticity

In addition to the primitive equations presented in Sec.1.3.1 also equations describing vorticty in a fluid field are of importance. In a horizontal flow, *i.e.* $\vec{v} = (u, v, 0)$, vorticity ξ is the vertical component of the rotation of the velocity field:

$$\vec{\xi} = \nabla \times \vec{v} = (\frac{\partial v}{\partial x} - \frac{\partial u}{\partial y})\hat{z} \tag{1.10}$$

with u and v being, respectively, the zonal and meridional velocity components. Vorticity is a measure of the local rotation or spin of the flow at any point in the flow. If we want to consider the absolute vorticity in an inertial reference system as the Earth, we have to add the part stemming from the planetary rotation, $f = 2\Omega \sin\lambda$, being λ the latitude and Ω the frequency of the Earth's rotation. Thus, absolute vorticity is $\eta = \xi + f$.

If we consider that the vertical advection and torque effects are negligible, the horizontal movements described by Eq.1.6 can be written by means of the absolute vorticity [7, 10, 11]:

$$\frac{\mathrm{d}}{\mathrm{d}t}(\xi + f) + (\xi + f)\nabla \cdot \vec{v} = 0 \tag{1.11}$$

Therefore, under these conditions, absolute vorticity is conserved.

1.3.3 Global large-scale circulation patterns

Tropical forcing

If we consider Eq. 1.5 in isobaric coordinates, continuity states that [7, 10, 11]:

$$\nabla \cdot \vec{v}_\chi + \frac{\partial \omega}{\partial p} = 0, \tag{1.12}$$

where ω is the vertical velocity component. Then, integrating from surface pressure p_s and a level p, we get:

$$\omega(p) = \int_p^{p_s} \nabla \cdot \vec{v}_\chi, \tag{1.13}$$

given that $\omega(p_s) = 0$ in the absence of topography. Therefore, horizontal divergence anomalies cause intense vertical movements. These anomalous ascents induce baroclinic structures, which in turn induce changes in the upper troposphere and in the divergent circulation. This feature describes thermally driven circulation such as the Walker and Handley cells presented in the next section.

The three-cell model

In the tropical oceans, particularly near the Equator where the insolation reaches a maximum, there is the Intertropical Convergence Zone (ITCZ). There, the moist air is warmed by

the insolation such that its density decreases and rises. This rising air creates a low pressure in the Equator generating convection. Then, when the air reaches upper tropospheric levels, the air masses move poleward up to 30 degrees of latitudes where the air gets cold enough to descend to surface levels. These latitudes are therefore characterized by high surface pressures.

This convective cell is named Hadley cell after the British scientist George Hadley who discovered it when he was studying trade winds resulting from subtropical air masses going back to the Equator. Thanks to the Coriolis force, these air masses are deflected to the west in both hemispheres, resulting in trade winds from the northeast in the Northern Hemisphere (NH) and from the southeast in the Southern Hemisphere (SH).

A similar phenomenon occurs in the polar regions where the energy radiative balance is negative. Indeed, in these regions, the atmosphere radiates more energy in the long wavelengths than in the short (corresponding to the solar radiation) wavelengths eliciting the formation of Polar cells. As well as the Hadley cell, the surface winds are deflected to the west in both hemispheres.

Finally, between the Polar and Hadley cells, there exist the Ferrel cell that is not driven by thermal forcing but by eddy (weather systems) forcing. In this case, as the circulation is opposite to the Polar and Hadley cells, the surface winds are deflected to the east. That is why in middle latitudes we have the westerlies.

This three-cell model is symmetric in both hemispheres and explains reasonably well the annual mean surface wind distribution in the atmosphere. However, except for the Hadley cell in the tropics, it cannot explain the circulation in the upper levels of the troposphere which requires taking into account of planetary wave motions.

In Figure 1-4 we show the climatology of the winds at 1000 mb in January and July. In this figure we can clearly distinguish the trade winds, the westerlies and the way that the continental areas affect wind distribution. The three-cell model explains in broad strokes the pattern observed there. Comparing both months considered, we can see that winds are more intense in the winter hemisphere, particularly the tropical winds due to the seasonal shift of the ITCZ.

As winds are much more intense on upper levels, we present in Fig. 1-5 the climatological

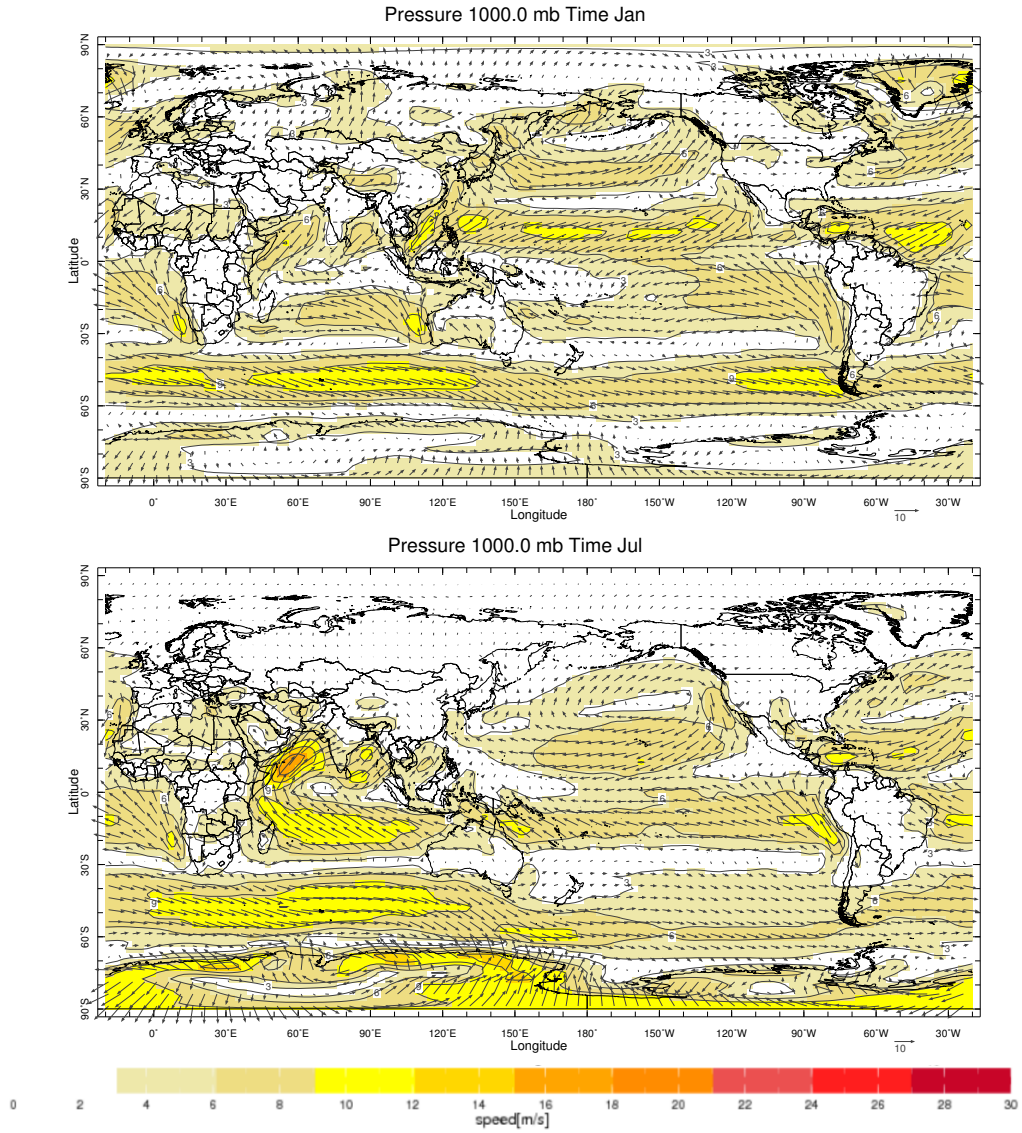

Figure 1-4: Climatology of near surface winds (1000 mb) in January (above) and July (below). Data from NCEP/NCAR Reanalysis [12].

winds at 200 mb both in January and July. To facilitate the comparison between Figures 1-4 and 1-5 in both we use the same color code. It can be seen that the general flow is westerly, characterized by the subtropical and subpolar jets. These jets shift latitudinally with the seasons and are stronger in winter. The subtropical jet is where the speed is most intense, between 20°N and 40°N in the northern hemisphere winter. In July, during the boreal summer, the subtropical jet moves north to about $40° - 45°$N. Due to the warming of

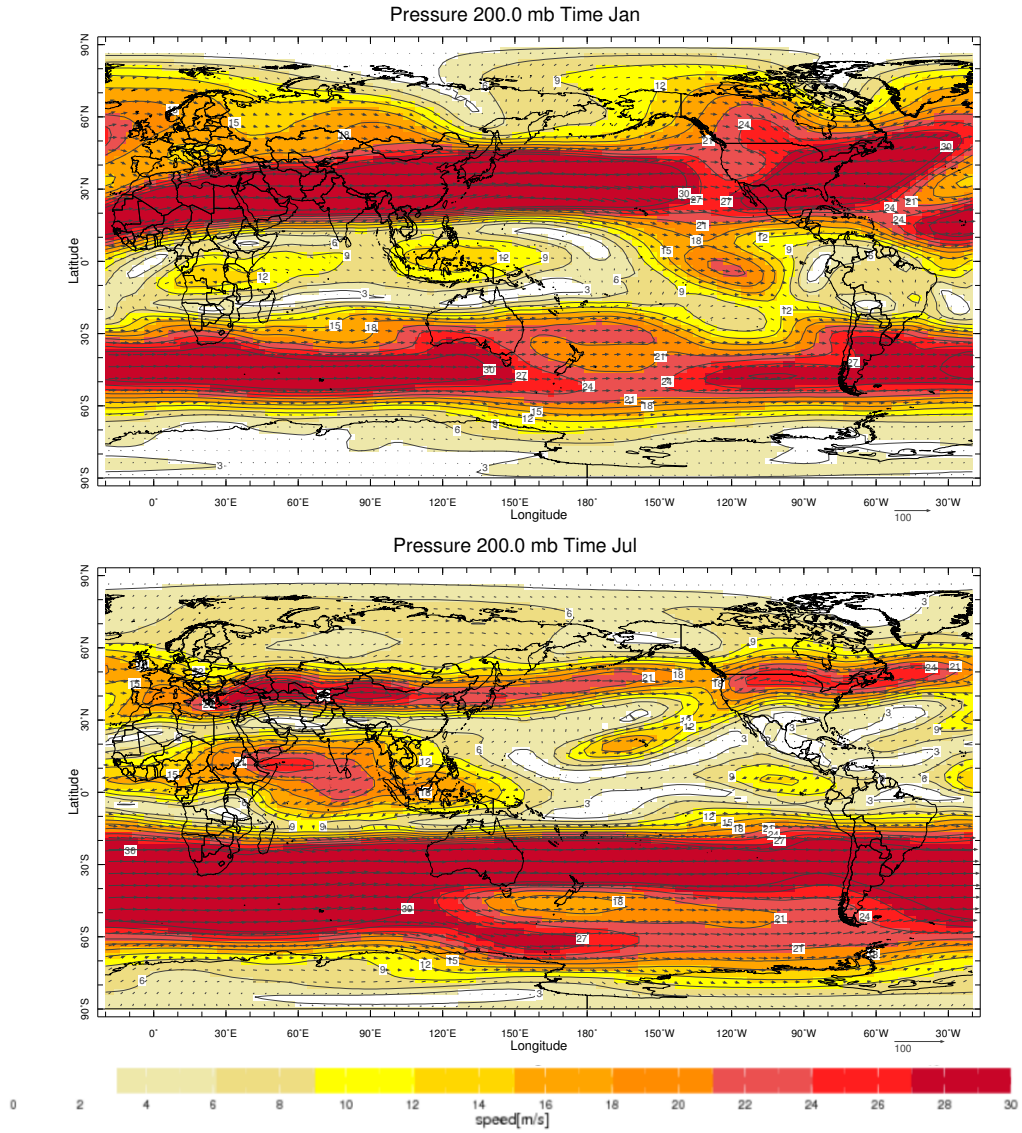

Figure 1-5: Climatology of upper troposphere winds (200 mb) in January (above) and July (below). Data from NCEP/NCAR Reanalysis [12].

the land, the latitudinal temperature gradients are weaker and the mean wind speed of the northern hemisphere jet stream is diminished. In the SH, the subtropical jet also moves with seasons and is located around 30°S during winter and 40°S during summer. Overall, the tropical region shows upper level easterly winds, differing markedly from the extratropical regions, and suggesting a different kind of dynamics.

Walker cell

The Walker circulation was first postulated by Bjerknes in 1969 [13]. Then, the focus was on the tropical Pacific, particularly related with ENSO phenomenom and was called Walker in honor to Gilbert Walker study on the Southern Oscillation (SO) in the 1920's decade [14, 15, 16]. Nowadays, it is known that the zonal circulation presents different cells along the tropical belt [17, 18].

The Pacific Walker circulation is characterized by rising air in the western Pacific, where there is usually warmer surface temperature that the east Pacific, with west to east flux in the upper levels, descending air in the eastern Pacific and easterlies in the surface levels. The Atlantic cell has the same circulation sense, with rising air in South America and descending air in the East Atlantic. These zonal circulation in the tropical atmosphere, that are thermally driven as the Hadley and Polar cells, are highly related to the topography and sea-land distribution on the Earth.

In Figure 1-6 we show the climatology of precipitation in January and July. The ITCZ, close to the Equator, is where the precipitation is maximum. As well as in the wind figures, it is clear the shift of the ITCZ following the sun. Also, in the Pacific Ocean, there is a significant difference between eastern and western sectors, with rainfall over the Maritime continent being far more stronger than close to the coast of South America, as it is expected from the Walker cell. Over midlatitudes, the winter hemisphere has more precipitation, especially over the Atlantic and Pacific oceans, consequence of the strong eddy activity. Furthermore, in January, there are large precipitation amounts over what is known as the South Atlantic Convergence Zone (SACZ) that covers part of South America, specifically oriented in a northwest-southeast manner across southeast Brazil into the southwestern Atlantic Ocean.

1.4 Teleconnections

The term *teleconnections* does not appear in the glossary of meteorology of 1959 [19]. However, it had been previously used in 1935 by the Swedish meteorologist Anders Ångström in his article on climate in the North Atlantic region [20]. Teleconnection means that there

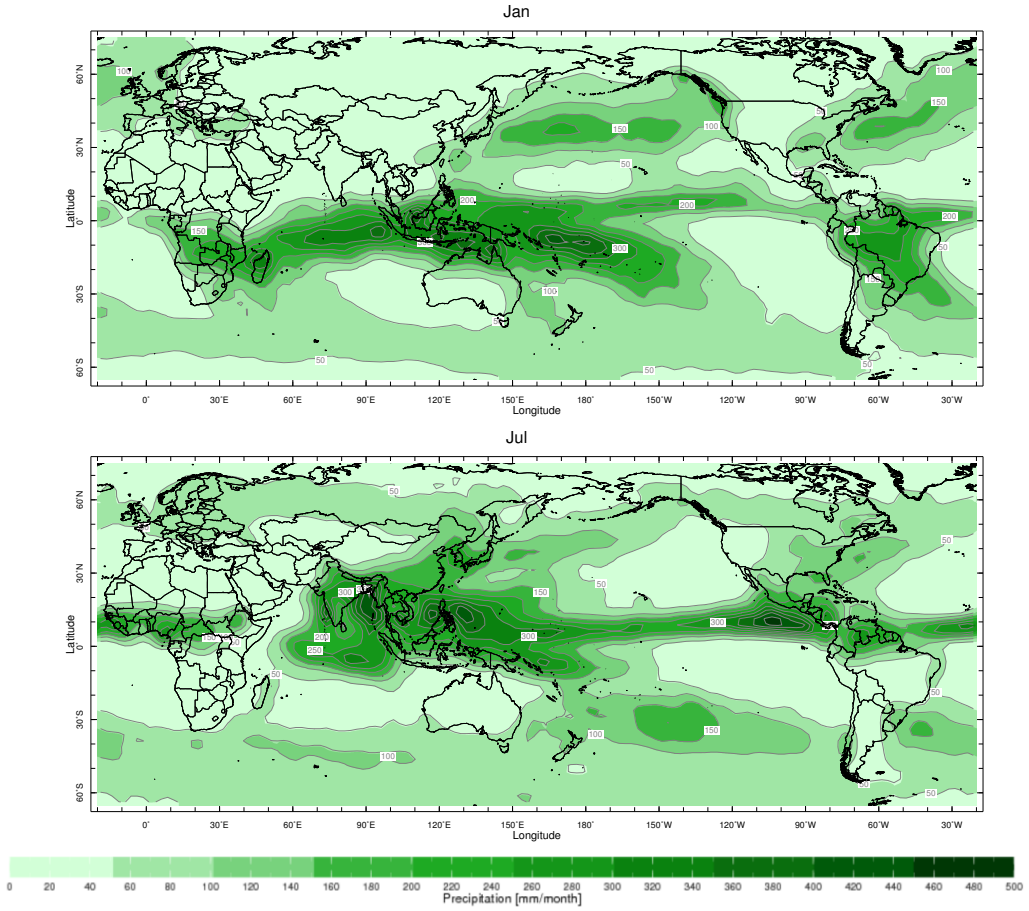

Figure 1-6: Climatology of precipitation in January (above) and July (below). Data from CAMS-OPI data.

is a physical reason for synchronized behavior, often of opposite sign, over distant parts of the globe. Typically, the distance between these regions can be of thousands of kilometers.

1.4.1 Rossby waves: source and propagation

Teleconnections are associated with atmospheric circulation anomalies which generally arise in zones with intense ocean-atmosphere interactions, mainly in the tropics, and propagate into extratropical regions by means of Rossby wave propagation. Using Helmholtz representation, vorticity equation can be represented as:

$$\frac{\partial \xi}{\partial t} + \vec{v}_\psi \cdot \nabla(\xi + f) = -(\xi + f)\nabla \cdot \vec{v}_\chi - \vec{v}_\chi \cdot \nabla(\xi + f). \tag{1.14}$$

The left-hand side of the equation, governed by the rotational component, describe the Rossby wave propagation while the right-hand side, with the divergent circulation component, represents the Rossby Wave Sources (RWS) [21].

On the one hand, in the extratropics, where there is no surface forcing, the anomaly propagation by means of Rossby waves is well suited by models of non-divergent barotropic vorticity [22, 23, 24]. On the other hand, in tropical latitudes, purely baroclinic models represent well the atmospheric response to the anomalies provoked by perturbations in the divergent circulation [25, 26]. The atmospheric response is characterized by zonal waves that propagates eastward (Kelvin waves) and westward (Rossby waves) from the forcing area.

1.4.2 Geopotential height

In this thesis, specifically in Chapter 3 and Chapter 4, we focus mainly in the 200 mb pressure level. This is because it is the height of the maximum tropical divergence that is crucial for the excitation and propagation of tropical and extratropical anomalies like Rossby wave trains [27]. Particularly, we analyse the geopotential height field Z, that is defined by:

$$Z = \frac{\Phi(z)}{g_0}, \tag{1.15}$$

where $\Phi(z) = \int_0^z g \mathrm{d}z$ is the geopotential function and g the gravity force (g_0 the gravity at surface level, $z = 0$).

In Fig. 1-7, we show an example of Z field at 200 mb ($Z200$) together with its eddy component ($Z200e$). It is the monthly mean field of December 1997, during a very strong El Niño, which is known for its global extreme consequences. It can be seen that $Z200$ has a meridional structure dominated by relatively large values in the tropical regions that decrease toward the poles. The eddy geopontential height $Z200e$, on the other hand, that is calculated by removing the zonal mean of each latitudinal band for this month considered, highlights deviations from this very strong zonal structure. The strong deviations with alternating opposite signs, that are found mainly in the extratropics, are signs of planetary waves.

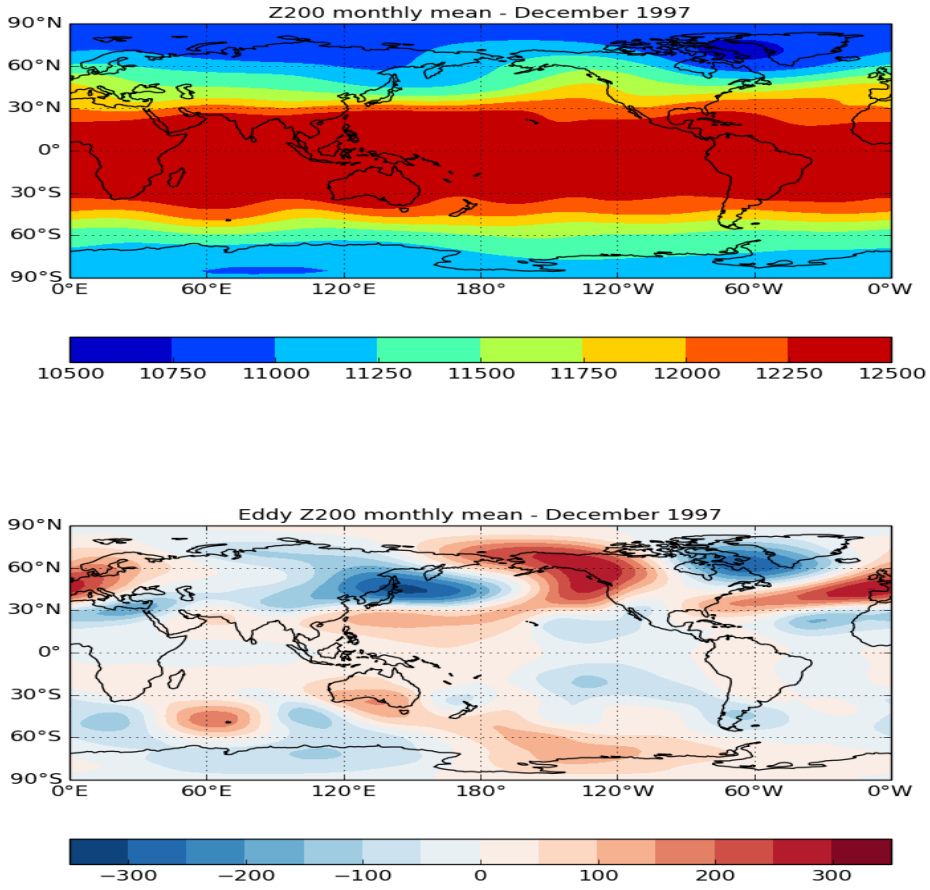

Figure 1-7: Geopotential height (above) and eddy geopotential height (below) fields at 200mb in December of 1997. A very strong El Niño year. Data from NCEP/NCAR Reanalysis [12].

1.5 Large scale variability patterns

The global atmospheric circulation presents a number of preferred patterns of variability, as we already mentioned in the subsection 1.1.1. In this section we introduce the oscillation patterns which strongly impacts on surface climate variations.

1.5.1 ENSO

El Niño-Southern Oscillation (ENSO) is a climate variability mode that develops in the tropical Pacific with a period that varies from 2 to 7 years. Its positive (negative) phase, called El Niño (La Niña), involves anomalous warm (cold) anomalies in the sea surface temperature (SST) of the eastern equatorial Pacific. In Figure 1-8, as an example, we show the warm anomalies of sea surface temperature of El Niño of 1997/1998 together with the cold anomalies of La Niña of 1999/2000.

El Niño name refers to Christ child, meaning *the boy* in spanish. Historically, the phenomenon was detected along the coast of Ecuador and Peru in late December, near Christmas, when local fishermen noted that the superficial waters of the eastern Pacific could be much warmer than usual. Besides the changes in the surface temperature, ENSO impacts strongly on atmospheric circulation.

Usually, over eastern Pacific ocean trade winds push waters from the Peruvian coast westward. This causes both an increase of sea level in the western Pacific and cold water upwelling in the eastern side. As a result from this dynamics, there are relatively cold waters in the east (cold tongue) and relatively hot waters in the west (warm pool) [28]. When there is an El Niño episode, the warmer waters extend to the eastern side of the basin, decreasing the zonal temperature gradient. This, in turn, weakens the Walker circulation shifting the rainfall bands to the East. There is a positive feedback between SST anomalies and surface winds that leads to the growth of El Niño maintaining it for several months. The contrary happens when there is a La Niña episode. The temperature gradient becomes larger and so the Walker circulation gets stronger. There is more upwelling than usual on the East Pacific and the West Pacific exhibits more intense convective activity.

As explained in Section 1.3.3, particularly with Eq. 1.13, changes in the horizontal divergence induce intense convective activity that ends up affecting the divergent circulation in the upper troposphere and generates Rossby waves that propagate to the extratropics. Through this mechanism ENSO can affect climate world wide. Even tough there have been many studies about ENSO teleconnections, there is still the need of a better understanding of its influences to improve seasonal regional forecasts. In Chapter 4, we tackle this problem by means of climate networks focusing particularly on the SH.

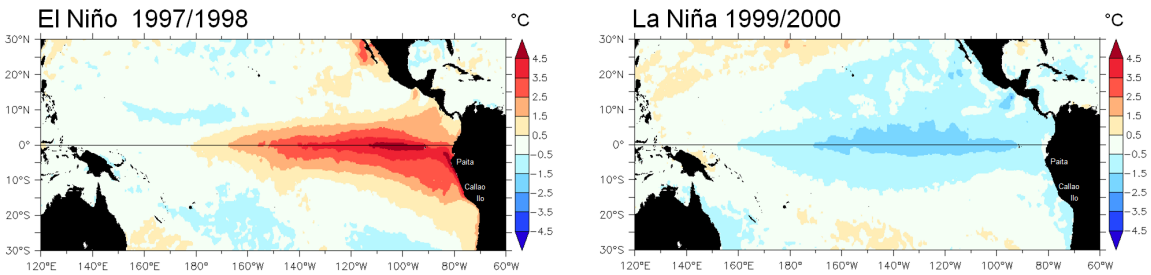

Figure 1-8: Average sea surface temperature anomalies of November, December and January of the years 1997/1998 (El Niño) and 1999/2000 (La Niña). Data from NOAA OISSTv2 [29].

1.5.2 PNA pattern

The Pacific North American (PNA) pattern is a quadripolar structure of pressure height anomalies, specifically at 500 mb level, located in the central Pacific ocean and centers of action over western Canada and the southeastern United States [30]. It tends to be most pronounced in boreal winter and is characterized by anomalies of similar sign located south of the Aleutian Islands and over the southeastern US. These anomalies with opposite sign are located in the vicinity of Hawaii, and over the intermountain region of North America (central Canada).

PNA is one of the most prominent modes of low frequency variability in the NH and strongly affects climates all over North America. The positive phase, with positive anomalies over Hawaii and central Canada, is associated with above-average temperatures over western Canada and US, and below-average temperatures across the south-central and southeastern US. The associated precipitation anomalies, on the other hand, are above average in the Gulf of Alaska extending into the Pacific northwestern US, and below average over the midwestern United States. Even being a natural internal mode of climate variability, the PNA pattern is also strongly influenced by ENSO. The positive phase of the PNA pattern tends to be associated with El Niño while the negative phase with La Niña.

1.5.3 PSA1 and PSA2 patterns

In 1986, Mo and Ghil [31] detected a wave train from the central Pacific to Argentina with large amplitudes in the Pacific South American (PSA) sector. Since its similarity with the

PNA wave train in the NH, they called it PSA mode (hereafter PSA1 mode). Szeredi and Karoly in 1987 [32], on the other hand, examined teleconnections in the SH using station data and found another wave train pattern with a 90° zonal phase lead with PSA1: the PSA2 pattern.

Nowadays, the PSA1 and PSA2 patterns are known as two pervasive modes of atmospheric variability that influence circulation and rainfall anomalies over South America [33]. They are, respectively, the second and third leading empirical orthogonal functions (EOFs) of the anomaly field of the geopotential heigth at 500 mb (Z500) and are found from intraseasonal to decadal time scales. Their spatial distribution can be seen in Fig. 1-9, where we plot the 2nd and 3rd EOF of the Z500 anomaly field focusing on SON season. The correspondent explained variances are 15.17% and 10.46%.

Both PSA1 and PSA2 exhibit a well-defined wave train pattern in the PSA sector, in mid to high latitudes with wavenumber 3. PSA1 is the response to ENSO at interannual time scales and is also related to SST anomalies over central and eastern Pacific at decadal scales. Therefore, similarly to ENSO, the associated rainfall pattern has negative rainfall anomalies over northeastern Brazil and positive over southeastern South America. PSA2, on the other hand, is related with the Quasi-Bieannial Oscillation (QBO) that is a quasiperiodic oscillation of the equatorial zonal wind between easterlies and westerlies in the tropical stratosphere with a mean period of 28 to 29 months. Contratry to the PSA1 pattern, which signal is present in all seasons, PSA2 strongest connections occur during the austral spring. The correspondent rainfall pattern shows a dipole pattern with anomalies out of phase between the South Atlantic Convergence Zone (SACZ) extending from central South America into the Atlantic and the subtropical plains centred at 35°S.

1.6 Objetives and thesis organization

In this thesis, we explore teleconnections and some aspects of climate dynamics from a complex system perspective. Our look is based on analyzing the possible insights of considering the emergence of climate patterns as the result of the collective behavior of interacting agents that are represented trough geographical regions. We aim to deepen our understanding of

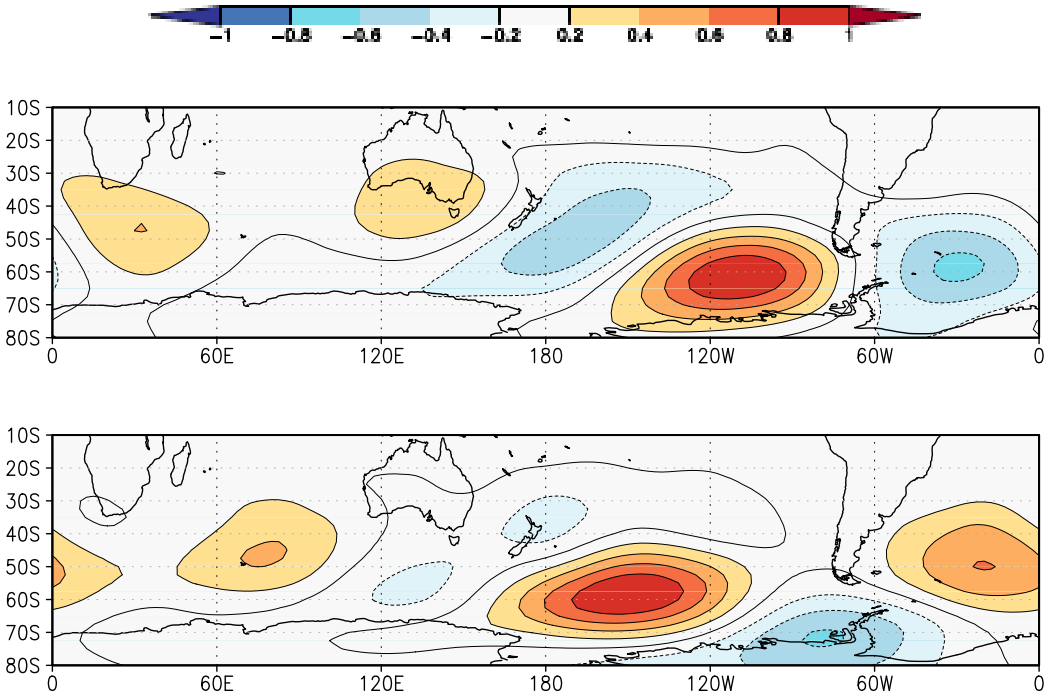

Figure 1-9: Second (above) and third (below) leading EOFs of the southern hemisphere Z500 anomalies representing, respectively, the PSA1 and PSA2 patterns. The EOFs are calculated during SON using data from NCEP/NCAR reanalysis [12].

teleconnections and different aspects of climate dynamics by means of functional network analysis and other more usual time series analysis techniques. To do that, in Chapter 2 we introduce the data sets and methods that we consider along this thesis. Particularly, we introduce the climate network paradigm that is the backbone of our work. In Chapter 3, we analyse the evolution of different eddy geopotential height networks in the upper levels of the atmosphere in order to charaterize this crucial field in the propagation of Rossby waves. We consider different time scales and both linear and nonlinear methodologies. Also we use outputs from an Atmospheric General Circulation Model (AGCM) to understand the connectivity differences associated to external forcing and internal variability. In Chapter 4, we discriminate our $Z200e$ networks through the different seasons and focus on austral spring, where the different phases of ENSO are found to have not opposite but different response in the propagation of teleconnections. In addition, in Chapter 5 we identify global patterns of unpredictability and response to insolation. Finally, in Chapter 6 we summarize

the results and discuss some possible problems to further explore in the future.

Chapter 2

Methods and datasets

2.1 Climate datasets

Historically, climate observations are strongly related with demography. For instance, the northern hemisphere has a geographically more detailed record than the southern hemisphere mainly because of two factors: there is more land in the northern hemisphere and it has always been more populated. Due to this asymmetry, the northern hemisphere has a stronger obsevational basis. Since 1979, however, the satellite era began and the coverage of climate information became more uniform. There are also indirect ways to acquire information about the history of a certain climate varibable. Ice cores, tree rings as well as rocks or sediments are used as proxies to determine past climates on Earth.

As we describe in the previous chapter, climate dynamics is full of different temporal and spatial variability features. Then, to get insights of a particular time scale process, we need to work this data by means of statistical tools. In this chapter, we review the sources of climate data and explain the different methods of statistical analysis that we use along this thesis.

2.1.1 Reanalysis data

As mentioned, observations are not uniformly placed over the globe, and the time period covered in different locations also varies. This is why the best global estimate of the historical evolution of any climate field is given by what is called *reanalysis data*, that is the blending

Table 2.1: Monthly databases used. N: number of grid points, T: number of months.

Source	Resolution	Time period	Dimensions	
NCEP CDAS1	2.5°	[1949, 2015]	$N = 10\,224$	$T = 792$
NOAA 20th Century	2°	[1901, 2009]	$N = 16\,020$	$T = 1308$
ERA Interim	1.5°	[1980, 2014]	$N = 28\,562$	$T = 408$
ICTP AGCM	3.75°	[1901, 2006]	$N = 4980$	$T = 1272$
EN34 index	–	[1901, 2006]	$N = 1$	$T = 1308$

Table 2.2: Daily databases. N: number of grid points, T: number of days.

Source	Resolution	Time period	Dimensions	
NCEP CDAS1	2.5°	[1949, 2014]	$N = 10\,224$	$T = 24024$

of all the observations available with an Atmospheric General Circulation Model (AGCM). There are many reanalyisis data bases and here we consider three of the most widely used. These are the NOAA 20th Century [29], the NCEP CDAS1 [12] and the ERA Interim [34] reanalysis data bases.

2.1.2 Atmospheric General Circulation Model (AGCM)

In addition to the reanalysis data bases, in order to explore the differences between intrinsic variability and external forcing, we also consider an ensemble of 10 AGCM experiments made with the International Center for Theoretical Physics (ICTP-AGCM) forced with global historical sea surface temperature (ERSSTv.2, [35]). The ICTP-AGCM is a full AGCM with simplified physics and a horizontal resolution of T30 (i.e. the horizontal resolution is limited by a triangular spectral cut-off to a total wave number of 30) with eight vertical levels [36, 37]. The performance of the ICTP-AGCM in representing the large-scale atmospheric dynamics is comparable to that of state-of-the-art AGCMs and has been used to study global climate variability, including the influence of the oceans on the circulation [38, 39, 40, 41].

The characteristics of each of the databases used are display in Table 1 (monthly data) and Table 2 (daily data).

2.1.3 Variables considered

In this section we describe the variables used along the thesis. In Chapter 3 and Chapter 4, we explore the atmospheric circulation through the building of geopotential height (Z) functional networks. To do so, given that our focus is to deepen our understanding of atmospheric circulation and teleconnections, we choose the 200 mb level ($Z200$) that is where the maximum tropical divergence takes place, crucial for the excitation of extratropical Rossby wave trains. As the geopotential height has a very strong meridional structure with high values in the tropics and decreases toward the poles, we consider the eddy component removing the corresponding zonal mean at each node for every time step (hereafter called Z200e). In this way the signatures of the planetary waves, which are deviations of this zonal structure, are highlighted.

In Chapter 3, we consider only monthly mean data and we focus on intraseasonal, intraannual and interannual time scales. We use data from NOAA 20th Century and NCEP CDAS1 reanalyisis databases for the analysis and an ensemble of runs made with ICTP AGCM model to distinguish between forced and intrinsic variability.

In Chapter 4, using NCEP CDAS1 reanalysis database, we extend the $Z200e$ network analysis separating between seasons and ENSO scenarios. Given that we found significant differences in the connectivity of the SH austral spring, we further calculate intraseasonal weather regimes in order to deepen the analysis in terms of the interactions of tropically forced waves with extratropical dynamics. To do that, we consider the daily mean of the same field (eddy geopotential height anomalies at 200 mb) and reanalysis data base (NCEP CDAS1). In addition, to improve our knowledge on the role of the tropical forcing in the SH austral spring, we also use the daily mean of 200 mb zonal and meridional winds from the same reanalysis database (NCEP CDAS1) to calculate the Rossby wave sources using equation 1.14.

Finally, in Chapter 5 we analyze global patterns of unpredictability and response to insolation in atmospheric data. Specifically, we focus on the Surface Air Temperature field (SAT) and we consider monthly mean data both from ERA Interim and NCEP CDAS1 reanalysis databases.

2.2 Climate network paradigm

A significant fraction of the advances in understanding climate dynamics is the result of the use of stastistical analysis tools such as composites, empirical orthogonal functions, clustering methods, to name a few. Besides the use of these tools, the challenge in this work, which was done in the context of an EU-funded project called Learning about Interacting Networks in Climate (LINC), is to implement the framework of complex networks to analyse the atmospheric circulation.

Since the late 1990's complex system theory have become a transversal field over many areas. The idea to improve our understanding of a dynamical processes by considering it as a set of interacting agents, has brought graph theory to the main stage. Biological, social, neuronal systems, among others, have been deeply study from this point of view in the last years [42]. In the following section we introduce main concepts of complex networks theory that we use along this thesis.

2.2.1 Complex networks

Euler and the seven bridges of Könisberg

In 1736, the great Swiss scientist Leonhard Euler published what is now known as the first problem of graph theory. The paper showed a picture of the Könisberg city -now Kaliningrad, Russia-, that is divided by the Pleger river and has two islands that are connected to each other and to the two mainland portions of the city by seven bridges[1] as it is schematized in Figure 2-1.

The problem was to devise a closed path in such a way that it would cross every one of the seven bridges once and only once. Euler proved that there was no solution. In his proof, he simplified the map focusing exclusively in the land regions and its connections. He considered each land region as a point -or node- and each bridge as a connection -or link- between these nodes. Then, he realized that in a circle path, each intermediary point necesarily needs an even number of links, that is the same amount to come as to leave, and the initial and last point. In addition, if we consider that the initial point need also to be

[1] That is why it is known as the seven bridges of Könisberg problem

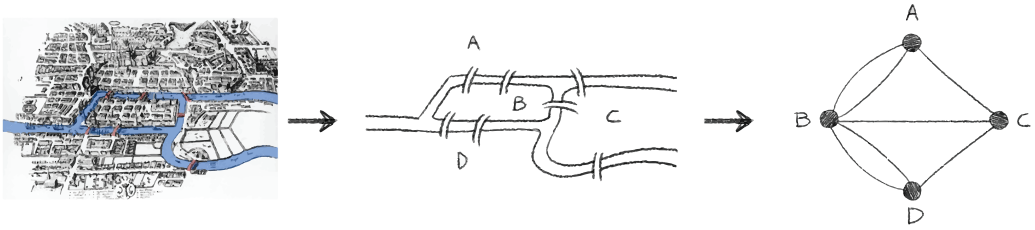

Figure 2-1: Euler's abstraction scheme of the Köninsberg bridge problem. Ilustration credits: Ana Revello.

the final, every point would need an even number of links in order to have a solution. In this particular problem we can see that there are four regions that have an odd number of bridges and therefore there is no solution. This is considered the birth of graph theory.

Adjacency matrix

The mathematical definition of a graph -or network- describe it as an ordered pair $N = (V, E)$ comprising a set V of vertices v_i -or nodes or points- together with a set E of edges e_{ij} -or arcs or links- [42]. The elements, e_{ij}, of E are a 2-element subset of V in the sense that each link is associated with nodes v_i and v_j. Note that there cannot be an edge without vertices while the opposite is possible, there can be vertices without edges.

There are many ways of considering the edges within a network -*i.e.* with different link intensity and/or with direction- but we will only focus in undirectional, without multiedges or self-edges and unweighted networks. This implies that $e_{ij} = e_{ji}$. In this case, a network of n nodes can be fully described by the $n \times n$ adjacency matrix $\mathbf{A}$ such that:

$$A_{ij} = \begin{cases} 1 & \text{iff } e_{ij} \in E \text{ (\textit{i.e.} there is an edge between } v_i \text{ and } v_j\text{).} \\ 0 & \text{otherwise.} \end{cases} \tag{2.1}$$

Topological properties of a network

Degree distribution The degree k_i is number of links connected to the node v_i. It can be defined by the following expression:

$$k_i = \sum_{j=1}^{n} A_{ij} \tag{2.2}$$

If we sum the degrees of every node, as every edge has two ends, we get twice the number of edges m,

$$\sum_{i=1}^{n} k_i = 2m. \tag{2.3}$$

An important property to characterize a network is the mean degree c,

$$c = \frac{1}{n} \sum_{i=1}^{n} k_i = \frac{2m}{n}. \tag{2.4}$$

The density ρ on the other hand is the fraction between the actual edges on the network and all the possible edges:

$$\rho = \frac{2m}{n(n-1)} = \frac{c}{n-1}, \tag{2.5}$$

being $\binom{n}{2} = n(n-1)$ the maximun possible number of edges in a simple graph.

We define the degree distribution $P(k)$ as the fraction of nodes with a degree k. This probability distribution gives valuable information about the topology of the network.

Average path length Another very important characteristic of a network is the efficiency in the information or mass transport. If we consider σ_{ij} as the distance between node v_i and v_j, where distance refers to the number of nodes within the shortest path that connects both nodes, we can define the average path length l_G by the following equation:

$$l_G = \frac{1}{n \cdot (n-1)} \cdot \sum_{i \neq j} \sigma_{ij} \tag{2.6}$$

Clustering coefficient The neighbourhood N_i of a vertex v_i is defined as all the nodes that are immediately connected to v_i,

$$N_i = \{v_j : e_{ij} \in E\}, \tag{2.7}$$

It is important usually to know the degree to which the nodes in the network tend to cluster together. To do that, the local clustering coefficient C_i of a node v_i, first introduced by Watts and Strogatz in 1998 [43], is defined by:

$$C_i = \frac{2|\{e_{jk} : v_j, v_k \in N_i, e_{jk} \in E\}|}{k_i(k_i - 1)} = \frac{\sum_{j,k=1}^{n} A_{ij}A_{ik}A_{jk}}{k_i(k_i - 1)}. \tag{2.8}$$

This is the fraction between of all the connections among the nodes within the neighbourhood N_i and all the possible connections between them, that is $k_i(k_i - 1)/2$. For example, if all the nodes in N_i are connected to each other, then $C_i = 1$. If there are some connections but some missing, then $0 > C_i > 1$. Otherwise, if there is no connection between the nodes in N_i, $C_i = 0$.

To characterize the whole network, $\bar{C}$ is the mean of all the local cluster coefficients and is used as a global clustering coefficient:

$$\bar{C} = \frac{1}{n}\sum_{i=1}^{n} C_i. \tag{2.9}$$

Furthermore, also as a global measure it is often used the transitivity T (sometimes referred to as the Barrat-Weigt clustering coefficient), that is the density of closed triangles in the network:

$$T = \frac{\sum_{i,j,k=1}^{n} A_{ij}A_{ik}A_{jk}}{\sum_{i,j,k=1, j\neq k}^{n} A_{ij}A_{ik}}. \tag{2.10}$$

Centrality measures

The importance of the nodes depends on the perspective. Here, we want to enumerate the main centrality measures:

Degree centrality and Area–weighted connectivity AWC Nodes with many connections are, in some way, more relevant in the network dynamics. To characterize the 'importance' of links several centrality measures are defined. The degree centrality is one of the most employed centrality measures.

In the context of climate networks, since data is usually defined on a grid which is regular in angular coordinates, the geographic distances of the grid points depends on latitude. Therefore, the related quantity known as the *area-weighted connectivity* AWC is defined:

$$\mathrm{AWC}_i = \frac{\sum_{j=1}^{N} A_{ij} \cos(\lambda_j)}{\sum_{j=1}^{N} \cos(\lambda_j)}, \tag{2.11}$$

where λ_j is the latitude of node j and the cosine factors are included to correct the differences in the area covered by the nodes given their geographical distribution over the regular grid. It can be easily proved that for angularly equidistant grids, as we use, the corresponding area on the Earth's surface of a node i is proportional to the cosine of the latitude λ_i.

Betweenness centrality The betweenness centrality is based on the shortest paths. For every pair of nodes, there is at least one shortest path between them such that the number of edges that the path passes through is minimized. The betweenness centrality B_{v_i} is then the number of shortest path that pass through the vertex v_i:

$$B_{v_i} = \sum_{j,k \neq i}^{n} \frac{\sigma_{jk}(v_i)}{\sigma_{jk}} \tag{2.12}$$

It is interpreted as a measure of the information flow through a node. Also it can be used to separate the nodes in clusters since if the nodes with highest values of B are removed the network will eventually be cut in many subnetworks.

Small world and free scale networks

Since the late 90's, as we mention before, the use of complex networks has been highly increased in several scientific areas. In 1998, Watts and Strogatz [43] introduce the concept

of small world networks. They notice that many real networks were not well represented neither by Erdos Renyi (ER) networks -also known as random networks- which have short average path length and low clustering values nor by regular networks, that are highly clustered but with large values of average path length.

Small–world networks In 1967, the psychologist Stanley Milgram published a work entitled 'The Small World Problem' [44]. There, he showed through an experiment that the average path lengh in the United States of America (US) was around 6. In the experiment, he selected 160 people from one coast side of the US and made them each to send a letter to an unknown person from the opposite coast side of US. The rules were that each individual, as well as each intermediary, had to send the letter to some known contact that was thought to be closer to the destination. Sorprisingly, the number of steps of all letters was between 2 and 10.

Motivated from this work among others, Watts and Strogatz created a network model that combine Erdos Renyi and regular networks. These networks, called 'small–world' networks, have both the characteristic of a high clustering coefficient as the regular networks and a short average path length as ER networks. After this pioneering work, many papers emerged showing the small–world feature in several real networks.

Climate is not an exception. In the first climate network paper, in 2003, Tsonis and Roebber [45] showed that the network that underlies the intermediate levels of the atmosphere has small–world characteristics. While the tropics are almost fully connected, the middle latitudes show a scale–free degree distribution.

Scale–free networks The scale–free networks exhibit a potential law in their degree distribution:

$$P(k) \sim k^{\gamma} \tag{2.13}$$

Usually, γ takes values in the range $2 < \gamma < 3$. This means that the number of nodes of k degree grows with the number of nodes as $\log n$. Scale–free networks have, therefore, few nodes with many connections -that are called hubs or supernodes- and many nodes with low degree. This feature, as well as the small world property, is universally found in many real

networks.

2.2.2 Functional networks

The idea of using network analysis for complex systems is quite straightforward in the sense that it is easy to define each agent as a node and the interactions through the links. In constrat, when the system to be analized is a continuos field such as the level of activity in the brain or the surface temperature of the Earth, small portions (or regions), arbitrarily chosen, are considered as individual dynamical systems that interact with each other. Using characteristic variables as oxygen consumption measured using fRMI tools or sea surface temperature revealed by satellite observations in climate, it is possible to construct a functional network.

In particular, in climate, any field evolution is usually represented by a regular grid where each node has an associated time series that describes its evolution. Then, by evaluating the normalized statistical similarity S_{ij} among the time series of nodes v_i and v_j, it is possible to analyze if the nodes are linked or not. Obviously, there is always a statistical threshold τ that has to be overcome to discern an actual link from an spurious one.

Therefore, the adjacency matrix can be defined by the following expression:

$$A_{ij} = \Theta\left(|S_{ij}| - \tau\right) - \delta_{ij}, \tag{2.14}$$

where $\Theta()$ is the Heaviside step function and δ is the Kronecker delta.

In order to set the threshold value, we use the same criterion for every network, randomly relabeling the elements within each time series such that the distributions remain unchanged, and taking a fixed quantile value (99 %) of the statistical similarity measure of every possible combination as a threshold. Thus, the null hypothesis of independence can be rejected if the statistical similarity value of the original time series is above this threshold at a 0.01 level of significance. Once the adjacency matrix is defined, all the tools of complex networks can be used to get insights of the underlying dynamics.

2.3 Time-series analysis

As we present in the beginning of this chapter, climate information usually comes in the form of time series. In this work, particularly, these time series are regularly spaced in time. This fact simplifies our goal to explain the basis of the standard time series analysis tools that we have used along this thesis.

In this section, we briefly explain the statistical analysis tools that we use in the following chapters. We follow an increasing order in the complexity of the tools, from univariate to multivariate time series analysis techniques.

2.3.1 Simple linear regression

It is a method to estimate the linear relationship among two variables. It is assumed that one variable has to be independent and the other dependent. However, it is not neccesary to assume a causality relation among them. Then, if we take t and x as the independent and dependent variables respectively, the simple linear regression model is:

$$x(t) = \beta_0 + \beta_1 t, \tag{2.15}$$

where β_0 and β_1 are the regression parameters. Specifically, as we handle time series along this work, we have discrete relationships. Therefore, at each time step t_i we have the following equality:

$$x_i(t_i) = \hat{\beta}_0 + \hat{\beta}_1 t_i + \varepsilon_i, \tag{2.16}$$

where $\hat{\beta}_0$ and $\hat{\beta}_1$ are the estimators of the regression parameters; and ε_i is the residual at t_i. To calculate the estimators, the most commonly used approach is the least-squares method. Basically, the method finds the parameters that minimize the sum of the square of the residuals.

In this thesis, we use it for detrending time series to eliminate spurious correlations given by the global warming. Also, in Chapter 3, we estimate the linear component of variability patterns as ENSO and PNA in order to remove their influence in the Z200e networks.

2.3.2 Pearson correlation

The Pearson correlation coefficient (ρ) is one of the most used statistical similarity measures. It is a measure of the linear correlation between two variables x and y. It is defined as

$$\rho_{xy} = \frac{\mathrm{cov}\,(x\,,y)}{\sigma_x\,\sigma_y} = \frac{E\left[(x\,-\,\mu_x)\,(y\,-\,\mu_y)\right]}{\sigma_x\,\sigma_y}, \tag{2.17}$$

where $\mu_{x,y}$ and $\sigma_{x,y}$ are, respectively, the mean and standard deviations of x, y and E is the expected value operator. It can easily proved that $\rho \in [-1, 1]$. On the one hand, in the case that $|\rho| \approx 1$ it means that there is a strong correlation between both variables (negative in case that $\rho < 0$). On the other hand, if $\rho \approx 0$ it means that both processes are independent. In this case, considering x and y as synchronized time series of lenght T:

$$\rho_{xy} = \frac{\sum_{i=1}^{T}(x_i - \mu_x)(y_i - \mu_y)}{\sqrt{\sum_{i=1}^{T}(x_i - \mu_x)^2}\sqrt{\sum_{i=1}^{T}(y_i - \mu_y)^2}}. \tag{2.18}$$

where the mean values can be explicitly calculated as $\mu_x = \sum_{i=1}^{T} x_i/T$ and $\mu_y = \sum_{i=1}^{T} y_i/T$.

2.3.3 Persistence

It is also of interest to measure the Pearson correlation of the same variable but with a temporal lag τ. This is known as autocorrelation $R(\tau)$, or serial correlation, and it is used to get information about the *memory* of the underlying process. Considering now x as a time series with zero mean and unit variance, the equation is

$$R(\tau) = \frac{1}{t-\tau}\sum_{i=1}^{t-\tau} x_i x_{i+\tau}. \tag{2.19}$$

Given that most climate variables have a red noise spectrum, persistence, let us call it τ_p, is defined as the needed lag such that the autocorrelation drops below $1/e$. Therefore, it can be written as

$$R(\tau) \sim e^{-\frac{\tau}{\tau_p}}. \tag{2.20}$$

The number of degrees of freedom of a time series is affected by its persistence. In case

the persistence is large, the number of freedom degrees is lower than in the case that of a poorly autocorrelated time series. This is important when considering the significance of a Pearson correlation value. Specifically, given a value of Pearson correlation ρ, the statistic t that follows the t-Student distribution can be calculated as

$$t = \frac{\rho}{\sqrt{1-\rho^2}}\sqrt{N_{\text{degrees of freedom}}}, \tag{2.21}$$

to test the null hypothesis, H_0: $\rho = 0$.

2.3.4 Taxicab distance

Another interdependency measure is the *taxicab* distance. Given two synchronized time series x and y of lenght t, it is defined as

$$d = \frac{1}{t}\sum_{i=1} |y(i+\tau) - x(i)|. \tag{2.22}$$

where τ parameter is used to consider phase shifts between the variables.

We consider this specific metric in Chapter 4 because we are interested in measuring the similarity of the shape of x and y waveforms. To do that, the two variables are normalized to zero mean and unit variance.

2.3.5 Composite Analysis

Composite analysis is a widely used technique to determine some of the basic structural characteristics of a meteorological or climatological phenomenon which occur over time (e.g., the weather/climate over a given geographic area). In studying climate, composites can be quite useful for exploring the large scale impacts of teleconnections from modes of atmospheric variability such as ENSO.

In this thesis, we use composite analysis in Chapter 3. There, we aim to compare the Z200e large scale patterns in the SH under the influence of the different stages of ENSO. To do that, we consider the average of the Z200e field during each of the ENSO phases: El Niño (EN), La Niña (LN) and Neutral (N) years. Basically, we test the differences of the average values during both EN and LN years with respect to the average values during N

years. To assess the differences of the mean values we consider the $t*$ statistic at each node i of the grid:

$$t* = \frac{\bar{x_i}^{EN,LN} - \bar{x_i}^{N}}{\sqrt{n_1^{-1} + n_2^{-1}}} \sqrt{\frac{(n_1 - 1)\sigma_i^{EN,LN2} + (n_2 - 1)\sigma_i^{N2}}{(n_1 + n_2 - 2)}}. \tag{2.23}$$

Here, n_1 correspond to the size of the sample of EN or LN years, while n_2 is that of the N years. The statistic $t*$ is compared with the Student's t distribution with a two-tailed confidence interval of 95%, with H_0: $\rho = 0$ being the null hypothesis. Then, if the null hypothesis is rejected, it can be inferred that the average values of both distribution cannot be considered different wiht a 95% significance level. It is worth to mention that, in order to use this $t*$ statistic, it is neccesary for the compared distributions to have equal variance.

2.3.6 Emphirical Orthogonal Functions

The Emphirical Orthogonal Functions (EOFs) analysis, also known as Principal Component Analysis (PCA), was first used in meteorology in 1956 by E. N. Lorenz [46]. There, he uses this technique for the US sea level pressure forecast with data given by 64 meteorological stations.

The method, as the name implies, is used to generate a new basis of orthogonal functions by which we can represent the evolution of the whole data set. Particularly, let us call $\mathbf{X}_{n\times t}$ to a set of synchronized time series recorded in n different geographical locations, not necessarily spatially equidistant. Then, if $\mathbf{X}$ is normalized to have zero mean and unitary variance, the covariance matrix $\mathbf{C}_{n\times n}$ can be written as

$$\mathbf{C} = \mathbf{X}^T\mathbf{X}, \tag{2.24}$$

$\mathbf{C}$ can be diagonalized by solving the eigenvalue problem

$$\mathbf{VC} = \mathbf{\Lambda C}, \tag{2.25}$$

where $\mathbf{V}$ is the eigenvector matrix and $\mathbf{\Lambda}$ is the diagonal matrix containing the covariance eigenvalues. Since $\mathbf{C}$ is symmetric, the eigenvectors form an orthogonal basis on which the

data matrix $\mathbf{X}$ can be projected such that

$$X_i(t) = \sum_{j=1}^{n} P_j(t) E_{ji}, \tag{2.26}$$

where $P_j(t)$ are called principal components (PCs) and $\vec{E_j}$ are the EOFs of the field $\mathbf{X}$. As the EOFs are of the same dimension that the spatial dimension of the database, they can be plotted as maps. The temporal evolution of these structural patterns are given by the PCs.

It can be proved that the trace of $\mathbf{\Lambda}$ is equal to the variance of $\mathbf{X}$, thus the coefficient

$$\nu_j = \frac{\Lambda_{jj}}{\mathrm{Tr}\,(\mathbf{\Lambda})}, \tag{2.27}$$

represent the variance fraction explained by each mode of variability $P_j(t).\vec{E_j}$. The first EOF is the combination of variables that maximizes the variance. The second EOF maximizes the remaining variance being perpendicular to the first one, and so on until all the dimension D of the problem is covered ($D = \min[n, t]$). The great advantage of this methodology is that the first EOFs account for a large percentage of the total system variance. Thus, by considering only a few EOFs, it allows to reduce the dimensionality of the problem without losing much information about the variability of the system.

2.4 Clustering technique

2.4.1 K-means method

The K-means clustering is an algorithm for separating N data points in an D-dimensional space into K clusters. Each cluster is characterized by its correspondent D-dimensional centroid $\mathbf{c}$.

Let us call the data points by $\{x^n\}_{n=1}^{N}$. Each point x has D components. The method considers an Euclidean metric such that the distance d between two points $\mathbf{x}$ and $\mathbf{y}$ is defined by

$$d(\mathbf{x}, \mathbf{y}) = \frac{1}{2} \sum_{i=1}^{D} (x_i - y_i)^2. \tag{2.28}$$

The algorithm begins by arbitrary chosing K centroids $\mathbf{c}$, usually randomly. Then, the iteration consists in two steps: assignment and update. The assignment step consists in assigning the closest centroid to each one of the N points. Once all the points are assigned to a centroid, each centroid is updated by modifying its position in order to decrease the overall distance of all the points within the cluster. These two steps are repeated until there is no modification in the assignment of the points.

A particular feature of this method is that the number K of clusters needs to be defined *a priori*. Therefore, it is necessary to implement some criterion to select the number of clusters that better suits the particular problem. In our case, we consider the algorithm in Chapter 3. There, we explain in detail our methodology based on previous works [47, 48].

2.5 Information theory measures

In 1948, Claude Shannon published a paper entitled 'A Mathematical Theory of Communication' [49]. This influential article was the founding work of the field of information theory. Basically, he set the basis of communication analyzing the route of a message from its original source to the destination, taking into account the aparition of noise.

Here, we consider some measures from information theory to analyze climate time series. In this section we briefly describe the mutual information and the Shannon entropy that we use respectively in Chapter 2 and Chapter 4. Furthermore, we explain different ways of building the probability density functions from discrete time series; using histograms (classical view) and using symbolic analysis, that is used to focus on different time scales in Chapter 2.

2.5.1 Shannon entropy

Shannon entropy is a measure of the degree of disorder or unpredictability of a given probabilistic stochastic process. As well as Boltzmann entropy, from which it takes the name, it depends on the possible states of the system. It is defined by the probability density function $p(x)$, where x represent a state of all the possible states of the system, say χ. Then, entropy

H is defined by:

$$H = -\int_{x\in\chi} p(x)\log p(x)dx. \tag{2.29}$$

In this case, as we only consider discrete time series, the definition of H is

$$H = -\sum_{x\in\chi} p_x \log p_x. \tag{2.30}$$

Then, a process with N equiprobable states, *i.e.* $p_x = 1/N$, has maximum entropy. This means that there is no *a priori* information to predict the evolution of the process. On the contrary, if there is clearly one state with higher probablity of occurrence $p_x \sim 1$, H tends to 0 given that the process is very likely to be in that state. Note that only states with non-zero probability can be considered given that the sum in Eq. 2.30 would be undefined otherwise.

2.5.2 Mutual information

Mutual information is a non linear interdependency measure between two processes, say x_i and x_j. According to information theory, it is a measure of how much information about x_i is gained by knowing x_j. It is defined by the following equation:

$$M_{ij} = \sum_{m,n} p_{ij}(m,\, n)\, \log \frac{p_{ij}(m,\, n)}{p_i(m)\, p_j(n)}, \tag{2.31}$$

where p_i and p_j are the marginal probability density functions of x_i and x_j, respectively, and p_{ij} the joint probability density function.

If the processes are independent of each other, $p_{ij}(m,\, n) = p_i(m)\, p_j(n)$, then $M_{ij} = 0$. It is important to notice that mutual information, as well as the linear correlation and taxicab distance, are symmetric quantities.

2.5.3 PDF building

Shannon entropy and mutual information are defined by the pdfs of the processes. Therefore, the way in which the pdf is built is crucial for the analysis.

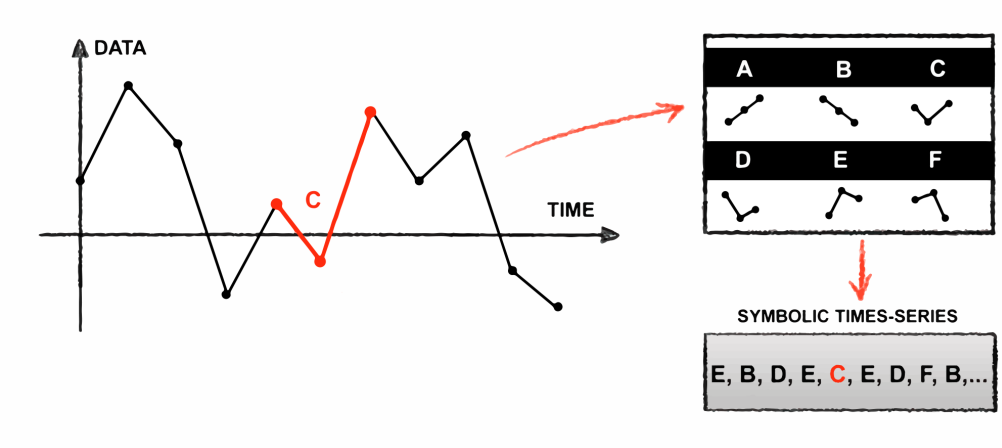

Figure 2-2: Ordinal pattern calculation. Ilustration credits: Ana Revello.

Histograms

Usually, the pdf of a time series is built using traditional histograms. The numbers of bins in the histogram are chosen according to the time series length to ensure good statistics. This way, the marginal probabilities don't have any temporal information about the process and there is no emphasis on any particular timescale.

Symbolic analysis

In order to focus on different time scales, we use a symbolic analysis method by Bandt and Pompe [50], called ordinal patterns. As its name refers, the method consists in building the pdfs by analyzing the patterns formed by consecutive values in the time series. Specifically, each time series $x_i(t)$ is divided into $T-D$ overlapping D dimensional vectors. The values within each vector are compared with each other and replaced by their correspondent ordinal position, 0 for the lowest and $D-1$ for the largest. The possible configurations are $D!$ and each one can be replaced by a symbol (i.e. letters: A, B, C,...). For instance, if $D=3$, the possible configurations are (012 021 102 120 201 210), and, for example, the sequence (2.7, 0.1, 8.3) corresponds to the 102 pattern because $x_i^3 > x_i^1 > x_i^2$, see Fig. 2-2.

The BP methodology needs only a weak stationary assumption that the probability for $x_i^t < x_i^{t+D}$ should not depend on t. This is satisfied by removing the linear trend in the data. Another condition for calculating the pdfs adequately is that $T \gg D!$; the number of vectors

must be much larger than the number of possible symbols.

There is an embedding dimension given by the D parameter of the BP method that fixes the causality timescale. This is the main difference from the classical histogram-based technique used to build the probability density functions. An important advantage is that it is possible to build sequences of patterns not only with consecutive values but with an arbitrary time interval Δt between them. In Chapter 2, we have chosen to work with $D = 3$ and $\Delta t = 1$-, 4- and 12-month intervals to characterize atmospheric processes on intra-seasonal, intra-annual and inter-annual timescales, respectively.

Chapter 3

Evolution in Z200e networks during the 20th century[1]

In this chapter, we aim to study the evolution of the upper atmosphere connectivity over the 20th century as well as to distinguish the connectivity associated with the oceanically forced component from the one due to atmospheric internal variability. For this purpose we build networks from two different reanalysis data sets using both linear and nonlinear statistical similarity measures to determine the existence of links between different regions of the world in the two halves of the last century. We furthermore use symbolic analysis to emphasize intra-seasonal, intra-annual and inter-annual timescales. Both linear and nonlinear networks have similar structures and evolution, showing that the most connected regions are in the tropics over the Pacific Ocean. Also, the Southern Hemisphere extratropics have more connectivity in the first half of the 20th century, particularly on intra-annual and intra-seasonal timescales.

Changes over the Pacific main connectivity regions are analyzed in more detail. Both linear and nonlinear networks show that the central and western Pacific regions have decreasing connectivity from early 1900 up to about 1940, when it starts increasing again until the present. The inter-annual network shows a similar behavior. However, this is not true for other timescales. On intra-annual timescales the minimum connectivity is around 1956, with a negative (positive) trend before (after) that date for both the central and western Pacific. While this is also true of the central Pacific on intra-seasonal timescales, the western

[1]The results of this chapter are published in F. Arizmendi, A. Martí and M. Barreiro, NPG, 21, 2014 [51].

Pacific shows a positive trend during the entire 20th century.

In order to separate the internal and forced connectivity networks and to study their evolution through time, an ensemble of atmospheric general circulation model outputs is used. These results suggest that the main connectivity patterns captured in the reanalysis networks are due to the oceanically forced component, particularly on inter-annual timescales. Moreover, the atmospheric internal variability seems to play an important role in determining the intra-seasonal timescale networks.

3.1 Introduction

In the last 15 years, since the pioneering works on complex networks [43, 52], the use of this statistical methodology of analysis has allowed significant advances in a broad variety of scientific disciplines, including social systems, the internet, neural networks, metabolic networks, and gene expressions, among many others [53, 42].

The study of the climate system has also benefited from the application of this new theoretical framework [45, 54, 55, 56, 57], providing valuable insights into climate dynamics. As it is well known, the climate system is characterized by phenomena with many different temporal and spatial scales, so that a variety of network approaches have been proposed to study different modes of variability. For example, a large number of articles has been dedicated to the analysis of the El Niño–Southern Oscillation (ENSO) phenomenon [58, 56, 59, 60] and how it affects the connectivity of the network [61].

El Niño has been shown to influence the climate of far-away regions through the excitation of anomalous wave trains that propagate longitudinally within the tropics and meridionally to the extratropical regions. The theory that tries to explain these remote connections, or teleconnections, is mainly based on linear wave dispersion of Kelvin and Rossby waves [62, 63]. Even though there have been several attempts to improve this description, including the influence of El Niño through other mechanisms [64], the understanding of atmospheric teleconnections is not yet complete.

Here we aim to deepen the understanding of atmospheric teleconnections, analyzing the Z200e field from a complex network perspective. This level is chosen because it is the height

of the maximum tropical divergence that is crucial for the excitation and propagation of tropical and extratropical anomalies like Rossby wave trains [27]. A second objective of this study is to analyze potential changes in the atmospheric connections during the last century. We focus on the connectivity of the tropical Pacific because it has been shown by several studies [55, 57] to be globally the most connected region.

In order to accomplish these objectives, we analyze two atmospheric reanalysis data sets, which represent the best estimate of the historical atmospheric evolution [65] and an ensemble of experiments performed with an atmospheric general circulation model (AGCM). The networks determining the connectivity among nodes are built in terms of the statistical similarity of the corresponding time series. We use both linear and nonlinear statistical measures [55, 57]: the linear Pearson correlation (PC) and the nonlinear mutual information (MI). Alternatively, as the PC and the MI obtained from histogram-based probability density functions (PDFs) do not take into account the temporal order of the time series, we introduce the ordinal patterns methodology proposed by [50] (BP). This method allows us to consider different timescales with the MI as in [57]. In general, the networks are represented graphically as two-dimensional maps by plotting the *area-weighted connectivity* AWC, which represents the fraction of the total area of the Earth to which each node is connected.

The total atmospheric variability can be decomposed in a component due to the intrinsic dynamics of the atmosphere that would occur even in the absence of changes to the boundary conditions, and another component due to the forcing of the surface ocean conditions. The latter is potentially more predictable, as it depends on the evolution of the sea surface temperature, which varies on much longer timescales than the atmospheric fields. In some regions, like the tropical Pacific, the ocean–atmosphere interaction allows the prediction of sea surface temperature anomalies (i.e., the development of El Niño) by up to 6–9 months in advance [66]. It is thus extremely important for seasonal climate prediction to be able to separate and determine the dynamics that characterizes the intrinsic and forced atmospheric variability. In this study we use the AGCM output to determine if the observed 20th century changes in connectivity are related to the forced or intrinsic dynamics.

Our results show that the most connected regions are in the tropical Pacific, and the overall network structure is broadly similar when using both the linear and nonlinear statis-

tical measures when no timescales are considered. However, using the BP methodology, the nonlinear symbolic analysis reveals several structural changes in the networks. Some regions become more relevant depending on the timescale considered, such as the Southern Hemisphere, relatively more important on the intra-seasonal timescales, or the northern Pacific, and more relevant on the intra-annual timescales.

Moreover, we show that the evolution of the connectivity in the Pacific depends on the timescale. For example, in the central Pacific the connectivity is smallest around 1956 for intra-seasonal and intra-annual timescales, with negative (positive) trends before (after) that date. On inter-annual timescales, however, the connectivity decreases from early 1900 up to around 1940 and increases afterwards.

3.2 Data and methodology

3.2.1 Data

We use monthly average geopotential height at 200 mb from the NOAA 20th Century Reanalysis [29] and from the NCEP CDAS1 reanalysis data [12]. It is important to keep in mind that a reanalysis is constructed blending observations into an AGCM, and thus the value of the reanalysis compared to a simple model run increases where there are enough observations to constrain the model solution. Prior to 1950, observations were very limited globally, and thus the reanalysis may not represent so much reality as it represents the model solution. In the Southern Hemisphere the situation is even more difficult because there were very few observations prior to the satellite era that began in 1979. Given these limitations the results of this study must be considered as a first approximation of reality based on the best information available to date about the evolution of the 20th century atmospheric circulation.

In addition, we consider the output of the atmospheric general circulation model from the International Center for Theoretical Physics (ICTP-AGCM) forced with global historical sea surface temperature (ERSSTv.2, [35]). The ICTP-AGCM is a full AGCM with simplified physics and a horizontal resolution of T30 with eight vertical levels [36, 37]. The performance of the ICTP-AGCM in representing the large-scale atmospheric dynamics is comparable

Table 3.1: Characteristics of data. N: number of grid points, T: number of months.

Source	Resolution	Time period	Dimensions	
NCEP CDAS1	2.5°	[1949, 2011]	$N = 10\,224$	$T = 756$
NOAA 20th Century	2°	[1901, 2009]	$N = 16\,020$	$T = 1308$
ICTP AGCM	3.75°	[1901, 2006]	$N = 4980$	$T = 1272$
EN34 index	–	[1901, 2006]	$N = 1$	$T = 1308$

to that of state-of-the-art AGCMs and has been used to study global climate variability, including the influence of the oceans on the circulation [38, 39, 40, 41].

Reanalysis and model output are distributed in regular latitude–longitude grids with every point characterized by a time series of the variable of interest. The characteristics of each data set are summarized in Table 3.1.

The average geopotential height at 200 mb has a meridional structure dominated by relatively large values in the tropical regions that decrease toward the poles. To highlight connections that are deviations from this zonal structure, we consider the eddy geopotential height that is calculated by removing the zonal mean of each latitudinal band for every month. We furthermore consider monthly anomalies; that is, the mean annual cycle is removed prior to analysis, and the time series are linearly detrended at each grid point.

3.2.2 Network building

To build the network, every pair of points over the grid is linked whenever a significant statistical similarity between the correspondent time series is detected. Let $\vec{x}(t) = \{x_i(t)\}_{i=1}^{N}$ with $t = 1, \ldots, T$ be the Z200e anomaly field and $S(x_i, x_j) = S_{ij}$ the similarity measure between the nodes i and j. The elements of the climate network adjacency matrix $\mathbf{A}$ are then

$$A_{ij} = \Theta\left(|S_{ij}| - \tau\right) - \delta_{ij}, \tag{3.1}$$

where $\Theta()$ is the Heaviside step function, δ is the Kronecker delta and τ is a statistical threshold introduced to avoid spurious connections.

In order to set the threshold value, we use the same criterion for every network, randomly relabeling the elements within each time series such that the distributions remain unchanged, and taking a fixed quantile value (99 %) of the statistical similarity measure of every possible

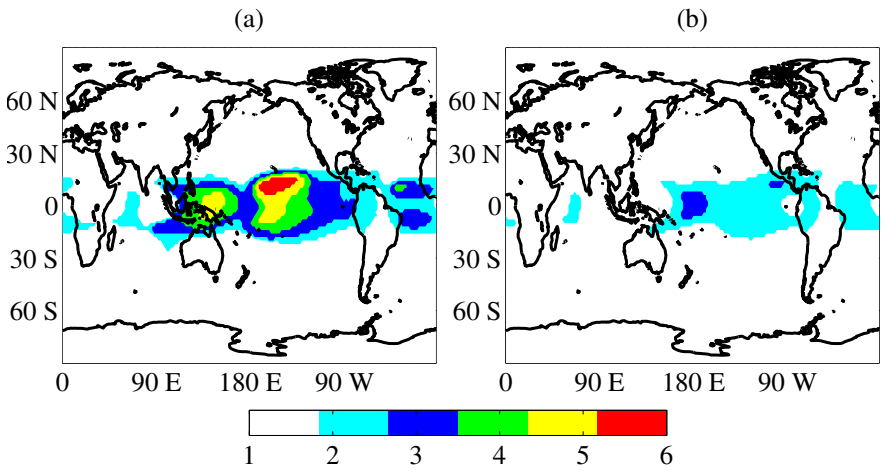

Figure 3-1: Persistence of the Z200e field of the 20th Century Reanalysis data from NOAA with **(a)** and without the EN34 component **(b)**.

combination as a threshold. Thus, the null hypothesis of independence can be rejected if the statistical similarity value of the original time series is above this threshold at a 0.01 level of significance.

Recently, Palus et al. [67] showed that two time series that have some autocorrelation, or persistence (defined as the lag time when the autocorrelation drops below $1/e$), induce a bias in the connectivity measure. To avoid this supposed bias, [67] suggest considering the persistence of the time series when calculating the connectivity threshold, and show that the network topology differs significantly when this is taken into account. In climate data, the persistence at a particular location is due to local and remote processes. The part of the persistence due to local processes is the one that should be considered in the connectivity threshold because it introduces a bias in the connectivity. However, the persistence due to remote processes is a real connection that occurs through atmospheric circulation anomalies and should be considered when calculating connectivity measures as the area-weighted connectivity used here (see Sect. 2.3).

The persistence of the Z200e of the NOAA 20th Century Reanalysis data is shown in Fig. 1a, where it can be seen that the persistence is 1 month for the extratropics and from 2 to 6 months in the tropical band, with a maximum in the Pacific. This result reinforces the large internal atmospheric variability present in the extratropics and the dominant role

of the ocean forcing in the tropical band. Figure 1b shows the persistence after the effect of El Niño (El Niño 3.4 index) has been linearly removed by a regression procedure. Clearly, most of the persistence in the tropical band is due to the El Niño phenomenon that is known to affect the global tropics through the propagation of Kelvin waves [68]. Given that El Niño persists for about 6 months, the connections from the tropical Pacific to the rest of the tropics are also maintained. After El Niño is removed, the persistence in the tropics is about 2 months, which might be because the linear procedure was not enough to remove the El Niño influence completely or due to the existence of local processes that slightly enhance persistence. Nevertheless, it is clear that a persistence of 1 month is characteristic of most of the world, and thus randomly mixing the time series on individual points of the grid provides a good test of the connectivity. This procedure ensures the recognition of the importance of the tropical Pacific in the global connectivity, as clearly seen in Fig. 2.

We constructed networks using linear and nonlinear measures of connectivity. Networks made of purely linear interaction are constructed by evaluating the statistical interdependency between the nodes with the Pearson correlation:

$$\rho_{ij} \;=\; \frac{\mathrm{cov}\,(x_i\,,x_j)}{\sigma_i\,\sigma_j} \;=\; \frac{E\,[(x_i\;-\;\mu_i)\,(x_j\;-\;\mu_j)]}{\sigma_i\,\sigma_j}, \tag{3.2}$$

where $\mu_{i,j}$ and $\sigma_{i,j}$ are the mean and standard deviations of the time series i,j. Since we consider anomalies, $\mu_i = 0$ for all $i \in [1,\, N]$.

Nonlinear connectivity is evaluated using a quantity from information theory, the mutual information:

$$M_{ij} \;=\; \sum_{m,n} p_{ij}(m,\,n)\;\log\;\frac{p_{ij}(m,\,n)}{p_i(m)\,p_j(n)}, \tag{3.3}$$

where p_i and p_j are the marginal probability density functions of x_i and x_j, respectively, and p_{ij} the joint probability density function. According to information theory, mutual information is a measure of how much information about x_i is gained by knowing x_j. It is easy to see that if the time series are independent, $p_{ij}(m,\,n) = p_i(m)\,p_j(n)$, and therefore $M_{ij} = 0$. It is important to notice that both mutual information and linear correlation are symmetric quantities.

The usual way of obtaining the PDF of a time series is by building traditional histograms.

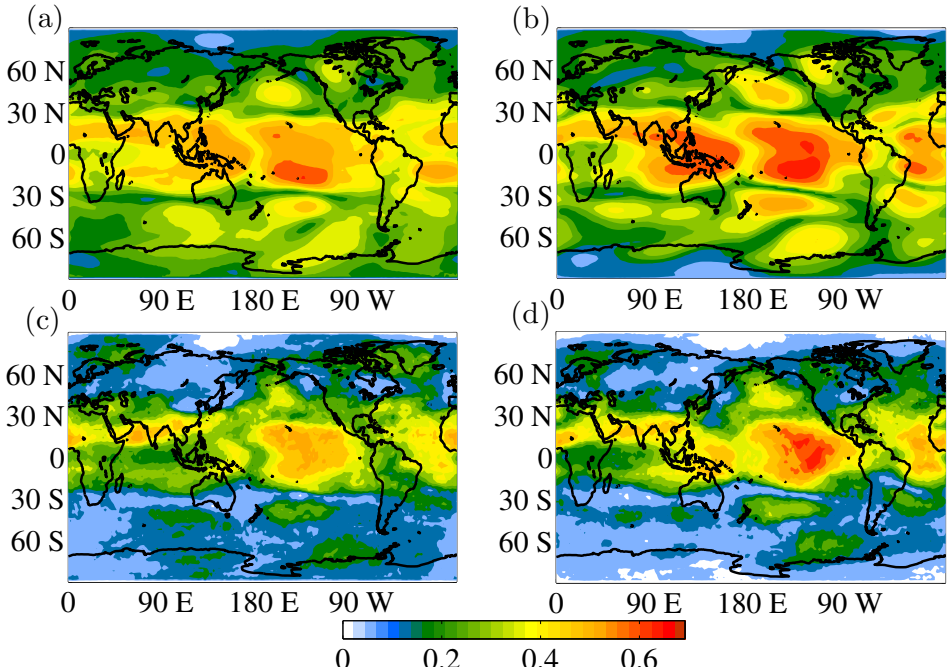

Figure 3-2: AWC maps of the Z200e field with the graph construction methodology applied to the 20th Century Reanalysis data from NOAA. Statistical similarity measures and periods taken: **(a)** PC (1901–1955), **(b)** PC (1955–2009), **(c)** MI with classical histograms (1901–1955) and **(d)** MI with classical histograms (1955–2009).

Generally, the numbers of bins in the histogram are chosen according to the time series length to ensure good statistics. One important characteristic of the use of histograms is that the time-related information is only considered in the joint probability because, when using histograms to represent the marginal probabilities, all the temporal information is neglected and, as with linear correlation, there is no emphasis on any particular timescale.

To overcome this limitation, the Bandt and Pompe (BP) methodology [50] based on symbolic analysis is employed to separate different timescales. This symbolic analysis consists of building the PDFs by analyzing the patterns formed by consecutive values in the time series. Specifically, each time series $x_i(t)$ is divided into $T-D$ overlapping D dimensional vectors. These vectors are represented as ordinary patterns (OPs), in the sense that each element is replaced by a number from 0 to $D-1$, depending on its ordinal position (i.e., 0 for the lowest and $D-1$ for the largest). Each vector is then assigned one of the $D!$ possible configurations. For instance, if $D=3$, the possible configurations are (012 021 102 120 201 210), and, for example, the sequence (2.3, 0.7, 8.7) corresponds to the 102 pattern

because $x_i^3 > x_i^1 > x_i^2$. The BP methodology needs only a weak stationary assumption that the probability for $x_i^t < x_i^{t+D}$ should not depend on t. This is satisfied by removing the linear trend in the data. Another condition for calculating the PDFs adequately is that $T \gg D!$; i.e., the number of vectors must be much larger than the number of possible OPs.

There is an embedding dimension given by the D parameter of the BP method that fixes the causality timescale. This is the main difference from the classical histogram-based technique used to build the probability density functions. An important advantage is that it is possible to build sequences of patterns not only with consecutive values but with an arbitrary time interval Δt between them. In this study we have chosen to work with $D = 3$ and $\Delta t = 1$-, 4- and 12-month intervals to characterize atmospheric processes on intra-seasonal, intra-annual and inter-annual timescales, respectively.

3.2.3 Area-weighted connectivity

The network representation of a climate field helps us to obtain valuable insight into the climate dynamics; however, it is clear that the specific procedure to build the network is decisive in the information it provides. Moreover, once the network is built, there are different analyses that can be done in order to deepen our understanding of the dynamics that is being analyzed, such as calculating the betweenness centrality for each node [55] or detecting the underlying communities in the system [69].

Following the procedure proposed in previous studies on climate networks [54, 55, 57], to analyze the changes in the atmospheric connectivity in the 20th century, we use the *area-weighted connectivity* (AWC) function

$$AWC_i = \frac{\sum_{j=1}^{N} A_{ij} \cos(\theta_j)}{\sum_{j=1}^{N} \cos(\theta_j)}, \tag{3.4}$$

where θ_j is the latitude of node j and the cosine factors are included to take into account the sphericity of the Earth; that is, for a regular planar grid embedded in an spherical surface, the nodes correspond to different areal regions. The AWC provides the fraction of the Earth that each node is connected to, similar to the k degree usual in complex networks, but area

weighted and normalized ($\mathrm{AWC}_i \in [0, 1]$).

This study is concerned with analyzing the changes in the atmospheric connectivity during the last hundred years, which, as will be shown below, is dominated by the tropical Pacific. Thus, to further understand the changes in these regions, we calculate the evolution of the AWC in two boxes in the central and western Pacific using the NOAA 20th Century Reanalysis data.

Specifically, we considered the nodes in the following regions:

- Western Pacific (WP): [3–7° S, 129–135° E]
- Central Pacific (CP): [3–7° S, 151–157° W].

The evolution is measured by taking 30-year sliding windows in the steps of 1 year and calculating the average AWC for each region.

3.2.4 Intrinsic and forced atmospheric variability

To aid the interpretation of the networks obtained with reanalysis data, we consider the output of the ICTP-AGCM. In particular, as explained in the introduction, we are interested in separating the networks associated with intrinsic (or internal) atmospheric variability and with oceanically forced atmospheric variability, so we constructed an ensemble of 9 ICTP-AGCM runs where the model is forced with global historical sea surface temperatures as boundary conditions and slightly different atmospheric initial conditions. Given the chaoticity of the atmospheric dynamics, this setup generates $n = 9$ equiprobable realizations of the atmospheric evolution.

The procedure to separate internal from forced atmospheric variability is as follows. We start with the model outputs, whose characteristics are detailed in Table 3.1 as $\vec{x}_j^m(t) = \{x_{ji}^m(t)\}_{i=1}^N$, with $j = 1, \ldots, n$ as the different realizations. The evolution for each realization j at each grid point can then be separated into two components

$$\vec{x}_j^m = \vec{x}_{\mathrm{F}}^m + \vec{x}_{\mathrm{I}j}^m, \tag{3.5}$$

where $\vec{x}_{\mathrm{F}}^m$ is the oceanically forced component and $\vec{x}_{\mathrm{I}j}^m$ is the evolution due to atmospheric

internal dynamics (independent of the surface boundary conditions). We consider that, to first order, the two components do not interact [39].

Thus, assuming that the internal variability is filtered by averaging, the forced component can be estimated by taking the ensemble mean of the runs:

$$\vec{x}_{\mathrm{F}}^{m} \approx \frac{1}{n} \sum_{j}^{n} \vec{x}_{j}^{m}. \tag{3.6}$$

Once the forced component is estimated, the internal variability in each run can be approximated by applying Eq. (3.5). To characterize the network of atmospheric internal variability, we calculated the AWC for every $\vec{x}_{\mathrm{I}j}^{m}$ and then computed the average over the n AWC values.

3.2.5 Principal component analysis

To aid further understanding of the role of the tropical Pacific in setting the connections, we removed its contribution and constructed the AWC maps again. To do so, we built the AWC maps after removing the first two leading empirical orthogonal functions (EOFs) calculated for the tropical Pacific. The difference between the AWC maps after these EOFs are removed allows us to associate regions of large connectivity with known modes of variability.

We calculated the EOFs of the NOAA 20th Century Reanalysis data in the region [90° E–70° W, 30° N–30° S] and considered the two eigenvectors ($\vec{v}_1$, $\vec{v}_2$) with the largest eigenvalues ($\lambda_1 > \lambda_2 > \dots$) of the covariance matrix, which together explain about a quarter of the total variance.

The principal components $p_i(t)$ are obtained by projecting the original field onto the eigenvectors:

$$p_i(t) = \vec{x}'(t) \cdot \vec{v}_i, \tag{3.7}$$

where $\vec{x}'(t)$ is the Z200e field within the tropical Pacific.

To calculate the pattern associated with each EOF over the whole globe, we calculated the linear regression of the original field $\vec{x}(t)$ onto $p_i(t)$, such that the parameters $\hat{a}$ and $\hat{b}$ that best estimate the relation of time series x_j with p_i,

$$\hat{x}_j = \hat{a}_j \quad p_i + \hat{b}_j, \tag{3.8}$$

are obtained with the least squares approach. The significance is assessed by evaluating the correlation and comparing the statistics $T^{\mathrm{s}} = |\rho|\sqrt{T-2}/\sqrt{1-\rho^2}$ with the Student's t distribution with a two-tailed confidence interval of 95 %, with H_0: $\rho = 0$ being the null hypothesis. Assuming a red noise type of spectrum, the effective number of degrees of freedom for each node is calculated as $T^* = (T/2)\,\log(a)$, where a is the correspondent serial correlation value at a lag equal to 1 month.

3.3 Results

3.3.1 Area-weighted connectivity maps

Being our goal to analyze changes in the structure of atmospheric connectivity on different timescales since the beginning of the 20th century, we divide the NOAA 20th Century Reanalysis data into two equal-length halves: January 1901–December 1955 and January 1955–December 2009. The resulting networks built with Pearson correlation (PC) and mutual information with the probability density functions based on classical histograms (MIH) are represented by the AWC function in Fig. 2.

The spatial structure of the maps obtained with the different measures of connectivity is globally very similar when using both statistical measures. The main difference lies in the number of connections that represent the number of significant links, which is greater in the PC networks than in the MIH networks. The links density δ,

$$\delta = \sum_{ij}^{N} \frac{A_{ij}}{N(N-1)} \; : \; \delta \in [0,\,1], \tag{3.9}$$

is defined to quantify this property. The networks constructed with PC present $\delta \approx 0.30$, while the corresponding value with MIH is $\delta \approx 0.20$.

In all the cases, the most connected areas are located in the tropics, particularly in the central Pacific and to a lesser extent over the Atlantic Ocean, southern Asia and over Indonesia between the Indian Ocean and the western Pacific. This larger connectivity in the tropics found with both methodologies agrees with our current understanding of tropical dynamics. Likewise, but weaker in the sense that they vanish for lower threshold values

than in the tropics, there are some extratropical regions that show a relatively large number of connections, such as the South and North Pacific. The wavy patterns suggest that they reflect the teleconnection patterns in both hemispheres through the propagation of Rossby waves.

The area of largest connectivity in the central Pacific shows differences according to the methodology: while using PC, the area of maximum connectivity is to the south of the Equator at about 170–130° W; using MIH the (less well-defined) maximum is centered on the Equator further to the east. Nevertheless, both methodologies show that the central Pacific connectivity is largest in the second half of the century.

Some additional observations can be made from Fig. 2. Firstly, we note that there are more significant links in the Southern Hemisphere extratropics than in the Northern Hemisphere, especially in the first period. A structure, displaying the signature of Rossby waves propagating from the tropics through the southern Pacific to the South American continent, can be distinguished in all the panels. The increase in the connectivity in the western Pacific and South Asia in the second period can also be distinguished. The opposite occurs in the southern Indian Ocean west of 90° E, which decreases its connectivity, especially with the MIH. These variations suggest that the excitation of Rossby wave trains that propagate from the Indian Ocean sector toward the South Pacific was stronger from 1901 to 1955.

3.3.2 Ordinal pattern analysis

Figure 3 presents the results of the AWC calculated with the BP methodology for intra-seasonal, intra-annual and inter-annual timescales using $D = 3$ as the embedding dimension. Irrespective of the period considered, the tropics always dominate the connectivity on all timescales. However, the connectivity of the tropical region increases with the timescale, while that of the extratropics does not vary to the same extent. In fact, connectivity in the northern extratropics is largest on intra-annual timescales (particularly in the northern Pacific). In the southern extratropics the regions of maximum connectivity are similar, but increase with the timescale, suggesting a connection with the tropics. Note the existence of a region of large connectivity centered at 60° S only on inter-annual timescales, also seen in the

AWC maps calculated with PC and MIH in Fig. 2, commented in the previous subsection. This suggests that the symbolic analysis is successful in separating processes according to their timescales and that connectivity in Southern Hemisphere extratropics is strongest on inter-annual timescales. Regarding links density, $\delta \approx 0.14$ for intra-seasonal networks and $\delta \approx 0.16$ for intra-annual and inter-annual networks.

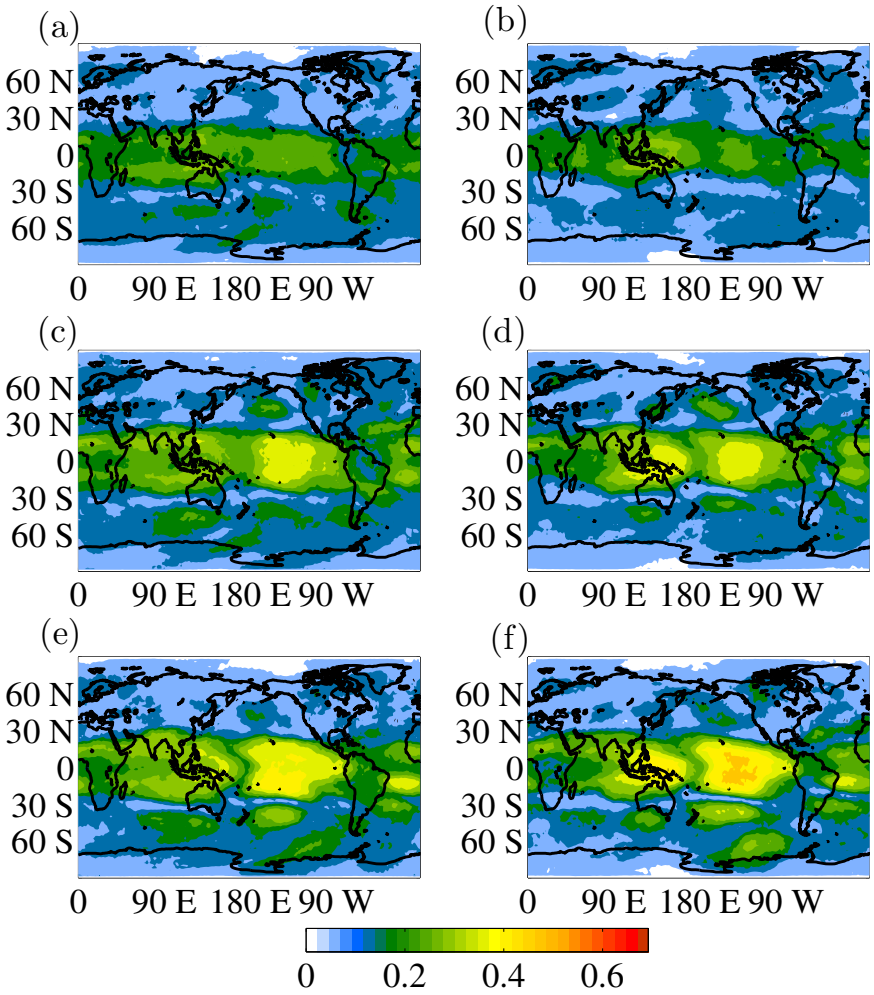

Figure 3-3: AWC maps of the Z200e field obtained from the MI using the BP symbolic analysis for different timescales: intra-seasonal ($\Delta t = 1$) (1901–1955) **(a)** and (1955–2009) **(b)**, intra-annual ($\Delta t = 4$) (1901–1955) **(c)** and (1955–2009) **(d)**, and inter-annual ($\Delta t = 12$) (1901–1955) **(e)** and (1955–2009) **(f)**.

Globally, on intra-seasonal timescales, $\Delta t = 1$ month, the area of larger connectivity in the first period is over the tropical Pacific and between southern Asia and northern Oceania,

while in the second half, the global maximum is in the western Pacific and the central Pacific region is less connected. Also, the first period shows larger connectivity areas in the southern Indian Ocean and southern extratropics, with similar differences in structure between the two periods as with PC and MIH (Fig. 2). This behavior is also valid for longer timescales.

In the case of intra-annual timescales, $\Delta t = 4$ months; there are more links in the Northern Hemisphere, particularly over oceanic areas, and there are also some differences in the connectivity of the tropical Pacific in both time periods. While in the eastern sector during the second period there is lower connectivity, the opposite is true in the western region, such as with intra-seasonal networks. On the other hand, both the central and western Pacific are more connected during the second half of the 20th century than during the first half on inter-annual timescales; $\Delta t = 12$ months.

This different behavior in the evolution of the connectivity in certain regions that depend on the timescale can also be noticed over Canada, with larger connectivity in the first period in the intra-annual networks, while in the inter-annual one it is the other way around. Intra-annual and intra-seasonal networks also show larger connectivity in the first half of the century over southern Brazil, contrary to what is seen in the PC and MIH networks in Fig. 2. These results suggest different evolutions of atmospheric processes on different timescales during the period of study, and suggest that the networks from PC and MIH mainly represent inter-annual connections in the tropical band and the Southern Hemisphere.

The hemispheric asymmetries in the AWC are quantified in Fig. 4, where the longitudinal sum of AWC versus latitude is plotted. All the curves present a local minimum on the Equator and two similar maxima near 10° S and 10° N. Clearly, the tropical connectivity increases with the timescale, and is larger in the second half of the century on long timescales. The plots also show that the relatively larger connectivity seen in the Southern Hemisphere during the 1901–1955 period is maximum on intra-seasonal timescales and decreases on longer timescales.

In order to test the robustness of the results, we have also constructed the AWC using the Z200e of the NCEP CDAS1 reanalysis [12] during the same time period as the second half of the NOAA data. Figure 5 shows the AWC maps derived from the Pearson correlation and the mutual information with the symbolic analysis. The values of δ obtained are in good

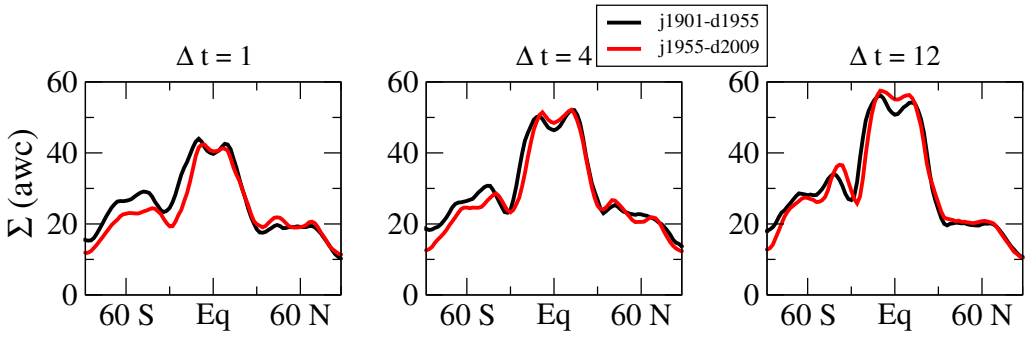

Figure 3-4: Zonal sum of the AWC for the NOAA networks built by using the BP methodology with $\Delta t = 1$, $\Delta t = 4$ and $\Delta t = 12$, for both time periods (black: 1901–1955, red: 1955–2009).

agreement with the maps shown in Figs. 2 and 3. The maps recover the same characteristics obtained for the NOAA 20th Century Reanalysis, for example relatively larger connectivity in the northern Pacific on intra-annual timescales, a weakly connected central Pacific on intra-seasonal timescales and a strongly connected Southern Hemisphere extratropical region on inter-annual timescales.

3.3.3 Temporal evolution of the AWC

We study further the time evolution of the connectivity for the western Pacific (WP) and central Pacific (CP) regions defined above. To enhance statistics, especially with the mutual information calculated with symbolic analysis, we took 30-year windows such that $T = 360$ months.

Figures 6 and 7 show plots of AWC versus time for all the different networks for the WP and CP zones, respectively. The behavior of the AWC for both regions depends on the methodology used. Basically, the overall evolution in the PC, MIH and inter-annual networks is very similar, but intra-annual and intra-seasonal networks behave differently. In the western Pacific, Fig. 6, the mean connectivity of the WP box presents a general decrease for the PC, MIH and intra-seasonal networks from the beginning of the century up to approximately 1940, when the connectivity starts an increasing trend until the end of the century. The connectivity of the intra-annual network shows a small negative trend during the first decades, and from 1956 onwards it starts increasing up to the end of the record.

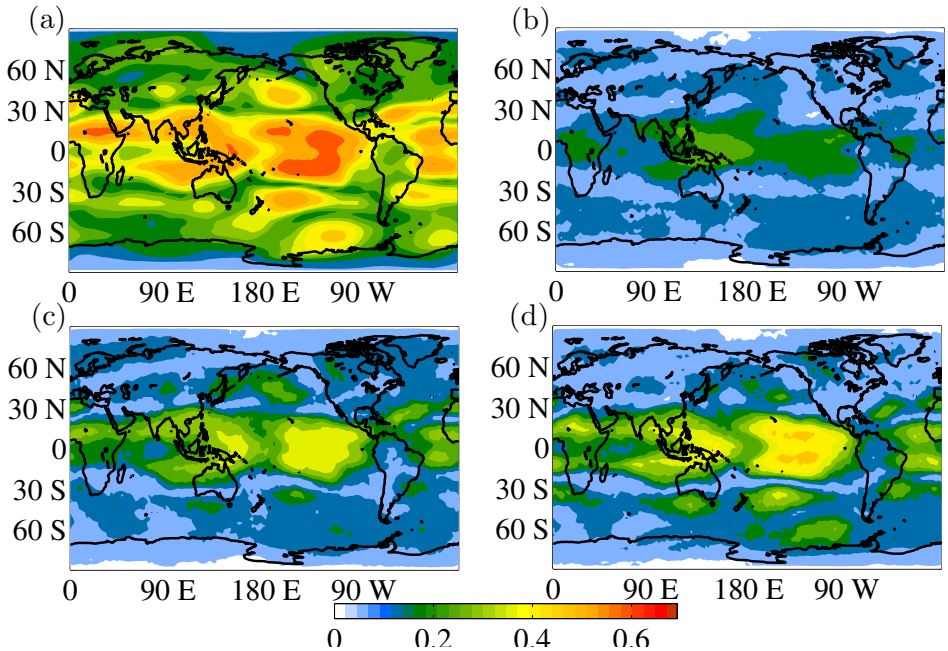

Figure 3-5: AWC maps obtained from NCEP/NCAR reanalysis data. The statistical similarity measures used: PC **(a)** and MI with the BP methodology: $\Delta t = 1$ **(b)**, $\Delta t = 4$ **(c)** and $\Delta t = 12$ **(d)**.

On the other hand, the AWC of the intra-seasonal networks has a small but positive trend during the entire period considered. The PC, MIH and inter-annual networks show similar behavior in the connectivity of the CP box, Fig. 7, starting with a negative trend until 1940 and then an increase in the connectivity to the end of the period. On the other hand, the intra-annual and intra-seasonal networks maintain the negative trend for a few more years, starting the increasing period in 1956, such as the WP box in the intra-annual case. Except in the intra-seasonal case, the connectivity values of the central Pacific are larger than in the western Pacific. It is important to mention that the evolution of the connectivity in the MIH networks presents two periods, 1920–1926 and 1956–1967, of anomalous higher values in both regions analyzed. Further analysis indicated that the new connections in both periods are located in the extratropics but have no clear spatial structure, and thus they might be related to noisy behavior.

Figure 8 shows the statistically significant mean Pearson correlation values of the CP and WP boxes for 1915, 1940 and 1995. In agreement with the time evolution shown in Figs. 6 and 7, it can be seen clearly that in the 30 years centered in 1940 the connectivity

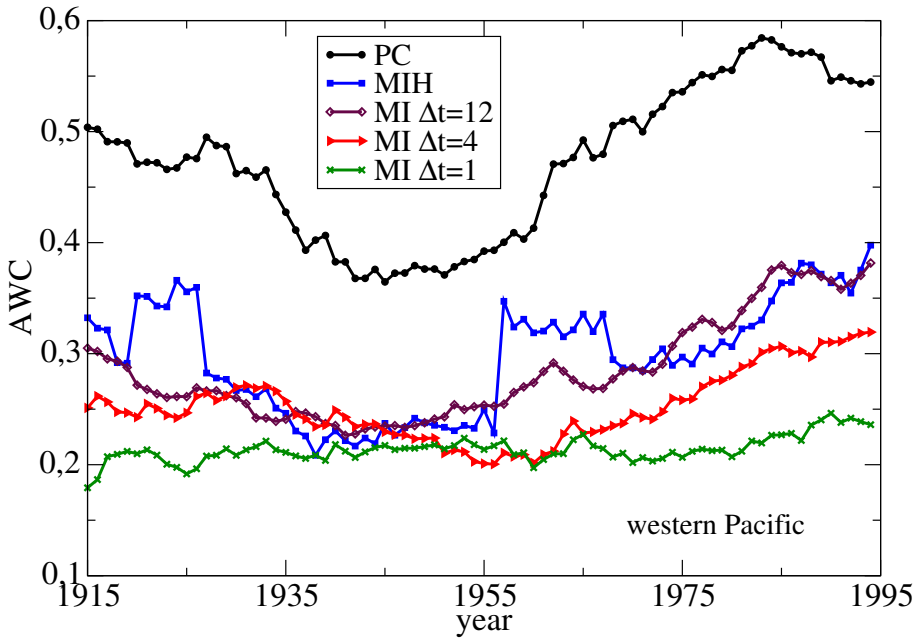

Figure 3-6: Temporal evolution of the AWC in the WP box. The networks are built by evaluating 30-year windows of the 200 mb eddy geopotential height field (NOAA) with every methodology used in this work. For each mean AWC value the time periods corresponding to 15 years before and 15 years after were considered.

of these regions is smallest. For the WP the increase in global connectivity seen in the last decades of the century is evident, particularly due to stronger connections with the tropical Atlantic. These maps also show that the CP and WP are not independent of each other, which is expected because phenomena like El Niño connect both regions through changes in the Walker circulation.

3.3.4 Intrinsic and forced variability

Nine ICTP-AGCM model runs with the same sea surface temperature boundary condition and slightly different atmospheric initial conditions were performed. As is explained in the Methodology Section, this allows for the separation of the AWC function of the forced and internal variability components. In this case, to continue comparing the behavior in the connectivity of the networks with time, and as the time interval covered by the model output is from 1901 until 2006, the resulting halves are [January 1901, December 1953] and [January 1954, December 2006], with $T = 636$. Even though the periods do not match

exactly with those of the reanalysis, they only differ in 2 out of 53 years, and thus the statistics should be fully comparable.

The forced component AWC maps obtained by the BP mutual information methodology are shown in Fig. 9. All timescales present similar spatial structures for both time periods considered, with maxima in the tropics and wavy patterns in the extratropics. The main difference is the intensity of the connectivity, such that during the second period the number of connections increases on all the timescales. Also, on intra-annual timescales Northern Hemisphere regions are more connected than on intra-seasonal timescales and are larger than in the Southern Hemisphere. On inter-annual timescales the connectivity is similar in the northern and southern extratropics, and the model recovers the highly connected region in the South Pacific located at about 35° S, 150° W.

Comparison with reanalysis reveals very similar structures, suggesting that most of the connectivity seen in Fig. 3 is due to oceanically forced variability. The largest differences occur on intra-seasonal timescales in the Indo-Pacific region (particularly in the second half), where the model shows weaker connectivity. Also regarding the reanalysis comparison, it can be noticed that the spatial structure is quite similar except for differences in strength, but there are fewer changes between the first and second halves of the 20th century. This might suggest that the reanalysis differences between both time intervals considered are mostly because of the internal variability.

Figure 10 shows the longitudinal sum of AWC versus latitude for different timescales. Compared with the reanalysis, the curves obtained are less smooth, but the general shapes are similar. The values of AWC are, as expected, substantially smaller than in the reanalysis, given that internal variability has been filtered. The differences in magnitudes are smaller on longer timescales, indicating the dominant role of the oceans in the inter-annual global connectivity.

As in the reanalysis data, the second period shows enhanced connectivity in the tropics for long timescales. There are asymmetries between the two maxima near the Equator that are, however, not present in the reanalysis data networks. Moreover, in the Southern Hemisphere extratropics, the intra-seasonal networks show largest connectivity in the second period, contrary to the results obtained in the reanalysis. This suggests that atmospheric

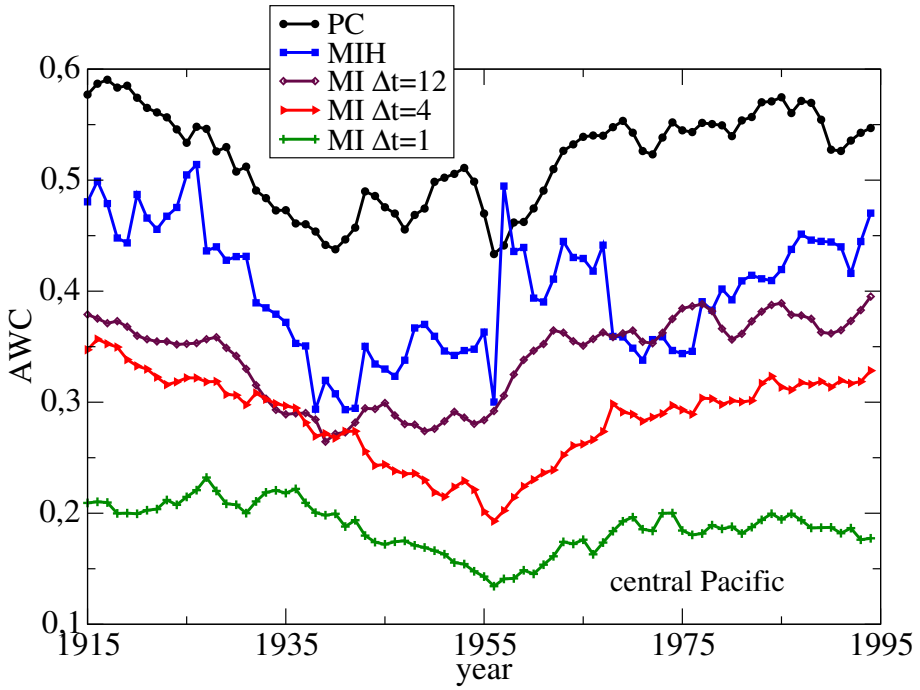

Figure 3-7: Temporal evolution of the AWC in the CP box. The networks are built by evaluating 30-year windows of the 200 mb eddy geopotential height field (NOAA) with every methodology used in this work. For each mean AWC value the time periods corresponding to 15 years before and 15 years after were considered.

internal variability plays an important role in the connectivity of the southern extratropics on intra-seasonal timescales.

The AWC maps of internal atmospheric variability are shown in Fig. 11. Both Pearson correlation and mutual information with $\Delta t = 1$ and 4 have mostly the same spatial structure but differ where global maxima are located. AWC maps are also strongly similar between periods.

On intra-seasonal timescales the regions of maximum connectivity are in the western Pacific and in the southern extratropics. There are clear connections in the Southern Hemisphere between the South Pacific, South America and the South Atlantic. There is also a small difference between the two periods: the connectivity in the western Pacific region increases, and it decreases in the central Pacific, similar to the NOAA 20th Century Reanalysis data. This result, together with the AWC maps of the forced atmospheric variability, suggests that the changes in connectivity observed in the reanalysis during the 20th century

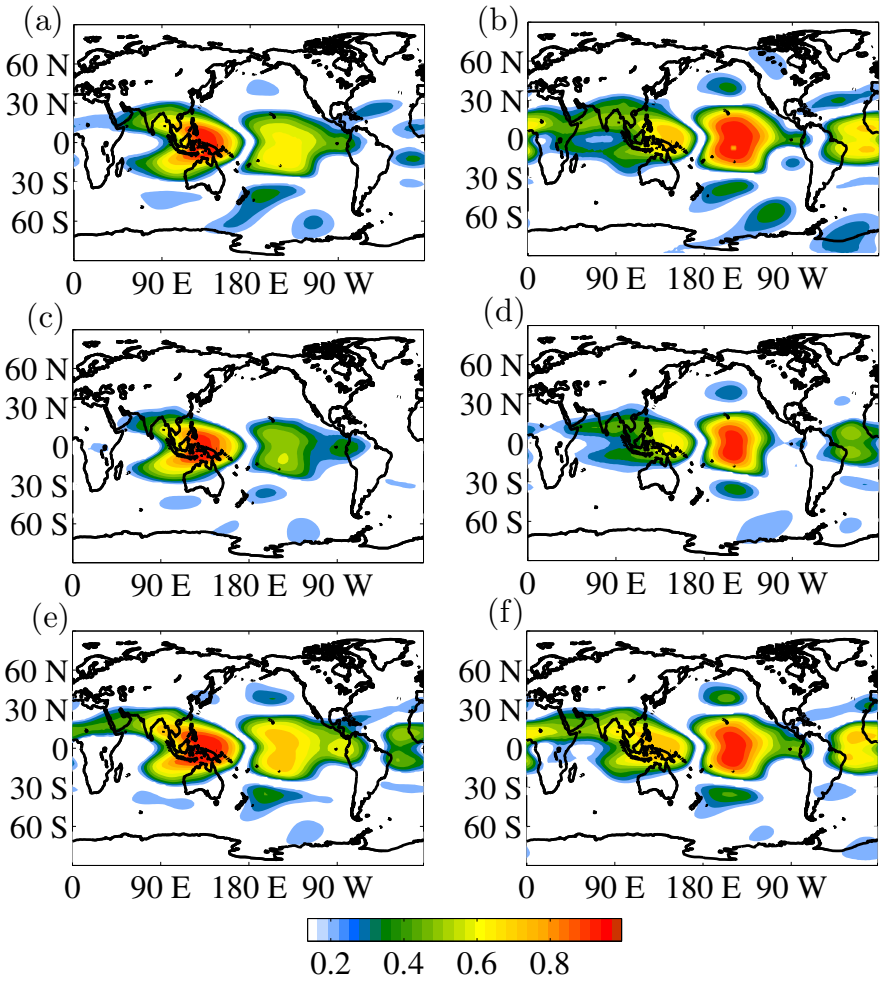

Figure 3-8: Maps of the cross-correlation using the NOAA data. The color represents the mean value of the PC of each location, with the WP and CP boxes (defined in the text) taking 30-year windows centered at 1915, 1940 and 1995. **(a)** and **(b)** correspond to 1915, WP and CP, respectively, **(c)** and **(d)** to 1940, WP and CP, respectively, and **(e)** and **(f)** to 1995, WP and CP, respectively.

are due to changes in both forced and intrinsic components.

On intra-annual timescales, on the other hand, the AWC maps present the same well-connected spots as the intra-seasonal case, but also additional regions mainly in the North Pacific and the North Atlantic. The wavy patterns suggest the existence of wave train propagation all over the globe. The Pearson correlation network has different link density, but the structure is quite similar to the intra-annual case.

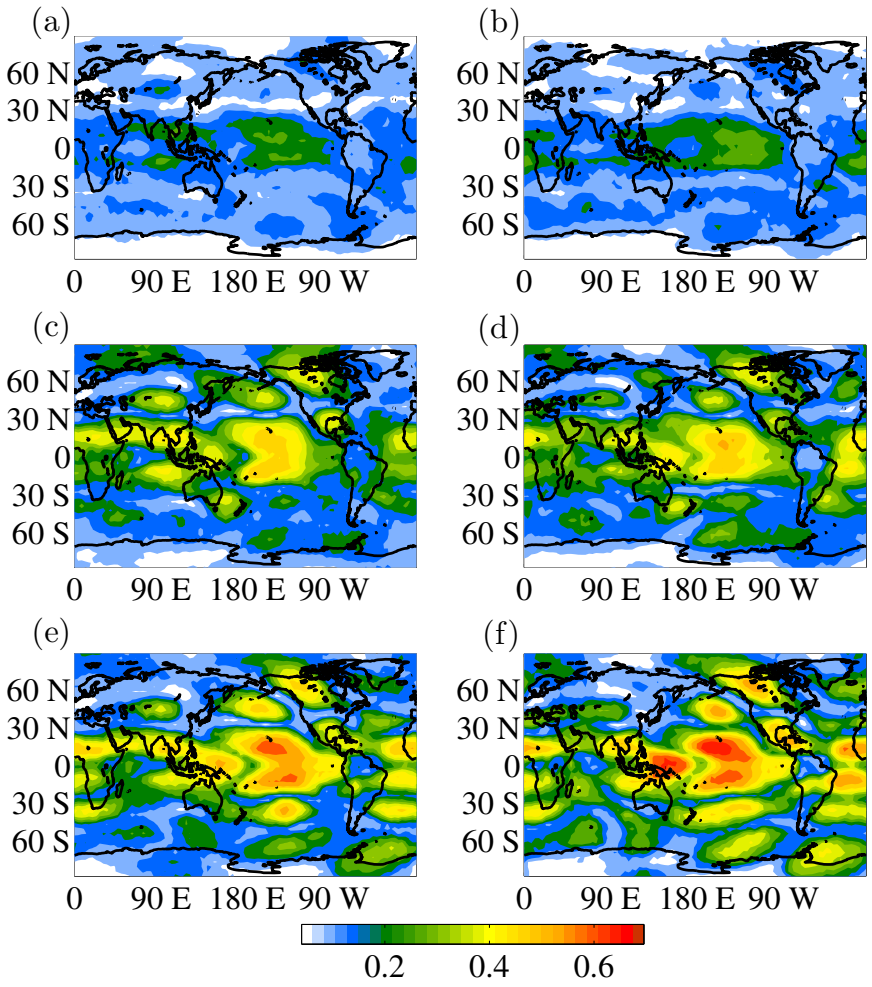

Figure 3-9: AWC maps of the ICTP-AGCM oceanically forced component. The networks are built by evaluating the mutual information with the BP methodology: $\Delta t = 1$: (1901–1953) **(a)** and (1954–2006) **(b)**, $\Delta t = 4$: (1901–1953) **(c)** and (1954–2006) **(d)**, and $\Delta t = 12$: (1901–1953) **(e)** and (1954–2006) **(f)**.

3.3.5 Principal component analysis

Considering the influence, or connectivity, of the tropical Pacific, a principal component analysis was done using the NOAA 20th Century Reanalysis data [29] in this region. In particular, the time series within the following coordinates were considered: [90° E, 110° W–30° S, 30° N]. As in the sections above, we worked with anomalies of the Z200e and considered the data in halves: [January 1901–December 1955] and [January 1955–December 2009].

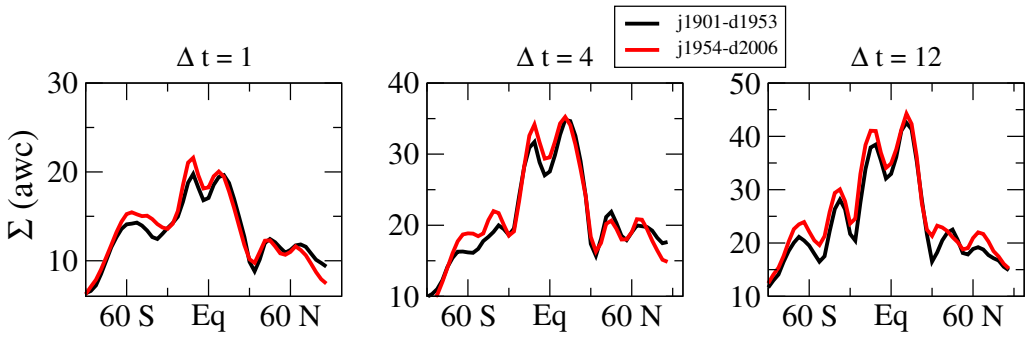

Figure 3-10: Zonal sum of the AWC function for the ICTP-AGCM networks built by using the BP methodology with $\Delta t = 1$, $\Delta t = 4$ and $\Delta t = 12$, for both time periods (black: 1901–1953, red: 1954–2006). It is a measure of how many connections each latitude has.

Figure 12 shows the first two eigenvectors of the 1955–2009 period, rotated following the simplicity (varimax) criterion. The percentages of explained variance are 14.9 and 9.1, respectively. It is worth mentioning that the leading eigenvector has the same structure as the connectivity degree [70]. The first EOF is consistently strongly related to the ENSO pattern.

The maps in Fig. 13 show the regressions of global eddy geopotential height anomalies onto the principal components of the first two EOFs. Clearly, the regression with the leading principal component has a very similar structure to the map of the AWC of the ICTP-AGCM forced component, indicating that the ocean forcing is to a large extent from the tropical Pacific. The second eigenvector is also related in both periods to the Pacific–North American (PNA) pattern. Note that for the leading EOF in the second period, the extratropical wave train that propagates from the central Pacific is much stronger than in the first period.

Taking into account that the first two patterns are related to well-known modes of variability such as ENSO and PNA, we build the AWC maps for the original data without these principal components (Fig. 14). The AWC function of the Pearson correlation network without the ENSO-related component (Fig. 14a) shows an important decrease mainly in the tropical regions (compare it with Fig. 2b). The overall appearance of the resulting structure is similar to the one obtained for the atmospheric internal variability of the ICTP-AGCM, showing again the important influence of ENSO on the atmospheric dynamics. On the other hand, Fig. 14b shows the AWC map of the mutual information, with BP

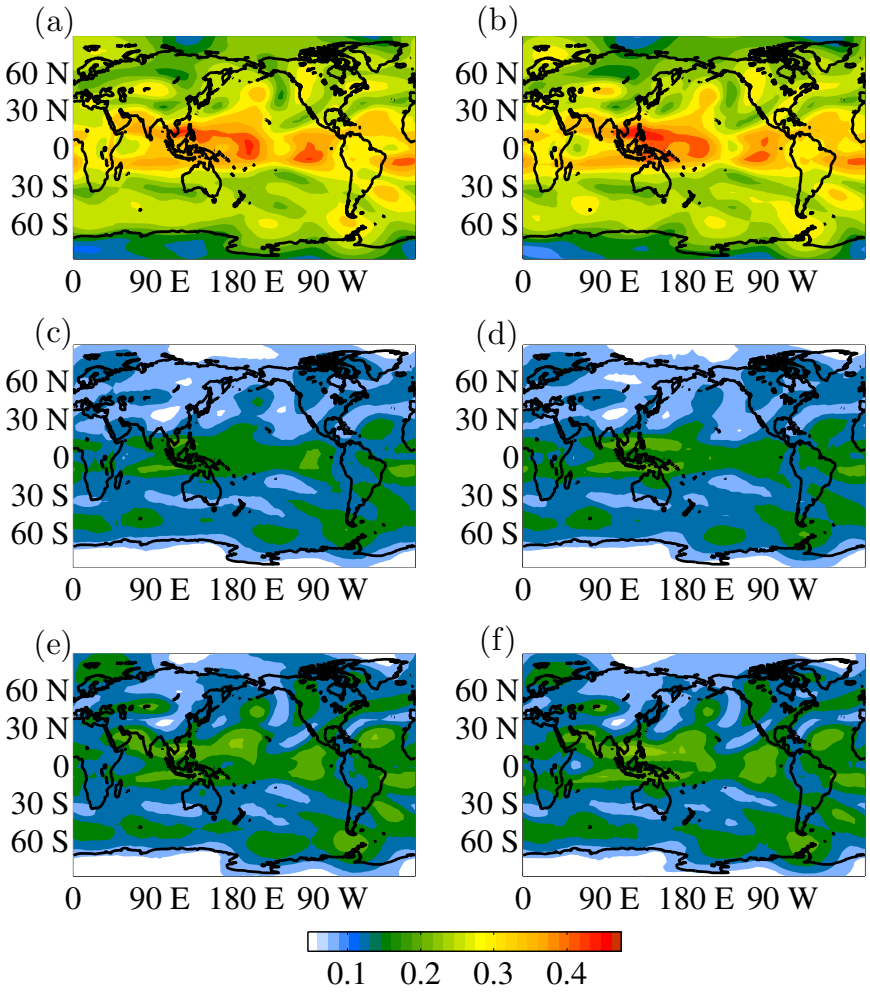

Figure 3-11: AWC maps of the ICTP-AGCM internal variability. The networks are built with PC (1901–1953) **(a)** and (1954–2006) **(b)**, BP methodology with $\Delta t = 1$ (1901–1953) **(c)** and (1954–2006) **(d)**, and BP with $\Delta t = 4$ (1901–1953) **(e)** and (1954–2006) **(f)**.

symbolic analysis on the intra-seasonal scale ($\Delta t = 1$) without the ENSO-related component revealing that the connectivity in the tropical Pacific decreases, but to a far lesser extent than in Fig. 14a. The general structure in the tropical connectivity is maintained, and some subpolar regions become more connected. This suggests that this network connectivity is less ENSO dependent, and thus less influenced by the oceanically forced component than the Pearson correlation network. In the bottom panels the AWC maps of the mutual information on the intra-annual scale network (BP with $\Delta t = 4$) without the ENSO-related component

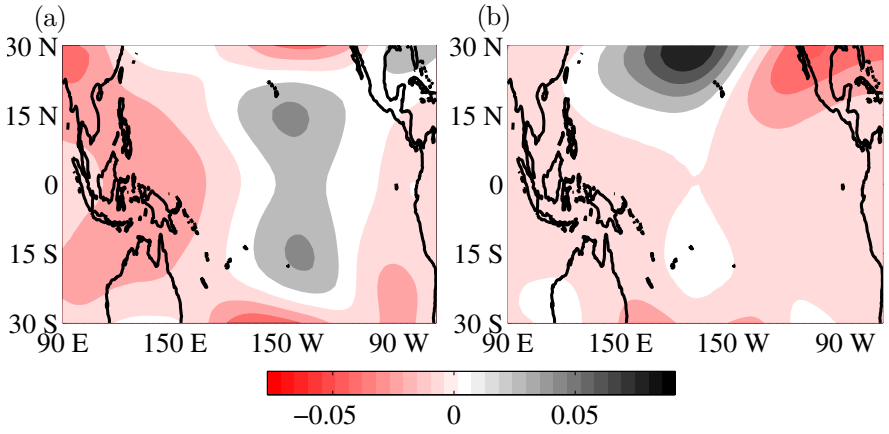

Figure 3-12: The first two eigenvectors of the covariance matrix (EOFs) of the NOAA reanalysis data considering the time period (1955–2009) rotated by the simplicity criterion. Variance explained: 14.9 % (ENSO pattern related) **(a)**, 9.1 % (PNA pattern related) **(b)**.

(Fig. 14c) and without the PNA-related component (Fig. 14d) are shown. Comparison reveals that in the northern Pacific, the region where both EOFs have weight, the maximum connectivity region seen in Fig. 3d is composed of a center due to the ENSO component located at 40° N, 160° W and a broadly connected region to the northwest related to the PNA.

3.4 Conclusion

We have studied the variability of the Z200e anomaly field by means of networks built using linear and nonlinear statistical similarity measures. The analysis considered two reanalysis data sets (NOAA 20th Century and NCEP CDAS1) and the output of an atmospheric general circulation model (ICTP-AGCM). This combination allowed us to study the evolution of the upper atmosphere connectivity over the 20th century as well as to disentangle the networks due to the forced and internal components.

Overall, the resulting connectivity structures built using Pearson correlation (linear) or the mutual information (nonlinear) have similar spatial features dominated by high connectivity in the tropics and smaller connectivity centers shaped as wave trains in the extratropics.

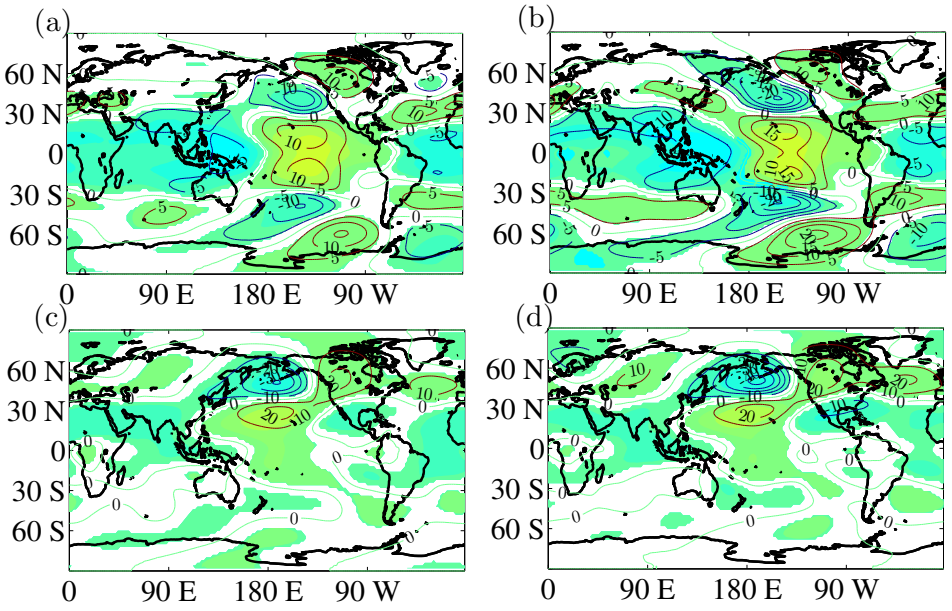

Figure 3-13: Regressions of the two principal components obtained with the global eddy geopotential height field where the correlation is statistically significant (these areas are colored for visualization only). **(a)** and **(c)** from 1901 to 1955; **(b)** and **(d)** from 1955 to 2009.

The use of mutual information calculated with the BP symbolic analysis allows for the separation of the connectivity, depending on the timescale. On intra-seasonal scales ($\Delta t = 1$), the connectivity of the central and western Pacific regions shows significant changes between both halves of the 20th century: while in the first half the tropical Pacific is fully connected, as with the Pearson correlation or mutual information with histograms, in the second period the eastern area appears less connected and the global maximum is in the western Pacific.

On the intra-annual timescale ($\Delta t = 4$), the main feature is the maximum in connection with the northern Pacific, which is larger than on the other two timescales considered. Finally, for the inter-annual timescale ($\Delta t = 12$), the southern subtropical Pacific arises as a connectivity hub, also seen in the AWC constructed with the Pearson correlation and mutual information with histograms, but which is not so clear in the other timescale networks. In other regions such as Canada, the connectivity decreases (increases) in the second period on intra-annual (inter-annual) timescales. These examples indicate that the BP methodology is useful in separating different timescale processes.

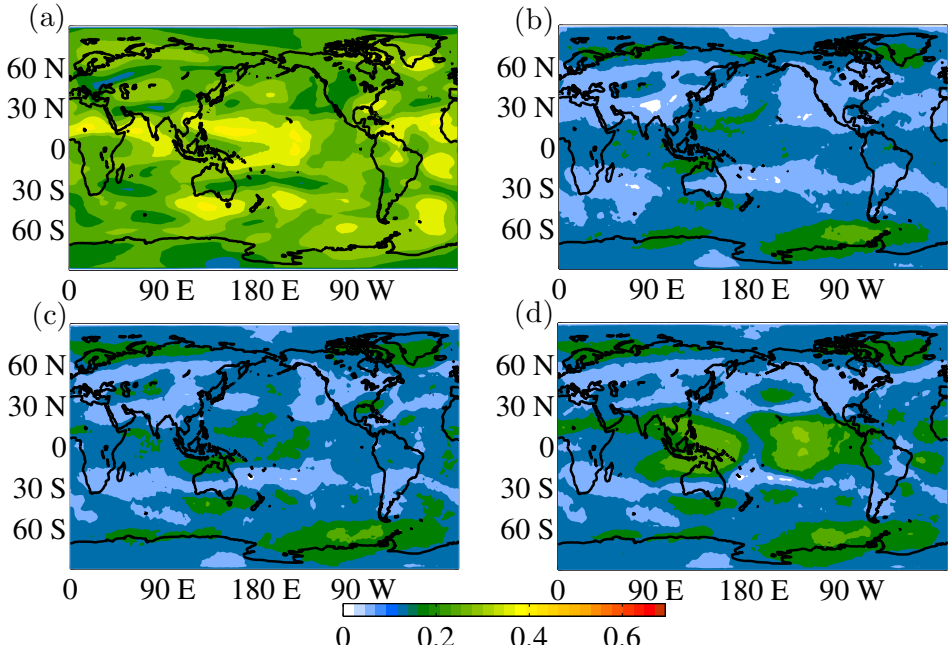

Figure 3-14: AWC maps of the NOAA reanalysis using data from the second period (1955–2009). The networks are built after removing one of the two principal components related to ENSO and PNA patterns. PC without ENSO-related component **(a)**, MI with intra-seasonal timescale without ENSO-related component **(b)**, and MI with intra-annual timescale without the ENSO and PNA components (**(c)** and **(d)**, respectively).

Furthermore, in order to deepen our understanding of the changes in the connectivity during the 20th century, we have studied the AWC time evolution by considering a 30-year sliding window of two main boxes located in the western Pacific (WP) and the central Pacific (CP). The temporal evolution in the connectivity of these regions indicates a decreasing behavior at the beginning of the century until 1940, when it starts increasing until the end of the century for the PC, MIH and inter-annual networks. Intra-annual networks, on the other hand, present a negative trend during the first decades, and a positive trend from 1956 onwards.

The connectivity thus evolves similarly in WP and CP for the PC, MIH, inter-annual and intra-annual networks. On the other hand, the connectivity of intra-seasonal networks is very similar to the intra-annual networks in CP, but shows a positive trend in WP during the entire 20th century. Even though is not trivial to interpret these results, we might assume that when the connectivity of these main regions decreases (increases) in a particular

timescale network, the teleconnections associated with these regions and with this particular timescale are weakened (strengthened) in the temporal period considered. In this case, both the mid-century minima of all the networks' connectivity in the central Pacific, as well as the continuously increasing connectivity in the intra-seasonal timescale networks in the western Pacific, deserve further investigation.

Finally, using the ICTP-AGCM we were able to separate the oceanically forced component from the atmospheric internal variability and calculate the respective connectivity networks. The results suggest that the main patterns of connectivity captured by the reanalysis networks are due to the oceanically forced component, particularly on inter-annual timescales. On intra-seasonal timescales the atmospheric internal variability seems to play an important role in determining the network structure.

This study showed a great richness in upper atmospheric connectivity during the 20th century in time and space, revealed through the use of linear and nonlinear similarity measures and a technique to separate timescales. We focused on the tropical Pacific due to its global relevance. Subsequent studies should focus on extratropical regions, where this technique can prove useful for improving climate predictions.

Chapter 4

ENSO teleconnections in the Southern Hemisphere[1]

Using functional network analysis we study the seasonality of atmospheric connectivity as well as its interannual variability depending on the different phases of El Niño-Southern Oscillation (ENSO) phenomenon. We find a strong variability of the connectivity on seasonal and interannual time scales both in the tropical and extratropical regions. In particular, there are significant changes in the southern hemisphere extratropical atmospheric connectivity during austral spring within the different stages of ENSO: We find that the connectivity patterns due to stationary Rossby waves differ during El Niño and La Niña, showing a very clear wave train originating close to Australia in the former case, as opposed to La Niña that seems to generate a wave train from the central Pacific. An attempt to understand these differences in terms of changes in the frequency of intraseasonal weather regimes cannot fully explain the differences in connectivity, even though we found prevalence of different intraseasonal regimes in each phase of ENSO. We conclude that the differential response to extreme phases of ENSO during austral springtime is related to the forcing of waves of different tropical origins.

4.1 Introduction

El Niño-Southern Oscillation (ENSO) is a climate variability mode that develops in the tropical Pacific with a period that varies between 2 and 7 years. Its positive (negative)

[1]The results of this chapter are published in F. Arizmendi and M. Barreiro, CHAOS, 27, 2017 [71].

phase, called El Niño (La Niña), involves anomalous warm (cold) anomalies in the sea surface temperature (SST) of the eastern equatorial Pacific. These SST anomalies shift the position of the tropical convection producing changes in the location and intensity of the diabatic heating which, in turn, is the main driving of tropical and extratropical circulation anomalies. These teleconnections remotely forced by the tropical oceans are the basis of seasonal climate prediction [72]. In particular, the extratropical response takes the form of waves describing large-circle paths, which has been understood in terms of the linear dispersion of Rossby waves [73, 74, 27] or as the result of the interaction of tropically-forced waves with midlatitude dynamics [75].

In the late 1990's there was a prevailing idea that the global atmospheric response to the extreme phases of ENSO was linear in the sense that the patterns induced all around the globe were equal but opposite for El Niño (EN) and La Niña (LN). However, since then, many studies have stressed the nonlinearity of the response of the atmosphere to tropical SST forcing, particularly during Northern Hemisphere (NH) winter [76, 77, 78]. These works focus mainly on the NH extratropics and specifically on the Pacific North American (PNA) oscillation pattern, which is a quadripolar pattern of pressure height anomalies in the central Pacific Ocean and centers of action over western Canada and the southeastern United States. The PNA is one of the most prominent modes of low frequency variability in the NH and strongly affects climates all over North America. It has been proposed that the nonlinear response is due to the fact that the maximum tropical rainfall anomalies are located east of the dateline during El Niño events, but west of the dateline during La Niña events [76]. Other authors have proposed that the transient eddy feedback plays a role in generating the nonlinearity. They suggest that the tropically forced low-frequency flow pattern in the midlatitudes modulates the storm track activity through changes in the jet stream, which changes then feed back onto the direct tropical response, leading to different extratropical circulation anomalies [79].

For the Southern Hemisphere (SH) it has been suggested that ENSO heating anomalies are able to trigger changes in the occurrence of midlatitude intraseasonal regimes during late spring through Rossby wave propagation, leading to the observed ENSO teleconnection [47]. However, little has been assessed about a possible differential response to tropical forcing

in the Southern Hemisphere (SH) extratropics. Moreover, there is no systematic study that addresses the global atmospheric connectivity in the different seasons of the year and how it varies interannually depending on the phase of ENSO.

Complex network theory has been used in many fields in the last two decades. Since the pioneering works on complex networks [43, 52], the use of this statistical methodology of analysis has allowed significant advances in a broad variety of scientific disciplines, including social systems, neural networks, and gene expressions, among many others [42]. If a system can be thought as a set of interacting agents, then it can be represented by a network where each node represents an agent and the links between them represent their interactions. The climate system is one example and recently there have been many insights on climate dynamics through the application of this theoretical framework [45, 54, 56, 80, 59, 81, 55, 57, 82, 51, 58, 60, 83, 84].

Usually, climate networks are built by evaluating the statistical similarity (links) among time series (nodes) that represent the temporal evolution of a certain variable of interest in specific geographical locations. Surface air temperature (SAT) as well as the geopotential height (Z) are some of the climate variables widely used. There are many ways of considering the interdependencies among nodes. In some cases, based on finding the maximum (or minimum) correlation value, temporal shifts are considered between the phases of the nodes [**?**, 80]. However, it was shown that the time lag parameter is quite sensitive and can strongly influence the topology of the network [59], which may be related to the non-stationarity of the climate network [81]. Also, there have been studies that compare different statistical measures of similarity such as the linear Pearson correlation and the nonlinear Mutual Information [55], which can also take into account different time scales by means of symbolic analysis [57, 82, 51].

In particular, being a major component of the climate system, ENSO has been the focus of many climate network studies [56, 80, 59, 60, 83, 84]. Using an evolving climate network analysis, it has been shown that it is possible to distinguish between different types of EN through the topology of the SAT anomalies network [**?**]. There have also been studies evaluating the global impact of ENSO on other climate variables such as the moisture [**?**, 84]. In this work, we use climate network analysis to improve our understanding of

the impact of different ENSO phases on the 200 hPa geopotential height climate network by means of the area weighted connectivity measure. This is done for all four seasons in an attempt to perform a systematic study of the ENSO teleconnections from a complex network point of view and determine a possible differential response to the different phases. Then, using also other commonly used climatological techniques, we focus our analysis on the SH atmospheric circulation in austral spring (September, October, November - SON). In the course of this chapter we show how the tools of complex networks are complementary to standard statistical techniques and allow to uncover new insights into the atmospheric response to ENSO.

The chapter is organized as follows. In the next section we introduce the data sets and methodology used, including the definition of the connectivity measure and the composite analysis. Having in mind that the extratropical connectivity during ENSO years can be considered as the result of the dispersion of tropically forced Rossby waves or as due to the interaction of these stationary waves with midlatitude dynamics we introduce the methodology to compute intraseasonal weather regimes in the SH extratropics. In the Results Section, we show that the atmospheric connectivity shows large seasonal and interannual variability. Moreover, we find that during austral spring the extratropical response to ENSO is highly dependent on the phase considered, and can be interpreted as the result of forcing of waves of different tropical origins. Finally, in the Conclusions Section, we present a discussion and summarize the results.

4.2 Data and methodology

4.2.1 Data sets

In order to study the global atmospheric connectivity on interannual time scales and the influence of ENSO on these emerging patterns, we use the 200 hPa geopotential height monthly mean (Z200) from the NCEP CDAS1 reanalysis data base [12]. It is worth mentioning that reanalysis data is the blending of observations with an Atmospheric General Circulation Model (AGCM) and is the best estimation of the historical atmospheric evolution. The data is distributed in a regular latitude-longitude grid with resolution of 2.5°,

leading to a total number of grid points N=10224, and covers the time-period [1949-2014], that is T=792 months. The election of 200 hPa is due to the fact that this level is where the maximum tropical divergence takes place, which is crucial for the excitation of extratropical Rossby wave trains [27]. As the geopotential height has a very strong meridional structure with high values in the tropics and decreasing toward the poles, we consider the eddy component removing the corresponding zonal mean at each node for every time step. In this way the signatures of the planetary waves, which are deviations of this zonal structure, are highlighted. In addition, as it is most commonly used in climate dynamic studies, we use the anomalies calculated by substracting the mean annual cycle from each time series.

Since we found important differences in the interannual connectivity patterns in austral spring between EN and LN scenarios, we also calculate intraseasonal weather regimes in the SH to try to understand whether these differences could be explained in terms of the interactions of tropically forced waves with extratropical dynamics. To do that, we consider the daily mean of the same field (eddy geopotential height anomalies at 200 hPa) and reanalysis data base (NCEP CDAS1). In addition, to deepen our knowledge on the role of the tropical forcing in the SH austral spring, we also use the daily mean of 200 hPa zonal and meridional winds from the same reanalysis database (NCEP CDAS1) to calculate the Rossby wave sources.

4.2.2 Climate network analysis

Let us consider the eddy Z200 anomaly field as $\vec{x}(t) = \{x_i(t)\}_{i=1}^{N}$ with $t = 1, \ldots, T$ being the temporal steps, monthly mean values in this case. Given this grid of time series, we establish an unweighted and nondirectional link among two nodes i and j if the Pearson correlation between the correspondent time series ρ_{ij} is above a threshold τ. Then, the elements of the adjancency matrix $\mathbf{A}$ result

$$A_{ij} = \Theta\left(|\rho_{ij}| - \tau\right) - \delta_{ij}, \tag{4.1}$$

where $\Theta()$ is the Heaviside step function and δ is the Kronecker delta. To keep only the strongest connections, we consider τ as the 99% quantile value of the distribution of the

correlation values between shuffle surrogates of every pair of nodes. This assures that the original distributions remain unchanged, and the null hypothesis of independence is rejected if the correlation value of the original time series is above this threshold at a 99% confidence level. The choice of the threshold is based on the study of the previous chapter, where we find that the 99% quantile is adequate for representing atmospheric teleconnections [51]. Moreover, it is shown that the eddy Z200 anomaly field has mostly auto-correlations that become significant already after one temporal step and thus the shuffle surrogates provides a good test of the connectivity.

In order to quantify the influence of every node of the network, we calculate the *area-weighted connectivity* (AWC) that can be interpreted as the fraction of Earth that each node is connected to [54, 55]. Specifically, this measure is defined as

$$\mathrm{AWC}_i = \frac{\sum_{j=1}^{N} A_{ij} \cos(\theta_j)}{\sum_{j=1}^{N} \cos(\theta_j)}, \tag{4.2}$$

where θ_j is the latitude of node j and the cosine factors are included to correct the differences in the area covered by the nodes given their geographical distribution over the regular grid. It can be easily proved that for angularly equidistant grids, as we use, the corresponding area of the Earth's surface of a node i is proportional to the cosine of the latitude θ_i.

Following the Oceanic Niño Index classification (ONI) maintained by the Climate Prediction Center of NOAA (`http://www.cpc.ncep.noaa.gov/products/analysis_monitoring/ensostuff/ensoyears.shtml`), we classify each year of our analysis depending on the correspondent ENSO phase. We consider the ENSO year y from June of year y to May of year $y+1$, since normally ENSO events start developing in the boreal summer, with a peak in December-January and decaying in austral fall [85]. Then, considering all the years of same ENSO phase (*i.e.* EN, LN or N years), we build eddy Z200 anomalies networks taking into account the different seasons as June-July-August (JJA), September-October-November (SON), December-January-February (DJF) and March-April-May (MAM). We further calculate the AWC for each ENSO phase and the differences between EN and LN years for every season. Since we analyze the period [1949,2014] we have 21 EN years and 22

years for both LN and N. Thus, the length of the time series T to construct the networks for each season and their corresponding threshold is 63 for EN and 66 for LN and N years.

Recently, it has been shown that the AWC can be similar to the leading Empirical Orthogonal Function (EOF1) of the field [70]. However, the authors also showed that this is the case only if there is one leading EOF that explains most of the variance. The eddy Z200 field used in this study does not fulfill this condition. As an example, we calculated the EOFs during SON and found that the leading mode explains only a minor fraction of the total variance (14% for EN, 18% for LN and 8% for N years). Thus the climate network analysis performed here, even though only considers the AWC, has the potential to add new information compared to the standard EOF technique. As shown below, the use of the AWC allowed us to uncover new characteristics of ENSO teleconnections.

4.2.3 Composite analysis

In addition to the climate network analysis, we also perform a composite study using the same monthly mean eddy Z200 anomaly field. In this case, following the classification of ONI, we distinguish years according to the three phases of ENSO (EN, LN and N years), and calculate the differences in the means (d) of the eddy Z200 anomalies for EN minus N and LN minus N. This allows comparing the mean changes of EN and LN years with respect to N years. The significance at each point i is assessed by evaluating the statistic t of the difference of means. For example, in the case of the composite EN minus N the test is calculated as

$$t* = \frac{\bar{x_i}^{EN,LN} - \bar{x_i}^{N}}{\sqrt{n_1^{-1} + n_2^{-1}}} \sqrt{\frac{(n_1 + n_2 - 2)}{(n_1 - 1)\sigma_i^{EN,LN2} + (n_2 - 1)\sigma_i^{N2}}}. \tag{4.3}$$

Here, n_1 corresponds to the size of the sample of EN years, while n_2 is that of the N years; σ_i is the standard deviation. The statistic $t*$ is compared with the Student's t distribution with a two tailed confidence interval of 95%, with $H_0 : d = 0$ being the null hypothesis.

To aid further understanding of the role of the tropical forcing in the SH atmospheric circulation, we also perform a composite analysis of the Rossby Wave Source (S) and compare differences between EN and LN scenarios. The sources S are the regions where atmospheric

baroclinic modes that dominate the response to tropical thermal anomalies are converted to barotropic anomalies that propagate into the extratropics as Rossby waves. In the extratropics the phase of the wave is independent of height and thus the perturbations are considered barotropic. S can be calculated from the barotropic vorticity equation, which depends on the divergent flow [74],

$$S = -\xi D - \vec{v}_\chi . \nabla \xi = -\nabla . (\vec{v}_\chi \xi), \tag{4.4}$$

where ξ is the absolute vorticity, $\vec{v}_\chi$ the irrotational wind component and D the divergence of the wind. The first term, $-\xi D$, represents the generation of vorticity by divergence while the second, $-\vec{v}_\chi . \nabla \xi$, corresponds to the advection of absolute vorticity by divergent flow.

To calculate S we use daily 200 mb zonal and meridional winds from NCEP/NCAR reanalysis [12]. We perform a composite study, as done for Z200, comparing EN and LN years with respect to N years using the classification of ONI and assessing the statistical significance with the statistic $t*$ and the Student's t distribution (two tails, 95%).

4.2.4 Weather regimes

As it is shown below, there are significant differences in the networks during EN and LN years, particularly in SON over the SH, suggesting a differential response of the extratropical atmospheric circulation to the tropical forcing induced by opposite phases of ENSO. This result could in principle be explained as the response of the extratropical atmosphere in terms of Rossby wave trains that emanate from different regions of the tropics and/or in terms of the interactions of tropically forced waves with extratropical dynamics. To assess the latter possibility we follow Cazes-Boezio et al (2003). The first step in the methodology is the identification of intraseasonal regimes in the SH extratropics, that is, determine preferential atmospheric circulation patterns with time scales between 10 and 90 days. Subsequently, we study how EN and LN favor the occurrence of those regimes. The characterization of intraseasonal regimes is done performing a K-means cluster analysis in a dimensionally reduced space after filtering the data as follows. In order to focus on intraseasonal timescales,

we preprocess the daily mean eddy Z200 field, considering days in SON, by applying first a low pass filter of 10 days. We further remove the seasonality by subtracting the mean seasonal cycle (averaged over all the years in the dataset), and remove interannual variability by subtracting from each year the corresponding seasonal mean. Then, EOFs are computed from this filtered data over the SH between 20°S and 70°S. Finally, considering the first 10 leading EOFs as our reduced dimensional space, we perform the cluster analysis. The Principal Components (PCs) associated with our 10-dimensional reduced subspace account for more than 70% of the total 300 subseasonal variance and each point in this 10-dimensional subspace represents the weather systems over the spatial domain on a single day.

The cluster analysis consists of separating all the points in the subspace (*i.e.* daily mean configurations of the eddy Z200 anomalies) into K clusters such that the total sum of the Euclidean distances between each point and the correspondent centroid is minimum. Since K-means method requires the number of K clusters to be defined a priori, we follow Michelangeli et al. (1995) and calculate the reproducibility index (R) for different values of K. The steps of the K-means and R index calculations for every value of K in the range $[2, 7]$ are as follows:

1. We separate our data in K clusters by applying the K-means method for the whole dataset (reference partition):
   - The clustering is repeated 50 times from different initial seeds,
   - each seed is chosen from the centroid of a random 10% subset of the considered data.
   - We obtain K cluster centroids which can be projected onto our spatial domain.
2. We start the reproducibility index R calculation as follow:
   - We generate 100 random subsamples by taking randomnly 50% of the data,
   - we calculate K-means for each subsample as in 1),
   - we project in our spatial domain the 100 K-dimensional cluster centroids.
3. We compare the cluster centroids obtained in 2) with those of the reference partition in terms of average pattern correlation (APC) [47, 48]. This means to calculate the

mean cross-correlation of all of the 100 cluster centrois from 2) with the centroids of the reference partition.

4. Then, we have K values of reproducibility index R, one for each cluster centroid.

Each of the 10 dimensional cluster centroids that results from the K-means analysis can be projected onto the spatial domain considered, in this case all the globe within [20° S, 70° S], representing a weather regime. As we show below, $K = 3$ is the best choice given that the reproducibility index R has the highest values for the respective cluster centroids. This mean that separating the weather regimes into 3 clusters is an good election for our analysis.

4.3 Results

4.3.1 Area weighted connectivity maps

Figure 4-1 shows the respective AWC maps separately for each season (JJA, SON, DJF and MAM) and ENSO phases.

In the SH winter, JJA, which usually marks the initial stage of development of El Niño and La Niña, the AWC shows the largest values in the tropical band, with maxima over the western Pacific and Maritime continent, as well as in the eastern Pacific (Figure 4-1, first row). Moreover, connectivity is largest in the SH compared to the NH, as expected, given that during austral winter the subtropical jet moves equatorward in the Pacific sector strengthening the Rossby wave source. In EN years the tropical areas of largest connectivity are similar to those of N years, but values are larger over the western Pacific. In the SH extratropics, the largest connectivity occurs during N years. During LN years the AWC is significantly larger in the central Pacific. In the SH extratropics the centers of maximum connectivity during EN and N years tend to coincide spatially, while they are displaced during LN years. The difference plot of AWC (EN-LN) stresses these differences and also shows the larger connectivity in the tropical Atlantic and Indian oceans during LN events.

Figure 4-1 shows that in SON the AWC patterns are very different for EN, LN and N years (Figure 4-1, second row). During N years the AWC is largest in the tropical band, but

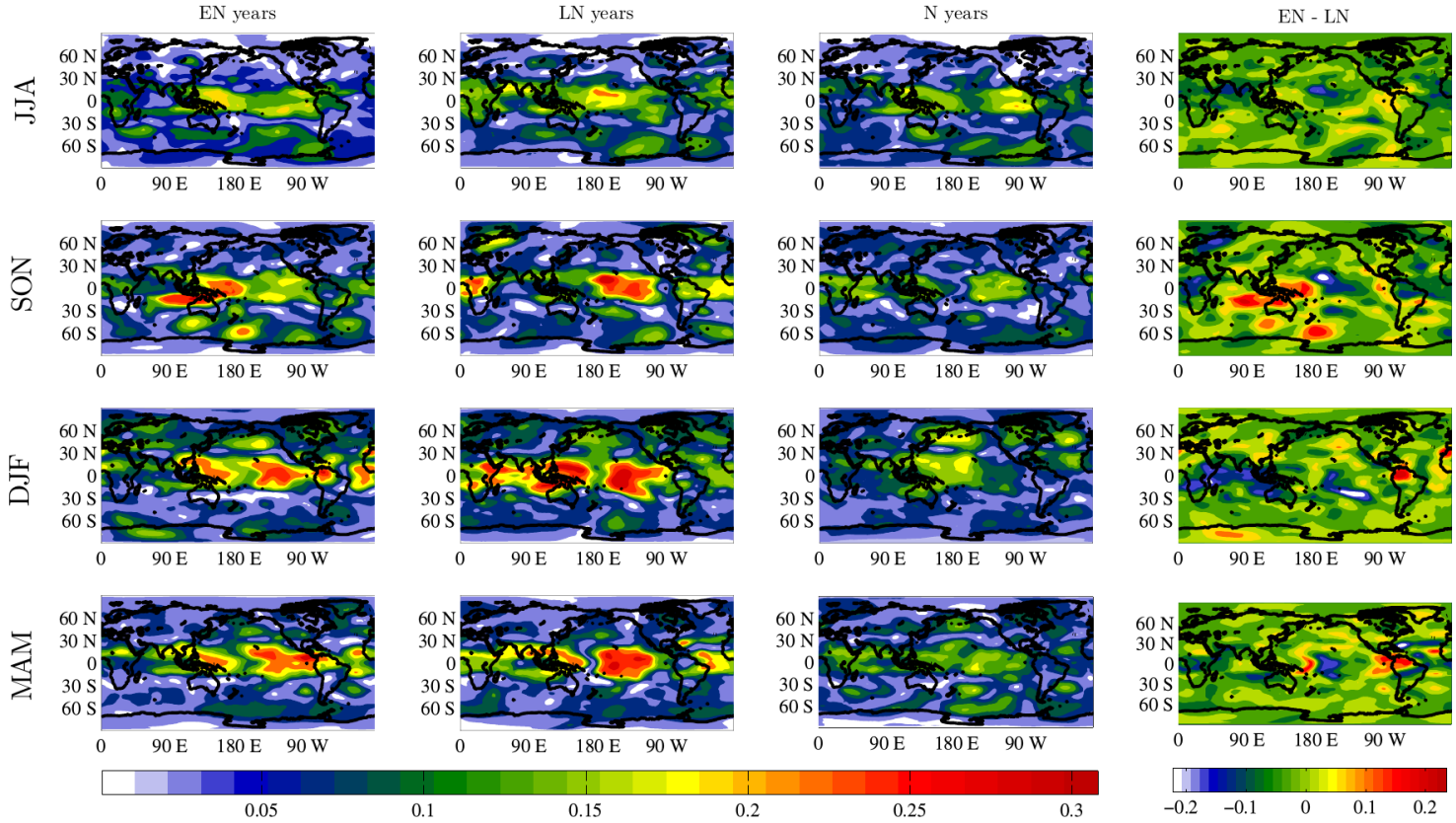

Figure 4-1: AWC maps from eddy Z200 anomalies for all seasons (rows) and different phases of ENSO (columns): El Niño years, La Niña years, Neutral years and differences in AWC obtained for El Niño and La Niña years.

shows no well-defined structure. During EN years the AWC is largest in the western tropical Pacific and over the Maritime continent extending toward the Indian ocean. Moreover, it is possible to detect a clear wave train that emanates from the southeastern Indian ocean and arches toward the South Pacific and then northeastward into the South Atlantic. During LN the maximum AWC is located in the central equatorial Pacific and the map suggests a wave train emanating from there and propagating into the South Pacific and then into the southwestern Atlantic. The difference map EN-LN clearly shows the shift in the maximum of tropical AWC, as well as the fact that connectivity is largest in the tropical Atlantic during LN. These two characteristics are also seen, but to a lesser extent in JJA. Moreover, the difference also shows the existence of maximum AWC in the shape of a wave train in the extratropical SH during EN years. Thus, different from JJA, during SON the SH extratropics is most connected during EN, and both phases of ENSO exhibit more connections than N years.

As regard to DJF, EN and LN years have similar patterns of connectivity in the tropical Pacific with two areas of large values to the east and west of the dateline, with LN ones being slightly stronger (Figure 4-1, third row). LN years also show larger values of AWC in the Indian sector. Connections in the NH extratropics become larger than in the SH, particularly over the North Pacific and western Atlantic oceans, probably as consequence of the excitation of the Pacific North American (PNA) pattern by the tropical forcing. Some continental areas become also much more connected during EN years, including the Mediterranean region and northern South America. As result, tropical South America shows a clear dipole of connectivity in the EN-LN map, such that during EN northern (central) South America is more (less) connected than during LN years. In the extratropical SH connectivity is low and there is no clear presence of wave trains as during SON, which we hypothesize is due to the climatological southern shift of the jet stream during summertime. It is worth noting that during N years the tropical connectivity is much lower than during EN and LN years, but is of similar magnitude in the NH extratropics.

During MAM (Figure 4-1, last row) the tropical connectivity is similar to that seen during DJF, although with smaller values. In particular, the western Pacific becomes much more disconnected, while the central-east Pacific tends to maintain its connectivity. The tropical

Atlantic is well connected during EN and LN years, with centers of maximum values shifted longitudinally. The main difference between LN and EN years is the strong connectivity seen in northern South America during EN, which is absent during LN, as is found in DJF. No significant connectivity is seen in the SH extratropics for either EN or LN years, with the only exception of two maxima straddling South America at about 35ÂřS. In N years the overall connectivity is small and there are no clear spatial structures in the AWC.

The eddy component of the SH atmospheric circulation and the extratropical response to ENSO are often analyzed in terms of the Pacific-South American (PSA) modes. The PSA patterns are common modes of atmospheric variability in the SH on several time scales, and exhibit wavenumber 3 structures in the midlatitudes with the largest weight in the Pacific-South American sector [86]. The PSA 1 mode has been associated with enhanced convection in the western Pacific and suppressed convection over the Indian Ocean. It is also thought to be the atmospheric response to ENSO in the SH extratropics during all seasons independently of the phase of ENSO [86]. On the other hand, the PSA 2 mode is linked to tropical heating anomalies in the central Pacific and suppressed convection in the western Pacific, and is excited by the tropical Pacific only during spring32. A visual inspection of the AWC maps suggests that the wave train of connectivity seen in SON during EN years has a pattern similar to the Pacific South American pattern 2 (PSA 2), while the shape of the connectivity seen during LN years in the SH reminds of the PSA1. In JJA, however, the response in the SH resembles during EN years the PSA1, while in other seasons the patterns of connectivity cannot simply be related to these patterns. Thus, our results suggest that the SH atmospheric response to El Niño and La Niña during the different seasons cannot be simply described in terms of PSA modes.

Figure 4-2 shows the zonal mean connectivity for each season. It can be seen that overall the tropics are most connected during LN years, particularly during JJA and DJF. Also, while the extratropical connectivity is larger during winter than in summer (for both hemispheres) for all phases of ENSO, there is a tendency for larger values during EN years particularly in the NH winter subtropics. On the other hand, during SON of EN (LN) years there are relatively more extratropical connections in the SH (NH) hemisphere compared to during LN (EN) years. From Figure 4-1, it is clear that the increase in connectivity in the

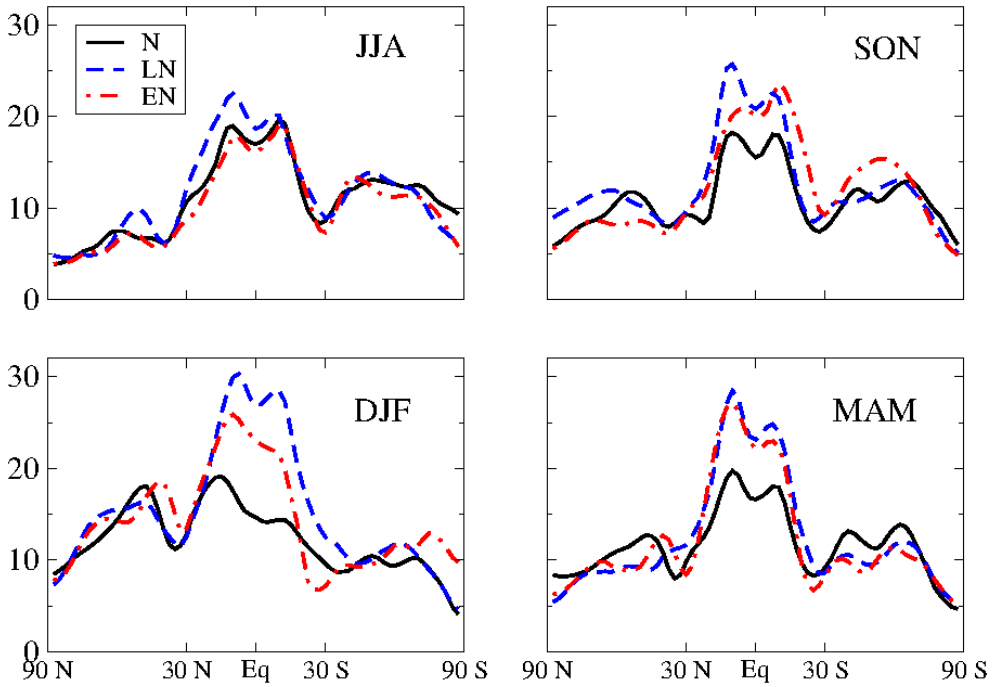

Figure 4-2: Zonal mean of AWC for the different phases of ENSO (EN, LN and N years).

SH during SON of EN years is due to the existence of a wave train that emanates from the Indian ocean, which is absent during LN years.

The connectivity wave train seen in the AWC for SON is physically consistent with the circulation anomalies present during EN events. Figure shows a composite analysis of the differences in the average values of EN and LN years with respect to N years for the eddy Z200 field. The results are statistically significant at 95% level, and show similar wave patterns as the AWC maps of Figure 4-1 considering EN and LN years, respectively. In particular, note that in the case of EN years the composite shows a wave emanating from the Indian ocean toward higher latitudes, which in the case of LN years does not exist. Moreover, anomalies are large in the tropical Atlantic during LN, but not during EN years, which is consistent with the map of AWC of Figure 4-1 (second row), and suggests that the large connectivity of the Atlantic basin is due to the LN influence on this basin and not due to possible impacts of other regions on the Atlantic. Finally, it can be seen that highly connected areas in the tropical Pacific correspond to regions where anomalies, on average, take negative values with respect to N years. These tropical regions with negative

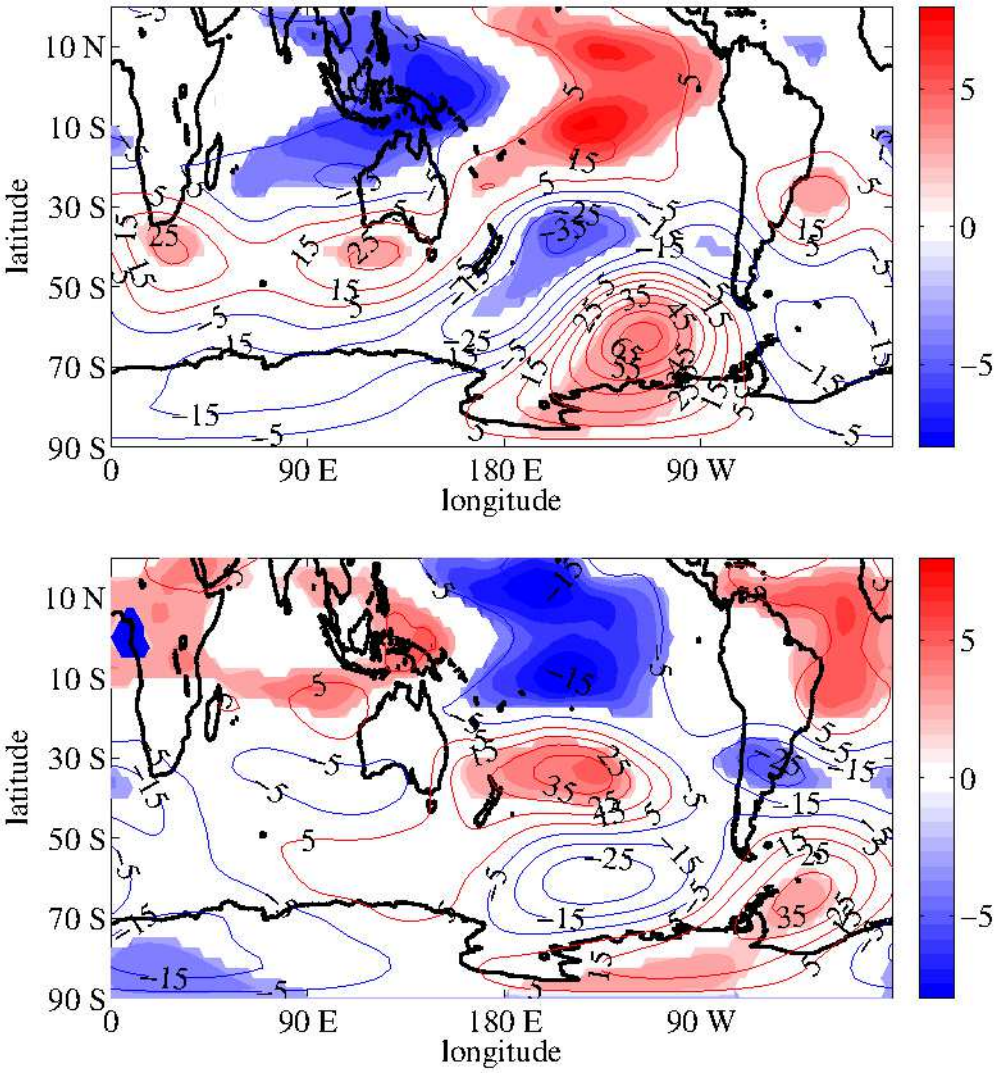

Figure 4-3: Composites of 200 hPa eddy geopotential height anomaly field. Average differences between EN (upper panel) and LN (lower panel) years with respect to N years. The areas where the differences are statistically significant (95%) are colored.

geopotential height anomalies have also less rainfall activity, suggesting that these areas are very important for the generation of Rossby waves.

In the following section we focus on the SON season in order to further understand the dynamics that causes the large SH connectivity and the different patterns found during EN and LN years. In particular, we assess whether the AWC patterns can be understood in terms of the dispersion of Rossby waves or as an enhacement of patterns of extratropical

intraseasonal variability during EN years forced by tropical anomalies.

4.3.2 Intraseasonal weather regimes

The first 10 leading EOFs of the eddy Z200 field are shown in Figure , and together account for more than 73% of the total variance. The first two EOFs, that explain about 14.9% and 12.1% of the total variance, are in quadrature with each other, consistent with eastward propagation. They are dominated by wavenumber 3 in midlatitudes and the wave trains describe an arch such that the main activity is over the Pacific-South American sector at approximately 60°S. The following pair of EOFs explain 8.4% and 8.1%, have also patterns with wavenumber 3 and are also in quadrature with each other, but in these cases anomalies occupy all the longitudinal domain having large amplitudes also to the South of the Indian ocean. The pattern is also mainly zonally distributed with maxima of amplitudes located between 45°S and 60°S. The 5^{th} and 6^{th} EOFs explain 7% and 6.4% of total variance, have largest amplitudes between 45°S and 70°S but also present structures in the subtropics. The 7^{th} and 8^{th} EOFs, that explain 5.5% and 4.9% of total variance, have wavenumber 4 and are also in quadrature. The waves are zonally distributed with maximum amplitudes between 45°S and 55°S. Finally, the 9^{th} and 10^{th}, explain 3.3% and 3% of the variance and are dominated by wavenumber 1 at 45°S.

Considering the field temporal evolution, i.e. all the points in the 10-dim subspace, we perform a K-means cluster analysis following the methodology explained above. As mentioned in the previous section, the number K of clusters has to be settled a priori. Thus, in order to obtain the best possible representation of weather regimes, we perform a reproducibility analysis explained in the Methodology Section. The result of the reproducibility analysis is presented in Figure 4-5 and shows that the best representation in terms of the reproducibility index is $K = 3$ as found in previous works [47, 48].

The weather regimes obtained from the cluster analysis are shown in Figure 4-6. In particular, we present the centroids of the three resulting clusters. However, as the centroids are computed from data filtered such that the timescale is intraseasonal and we want to study the interannual variability in terms of these intraseasonal regimes, we assigned each day of the daily mean 200-hPa height maps to each cluster by projecting the dataset with

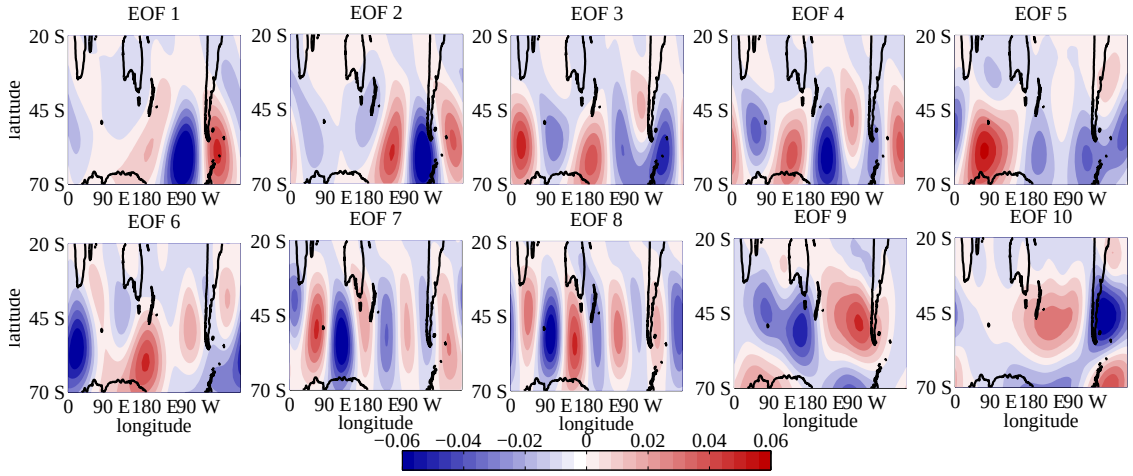

Figure 4-4: First 10 EOFs of daily mean eddy geopotential height at 200 hPa calculated between 20°S and 70°S. They represent the reduced dimensional space.

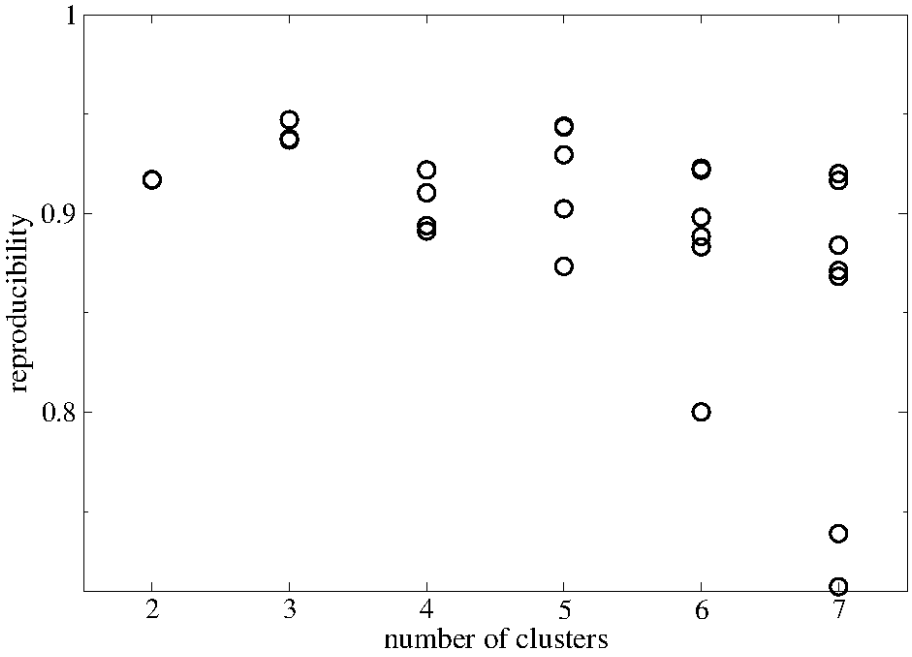

Figure 4-5: Reproducibility index.

interannual variations retained (low pass filtered and deseasonalized) into the intraseasonal EOF subspace, and assigned each day to the nearest centroid. All three regimes show

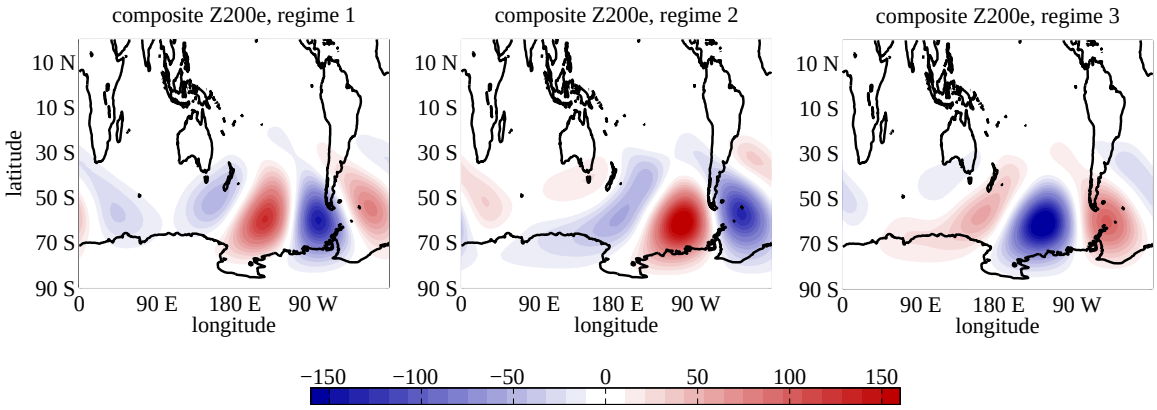

Figure 4-6: Weather regimes during SON: weather regime 1 (WR1, left panel), weather regime 2 (WR2, middle panel), weather regime 3 (WR3, right panel).

Table 4.1: Frequency of occurrence of weather regimes (WR1, WR2, WR3) depending on the phase of ENSO.

	WR1	WR2	WR3
El Niño	42.8%	28.9%	28.3%
La Niña	25.2%	30.1%	44.7%
Neutro	30.5%	33.1%	36.4%

wave trains dominated by wavenumber 3 with maximum weight in the South Pacific-South Atlantic sector, South of 30°S, like the PSA patterns and are consistent with previous studies [47]. They bear some resemblance with the composite of the eddy geopotential height during EN and LN shown in Figure 4-3, and thus part of the SH extratropical anomalies seen during the different phases of ENSO might be related to a change in the frequency of occurrence of intraseasonal weather regimes, as suggested previously [47].

Table 1 shows the dependence of the frequency of occurrence of the weather regimes on the ENSO phase. During EN years the first weather regime (WR1) occurs significantly more frequently than the other two regimes. During LN years, instead, the third weather regime (WR3) occurs most frequently. During N years WR3 also tends to occur more often than the others, but the differences are smaller. The second weather regime (WR2) does not have a preference of occurrence depending on the ENSO phase. Having in mind that a circulation regime that occurs more frequently leaves a residue after seasonal averaging, it is possible that part of the teleconnections observed during EN and LN can be explained in

these terms. Figure 4-7 shows the interannual anomalies constructed based on the frequency of occurrence of the weather regimes for EN and LN years. An inspection of the composites indicates that they have very similar spatial structures with anomalies of opposite polarities, thus suggesting that it is not possible to explain the different SH connectivity during EN and LN years during SON in terms of changes in the frequency of midlatitude intraseasonal weather regimes.

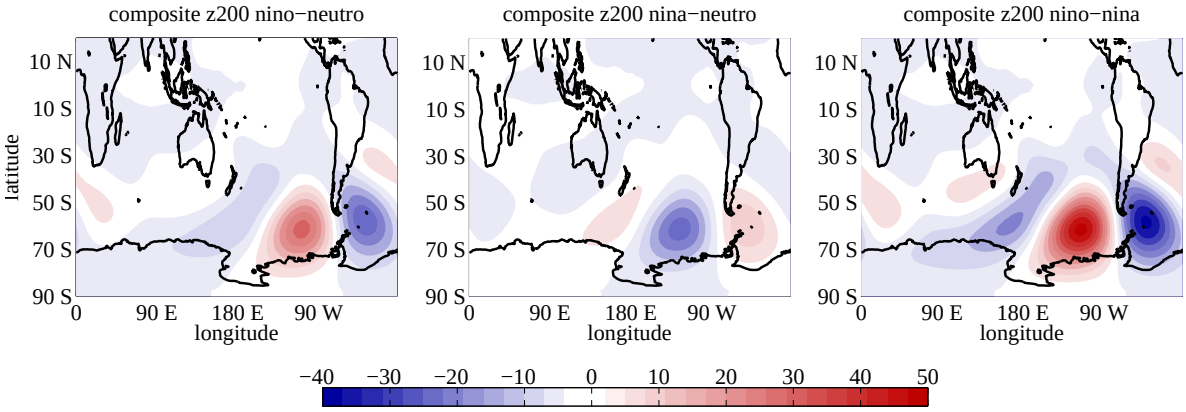

Figure 4-7: Composites of EN-N (left panel), LN-N (middle panel) and EN-LN (right panel) built from the frequency of occurrence of weather regimes.

Based on the above results we speculate that changes in the tropically-forced waves are able to explain the observed interannual changes in AWC. To further support this view we calculate the Rossby wave sources S for EN, LN and N years and perform the composite of EN-N and LN-N shown in Figure 4-8, analogously to Figure 4-3. Overall, it is clear that the S anomalies have opposite signs in the subtropical Pacific during EN and LN years, consistent with the atmospheric circulation anomalies that characterize each phase of ENSO, in particular those related with changes in the subtropical jet. Moreover, during EN years there is a significant S to the west of Australia that is not present in LN composite. This anomalous S lies at the right location to be acting as the forcing of the stationary wave that emanates from the Indian ocean and propagates eastward and poleward, as seen in Figure 4-3 and in the connectivity map of Figure 4-1 (second row). This result supports the view that the different connectivity during EN and LN in the SH extratropics during springtime can be understood in terms of tropically forced Rossby waves of different origins.

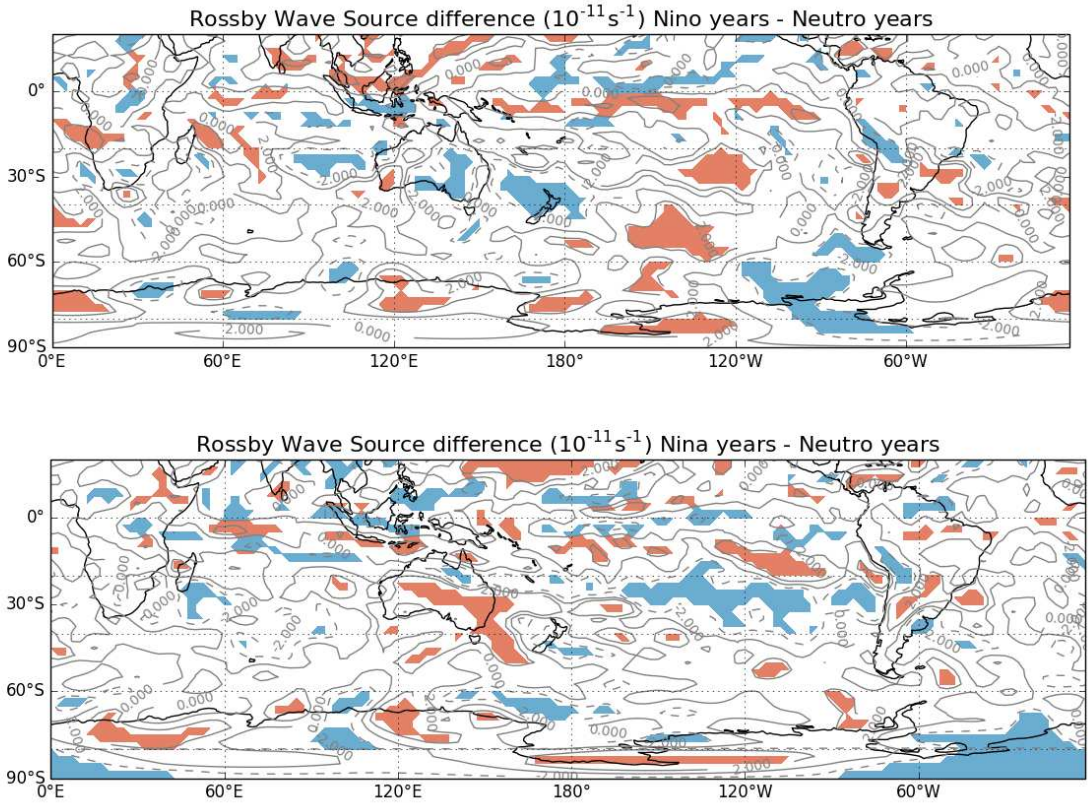

Figure 4-8: Composites of Rossby Wave Source. Average differences between EN (upper panel) and LN (lower panel) years with respect to N years. The areas where the differences are statistically significant (95%) are colored.

4.4 Conclusions

We have studied the upper level atmospheric connectivity in different seasons of the year using techniques from climate network analysis as well as the more commonly used composite analysis. Given that ENSO is the largest climate phenomenon on interannual time scales, we classified the years of the studied period according to the occurrence of El Niño, La Niña and Neutral years in order to study interannual changes in connectivity. Overall, results show that the AWC calculated for the eddy Z200 presents strong seasonal and interannual variability.

We have found that during Neutral years the tropical band is the most connected region of the world, and the Pacific ocean is the main hub. Note that the AWC is largest in the tropical

band because it acts as a wave guide to any disturbance. That is, atmospheric perturbations in the tropics tend to propagate to the east as Kelvin waves and to the west as Rossby waves, thus connecting the whole band. Moreover, Neutral years show a large seasonal dependence: SON is the season with overall least connectivity and has the maximum in the central-east Pacific; DJF (JJA) is the season of maximum connectivity in the NH (SH) probably because the strong winter jet stream acts as a source of tropically-induced perturbations as well as due to the presence of well known intrinsic modes of variability such as the NAO and PNA in the NH that are more active during wintertime.

Except in JJA season, which typically coincides with the beginning of the ENSO phenomenon, during the rest of the year the atmospheric connectivity is much larger during El Niño and La Niña years, in particular in the tropical region. During SON the EN AWC map shows large values in the western Pacific and South Indian Ocean and a wave like feature that describes an arch from the Indian ocean toward the southern pole and then back to the South Atlantic. During LN the AWC map shows a maximum in the central-eastern Pacific and Atlantic basin as well as a wave-like feature that appears to emanate from the central Pacific. This result supports a differential response of the SH extratropics to ENSO during austral springtime, that is, the circulation anomalies associated to extreme phases of ENSO are not opposite. This has been previously suggested by other authors [76, 78] for the NH winter, but to our knowledge this is the first time it has been shown for the SH spring. Note that our results also show that during DJF the NH extratropical connectivity is different for EN and LN cases, with the former having the largest values in the central Pacific and the latter in the western Pacific and Atlantic oceans. In MAM the main difference between EN and LN occurs in South America which shows larger (smaller) connectivity to the North (south) of 5°S during EN years. Finally, our results suggest that the extratropical atmospheric response to ENSO cannot simply be described in terms of PSA patterns.

In order to deepen our understanding of the differential response to ENSO in the SH extratropics during springtime we assessed whether it can be understood in terms of tropical Rossby wave forcing or an enhancement of patterns of extratropical intraseasonal variability during EN years. The calculation of the intraseasonal weather regimes for the SH extratropics lead to a classification into three clusters characterizing wave trains of wavenumber 3.

Considering the mean flow, these wave trains propagate from the Pacific into the Atlantic basin. Even though the frequency of ocurrence of these weather regimes changes depending on the phase of ENSO, consistent with previous studies [47], we found that shifts in the distribution of weather regimes can not explain the different connectivity patterns found in the AWC maps during EN and LN years. On the other hand, the composite of 200 hPa eddy geopotential height together with the Rossby Wave Source supports the view that the extratropical differential response to ENSO is a result of different waves forced from different tropical regions during EN and LN years. This result is not inconsistent with previous studies that suggest that at least part of the SH extratropical response to ENSO can be explained in terms of shifts in the frequency of ocurrence of weather regimes [47], because that study focuses on the linear component of the response to EN and LN events.

To finalize, we would like to stress that the use of complex network techniques complement and expand the results found using standard composite or EOF techniques. The use of the AWC allowed a systematic study and concise representation of global atmospheric connectivity in upper levels during EN, LN and N years. In this study we focus on SON, which showed distinct extratropical responses to tropical forcing, but other seasons show different connectivities in the tropical band during EN, LN and N years that are also uncovered by the methodology. Also, as shown in Figure 4-3, in SON composite maps during EN and LN show similar features as the corresponding AWC maps. However, from the composites it is not possible to determine which tropical regions are the most connected as anomalies have similar magnitudes particularly for EN years. Also, in the extratropics the centers of maximum connectivity do not coincide with those having largest eddy Z200 anomalies in the composites. A similar result is found comparing the leading EOF (that explain a small amount of variance, see Section 4.2.2) with the AWC patterns for different ENSO phases (not shown), and is consequence of the fact that higher order EOFs are needed to explain the teleconnections during EN and LN years.

Chapter 5

Large-scale patterns of unpredictability and response to insolation in atmospheric data[1]

Understanding the complex dynamics of the atmosphere is of paramount interest due to its impact in the entire climate system and in human society. Here we focus on identifying, using nonlinear dynamics tools, the geographical regions which have similar atmospheric properties. To analyze this problem we use surface air temperature (SAT) time series with monthly resolution, recorded at a regular grid covering the Earth surface. We consider two datasets: NCEP CDAS1 and ERA Interim reanalysis. We show that two surprisingly simple measures are able to extract meaningful information: i) the distance between the lagged SAT and the incoming solar radiation and ii) the Shannon entropy of SAT and SAT anomalies. The distance uncovers well-defined spatial patterns formed by regions with similar SAT response to solar forcing while the entropy uncovers regions with similar degree of SAT unpredictability. The entropy analysis also allows us identifying regions in which SAT has extreme values. Importantly, we uncover differences between the two datasets which are due to the presence of extreme values in one dataset but not in the other. Our results indicate that the distance and entropy measures can be valuable tools for the study of other climatological variables, in particular to detect anomalies and perform model intercomparisons.

[1] The results of this chapter are published in F. Arizmendi, M. Barreiro and C. Masoller, Scientific Reports, 7, 2017 [87].

5.1 Introduction

Large-scale climate phenomena have attracted great interest in the last decades, as improving the understanding of climate interactions is crucial for advancing long-term forecasts. A lot of research is nowadays focused in the development of appropriated data analysis tools, and a methodology that has been proven to be valuable is based on complex networks [54, 58, 56, 55]. Within this approach, a climate network is defined over a regular grid of geographical locations (nodes), and the links between pairs of nodes are defined by performing bi-variate analysis of the time series of climate variables recorded at the nodes. Different measures have been used to infer the links (cross-correlation, the mutual information, the conditional mutual entropy, Granger causality [57, 59, 88, 82, 89, 90, 91, 92]), and similar, or different, network structures have been unveiled, depending on the measure employed and the significance confidence level used.

Important challenges for inferring the connectivity of the climate system include the role of external solar forcing and the role of climate variability with different time-scale with respect to that of the analysis. For example, when the analysis is focused on phenomena at inter-annual or longer time-scales, weather systems concentrated at 3-7 days may be regarded as noise.

A potential drawback of the climate network method is that, if the time series in two nodes have similar characteristics, these regions can appear as 'linked', in spite of the fact that there might not be genuine underlying interaction between the climate variables in the two regions. Climatic similarities can be due to physical processes that act in a similar way in distant regions, and produce similar effects. Such climatic similarities might be reflected as network links, depending on the similarity measure used to construct the network [92] . On the other hand, regions with different climate characteristics may as well be linked by genuine long-range interactions. It is therefore important to perform uni-variate time series analysis to yield insight into the interpretation of the inferred links.

Here we consider two reanalysis datasets (NCEP CDAS1 and ERA Interim) and analyze the properties of surface air temperature (SAT) time series with monthly resolution, focusing on quantifying the SAT response to solar forcing and the degree of disorder or unpredictabil-

ity of SAT variability. We chose the SAT field because it is an important climatological field that has been commonly used to define climate networks [55, 56, 57, 58, 59, 82, 88, 90]. We address the following questions: is it possible to identify geographical regions with similar SAT response to solar forcing? Where are the regions with strongest distortion? Is it possible to identify regions with similar degree of SAT unpredictability? Where are the most unpredictable regions? We are interested in the relationship between solar forcing and SAT unpredictability because is important to characterize the geographical regions where this field is more/less predictable. A main contribution to SAT predictability comes from solar forcing that induces, in many regions, an oscillatory behavior with a well-defined periodicity.

To study SAT response to solar forcing we use the insolation (i.e., the local top-of-atmosphere incoming solar radiation) as a proxy of solar forcing and compare the SAT and the insolation waveforms by means of a distance measure between time series. As we are interested in assessing the similarity of the two waveforms, the two time series are normalized to zero mean and unit variance and the SAT time series (response) is shifted by an appropriated number of months, $\tau > 0$, with respect to the insolation (forcing). To quantify the degree of unpredictability or disorder of SAT time series we use the standard measure of information theory, Shannon entropy [49], computed from the probability distribution function (PDF) of SAT and SAT anomaly values (the anomalies are calculated from SAT time series by removing the seasonal cycle).

The distance and entropy measures employed here are well known. A drawback of the distance measure is that it is non-zero for general forms of linear convolutions [i.e., it returns an non-zero value for the distance between x and y, if they are linearly related as $y(t) = \int_0^t g(t-t')x(t')dt$]. In spite of this drawback, our results demonstrate that, at least for the analysis of monthly SAT time series, the distance yields meaningful results. Regarding the entropy measure, a main drawback is that it does not capture the information about the ordering of the data values in the time series: the entropy returns the same value if the data is shuffled randomly. This is because randomly shuffling the data do not modify the PDF of SAT values. In spite of these important drawbacks, we demonstrate that both, the distance and the entropy provide meaningful information about large-scale atmospheric phenomena. Specifically, in tropical oceanic regions, well defined spatial patterns are uncovered, with

large distance and low entropy values; in the extra-tropics, localized regions associated with sea ice are detected; in the continents, rainforest regions are also detected. In addition, in specific geographical regions, differences between the two reanalysis dataset are identified, which are due to the presence of extreme values in the data (the quantitative definition of extreme values depends on the specific system extreme fluctuations in hydrodynamics, climate or optics are of very different magnitude; however, a general criterion is that the PDF of the data values in the time-series has a non-Gaussian, long-tailed shape).

5.2 Data sets and measures

We consider monthly mean SAT data from two reanalysis data sets: NCEP CDAS1 [93] and ERA Interim [34]. The spatial resolution is 2.5° /1.5° and cover the time-period [1949-2015]/[1980-2014] respectively. The NCEP CDAS1 reanalysis has $N = 10224$ time series of length $L = 792$ while the ERA Interim, $N = 28562$ and $L = 408$. The insolation at the top of the atmosphere is calculated as in Berger [94], as a function of day of year and latitude. Then, monthly averaged values for every latitude are calculated.

For each raw SAT time series, $r_i(t)$, we first normalize to zero-mean and unit variance, obtaining the normalized SAT, $y_i(t)$ [where $i \in [1, N]$ and $t \in [1, L]$]. Then, we calculate the climatology time series, $c_i(t)$, and the anomaly time series, $z_i(t)$ as follows. The climatology (or seasonal cycle) is computed as the SAT value each month, $y_i(t)$, averaged over the SAT values in that month in all years,

$$c_i(t) = (1/Y) \sum_{n=0}^{Y} y_i(t + 12n), \tag{5.1}$$

where $Y = L/12$ is the number of years (66 or 34 depending on the dataset). In this way we obtain a time series of length 12 months which is extended to cover the whole period $t \in [1, L]$ months by considering the climatology a periodic function with period equal to one year. The anomaly time series is the difference between the SAT and the climatology, i.e., $y_i(t) - c_i(t)$.

To study SAT response to solar forcing we compute the distance between SAT time series in grid point i, $y_i(t)$, and the insolation time series in the same point, $x_i(t)$. The distance is

calculated as

$$d_i = \frac{1}{T}\sum_{t=1} |y_i(t+\tau_i) - x_i(t)|. \tag{5.2}$$

which is known as *taxicab* metric. We further demonstrate that similar results are obtained by using the Euclidean distance, $\sum_t [x(t) - y(t)]^2$. Because we are interested in measuring the similarity of the shape of x and y waveforms, the two time series are normalized to zero mean and unit variance.

In Eq. 5.2 $\tau_i > 0$ is a lag that allows to take into account inertia and/or memory effects, and needs to be appropriately chosen. This lag is expected to be more important in the oceans in comparison with land masses, because of the larger heat capacity of water. A natural choice is the value of τ_i that minimizes the distance between the insolation and the climatology (the averaged monthly SAT); similar results were found when considering the raw SAT time series instead. We search the minimum of the distance considering τ_i values in the interval of [0-4] months because of the lack of physical mechanisms that could result in a longer delayed response of the climatology to the insolation. The robustness of the results with respect to the maximum lag was checked by performing a sensitive analysis. While SAT time-series has a strong deterministic component due to solar forcing (which imposes, in many regions a well-defined periodicity and thus, provides some predictability), SAT anomaly (SATA) values are much more unpredictable, because the deterministic part (the seasonal cycle) is essentially removed. To quantify the degree of disorder or unpredictability of both, SAT and SATA time series we calculate Shannon entropy [49]. The entropy is normalized to the maximum entropy, of the uniform distribution:

$$H_i = -\sum_{1}^{M} p_n^i \log p_n^i / \log M \tag{5.3}$$

Here p_n^i, $n \in [1, M]$ is the probability distribution of the values in the ith time series ($i \in [1, N]$) and M is the number of bins. M is the same for all time series within a reanalysis database, but is adjusted in each database to take into account the different length of the time series: we use $M = 20$ for ERA Interim and $M = 40$ for NCEP CDAS1. This gives similar average data points per bin (20). Similar results are obtained with lower M.

The bin size is determined by the local extreme values in each time series, i.e., $dy_i = (\max y_i - \min y_i)/M$. While this choice, at first sight, might seem contradictory with performing 'inter-site comparisons', it allows to resolve with adequate precision the shape of the PDF in each site. With a uniform bin size a good resolution of the shape of the PDF in each site is not possible because there are regions where SAT/SATA has large variations, and others in which the variations are considerably smaller.

5.3 Results

We begin by analyzing the distance between SAT and insolation time series. First, we consider the situation in which the SAT is not shifted with respect to the insolation, i.e., all lag times τ_i are taken equal to zero. Fig. 5-1(a) displays the map of d_i values obtained from the analysis of the ERA Interim dataset. Similar results were obtained from the NCEP CDAS1 dataset.

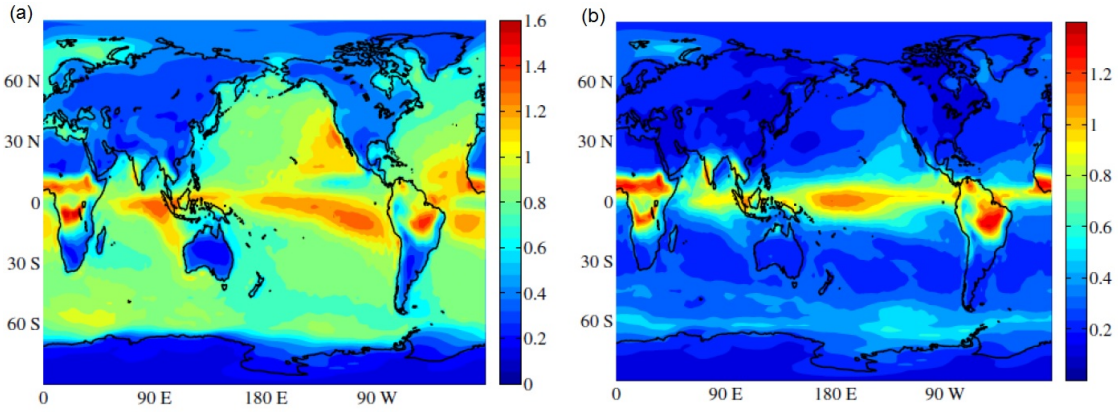

Figure 5-1: (a) Map of distances d_i calculated from Eq. 5.2 when the forcing (insolation) and the response (SAT) are not shifted ($\tau_i = 0$). (b) Map of distances, when the forcing and the response are shifted τ_i , with $\tau_i \in [0, 4]$.

We observe that over continental areas the distance values are smaller than over the oceans, but there are exceptions, as in the Amazon and in the African rainforest, which show high d_i values. Well defined spatial structures are seen over the oceans, in the cold tongues and areas associated with easterly trades and upwelling processes.

The large difference between the d_i values in the oceans and in the continents is mainly

due to the greater heat capacity of water, that results in long temperature memory in the oceans. When this effect is taken into account, at least in part, by shifting the SAT time series, the map of d_i values is strongly modified, as shown in Fig. 5-1(b). We observe that now the distance values over non-tropical continental areas and oceans are similar.

In the continents the largest d_i values appear over the tropical rainforests of Africa and South America. These are regions dominated by monsoons and can be expected to have large d_i values because during the summer rainy season, when the insolation has its highest values, the solar energy is used mainly for evaporation instead of for heating. In the oceans there are coherent spatial structures, with high d_i values, in regions which tend to coincide with regions of deep convection in the Atlantic, Pacific and Indian oceans, including the Intertropical Convergence Zone (ITCZ). In these regions the SAT and rainfall are strongly coupled so that relatively small changes in SAT gradients modulate and shift the ITCZ. In particular, air-sea coupling in the eastern basins induce oceanic cold tongues which together with the continental geometry maintain warm waters and the ITCZ to the North of the equator. Thus, the high distance values in this region can be interpreted as being due to the strong air-sea coupling. Outside the 10°S-10°N band, the higher d_i values over the eastern subtropical North Pacific, as compared to the western basin, can be due to the influence of the semi-permanent anticyclone and associated stratus clouds which shield solar radiation. High latitude oceans (southern Ocean, Labrador sea, Greenland sea) also have relatively large d_i values, which can be explained by the existence of seasonal sea ice in the regions.

The map of τ_i values shown in Fig. 5-2 uncovers a rather symmetric behavior between both hemispheres. Overall, extratropical land masses have a lag of about 1 month, while extratropical oceans present a lag of 2 months, in agreement with McKinnon et al. [95]. In tropical oceanic areas τ_i displays values in the [0-4] range, with values close to 0 and 1 in the ITCZ region and $\tau_i = 3$ in the eastern basins dominated by stratus clouds. Two continental regions, the African tropical rainforest and the Amazon rainforest, have a lag of 4 months, which could be due to the fact that in these regions SAT is colder in the summer compared to the spring because of intense rainfall.

Next, we present the results of the analysis of SAT unpredictability. Figures 3(a) and 3(b) display the entropy of SAT time series calculated from NCEP CDAS1 and ERA Interim

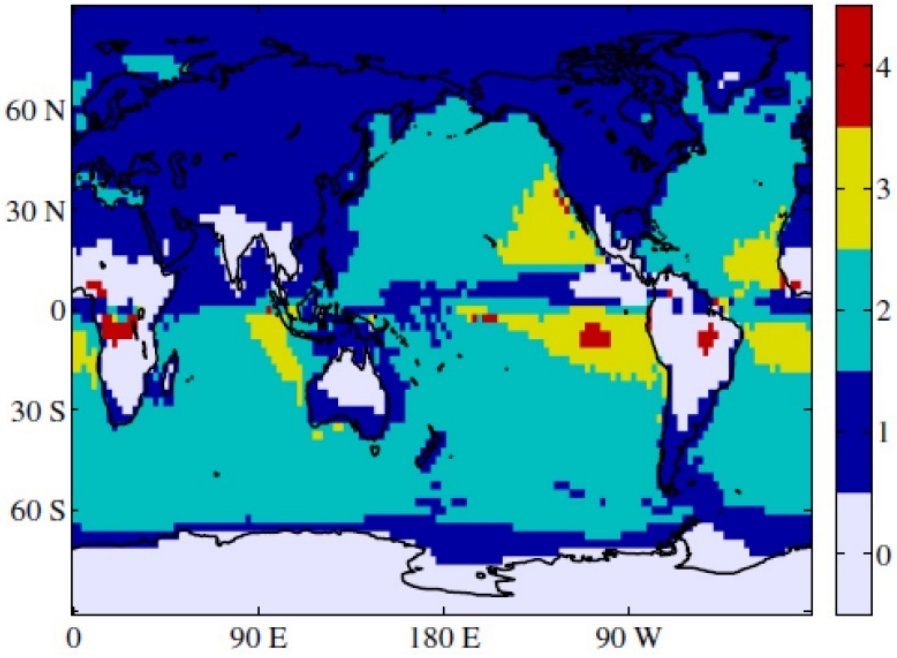

Figure 5-2: Map of τ_i values computed from ERA Interim reanalysis, with $\tau_i \in [0, 4]$.

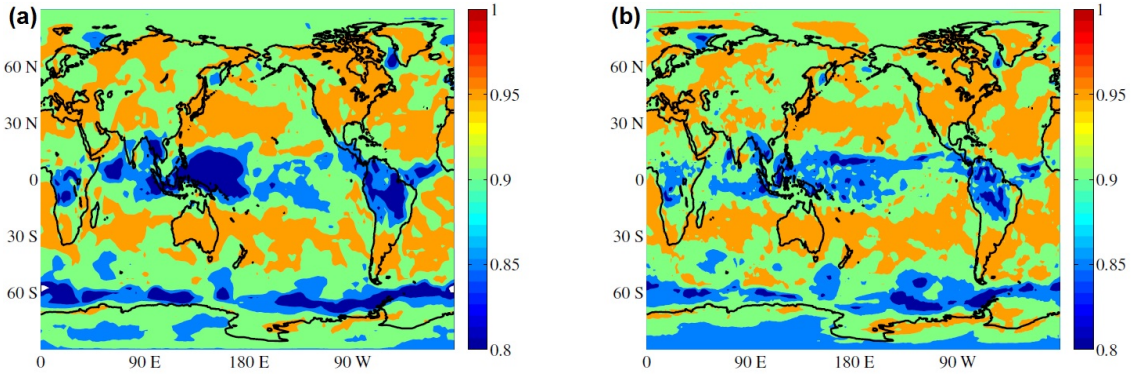

Figure 5-3: Shannon entropy of the SAT time series. (a) NCEP CDAS1 reanalysis, (b) ERA Interim reanalysis.

reanalysis respectively. The main spatial patterns in the tropics resemble those in the distance map, Fig. 1(b), however, regions with large d_i values tend to have low entropy values. The linear correlation coefficient between H_i and d_i values is -0.45.

In NCEP CDAS1 reanalysis data, Fig. 5-3(a), there is a difference with ERA Interim reanalysis, Fig. 5-3(b), in the western Pacific. In this region, as will be discussed latter, the entropy values are relatively low because of the presence of extreme values (outliers) in the

time series which render the PDF to be long-tailed. The entropy analysis of SATA time-series is presented in Fig. 5-3: panels (a) and (b) correspond to NCEP CDAS1 and ERA Interim reanalysis respectively, while panels (c) and (d) display a detail of the region where differences between the two datasets are found. In both SATA entropy maps we observe that the main spatial patterns in the tropical areas disappear and only those associated with sea ice remain.

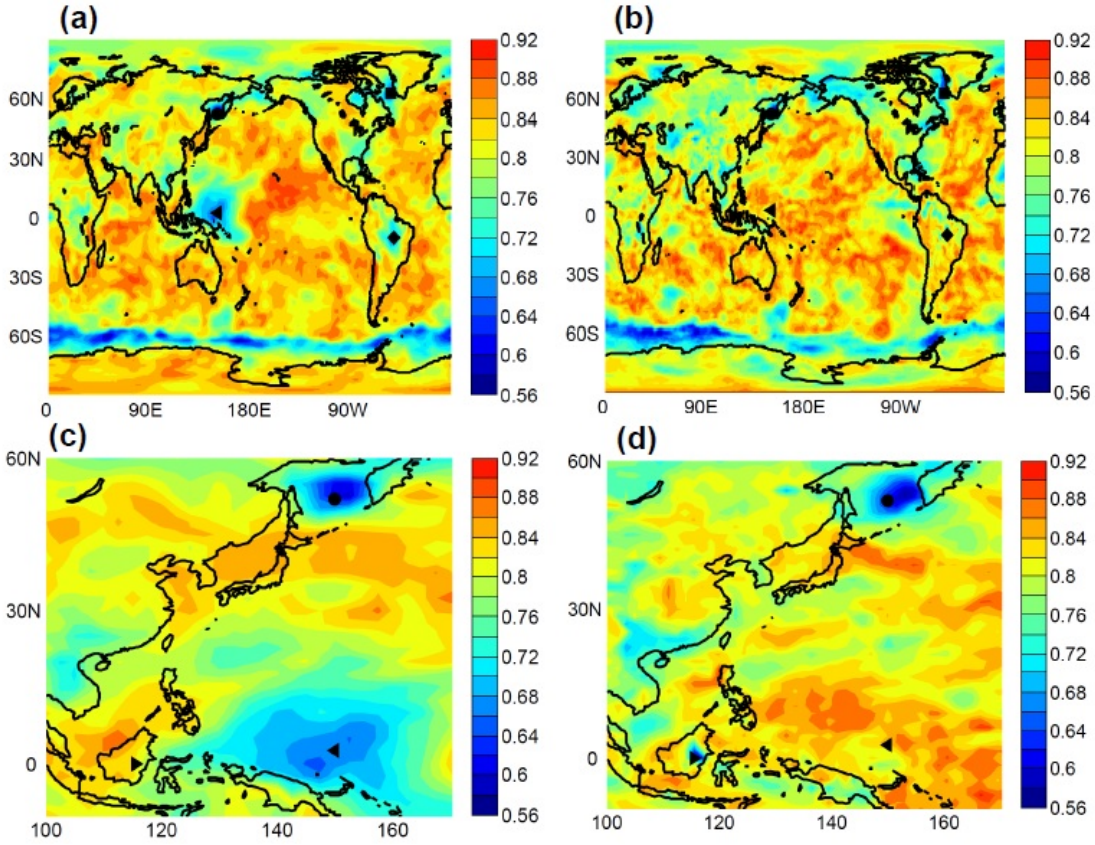

Figure 5-4: Map of Shannon entropy of SAT anomalies calculated from (a,c) NCEP CDAS1 reanalysis and from (b,d) ERA Interim reanalysis. The symbols indicate the geographical locations considered in Fig. 5-5: Sea of Okhotsk (northwest Pacific, indicated with a circle), Labrador Sea (square), Borneo (right triangle), western equatorial Pacific (left triangle) and the Amazonian region (diamond). Panels (c) and (d) show a detail of the western Pacific area: it can be observed that the two reanalysis have similar low entropy values in the region indicated with a circle, but have different entropy values in the regions indicated with left and right triangles. As shown in Fig. 5, the difference is due to the presence of extreme values in one dataset but not in the other.

This observation can be explained by the fact that in regions associated with sea ice

there is a strong seasonality in SAT variance that is not removed with the mean seasonal cycle. Examples of the SAT time series in these regions [indicated with a circle and with a square in Figs. 5-4(a) and 5-4(b)], are shown in Figs. 5(a) and 5(b). In these regions, the winter temperature can decrease considerably as the ice caps insulate the atmosphere from the ocean. In consequence, SATA PDF has a long tail in low temperature values, which is captured as a smaller entropy.

We also note that the map of the entropy computed from NCEP CDAS 1 reanalysis is similar to that obtained from ERA Interim except in the western tropical Pacific where NCEP CDAS 1 data shows lower entropy. A difference is also observed in the Amazon region, where NCEP CDAS 1 data also has lower entropy. Figures 5-5(c)-(e) show the SATA time series in these regions, which are indicated with the symbols left triangle, right triangle and diamond in Fig. 4. We observe that there are some extreme values which occur in one reanalysis but not in the other.

5.4 Discussions

In this section we compare our findings with results obtained within climate network approach. First, we can relate our findings with those of Hlinka et al. [90] that presented a methodology for the identification and quantification of the non-linear contribution to the interaction information, able to identify main sources of nonlinearity in the nodes couplings. A comparison of the spatial structures uncovered in Fig. 5-1(b) here, with those in Fig. 5 of Ref. [90] (central panel), reveals that some of the areas with large distance values in Fig. 5-1(b) tend to coincide with the areas with nonlinear contribution to the mutual information. For example, the African and the Amazon rainforests are clearly seen in both maps. However, there are also differences: a ring in the Ocean around the Antarctica and a well defined region near the North pole (Greenland sea) that are strong signals in Fig. 5 of Ref. [90] are not strong in Fig. 5-1(b) here. However, in Fig. 5-4(b) here, we note that these regions are regions of low SAT anomaly entropy. Therefore, the regions with nonlinear contribution to mutual information seen in Fig. 5 of Ref. [90] are seen either in Fig. 5-1(b) or in Fig. 5-4(b) here. This observation suggests that the connectivity of these regions

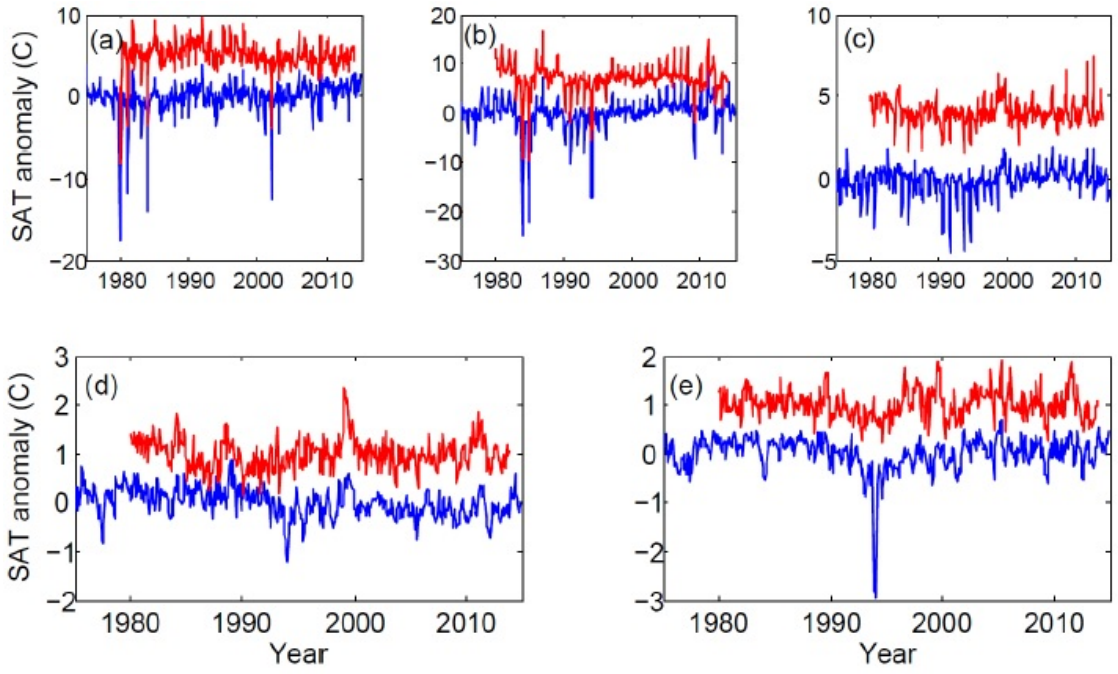

Figure 5-5: SAT anomaly time series from NCEP CDAS1 reanalysis (blue) and ERA Interim (red, displaced vertically by 1°C for clarity). Panels (a)-(e) correspond to the geographical locations indicated in Fig. 5-4 with circle (Sea of Okhotsk, northwest Pacific), square (Labrador Sea), diamond (Amazonian region), right triangle (Borneo) and left triangle (western equatorial Pacific), respectively. In panels (d) and (e) some extreme values occur in one reanalysis but not in the other.

might reflect, in part, similar response to solar forcing, and/or similar SAT variability, with long-tailed distribution of SAT anomalies.

We can also relate our findings with those of Tirabassi and Masoller [88], who analyzed the effects of lag-times in networks constructed from monthly SAT reanalysis. The maps of the lag-times between SAT time series in different regions, Fig. 2 of Ref. [88], for the three regions considered (Australia, El Niño basin and Mongolia) have a structure that is remarkably similar to the map presented in Fig. 5-2 here; however, a different procedure was used to compute the maps: the lag between any two regions was determined such as to superpose the two seasonal cycles (i.e., the lag was chosen to maximize the cross-correlation between the two SAT time series). In contrast, here the lag is determined by superposing the SAT time series and the insolation (by minimizing the distance between SAT time series and the insolation). The two approaches give several consistent observations. For example,

i) in Fig. 2, right panel of Ref. [88] the large red area in the North Hemisphere represents the geographical regions where the seasonal cycle is in-phase with that in Mongolia; comparing with Fig. 5-2 here we note that these are regions with a one month lag between the seasonal cycle and the isolation;
ii) in Fig. 2, left panel of Ref. [88] the large red area in the Southern Hemisphere represents the regions where the seasonal cycle is in-phase with that in Australia; comparing with Fig. 5-2 here we note that in those regions, the lag between the seasonal cycle and the isolation is 0 or 1 month;
iii) in Fig. 2, center panel of Ref. [88] the red area in the equator represents the regions where the seasonal cycle is in-phase with that in El Niño basin; comparing with Fig. 5-2 here we note that several well-defined structures appear in both maps, in particular the regions in yellow in Fig. 5-1, where the lag between the seasonal cycle and the isolation is 3 months, correspond to regions either in red or in light blue in Fig. 2, center panel of Ref. [88].

The mutual lags between SAT time series in different regions were then used by Tirabassi and Masoller [92] to infer climate communities, defined as regions that share similar properties of SAT time series. The map of the communities obtained, which are the regions which have a synchronous seasonal cycle, Fig. 1 of Ref. [92], has also several features in common with the map presented in Fig. 5-2 here. For example, the large communities formed by the oceans in the northern and in the southern hemispheres, represented with orange and blue in Fig. 1 of Ref. [92], have in Fig. 5-2 here a 2 month lag between SAT and the insolation. The similarity of these maps, despite the different way they were obtained, suggests that, at monthly scale, the mutual lag times are mainly determined by the seasonality induced by solar forcing.

The similarity of the ERA Interim maps and NCEP CDAS1 maps provides convincing evidence of the robustness of the results. The similarity is particularly remarkable considering that the datasets have different spatial resolution and temporal coverage. The different spatial resolution leads to more/less detail in the structures but has, in principle, no other influence in the results of the analysis. In contrast, some of the differences found when comparing the maps obtained from the two datasets could be attributed to their different

temporal coverage.

A relevant problem is to estimate the uncertainty of the results. This could be done by using a subsampling approach; however, SAT/SATA non-stationarity may hamper the uncertainty estimation. An alternative approach is to perform the analysis using different runs of the same atmospheric general circulation model, with slightly different initial conditions. However, this study is out of the scope of the present work and is left for future work.

5.5 Conclusions

In this work we have investigated the statistical properties of a climatological field (the surface air temperature, SAT) using two monthly reanalysis datasets: ERA Interim and NCEP CDAS1. We have quantified SAT unpredictability by means of Shannon entropy, H_i, and we have analyzed the response to solar forcing, by means of a distance measure, d_i, that assesses the similarity in the shape of the time series of the top-of-atmosphere incoming solar radiation (the insolation, x_i) and the SAT, y_i, both having been previously normalized to have zero mean and unit variance. A delay in the response of SAT to solar forcing was taken into account by lagging SAT with respect to the insolation.

We found that these two measures provide meaningful insight into global properties of SAT time series. In the distance map the tropics have considerable larger distance values, in comparison with the extratropics. There are well-defined structures in the oceans, over the Intertropical Convergence Zones, and over some continental areas, especially in regions largely dominated by monsoons, such as the tropical rainforest in Africa and South America, as well as over India. These regions do not respond to local insolation, but are characterized by strong air-sea coupling or land-atmosphere interaction which involve non-local processes.

In the SAT entropy maps we found qualitatively similar spatial structures, but with opposite high/low values: regions with high d_i values tend to correspond to regions of low entropy. When the entropy was calculated from SATA, the tropical spatial patterns disappeared but those in the high latitudes remained. This was interpreted as due to the fact that in high latitudes, mainly because of the presence of sea ice, there is a strong seasonality in variance that remains even when the annual cycle is removed.

While we found that the ERA Interim maps and the NCEP CDAS1 maps are remarkably similar, the entropy analysis also allowed to identify, in a well-defined region of the tropical western Pacific, relevant differences between the two datasets, which are likely due to the fact that there are some extreme values which occur in one reanalysis but not in the other. The different temporal coverage of ERA Interim and NCEP CDAS1 datasets might be another reason of these differences. A detailed comparison is, however, outside of the scope of the present work, which is aimed at demonstrating the suitability of two measures (the distance and the entropy) for quantifying similar properties of SAT time series in different regions.

Chapter 6

Summary and perspectives

This thesis is the result of an exploration of the atmospheric circulation using the framework of complex systems. The study of atmospheric teleconnections in the SH, how they have evolved in time during the 20th century and how they vary on different time scales are the main questions we address in this study. Another important topic we have addressed in this thesis, is how is the response of the climatic system to the incoming solar radiation. To do that we consider standard climate analysis tools complement with the addition of complex networks and information theory techniques. The combination of classical and novel techniques allows to better understand the additional information that the complex network point of view can provide to improve our understanding of atmospheric teleconnections.

Chapters 3, 4 and 5 comprise the main results of this PhD thesis which are already published [51, 71, 87].

In Chapter 3, we presented a study of the $Z200e$ climate networks during the 20th century [51]. Using data from the NOAA 20th Century reanalysis [29], we have built functional networks of different nature, considering both linear and nonlinear techniques, and focusing on different time scales by means of symbolic analysis. We compared the two halves of the 20th century together with the evolution of the connectivity of core teleconnection areas such as Eastern and Western Pacific using a sliding temporal window of 30 years, finding interdecadal variability and different evolution depending of the nature of the network. Furthermore, we considered an ensamble of outputs of a AGCM model to disentangle the networks due to their forced and internal components. The results suggest that the main

patterns of connectivity captured by the reanalysis networks are due to the oceanically forced component, particularly on inter-annual timescales. On the other hand, on intra-seasonal timescales the atmospheric internal variability seems to play an important role in determining the network structure. Beyond the particular results of this exploratory work, that are fully described in the conclusions of chapter 3, it would be interesting to deepen some aspects of the findings and also to extend the use of these tools. Specifically, what is the role for the MJO in generating the connectivity on intraseasonal time scales? Additionally, it would be interesting to perform the same analysis of the connectivity and the sliding 30 years window but to focus in key regions of the extratropics.

In Chapter 4, we concentrated our efforts on the ENSO teleconnections in the Southern Hemisphere. With this objective we analyzed the interannual variability of the area weighted connectivity by considering different seasons separately. In addition, we analyzed the impact of the different stages of ENSO: El Niño, La Niña and Neutral years, finding strong seasonal and interannual variability. Given the results of the analysis of the $Z200e$ climate networks in Chapter 3 [71], we only considered linear interdependency measures among the nodes of the network.

Particularly, we have found that the most connected regions vary among seasons and ENSO phases. The details are described in detail in the Conclusions of Chapter 4. We further focused on the SON season, where El Niño AWC map shows large values in the western Pacific and south Indian Ocean and a wave-like feature that describes and arch from the Indian ocean toward the southern pole and then back to the south Atlantic. During La Niña, instead, the AWC map shows a maximum in the central-eastern Pacific and Atlantic basin as well as a wave-like feature that appears to emanate from the central Pacific. This result supports that the SH extratropical atmospheric response to ENSO is nonlinear during austral springtime in the sense of that the circulation anomalies associated to extreme phases of ENSO are not opposite. This has been previously suggested by other authors for the NH winter [76, 78], but to our knowledge this is the first time it has been shown for the SH spring. Moreover, our results also show that during DJF the NH extratropical connectivity is different for EN and LN cases, with the former having largest values in the central Pacific and the latter in the western Pacific and Atlantic oceans. In MAM the main difference

between EN and LN occurs in South America which shows larger (smaller) connectivity to the north (south) of 5°S during EN years. These different network structures and their associated teleconnections, may be better understood by using other graph theory measures such as the the betweenness centrality, clustering coefficient or the transitivity, to mention a few. This would be a very interesting future work.

We further deepen our analysis on the differential response to ENSO in the SH extratropics during SON. Specifically, by means of the K-means clustering technique and following a previously done procedure [48, 47], we calculated the SH weather regimes and their changes in occurrence depending on the ENSO phase. We found that a well representation of the weather regimes are three wave trains of wavenumber 3 that, taking into account the mean flow, propagate from the Pacific into the Atlantic basin. Even though the occurence of these weather regimes changes with the ENSO phase, it can not explain the differences in the connectivity patterns. On the other hand, composites analysis of the $Z200e$ and that of the Rossby Wave Source suggest that the differences during EN and LN years are a result of wave trains of different tropical regions.

These results increase our understanding of the regional seasonal climate predictability during different seasons as the latter strongly depends on the tropical extratropical connections. Moreover, in the case of the SH extratropics during springtime, the observed differential response to tropical forcing can be used as a test for climate models used for operational seasonal forecasting. Given that the current understanding is that depending on the evolution of the SST anomalies there are 2 kinds of El Niño, canonical and Modoki [96, 97], future work may address how this response vary according to the nature of the ENSO event.

Finally, in Chapter 5, we focused on the dynamics of the SAT field using complex systems techniques considering two of the most used reanalysis datasets [87]. Specifically, we have quantified SAT unpredictability by means of Shannon entropy, H_i, and we have analyzed the response to solar forcing, by means of a distance measure, d_i, that assesses the similarity in the shape of the normalized time series of the top-of-atmosphere incoming solar radiation (insolation) and the SAT. Given surfaces with different temporal response to the solar forcing, both fields were previouly synchronized by locally lagging the SAT time series.

Overall, we have found that the global patterns of both the Shannon entropy and the distance measure considered here are related and some interesting features of the SAT field dynamics stand out. Specifically, as expected, the structures obtained with the distance measure are highly related with precipitation. This is because, as we explained in section 1.2.2, in moist environments, most of the incoming energy is used for evaporation such that there is little impact on local temperature and therefore insolation and SAT normalized time series have dissimilar shapes, resulting in large distance values. Several areas as the ITCZ and the regions dominated by monsoons such as India or the tropical rainforest in Africa and South America are highlighted. The SAT dynamics in these regions are characterized by strong air-sea coupling or land-atmosphere interaction which involve non-local processes.

Surprisingly, we found qualitatively similar spatial structures in the SAT entropy maps but with opposite high/low values: regions with high distance values tend to correspond to regions of low entropy. However, when the entropy was calculated from SAT anomalies, the tropical spatial patterns disappeared but those in the high latitudes remained. This was interpreted as due to the fact that in high latitudes, mainly because of the presence of sea ice, there is a strong seasonality in variance that remains even when the annual cycle is removed. In addition, while we found that the ERA Interim maps and the NCEP CDAS1 maps are remarkably similar, the entropy analysis also allowed to identify, in a well-defined region of the tropical western Pacific, relevant differences between the two datasets, which are likely due to the fact that there are some extreme values which occur in one reanalysis but not in the other.

It will be interesting, for future work, to compare the results presented here with those obtained by using other distance and entropy measures that lack the drawbacks of the measures used here, which were discussed in the Introduction of Chapter 5. For computing the entropy, a promising approach for gaining more information is based in symbolic analysis [98]. In this approach a time series is first transformed into a sequence of symbols, and then, the entropy is computed from the probabilities of the symbols. In this case, depending on the specific rule employed to define the symbols, the entropy will capture different properties of the ordering of the values in the time series, and will give a different result when the data values are shuffled randomly. For computing the distance between SAT and insolation

time series, several advanced approaches can be used, for example, each time-series can be mapped into a network (by using, e.g., recurrence [99], visibility [100], or symbolic networks [101]) and then, the dissimilarity of the two networks obtained can be computed [102].

Bibliography

[1] E. N. Lorenz, "Deterministic nonperiodic flow," *Journal of the Atmospheric Science*, vol. 20, pp. 130–141, 1963.

[2] H. D. Dijkstra and M. Ghil, "Low-frequency variability of the large-scale cean circulation: A dynamical systems approach," *Rev. Geophys.*, vol. 43, no. 3, p. RG3002, 2005.

[3] M. Planck, *The Theory of Heat Radiation.* Blackiston's Son and Co, 1914.

[4] H. A. Brigman and J. E. Oliver, *The Global Climate System: Patterns, Processes, and Teleconnections.* Cambridge University Press, 2006.

[5] J. M. Lewis, "The story behind the bowen ratio," *Bulletin of the American Meteorological Society*, vol. 76, pp. 2433–2443, 1995.

[6] V. Bjerknes, "The problem of weather forecasting from the standpoint of mechanics and physics," *Meteorologische Zeitschrift*, vol. 21, pp. 1–7, 1904.

[7] J. R. Holton, *An Introduction to Dynamic Meteorology.* Academic Press, 1992.

[8] P. D. Sardeshmuhk and B. J. Hoskins, "Vorticity balances in the tropics during the 1982-1983 el niño-southern oscillation event," *Quart. J. R. Met. Soc.*, vol. 111, pp. 261–278, 1985.

[9] ——, "On the derivation of the divergent flow from the rotational flow: the χ problem," *Quart. J. R. Met. Soc.*, vol. 113, pp. 339–360, 1987.

[10] I. N. James, *Introduction to circulating atmospheres.* Cambridge University Press, 1994.

[11] G. K. Vallis, *Atmospheric and Oceanic Fluid Dynamics: Fundamentals and Large-Scale Circulation.* Cambridge University Press, 2006.

[12] E. Kalnay *et al.*, "The ncep/ncar 40-year reanalysis project," *B. Am. Meteorol. Soc.*, vol. 77, pp. 437–471, 1996.

[13] J. Bjerknes, "Atmospheric teleconnections from the equatorial pacific," *Monthly Weather Review*, vol. 97, no. 3, pp. 163–172, 1969.

[14] G. T. Walker, "Correlation in seasonal variations of weather, viii: A preliminary study of world weather," *Memoirs of the Indian Meteorological Department*, vol. 24, no. 4, pp. 75–131, 1923.

[15] ——, "Correlation in seasonal variations of weather, ix: A further study of world weather," *Memoirs of the Indian Meteorological Department*, vol. 24, no. 9, pp. 275–332, 1924.

[16] ——, "World weather iii," *Memoirs of the Royal Meteorological Society*, vol. 2, no. 17, pp. 97–106, 1928.

[17] T. N. Krishnamurti, "Tropical east-west circulations during northern summer," *Journal of the Atmospheric Science*, vol. 28, pp. 1342–1347, 1971.

[18] T. N. Krishnamurti, M. Kanamitsu, W. J. Toss, and J. D. Lee, "Tropical east-west circulations during northern summer," *Journal of the Atmospheric Science*, vol. 30, pp. 780–787, 1973.

[19] R. E. Huschke, *Glossary of Meteorology*. American Meteorological Society, 1959.

[20] A. Ångström, "Teleconnections of climate changes in present time," *Geogr. Ann.*, vol. 17, pp. 242–258, 1935.

[21] P. D. Sardeshmuhk and B. J. Hoskins, "The generation of global rotational flow by steady tropical divergence," *J. Atmos. Sci.*, vol. 45, pp. 1228–1251, 1988.

[22] B. J. Hoskins and D. J. Karoly, "The steady linear response of a spherical atmosphere to thermal and orographic forcing," *J. Atmos. Sci.*, vol. 38, pp. 1179–1196, 1981.

[23] G. Branstator, "Horizontal energy propagation in a barotropic atmosphere with meridional and zonal structure," *J. Atmos. Sci.*, vol. 40, pp. 1689–1708, 1983.

[24] B. J. Hoskins and T. Ambrizzi, "Rossby wave propagation on a realistic longitudinally varying flow," *J. Atmos. Sci.*, vol. 50, no. 12, pp. 1661–1672, 1993.

[25] T. Matsuno, "Quasi-geostrophic motions in the equatorial area," *Journal of the Meteorological Society of Japan*, vol. 44, no. 1, pp. 25–43, 1966.

[26] A. E. Gill, "Some simple solutions for heat-induced tropical circulation," *Quart. J. Roy. Meteor. Soc.*, vol. 106, pp. 447–462, 1980.

[27] K. E. Trenberth, G. W. Branstator, D. Karoly, A. Kumar, N. C. Lau, and C. Ropelewski, "Progress during toga in understanding and modeling global teleconnections associated with tropical sea surface temperatures," *J. Geophys. Res.*, vol. 103, pp. 14 291–14 324, 1998.

[28] H. A. Dijkstra, *Nonlinear climate dynamics*. Cambridge University Press, 2013.

[29] G. P. C. et al., "The twentieth century reanalysis project," *Q. J. Roy. Meteorol. Soc.*, vol. 137, pp. 1–28, 2011.

[30] J. M. Wallace and D. S. Gutzler, "Teleconnections in the geopotential height field during the northern hemisphere winter," *Monthly Weather Review*, vol. 109, pp. 784–812, 1981.

[31] K. C. Mo and M. Ghil, "Statistics and dynamics of persistent anomalies," *Journal of Atmospheric Science*, vol. 44, pp. 877–901, 1986.

[32] I. Szeredi and D. Karoly, "The horizontal structure of monthly fluctuations of the southern hemisphere troposphere. from station data." *Australian Meteorological Magazine*, vol. 35, pp. 119–129, 1987.

[33] K. C. Mo and J. N. Paegle, "The pacific-south american modes and their downstream effects," *Int. J. Climatology*, vol. 21, pp. 1211–1229, 2001.

[34] E. C. F. M.-R. W. F. (ECMWF), Shinfield Park, Reading, RG2 9AX, United Kindom.

[35] T. M. Smith and R. W. Reynolds, "Improved extended reconstruction of sst (1854–1997)," *J. Climate*, vol. 17, pp. 2466–2477, 2004.

[36] F. Molteni, "Atmospheric simulations using a gcm with simplified physical parametrizations, i. model climatology and variability in multi-decadal experiments," *Clim. Dynam.*, vol. 20, pp. 175–191, 2003.

[37] F. Kucharski, F. Molteni, and A. Bracco, "Decadal interactions between the western tropical pacific and the north atlantic oscillation," *Clim. Dynam.*, vol. 77, pp. 79–91, 26.

[38] M. Barreiro, A. Fedorov, R. Pacanowski, and S. G. Philander, "Abrupt climate changes: How freshening of the northern atlantic affects the thermohaline and wind-driven oceanic circulations," *Annu. Rev. Earth Planet. Sci.*, vol. 36, pp. 33–58, 2008.

[39] M. Barreiro, "Influence of enso and the south atlantic ocean on climate predictability over southeastern south america," *Clim. Dynam.*, vol. 35, pp. 1493–1508, 2010.

[40] F. Kucharski, A. Bracco, J. H. Yoo, A. Tompkins, L. Feudale, P. Ruti, and F. Molteni, "Atlantic forced component of the indian monsoon interannual variability," *Geophys. Res. Lett.*, vol. 35, p. L04706, 2008.

[41] F. Kucharski, A. Bracco, J. H. Yoo, A. Tompkins, L. Feudale, P. Ruti, and A. Dell'Aquila, "A gill-matsun-type mechanism explains the tropical atlantic influence on african and indian monsoon rainfall," *Q. J. Roy. Meteorol. Soc*, vol. 135, pp. 569–579, 2009.

[42] M. E. J. Newman, *Networks: An Introduction*. Oxford University Press, 2010.

[43] D. J. Watts and S. H. Strogatz, "Collective dynamics of 'small-world' networks," *Nature*, vol. 393, pp. 440–442, 1998.

[44] S. Milgram, "The small world problem," *Psychology Today*, vol. 2, pp. 60–67, 1967.

[45] A. A. Tsonis and P. J. Roebber, "The architecture of climate networks," *Physica A*, vol. 103, pp. 14 291–14 324, 2004.

[46] E. N. Lorenz, "Empirical orthogonal functions and statistical weather prediction," *Sci. Rep. No. 1 Statistic. Forecasting Proj., Dept. Meteor., MIT*, p. 49pp, 1956.

[47] G. Cazes-Boezio, A. W. Robertson, and C. R. Mechoso, "Seasonal dependence of enso teleconnections over south america and relationships with precipitation in uruguay," *J. Climate*, vol. 16, pp. 1159–1176, 2003.

[48] P. A. Michelangeli, R. Vautard, and B. Legras, "Weather regimes: Recurrence and quasi stationarity," *Journal of the Atmospheric Sciences*, vol. 52, 1995.

[49] C. E. Shannon, "A mathematical theory of communication," *Bell Syst. Tech. J.*, vol. 27, pp. 623–656, 1948.

[50] C. Bandt and B. Pompe, "Permutation entropy: A natural complexity measure for time series," *Phys. Rev. Lett.*, vol. 88, p. 174102, 2002.

[51] F. Arizmendi, A. C. Martí, and M. Barreiro, "Evolution of the atmospheric connectivity during the 20th century," *Nonlin. Processes Geophys.*, vol. 21, pp. 825–839, 2014.

[52] A. L. Barabási and R. Albert, "Emergence of scaling in random networks," *Science*, vol. 286, pp. 509–512, 1999.

[53] R. Albert and A. L. Barabási, "Statistical mechanics of complex networks," *Rev. Mod. Phys.*, vol. 74, pp. 47–97, 2002.

[54] A. A. Tsonis, K. L. Swanson, and P. J. Roebber, "What do networks have to do with climate?" *Bull. Amer. Meteorol. Soc.*, vol. 87, pp. 585–596, 2006.

[55] J. F. Donges, Y. Zou, N. Marwan, and J. Kurths, "Complex networks in climate dynamics," *Eur. Phys. J.-Spec. Top.*, vol. 174, pp. 157–179, 2009.

[56] K. Yamasaki, A. Gozolchiani, and S. Havlin, "Climate networks around the globe are significantly affected by el niño," *Phys. Rev Lett.*, vol. 100, p. 228501, 2008.

[57] M. Barreiro, A. C. Martí, and C. Masoller, "Inferring long memory processes in the climate network via ordinal pattern analysis," *Chaos*, vol. 21, p. 013101, 2011.

[58] A. A. Tsonis and K. L. Swanson, "Topology and predictability of el niño and la niña networks," *Phys. Rev. Lett.*, vol. 100, p. 228502, 2008.

[59] E. A. Martin, M. Paczuski, and F. Davidsen, "Interpretation of link fluctuations in climate networks during el niño periods," *Eur. Phys. Lett.*, vol. 102, p. 48003, 2013.

[60] A. Radebach, R. V. Donner, J. Runge, J. F. Donges, and J. Kurths, "Disentangling different types of el niño episodes by evolving climate," *Phys. Rev. E*, vol. 88, p. 052807, 2013.

[61] J. I. Deza, C. Masoller, and M. Barreiro, "Distinguishing the effects of internal and forced atmospheric variability in climate networks," *Nonlin. Processes Geophys.*, vol. 21, pp. 617–631, 2014.

[62] I. M. Held, *Stationary and quasi-stationary eddies in the extratropical troposphere: Theory. Large-scale Processes in the Atmosphere.* Academic Press, 1983.

[63] I. N. James, *Introduction to circulating atmospheres.* Cambridge University Press, 1994.

[64] R. Seager, N. Harnik, Y. Kushnir, W. Robinson, and J. Miller, "Mechanisms of hemispherically symmetric climate variability," *J. Climate*, vol. 16, pp. 2960–2978, 2013.

[65] S. Broennimann, A. Stickler, T. Griesser, A. M. Grant, T. Ewen, T. Zhou, M. Scharner, E. Rozanov, and T. Peter, "Variability of large-scale atmospheric circulation indices," *Meteorol. Z.*, vol. 18, pp. 379–396, 2009.

[66] J. D. Neelin, D. S. Battisti, A. C. Hirst, F. F. Jin, Y. Wakata, T. Yamagata, and S. E. Zebiak, "Enso theory," *J. Geophys. Res.-Oceans*, vol. 103, pp. 14 261–14 290, 1998.

[67] M. Paluš, D. Hartman, J. Hlinka, and M. Vejmelka, "Discerning connectivity from dynamics in climate networks," *Nonlin. Processes Geophys.*, vol. 18, pp. 751–763, 2011.

[68] B. R. Lintner and J. C. H. Chiang, "Adjustment of remote tropical climate of el niño conditions," *J. Climate*, vol. 20, pp. 2544–2557, 2006.

[69] A. A. Tsonis, G. Wang, K. L. Swanson, F. A. Rodrigues, and L. da Fontura Costa, "Community structure and dynamics in climate networks," *Clim. Dynam.*, vol. 37, pp. 933–940, 2011.

[70] J. F. Donges, I. Petrova, A. Loew, N. Marwan, and J. Kurths, "How complex climate networks complement eigen techniques for the statistical analysis of climatological data," *Clim. Dynam.*, vol. 54, pp. 2407–2424, 2015.

[71] F. Arizmendi and M. Barreiro, "Enso teleconnections in the southern hemisphere: A climate network view," *Chaos*, vol. 27, p. 093109, 2017.

[72] J. Shukla, "Predictability in the midst of chaos: a scientific basis for climate forecasting," *Science*, vol. 282, pp. 728–731, 1998.

[73] B. J. Hoskins and D. J. Karoly, "The steady linear response of a spherical atmosphere to thermal and orographic forcing," *J. Atmos. Sci.*, vol. 38, pp. 1179–1196, 1981.

[74] P. D. Sardesmukh and B. J. Hoskins, "The generation of global rotation flow by steady idealized tropical divergence," *J. Atmos. Sci.*, vol. 45, pp. 1228–1251, 1988.

[75] I. M. Held, S. Lyons, and S. Nigam, "Transients and the extratropical response to el niño," *J. Atmos. Sci.*, vol. 6, pp. 163–174, 1989.

[76] M. Hoerling, A. Kumar, and M. Zhong, "El niño, la niña, and the nonlinearity of their 510 teleconnections," *J. Climate*, vol. 10, pp. 1769–1786, 1997.

[77] H. Lin and J. Derome, "On the modification of the high and low-frequency eddies associated with pna anomaly: An observational study," *Tellus*, vol. 49, 1997.

[78] A. Wu and W. W. Hsieh, "The nonlinear northern hemisphere atmospheric response to enso," *Geophys. Res. Lett.*, vol. 31, p. L02203, 2004.

[79] J. Sheng, J. Derome, and M. Klasa, "The role of transient disturbances in the dynamics of the pacificâĂŞnorth american pattern," *J. Climate*, vol. 11, p. 523âĂŞ536, 1998.

[80] J. Ludescher, A. Gozolchiani, M. I. Bogachev, A. Bunde, S. Havlin, and H. J. Schellnhuber, "Very early warning of next el niño," *Proc. Natl. Acad. Sci. U.S.A.*, vol. 111, 2014.

[81] A. Ramos, Y. Zou, G. Sampaio, J. Kurths, and E. E. N. Macau, "Unveiling non-stationarity coupling between amazon and ocean during recent extreme events," *Clim. Dyn.*, 2017.

[82] J. I. Deza, M. Barreiro, and C. Masoller, "Assessing the direction of climate interactions by means of complex networks and information theoretic tools," *Chaos*, vol. 25, p. 033105, 2015.

[83] M. Wiedermann, A. Radebach, J. F. Donges, J. Kurths, and R. V. Donner, "A climate network based index to discriminate different types of el niño and la niña," *Geophys. Res. Lett.*, vol. 43, pp. 7176–7185, 2016.

[84] N. Boers, R. V. Donner, B. Bookhagen, and J. Kurths, "Complex network analysis helps to identify impacts of the el niño-southern oscillation on moisture divergence in south america," *Clim. Dyn.*, vol. 45, 2015.

[85] S. G. Philander, *El Niño, La Niña and the Southern Oscillation.* Internat. Geophys. Series 46, Academic Press, 1990.

[86] K. C. Mo and J. N. Paegle, "The pacific-south american modes and their downstream effects," *Int. J. Climatology*, vol. 21, pp. 1211–1229, 2001.

[87] F. Arizmendi, M. Barreiro, and C. Masoller, "Identifying large-scale patterns of unpredictability and response to insolation in atmospheric data," *Scientific Reports*, vol. 7, p. 45676, 2017.

[88] G. Tirabassi and C. Masoller, "On the effects of lag-times in networks constructed from similarities of monthly fluctuations of climate fields," *EPL*, vol. 102, p. 59003, 2013.

[89] A. Tantet and H. A. Dijkstra, "An interaction network perspective on the relation between patterns of sea surface temperature variability and global mean surface temperature," *Earth Syst. Dynam.*, pp. 1–14, 2014.

[90] J. Hlinka, D. Hartman, M. Vejmelka, D. Novotna, and M. Palŭs, "Non-linear dependence and teleconnections in climate data: sources, relevance, nonstationarity," *Clim. Dyn.*, vol. 42, pp. 1873–1886, 2014.

[91] I. Fountalis, A. Bracco, and C. Dovrolis, "Enso in cmip5 simulations: network connectivity from the recent past to the twenty-third century," *Clim. Dyn.*, vol. 45, p. 511âĂŞ538, 2015.

[92] G. Tirabassi and C. Masoller, "Unravelling the community structure of the climate system by using lags and symbolic time-series analysis," *Scientific Reports*, vol. 6, p. 29804, 2016.

[93] R. Kistler, E. Kalnay, and W. Collins, "The ncep-ncar 50-year reanalysis: Monthly means cd-rom and documentation," *B. Am. Meteorol. Soc.*, vol. 82, pp. 247–268, 2001.

[94] A. L. Berger, "Long-term variations of daily insolation and quaternary climatic changes."

[95] K. A. McKinnon, A. R. Stine, and P. Huybers, "Spatial structure of the annual cycle in surface temperature: amplitude, phase, and lagrangian history," *J. Climate*, vol. 26, pp. 7852–7862, 2013.

[96] R. G. Tedeschi, I. F. A. Cavalcanti, and A. M. Grimm, "Influences of two types of enso on south american precipitation," *International Journal of Climatology*, vol. 33, pp. 1382–1400, 2013.

[97] R. G. Tedeschi, A. M. Grimm, and I. F. A. Cavalcanti, "Influence of central and east enso on precipitation and its extreme events in south america during austral autumn and winter," *International Journal of Climatology*, vol. 36, pp. 4797–4814, 2013.

[98] H. Kantz and T. Schreiber, *Nonlinear time series analysis*, 2nd ed. New York, USA: Cambridge University Press, 2003.

[99] R. V. Donner, Y. Zou, J. F. Donges, N. Marwan, and J. Kurths, "Recurrence networks: A novel paradigm for nonlinear time series analysis," *New J. Phys.*, vol. 12, p. 033025, 2010.

[100] L. Lacasa and R. Toral, "Description of stochastic and chaotic series using visibility graphs," *Phys. Rev. E*, vol. 82, p. 036120, 2010.

[101] X. Sun, M. Small, Y. Zhao, and X. Xue, "Characterizing system dynamics with a weighted and directed network constructed from time series data," *Chaos*, vol. 24, p. 024402, 2014.

[102] T. A. Schieber, L. Carpi, A. Diaz-Guilera, P. M. Pardalos, C. Masoller, and M. G. Ravetti, "Quantification of network structural dissimilarities," *Nat. Comm.*, vol. 8, p. 13928, 2017.