Disentangling Climatic Interactions and Detecting Tipping Points by Means of Complex Networks

Giulio Tirabassi

Prof Cristina Masoller, thesis director

Terrassa, Spain, 2015

Contents

Abbreviations

AMO	Atlantic Multidecadal Oscillation
AWC	Area Weighted Connectivity
CC	Cross Correlation
CSD	Critical Slowing Down
ENSO	El Niño - Southern Oscillation
EOF	Empirical Orthogonal Function
FPR	False Positive Ratio
GC	Granger Causality
GCE	Granger Causality Estimator
ITCZ	Inter-Tropical Convergence Zone
KLD	Kullback–Leibler Distance
LPF	Local Positive Feedback
MCA	Maximum Covariance Analysis
MI	Mutual Information
MIOP	Mutual Information from Ordinal Patterns
MJO	Madden-Julian Oscillation
NAO	North Atlantic Oscillation
OP	Ordinal Pattern
PCA	Principal Component Analysis
PDO	Pacific Decadal Oscillation
TPR	True Positive Ratio

SACZ	South Atlantic Convergence Zone
SAT	Surface Air Temperature
SATA	Surface Air Temperature Anomalies
SOI	Southern Oscillation Index
SSM	Statistical Similarity Measure
SST	Sea Surface Temperature
WAWC	Weighted Area Weighted Connectivity

Preface

In a scenario of major global climatic changes, the understanding of the Earth system has become in recent years an impelling task of the scientific community. In the last decade, an increased performance of supercomputers as well as a better understanding of the physical processes underlying Earth dynamics (such as cloud formation, moist exchange between soil and atmosphere, and so on) improved dramatically the quality of climate models. Moreover, an increasing satellite coverage allowed the generation of very detailed databases, called reanalysis data, that give with good precision the state of the most important dynamical variables at high spatio-temporal resolution for at least the past 40 years.

This huge amount of data is being used by the climatological scientific community to investigate the nature of the ongoing physical processes in the Earth system. The increasing of such datasets motivates the development of new analysis. In particular, since in first approximation the climatological behaviour of the atmosphere can be described by relatively simply linear models, the non-linear data analysis has been largely overlooked so far.

However, many relevant climatic processes have a strong non-linear component. A paradigmatic example is the so-called El Niño-Southern Oscillation (ENSO), a coupled ocean-atmosphere process that can be sketched in first approximation as a non-linear oscillator with delayed feedback [1].

Another phenomenon of the Earth system in which the non-linearities play a major role is doubtlessly the barotropic polar jet, that is maintained by the non-linear stress produced by its own eddies [2]. In this way a complex system of positive and negative feedbacks is established, influencing the dynamics of the major actor in the synoptical variability (weather) in the high latitudes.

Given these considerations, it is important to study such phenomena with tools coming from the field of complex systems.

In particular, in this thesis we present new data analysis techniques based on information theory and complex networks that take into account the non-linear nature of the climate processes under examination, as well as give a new perspective to climatological data analysis.

Complex networks have emerged in the recent years as a useful and powerful tool to investigate a large variety of phenomena in which it is possible to identify a discrete number of components among which relations can be established. Being this a very general concept, it is not surprising that it found successful application in very different field such as sociology, biology, epidemics, geophysics, and so on [3, 4].

In particular in this thesis we will focus in the construction of what are called *climate networks*, *i. e.* networks in which the nodes are composed by geographical locations on Earth and the links are given by relations among them. We will generally consider links computed from the statistical similarity of the dynamics of a climatological variable available at each geographical location, such as the surface air temperature (SAT). This is a quite general approach that, depending on the definition of the statistical similarity employed, allows to investigate different characteristics of the variable under examination. We will define the statistical similarity using concepts that are adopted from information theory.

1 The Climate System

In this chapter we will introduce the main characteristic of the climate system, focusing on the main sources of variability. The discussion will range from atmospheric to oceanic processes, mainly focusing in the large spatio-temporal scales.

1.1 Climate Variability

The climate system can be understood as the result of complex interactions among different complex subsystems, composed in turn by a numerous amount of smaller components, in a hierarchical cascade towards the micro-physics.

As an example, cloud formation is one of those phenomena in which this feature is clearly evident. A cloud is formed by the aggregation of water, by means of bigger saline particles that act as nucleation centres. The dimension of this components is usually smaller than that of the typical grain of dust. The type and the extent of the resulting clouds will depend on the nature, concentration and distribution of such agents, together with the termodynamical state of the external environment. This mesoscopic characteristic will then impact by means of rain, longwave absorption, and shortwave reflection over a certain number of systems: atmosphere, biosphere, criosphere, and so on. In this way a clear bridge is built, connecting large systems, with molecular ones, in a broad interaction of different scales.

A very schematic – and somehow reductive – review of the most important interactions among the main actors of the climate system is reported in Fig. 1.1. As it can be seen, the relations among the various parts are very often bidirectional, that is, if one subsystem affects another subsystem, with high probability it will undergo a feedback from that second subsystem. The complex feedback web that is created in this way make the analysis of the climate dynamics extremely difficult.

Another important characteristic derivable from Fig. 1.1 is that there are few sub-systems presenting only outgoing fluxes. Those are the external forcings, among which the most important is, doubtlessly, the insolation from the Sun, that is the amount of energy the Earth

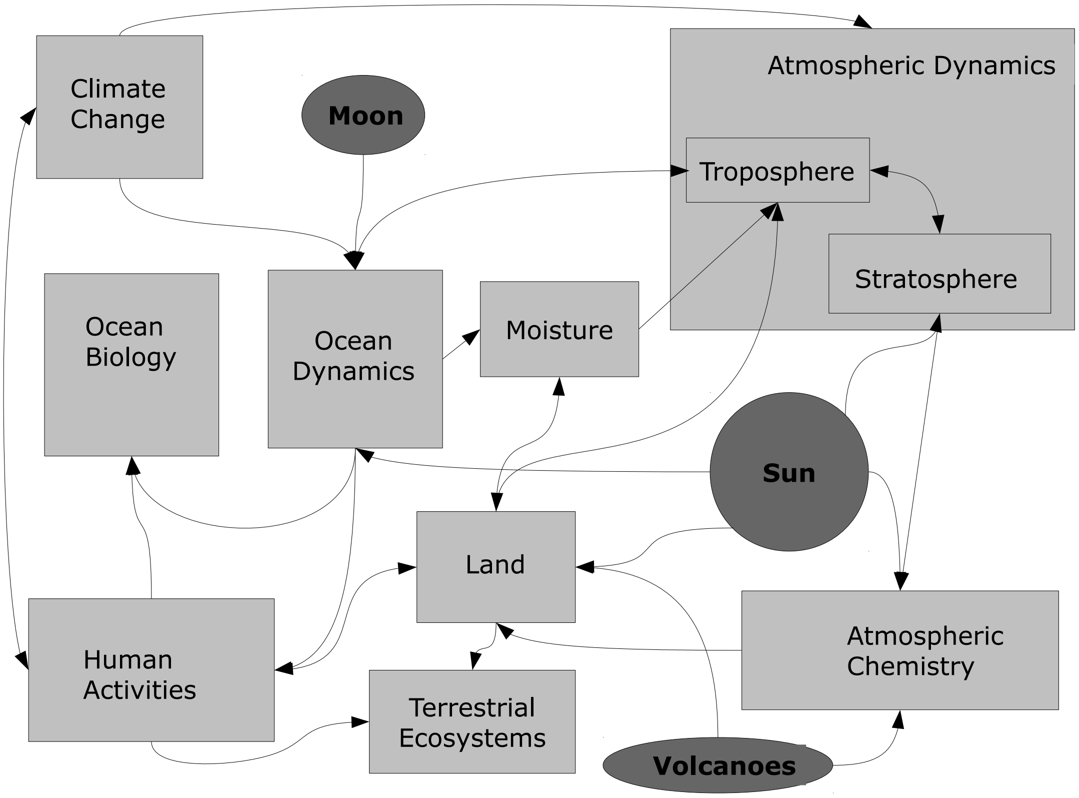


Figure 1.1: Scheme of the main components of the climatic system and of their interactions. The main forcing for the whole system is the Sun, that acts directly over a number of subsystems. The energy then flows to the other subsystems by means of the mutual interactions among them. Note how many of the arrows are double-headed, meaning that the two subsystems can mutually influence each other. Figure adapted from [5].

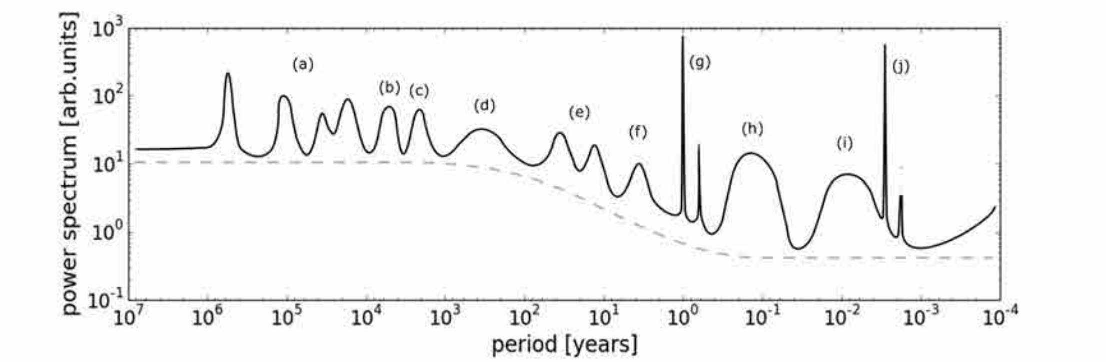


Figure 1.2: Artistic rendering of the energy density spectrum of an idealised climatological time-series. This idealisation is obtained taking into account the various time scales typically displayed by different climatological datasets, with a broad range of sampling frequencies and time-coverage. We can identify various peaks associated with different known processes: (a) Milancović Cycles, (b) Heinrich Events, (c) Dansgaard-Oeschger Cycles, (d) Centennial Variability, (e) Interdecadal Variability, (f) ENSO, (g) Seasonal Cycle and its harmonics, (h) Intraseasonal Variability, (i) Weather, (j) Daily Cycle and its harmonics. Adapted from [6].

receives in form of shortwave radiation.

The energy difference between the incoming solar radiation and the outgoing longwave emissions from the Earth is the main source of the energy flux spreading along the climate system. Only a minimal component is given by tides[1], and even less is the influence exercised by volcanoes.

All the processes sketched in Fig. 1.1 take place over a broad range of time-scales, that can range from seconds (the dissipation scale of the three-dimensional atmospheric turbulence) up to thousands of years (the scale of the changes in Earth's orbit parameters). Among these two extremes we can find a large ensemble of processes and phenomena, as depicted in Fig. 1.2

The most notable are doubtlessly the two peaks at 1 year^{-1} and 1 day^{-1} with their higher order harmonics. These peaks are induced by insolation variability, that, as mentioned before, is the main forcing of the climate system. There are also a whole number of broad-band processes, mainly due to large scale patterns (frequency smaller than 1 year^{-1}) as El Niño - Southern Oscillation (ENSO) [7], Pacific Decadal Oscillation (PDO) [8] or Atlantic Multidecadal Oscillation (AMO) [9] to name a few. Another broad-band peak is placed at the monthly scale, and represents the so called intraseasonal variability, among which is notable the Madden-Julian Oscillation (MJO)[10]. Lastly we can observe another broad range of frequencies centred around 1 week^{-1}, representing the synoptic weather evolution.

---

[1] Despite in the framework of ocean dynamics the tidal injection of momentum in the ocean is not as relevant as the presence of a strong meridional thermal gradient, or of surface winds, it can play a major role in problems such as ocean mixing, in which it is roughly as important as the wind stress forcing.

It is also interesting to note that, between the tame scale of 10^5 years and 1 month, the spectrum presents a skewed continuous base, namely a red background. The red background is an ubiquitous feature of climatological time-series [11, 12], and for a long period it remained unexplained. With the understanding of this behaviour it comes one of the first "climate model" in history: the Hasselmann Model [13].

1.2 Red Spectra in Climate Time-Series

One of the most prominent features of geophysical time-series in a climatological time-scale is that very often they display a red spectrum, that is, a spectrum in which the energy carried by a certain frequency is a decreasing function of the frequency itself.

This feature, detected in time-scales usually ranging from 1 month to 10^5 years, has been found in different datasets, such as meteorological and proxy data, including ice cores, sediments, ice volumes, etc. [12]

An explanation of such behaviour was missing until 1976, when K. Hasselmann described it in a very simple and straightforward way in his seminal paper [13].

Hesselmann proposed to divide the climatological processes in two components, a fast one, mainly atmospheric, acting as a forcing over a slower one, the real climate component (oceans, glaciers, etc.). As an example let us suppose that the main part of the slow variables is given by ocean temperature, T, and the fast one by atmospheric forcing at the ocean boundary, W. We can write

$$\dot{T} = -\, C\, (T - \bar{T}) + W, \tag{1.1}$$

where C models the thermal inertia of the oceans and the $\bar{T}$ is the average temperature.

Hasselmann though to model the atmospheric forcing as simple white noise. This way Eq.1.1 can be transformed in a stochastic differential equation:

$$\mathrm{d}\theta = -\, C\, \theta\, \mathrm{d}t + \sigma\, \mathrm{d}w, \tag{1.2}$$

where $\theta = T - \bar{T}$, σ is the white noise variance and w is a Wiener process. From Eq. 1.2 we can see that the variable θ, modelling the ocean, is integrating in its evolution the white noise given by the atmosphere. These kind of processes are called Ornstein-Uhlenbeck processes, and display a red spectrum [12] In this way we are able to understand, with almost no special assumption, such general characteristic of climate data.

As it will be discussed in the next chapter, the presence of a red background in a time-series implies the presence of a certain amount of autocorrelation. This in turn affect the confidence intervals of measures such as the cross correlation and the mutual information (cfr. Chapter 2), because it reduces the number of degrees of freedom of the time-series.

1.3 Large-Scale Patterns

It is known since a long time that the Earth climate exhibits large scale correlation patterns. As an example the Norse, due to their fine knowledge of the north Atlantic basin, already noted in the 12th century long range correlations between weather in Greenland and the continental Europe [14]. This correlation was just part of a bigger pattern identified by sir Walker in the first part of the 20th as a see-saw of the pressure difference between Iceland and Azores, that is, a dipolar structure of low pressure over Iceland and high pressure over the Azores modulated in time. This pattern is called North Atlantic Oscillation (NAO) [15] and it is a mainly atmospheric pattern that impacts greatly the weather conditions over Europe, Northern Asia and North America.

1.3.1 NAO

Given the area of impact, the socio-economical effects of NAO are huge and a lot of effort has been devoted to its characterisation. Several methods have been proposed to describe the oscillation with a single time series. One is the aforementioned (normalized) pressure difference between two locations in Iceland and Azores, another involves the use of empirical orthogonal functions (EOF, cfr sec. 2.2.4), and generally all these methods give highly correlated results [16].

It is referenced as positive NAO phase a situation in which the pressure difference between Azores and Iceland is higher than normal, and negative phase a situation in which the opposite is true. When we deal with a positive NAO phase the higher than normal meridional pressure gradient moves the barotropic polar jet northwards. Since along the jets we have the passage of the storm systems (storm-track), it means that the weather perturbations will move eastwards at latitudes higher than normal, at least in Europe. Thus northern Europe will be wetter than normal, and southern Europe dryer. The opposite holds for the negative NAO phase, in which is southern Europe to be wetter than normal.

The prominent feature of NAO is its high variability. In Fig. 1.3 we report the time-series of daily and monthly NAO indices. The series can fluctuate among positive and negative extremes within few days, and there is no sign of clear periodicity or of a preferential time-scale.

This is also evident from the spectrum, depicted in Fig. 1.4, that is fairly flat, resembling that of white noise.

The main reason for such high variability is that NAO is mainly an internal atmospheric mode. This means that the high level of turbulence present in the high latitude atmosphere affects its

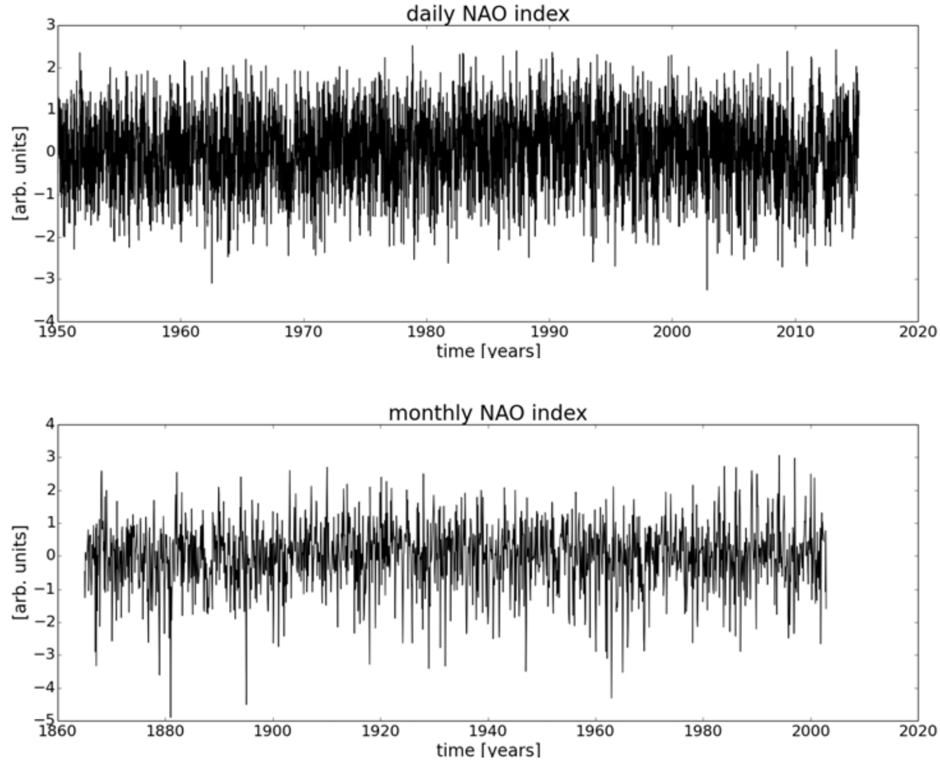


Figure 1.3: (**top**): daily NAO index time-series computed from principal component analysis. (**bottom**): monthly NAO index time-series computed as the normalized time-difference between pressure in Azores and Stykkishólmur (Iceland). Data from [climatexplorer]

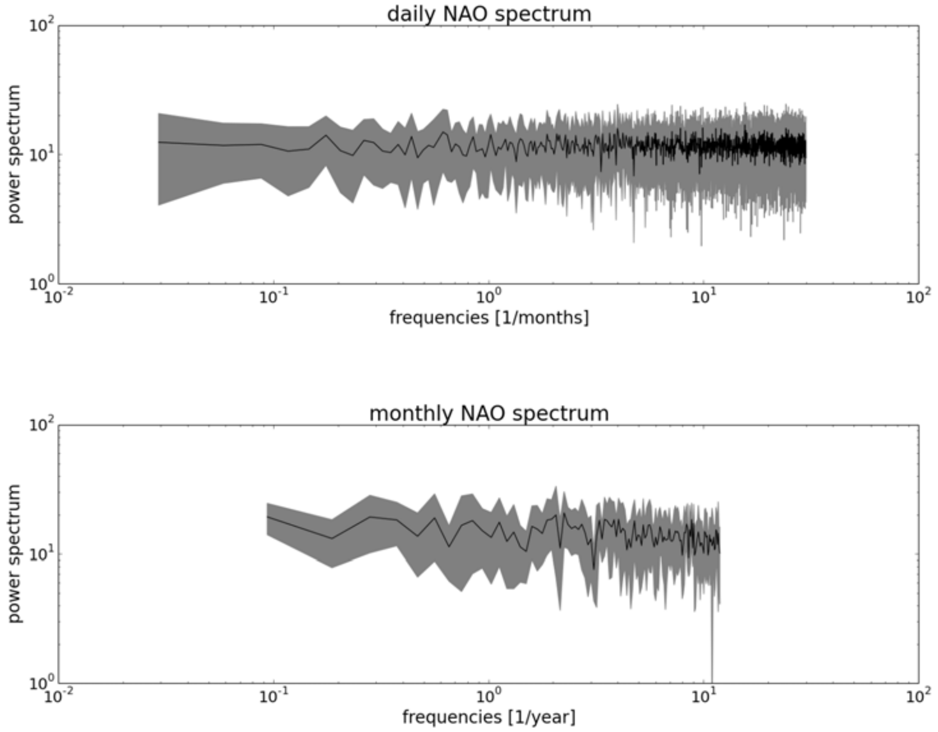


Figure 1.4: (**top**): daily NAO index spectrum computed from the time-series depicted in Fig. 1.3. (**bottom**): monthly NAO index spectrum computed from the time-series depicted in Fig. 1.3. In gray is shaded the standard deviation of the spectrum. As it can be seen, hese spectra are almost flat, meaning that there is not a preferential time-scale for the evolution of the time-series.

evolution, resulting in an erratic behaviour. In fact, it is very hard to predict NAO on a long time-scale, while it is possible to predict its behaviour within a 2 weeks window using weather forecast models.

1.3.2 ENSO

Another pattern known since a long time takes place mainly over Eastern Pacific Ocean and is called El Niño - Southern Oscillation (ENSO). Peruvian fishermen already noted that recurrently, around Christmas[2], superficial waters of the Eastern Pacific could be much warmer than usual. Indeed a sea surface temperature (SST) anomaly during an El Niño event can easily reach 5 °C. Such strong variation over a very large basin as the Eastern Pacific has an enormous impact over many regions of the world, especially the Pacific ones. In particular, strong fluctuations of the SST along the equator, a zone in which the atmosphere is very sensitive to the ocean's temperature, greatly affect the atmospheric circulation patterns, that in turn modify the average distribution and intensity of rains and droughts over the tropical and sub-tropical Pacific. These changes can be so intense to affect significantly human activities, such as agriculture, that strongly depends on rainfall. Moreover, such big anomalies of the sea temperature cause fish migrations, with subsequent impact over fishery. Indeed ENSO is the Earth system phenomena that most affects world economy, so it is not surprising the huge effort put in the last 30 years in the understanding of its mechanisms.

ENSO is a coupled ocean-atmosphere phenomenon, in which both systems are influenced and influence each other by means of feedbacks.

The climatological state over Eastern Pacific Ocean see the continuous action of easterly trade winds that push equatorial waters from the Peruvian coasts towards the western limits of the basin. This causes both an increase of sea level in the western Pacific boundary and a cold water upwelling in the eastern one, the latter being also favoured by the tendency of the ocean to adjust its surface level. The net result is the formation of two distinct areas (Fig. 1.5), one of relatively cold waters in the east (called *cold tongue*) and one of relatively hot waters in the west (called *warm pool*) [17].

When the easterly trade winds change, the pressure difference among the two extremes of the basin and the eastern upwelling are affected. In particular, a weakening of the trade winds result in a warmer than normal cold tongue area (since there is less cold water coming from the upwelling) and a warm pool shifted to the east (because less superficial warm water is pushed westward). If the warming of the cold tongue is above a certain threshold, the resulting configuration is called El-Niño (Fig. 1.6, top). Conversely, when the trade winds are stronger than normal, the cold tongue as well will be more intense. If the cooling of the cold tongue is below a certain value the configuration is called La Niña (Fig. 1.6, bottom). However this threshold is not univocally defined, for example the U.S NOAA definition of El-Niño is a 3 month average warming of at least 0.5 degrees, while other organizations prefere other

---

[2] That is where the name "El-Niño" comes from, meaning, in Spanish, "baby born".

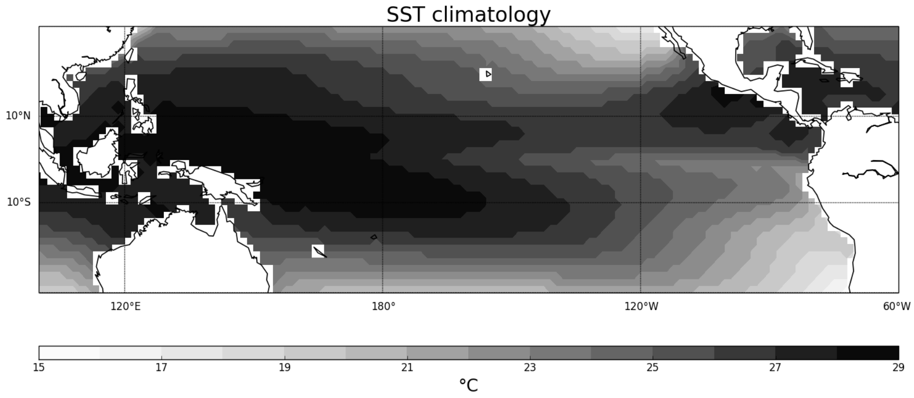


Figure 1.5: SST mean state in the tropical Pacific Ocean. The cold tongue on the right and the warm pool on the left are clearly distinguishable. Data from ERA Interim Reanalysis.

definitions.

Since the changes in SST distributions given by the winds affect the heat exchange with the atmosphere, the air circulation will be affected as well, closing this way a feedback loop from the atmosphere to the ocean and then back to the atmosphere.

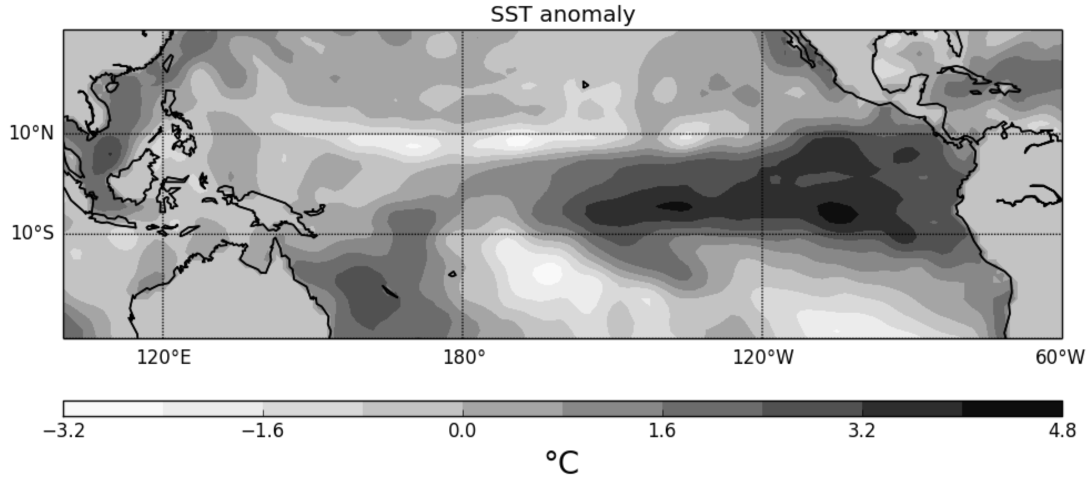
There are several ways to quantify these fluctuations with only one index time-series. One of the most common is an oceanic index called NINO3, defined as the average SST anomaly in the area [5S,5N] × [150W,90W]. Another common index is an atmospheric one, and it is called Southern Oscillation Index (SOI) defined as the normalised pressure difference between the island of Tahiti (central Pacific) and the island of Darwin (eastern Pacific). This pressure difference allows to quantify changes in the easterly trade-winds, and it is strongly anticorrelated with NINO3 index. For this reason the whole phenomenon takes the joint name of El Niño-Southern Oscillation.

Unlike the NAO, ENSO displays a short certain of preferential time-scales, ranging roughly from 1 to 10 years with significative periodicity of about 3-4 years (cfr. Fig. 2.2).

The ocean can be divided in two main layers, upper and lower ocean, that show almost constant salinity and temperature, and are separated by a region, called *thermocline* in which the two quantities change abruptly[3]. The position of the thermocline is not constant and varies in time and space. In particular in the pacific region, the height of the thermocline changes because of sea level adjustments and because of the upwelling intensity, and thus partecipates to the evolution of ENSO. It is clear that the higher the thermocline will be, the

---

[3] The reasons for which this happens are many and complex, and they are beyond the scope of this chapter. For an introduction to the thermocline formation theory see, for example, [2].



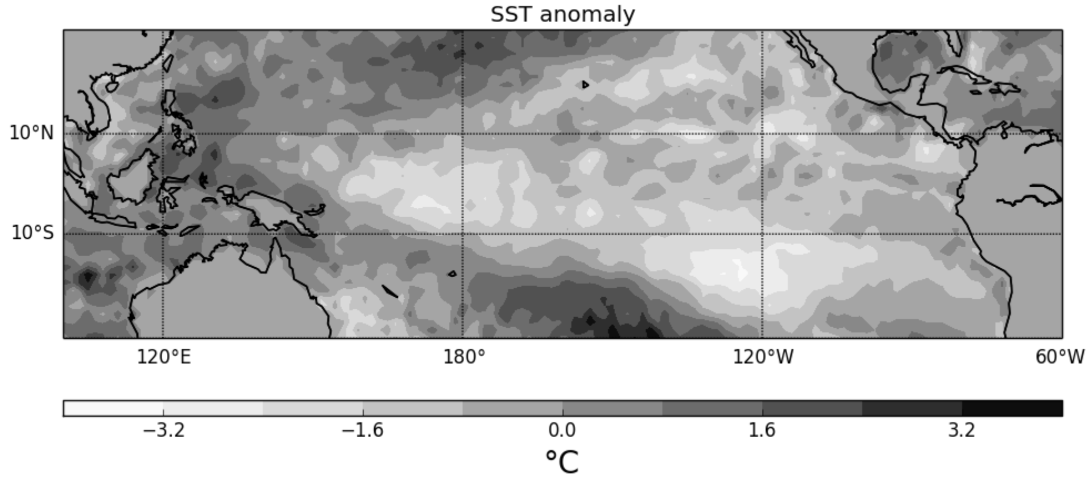


Figure 1.6: Sea surface temperature anomalies during a strong El Niño event (top) and during a strong La Niña event (bottom). Data from ERA Interim Ranalysis.

colder the superficial temperature of the ocean will be, and vice versa.

When the wind stress changes, the ocean reacts, to adjust to the new forcing, by means of waves propagation, waves that are amplified at the thermocline due to the reduced gravity effect [2]. We can distinguish two main type of waves allowed by equatorial ocean dynamics. The first ones are the equatorial *Kelvin waves*. They ride eastward over the equator and have a meridional width of around 300 km far from it [17]. They are quite fast and can cross the Pacific in about $\delta_{\mathrm{K}} = 3$ months. The second ones of wave are *Rossby waves* travelling westward with an off-equator maximum. The longest are also the fastest, and can cross the Pacific basin in $\delta_{\mathrm{R}} = 9$ months. When one of this two types of waves meets the basin boundaries (that is the coasts) it is partially reflected and transformed into the other type. When it comes back to the place of origin it will carry the information of the state of the thermocline of months before, providing a memory effect. In this way not only the ocean must adjust to equilibrate the new wind stress, but it has also to take into account its previous state. Since Rossby waves and Kelvin waves have the maximum meridionally out of phase, the reflected wave will be roughly in zonal quadrature with the incident one. Thus, when a deeper than normal thermocline wave is reflected, the new backpropagating wave will produce a shallower than normal thermocline.

All those features (complex ocean-wind feedbacks, memory effects, etc.) once put together give rise to an incredibly complex system, and it is not surprising that we still lack an exhaustive predictive model of ENSO. However, many conceptual models able to reproduce its main characteristic have been proposed. An interesting conceptual model is the so-called delayed oscillator theory [18]. In this model the ocean-atmosphere interaction and the memory induced by the wave reflection at the boundaries are the main ingredients.

1.3.3 Steady Atmospheric Patterns

The NAO and the SOI are just a couple of the many long range correlations that can be found in the atmosphere. These patterns take the name of *teleconnections*, and they are an important component of the climate variability [19].

Teleconnections usually fluctuate over a broad range of scales, generally interdecadal time-scales, and form patterns that are extended over large portions of the Earth's surface. They are usually detected through analysis of the covariance matrix of a climatic variable over a certain area, and thus they are often closely related with EOF technique (see Sec. 2.2.4). The fact that their detection is in general performed with a linear analysis is not a case. Indeed many of these patterns can be explained by means of linear theory in first approximation.

In fact, calculating numerically the steady state for the anomalous circulation of the atmosphere in a linear model, allows one to compare it to the observed large-scale circulation. In [20] we have an example of such calculation, showing very good agreement between the linear response of the model and the observations.This suggests that the atmosphere is almost a

Figure 1.7: Image taken by GOES 12 satellite of the equatorial South America. The ITCZ and the SACZ are clearly visible respectively as a white band along the equator and a diffuse cloudiness between the southern Amazon forest and the subtropical Southern Atlantic ocean.

linear system on the climatological time-scale, with spatial stationary patterns given by the resonance of the circulation with the topographic forces and heat sources. In this framework the teleconnections are fluctuation over various time scales of the amplitude of these stationary waves, given by changes in the forcing intensities.

However, it is clear that for a more precise analysis, non linear effects should be taken into account. Especially in the extratropics, in fact, where the atmosphere is more active and the eddy forcing is more intense, the non-linearities would play a more important role.

1.4 Large Scale Convective Patterns: ITCZ and SACZ

Along the equator, on the region of maximum insolation, it is located a zone of deep convection named Inter-Tropical Convergence Zone (ITCZ). This pattern encircles the Earth and swings northward and southward following the interface between winter and summer Hadley cells. This swings are more intense over land, which responds to insolation changes more rapidly than the ocean.

Along the main path of the ITCZ numerous convective subsystems are located. The Monsoon systems (that is, systems in which the convection is triggered by a strong air-sea contrast) in India and northern South America are examples, as well as the South Pacific Convergence Zone in the south-western Pacific ocean.

A particularly relevant branch of the ITCZ is the so-called South Atlantic Convergence Zone (SACZ) [21] which extends between southern-central Brazil and the subtropical Atlantic Ocean.

The South Atlantic Convergence Zone (SACZ) is one of the main components of the South American Monsoon: it is a convective pattern that extends from the Amazon forest to the subtropical South Atlantic ocean, oriented in northwest-southeast direction [22, 23, 24]. When active, the SACZ is associated with heavy precipitation over the Amazon forest and southeastern Brazil, causing floods and landslides over the densely populated areas of Minas Gerais, Sao Paulo and Rio de Janeiro.

The main mode of SACZ variability consists of a precipitation dipole with centres over the SACZ and over Uruguay, such that when the SACZ is active there is decreased rainfall over Uruguay and vice versa. This mode varies on several time scales ranging from day-to-day due to the passage of fronts, to intraseasonal, interannual and even longer time scales having been associated with the observed summertime rainfall trend in eastern South America during the 20$^{\text{th}}$ century [25, 26, 27, 28].

Due to its socio-economic importance there have been several attempts to improve our understanding and predictability of rainfall over the SACZ on seasonal to interannual time scales, particularly focusing on the role of upper ocean in modulating its behaviour. Studies have shown that the subtropical South Atlantic Ocean (SAO) may influence the evolution of the SACZ [29, 27, 30]. For example, [27, 30] show that even though the region is dominated by internal atmospheric variability, SST anomalies can force a dipole of precipitation anomalies located mainly over the oceanic portion of the SACZ. Subsequent studies have suggested that the air-sea interaction is such that an initially stronger SACZ – due to internal atmospheric variability – induces an oceanic cooling that in turn negatively affects the convective precipitation, resulting in a negative feedback loop [31, 32].

However, air-sea interaction has been difficult to quantify both in observations and in model simulations and to date it is unclear how the circulation associated with the SACZ is influenced by surface ocean conditions. Disentangling air-sea interaction in the subtropics is a challenging task, and up to now to the best of our knowledge, no method has allowed a robust identification of this interaction in observational data. To tackle this problem, we propose a new methodology based on the combined use of Granger Causality (GC) together with a new measure from climate network theory (Area Weighted Connectivity) and Maximum Covariance analysis; this methodology aims to assess the presence of air-sea interaction, and to distinguish among different interaction regimes.

In Chap. 6 we will analyse this problem in detail, taking into account this double nature of the SACZ variability, by means of climate networks (cfr Sec.2.4) and Granger causality (GC) analysis (cfr Sec. 2.3.1). We will show that GC is capable of disentangle air-sea interactions in this area, classifying the SACZ years in four distinct categories depending by the nature of the interaction.

In fact we find that for the leading SACZ mode of variability the air-sea coupling is significantly active only in 50% of the cases, and that, when active, it manifests itself in three distinct "flavours": there are years in which the forcing is mainly directed from the atmosphere to the ocean, years in which the ocean forces the atmosphere and years in which the influence is mutual. Moreover, we find that the conditions in the upper ocean can modulate rainfall events in the SACZ, affecting the position, the intensity and the persistence of the precipitation anomalies.

1.5 Tipping Points

It might happen that changes in the climate system could lead to an abrupt modification of the climatic mean state, causing this way a climatic shift. From the point of view of the dynamical systems theory we say that the climate system undergoes a bifurcation; this kind of phenomena take the very general name of *tipping points* [33]. The main ingredient of a tipping point is a multistable system, in which a certain order parameters can make the system collapse from a stable state to another. We can find multistability examples in many components of the climate system, such as oceanic, atmospheric and vegetation dynamics among the others.

1.5.1 Moisture - Vegetation Feedback

Here we present an example of a tipping point in a land model that we will study more in detail in Chapter 7.

In the subtropical regions, such for example the Sahel savannah in subsaharian Africa, the amount of vegetation experiences strong variations, with consequent disastrous effects for the human and animal populations living in those areas. This strong fluctuations can be explained by means of a multistable model similar to the one described before. In particular we can consider the fact that the presence of vegetation is favourished by soil moisture (positive feedback), and contrasted in a logistic way by competition for the resources. In turn moisture, feeded by rain, is deplected by runoff and by the consumption from the plants (negative feedback). This two opposite feedbacks are the ingredients for the following multistable model [34]

$$\dot{w} = R - w - \Lambda w B, \tag{1.3a}$$

$$\dot{B} = rB(w - B) - \mu \frac{B}{B + B_0} \tag{1.3b}$$

where w and B are respectively the moisture and the vegetation amount and Λ, r, μ and B_0 are real constants; R is the order parameter and represents the amount of rainfall that feeds the soil moisture.

This model present two tipping points, the main one is given by a saddle node that, for R low enough, cause a collapse of the system in the desertic state $B = 0$, as sketched in Fig. 1.8.

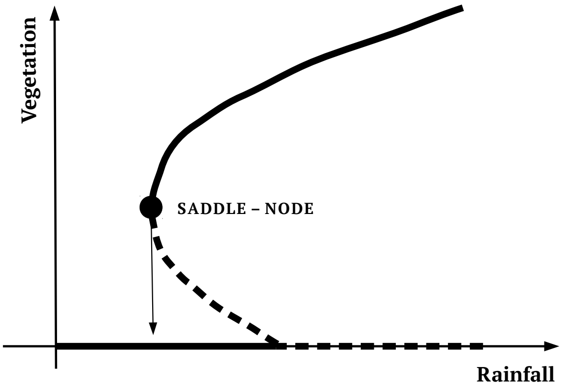


Figure 1.8: Saddle-node bifurcation diagram for the system 1.3. If the systems is on the vegetated stable branch (solid line), when the rainfall is decreased down to certain value, it undergoes a saddle node bifurcation, causing the collapse of the vegetation into a desert state.

Since a saddle node usually implies the presence of hysteresis, to recover the vegetated state the system will need more rain than the one at which the first transition occurred.

1.5.2 Early Warnings of Tipping Point in Vegetation Models

There are now numerous indications, from observations and models, that transitions between different vegetation states can occur due to the existence of multiple equilibrium states [35, 36, 37, 38]. One of these transitions is between vegetated and bare soil states in semi-arid regions and is associated with desertification. Other transitions are, for example, those between savannah and forest states as found in observations of woody cover [39] and above ground biomass [40]. In the latter case, the equilibrium states appear as different maxima in the probability density function (PDF) of these quantities. Although transient growth effects in the relatively short length of the time series obscure the interpretation of the PDFs, it is plausible that multiple equilibrium states exist [40].

The existence of multiple equilibria is supported by models of vegetation dynamics of various levels of complexity. These models can be mainly classified into spatial (PDEs) and non-spatial (ODEs) models, and both types can display multiple equilibria and catastrophic transitions to desertification. Looking at the spatial models, most of them are of reaction-diffusion type, contain positive feedbacks at different scales, represent the dissipation mechanism as diffusion, and use precipitation as stressor. Analysis of three basic spatial models which describe desertification due to decreasing precipitation [41] indicates that the route to desertification occurs through a saddle-node bifurcation which gives rise to hysteresis.

From a land management point of view, these theoretical results suggest that if transitions to desert states occur in a sudden and unexpected way, it would be crucial to develop early-warning signals in order to prevent or at least prepare for such transitions. Previous studies [42] proposed that such indicators could be developed based on critical slowing down (CSD), that is, the slow recovery of a system to small perturbations which appears to be a generic phenomenon in the vicinity of bifurcation points [43]. However, CSD can only be detected indirectly by specific statistical properties of the dynamics of a system such as an increase in spatial and temporal correlation as well as variance [42, 44, 45]. These CSD-based statistical properties have been suggested to act as early-warning signals for critical transitions [42], and they have been experimentally demonstrated to exist in various living systems [46, 47, 48].

1.6 Objectives and thesis organisation

As we have seen in many components of the climate system the relevant effects are due to the presence of nonlinear interactions, that enrich the dynamics and increase its complexity.

A careful nonlinear analysis of the available dataset is thus an important aspect the climatological community should focus on, and the main goal of this thesis is to exploit nonlinear tools of data analysis to yield insight into specific aspects of the climate system.

The thesis is structured as follows:

In the first chapter we will introduce the main characteristic of the climate system as well as the traditional and most employed techniques of climate data analysis. At the end of the chapter we will also present the basis of network theory that will be needed in the subsequent chapters.

In Chapter 3 we will focus mainly in the problem of constructing a climate network that could take into account the lag-times between the SAT time-series.

In Chapter 4 we will analyse the validity of the construction method applying it to a system of oscillators with controlled interactions.

In Chapter 5 we will present two new methods to identify community patterns over the Earth that are compatible with the major characteristic of the climate system.

In Chapter 6 we will study the directionality of climate interactions, specifically investigating the direction of air-sea interaction in subtropical South Atlantic through Granger Causality.

The results presented in Chapter 7 are not directly related to climate networks, but they are an application of complex networks in the framework of catastrophic transitions and tipping points, to construct indicators of vegetation collapse.

Finally, in Chapter 8 we will resume the main results achieved and discuss open problems for future work.

2 Climate Datasets and Methods of Data Analysis

In this chapter we will review the main sources of climate data as well as the standard linear and nonlinear techniques used to analyse them.

2.1 Climatic Data

The first step to understand the climate system is obviously to observe it. By means of measurements we can built datasets that, through a careful analysis, can uncover the features of the system producing that data.

Climatic data can be divided in three families: *direct data*, obtained from instrumental measurements, *proxies*, that is indirect data obtained by a whole number of different sources called "archives", and *model data*, coming from models simulations.

2.1.1 Models and Reanalysis

In this chapter as well as through the all thesis we will deal with only with models and reanalysis data.

Models are very useful to integrate the information of directly measured data. For example, a satellite observing the temperature of the Earth's surface cannot measure at the same time the temperature of the whole world. It can at most span continuously in time an area of a certain width. The temperature in the other regions has to be inferred in some way using the information provided by the measured data. This procedure is called *data assimilation*, and it can be done in several ways, by means of different statistical techniques, such as Bayesian data assimilation [49].

The data produced in this way is a mixture of observation and synthetic data. In particular the synthetic part is chosen in a way that displays the best possible accordance with the measured one. These datasets are called *reanalysis data*, and we will use them extensively in this thesis.

The two mostly used reanalysis datasets are those produced by NCEP/NCAR reanalysis project [50], that is available at a resolution of 2.5° in space and daily in time, and ERA Interim reanalysis [51], which is available with a resolution down to 0.25° in space and daily in time.

The NCEP/NCAR reanalysis starts in 1948 while the ERA Interim one starts in 1979. However the latter is derived from a more modern and complete model, moreover, using only recent data (the database starts at the beginning of the satellite era) we can consider it a bit more reliable.

In Chapters 3 and 5 we will use NCEP/NCAR reanalysis products, while we will use ERA Interim for Chapter 6.

To develop the analysis of the databases we will separate the raw data in a seasonal part, that takes into account the harmonics of the seasonal cycle, and the anomalies respect to it. Supposing to have a monthly sampled database, the procedure to compute the seasonal part is the following: for each month the average of the field is computed separately, obtaining this way a field with only twelve samples. The anomalies are obtaining subtracting this field to the raw one. There are other ways to compute the seasonal cycle – for example using EOF decomposition (cfr. Section 2.2.4) or Fourier spectral analysis – but we will note use them in these thesis.

2.2 Climate Data Analysis Tools

Once we have the data in terms of time-series representing the evolution of a climatic variable in a specific geographical location, we next present the tools to analyse the data to extract the relevant information about the processes we want to characterize.

In general, to asses the underlying linear climate properties of a time-series is often the first step of the analysis. In fact, its linear features and properties make easier to understand the outcomes respect to more advanced non-linear analysis. In this section we will briefly review the most common techniques used to extract the linear properties of time-series.

2.2.1 Cross Correlation

The Cross Correlation is a measure of the linear interdependence between two time series.

Given two time-series X_t and Y_t synchronously sampled, the cross correlation is defined as

$$\rho(\tau) = \frac{\mathrm{cov}(X(t), Y(t+\tau))}{\sqrt{\sigma_X^2 \sigma_Y^2}}, \tag{2.1}$$

where $\mathrm{cov}(X, Y)$ is the covariance of the two series, τ a certain lag-time among them, and

σ^2 their variance. This expression is quite general, and can be applied also to continuous series. From the definition of the covariance ρ is bounded between 1 and -1. In particular 1 (-1) means perfect linear correlation (anticorrelation), that implies $X_t \propto Y_t$. Instead, if $\rho = 0$, the two series are linearly uncorrelated.

For two time-series of finite length N, the Pearson coefficient can be estimated by the unbiased quantity

$$r = \frac{\sum_{t=1}^{N-\tau}(X_t - \bar{X})(Y_{t+\tau} - \bar{Y})}{\sqrt{\sum_{t=1}^{N-\tau}(X_t - \bar{X})^2}\sqrt{\sum_{t=1}^{N-\tau}(Y_{t+\tau} - \bar{Y})^2}}, \tag{2.2}$$

where $\bar{X}$ and $\bar{Y}$ represent the average value of the two series.

Lagged cross correlation is sometimes used also for assessing a direction of interaction between two systems, depending on the sign of τ for which ρ has a local maximum [52, 53, 54].

However more fine instruments for detecting directionality among two processes can be used, as we will see in Sec. 2.3.

2.2.2 Autocorrelation

The autocorrelation is defined as the cross correlation between a time-series and itself with a certain time lag. Namely, given a time-series $X(t) = \{X(t_0), X(t_1), ..., X(t_N)\}$ normalized to zero mean and unitary variance we can write this as

$$c(\tau) = \langle X(t)X(t+\tau)\rangle = \frac{1}{N-\tau}\sum_{i=0}^{N-\tau} X(t_i)X(t_{i+\tau}). \tag{2.3}$$

The autocorrelation of a time-series gives useful information about the "memory" of the processes that generated it. For climate data, if the series has not a specific periodicity, the auto-correlation drops to zero exponentially with increasing τ. Writing

$$c(\tau) \sim \exp\left\{-\frac{\tau}{\tau_d}\right\} \tag{2.4}$$

we define τ_d as the decorrelation time of the time-series.

If the time-series under examination presents a preferred frequency of oscillation and the sampling interval is small enough to sample it, the autocorrelation will present an oscillating behaviour enveloped in an exponential decay.

In Fig. 2.1 we present two examples of autocorrelation functions for two timeseries of zonal wind at 850 hPa anomalies at two latitudes. The solid line represents the autocorrelation of an equatorial time-series, while the dashed one is the autocorrelation function of an mid-latitudes one. In the inset the two time-series are also reported. In the extra-tropics the

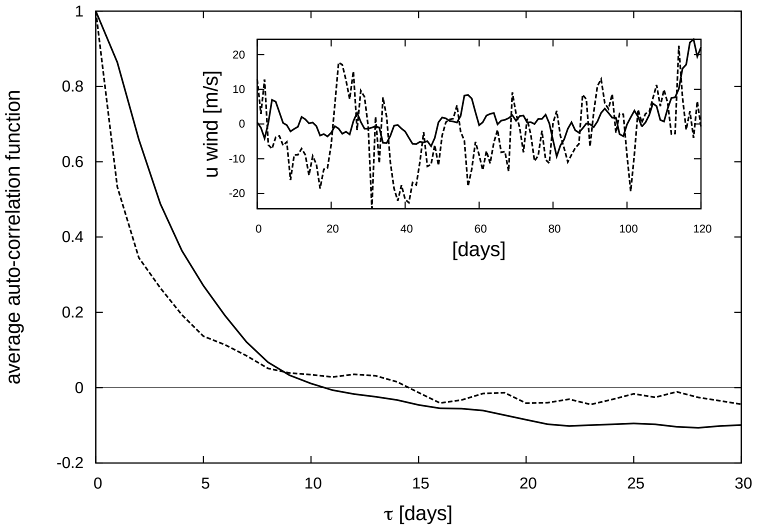


Figure 2.1: Autocorrelation function of two time series (in the inset) of zonal wind at 850 hPa anomalies at the equator (solid) and at mid-latitudes (dashed). In mid-latitudes, where the atmosphere is more turbulent, the fluctuations of the time series are faster and more intense. This is reflected by the autocorrelation, that drops to zero quicker than the equatorial one.

atmosphere is more energetic and turbulent respect to the tropics, which are dominated by very regular trade winds. These features are reflected by the aspect of the two series, one being much more noisy than the other. This in turn reflects into the autocorrelation functions, with the tropical one having a slower decay (longer decorrelation time) than the extra-tropical one.

The autocorrelation of a series affects the number of degrees of freedom of the series itself. In fact, a highly autocorrelated series will have less degree of freedom of a poorly autocorrelated one. This in turns has consequences when determining the significance of a cross-correlation value. Given a cross-correlation r one can construct the t-Student variable

$$t = \frac{r}{\sqrt{1-r^2}}\sqrt{N_{\text{degrees of freedom}}} \tag{2.5}$$

to test the null hypothesis $r = 0$. In this way the higher the autocorrelations the stricter will be the test.

2.2.3 Spectral analysis

We have seen that the autocorrelation displays the variability in the time-series. A well known relationship links the autocorrelation of a signal and its power spectrum; in particular this link is a Fourier transform, a result that is known as Wiener-Khinchin theorem [55].

To investigate in detail the spectral properties of a time-series the Fourier decomposition is the most natural choice. In the case of a discrete signal the Fourier power spectral density is usually called *periodogram* and is computed via a simple Fast Fourier Transform algorithm (FFT) [55].

We can write

$$X(t) = \sum_{j=0}^{N} P(\omega_j) e^{i\omega_j t} \qquad \text{with} \quad \omega_j = \frac{2\pi j}{T} \tag{2.6}$$

where $|P|^2$ is the periodogram and N is fixed by the sampling resolution.

The periodogram gives us useful information about the time-scales involved in the processes we are studying. In fact, given a spectrum $P(\omega)$, the energy carried by the time-series at the ω frequency is defined as

$$E(\omega) = |P(\omega)|^2. \tag{2.7}$$

As in all statistical analysis is good to provide a way to check the significance of the obtained result. Usually climatological time-series display a *red spectrum*, that is a power spectrum in which the largest fraction of the energy is kept by the slow time-scales (low frequencies) [12]. This feature is well explained by the Hasselmann model presented in Chap. 1. For this reason, once computed a spectrum of a certain climatic variable, one should at least test it against the red spectrum given by an autoregressive process of order 1 with same mean, variance and autocorrelation at lag 1 of the original time-series. In this way, eventual peaks at low frequencies could be explained as just a product of integrated noise, not representing real cycles. In fig. 2.2 we report a practical example. The spectrum of a time series of NINO3.4[1] (solid line) is reported together with the 5% highest and lowest percentiles (dashed lines) of the power spectrum distribution of 100 red noise processes with same autocorrelation. As we can see only an interval of frequencies is bigger than the 5% biggest red spectrum, ranging roughly from 1 to 10 years with significative peaks around 3-4 years (cfr. Chapter 1).

2.2.4 Empirical Orthogonal Functions

If, instead of a single time-series, the data to analyse is made by a whole field, one might be interested in the identification of the spatial patterns that the data display, and among them, to select those that carry the greater amount of variance of the field. To do so we can use the Empirical Orthogonal Functions (EOF) technique, also called Principal Component Analysis (PCA).

---

[1] the average SST anomaly in the area [5S,5N] × [170W,120W]

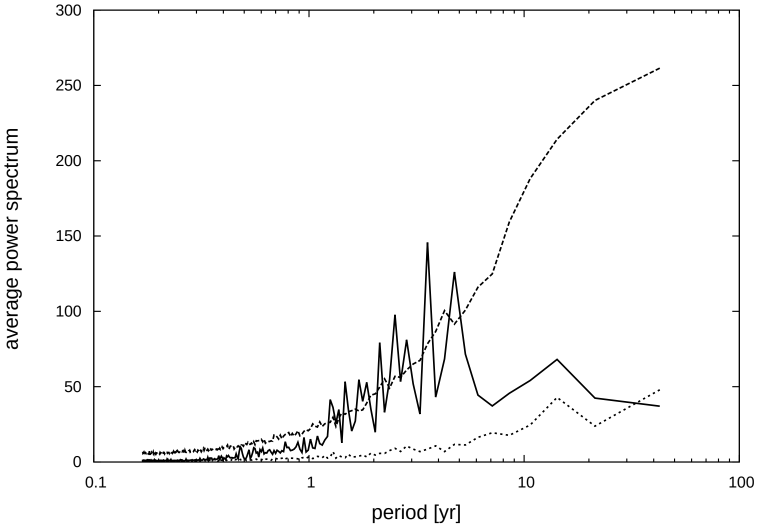


Figure 2.2: In solid line: power spectrum of NINO3.4 monthly time-series. In dashed lines: 5% highest and lowest percentiles of the power spectrum distribution of 100 noise processes with same lag-1 autocorrelation and variance of the NINO3.4 time-series.

Let us suppose to have a database of synchronous time-series X_t^i recorded in N different geographical location (as we will see, the spatial sampling does not need to be regular). If the X^i are normalised to 0 mean and unitary variance we can write their covariance matrix as

$$\mathbf{C} = \mathbf{X}^T\mathbf{X}. \tag{2.8}$$

Since **C** is symmetric the eigenvectors form an orthogonal basis on which we can project the data matrix **X**:

$$X_i^t = \sum_{j=1}^{N} P_j^t A_{ji}, \tag{2.9}$$

where the $\vec{P}_j$ are called principal components (PCs) and the $\vec{A}_j$ are the empirical orthogonal functions (EOFs) of the field **X**. The EOFs have the same dimension of the spatial dimension of the database, and they can thus be seen as maps. Looking at Eq. 2.9 we can then see in the PCs the time evolution of the maps represented by the EOFs (Fig. 2.3). The sum of the oscillation of all the EOFs give the original field **X**. It can be proved that the trace of D is equal to the variance of X, thus the coefficient

$$\lambda_j = \frac{D_{jj}}{\mathrm{Tr}(\mathbf{D})}, \tag{2.10}$$

represent the variance fraction explained by each empirical mode of variability $\vec{P}_j * \vec{A}_j$. Usually it is common to consider as significative the first modes only, because they can explain alone great part of the field variance. However, among all the main modes, the first one is the most reliable, since is not constrained by the orthogonality requirement. To overcome this issue non-orthogonal EOFs can be derived through opportune algorithms.

As an example in fig. 2.3 we report the first EOFs of the North Atlantic geopotential height field at 500 mbar. As we can see the first map displays a dipole centred over high and mid latitudes. This pattern is closely related to the North Atlantic Oscillation (NAO), the leading mode of variability for the North Atlantic sector of geopotential field.

We will use EOF analysis in Chapter 6 to analyse the pattern of variability in the South Atlantic Convergence Zone (SACZ) region.

2.2.5 Entropy

Autocorrelation measures how much the value of a time-series X_{t+1} depends (linearly) by its predecessor X_t. Another, more general way to describe this concept is to measure the amount of "surprise" the value X_{t+1} causes given all the previous values. This quantity is called *Shannon entropy* and it is defined as [56]

$$H(X) = -\sum_{i=1}^{S} p(x_i) \, \log p(x_i) \tag{2.11}$$

where S denotes the number of possible values x_i the series X_t can take and $p(x_i)$ are their probabilities of occurrence.

The entropy H measures the degree of disorder, or unpredictability, of the series and it is maximum if the $p(x_i)$ are all equal.

2.2.6 Mutual Information

The mutual information (MI) among two time series, X and Y, is defined as the amount of information related to the time-series X contained in the time-series Y (and vice versa). If the time-series are discrete the MI reduces to a sum of the kind:

$$MI(X,Y) = \sum_{ij}^{S} p_{XY}^{ij} \, \log\left(\frac{p_{XY}^{ij}}{p_X^i p_Y^j}\right), \tag{2.12}$$

where S again is the number of possible values.

We can think to the MI as the non-linear partner of the Cross-correlation. In fact, if two variables are independent, we have $p_{XY}(x,y) = p_X(x)p_Y(y)$ and the logarithm in the integral is identically 0. The MI can also being computed with a lag-time.

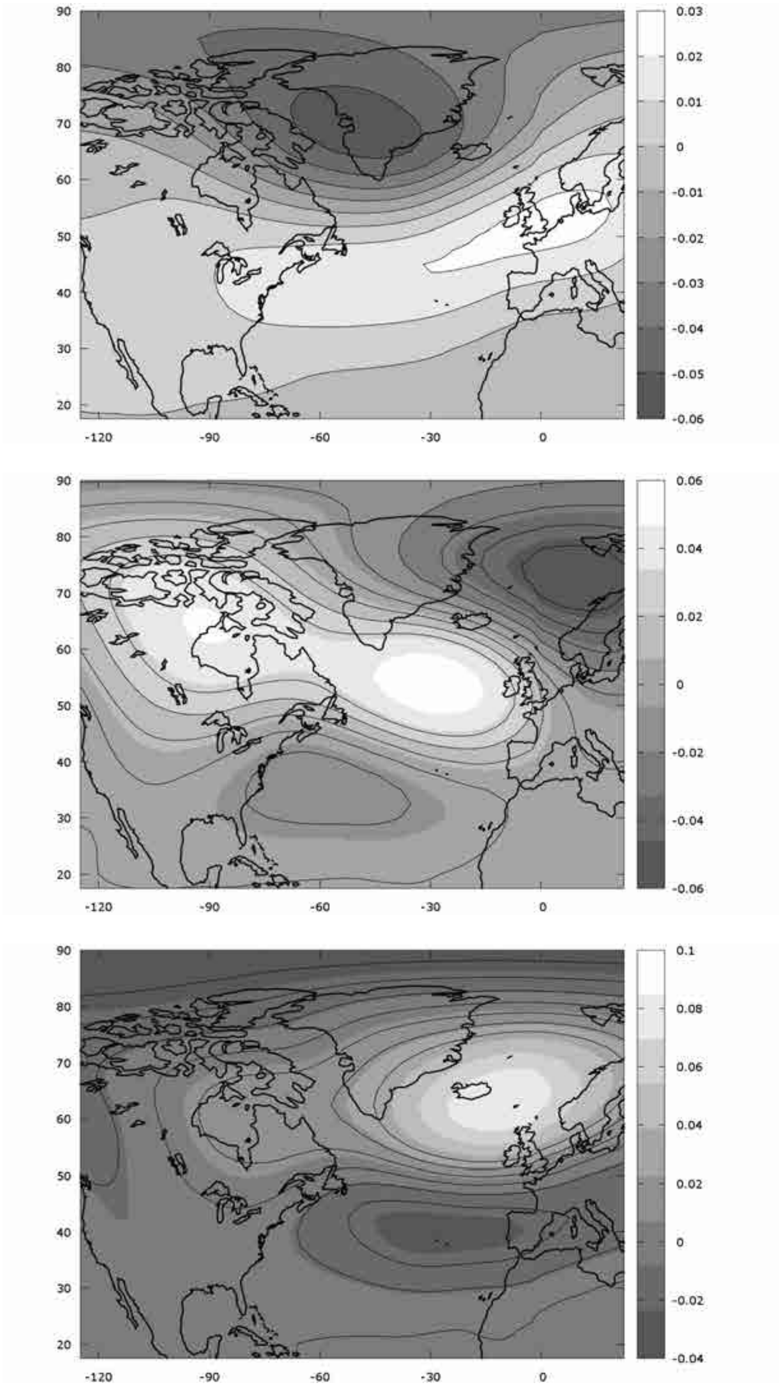


Figure 2.3: From top to bottom, first three EOFs for geopotential height at 500 mbar. The time series of geopotential height have a monthly time resolution and a 2.5° spatial resolution. Data from NCEP/NCAR reanalysis.

2.2.7 Ordinal Analysis

If we are interested in the trends displayed by the data, more than to their absolute values, we can use the so-called *ordinal analysis*, that takes into account only the sign of relative differences between the values of a time-series [57].

For example, let us imagine to have a time series of temperature records T_i daily sampled. If the value of the time series at a certain day is larger than the one of the day before, we assign to that day the symbol "+", while if it is lower we assign it the symbol "-". In this way we have reduced the time series to a sequence of two symbols that we can study through the information theory tools that we have presented in the previous sections. For example, once we have reduced two time series into two sequences of ordinal patterns, we can compute their ordinal mutual information (MIOP) using Eq. 2.12.

In this way, however, we lose the information about the real magnitude of the temperature values, thus we lose the information about the trends that are not at a daily scale. Nevertheless this coin has also a positive face; we are in fact able using ordinal analysis to perform time-scale selection.

If we want to study longer time scales than the sampling one, we have just to compare the value of the time series at a certain day with that of the time series at a certain time in the past. Repeating it for many different lags we are able to construct different ordinal series that focus on different time-scales.

Moreover, by using more than two symbols, one can investigate more complicated behaviours than just two-points linear trends. In Fig. 2.4 we report an example with 6 possible symbols (also called "words"). The number of the symbols depends on how many points we want to use to describe the trends. For three-points-trends we have 6 possible different ordinal patterns, while for 4 points ones the number grows to 24. Using combinatory calculus it is easy to see than the number N of words with length K grows as $N = K!$. Thus N quickly diverges with the number of points we use to define the patterns. This is a severe limitation if we want to study finite length time-series. For a standard monthly reanalysis time-series, the typical length of the data is $L \sim 10^3$. To estimate correctly the joint probability distribution of the mutual information, one should not go below the heuristic value of 5 data per possible symbol. The number of possible different symbols in a joint probability calculation is N^2. Thus we have a rule of thumb for N of the kind $N \sim \sqrt{L/5}$. For a monthly time-series this limits the analysis to $K = 3$ or – with large uncertainties – to $K = 4$.

2.3 Directionality

If we know that the time-series that we are analysing come from two systems that might be interacting, it can be relevant to check whether it is possible to determine the direction of the interaction from the time series behaviour.

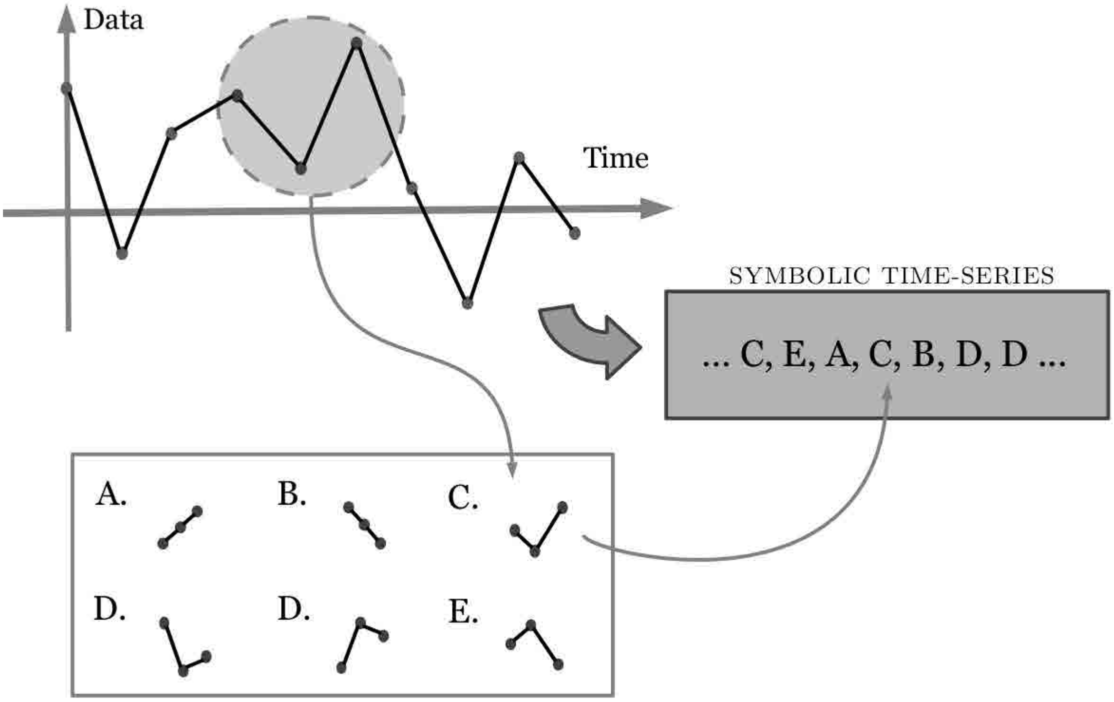


Figure 2.4: Example of ordinal pattern reduction, through symbols computed from 3 consecutive values of the original time-series. The number of possible patterns is 3! = 6.

The similarity measures presented before (cross correlation and mutual information) are symmetric, and thus do not provide information about a possible presence of a preferential direction of interaction, thus we have to use new tools to investigate this aspect.

A powerful tool that can help us in this task is the so-called Conditional Mutual Information (CMI) [58],

$$CMI(X,Y|Z) = \sum_k p_Z^k \sum_{ij} p_{X,Y|Z}^{ijk} \log\left(\frac{p_{X,Y|Z}^{ijk}}{p_{X|Z}^{ik} p_{Y|Z}^{jk}}\right) \tag{2.13}$$

To see how this measure can be useful to our scope, let us suppose to have computed the quantity $MI(X(t), Y(t+\tau))$ that is, the amount of information that the series X contains about the future of the time-series Y after a time τ. Since this measure contains a certain amount of information from $Y(t)$ to $Y(t+\tau)$, to eliminate this effect we compute the so called *transfer entropy* [58, 59]

$$TE_\tau(X,Y) = CMI(X(t), Y(t+\tau)|Y(t)). \tag{2.14}$$

Then, if $TE_\tau(X,Y) > TE_\tau(Y,X)$ we infer that the net transfer of information is from the X system to the Y one, and vice versa. Thus, the difference $TE_\tau(X,Y) - TE_\tau(Y,X)$ can be thought as a directionality index for the interaction among the systems X and Y [60, 61].

2.3.1 Granger Causality

Another way to infer the interaction direction among two processes X and Y relies in the possibility of modelling them as autoregressive processes. Let us suppose to study the influence of the time-series X over the time-series Y. We model Y as an autoregressive processes of order D forced by the elements of X:

$$Y_t = \sum_{i=1}^{D} a_i Y_{t-i} + \sum_{i=1}^{D} b_i X_{t-i} + \epsilon_t; \tag{2.15}$$

where $\vec{a}$ and $\vec{b}$ are vectors of constant coefficients and ϵ_t are the noise residuals.

The idea is to test the hypothesis $\vec{b} \neq 0$ against the null one $\vec{b} = 0$. To do so we proceed in three steps. In the first one we fit with a linear regression $\vec{a}$ and $\vec{b}$, and compute the associated variance of the residuals, $\sigma^2_{\text{coupled}}$.

In the second step the fit is repeated setting $b_i = 0$, and again the variance of the residuals is computed, namely $\sigma^2_{\text{uncoupled}}$.

The last step involves comparing the two residual variances: if the variance in the coupled case is smaller than the one in the uncoupled case, it means that the predictive power of the coupled model is higher. Then, the Granger causality estimator (GCE) is calculated as [62]

$$GCE = \frac{\sigma^2_{\text{uncoupled}} - \sigma^2_{\text{coupled}}}{\sigma^2_{\text{uncoupled}}}. \tag{2.16}$$

From this equation is clear that if $GCE > 0$ the information given by X allowed for a more precise prediction of Y. Thus, X is said to be Granger causal of Y.

To test the statistical significance of this result, is possible to use a variable computed from GCE having a known distribution. In particular we can write the quantity F as

$$F = \left(\frac{T - 2D - 1}{D}\right) \frac{\sigma^2_{\text{uncoupled}} - \sigma^2_{\text{coupled}}}{\sigma^2_{\text{coupled}}}, \tag{2.17}$$

where T is the length of the time-series, which distributed according the Fisher-Snedecor distribution with D and $T - 2D - 1$ degrees of freedom [63].

Let us now suppose that we have not just a couple of time-series X and Y but a whole ensemble of couples of series X_i and Y_i, being independent realisations of the same process described by Eq. 2.15. In this case we face a problem of multiple trials Granger causality. The procedure to compute the GCE is as seen before, the only difference being in the fit procedure. Writing Eq. 2.15 as

$$\vec{Z} = \mathbf{W}\vec{\gamma} + \vec{\epsilon}, \tag{2.18}$$

where $\vec{Z}$ is the time-series Y^T_{t+1}, $\vec{\gamma} = \{a_1, ..., a_D, b_1, ..., b_D\}^T$ and $\mathbf{W}$ is $2D \times T$ matrix containing the elements of X and Y, we can obtain first

$$\mathbf{W}^T \vec{Z} = \mathbf{W}^T\mathbf{W}\vec{\gamma} + \mathbf{W}^T\vec{\epsilon}. \tag{2.19}$$

Then, by averaging over all the trials we obtain

$$\vec{\gamma} = \left(\langle \mathbf{W}^T\mathbf{W} \rangle\right)^{-1} \langle \mathbf{W}^T \vec{Z} \rangle, \tag{2.20}$$

where we have supposed $\vec{\gamma}$ to be the same in all the trials.

As final note, the order D of the autoregressive model in Eq. 2.15 is chosen in order to minimize the function,

$$S_i(D) = \frac{T}{2} \ln\left(\frac{\sigma^2_{\text{uncoupled}}}{\sigma^2_Y}\right) + \frac{D}{2}\ln(T), \tag{2.21}$$

that is a good compromise between obtaining a good fitting and avoiding over-fitting [64].

The Granger methodology has been successfully applied to “disentangle” interactions in many complex systems from purely observational time-series, such as finance [65], neuroscience [66, 67], but also climate [68, 69].

2.3.2 Application to climate data

In [62] Granger causality has been employed to understand the interplay between ENSO and the Indian Monsoon. It has been found that the two systems influence each other mutually, although the interaction in one direction is different from the other one in the opposite direction. In particular, using non linear regressions as a generalisation of Eq. 2.15, it was shown that the influence of El Niño over the Monsoon is non-linear and active while the influence of the Monsoon over El Niño is mainly linear and inertial.

Moreover, using moving windows, it is possible to investigate time variations of the couplings. It has been shown that the impact of El Niño over the Indian Monsoon has a long scale time variation, with peaks in the 1890—1920 and 1950—1980 periods.

In Chapter 6 we will use this technique to study in detail the air-sea coupling in the South Atlantic Convergence Zone (cfr. 1.4).

2.4 Climate Networks

In the last decade, a new field of climate data analysis arose, based in the theory of complex networks, a tool already successful in the analysis of biological, social, and economical systems. Here we will review some concepts of network theory as well as their main application to climatological data analysis.

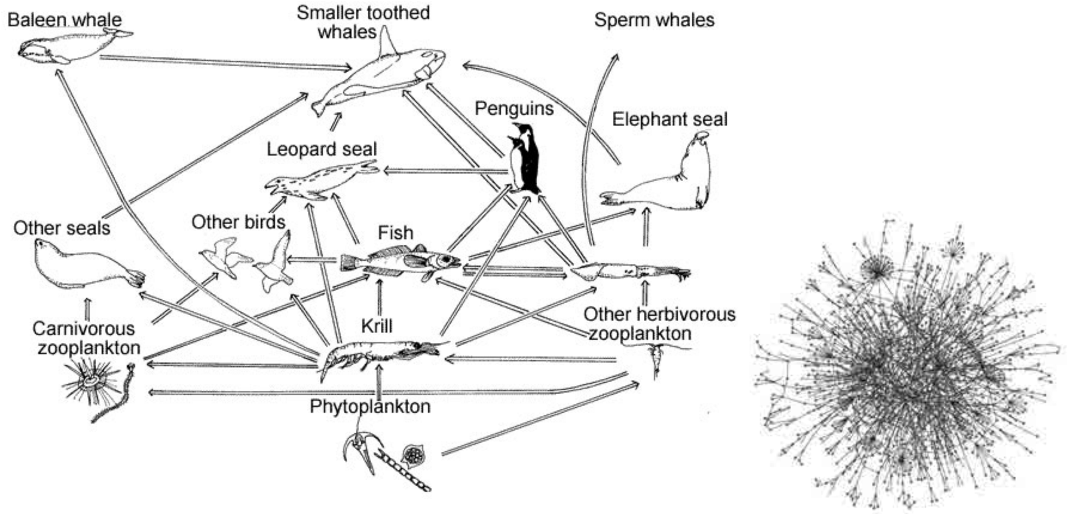
2.4.1 Complex Networks

Complex Network theory begins in the 18th century, with the work of Euler about the so-called Seven Bridges of Königsberg problem. To the mathematical object that he introduced was given the name of *graphs* by Silvester in the 19th century [70].

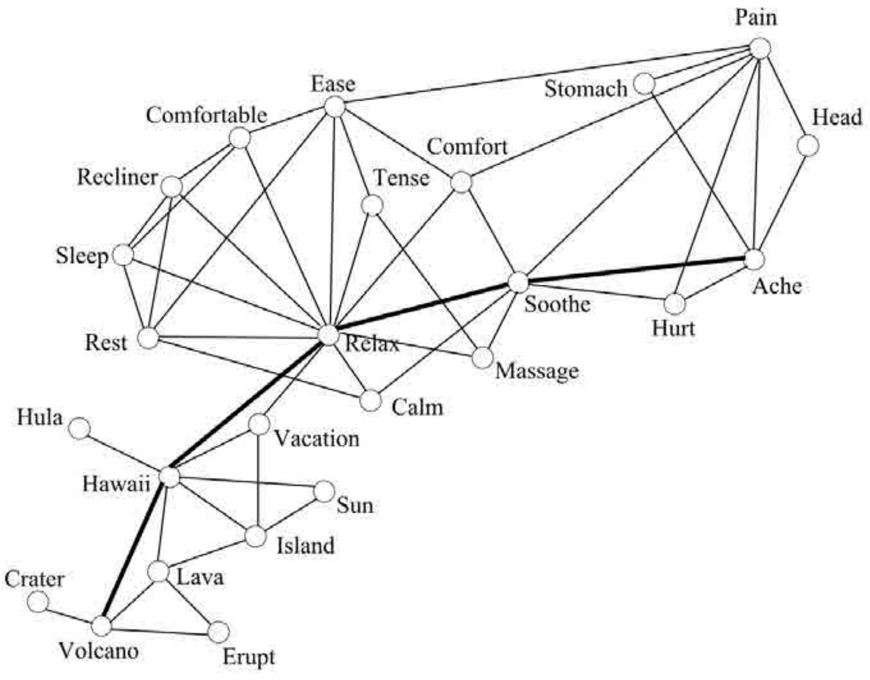
A graph G is composed by two sets. The first one, V, is the set of its "nodes" or "vertices", the second one, E, is the set of its "edges" or "links". Each edge is in relation with two elements of V and this relationship is what defines the graph topology [70].

When a graph is used to describe the relations among a certain number of discrete objects it is referred as *network*.

The network definition is very general, which makes them applicable to the description of numerous systems. In Fig. 2.5 we report three example of networks from different fields of research.



(a) food web (b) proteins web



(c) words web

Figure 2.5: Three example of complex networks coming from very different research fields. (**a**): Oceanic food web. The links represent prey → predator relations. [71] (**b**): Network of interactions between the proteins in *Saccharomyces cereat* better known as baker's yeast [4]. (**c**): Word netwok, in which words are linked if they occur in the same sentence [3].

2.4.2 Networks Properties

The relationship among nodes given by the vertices is the characteristic feature of a network. A straightforward way to describe this relationship is by mean of a matrix, the so-called *adjacency matrix*, defined as [70]

$$A_{ij} = \begin{cases} 1, & \text{if } i \text{ and } j \text{ are linked nodes,} \\ 0, & \text{otherwise.} \end{cases} \tag{2.22}$$

If the network contains N nodes the adjacency matrix will be a $N \times N$ one. If the relations among the nodes are non-directional (that is, having i connected to j implies that also j is connected to i) the adjacency matrix will be a symmetric one, whereas, if the relationships are directional, the matrix will be non-symmetric.

As we can see from Fig. 2.5 one of the main features of a complex network topology is that the number of the links varies from node to node. To quantify this feature we can compute the so called *degree*, that is, the number of nodes each node is connected to (also called *neighbours*):

$$k_i = \sum_j A_{ij}. \tag{2.23}$$

From the properties of the degree distribution we can infer different features of the networks. One of the most notable is the scale-free character, in which the degree distribution follows a power law; these networks are interesting since they reveal autorganization of the system [70].

The degree is just one of the many quantities that can be computed to analyse the properties of a network [70], and many of them have also been applied to climate data analysis [72]. Popular measures are the *clustering*, that measure the fraction of a node's neighbours that are neighbours also among them [73], and the *betweenness coefficient*, which is the fraction of shortest paths crossing a node over all the possible shortest paths in the network [72].

Several of these network properties will be described in more detail in Chapter 7 as they will be used as early warning indicators of desertification transition in a vegetation model.

Another important characteristic, that we will analyse in Chapter 5, is the presence of nodes that share some kind of property. These group of nodes are referred as *communities* [74]. Communities are very interesting in the framework of complexity reduction or system coarse graining. In fact, nodes that belong to a same community can be though into a first rough approximation as coherent parts of a bigger macrosystem, that will be sketched by the community itself. The connections among all the nodes between the various communities can be coarse grained into links among communities, and in this way one can build a network of communities, that encode, with a smaller complexity, great part of the information of the original network.

In Chapter 5 we introduce two methods for detecting the community structure of the global climate system: the first one is based on estimating time-lags between climate variables recorded at different locations; the second one exploits a symbolic technique, to estimate climate similarities in different locations. These two techniques reveal complementary information about the underlying climate dynamics: while the communities identified via time-lags reveal the main differences in thermal inertia of the various regions of Earth, symbolic analysis reveals a division in four main communities with similar dynamics: the ENSO basin, extra-tropical oceans, tropical oceans and extra-tropical landmasses. Thus, these two techniques yield light into main physical and dynamical aspects of climate variability.

Now we want to highlight a link between the classical linear analysis of the EOFs (Sec. 2.2.4), and a climate network built by the thresholding of the covariance matrix. This is one of the most used type of network, and also along this thesis is going to be used extensively.

We will show that the AWC of a network inferred by using the cross correlation is closely related to the EOFs.

As we have seen, the adjacency matrix of the network is a binarisation of the covariance matrix, in which only the strongest values are kept and set to 1. Therefore,

$$A_{ij} = \mathrm{H}(CC_{ij} - \Theta). \tag{2.24}$$

Given a database $\mathbf{X} = X_i^t$ of N time-series of length T, normalised at zero mean and unitary variance, we can decompose it through the EOFs as

$$X_i^t = \sum_n PC_n^t e_i^n, \tag{2.25}$$

where we denote with PC_i the principal components and with e_i the orthonormal patterns. Then, we can write the covariance matrix as

$$\begin{aligned}
CC_{ij} &= \frac{1}{T}\sum_t X_i^t X_j^t = \\
&= \frac{1}{T}\sum_t \left(\sum_n PC_n^t e_i^n\right)\left(\sum_m PC_m^t e_j^m\right) \\
&= \frac{1}{T}\left(\sum_{n=m} + \sum_{n\neq m}\right)\sum_t \left(PC_n^t e_i^n\right)\left(PC_m^t e_i^m\right) = \\
&= \sum_n \mathrm{var}\left[PC_n\right] e_i^n e_j^n + \sum_{n\neq m} \mathrm{cov}\left[PC_n, PC_m\right] e_i^n e_j^m, \\
&= \sum_n \lambda_n e_i^n e_j^n + \sum_{n\neq m} \mathrm{cov}\left[PC_n, PC_m\right] e_i^n e_j^m, \\
&= \sum_n \lambda_n e_i^n e_j^n,
\end{aligned} \tag{2.26}$$

since the principal components covariances are null; here λ are the eigenvalues of the covariance matrix.

Now, if the data **X** has a dominant pattern of variance, we will have $\lambda_1 \gg \lambda_{n\neq 1}$, whereas, since the e_i vectors are all of the same magnitude, their external products will be roughly all of the same order of magnitude. Thus Eq. 2.26 will reduce to

$$CC_{ij} \approx \lambda_1 e_i^1 e_j^1. \tag{2.27}$$

The adjacency matrix will be composed by only the elements of this new matrix that are large enough to trespass the threshold. Computing the quantity $A_i = \frac{1}{N}\sum_j \left|CC_{ij}\right| = \alpha \left|e_i^1\right|$ where α is a proportionality factor, and approximating:

$$A_i \approx \frac{1}{N}\sum_j \left|CC_{ij}\right|_{\text{big elements}} \sim AWC_i \tag{2.28}$$

and thus in first, crude approximation, we can link the area weighted connectivity and the EOFs as

$$AWC_i \sim \left|e_i^1\right| \tag{2.29}$$

that is the spatial pattern displayed by the AWC is dominated by the spatial pattern of the first EOF of the field.

2.4.3 Functional and Structural Networks

Systems composed by interacting dynamical elements are ubiquitous in nature. In many situations, it is desirable to model such systems as networks of coupled oscillators, where the nodes represent the individual units and the links represent the interactions among them. These interactions, which arise due to a variety of physical mechanisms, are often unknown, or only partially understood, and thus, a complete knowledge of the network topology is lacking. Technical limitations, or the nature of the network itself, make sometimes impossible to infer how the nodes are linked among each other. Within this framework, it is important to address the problem of how to optimally infer the connectivity of a system from the observation of the dynamics of its interacting elements[75, 76, 77, 78, 79, 80, 81, 82]. Two paradigmatic examples of this situation are brain functional networks and climate networks [83, 84, 85, 86, 87, 88, 89, 90].

In these cases, networks are built through the statistical study of the correlations between the time series associated to different physical regions (areas of the brain or geographical areas of Earth), which, in turn, are the nodes of the network. First, a covariance matrix representing the coordinated activity between all pairs of nodes is calculated, and then, a functional network is obtained by thresholding the covariance matrix: if two nodes display a high statistical similarity they are considered linked, otherwise not [91, 87]. In this way, from the filtered covariance matrix an adjacency matrix is obtained. In the following, we will refer to this inferred matrix as the functional network associated with the system. The functional network does not necessarily correspond to the real physical connectivity: In fact, two or more nodes are influenced by the same noise source, or some common external forcing, their dynamics might result highly correlated even if they are not directly interacting.

However, as one aims at finding a functional network that resembles as close as possible the physical connectivity of the system, it is crucial to develop reliable techniques capable of unveiling interdependencies, based on the observation of the dynamics of the units that compose the system.

In the recent years, numerous approaches of network inference have been proposed [77, 78, 79, 80, 81, 92, 93]. For example, one method was based on studying the response of coupled oscillators when they are perturbed by an external forcing [77]; in this case the network was inferred after a sufficiently large amount of tests with different driving intensities. In [80] the functional network was retrieved by using a delayed feedback control that drives the network into steady states. Other methods rely on the direct observation of the nodes' dynamics. For example, in [79] and [82], the Laplacian matrix is reconstructed using the covariance matrix of the nodes' time-series, and from the properties of the Laplacian, the links are inferred.

In [81] a system composed by discrete-time units was studied (various maps were considered, including the Logistic and Circle maps) and it was found that (i) under appropriated coupling conditions the network can be perfectly inferred, these conditions being a weak coupling regime where the network is neither fully synchronized, nor completely desynchronized and (ii) regarding the statistical similarity measure used for inferring the system structural connectivity, the mutual information in general outperforms the cross-correlation.

In Chapter 4 we test these observations by considering a set of phase oscillators and a set of chaotic three-dimensional oscillators. We chose the Kuramoto system because it is a paradigmatic model of many physical, biological and chemical systems [94]. To test the inference method under controllable experimental conditions, we also consider Rössler oscillators, operating in the chaotic regime, which are implemented electronically. I the case of the Rössler oscillators we will use only one component of the three-dimensional time-series of the oscillators instead of using a multivariate analysis. In this way we can test what happen with those systems in which only a projection of a multidimensional dynamics is available for being monitored, as in brain and climate network.

We show that the inference method performs well for both, Kuramoto phase oscillators and Rössler chaotic oscillators. In particular we find that, contrarily to what was found in [81] for discrete-time maps, in the Kuramoto oscillators' case the cross correlation is usually the best performing similarity measure. We also show that the inference method is able to reconstruct the physical connectivity of the system (in the following, referred to as the structural network) for a broad range of parameters, even when the oscillators are synchronized. This is due to the presence of independent noise sources in the individual oscillators, which prevent them to assume identical states.

2.4.4 Application to climate data

The tool of functional network can be applied to the study of climate data. Let us suppose to have a dataset relative to a climate variable discretised in time and space $\mathbf{X} = X_i^t$, where i is a spatial grid index and t a temporal one. Then we can compute a certain statistical similarity measure (SSM) among all the possible pair of points and obtain a matrix SSM_{ij}. The SSM can be the absolute cross correlation [87], the mutual information [95], the mutual information computed via ordinal patterns [96], or more complicated objects [97, 98].

If we want to focus only in the strongest values of this similarity we can binarise the SSM_{ij} using a threshold[87]:

$$A_{ij} = \mathrm{H}(SSM_{ij} - \Theta), \tag{2.30}$$

where Θ is the threshold value and H is the Heavyside step-function. The value of Θ can be either global (that is, a constant) or local (that is, depending on the node or on the link).

The matrix **A** can then be interpreted as an adjacency matrix and thus as a network. For example we can think to compute the weighted degree of such a network as [87]

$$AWC_i = \frac{\sum_j w_j A_{ij}}{\sum_j w_j}, \tag{2.31}$$

where w_j is a node weight to take into account that different grid points could refer to regions with different surfaces. In this way this quantity represents the fraction of the total area the node i is connected to; for this reason it is called *area weighted connectivity*. For example, in this thesis we will use data with a fix degree resolution, this means that nodes close to the Earth's poles will describe smaller areas than those close to the equator. In this case the weights are defined as

$$w_i = \cos(\Lambda_i), \tag{2.32}$$

where Λ is the node's latitude.

The first application of such a procedure was performed by the pioneeristic work of Tsonis and Roebber [87] in which, using as SSM the cross-correlation, they were able to find that the geopotential winter anomalies at 500 hPa could be associated with a climate network composed of two main sub-networks, a tropical one, highly interconnected, and an extratropical one, less connected, and with an almost scale-free structure.

From this first work many others arose. In [72] the work of Tsonis is generalised with the introduction of the mutual information aside the cross correlation, to take into account also the nonlinear component of the dataset. The analysis is then focused in the determination of the betweenness and of the centrality field, discovering that many areas of high betweenness are associated with the presence of oceanic currents.

In [99], instead, a set of networks is constructed using cross-correlations for geopotential heights at various pressure levels. Then, links between the networks are built, again using the cross correlation, producing in this way a multilayer network for the study of the interaction among the various levels of the atmosphere. The network is studied through various indicators, like betweenness, clustering and centrality, and many features of atmospheric circulation are retrieved, such as Ekmann effect, Hadley, Ferrel and Polar Cells, Circumpolar Vortex, Tropopause Effect and Hot Towers.

Other studies have been focused on localised regions, instead than on the whole world. For example in [98, 100, 53] the ENSO region is deeply investigated, and in particular it is pointed out that, during an El-Niño events, the central Pacific becomes an isolated system, not being forced any more by the surrounding areas. In [101], instead, the dependence of the climate network of Northern Atlantic by the NAO phase is studied, finding a strengthening (weakening) of the links during positive (negative) NAO phases.

Climate networks have been also useful in the studying of phenomenon as the Indian Summer Monsoon [102] in which through the analysis of various complex-network metrics, they can describe typical repetitive patterns of organised rainfall, in various regions of the Indian subcontinent.

In [74] multidecadal variability of sea surface temperature is analysed, through a community detection algorithm applied to a network built from cross-correlations values, finding out that, while El Niño–Southern Oscillation is most dominant on interannual timescales, the Indian West Pacific and North Atlantic may also play a key role on decadal timescales.

Another very interesting application of climate networks, are the so-called *lagrangian networks* [103], in which the links in a fluid (in this case the Mediterranean basin) are established by the transport of some tracers by the flux. In this way it is possible to define coherent structures in the circulation (like, for example, gyres or fronts) called "provinces", that can help to understand plancton or pollutant dynamics.

3 Effects of Seasonal Lag-Times in Climate Networks

As discussed in the previous chapter, the complex network framework has been successfully applied to the analysis of climatological data, providing, for example, a better understanding of the mechanisms underlying reduced predictability during *El Niño* or *La Niña* years [98, 100, 53]. Despite the large interest that climate networks have attracted, several issues remain to be investigated. In this chapter we focus in the influence of the periodic solar forcing in climate networks constructed via similarities of monthly averaged surface air temperature (SAT) anomalies. We shift the time series in each pair of nodes such as to superpose their seasonal cycles. In this way, when two nodes are located in different hemispheres we are able to quantify the similarity of SAT anomalies during the winters and during the summers. We find that data time-shifting does not significantly modify the network area weighted connectivity (AWC). This network property can be understood in terms of how time-shifting modifies the strength of the links connecting geographical regions in different hemispheres, and how these modifications are washed out by averaging the AWC.

The results presented in this chapter have been summarised in [85].

3.1 Dataset

We consider monthly-averaged surface air temperature (SAT), re-analysis data from the Center for Environmental Prediction/National Center for Atmospheric Research (NCEP/NCAR) [50]. The data covers the period from January 1948 to May 2012 ($N = 773$ months) and have a spacial grid resolution of 2.5° ($M = 10226$ nodes). Removing the annual cycle in each node results in zero-mean anomalies (SATA), which are then normalized to have unitary variance.

3.2 Identification of Lag-Times

To compare the climate in any two nodes, i and j, under the same stage of the annual solar cycle, we shift the time series in one node, and find the time shift that gives maximum similarity, which is referred to as the *lag-time*, τ_{ij}. Specifically, we calculated the Cross Correlation coefficient ($CC_{ij}(\tau)$) between the SAT time-series (including the annual cycle) in i and j, shifting one series τ months with respect to the other. Then, τ_{ij} is chosen as the value of τ in the interval [0,11] where $CC_{ij}(\tau)$ is maximum (see Fig. 3.1). There are pairs of nodes for which $CC_{ij}(\tau)$ has two maxima in the interval [0,11]; this can be frequent when one of the nodes is in the tropical regions, were there is a well known semi-annual periodicity. In these cases the lag-time was chosen as the smallest of the two possible lag-times.

To demonstrate that this procedure indeed gives meaningful lag-times, in Fig. 3.2 we display in color code the lag-times of three nodes, one in the north hemisphere (NH), one in the south hemisphere (SH), and one in the tropics. The maps reveal clear characteristic patterns, signatures of climatic regions. We can also note the memory effect induced by the oceans, and the almost perfect 6 months symmetry between the NH and SH.

For the sake of clarity we remark that τ_{ij} if determined from the similarity of the SAT field, while the network is constructed from the similarity of anomalies (SATA field).

3.3 Construction of Climate Networks

We measure the degree of *statistical similarity* of the time-series in two nodes by the absolute value of the cross correlation coefficient (CC) and by the mutual information (MI), using the identified lag-times:

$$CC_{ij} = \left| \frac{1}{N-\tau_{ij}} \sum_{t=0}^{N-\tau_{ij}} a_i(t)\; a_j(t+\tau_{ij}) \right|, \tag{3.1}$$

where the a_i and a_j are the SATA time series in nodes i and j, τ_{ij} is their lag-time, and $N = 773$ months;

$$MI_{ij} = \sum_{m,n} p_{ij}(m,n) \log_2 \left(\frac{p_{ij}(m,n)}{p_i(m) p_j(n)} \right), \tag{3.2}$$

where the p_i, p_j are the probability distributions associated to $a_i(t)$ and $a_j(t)$, and p_{ij} is the joint probability of $a_i(t)$, $a_j(t+\tau_{ij})$. These probabilities were estimated by 8-bin frequency histograms [72].

We also used the *Ordinal Pattern Mutual Information* (MIOP, see Sec. 2.2.7) This symbolic transformation allows to compute the MI – and from that to obtain a network – that takes into account memory effects at different time-scales, depending on the way the ordinal patterns (OPs) are constructed.

We consider OPs of length $D = 4$, and thus the number of possible different OPs is 24. In this way, the SATA data in each node i, $a_i(t)$, is transformed into a sequence of integer numbers, $n_i(t)$ with $n_i \in [1,24]$. From the symbolic sequences in two nodes we compute the probabilities

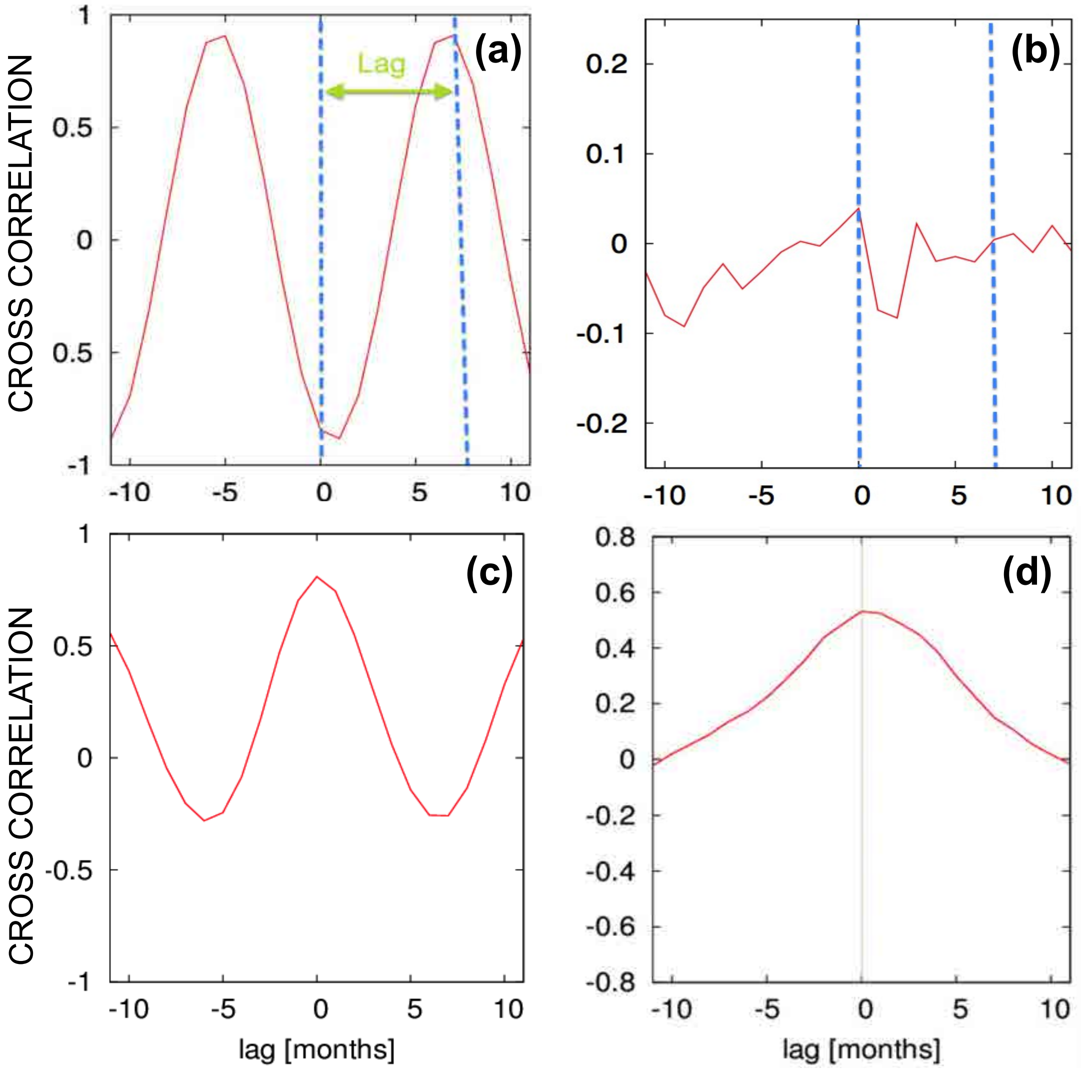


Figure 3.1: Representation of the method used to identify the lag-time, τ_{ij}, between nodes i and j. As an example we plot in panel (a) the cross-correlation of the SAT time-series of nodes close Rome and Buenos Aires, there the maximum in the interval [0,11] occurs at $\tau_{ij} = 7$ months. (b) CC between the same nodes but now computed from SATA data (notice that there is no significant maximum). (c), (d): as (a), (b) but for the teleconnection between a node in the equatorial Indian Ocean (7 S, 65 E) and a node in *El Niño* basin (12 S, 145 W). In this case there is a pronounced maximum at zero-lag in (d) (vertical line).

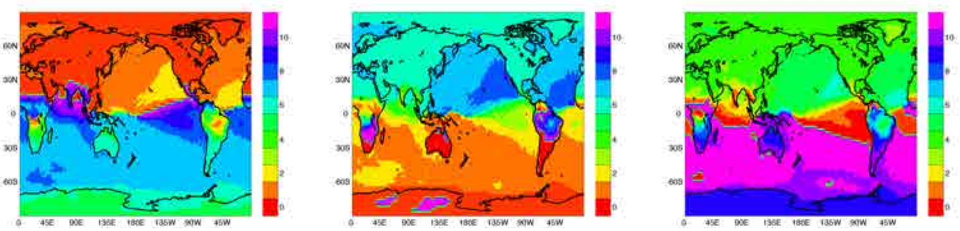


Figure 3.2: Lag-times of a node in Mongolia (left), in Australia (center) and in El Niño basin (right).

of the OPs, and then, the MIOP from Eq. 3.2.

The advantage of the symbolic method is that the OPs can be constructed with either consecutive or non-consecutive months. We constructed the OPs with four consecutive months as well as with four equally spaced months covering a one-year period. In the following sections we will present the results only in the former case because the main conclusions are the same for both ways of constructing the OPs.

3.4 Distribution of link strengths and lag-times

A first test of the influence of the lag-times in the climate network is provided by the analysis of the degree of similarity, by comparing the distributions of the values of CC and MI computed with and without lag-times. This is presented in figs. 3.3a, 3.3c (CC) and figs. 3.3b, 3.3d (MI), where one can observe that the distributions are almost identical, with no significant influence of the lag-times in the mean values or in the shape of the distributions.

Further insight can be obtained by analysing the relation between the strength of a link and its associated lag-time. As discussed in the introduction, climate networks are defined over a regular grid, in which some long-range teleconnections occurs. This fact is reflected in the two-dimensional probability distributions of similarity values in the plane (lag-time, CC/MI value) displayed in fig. 3.3e (CC) and in fig. 3.3f (MI); in both cases the probability is represented in logarithmic scale. As it was expected, the higher occurrence of high correlations/mutual information values is at zero lag, confirming the local character of the vast majority of the connections.

3.5 Extracting relevant links with non-zero lag-times

In order to build climate networks we need to define a significance criterion, such that, if the similarity measure of nodes i and j exceeds the threshold they are linked, otherwise they are not.

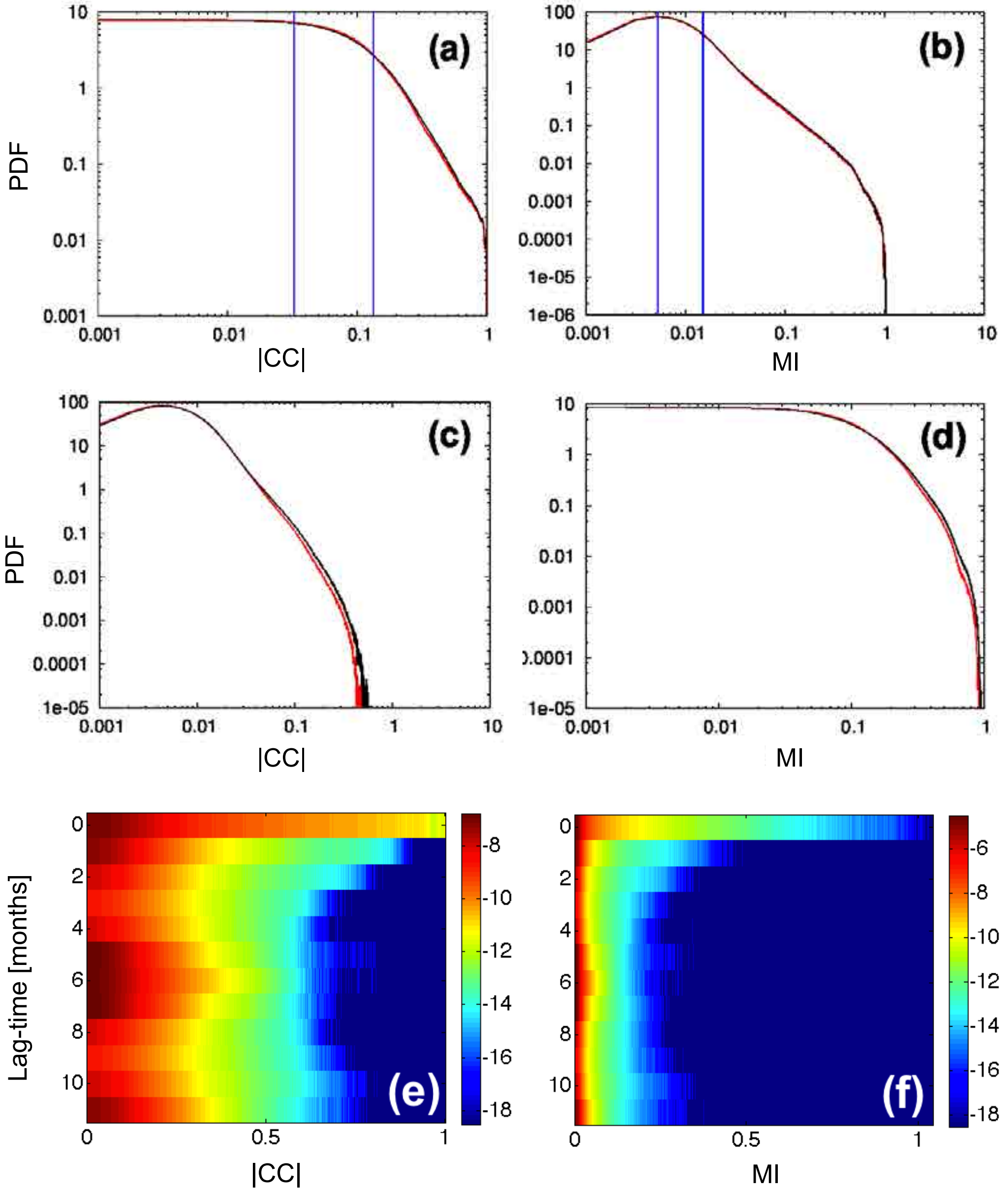


Figure 3.3: **(a)**,**(b)**: Histograms (in log-log scales) of the absolute value of the Cross Correlation (|CC|) and of the Mutual Information (MI), computed with (red) and without (black) lag-time shifting of the time-series. as can be seen the histograms are nearly identical. The blue lines indicate the two thresholds with which 50% of the links are extracted (see text for details). **(c)**,**(d)**: Histograms of CC_{ij} and MI_{ij} such that $\tau_{ij} \neq 0$. The difference with the histograms in **(a)** and **(b)** is a decrease of the maximum value due to the removal of local links, which are at $\tau_{ij} = 0$. **(e)**,**(f)**: Two-dimensional plots of the probability (in color code in logarithmic scale) vs. the value of |CC| or MI and the lag-time. It can be seen that most of the links with high |CC| or MI values have $\tau_{ij} = 0$, which can be expected as the strongest links are short-range connections.

However, since we are interested in observing the effects of the lag-times in the network topology, we can not keep only the strongest links, as these are mainly local and thus, have zero-lag. The influence of the lag-times in the network connectivity can be observed only if the network contains weak links. However, when links that have very low CC or MI values are included, they might not be relevant as these low correlations might occur just by chance.

As a compromise solution we decided to use two thresholds, in the following way: we chose a high threshold to remove the 25% strongest links (which are at zero-lag and will obscure the influence of the lag-times), and a low threshold to remove the 25% weakest links (which are considered noise). In this way the network extracted preserves 50% of the total links, those that are in the second and third "quartiles" (see figs. 3.3a and 3.3b). To justify the use of a low threshold, we point out that, as the CC value in Eq. 3.1 is the sum of about 770 terms, it can, for simplicity be assumed to be the sum of independent identically distributed random variables with zero mean and standard deviation 1. Such sum can then be expected to follow the central limit theorem, thus the cross-correlation –without the absolute value as in Eq. 3.1— can be expected a to be zero-mean Gaussian distributed with standard deviation around 0.04. Then, correlation values above 0.03 are likely to occur by chance with a probability of about 0.4.

3.6 Results

The networks obtained after computing the three similarity matrices, CC_{ij}, MI_{ij} and $MIOP_{ij}$ (with the ordinal patterns constructed with four consecutive months), and after filtering with the thresholding technique described previously, are then graphically represented with the Area Weighted Connectivity (AWC) in each node i [104],

$$AWC_i = \frac{\sum_j^M A_{ij}\cos(\lambda_j)}{\sum_j^M \cos(\lambda_j)}. \tag{3.3}$$

where A_{ij} is the adjacency matrix and λ_j is the latitude of the node j.

The results are presented in Fig. 3.4: in the bottom row the lag-times were included when we computed the similarity matrices; in the top row, instead, the similarity matrices were computed at zero-lag.

as can be noticed, the introduction of the lag-times results only in tiny changes in the network, although the network constructed via CC values seems to be more influenced. This is the main and surprising result of our analysis, since it means that the synchronization of the time series according to the annual seasonal cycle (that is, comparing winters with winters, summers with summers and so on) does not increase the degree of connectivity.

To investigate if the influence of the lag-times could be hidden by noise, i.e., by the presence of weak links that are not significant, or by strong links (those in the third quartile, as the strongest quartile has already been removed), we applied an extra filtering technique. To each

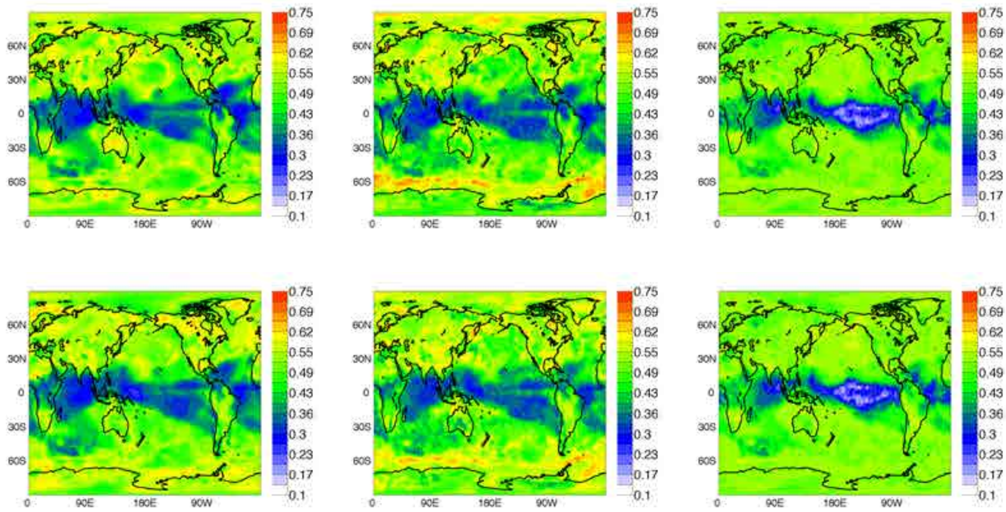


Figure 3.4: AWC plots: the degree of statistical similarity is quantified with the absolute value of the CC (left), the MI (center) and the MIOP (right), with and without lag-times (lower and upper row respectively). The thresholds to construct the network are such that the link density is 50%, with the strongest and weakest links removed.

link two labels have been assigned: a CC-label and a MI-label, each of them ranging from 1 to 4, representing the quartile the CC or MI value belongs to. Then, we filtered out the links that had one of the two labels equal to 4 (i.e., they were in the strongest quartile, either in terms of CC or MI value). In this way most of the local connections were eliminated. Then, we also eliminated the links that were in the first quartile of CC and in the first two quartiles of MI. In other words, we retained links with small CC values but with intermediate MI values. In this way we obtained networks with density of the order of 50%, but also their AWC maps (not shown) did not reveal any significant influence of lag times.

We also considered the influence of a random choice of the lag-times in the interval [0,11], and to our surprise, again the AWC maps did not reveal any clear influence of the random data-shifts.

It is possible that the introduction of lag-times results in small variations of the strength of the links, that can not be observed due to the thresholding process, by which the similarity matrices, CC_{ij} and MI_{ij}, are transformed into adjacency matrices, A_{ij}, of 0s and 1s. To investigate this issue, in each node we computed a *Weighted* AWC (WAWC) defined as in Eq. 3.3, but replacing the adjacency matrix, A_{ij}, with the similarity measure (CC or MI). Then, we plotted the difference between the WAWC calculated with lag-times and the WAWC calculated with zero-lags. The results are presented in the first row of Fig. 3.5, where a positive difference indicates an increase of the average correlation.

As we can see the changes are indeed very small, and this observation is consistent in the three

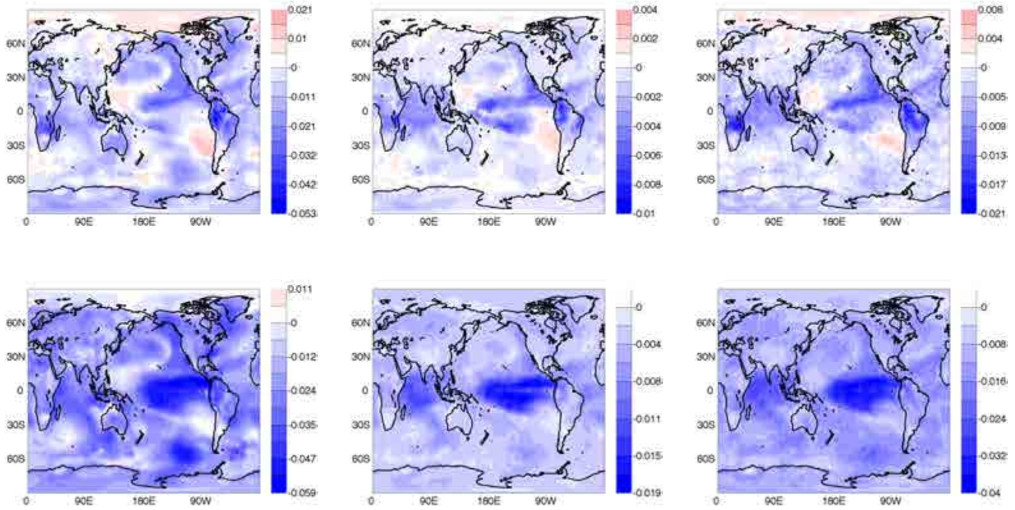


Figure 3.5: Variation of the average degree of similarity (WAWC, see text). The top row displays the variation when the lag-times used are those computed from SAT data, the bottom row displays the variation when the lag-times are chosen randomly in the interval [0, 11]. The degree of similarity is quantified with the absolute value of the CC (left), the MI (center) and the MIOP (right).

similarity measures, with a localized region of weak correlation enhancement, which does not occur in the mid-latitudes. On the contrary, it seems that an average loss of correlation influences these regions and, more in general, the global AWC map. Therefore, not only the introduction of lag-times produces only very small effects, but these occur in unexpected regions, such as western equatorial and south-eastern Pacific.

We tested these results against a random choice of lag-times in the interval [0,11] (Fig. 3.5 second row). As can be seen in this case there is only a loss of correlation, especially in the ENSO basin. This latter effect is due to the fact that El Niño region is connected with almost all the tropical belt, that is a region at zero-lag; shifting randomly the series partially destroys the correlations (see the correlation plot in fig. 3.1d), resulting in a decrease in WAWC.

We performed further checks of the analysis. We removed the local links by a spatial threshold of 3000 Km, searching for lag effects in the strongest fourth quartile. In this way only the strong teleconnections have been considered. However the inclusion of lag-times did not results in increased in mid-latitude connectivity. We also considered longer lag-times (all lags equal to 5 years) which resulted in the loss of the network architecture and the emergence of noisy structures, as expected.

We point out that, since the analysis was done with SATA data of 773 months, the network obtained is an average of the climate network over more than 64 years. If we use shorter windows the results might be different and the network topology might show a different sensitivity to time-shifts. However, by dividing the 773 months in shorter data sets we can compromise the robustness of the analysis as it would be performed over short data sets, and the variability of the results can then be attributed either to i) insufficient statistics, or ii) the evolution of the network topology. Nevertheless, the possible influence of time-shifts in networks defined over shorter time intervals can be an interesting study when performed over data set consisting, for example, of daily or weekly averaged SAT values (instead of averaged monthly as here).

We also carried out an explorative analysis of the geopotential height field at 1000 and 500 hPa, and also these field seems to show the same behaviour as the SATA field here described.

3.7 Conclusions

We have shown that the introduction of time-shifts when computing statistical similarity measures does not produce significant changes in the AWC of climate networks built from monthly-averaged SAT anomalies. This is an unexpected finding since the shifts were aimed at comparing the time series in two nodes at the same phase of the annual solar cycle (i.e., comparing winters with winters, summers with summers). This observation is robust with respect to the choice of statistical similarity measure, and is robust with respect to a random choice of time-shifts. Our results can be understood in terms of the lag-time distribution of the links, and the fact that most of the links with non-zero lags are weak links [see the 2D histograms in Figs. 3(e), 3(f)]. For these links the time-shifting indeed changes their similarity values; however, these changes appear to be random (see fig. 3.1b) and tend to be washed out when computing the AWC.

4 Inferring the connectivity of coupled oscillators from data

In this chapter, by considering synthetic and experimental data generated from two systems of coupled oscillators (Kuramoto phase oscillators and Rössler chaotic electronic oscillators) with known and controllable coupling conditions, we aim at testing the performance of the network inference method, by using the same linear and non linear statistical similarity measures used in the previous chapter.

The results presented in this chapter are resumed in [105].

4.1 Methods

In this section we first describe the two systems of coupled oscillators analysed: Kuramoto phase oscillators and Rössler chaotic oscillators, which were implemented experimentally via electronic circuits. Then, we describe the statistical similarity measures (SSMs) used to infer the system's physical connectivity: the absolute value of the cross-correlation and the mutual information. The latter was computed in two ways: i) via the usual estimation of probabilities from histogram of time-series values and ii) via the probabilities of symbolic patterns, using the ordinal transformation of time-series.

Kuramoto Phase Oscillators

The rate equations for N coupled Kuramoto phase oscillators are [106]

$$\mathrm{d}\theta_i = \omega_i \mathrm{d}t + \frac{K}{N}\sum_{j=1}^{N} A_{ij}\sin(\theta_j - \theta_i)\mathrm{d}t + D\,\mathrm{d}W_t^i, \tag{4.1}$$

where θ_i and ω_i are respectively the phase and the natural frequency of the oscillator i, and K is the coupling constant between the oscillators. $\mathrm{d}W_t^i$ is a Weiner process having 0 mean and variance tuned by the parameter D. The matrix $\mathbf{A}$, usually referred to as the adjacency matrix, defines the structural network; it indicates the physical connectivity of the system, i.e., the existing interactions among pairs of nodes. In detail: $A_{ij} = 1$ if two oscillators are linked and $A_{ij} = 0$ otherwise. We consider symmetric bidirectional coupling, thus **A** is a symmetric matrix.

The distribution of the natural frequencies is commonly considered to be centered in 0, since the equations are invariant under a phase translation, $\theta_i \to \theta_i - \langle \omega_i \rangle\, t$. Moreover, the distribution is often taken symmetric [94], so we assume here that the oscillators' frequencies are uniformly distributed in the interval $[-\Omega, \Omega]$.

The collective dynamics of the oscillators is usually monitored by means of a complex order parameter, R, defined as [94]

$$R(t) \equiv \rho(t)\; e^{i\psi(t)} = \frac{1}{N}\sum_{k=0}^{N} e^{i\theta_k(t)}, \tag{4.2}$$

where $\psi(t)$ represents the average phase over the ensemble of oscillators and $\rho(t)$ quantifies the degree of synchronisation, ranging from 0 (no synchronisation) to 1 (perfect synchronisation).

From the oscillator phases, $\theta_i(t)$, we derive another two quantities: i) the instantaneous frequencies,

$$\nu_i(t) = \dot{\theta}_i(t), \tag{4.3}$$

and ii) the Y-projections

$$Y_i(t) = \sin(\theta_i(t)), \tag{4.4}$$

i.e., the vertical projection of the unitary vector associated to the phase θ_i. In this way, despite the interaction involves only the oscillator phases, we have three different magnitudes to infer the underlying links: the phases, the frequencies and the Y-projections. We follow this procedure because in real-world systems it is frequent to use proxies of the quantities of interest.

Electronic Rössler Oscillators

As stated in the Introduction, we test the network inference method with experimental data by constructing a set of coupled electronic oscillators. Specifically, the experiment consists of 12 identical piecewise Rössler electronic circuits [107, 108, 109] whose dynamics can by modelled by:

$$\hat{\tau}\dot{x}_i = -\Gamma x_i - \beta y_i - \lambda z_i + K\sum_{j=1}^{N} A_{ij}\,(x_j - x_i) \tag{4.5}$$

$$\hat{\tau}\dot{y}_i = x_i - \upsilon y_i \tag{4.6}$$

$$\hat{\tau}\dot{z}_i = g(x_i) - z_i \tag{4.7}$$

where $g(x)$ is the piecewise linear function

$$g(x) = \begin{cases} 0 & \text{if } x \leq 3, \\ \mu(x-3) & \text{if } x > 3 \end{cases} \tag{4.8}$$

x_i, y_i and z_i are the oscillator state variables and $\mathbf{A}\{A_{ij}\}$ is the structural connectivity matrix. We consider a symmetric bidirectional coupling, thus $\mathbf{A}$ being a symmetric matrix.

Figure 4.1 displays a schematic diagram of the experimental setup. The parameters of the individual circuits are such that their dynamics is chaotic, and the coupling parameter is varied in a range such that the system is not synchronized.

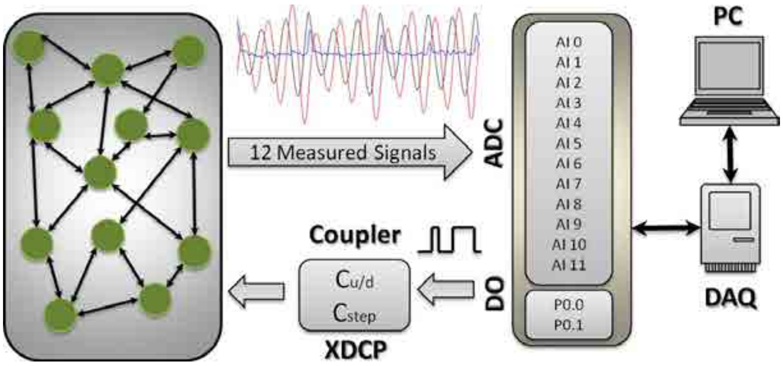


Figure 4.1: Diagram of the experimental setup representing a network of 12 electronic Rössler oscillators. The bidirectional coupling is adjusted by means of 12 digital potentiometers X9C104 whose parameters $C_{u/d}$ (control of increment/decrement of the resistance) and C_{step} (time-step to change the value of the resistance) are controlled by a digital signal via a DAQ Card. The output of each circuit is sent to a voltage follower that acts as a buffer and, then, sent to the analog ports (AI 0; AI 1; ... ; AI 11) of the same DAQ Card. The whole experiment is controlled from a PC with a Labview Software.

Since the most common scenario in real systems is not having a complete knowledge of the dynamical state of the system and, typically, not all relevant variables can be observed, we only took into account one of the three variables of the oscillators, $x(t)$, where the interaction actually takes place, and obtained an univariate time-series for each oscillator. Next, from the time-series of $x_i(t)$ we obtained the time-series of phases, $\theta_i(t)$, through a Hilbert transformation, and from the phases we computed the instantaneous frequencies, $\nu_i(t)$. In this way, also for Rössler oscillators we have three magnitudes from which we can attempt to infer the underlying links: x, θ and ν.

Statistical similarity measures

The construction of the functional network is based on evaluating the similarity of the dynamics of the oscillators through the computation of a statistical similarity measure (SSM). In this work we used three SSMs, namely the absolute value of the cross correlation (also known as Pearson's coefficient) CC, the mutual information MI and the mutual information of the time series ordinal patterns MIOP (see Chap. 2).

Before computing these SSM values, every time-series are normalized to have zero mean and unitary variance. In addition, for the time-series of the oscillator phases we remove the linear trend.

In the following, for computing the MI measure, the probability distributions of values will be estimated via histograms with a number of bins given as $N_{bin} \sim \mathcal{O}(\sqrt{T/5})$ where T is the length of the time-series. This allows for the robust estimation of the joint probabilities. The MIOP measure was computed with ordinal patterns of $Q = 4$. Using larger Q values requires extremely long time series, since the number of possible ordinal patterns increases as $Q!$. In real systems, the length of observed time series is normally limited, and thus we restricted our simulations to the same constraints in order to obtain results that can be exported to real systems.

In order to reduce the effect of noise, the SSM values were averaged over 10 windows: each time series was divided in 10 non-overlapping sections each of length T and in each section the SSM was computed as a function of the lag time τ_{ij}. Then, the statistical similarity measure is defined as the maximum value with respect to the lag τ_{ij},

$$SSM_{ij} = \max_{\tau_{ij}} SSM_{ij}(\tau_{ij}), \tag{4.9}$$

being SSM either CC, MI or $MIOP$.

We use the maximum SSM value to allow for the existence of lag-time between the time series. While we could have estimated the SSM via a moment of the SSM(τ_{ij}) distribution (e.g., the mean value plus a certain number of standard deviations) we chose this approach because it has been used to infer functional climate networks (see, e.g.,[52]). The lag time distribution in our case is usually peaked around 0 (not shown).

When computing the SSM, the maximum value can depend significantly on the interval of values allowed for τ_{ij} [110]. Here we consider $\tau_{\max} = T/5$. Since lagging two time series reduces the number of available data to compute the SSM value, $\tau_{\max}$ was chosen as a compromise, in order to allow for sufficiently long lag-times while simultaneously keeping enough data to evaluate reliably the SSMs.

Network reconstruction

The computation of the SSM_{ij} between all pair of nodes leads to a $N \times N$ matrix from which the structural links can try to be inferred by a suitable thresholding of the matrix. In order to chose an adequate threshold, we compute the mean value of the matrix elements, M_S, and their associated standard deviation, σ as

$$M_S = \frac{1}{N(N-1)} \sum_{i \neq j} SSM_{ij}, \tag{4.10}$$

and

$$\sigma = \sqrt{\frac{1}{N(N-1)} \sum_{i \neq j} (M_S - SSM_{ij})^2}. \tag{4.11}$$

With these two quantities we define the threshold TH as

$$TH = M_S + n\,\sigma; \tag{4.12}$$

where n is a number that was heuristically tuned to obtain the best possible reconstruction, that is to obtain values of inferred links as high as possible keeping low the values of mistakes. With this threshold we prune the **SSM** matrix to reconstruct the network topology in the following way:

$$\hat{A}_{ij} = H(SSM_{ij} - TH_{SSM}); \tag{4.13}$$

where $\hat{\mathbf{A}}$ denotes the reconstructed adjacency matrix and H is the Heavyside step-function. The matrix $\hat{\mathbf{A}}$ is the inferred network, which is usually referred to as the functional network. Here we aim at analysing up to which extent the functional connectivity (represented by $\hat{\mathbf{A}}$) resembles the physical connectivity (represented by the adjacency matrix, $\mathbf{A}$). The strong hypothesis that we make here is that high values of SSM correspond to structural couplings. Of course this may be not always the case. For example, a high synchronisation among the oscillators will imply high values of correlations, regardless whether they are directly coupled or not. As another example, oscillators that are under common external forcing will have correlated dynamics, and thus they will show high degree of statistical similarity, even if there are no structural connections among them.

4.2 Results

To test the inference method described in the previous section, we consider a structural network **A** that is a fully random (Erdös-Rényi) network. In such a network the nodes are connected with a probability p, and the link density is equal to p.

We highlight at this point that the model presents five parameters: the number of oscillators, N, the coupling constant, K, the network density, p, the noise intensity D, and the width of the distribution of natural frequencies, Ω. In the simulations, unless otherwise specifically stated, $N = 12$, $p = 0.27$, $D = 0.05$ and $\Omega = 20p/N$. These values of N and p are chosen to fit the experiment with Rössler oscillators. Moreover, the value of p is large enough that, for $N = 12$,there are few disconnected nodes. The issue of having *few* disconnected nodes is not a problem, because we are also interested in checking if we are able to detect them (and in fact we do, as we show below we achieve perfect reconstruction of the Kuramoto network applying the method to the oscillators' frequencies).

The only free parameter left is the coupling strength K. For each value of K, we studied the system's dynamics by performing 100 stochastic simulations with different values of the oscillators' frequencies, network realizations, noise realizations, and initial conditions. The total simulation time was $2 \times 10^3 N/p$; the first half of the simulation was disregarded as a transient time, while the second half was used for the calculations. The data obtained in the second half was subsequently coarse-grained to mimic a measurement process that results in time-series of 10^4 data points which were divided in 10 non overlapping sectors of $T = 10^3$ data points.

Varying K we studied the performance of the reconstruction method in the nine cases that result from the combination of the three SSMs (CC, MI, MIOP) computed from the three time-series in each node: phases, frequencies and Y-projections (Kuramoto) or x_i variable (Rössler).

We expect the performance of the algorithm to be significantly affected by the value of the coupling parameter K [81]. On the one hand, sufficiently low values of K will lead the system to behave as an ensemble of uncoupled oscillators. On the other hand, when K is large enough the system will synchronize and the trajectories will not be distinguished leading to an all-to-all fully connected matrix. In both limits, the functional network will not recover the organization of the structural connections. Thus, it is for intermediate values of K where the **SSM** matrix can be expected to contain information about the physical connectivity of the system.

Before presenting the analysis of the network reconstruction, it is relevant to discuss the behaviour of uncoupled ($K = 0$) Kuramoto oscillators in the presence of Gaussian noise.

Stochastic Uncoupled Kuramoto Oscillators

With $K = 0$ the dynamics of the nodes are independent Brownian processes with different trends but with the same noise strength. As it can be seen in Fig. 4.2, in this limit it is possible to obtain high values of CC between the time series of the oscillators' phases, despite the fact that the oscillators are uncoupled. From Fig. 4.2, it is also evident that the maximum CC value is independent of the integration time. It is also independent of the noise strength, because the cross correlation is normalized with respect to the variance of the time-series, that is, the intensity of the noise. An explanation of this behaviour is provided in the Appendix. Thus high correlation values between the phases of two nodes, even with a large amount of data, cannot be regarded as a reliable signature of structural interactions.

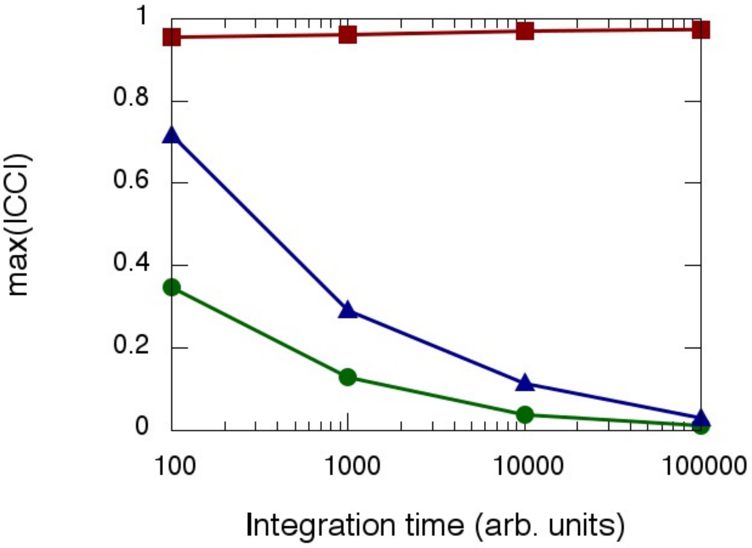


Figure 4.2: Maximum cross-correlation values between any two pair of uncoupled Kuramoto oscillators ($K = 0$). We analyze 120 stochastic simulation of (4.1) and compute max($|CC|$) as a function of the number of integration steps. We plot the values obtained from the analysis of the time series of: phases θ (red squares), instantaneous frequencies $\nu = \dot{\theta}$ (green circles), and vertical projections, $Y = \sin(\theta)$ (blue triangles).

For the other two observables, the instantaneous frequencies ν and the Y-projections, instead, high correlations vanish as the length of the time-series increases.

Stochastic Coupled Kuramoto Oscillators

When switching on the interactions between the nodes, i.e., when using $K > 0$, we can try to infer the structural network via the thresholding of the SSM matrices.

Figure 4.3 displays the results, when the phase θ is the observed variable. We evaluated the performance of the method as a function of K by plotting the fraction of correctly inferred links (true positive ratio, TPR), the fraction of wrongly inferred links that are not present in the structural network (false positive ratio, FPR), and the difference between the density of the reconstructed functional network and the density of the structural network. Each value is the average performed over 100 stochastic simulations, and the error bars are computed as left and right standard deviations respect to the mean.

The K range has been chosen in order to cover the transition from non synchronisation to almost full synchronisation.

Figure 4.4 presents the results obtained from the analysis of the instantaneous frequencies, ν_i. In this case the three SSMs present a critical K above which they are able to infer correctly the structural network, with the fraction true positives reaching one and the fraction of false negatives being zero, both of them with null dispersion over many stochastic simulations. However, CC still reveals to be the best SSM, since it has a lower critical K together with both higher values of TPR and lower values of FPR. Most importantly, we note that, in spite of the fact that the oscillators interact through their relative phases (and not their frequencies), the network recovery can be indeed perfect.

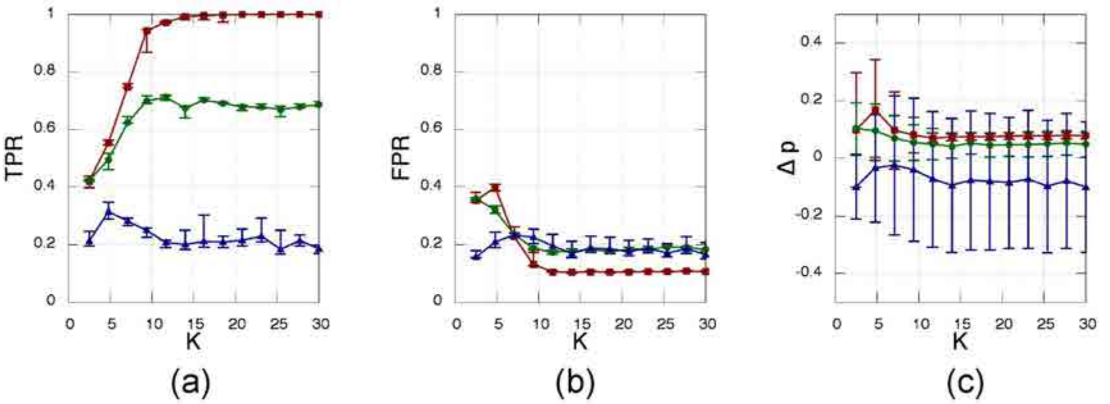


Figure 4.3: Performance of the inference algorithm for the three similarity measures considered (CC based quantities are depicted in red squares, MI in green circles and MIOP in blue triangles), applied tp the phase time-series, θ_i. (**a**) Fraction of true positives, (**b**) fraction of false negatives, and (**c**) difference between the inferred density and the actual density of the network, Δp, as a function of the coupling strength, K. Error bars are computed as the left and right standard deviation of the values obtained from over 100 stochastic simulations with different noise realisations, initial conditions, structural network realisations and natural frequencies.

In Fig. 4.5(a) we show the values of the CC matrix in increasing order for one realisation of the frequencies of Kuramoto dynamics. We report the values for two different K values, one for which we have poor reconstruction (lower) and one for which have perfect reconstruction (higher). We also report the associated threshold with the same color. As it can be seen, in the case of perfect reconstruction (high K), the CC values develop a gap that is not present in the poor reconstruction (low K) case. The fact that the threshold for the high K is placed just in the middle of the gap show that this abrupt change in the CC values separates links from non-links. This behaviour is consistent with what found in Ref. [81]. In 4.5(b) and (c) we show how this gap disappears if the MI is used instead of CC or if the length of the series is reduced, all cases of non-optimal reconstruction of the network.

In Fig. 4.6 we test the inference method with the other indirect variable, the Y-projection.

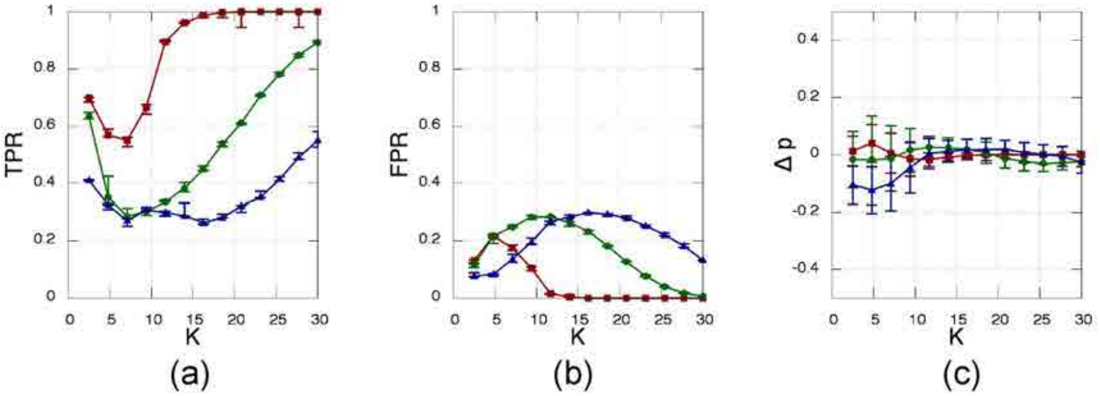


Figure 4.4: Same as in Fig. 4.3 but the statistical similarity measures are applied to time series of the instantaneous frequencies, ν_i.

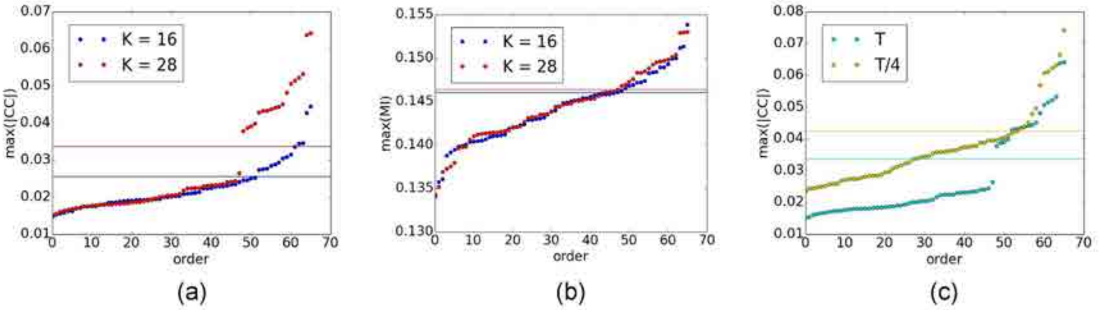


Figure 4.5: **(a)**: CC values calculated from the frequencies of the Kuramoto oscillators for one realisation of the dynamics with coupling strength $K = 16$ (blue) and $K = 28$ (red). The values are ordered from the smallest to the biggest. The threshold used for computing the functional connectivity is reported as a horizontal line of the same colour of the associated CC values. As it can be seen in the case of high K (*i. e.* in the case of perfect reconstruction) the CC develops a gap in the middle of which is placed the threshold. **(b)**: same as (a) but for MI values. No gap is present now. In this case the network reconstruction is worse than in the CC case. **(c)**: same as in (a) for $K = 28$. In green the original time-series are used, while in yellow the sime series are just 1/4 than the original ones. Also in this case tha gap disappears, and the network reconstrucion get worse. These results are in agreement with [81].

In this case the method fails almost completely, being unable to efficiently reconstruct the structural network. However the inferred density is often consistent with that of the target network – although a certain dispersion in these values is present. Table 4.1 summarizes the best performance for all the possible combinations of SSMs and observables.

It is important to note that the presence of noise is fundamental, preventing the complete synchronisation of the oscillators that would make the trajectories indistinguishable. Thus, even with a relatively high value of the synchronisation parameter r, we can still achieve perfect reconstruction.

Table 4.1: Best performance for a network of 12 randomly coupled Kuramoto oscillators for $p = 0.27$ and $\Omega = 1$. The best true positive fraction (BTP) and the associated best false positive fraction (BFP) are chosen as the TPR and the FPR closest to the (1,0) corner in the (TPR,FPR) plane. The associated K value is reported togehter with the corresponding SSM to which these best results are associated.

Observable	BTP	BFP	K	SSM
Phases	0.992	0.104	14.0	CC
Frequency	1.000	0.000	20.8	CC
Projections	0.583	0.251	4.8	MIOP

Figure 4.7 displays the influence of increasing the system's size, N, while keeping the link density constant ($p = 0.27$). As it can be seen, the only SSM showing a certain degree of resilience to the network size is the cross-correlation, which keeps decent values of TPR and FPR also for higher N.

In Fig. 4.8 we report effect of varying the network density while keeping constant the network size ($N = 12$). As it an be seen from the figure, the quality of the reconstruction decreases when increasing the network connectivity.

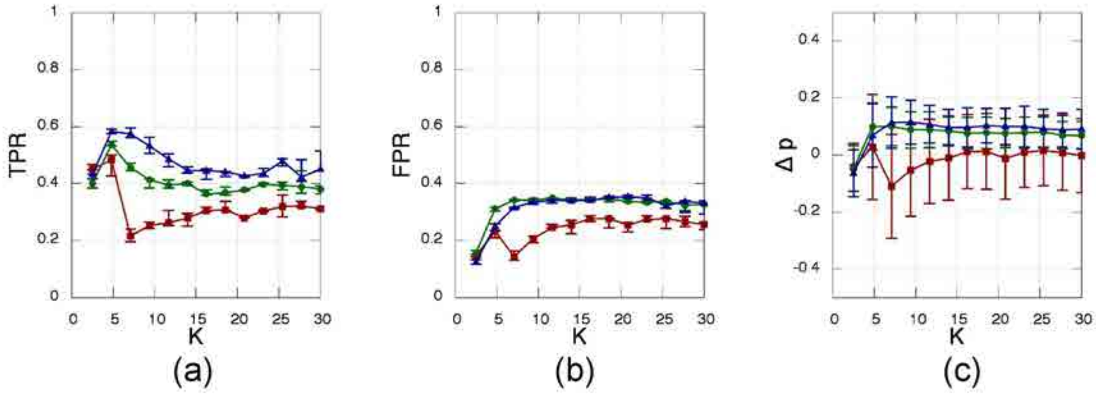


Figure 4.6: Same as in Fig. 4.3 but for Y_i time-series.

Experimental electronic Rössler oscillators

The results of the inference method applied to the data generated from the Rössler electronic oscillators are plotted in Fig. 4.9 (experimentally recorded time-series of the oscillator's variable $x_i(t)$ with $i = 1\dots12$), Fig. 4.10 (oscillator's phases computed via the Hilbert transform) and Fig. 4.11 (oscillator's instantaneous frequencies computed from their phases). As we can see, the three SSMs behave very similar when compared among them, but have different performance depending on the observable variable.

Figure 4.9 shows the results when the observable is $x(t)$. We can see how, after a critical coupling around $K_c = 0.02$, the TPR and FPR quickly converge to 0.9 and 0.15 respectively,

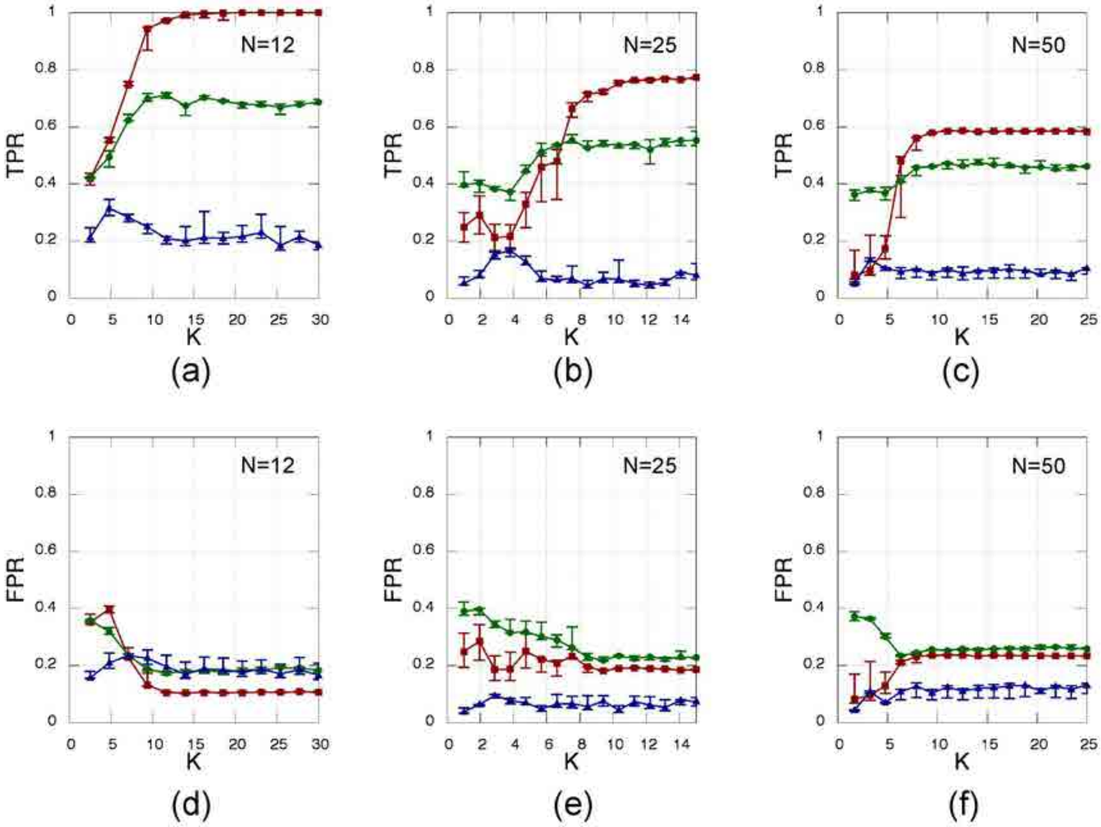


Figure 4.7: Same as Fig. 4.3 but for three different network sizes. The left column, **(a, d)**, is computed using the same amount of oscillators as in Fig. 4.3, while the other two columns refer to higher network dimensions, namely 25 oscillators **(b, e)** and 50 oscillators **(c, f)**.

leading to a good network reconstruction, although it is not possible to fully retrieve the structural network, unlike in the Kuramoto case.

When using the phases (Fig. 4.10) or the instantaneous frequencies (Fig. 4.11) to infer the network, the performance is slightly lower than the one obtained with $x(t)$. Table 4.2 summarizes the best performance for all possible combinations of SSMs and observables.

Table 4.2: Best performance results of network inference of 12 randomly coupled Rössler chaotic oscillators. The best true positive fraction (BTP) and the associated best false positive fraction (BFP) are chosen as the TPR and FPR closest to the (1,0) corner in the (TPR,FPR) plane. The associated K value is reported together with the corresponding SSM value.

Observable	BTP	BFP	K	SSM
Phases	0.950	0.170	0.110	CC
Frequency	0.89	0.19	0.040	CC
Projections	0.890	0.130	0.025	MI

Since the dynamical properties of the three Rössler oscillators components are different –

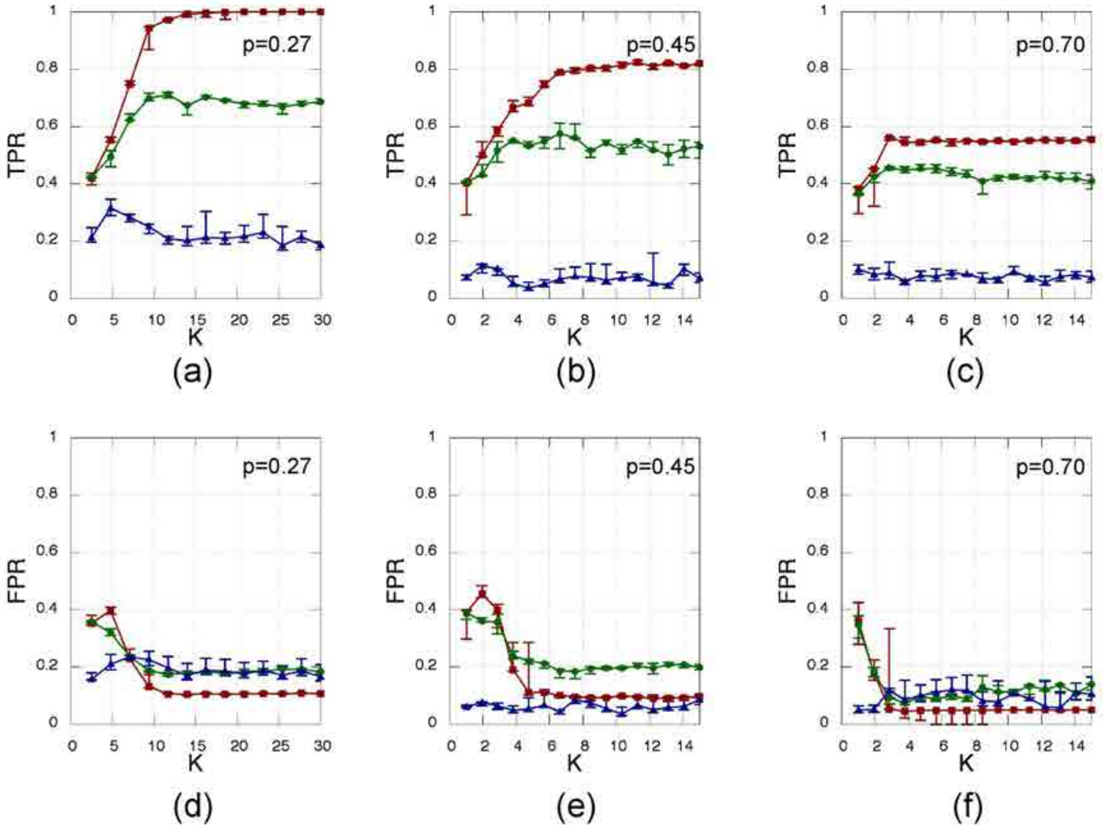


Figure 4.8: Same as Fig. 4.3 but for three different network densities. The left column, **(a, d)**, is computed using the same amount of links as in Fig. 4.3, while the other two columns refer to higher network connctivity, namely $p = 0.45$ **(b, e)** and $p = 0.70$ **(c, f)**.

especially those of the z variable – we speculate that the algorithm performance could be different when applied to y or z variables, although we cannot check this it because for technical limitations it was possible to record only the x_i time-series.

4.3 Conclusions

We have analysed the relation between inferred functional networks and the underlying structural networks in two different systems, a simulated set of Kuramoto oscillators and an experimental set of Rössler chaotic electronic circuits. We focused in the dependence on the strength of the interaction among the system components, in the similarity measure used, and in the type of variable analysed.

For the first time we have contrasted nine inference methods (using three types of variables — amplitude, phase and frequency, and three similarity measures — cross correlation, mutual information, and ordinal mutual information), and we presented a detailed comparison of their performance using both synthetic and empirical data.

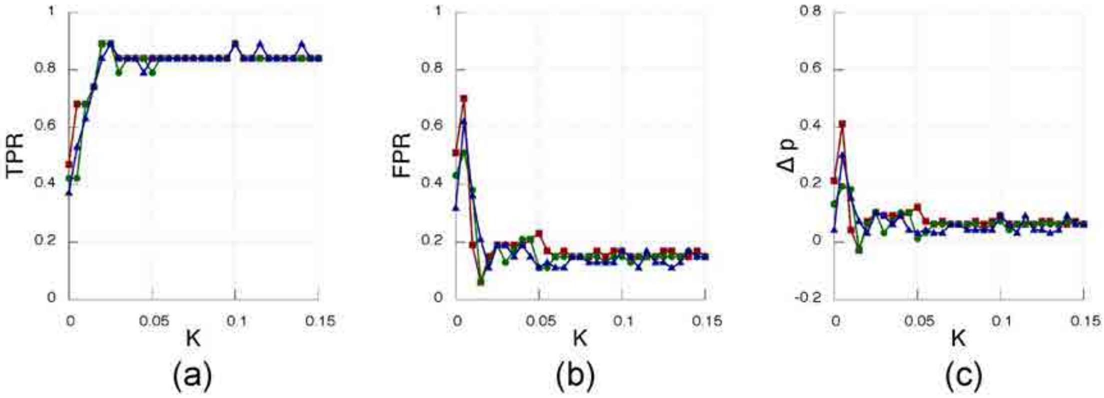


Figure 4.9: Performance of the inference algorithm for experimental data corresponding to a random network of 12 Rössler chaotic electronic circuits, for the three SSMs (CC based quantities are depicted in red squares, MI in green circles and MIOP in blue triangles). Time series contain the evolution of the x variable. (**a**) fraction of true positives, (**b**) fraction of false positives, (**c**) difference between the inferred density and the actual density of the network, Δp, as a function of the coupling strength K.

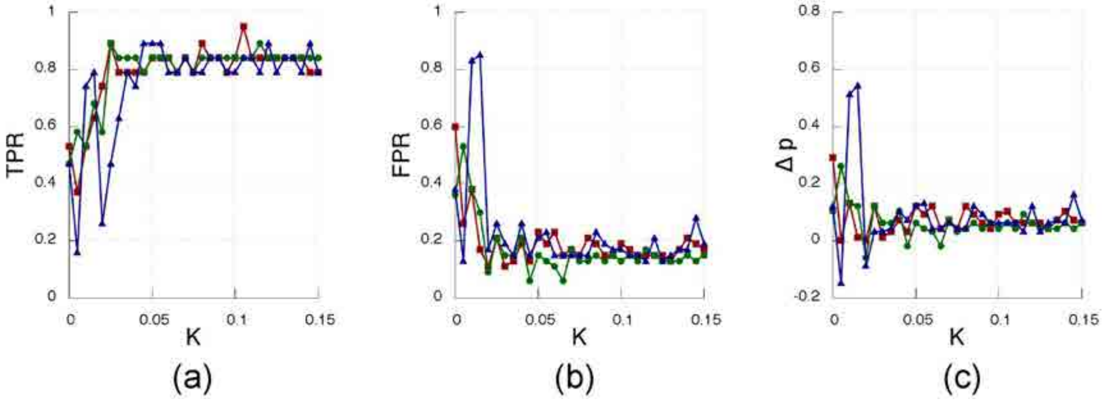


Figure 4.10: As in Fig. 4.9 but for the phases computed via a Hilbert transformation.

The first important result that we show here is that it is indeed possible to find regimes of the dynamics in which the functional connectivity of the system perfectly mimics the structural one (Kuramoto oscillators), or in which the former is a very good approximation of the latter (Rössler oscillators).

A second conclusion of our study is that functional networks obtained from thresholding the cross-correlation matrix perform better with respect to the two other non-linear measures used (the mutual information computed from histograms of values and the mutual information computed from histograms of ordinal patterns). We can see that, for Kuramoto oscillators, CC gives better results than both, MI and MIOP, thus the coarse grained covariance matrix, in a range of K values, is a good approximation of the adjacency matrix of the network. The fact that the linear measure, CC performs much better than the non-linear one is explainable

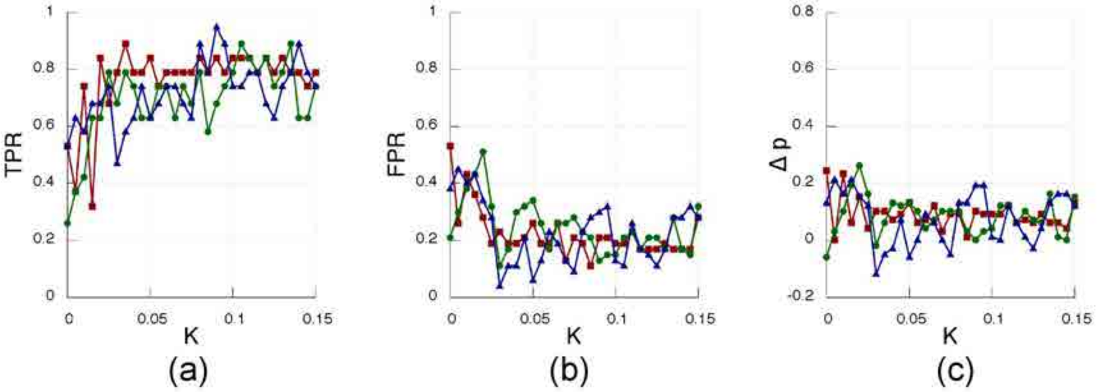


Figure 4.11: Same as Fig. 4.10 but for the instantaneous frequencies.

at least for the high values of K. In fact, for high K the dynamics is almost synchronised, it means that the interaction term can be considered approximately linear. In fact, if K is high enough such that $R(t) \sim 1$, we have $\theta_i \simeq \theta_{j \neq i}$ and the Kuramoto equation reads

$$\mathrm{d}\theta_i \simeq \left(\omega_i - K\, d_i\, \theta_i + \frac{K}{N} \sum_j A_{ij} \theta_j \right) \mathrm{d}t + D\, \mathrm{d}W_t^i. \tag{4.14}$$

where d_i is the fraction of nodes connected with node i. It is clear that in general, in a linear system, the covariance matrix contains a great amount of information regarding the system.

Also for the frequencies the CC performs better than the non-linear measures and the reason for this can be found again in eq. 4.16. When we correlate two frequencies ν, we multiply directly the interaction terms in the r. h. s. of eq. 4.16, where the connectivity appears in a linear way. In this way the covariance matrix reflects directly the adjacency one and we have an accurate reconstruction. Moreover, we shown that the CC for the frequencies develops a gap between the values of connected and unconnected pair of nodes, and this is in ultimate analysis the reason for which the thresholding procedure retrieves the network without any mistake.

We also shown that, for low values of K the FPR is higher than for high values of K, an effect that can be understood in terms of the results presented in the previous sections about the possibility of high cross-correlation values for the phases of uncoupled oscillators.

A third conclusion of our study refers to the influence of inferring the functional network by means of an indirect variable. Specifically, we found that in the Kuramoto oscillators the use of an indirect variable, such as the Y projection, can lead to errors in the network reconstruction, although the density of the inferred network is consistent with that of the structural network. On the other hand the use of the frequencies led to a better result than the one sing the phases. For the Rössles, instead, the indirect and direct variables give similar results.

We have also shown that the system's size and its connectivity density can compromise the network reconstruction. When exporting the methodology to large networks, such as climate networks, we do not expect the network density to be a serious issue, since these networks are usually sparse; however, their large size could result in poor performance.

The problem of extending these results to real-world systems, however, remains an issue. We stress the fact that the use of a uniform, global threshold of the SSMs values could be a drawback if exporting the methodology to systems in which the dynamical units (or the strengths of their interaction) are heterogeneous; however, a possible way to overcome this drawback is to use a local threshold instead of a global one, i.e., a threshold defined for each pair of time-series. Namely, for each pair of time-series, a certain number of surrogate time-series can be built (for example with bootstrap techniques) and a distribution of surrogate SSMs for each pair of time-series is obtained. The threshold then can be defined locally as a certain percentile of the surrogate distribution. Therefore, it will be interesting for future work to use local thresholds to study the reconstruction of the physical connectivity when the system displays heterogeneity (in the dynamical units or in their interactions) and to develop a reliable, unambiguous way to determinate the threshold, either local or global.

For future work it will also be interesting to analyse how the network topology affects the performance of the inference method. We suppose that in the scale free case, the architecture detection could be more difficult. The reason is that, in our study, we are using a method that assumes implicitly an homogeneous structural architecture – in fact, as stated before, we used a constant threshold to prune the SSMs matrix – an hypothesis clearly violated in a scale free network, which presents highly connected nodes as well as poorly connected ones. Instead, in small world networks (that is in networks that display a high clustering coefficient but a low average path length, as social networks) it is harder to guess which might be the performance of the functional network approach. For sure, in the Kuramoto case, since we shown that the good performance of the algorithm relies in the quasi-synchronisation of the nodes, we expect to experience good reconstruction for K values bigger than the ones presented here because it is well known that random networks are more easily synchronisable than small-world ones. [111]

It would also be very interesting to consider the possibility of improving the performance of the inference methodology by using a “multi-layer” approach, that is by considering the functional networks obtained from the intersection of those obtained from different variables.

Also link dependence on time is an issue that deserves attention, especially in the framework of climate and neuronal dynamics, where climate tipping points or neuronal plasticity change the structural connectivity in time; it would be interesting to understand how this behaviour if reflected in the functional connectivity, and which property of the structural one could be inferred using the associated functional network.

5 Unravelling the Community Structure of the Climate System

In this chapter we propose two novel methods for detecting the community structure of the global climate system: the first one is based on time-lags between climate variables recorded at different locations (calculated as in Chapter 3); the second one exploits symbolic analysis to estimate climate similarities in different locations. These two methods reveal complementary information about the community structure underlying climate dynamics: while the communities identified via time-lags reveal the main differences in thermal inertia of the various regions of Earth, symbolic analysis reveals a division in four main communities with similar dynamics: the ENSO basin, extra-tropical oceans, tropical oceans and extra-tropical land masses.

To perform the analysis we consider the same dataset as in Chapter 3.

The results presented in this chapter are summarised in [112].

5.1 Seasonal Cycle Communities

In the first method of community detection, we will analyse the properties of the seasonal cycle of the database. As in Chapter 3 we first determine the mutual lag between the seasonal cycles of each pair of points, ℓ_{ij}. Because the seasonal cycle is by definition periodic, we search the maximum in the τ window $[0, 11]$ [85, 54].

In principle, all the information about mutual lags can be displayed in just one map of the lags from one region k to any other region in the world. In fact, the lag among any two regions i and j is computable through the use of a third region k as

$$\ell_{ij} = (\ell_{ik} + \ell_{kj}) \bmod 12, \tag{5.1}$$

and thus all the information is contained $\vec{\ell}_k$, the other maps being obtainable through eq. 5.1. However, because we consider monthly-averaged data, ℓ_{ij}, ℓ_{ik} and ℓ_{kj} are integer number of months, and thus, because of round-off error (the real lags are not integer number of months) eq. 5.1 will not hold for all the triples (i, j, k).

In order to identify the regions that have well-determined lags among them, we chose a reference node i, and, for each other node j, we test eq. 5.1 for all the possible ks. If the relation is satisfied in more than 50% of the cases we consider the ℓ_{ij} a well determined lag, otherwise we do not assign it any value. The resulting map is plotted in Fig. 5.1a for a reference region in continental Europe and in Fig. 5.1b for one in southern South America. In this way, all the areas sharing the same color present a seasonal cycle in phase, and the lag among the different areas is easily computable form eq. 5.1. The white areas are those in which the lag with the reference point is not well-defined, and in fact they lay at the border between areas at different mutual lag.

As we can see, while the insolation (that is the responsible of the seasonal cycle) is zonally symmetric, the map of the lag times presents a certain zonal heterogeneity. In particular the oceans are delayed of one month respect to the landmasses. In the eastern boundaries of the oceans this delay reach two months (even 3 months in the NINO3.4 area). While the one month delay is easily explained in terms of thermal inertia of the water respect to the land, the large delay in the eastern oceans has no immediate explanation.

The method described here can be used also for other variables. In Fig. 5.2 for example we report the result obtained for the geopotential height at 500 hPa. In this case, due to the fact that the seasonal cycle of the geopotential height is highly non-linear and very heterogeneous, the areas with not well-defined lags among them increase. However some leading features can be found, as the 6 months symmetry between the two hemispheres, or the presence of the stationary atmospheric responses in the southern hemisphere. Also, in the northern hemisphere, two regions in the North Atlantic have a not well defined lag; these regions are consistent with the NAO pattern that on long scales can act as a source of noise for the lag determination.

5.2 Anomalies Communities

Removing the annual cycle from each time-series, results in zero-mean anomalies (SATA). In this case the time-series will show very different statistical properties, with the tropics having a very slow evolution, respect to the noisier extratropics. Thus as in [74], to make the statistical properties of the database more uniform, we remove the fastest fluctuations of the anomalies through a running mean having a two years window. Then the time-series are converted in ordinal patterns (see Sec. 2.2.7).

Using this symbolic time series, we would like to compute the relevant climatic properties of the series at each geographical location. Since the anomalies are usually very noisy, the simple probability distribution of the symbols is not enough, because they tend to be fairly uniform. In contrast, the number of transitions among the symbols can display non trivial

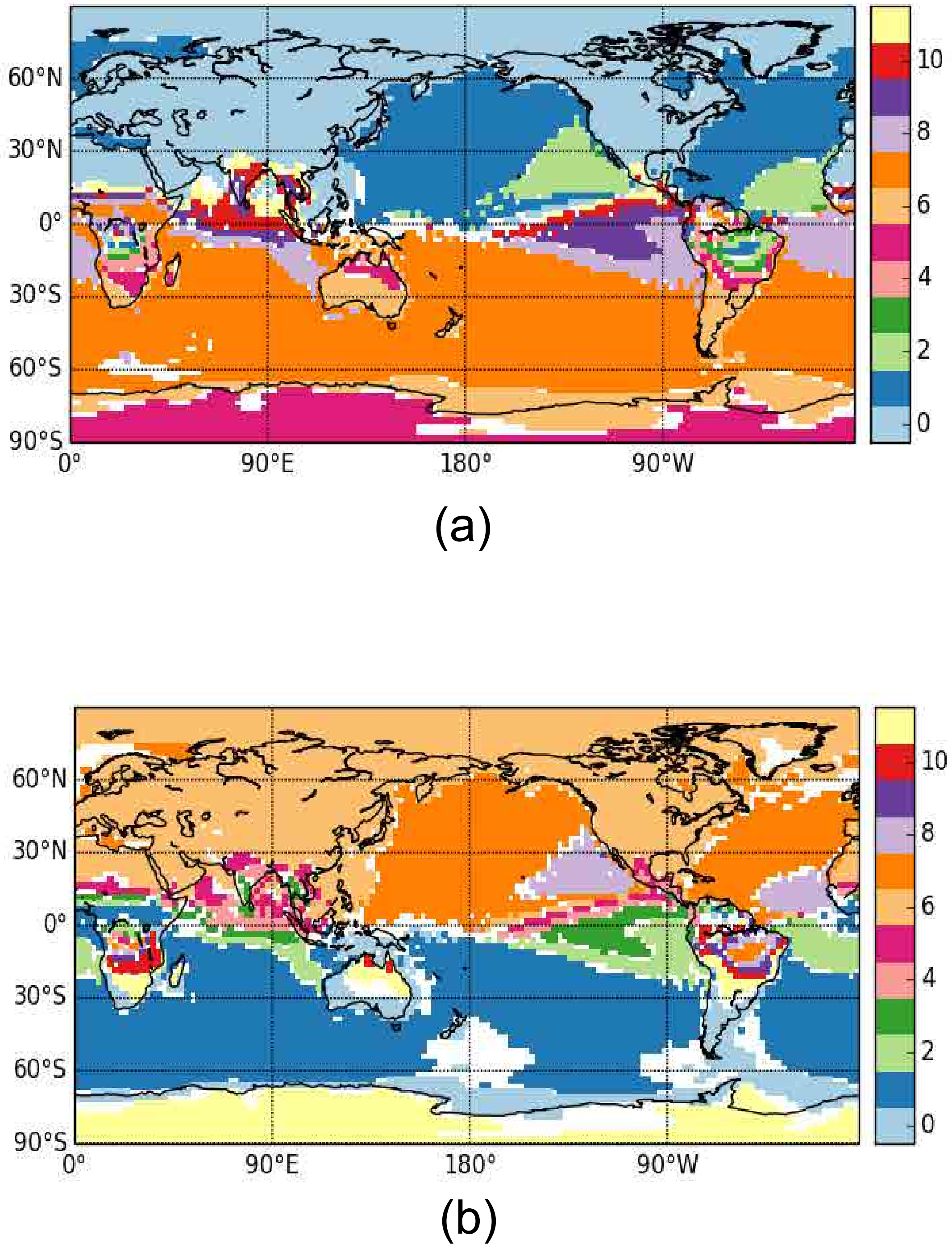


Figure 5.1: Maps of seasonal cycle communities obtained from SAT time-series. Regions that share the same color have a synchronous seasonal cycle, while the mutual lag between each region can be computed subtracting the numbers associated with each color. The first map is obtained using as a reference point k located in continental Europe, while the second one is obtained using a reference node in southern South America. As it can be seen the community structures in the two maps are very similar.

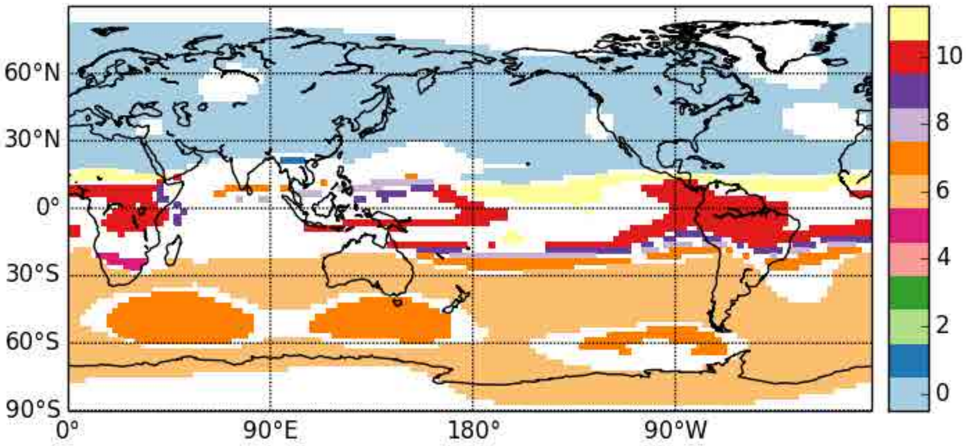


Figure 5.2: As Fig. 5.1a but computing the lag times from timeseries of geopotential height at 500 hPa.

properties and encode much information about the climate evolution in a single location. Thus, we convert all the time-series in transition matrices among the $Q!$ possible symbols [113]. In particular, given two symbols α and β, for a certain symbolic time-series $s(t_i)$ we have

$$M^s_{\alpha\beta} = \frac{\#(s(t_i) = \alpha \text{ and } s(t_{i+1}) = \beta)}{T-1}, \tag{5.2}$$

that is the number of $\alpha \rightarrow \beta$ transitions over all the possible transition in the time-series. Of course to have a good estimation of the transition frequencies the length T of the time-series must be much longer than the number of all the possible transitions, $Q!^2$. Thus, with $T = 773$ months, we use $Q = 3$.

In this way we are encoding all the information about the seasonal evolution of the time-series in N matrices that we can now investigate with climate networks tools.

To do so we define a weighted network in which the weights of the links are defined as

$$w_{ij} = \left(\sum_{\alpha\ \beta}^{Q!} (M^i_{\alpha\beta} - M^j_{\alpha\beta})^2 \right)^{-1}, \tag{5.3}$$

in this way nodes with similar transition matrices are strongly connected, and vice versa.

As usual in climate networks, to reduce the influence of noise, only the strongest links are considered, *i.e.*, we disregard the links with a weight below a certain threshold, defining a binary adjacency matrix as

$$A_{ij} = \mathrm{H}(w_{ij} - W), \tag{5.4}$$

where H is the Heavyside step-function.

We chose the value of W in a way that the final network is not too sparse or too connected, in particular each node is connected on average to the 5% of the globe.

Ath this point, we apply the algorithm of community identification known as Infomap [114], also used in [74]. With this algorithm we obtain the community structure presented in Fig. 5.3. As it can be seen the algorithm divides the world in 8 areas, labelled with different colors. These areas share the same statistical properties of the time-series evolution. In fact, it can be seen that continents in the two hemispheres share the same label ("0") even if there is not direct coupling among them. Also, a huge coherent area is detected by the algorithm in the ENSO basin, while the oceans are divided in tropical and extratropical.

It has to be noted that, using the measures of statistical similarities previously used in the field of climate networks (such as cross-correlation and mutual information) we would not have obtained such community structure. In fact, to belong to the same communities two nodes must belong to the same group of strongly interconnected nodes, and in traditional approach the links are prominently local, thus connections across hemispheres are impelled. In this way it would be impossible, for example, for points in the northern and southern extra-tropics to belong to the same climatological communities. Our approach instead overcomes this problem.

It is interesting to see how this communities are related to those found in Fig. 5.1 through the seasonal cycle. There are come border among different communities that are indeed shared by the two sets, such as the extra-tropical coastlines, or the separation of northern from southern Australia and of southern South America from the rest of the continent.

The Infomap algorithm automatically converges towards a certain number of communities that in principle cannot be controlled, since they are defined by the network structure. In particular, on our case, this is crucially dependent on the network density that is modified by the threshold used in Eq. 5.4 for the adjacency matrix definition, W. Increasing the network density makes the network itself to look more like a giant coherent cluster, and the Infomap algorithm will converge to a smaller number of communities. Decreasing the density will break the network in many small parts, and Infomap detects them as many separate communities. As an example, decreasing the network density such that each region is, on average, connected to the 2% of the globe, results in 29 different communities detected by the algorithm.

This method can also be applied to other variables, such as the geopotential height. Figure 5.4 displays the result for the geopotential height anomalies at 1000 and 300 hPa. As it can be seen, increasing the height of the field implies a more zonal distribution of the communities: at 300 hPa the tropics form a belt that is well distinct from the extratropical areas, which belong to the same community, the two are separated by strip-like communities probably a signature of the subtropical jet. At 1000 hPa, instead, the effect of tropical convection is dominant, and breaks the low latitudes in two areas, the Maritime Continent together with the ENSO basin (perhaps a signature of the Walker circulation), and the rest of the tropics. The extratropics instead are all grouped in the same community, regardless the presence of landmasses.

5.3 Conclusions

In this work we presented two new methods to identify climate communities using the properties of climatological time-series. These methods take into account the seasonal cycle and the anomalies with respect to it.

Regarding the first method, we divided the world in 12 areas of synchronous seasonal cycle. Nodes belonging to areas of the same colour posses their seasonal cycle on phase. As we can see from Fig. 5.1, the spatial distribution of these regions does not reflect completely the insolation spatial distribution. In particular, areas at the same latitude are out-of-phase even of two months (over the oceans' eastern boundaries). In this way we obtained a map of the

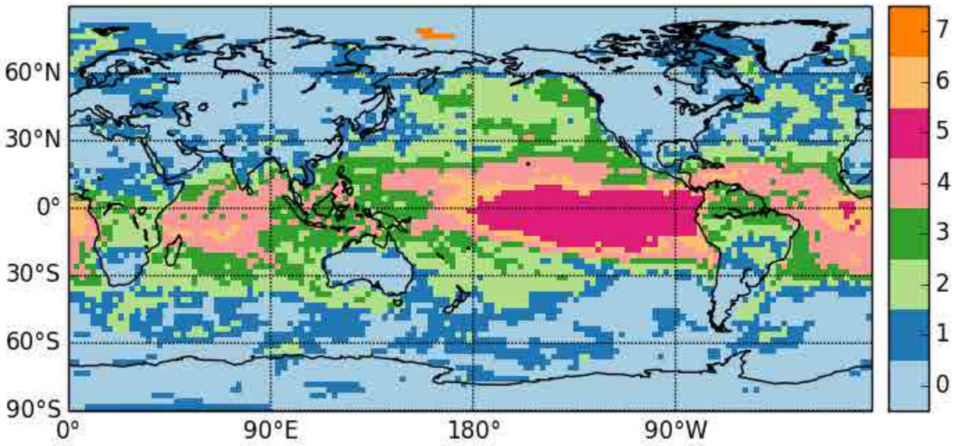
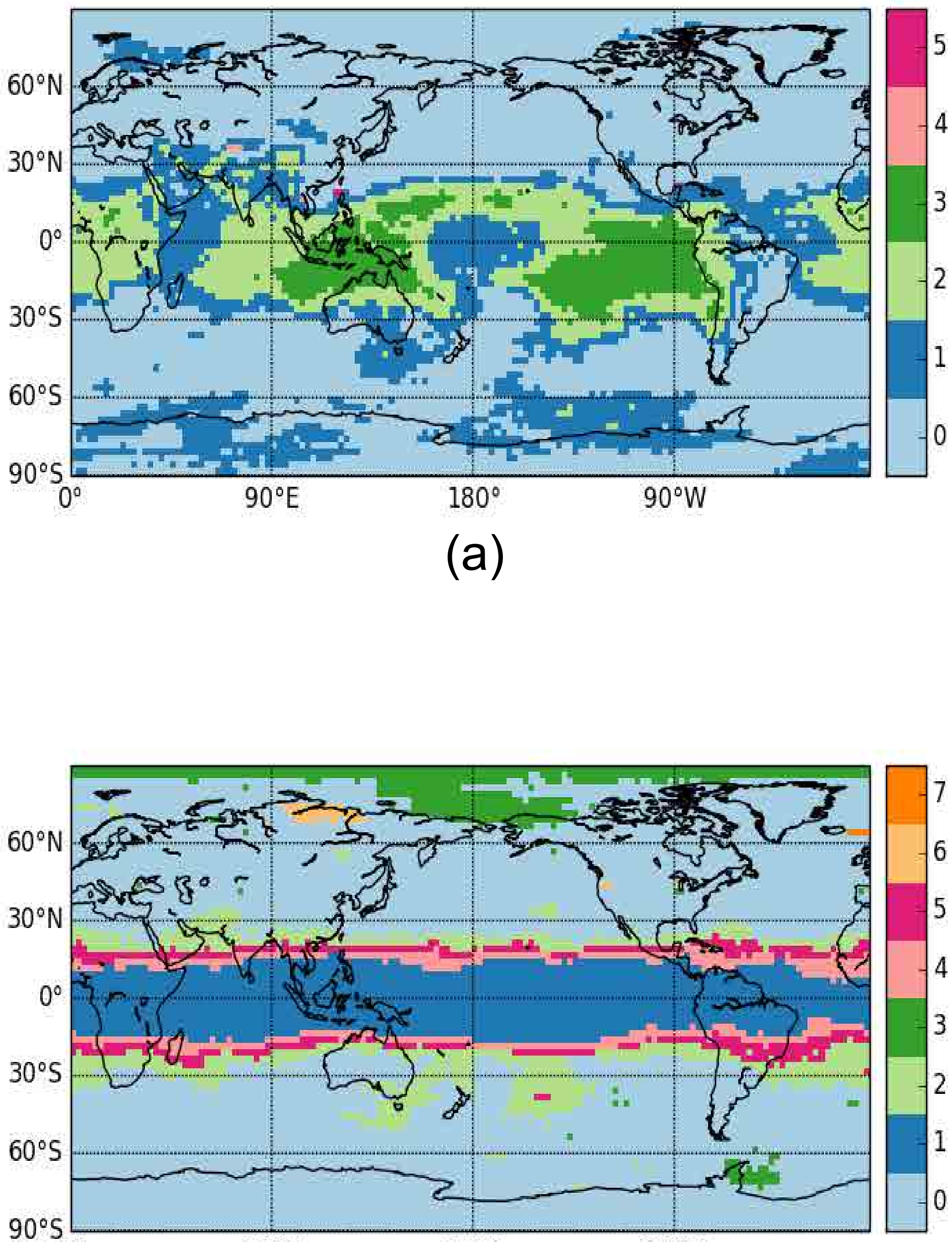


Figure 5.3: Community structure from SAT anomalies. Nodes depicted with the same color belong to the same community. The world is divided in 4 macro-communities: extratropical continents and oceans, tropical oceans and ENSO basin.



(a)

(b)

Figure 5.4: As Fig. 5.3 but for geopotential height anomalies at 1000 hpa (top) and 300 hPa (bottom).

thermal inertia of the various regions of Earth.

In the second method we analysed the statistical properties of each time-series by means of symbolic analysis, constructing for each time-series a transition matrix. Then we constructed a climate network linking nodes with similar matrices and we analysed the network with a community detection algorithm. Regions belonging to the same community are strongly interconnected, thus meaning that they have a similar climate. In fact, is clear from the communities inspection, that the world is separated in four macro-areas: ENSO basin, extratropical and tropical oceans and extratropical big landmasses.

We also tested the method applying it to the geopotential height at various pressure levels, uncovering community structures that are consistent with the main atmospheric large scale phenomena.

6 Granger causality analysis of air-sea interaction in the South Atlantic Convergence Zone

In this chapter we use the network approach to study the air-sea interaction in the region of the South Atlantic Convergence Zone (SACZ) (cfr. Sec. 1.4). Granger causality (see Sec. 2.3.1) will be used as a measure of directional coupling to build the network between the ocean and the atmosphere.

The results presented in this paper are collected in [115].

6.1 Data

As a proxy for precipitation we consider the vertical velocity at 500 hPa, ω, while the ocean state is characterized by the sea surface temperature (SST). Both datasets are daily mean data provided by ERA INTERIM Reanalysis [51], ranging from Dec 1979 to Mar 2013 on a horizontal grid with 1.5 degrees of resolution. We restricted the analysis to austral summer months (Dec-Mar) when the SACZ shows strongest activity. Anomalies were calculated first removing the annual cycle by subtracting the average value for each day, and then normalizing the series to have unit variance.

6.2 Air-sea connectivity

We start by focusing on the region $[20N, 40S] \times [70W, 20E]$, and compute the GCE (see Sec. 2.3.1) at each grid-node, both in the direction SST $\rightarrow \omega$ and $\omega \rightarrow$ SST. In this way we want to estimate the local interaction between the two systems, *i. e.* the interaction between a grid point in the ocean and the one placed in the atmosphere just above it. The GCE values at 99% confidence level are plotted in fig. 6.1. As it can be observed in (**a**), the main two zones in which the ocean forces the atmosphere are the deep tropics and the subtropical waters off Brazil. The deep tropics is expected to be a region of oceanic forcing (e.g. [116]), and the methodology consistently captures this fact. The ocean off southeast Brazil corresponds to the SACZ region and is within our region of interest. Moreover, the right panel of fig. 6.1 clearly shows that the atmosphere strongly forces the surface ocean also off Brazil. Thus, this ocean region can play both an active and a passive role, providing a hint of the presence of complex feedback loops in the area [31, 32].

Since the atmosphere and the ocean interact on larger spatial scales – not only locally – the calculation of the local GCE value provides limited information. Therefore, we computed the cross-GCE, in which the GCE is evaluated, in the two directions, for every couple of ocean-atmosphere nodes. In this way we are investigating also the presence of non-local forcing.

To represent this result we built a two-layer network [87] based on significant GCEs: every node of the network is a grid point of the oceanic or of the atmospheric layer, and it is linked to points in the other layer only if their GCE is significant at 99% confidence level. The resulting network is a bipartite one, that is, a network composed of two set of nodes in which there are links only from one set to the other. Furthermore, those links are directed and allow for feedback loops.

An easy way to represent such a network is through the so-called *adjacency matrix*, **A**. The dimension of **A** is $N \times N$, where N is the number of grid points for each layer, and it is composed of only 0s and 1s. Namely, if two nodes i and j are connected then $A_{i\to j} = 1$ and it is 0 otherwise. Since the GCE is a directional measure, **A** is a non-symmetric matrix, that is, $A_{i\to j}$ may be different from $A_{j\to i}$. For example, a location i in the ocean can be forcing a point j in the atmosphere without being forced by j itself.

Using the adjacency matrix, we can compute the area weighed connectivity (AWC) for the oceanic and the atmospheric layers:

$$AWC_i = \frac{\sum_{j=1}^{N} A_{i\to j}\cos(\Lambda_j)}{\sum_{j=1}^{N}\cos(\Lambda_j)}, \tag{6.1}$$

where i corresponds to the nodes in one layer, j to the nodes in the other layer, Λ_j is the latitude of node j and N is the number of grid points in the studied region. The links were weighted with $\cos(\Lambda_i)$ to correct the bias induced by the different areas represented by the different grid points.

Thus, nodes with high AWC correspond to points of one layer that exercise a broad forcing on the other layer, and this way, the AWC reveals to be a powerful way to quantify the spatial ranges of interaction of the two systems. In particular, high values of AWC indicate points that have a significant Granger causality with many other points in the other layer. However, the AWC does not provide information about the actual value of this Granger causality. A very weak forcing but acting over a large area, in fact, has a high value of AWC but a very low value of local GCE, while a strong forcing acting over a very narrow area, has a low AWC together with a high local value of GCE. Note, however, that the AWC depends on the chosen region, because atmospheric long-range teleconnections, that extend beyond the region considered, are not taken into account.

The AWC maps are depicted in fig. 6.2. The AWC map associated to SST → ω forcing (fig.6.2 a) presents two maxima. The first one is on the equator, at about 20W and is coincident with the tropical component of the map of local GCE. However, the largest AWC is placed offshore the southern Brazil coast and corresponds to the second maximum in the map of local GCE (fig. 6.1a). Thus, the relative magnitude of the signal in the equatorial and subtropical regions is opposed with respect to the local measure, suggesting that the SST forcing in the tropics is stronger but mainly local (vertical), while in the SACZ region it is weaker but non-local. In fact, while in the deep tropical Atlantic we have small values of AWC and high values of local GCE, in the subtropical part of the ocean the situation is the opposite, with many points presenting high AWC but relatively low values of GCE (low, compared with the tropical ones). This might be related to the fact that SST anomalies in the subtropics vary coherently on a larger spatial scale than in the equatorial Atlantic. We remark that the GCE values used to build this maps are statistically significant at 99% confidence level.

Regarding the AWC map of ω → SST (fig.6.2 b), the connectivity is generally large over the whole region, stressing the dominant role of atmospheric circulation in forcing SST anomalies. Moreover, there is a large highly connected region south of 30S which corresponds to a largely non local pattern of connectivity (we can see a very feeble signal in the local map in fig.6.1b). The location of the maximum suggests that it could be related to the influence of the south Atlantic anticyclone on SST through, for example, wind-induced changes in latent heat fluxes.

Taken together, these maps reveal that the SACZ region displays complex air-sea interactions in agreement with previous studies, e.g. [23] and [32]. Nonetheless, it is important to note that the AWC maps provide information only about the spatial range of the interactions but not about their strengths. Moreover, the ω field is very noisy including several types of phenomena that need not be directly related to the SACZ dynamics. Thus, in order to gain more insight about air-sea interactions in the region the next section focuses on the leading mode of coupled variability.

6.3 Directionality of air-sea coupling

The leading mode of covariability between SST and ω was computed using Maximum Covariance Analysis (MCA, [117]). It explains 8% of the cross covariance and is depicted in fig. 6.3, left column. It represents the intensification and northward shifting of the SACZ pattern respect to the climatology, accompanied by a cooling of the SST in between the two lobes of the dipole, with maximum cooling below increased convection. As mentioned in the introduction this dipole pattern has been shown to be the leading mode of variability in the region on several time scales [24]. Note that the region of ocean cooling is consistent with the one displayed in the AWC and in local GCE maps in figs. 6.1 and 6.2. Moreover, the map of points connected to the region of maximum AWC off Brazil reveals a structure consisting in two parallel bands oriented NW-SE, very similar in shape to fig. 6.3 (not shown). This suggests that this mode represents a bi-directional interaction of the two fields.

Focusing on the dynamics of the first mode, which represents the SACZ variability, allows to filter a good fraction of the weather noise and to truly concentrate on the interannual variability of the SACZ. In particular, SACZ events will evolve differently depending on the air-sea interaction.

To find the direction of the coupling between the oceanic and atmospheric patterns in the leading mode, displayed in fig. 6.3, we calculated the GCE using the corresponding Principal Components (PCs) over the whole period of study. The similarity of the values obtained (0.019 in the SST $\rightarrow \omega$ direction and 0.021 in the $\omega \rightarrow$ SST one, with a relative difference of less than 10%) indicates that the air-sea interaction is very important in this mode, and confirms the results of the previous analysis, once again showing the presence of feedbacks between the atmosphere and the subtropical south Atlantic. We stress that both these values are significant at the 99% confidence level.

The second mode of covariability (7%), instead, shows the imprint of the south Atlantic anticyclone, especially in the SST component, suggesting that this is likely an atmospheric driven mode (not shown). This hypothesis was confirmed by the calculation of the GCE of the corresponding Principal Components (0.021 and 0.027, respectively), and the mode is not studied further here.

We next calculated the GCE for every year of the PCs (fig. 6.4), which allowed to classify individual years into four regimes: ω forcing SST (9 years) and SST forcing ω (4 years) if only one GCE is statistically significant, double forcing (4 years) if both GCE values are significant, and neutral (17 years) if no GCE value is significant at 90% confidence level. The patterns associated with the SST forcing and neutral cases are shown in fig. 6.3. In the neutral case SST anomalies are small and rainfall anomalies are strongest over land. In the SST forced case, on the contrary, there are large SST anomalies with a dipolar structure and rainfall anomalies are largest over the oceanic portion of the SACZ. The next section will study the evolution of these events.

From the analysis it remains unclear whether the high number of Neutral years is a consequence of the small number of data that compose the annual datasets (four months, that is 121 days), or indeed, there are no significant interactions between the atmosphere and ocean. The composite analysis presented below suggest the latter.

Considering that El Niño can significantly affect the SACZ region [27, 118], we tried to relate the occurrences of the four different regimes with the occurrence of El Niño as characterized by the NINO3.4 index. Separating the whole period in El Niño, La Niña and neutral years – according to the Oceanic Niño Index (ONI) computed with NOAA ERSSTv3b data – we can observe that the only significant combination ($p < 0.1$) is between neutral years and pure atmospheric forcing (see tab. 6.1). In the table the p-values are computed using a binomial distribution, taking into account both the cardinality of ENSO and PC1 classification sets.

Thus, the analysis suggests that the circulation anomalies induced by El Niño and La Niña modify the regional dynamics such that the number of events of pure atmospheric forcing in the SACZ region decreases. At the same time, most of the years in which the SST forcing plays a role in the dynamics of the events (cases SST→ ω and SST↔ ω) correspond to either El Niño or La Niña years (6 out of 8 cases). This is discussed further in the next section.

6.4 Evolution of events

To study the evolution of the events in each of the four categories defined before, we constructed composites using the PC of ω corresponding to the leading mode. Thus, by construction all cases present a dipole in the vertical velocity field in the SACZ region, but the evolution of anomalies will be different. To smooth out the fields the MCA is performed after a running-mean filtering over three days. The composites are calculated as the difference between the average of the 5 days presenting the largest positive peaks in the ω PC, and the average of the 5 days presenting the largest negative peaks taken from the subset of years corresponding to each category. We then repeated this computation for [-4:4] days preceding and following the day of the maximum value. The composites for the four cases (figs. 6.5, 6.6, 6.7 and 6.9) are constructed for ω, SST, net surface heat fluxes, 850 hPa winds and geopotential height at 200 hPa.

6.4.1 Case ω →SST

Composites of the ω field (fig. 6.5, last row) clearly show that the dipole starts developing at time -4 days and has almost disappeared at time +4 days fig. 6.5, last row. In upper levels the development of the dipole is accompanied by a travelling wave train from the extratropical Pacific sector that bends into the SACZ region as the geopotential composites show(fig. 6.5, second row). The lower level wind anomalies are consistent with this wave train and there is clear cyclonic circulation in the SACZ that generates convergence (divergence) in the region of upward (downward) motion. This behaviour is consistent with the literature which suggests that midlatitude Rossby wave trains modulate the activity of the SACZ on time scales of a few days, such that intensified SACZ events are accompanied by upper-level throughs to the southwest [119, 120, 23]. The atmospheric forcing is clearly seen in the maps of net surface heat flux, which show large heat losses from day -4 to the peak cooling the SST below next to the brazilian coast (fig. 6.5, first row). Regions with southeasterlies induce strong latent and sensible heat flux loses between [55W-30W,30S-50S] up to day 0; after the peak the latent and sensible heat flux anomalies change sign (not shown). The pattern of shortwave and long wave radiative fluxes follow the shape of the omega anomalies such that increased ascent decreases shortwave and increases longwave fluxes at the surface through enhanced cloudiness (not shown). Note that after the peak the net surface heat flux changes sign such that at day +4 the ocean is gaining a significant amount of heat which weakens the created SST anomaly off southeast Brazil (fig. 6.5, third row). The net heat flux anomaly is dominated by the latent heat because the oceanic extension of the SACZ weakens rapidly after the peak.

6.4.2 Case Neutral

The evolution of atmospheric anomalies in this case is similar to the previous one, and it is reported in fig. 6.6. In the geopotential composites there is an upper level wave train that progresses from the extratropical Pacific and bends into South America. However, the path of the wave in this latter case is more meridionally oriented so that the center of the cyclonic anomaly in the SACZ region is closer to the coast allowing surface level winds to penetrate further into the continent. For example, at the peak there are southerly winds up to the Equator while in the $\omega \rightarrow$SST case the southerly winds reach only to 10S. As a consequence the low level wind anomalies are able to transport moisture more effectively from the Amazon basin toward the northern sector of the SACZ. This results in ω anomalies that have largest values over the continent and weaker over the oceanic portion of the SACZ (fig. 6.6, last row). Surface heat fluxes are also negative below the SACZ, although are weaker than in the previous case, and cool the SST off Brazil (fig. 6.6, first and third rows). However, in this case the net surface heat flux does not change sign after the peak below the SACZ. The negative heat fluxes below the enhanced portion of the SACZ are due primarily due to decreased shortwave radiation and increased latent heat flux (fig. 6.10). The oceanic loss of latent heat below an enhanced SACZ provides more moisture for rainfall and thus supports the maintenance of the oceanic extension of the SACZ. Moreover, the surface wind anomalies after the peak keep blowing from the ocean to the continent providing enhanced moisture supply. Thus, in this case the persistence of the oceanic portion of the SACZ is larger than in the $\omega \rightarrow$SST case (see also fig. 6.10). Therefore, in the neutral case there seems to exist two way air-sea interaction. However, since the largest anomalies occur over land this process does not significantly impact the evolution of the event and thus this category does not show any significant air-sea interaction according to the GC measure. It would be very interesting to study the land-atmosphere interaction in this category.

6.4.3 Case SST$\rightarrow \omega$

As shown in table 6.1, most of these cases occur during El Niño years. According to the literature El Niño induces positive upper level geopotential height anomalies over tropical South America and positive sea level pressure anomalies in the tropical Atlantic during this season (e.g. [121]). In addition El Niño forces extratropical wave trains that propagate from the Pacific toward South America in arch-like paths, a mechanism that weakens during summertime – for a review see [122]. In the events of this category we see mainly evidence for an influence of the tropical Pacific via extratropical wave forcing (fig. 6.7, second row). We found that cases with decreased rainfall in the northern side of the SACZ dipole correspond to El Niño, thus the SST composite constructed here shows negative anomalies in the eastern Pacific (fig. 6.7, third row). The composite of the geopotential height at 200 hPa shows evidence of a wave train coming from the extratropical Pacific with negative and positive centres south of 50S that arches toward the tropical Atlantic to the east of South America. This wave train has a stationary component since the centres of the anomalies move slowly from day -4 to day +4 (fig. 6.7, second row). In addition to the wave train, there is a positive geopotential height anomaly centred at (40S,90W) that persist during the whole event. This is in clear contrast to the previous two cases (e.g. $\omega \rightarrow$ SST and neutral, figs. 6.5 and 6.6) and indicates the influence of the tropical Pacific forcing. Superposed on the stationary wave component there is a travelling wave train that shifts the pattern westward and intensifies the atmospheric anomalies. The resulting path of the wave is quite south and the surface wind anomalies in the SACZ region are located mainly over the ocean. Thus, the ω anomalies in fig. 6.7 are strongest in the oceanic portion of the SACZ with only weak continental extension. Moreover, the oceanic dipole of ω anomalies persists up to day +4, albeit weakening with time

To further illustrate the distinct behaviour of this case with respect to the previous one, in fig. 6.8 we report the difference in atmospheric circulation anomalies between SST$\rightarrow \omega$ and neutral cases at the peak of the event. Consistent with the previous description, the geopotential difference shows a ridge over the continent and a trough over the ocean indicating that the negative geopotential anomalies in the SST$\rightarrow \omega$ are shifted over the ocean. Furthermore, due to its barotropic structure the ridge over eastern Brazil is associated with increased northerlies into the subtropics and easterlies in northeast Brazil. This induces moisture flux anomalies and the continental convection dipole observed in fig. 6.8 (right panel), that indicates smaller continental rainfall anomalies in the SST$\rightarrow \omega$ compared to the neutral case. In higher latitudes there are large differences in the geopotential field reflecting differences in the behavior of the wave associated with the two cases and mentioned before.

On the oceanic side, the Atlantic SST anomaly field is dominated by a strong dipole with cold anomalies north of 30S and warm anomalies to the south, except to the south of Uruguay. This SST pattern is conspicuous during summertime and is related to a strengthening of the south Atlantic anticyclone on seasonal time scales [123]. As in the previous cases there is strong heat loss off Brazil from day -4 to the peak, which cools the SST below. The southeasterly component of the surface winds provide more moisture to the region of enhanced ascent. Moreover, in the days following the peak there is oceanic latent heat loss below the region of anomalous ascent which, as in the Neutral case, implies anomalous moisture flux into the atmosphere which helps maintaining the ω (rainfall) anomalies over the oceanic portion of the SACZ up to day +4 (figs. 6.6 and 6.7). In this case, since most of the atmospheric anomalies are located over the ocean, the evolution of the rainfall dipole is influenced by the SST. These results suggest that of central importance for the SST to influence the evolution of SACZ events is that atmospheric anomalies have to be located mainly over the ocean. According to our results this means that the extratropical wave trains have to propagate mainly zonally and bend only slightly into the tropical Atlantic. Since 3 out of 4 years in this category correspond to El Niño years, it suggests that changes in the mean state caused by a warming in the equatorial Pacific may change the propagation properties of the wave trains through changes in the jet streams or blocking as found during winter by [124]. More research is needed to fully address this issue , including understanding what is special about these three El Niño events.

6.4.4 Case SST↔ ω

According to table 6.1, most of these cases occur during La Niña years and correspond to years with dipole events that have negative rainfall anomalies in the northern portion of the SACZ (fig.6.9). Contrary to the case SST→ ω, here the tropical anomalies induced by the equatorial Pacific ocean are more significant than the extratropical wave forcing (compare composites for the geopotential height in 200 hPa and for the surface winds in figs.6.7 and 6.9). As consequence there are northwesterly surface wind anomalies in northeastern Brazil that last the whole event and help the development of convergence there (Fig. 6.9, second row). In addition, over the continental SACZ region there is a positive geopotential height anomaly that, due to its baroclinic nature, has a low pressure associated at the surface that seems to play an important role in the event. Together with the low pressure center associated with a weak extratropical wave train they generate the cyclonic wind anomaly associated with the peak rainfall dipole in the SACZ. Note that at day -4 the convective anomalies over land and over the ocean develop separately and merge at day -2 when the surface low pressure center over the ocean moves westward. The travelling wave train in the geopotential composites is the weakest of all four categories having the maximum at day -4. At the peak the surface wind anomalies and oceanic heat loss are relatively weak (Fig. 6.9, first row). At day +2 there is a dipole in surface net heat flux anomalies with similar structure but weaker in magnitude than the one in SST→ ω case, but at day +4 heat fluxes are no longer significant. Finally, the region of ascent motion over the continent last from day -4 to day +4, the longest of all categories, probably due to the favourable conditions maintained by El Niño (Fig. 6.9, last row). Since anomalies are relatively weak and El Niño seems to play an important role it is hard to pinpoint how the ocean and atmosphere interact and more work is needed to understand this case. One possibility would be to repeat the Granger analysis after removing El Niño influence, but that is beyond the scope of this study.

As an additional analysis to compare the four regimes we report in Fig. 6.11 the time evolution of ω anomaly in the composites computed in two different boxes: one in the continental box [22S; 6S] × [55W; 35W], and another in an oceanic box at [40S; 30S] × [40W; 1W]. The two boxes, depicted in fig.6.3, were chosen to characterize the strength of the anomalies over the continent and the oceanic extension of the SACZ. The lag times in Fig. 6.11 are the same of figs. 6.5, 6.6, 6.7 and 6.9, thus the 0 lag corresponds to the convective peak. It can be seen that the overall evolution of anomalies over the continental region are quite symmetric about day 0, except in the case of ocean forcing. In this latter case the peak occurs at day +1 and anomalies tend to persist. Over the oceanic region anomalies are generally smaller in magnitude and have similar values at day -1 and day 0. The neutral case is the one characterized by weakest positive precipitation anomalies over the ocean and largest over the continent. The SST → ω case, on the other hand, shows weakest rainfall anomalies over the continent, but has largest persistence of anomalies over the ocean. Cases ω → SST and SST ↔ ω show a very similar evolution over the continent, except at the end of the event when cases SST ↔ ω have larger persistence. Over the ocean case ω → SST shows the largest anomalies, but do not persist for long: after the peak the vertical velocity rapidly goes back to zero at day +2. This is in sharp

Table 6.1: Division of the various years composing the database, respecting to the GC and ENSO classification. Some p-values are reported. The p-values are computed using a binomial distribution, taking into account both the cardinality of ENSO and PC1 classification sets.

	El Niño	La Niña	Neutral	Total
Neutral	4 ($p = 0.62$)	8 ($p = 0.19$)	5 ($p = 0.8$)	17
SST → ω	3 ($p = 0.10$)	0	1	4
ω → SST	2	1	6 ($p = 0.08$)	9
SST ↔ ω	1	3 ($p = 0.16$)	0	4

contrast with the case SST → ω where anomalies at day +2 are still large.

6.5 Conclusions

Air-sea interactions in the region of the South Atlantic Convergence Zone (SACZ) are studied using the Granger Causality Estimator (GCE) as a measure of directional coupling. A first important result is that this region has a large mutual air-sea connectivity during summertime. Introducing a new method that combines Granger Causality with Maximum Covariance Analysis, we have been able to identify four regimes, depending on the direction of the air-sea interaction. In particular, we identified years in which the coupling is mainly directed from the ocean to the atmosphere, years in which the coupling is from the atmosphere to the ocean, years of mutual interactions and years of no significant coupling.

We studied the evolution of the convective events related to these different regimes looking at atmospheric and surface oceanic fields and found that in all cases an extratropical wave train plays a major role in forcing convection in the SACZ, as suggested previously by several authors. Moreover, we found that the path of the wave seems to be instrumental in determining the evolution of the SACZ event. For example, if the path of the wave bends significantly over South America the largest anomalies develop first and are more intense over the continent, the oceanic extension of the SACZ is weak and there is no air-sea coupling (Neutral cases). On the other hand, in ocean-forced cases, the extratropical wave train is further to the south, so that the strongest convective anomalies develop over the oceanic portion of the SACZ, allowing the ocean to force atmospheric anomalies that tend to persist longer than in other regimes. We also found that these events are more frequent during El Niño years.

Further study of the SACZ using this methodology combined with tailored numerical experiments that, for example, isolate the effects of El Niño are needed to further understand the complex air-sea interactions in this region.

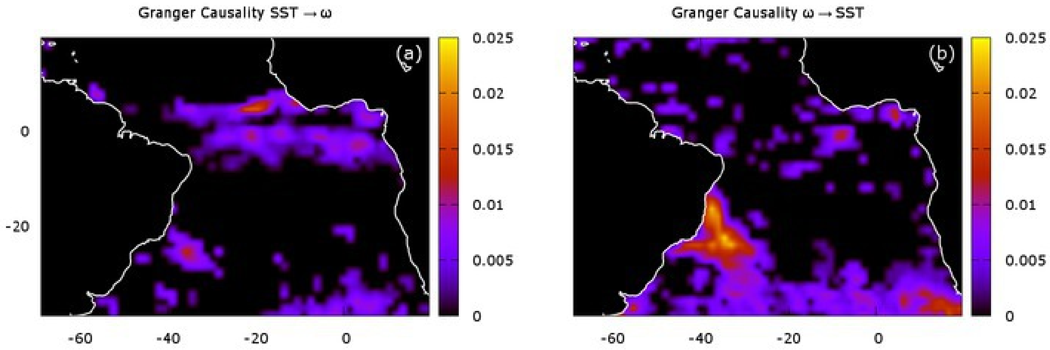


Figure 6.1: Local Granger Causality Estimator (GCE) for SST → ω (**a**) and ω → SST (**b**). Only values significant at 99% confidence level are reported.

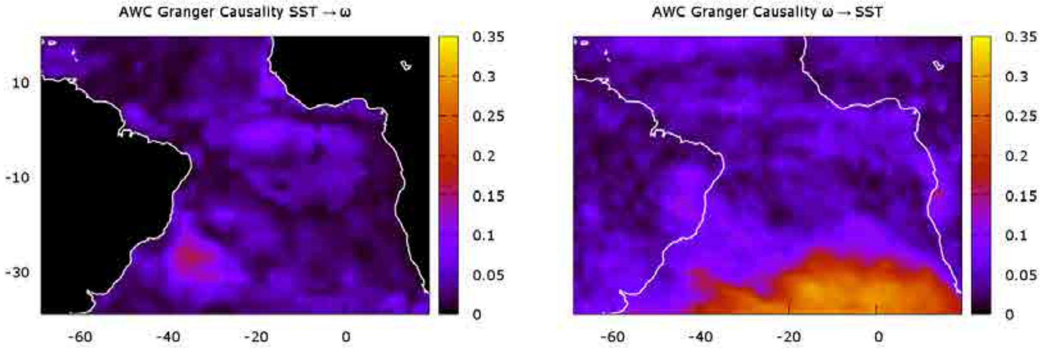


Figure 6.2: AWC for GC-based bilayer climate network composed of SST and ω datasets. (**a**) AWC computed with SST → ω GCE. (**b**) AWC computed via ω → SST GCE. Only GCE values significant at 99% confidence level have been used to compute the AWC.

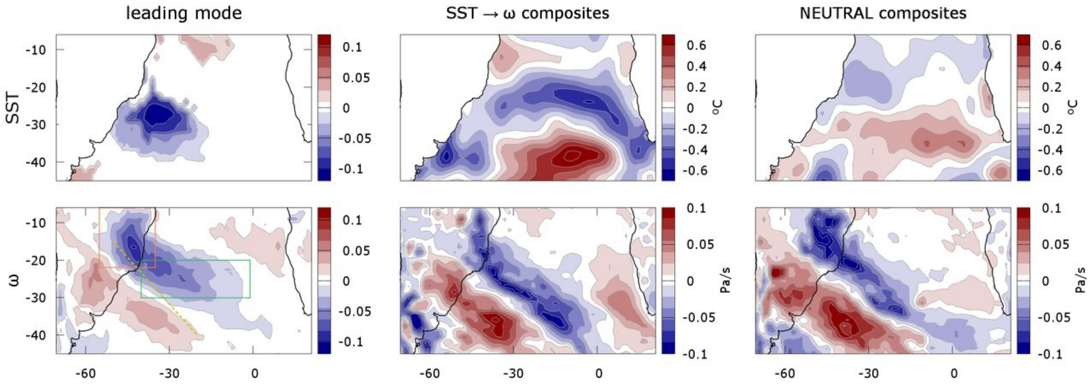


Figure 6.3: (**left column**): Leading mode of coupled variability related to the ω and SST fields. Only significant values at 95% are reported. The dashed line represents the climatological mean position of the SACZ. (**central column**): Composites for convective maxima of oceanic forcing years (SST→ ω regime, see section 5 for details). (**right column**): Composites for convective maxima of years without significant air-ocean coupling (neutral regime, see section 5 for details). In the bottom left panel the two boxes used for fig.6.11 are depicted.

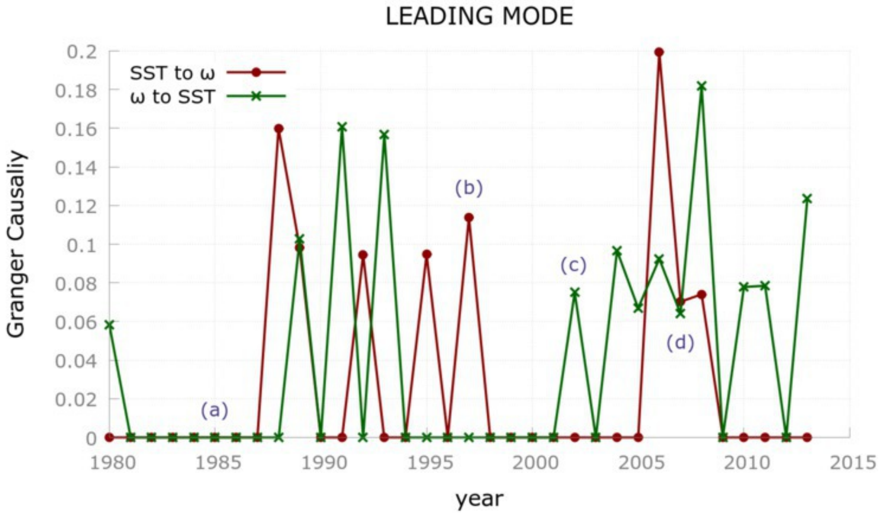


Figure 6.4: GCE between ω and SST PCs, relative to the leading mode of covariability. Non-significant (at 90%) GCE values are set to 0. (**a**): example of years displaying of non-significant coupling, *i. e.* none of the two GCE values is significant. (**b**): example of a year of oceanic forcing; only the SST to ω GCE value is significant. (**c**): example of a year of atmospheric forcing; just the ω to SST value is significant. (**d**): a year of mutual forcing; both the values are significant.

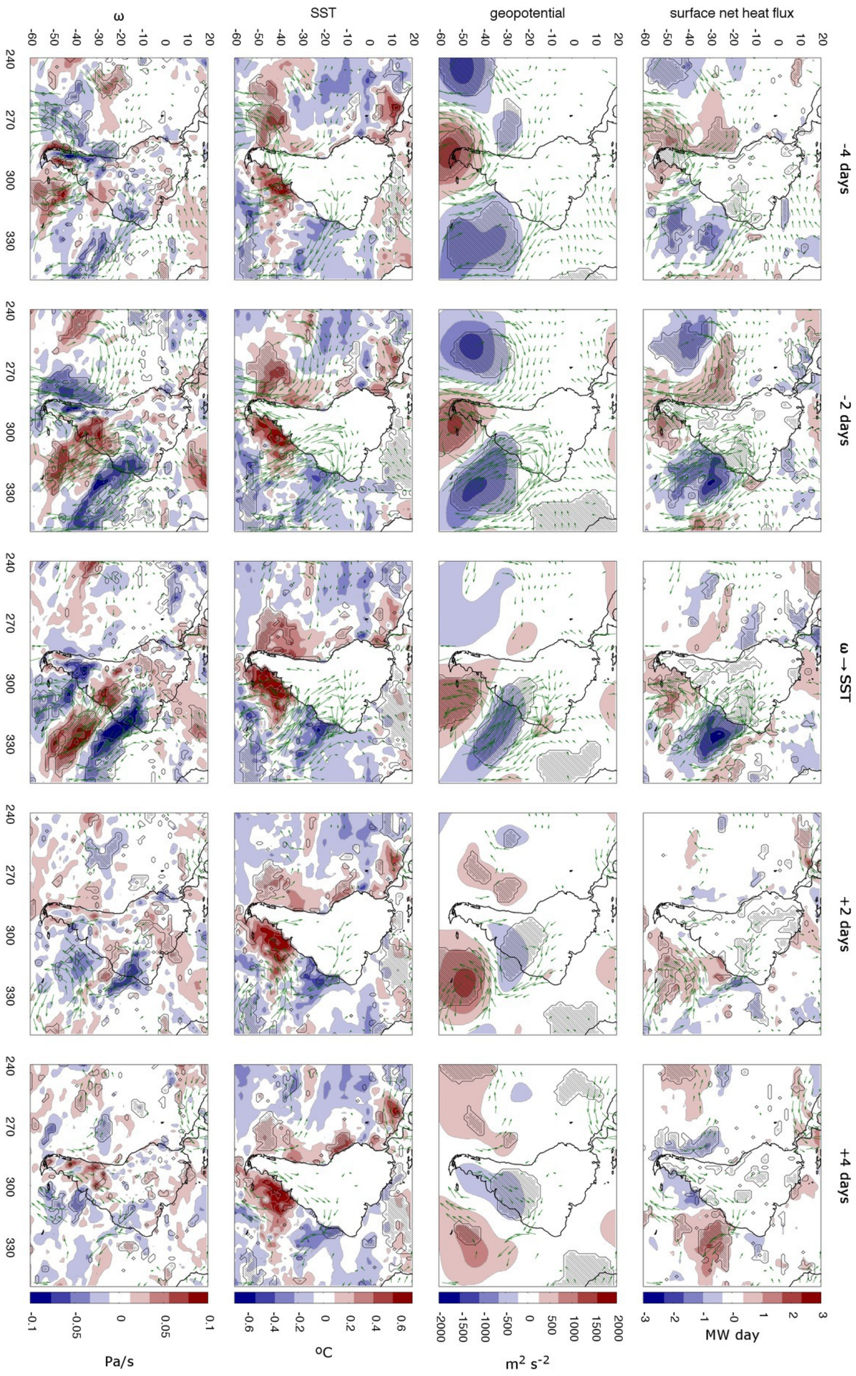


Figure 6.5: Lagged composites of ω anomalies at 500 hPa, SST anomalies, geopotential anomalies at 200 hPa, and surface net heat flux anomalies, for $\omega \rightarrow$ SST cases. Wind anomalies at 850 hPa are also included as vector field in all the maps. Black shading marks areas of 90% significance, moreover only arrows related to winds anomalies significant at 90% are reported.

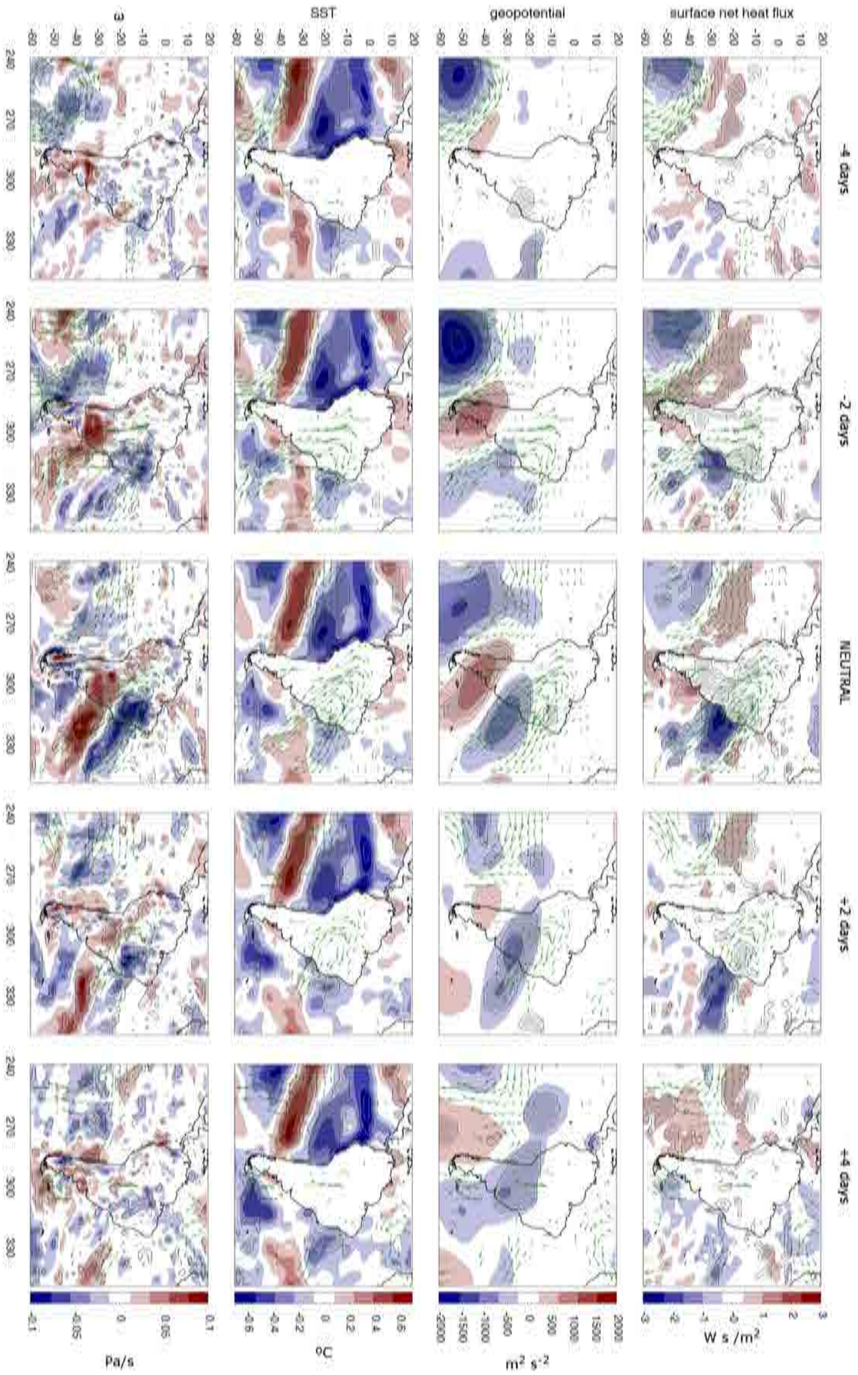


Figure 6.6: Same as Fig. 6.5 but for neutral case.

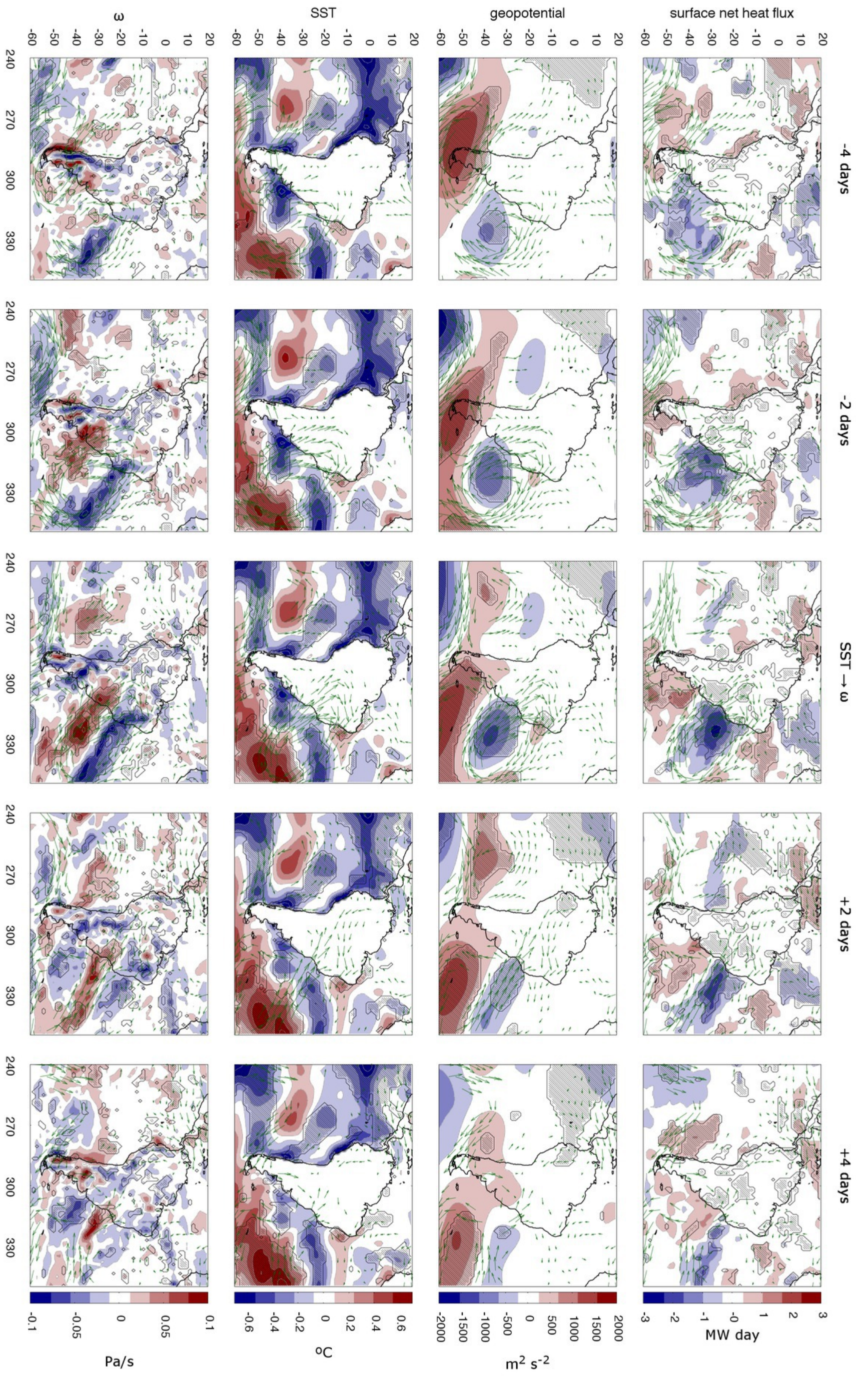


Figure 6.7: Same as Fig. 6.5 but for SST → ω case.

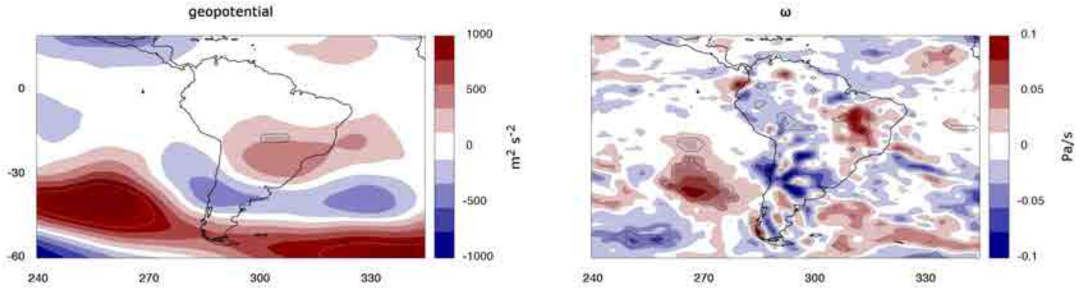


Figure 6.8: Differences among the lag-0 composite maps of the two cases SST → ω and neutral. (**Left**): differences between the maps of geopotential anomalies at 200 hPa composites. (**Right**): differences between the maps of ω anomalies at 500 hPa composites. Black shading marks areas of 90% significance.

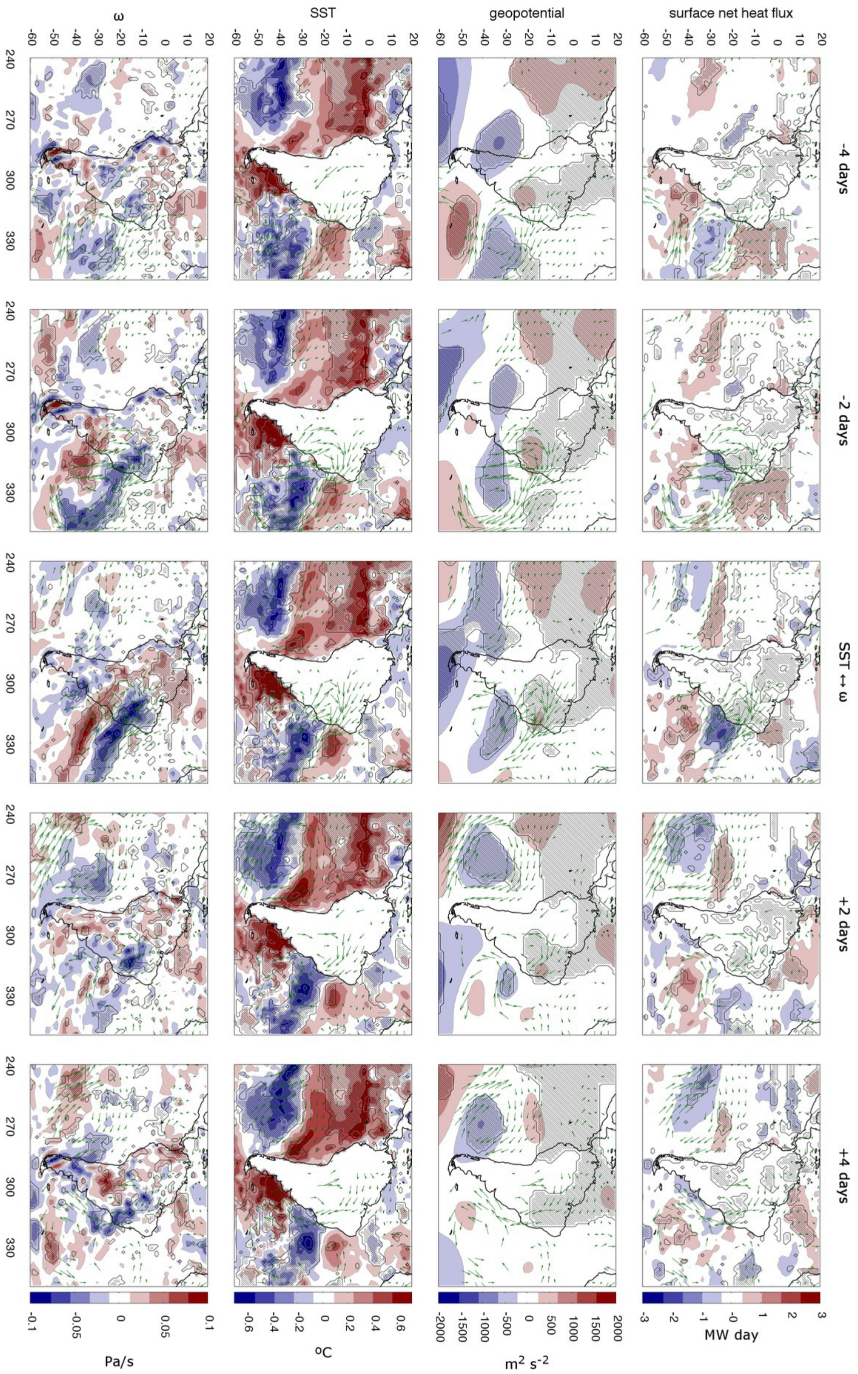


Figure 6.9: Same as Fig. 6.5 but for SST ↔ ω case.

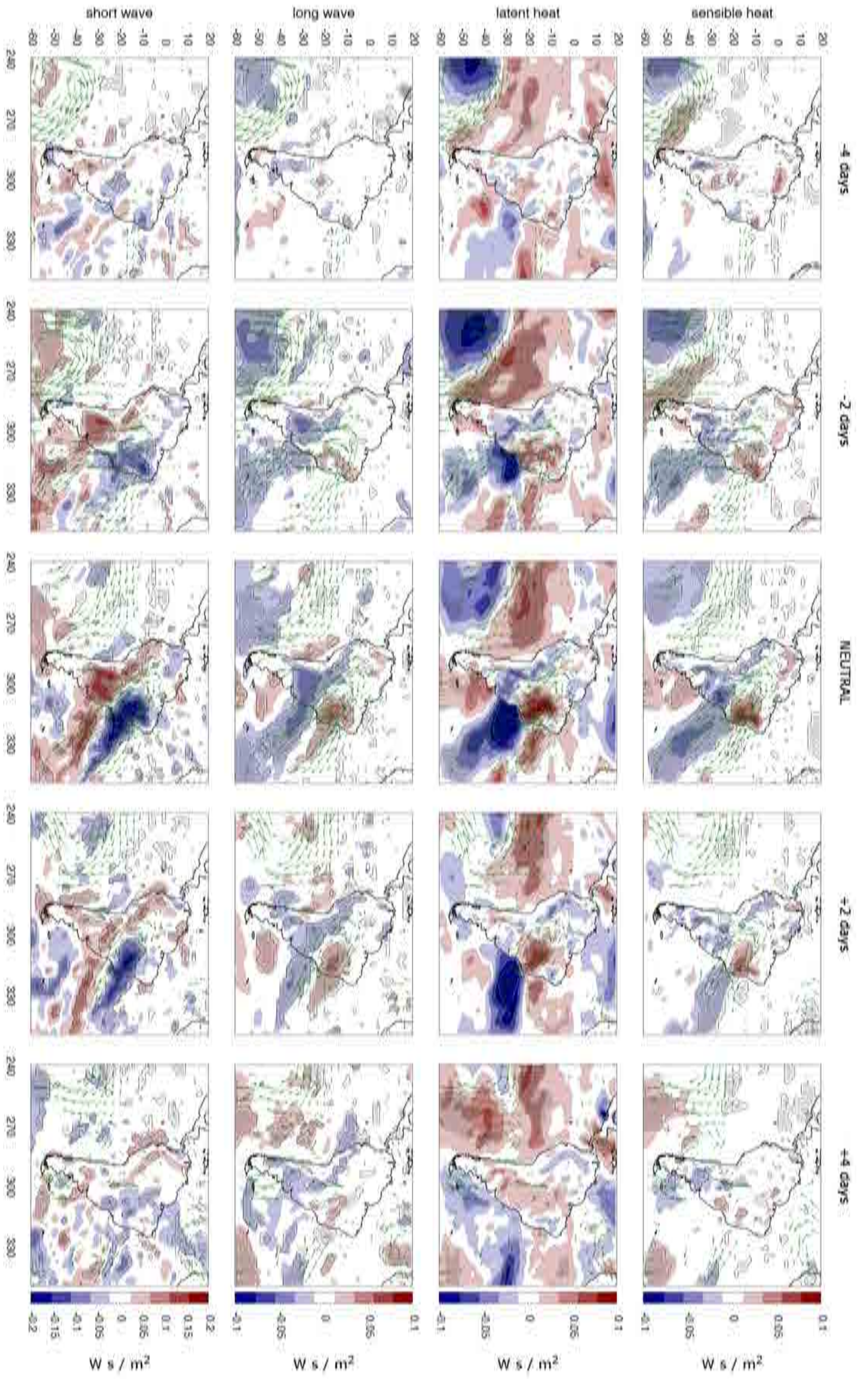


Figure 6.10: Composites of surface heat components and and wind anomalies at 850 hPa for the neutral case. Black shading marks areas of 90% significance, only arrows related to winds anomalies significant at 90% are reported.

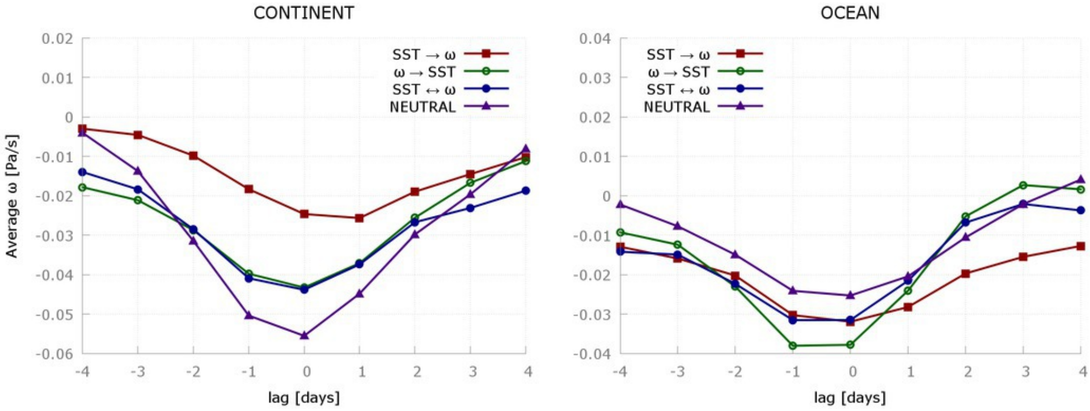


Figure 6.11: Composites of average ω anomalies in Continental ([22S;6S] × [55W;35W], **left**) and Oceanic ([40S;30S] × [40W,1W], **right**) boxes (see text for details).

7 Network-based early-warning indicators of vegetation transitions

In this chapter we deal with an important consequence of climate variability and dynamics. We consider vegetation patterns in semi-arid regions which precede the abrupt transition to bare soil (see Sec. 1.5). Complex network techniques are used to develop novel early-warning indicators for these desertification transitions. These indicators are applied to results from a local positive feedback vegetation model and are compared to classical indicators, such as the autocorrelation and variance of biomass time series. A quantitative measure is also introduced to evaluate the quality of the early-warning indicators. Based on this measure, we show that the network-based indicators outperform to the classical ones, being more sensitive to the presence of the transition point.

The results of this chapter are resumed in [125]

7.1 Model and Methods

In this section, we first present the spatial vegetation model used to produce biomass time series, and then present the methods of network construction and analysis.

7.1.1 A spatial model of vegetation dynamics with a local positive feedback

The local positive feedback model (LPF) is described by the following set of stochastic differential equations [34, 126, 41],

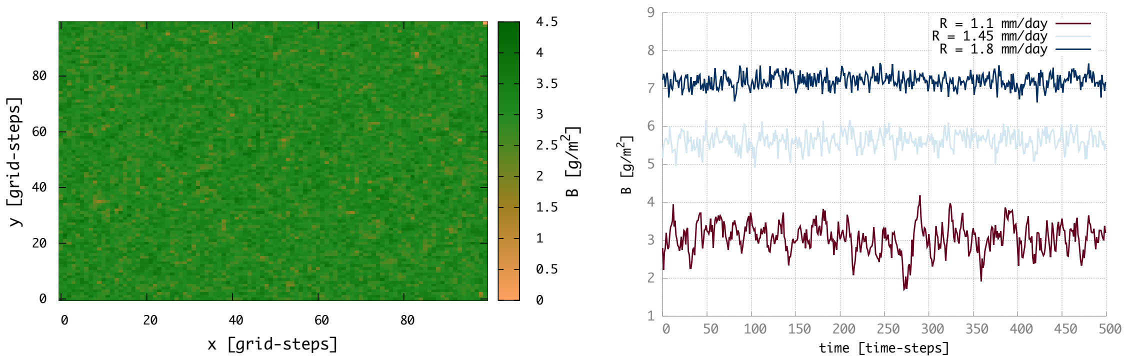
$$\frac{\partial w}{\partial t} = R - \frac{w}{\tau_w} - \Lambda w B + D\nabla^2 w + \sigma_w w_0 \xi^w(t) \,, \tag{7.1a}$$

$$\frac{\partial B}{\partial t} = \rho B\left(\frac{w}{w_0} - \frac{B}{B_c}\right) - \mu\frac{B}{B+B_O} + D\nabla^2 B + \sigma_B B_0 \xi^B(t) \,, \tag{7.1b}$$

where w (in mm) is the soil water amount and B (in g/m^2) is the vegetation biomass. The quantity D is the diffusivity and τ_w, μ, ρ, Λ, w_0, B_O, B_c are additional constants explained in

Table 7.1: Parameters of the local positive feedback model (LPF) given by Eq. (7.1b) and same values as in [41].

Parameter	Meaning	Value
D	Exchange rate	$0.5\ \mathrm{m^2 day^{-1}}$
Λ	Water consumption rate by vegetation	$0.12\ \mathrm{m^2 g^{-1}\ day^{-1}}$
ρ	Maximum vegetation growth rate	$1\ \mathrm{day^{-1}}$
B_c	Vegetation carrying capacity	$10\ \mathrm{g\,m^{-2}}$
μ	Maximum grazing rate	$2\ \mathrm{g\ day^{-1}\ m^{-2}}$
B_O	Half-saturation constant of vegetation consumption	$1\ \mathrm{g\,m^{-2}}$
σ_w	Standard deviation of white noise in water moisture	0.1
σ_B	Standard deviation of white noise in vegetation biomass	0.25
w_0	Water moisture scale value	1 mm
B_0	Biomass density scale value	$1\ \mathrm{g\,m^{-2}}$
τ_w	Water moisture scale time	1 day



(a) Snapshot of B field at $R = 1.1$ mm/day

(b) Example of B time series

Figure 7.1: (a) A snapshot at $t = 500$ time-steps of the biomass field B at $R = 1.1$ mm/day. (b): Example of biomass time series for a node (single grid-cell) i and different values of R.

Tab. 7.1. Finally, R is the amount of rainfall which is used as the bifurcation parameter of the system. Additive Gaussian white noise, ξ, for which

$$\langle \xi(t)\xi(t')\rangle = \delta(t-t'), \tag{7.2}$$

is prescribed with amplitudes σ_w and σ_B for soil water and biomass, respectively. A characteristic spatial pattern of biomass as well as time series for different values of R are shown in Fig. 7.1.

The important mechanism in this model is a positive feedback that causes each patch to have alternative stable states. This is demonstrated in the bifurcation diagram of the LPF model which is sketched in Fig. 7.2. The deterministic homogeneous solutions of the LPF model and their linear stability can be determined analytically. For all values of R, the trivial solution ($B = 0$, $w = \tau_w R$) exists. For the standard parameter values shown Tab. 7.1, the trivial solution is linearly stable for $R < 2$ mm/day and unstable for $R > 2$ mm/day (see Fig. 7.2). At $R = 2$ mm/day, a transcritical bifurcation occurs and two additional branches of steady solutions emerge. Solutions on the lower branch are not considered here because they have $B < 0$, i.e., they are physically non-realistic. Solutions on the upper branch are unstable for values of R down to $R_c = 1.067$ mm/day. At this R-value a saddle-node bifurcation occurs which provides a linearly stable upper branch of solutions for $R > 1.067$ mm/day. Finally, a fourth homogeneous solution exists but it has also values of $B < 0$ for every R value and hence is not further considered in this study.

In order to determine inhomogeneous vegetation patterns in the stochastic case, the model equations (7.1b) are numerically solved on a periodic square grid composed of $100 \times 100 = 10^4$ grid cells on a regular lattice with dimension $L = 100$ m. The evaluated model data consists of a set of time series (500 time steps with $\Delta t = 0.01$ days) of statistically equilibrated biomass fields B for different fixed rainfall parameters R. Time series related to 10 different values of R with $1.1 \leq R \leq 1.8$ mm/day are analysed. For $R < R_c$ only the desert-like solution, with $B = 0$ over the whole domain, is found.

The temporal and spatial mean values of the biomass distribution of each of these spatially inhomogeneous solutions are plotted as the red dots in Fig. 7.2. Obviously, the average values of the inhomogeneous solutions are similar to those of the homogeneous solutions, hence, diffusion and noise do not impact on the average state of the system.

7.1.2 Network approach and analysis

In the following, we will associate a network to the simulated field of biomass. The nodes are defined as the $N = 100 \times 100 = 10^4$ grid cells of the discretised LPF model. In order to define the links between the nodes, the zero-lag temporal correlations between the biomass time series at the different nodes are considered. More precisely, two nodes i and j are linked if the temporal cross-correlation $\mathscr{C}(B_i, B_j)$ of the time series B_i and B_j is statistically significant. A sketch of the network formation is depicted in Fig. 7.3. Note that in this way the number of nodes of the network is kept fixed, and changes in the network structure due to varying R are solely related to the links, that is, to changes in the values of the cross-correlation matrix $\mathscr{C}(B_i, B_j)$.

As usual we keep a cross correlation value as a link only if it trespass a certain threshold Θ To determine the value of Θ we build the following test variable for the Student's t-test, i.e.,

$$t = \frac{\Theta}{\sqrt{1-\Theta^2}} \sqrt{N_{\text{time-steps}} \frac{1-r}{1+r}}\,, \tag{7.3}$$

with the null hypothesis $\Theta = 0$. Here $r = r(R)$ and $N_{\text{time-steps}}$ are the autocorrelation and the length of the time series, respectively. The test variable takes the effective number of degrees of freedom of the time series into account. From this we can compute the value of Θ which ensures statistical significance of correlations larger than Θ. A value of $\Theta = 0.2$ guarantees that,

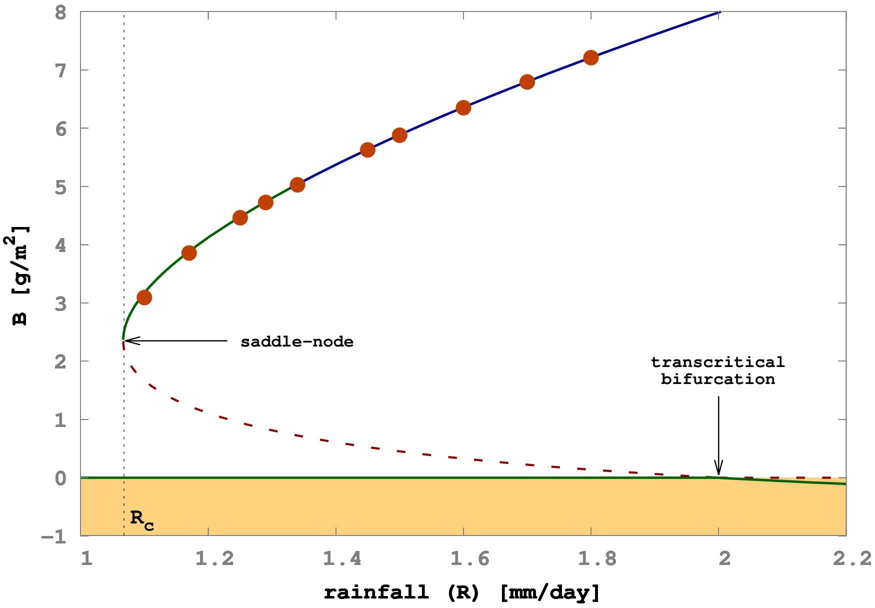


Figure 7.2: Bifurcation diagram of the local positive feedback model (LPF) given by Eq. (7.1b). Curves depict steady homogeneous states, that is, determined under vanishing diffusion and noise. Linearly stable branches are denoted by solid lines, whereas linearly unstable branches are indicated by dashed lines. The blue lines mark states for which the Jacobian matrix has eigenvalues with a non-zero imaginary part. The shaded region indicates to the non-physical negative B values. The average value of B for each simulation of biomass evolution via the full Eq. (7.1b) is depicted by dots. Obviously, the average values of the inhomogeneous solutions are similar to those of the homogeneous solutions, hence, diffusion and noise do not impact on the average state of the system.

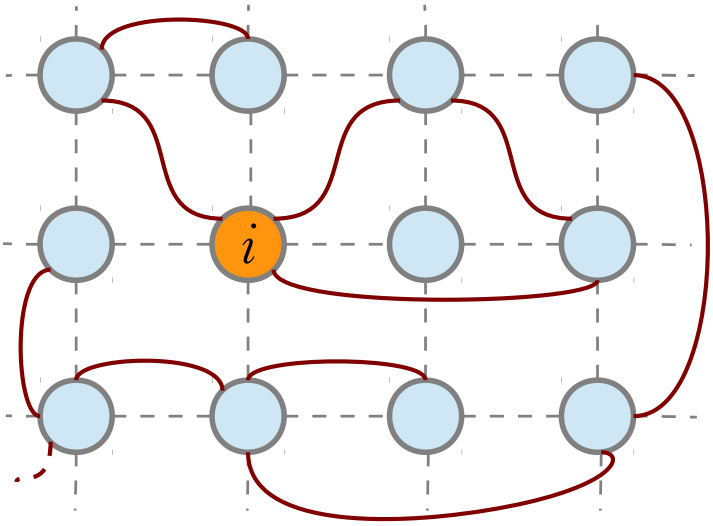


Figure 7.3: Sketch of the network associated to the local positive feedback model (LPF) given by Eq. (7.1b). The circles represent the biomass grid-cells, that is, the nodes of the network. The dashed lines represent the computational grid, whereas the solid lines represent the network connections, that is, the links. The node i (marked orange) has degree $k_i = 3$, clustering coefficient $c_i = 0.33$, and assortativity $a_i = 2$. The link length of its three links is $\sqrt{2}$, $\sqrt{2}$, and 2 grid-steps (see text for the definition of these quantities).

for each value of R, the zero-lag correlation between linked nodes is statistically significant with a p-value smaller than 0.05 and this value is taken in all results below.

A central advantage of this approach is that by taking into account only the statistically significant correlations to establish links between nodes, we can remove part of the noise and extract only the relevant information. This coarse-graining results in a more precise measure and identification of the essential properties of the system under study.

It is important to note that using the temporal cross-correlation among nodes does not directly take their spatial distribution into account. It is only the temporal part of the data which is used to build the network, but the spatial information is kept because the nodes have a definite spatial location in physical space (that is, on a grid). Thus, within this network approach we exploit both the systems temporal information and spatial information which are treated separately in the classical early-warning approaches.

After the construction of the interaction network of the biomass data, we can study changes in the topology of the network due to varying R. The most basic characteristic of a network is its degree distribution (see Sec 2.4.2).

Another basic network measure is the assortativity a_i of a node i, which is the average degree of its neighbours, that is, of all the nodes to which node i is linked to. In Fig. 7.3, all three neighbours of node i have degree 2 and hence its assortativity $a_i = 2$. In general, the assortativity coefficient of a node can be computed from the adjacency matrix via

$$a_i \equiv \frac{1}{k_i} \sum_{j=1}^{N} A_{ij} k_j \; . \tag{7.4}$$

The assortativity characterizes the tendency of a node to be connected to nodes with high degree. Note that a node can have low degree but at the same time high assortativity. As the degree, the assortativity coefficient can range from 0 to $N-1$.

As a third basic network measure we consider the clustering coefficient c_i that measures the number of links among the neighbours of a node i, weighted by the possible number of links among its neighbours (see Sec 2.4.2). In the example shown in Fig. 7.3, there is only one link between the neighbours of node i and there are three possible links among these neighbours and hence the clustering coefficient $c_i = 1/3$. In general, the clustering coefficient of a node can be computed from the adjacency matrix via

$$c_i \equiv \frac{1}{k_i(k_i-1)} \sum_{j=1}^{N} \sum_{l=1}^{N} A_{ij} A_{jl} A_{li} \; . \tag{7.5}$$

which can take values between 0 and 1. The average clustering coefficient of a network quantifies the presence of strongly connected groups (clusters) in the network.

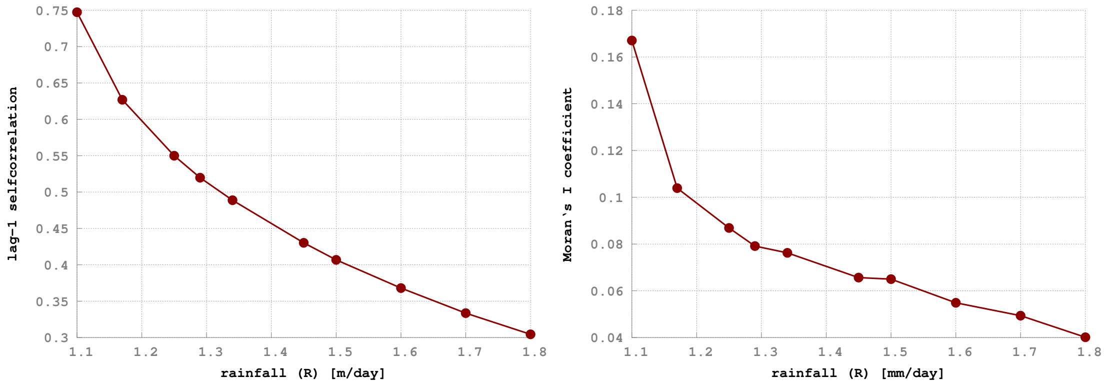
7.2 Results

In subsection 7.2.1, the time series of the LPF model related to different values of R are analysed by computing the classical early-warning indicators and focussing on the behaviour of these measures when R approaches the critical value R_c. In subsection 7.2.2, we investigate whether changes in the network topology (diagnosed via the distributions of degree k_i, assortativity a_i and clustering c_i) can be used as alternative network-based indicators of R approaching R_c.

7.2.1 Classical indicators

The most prominent statistical measure to infer CSD from data is the lag-1 autocorrelation. The spatially averaged autocorrelation of the LPF model data is shown in Fig. 7.4a for each value of R. As R decreases and approaches R_c, the system indeed experiences an increase in autocorrelation, a distinct fingerprint of CSD.

It has also been suggested that spatial statistics may be used to detect CSD [41]. In particular, an increase in spatial correlation of the system is expected when the system experiences CSD close to the transition. A typical measure of spatial correlation that has been used is Moran's



(a) (b)

Figure 7.4: (a) Spatially averaged lag-1 autocorrelation and (b) Moran's coefficient I computed using the spatial biomass distribution at the last time-step of the simulation for different values of the rainfall parameter R.

coefficient, I, defined here through the biomass mass field B as

$$I \equiv \frac{N}{\sum_{ij} g_{ij}} \frac{\sum_{ij} g_{ij}(B_i - \bar{B})(B_j - \bar{B})}{\sum_i (B_i - \bar{B})^2}\,, \tag{7.6}$$

where $g_{ij} = 1$ if i and j are two adjacent grid cells and $g_{ij} = 0$ otherwise. The spatially averaged biomass is indicated by $\bar{B}$. Moran's coefficient I is shown in Fig. 7.4b and demonstrates that the spatial correlation increases in a similar way – although stronger – to the temporal correlation as R decreases and approaches R_c.

However, despite the fact that the classical indicators are able to reflect CSD, they change in a very smooth, gradual and monotonic way. From a strictly local point of view, that is, based on a few closely spaced R values, it is not possible to estimate the proximity of the system to R_c. In other words, the correlation coefficients suffer from a lack of distinct features necessary to provide a pronounced early-warning signal of desertification.

7.2.2 Interaction network based indicators

In order to give a general characterization of the interaction network of the LPF model biomass data, we briefly discuss its link length distribution before studying the changes in degree, assortativity and clustering due to varying R.

The link length measures how close two linked nodes are in physical space, that is, the link length is the Euclidean distance between two linked nodes (for an example see again Fig. 7.3), where the grid spacing is taken as a unit length. The distribution of link lengths for different R is shown in Fig. 7.5. Apparently, the link length distribution does not depend on R. The smaller spike at link length of one grid spacing represents the underlying grid structure and

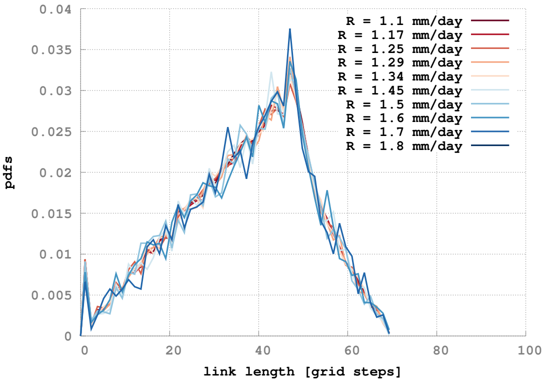


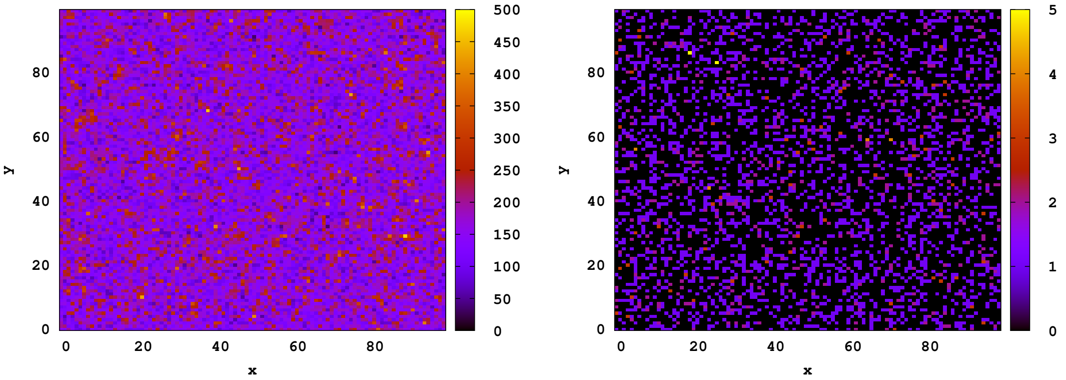
Figure 7.5: Link length PDFs for various values of rainfall R. Link length is computed as the Euclidean distance between two linked grid-cells, taking into account periodic boundaries.

is related to local correlations due to diffusion. The linear behaviour at larger distances can be attributed to the random structure of the network: If the links are randomly distributed in physical space, the probability of two nodes i and j to be connected is independent of their distance. In fact, the number of links of length d of any node i is solely proportional to the number of nodes at distance d from node i. This number scales with the circular area $2\pi d$. Hence, the probability of a certain link length is directly proportional to the link length itself, until the appearance of the boundary reduces the possible number of links again (here at a distance of 50 m, that is $L/2$). Consequently, the interaction network of the simulated biomass field appears to be composed of a random architecture superimposed on the grid structure.

Figure 7.6 shows the spatial field of the degree distribution for two values of R (at the transition and far away from the transition), together with the degree density distribution (i.e., the PDF of k_i) for different R. With decreasing rainfall R the mean number of links per node in the network increases. Consequently, the degree PDF shifts to the higher degrees when the system approaches the critical transition. The increase of network connectivity is related to higher values of the cross-correlation among the nodal time series which occurs near the saddle-node bifurcation [127].

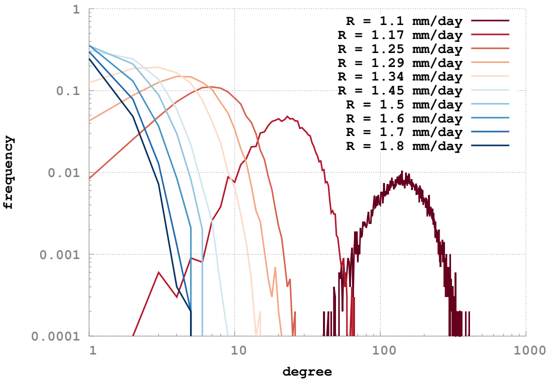
This feature can also be seen in the spatial patterns of the degree field. For $R = 1.8$ mm/day (Fig. 7.6b), the network is almost disconnected and the majority of the nodes has zero or one link. With decreasing rainfall the number of connections increases, and the disconnected nodes eventually join the network. At the transition, the spatial pattern of the degree distribution shows a granular structure, with patches of highly connected nodes (Fig. 7.6a).

The shift of the degree PDF to higher degrees (Fig. 7.6c) can be used to develop an indicator of the upcoming transition. Figure 7.7a shows the mean of the degree distribution as a function of R. This network measure is highly sensitive to R near criticality, showing a steep increase close to the transition. Additionally, Fig. 7.7b shows the variance of the degree distribution as a function of R. The increase in variance when approaching the transition point is even



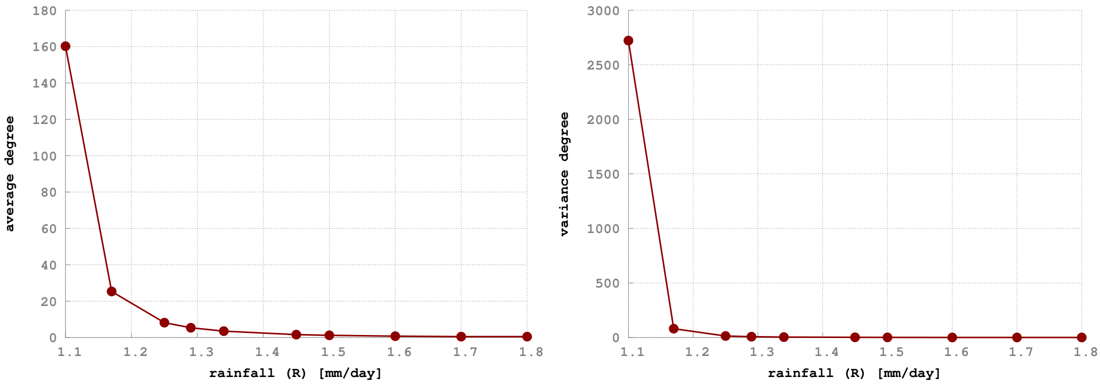
(a) $R = 1.1$ mm/day

(b) $R = 1.8$ mm/day



(c) degree distributions

Figure 7.6: (a-b) Spatial field and (c) PDF of node degree for different values of R.



(a) mean

(b) variance

Figure 7.7: (a) Mean and (b) variance of network degree distribution as function of the bifurcation parameter R.

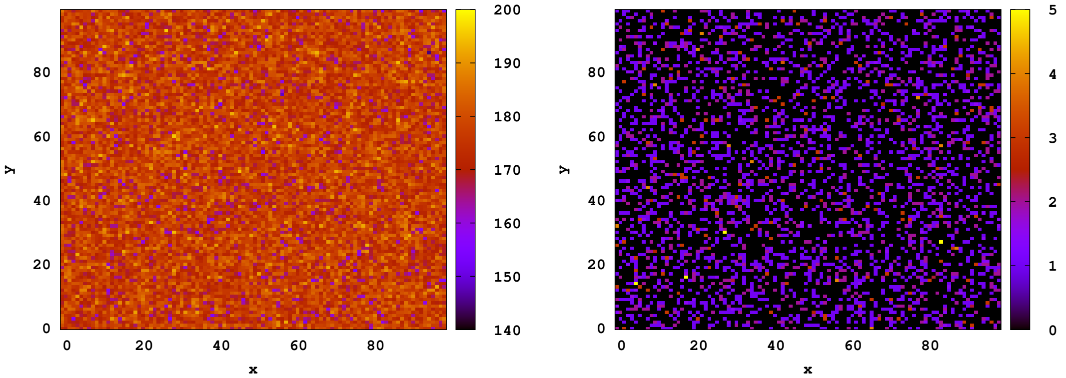
more abrupt. The behaviour of the average degree is directly related to high degree values occurring near the saddle-node bifurcation. Close to the transition, the vegetation variability synchronizes over the domain, which can be seen as the spatial expression of CSD. This apparent synchronization produces an increasing number of connections approaching the tipping point. As it can be seen, the panels in Fig. 7.7 allow quantifying the behaviour displayed by the spatial degree field in Fig. 7.6. In this way we capture the bidimensional information of the size of the patches and their degree values in a clear way, which monitors quantitatively the presence of the upcoming transition.

Figure 7.8 shows the assortativity field of the interaction network and the corresponding PDF for different rainfalls R. The spatial pattern of assortativity is more homogeneous than that of the degree field and hence its PDFs are narrower. Similar to the degree field, there is a substantial shift to larger values of the PDF when R approaches R_c. The spatial patterns of assortativity display a very noisy structure, compatible with the random network architecture previously inferred from the link length distributions. Furthermore, the average assortativity (not shown) increases sharply when approaching the transition as can also be seen in the PDFs.

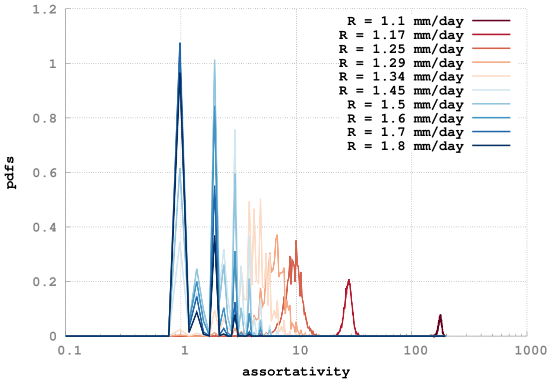
Figure 7.9a-b shows the skewness and kurtosis of the assortativity distribution[1] for different R. The switch in sign of the skewness (Fig. 7.9a) is a feature that could be related to the forthcoming transition. This switch in sign is unique (a distinct qualitative feature), and it is not prone to false alarms. Furthermore, the kurtosis (Fig. 7.9b) shows a huge and quick drop just *before* the transition, providing a distinct warning related to the upcoming transition point, especially if considered together with the behaviour of the skewness. However, the combined analysis of these two quantities reveals a more pronounced change in the assortativity distribution:

---

[1] We exclude the mean and variance of the assortativity distribution from the presentation because they are directly connected to the mean and variance of the degree distribution.

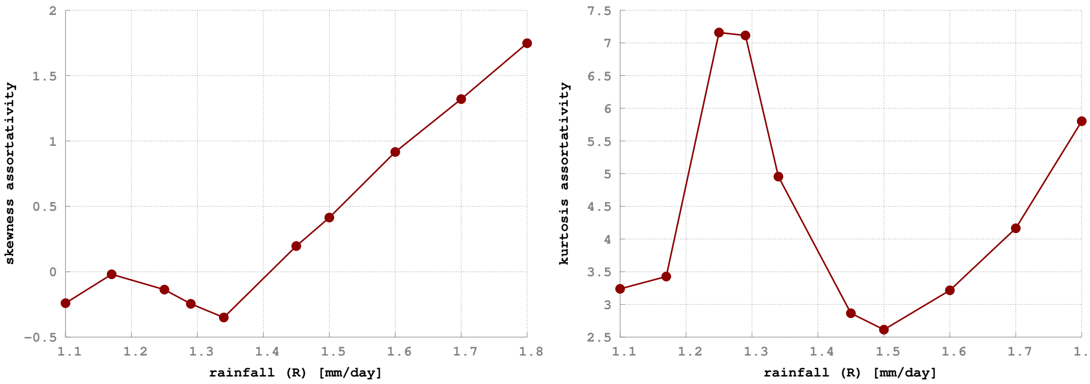


(a) $R = 1.1$ mm/day (b) $R = 1.8$ mm/day



(c) assortativity PDFs

Figure 7.8: (a-b) Spatial field and (c) PDF of the assortativity for different values of R. The sharp increase of the assortativity approaching the tipping point is directly related at the behaviour of the corresponding degree field.



(a) assortativity skewness

(b) assortativity kurtosis

Figure 7.9: (a) Skewness and (b) kurtosis of the assortativity distribution for different R.

Just before the transition the skewness approaches 0, while the kurtosis is close to 3; thus the assortativity distribution is close to a Gaussian distribution near R_c. This “Gaussianisation” may therefore be used as an early-warning signal of the transition point.

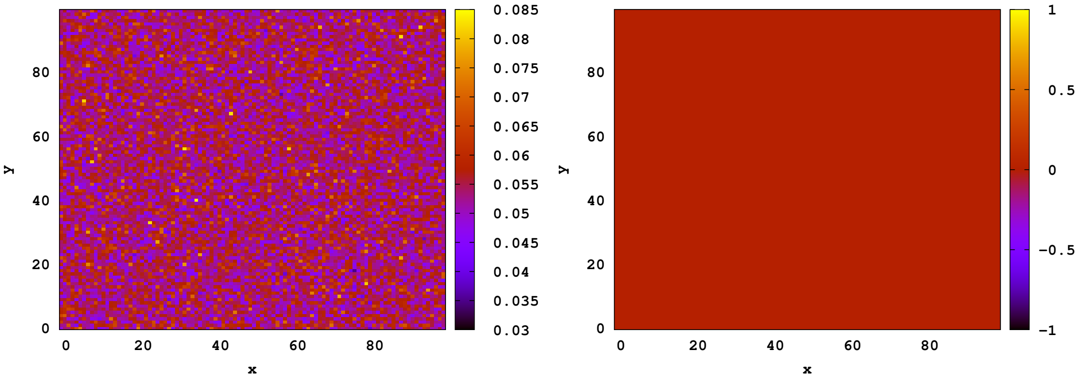
Finally, Fig. 7.10 shows the spatial field of the node clustering coefficient and the corresponding PDF for different rainfall parameter R. At high R, the clustering coefficient is zero everywhere. The average node degree is very low such that there are not sufficient links to create clustered structures. With decreasing R the first small clusters start to appear and a non-vanishing average value of the clustering coefficient emerges. The more links are being added to the network with decreasing R, the more clusters are formed and mean clustering coefficient continues to increase. Close to the transition, a richer network structure provides the noisy pattern displayed in Fig. 7.10a.

Regarding the transition, the mean clustering (not shown) does not provide more information than the mean degree because these two quantities are related: in a completely random network the clustering scales linearly with the number of links[2]. The variance of the clustering distributions (not shown) provides an early-warning indicator of the transition by drastically dropping just before R_c. The combined increase in mean clustering and reduction in variance can be explained as an effect of the increase in spatial coherence displayed by the system. Approaching the transition, patches of vegetation increase their synchronization forming well defined groups of strongly connected nodes. However, these indicators do not perform better than the corresponding degree measures.

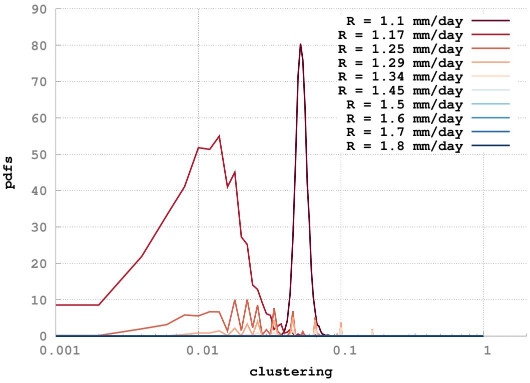
In contrast, the skewness and kurtosis of the clustering distribution, shown in Fig. 7.11c-d, can be used as an early-warning indicator if the two quantities are monitored together. They

---

[2] This is easy to see considering the clustering as the probability of two neighbours to be also connected among them. In a random network this probability is independent by the nodes and it is equal to the network density p. Thus the average clustering of a fully random network will be exactly p, that depends linearly by the number of the total links.

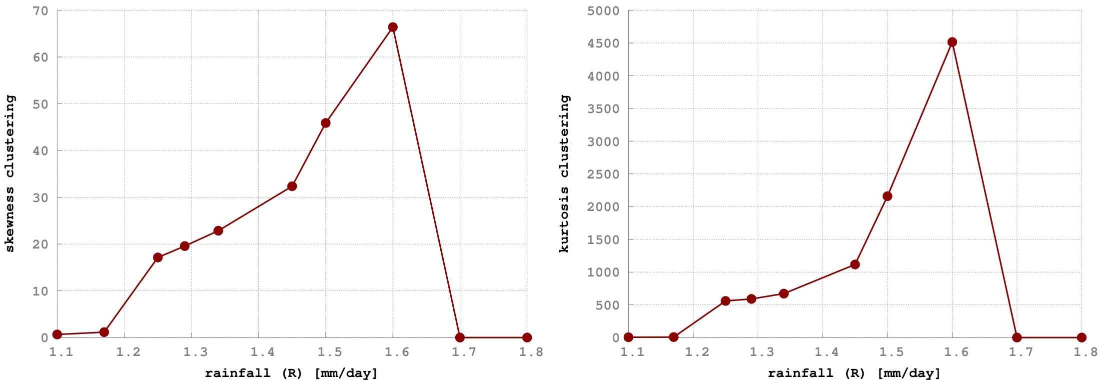


(a) $R = 1.1$ mm/day (b) $R = 1.8$ mm/day



(c) clustering PDFs

Figure 7.10: (a-b) Spatial field and (c) PDF of node clustering for different values of R.



(a) clustering skewness

(b) clustering kurtosis

Figure 7.11: (a) Skewness and (b) kurtosis of the clustering distribution for different R.

are displaying a “Gaussianisation” of the clustering distribution, similar to the assortativity distribution when approaching the tipping point. However, up to this point the “Gaussianisation” of the assortativity and clustering PDFs is only a qualitative feature. In the following section we give a more quantitative measure of this class of early-warning indicators.

7.2.3 Quality assessment of early-warning indicators

Based on the results of the previous section, we can divide the early-warning indicators considered into two main classes. The first class consists of *scalar-based* indicators, that is, scalar measures which generally change monotonically when the transition point is approached. For example, the mean and variance of degree, assortativity and clustering belong to this class. For these indicators it is generally important to monitor their derivative with respect to the bifurcation parameter R in order to sign proximity of the system to the transition point. The absolute value of the indicator itself gives less information than its abrupt change close to the transition [128, 42]. The classical indicators are also included in the scalar-based class.

The second class consists of *distribution-based* indicators. These indicators monitor changes in the distribution of key quantities when the system approaches the tipping point. In our case, the assortativity and the clustering belong to this class because their distributions approach Gaussian distributions close to R_c. The degree distribution, instead, does not approach a Gaussian (not shown). The critical normalization of clustering and assortativity can be quantified numerically through the Kullback–Leibler Distance (KLD), also called relative entropy, which measures the distance between two PDFs. Given two one-dimensional distributions $P(x)$ and $Z(x)$, their relative entropy is defined as

$$KLD \equiv \int_{-\infty}^{\infty} \ln\left(\frac{P(x)}{Z(x)}\right) P(x)\,\mathrm{d}x\,. \tag{7.7}$$

Measuring KLD between assortativity/clustering distributions and Gaussians with the same

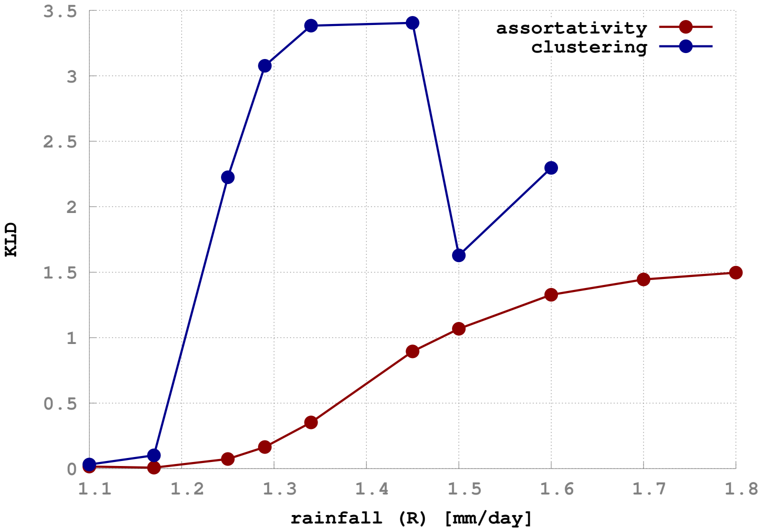


Figure 7.12: KLD values for clustering and assortativity PDFs with respect to Gaussian distributions with same mean and variance.

mean and variance allows to quantify the “Gaussianisation” of these PDFs when the tipping point is approached. We note that the Gaussianisation might a model-specific feature and further analysis is required to assess the generality of this result.

The KLD measure for both assortativity and clustering distributions is plotted in Fig. 7.12. Obviously, KLD quickly drops to zero when approaching the transition point for both the assortativity and clustering distributions. Small KLD means that the distribution under examination is comparable to a Gaussian, whereas for large values it deviates from Gaussianity. Unlike the scalar-based case, the absolute value of the indicator is more important than its derivative.

For both indicator classes a quality measure can be defined. First, we define an ϵ-environment around the bifurcation point by all the R values for which $(R-R_c)/R_c < 0.1$. For a scalar-based indicator J, we then define the normalized quality measure Q_{s}^{J} by

$$Q_{\mathrm{s}}^{J} \equiv \frac{\left\langle \frac{\partial J}{\partial R} \right\rangle_{R<\epsilon} - \left\langle \frac{\partial J}{\partial R} \right\rangle_{R>\epsilon}}{\left\langle \frac{\partial J}{\partial R} \right\rangle_{R<\epsilon} + \left\langle \frac{\partial J}{\partial R} \right\rangle_{R>\epsilon}} \,. \tag{7.8}$$

where the brackets indicate the mean over the interval indicated. In this way we achieve that if J shows an abrupt change in its derivative close to the transition then we have $Q_s^J \approx 1$. In contrast, if the change of J is merely linear when approaching the tipping point then we obtain $Q_s^J \approx 0$.

In case of distribution-based indicators we can define a similar quality measure Q_{d}^{J} by taking

Table 7.2: Quality values of different early-warning indicators

Indicator J	**class**	**type**	**$\mathbf{Q}^J$**
Average Degree	scalar	network	0.963
Variance of Degree	scalar	network	0.996
Variance of Assortativity	scalar	network	0.971
Average Clustering	scalar	network	0.904
Assortativity Gaussianity	distribution	network	0.973
Clustering Gaussianity	distribution	network	0.951
Lag-1 Autocorrelation	scalar	classical	0.468
Spatial Correlation	scalar	classical	0.730

into account J itself instead of its derivative, that is,

$$Q_{\mathrm{d}}^{J} \equiv \frac{\langle J\rangle_{R>\epsilon} - \langle J\rangle_{R<\epsilon}}{\langle J\rangle_{R<\epsilon} + \langle J\rangle_{R>\epsilon}} \,. \tag{7.9}$$

In Tab. 7.2 the quality values for both network-based and classical indicators are shown. Obviously, the network-based measures have significantly higher early-warning quality than the classical measures.

7.3 Conclusions

In this study the possibility of anticipating desertification transitions in a simple vegetation model using novel interaction network techniques is investigated. Interaction networks are constructed from time series of biomass fields and the topological changes in the resulting networks are studied along a gradient of decreasing rainfall. We find that network measures like degree, assortativity and clustering may offer novel indicators for identifying an upcoming desertification in semi-arid ecosystems.

Our results are consistent with previous studies that have used network measures as early-warning indicators of critical transitions in models. For example, [127] used an interaction network approach to obtain an early-warning signal of the Atlantic Meridional Overturning Circulation (AMOC) collapse. In that study the network is built using temperature time series, and the behaviour of the average degree is monitored as function of freshwater input. Similar to the results in this study, the average degree increases sharply approaching the transition. In contrast, [129] analysed the flow field of the wind-driven ocean circulation introducing a flux-based network approach. Also in that context, the degree of the network increases while the system approaches the transition, but a more precise early-warning indicator is given by the network's closeness which shows a big drop near the tipping point due to a local regime change in the flow field.

In this study we have introduced measures to assess the quality of different early-warning indicators. Using these quality measures we compared the performance of the novel network-based indicators with the classical indicators based on variance and autocorrelation. We find that the scalar network-based indicators have a higher quality value than the classical indicators. Moreover, distribution-based indicators, here calculated from the assortativity and clustering distributions, have also a high quality value. When these distributions become close to Gaussians, there is an early-warning indication of the upcoming transition. Although these observations hint that the indicators we developed here may offer a strong indirect measure of proximity to critical transitions, they may still be prone to similar limitations that classical indicators face, like producing false positives [130], or requiring a lot of information for their practical application [131].

Regarding the possibility of false positives, it is well known that CSD is a characteristic feature of transitions related to an eigenvalue going to zero. However, not all the transitions accompanied by an eigenvalue going to zero are also catastrophic – that is implying an abrupt discontinuity in the stable branches [38, 132]. Thus, CSD can be prone to false alarms, and it will be interesting to test, in future work, the new network indicators against this possible pitfall.

Regarding the sample size of the data, instead, it is clear that the network construction requires a sufficient spatial sampling which temporally based classical indicators do not need. As a test, we computed all the indicators presented in this paper also for other two networks, obtained by spatially down-sampling the original 100×100 dataset considered here. In particular we considered 10×10 and 50×50 sub-sets. We find that the scalar-based indicators still perform quite well, whereas the distribution-based indicators do not. In particular, the assortativity Gaussianity still anticipates the transition in the 50x50 data set, whereas the clustering measure is unable to detect the transition in both data sets. This is a reasonable feature of the proposed distribution-based indicators; in fact a good estimation of a PDF requires a big amount of data, much more than simple low-order moments estimators, as the one used here to build the scalar-based indicators.

While the average degree behaviour as a function of R can be easily linked to the spatial expression of the enhancement of correlations of the time series between the nodes, the interpretation of the other indicators, especially that of the distribution-based indicators, is not straightforward. Nevertheless, the network-based indicators seem to offer a better measure of proximity to the tipping point than the classical ones. This may be attributed to the thresholding of the correlation matrix when constructing the interaction network. This coarse-graining of the information eliminates most of the noise, producing a better signal to noise ratio of the spatial signature of CSD.

The network based indicators hence offer a promising alternative to detect critical transitions. An interesting next step is to analyse the utility of these indicators in a scale dependent feedback model [133, 41], for which basic bifurcation diagrams were presented in [132]. Also in such a model, the desertification transition is a saddle-node bifurcation but the approach to this saddle-node with decreasing rainfall is not as smooth as in Fig. 7.2 due to presence of multiple branches of stable steady states.

8 Conclusions and future perspectives

In this thesis we proposed new ways to analyse climatological data and to extract new information about the climate system from them. In particular we applied nonlinear techniques imported from information theory and dynamical systems, such as ordinal time series analysis. Another tool we extensively used is the reduction of the data into functional complex networks. We performed this reduction using several techniques (Granger causality, cross correlation, etc.), and we used the networks to analyse various climatological phenomena as well as for defining transitions indicators.

In Chapter 3 we have shown that the average connectivity of the climate networks constructed from monthly-averaged SAT anomalies is not significantly affected by the introduction of time-shifts when computing statistical similarity measures. This feature is quite unexpected, since the shifts were such as to produce a synchronisation of the annual solar cycle in the network nodes (i.e., comparing winters with winters, summers with summers). This finding was interpreted as mainly due to the fact that most of the links with non-zero lags are weak links, and for theselinks the time-shifting indeed changes their similarity values; however, these changes appear to be random and tend to be washed out when computing the AWC.

in Chapter 4 we have tested the statistical similarity techniques that are typically used in climate networks studies, to investigate up to which extent the functional connectivity reflects the structural one underlying the dynamics. We found that, in the studied cases, it is possible to find regimes in which the structural and functional networks are equal or highly similar. We also found that functional networks obtained from thresholding the cross-correlation matrix perform better with respect to the two other non-linear measures. Finally, we found that using derived variables can lead to a better reconstruction performance than using the dynamical ones, that is a suggestive hint for future network analysis of real world complex systems.

In Chapter 5 we presented two new methods to identify climate communities, one based in the seasonal cycle, and the other one based in the anomalies with respect to it. With the first method we obtained a map of the thermal inertia of the various regions of Earth. With the second method, instead, we classified regions of the Earth in terms of their anomalies statistical properties, in this way regions belonging to the same community display a similar climate. In fact our algorithm produces a division of the world in four macro-areas: ENSO basin, extratropical and tropical oceans and extratropical large landmasses.

In Chapter 6 we studied the air-sea interactions in the region of the South Atlantic Convergence Zone using the Granger Causality Estimator as a measure of directional coupling. A first result is that this region shows a large air-sea connectivity during summers. We have also been able to classify years in four types, depending on the direction of the air-sea interaction. In particular, we identified years in which the coupling is mainly directed from the ocean to the atmosphere, years in which the coupling is from the atmosphere to the ocean, years of mutual interactions and years of no significant coupling. We studied the evolution of the convective events related to these different regimes looking at atmospheric and surface oceanic fields and found that, in all cases, an extratropical wave train plays a major role in forcing convection in the SACZ; in particular it seems the path of the waves to be the important factor that differentiates the four kind of convective events. We also pointed out a possible role of El Niño - Southern Oscillation, noting that ocean-forcing cases are more probable during El Niño years.

In Chapter 7 we studied the possibility of anticipating desertification transitions in a simple vegetation model using interaction network techniques. We found that network measures like degree, assortativity and clustering may identify an upcoming desertification in semi-arid ecosystems. Introducing a quality measure we compared the performance of the novel network-based indicators with the classical indicators based on variance and autocorrelation. We found that the scalar network-based indicators have a higher quality value than the classical indicators, thus the network based indicators offer a promising alternative to detect critical transitions. An interesting next step is to analyse the utility of these indicators in more complicated models, as well as in real data, although this could be very difficult to implement given the spatio-temporal resolution required by the model.

This thesis also points into several direction in terms of possible future works. From the networks point of view, here we have only studied a classical quantities, such as the weighted degree, and, in some sections, also the clustering coefficient and the assortativity; however many quantities that could have a climatic interpretation could be analysed, as the betweenness, the centrality, the modularity, and so on.

Also, these techniques could be applied just to other datasets. For example, the analysis performed in Chapter 6 could be not only extended to other areas where air-sea interaction can be relevant, but also to other variables, such as the outgoing longwave radiation (OLR) – a good proxy for precipitation – used instead of the vertical velocity. As this tool allows detecting nonlocal interactions, it can be useful for studying the role of the sea surface temperature (SST) in the monsonic systems, where the bulk of the deep convection is over land.

In Chapter 6 we also introduced bi-layer networks. A generalisation to multilayer networks is quite natural and can be useful to study the impact of the ocean forcing in the various atmospheric levels. It can be also extended to other variables, such as to study the interplay between circulation, temperature, and precipitation, and so on. It could be also very interesting to study with these framework the relation among networks constructed in different states of the system under examination, such as El-Niño/La-Niña years or winter/summer.

Another line of investigation related to the SACZ could be that of understanding why ocean-forcing cases seem to be more frequent during El-Niño years.

Also, other directionality measures could be checked against the results given by the application of Granger Causality. In particular the Transfer Entropy, already successfully applied in [134], could be used to investigate possible nonlinearities in the coupling, analysing the difference from Granger Causality that is essentially a linear quantity.

Another possible direction of future work is that of testing the robustness of the transition indicators presented in Chapter 7 also in other more complex models. In fact, in arid vegetation ecosystems the formation of pattern through a Turing instability is a common feature, a characteristic that was not captured by the simple toy model we used. It could be also very interesting to be able to apply such methods to real data. However, the vegetation is these areas is strongly affected by seasonality, a feature that is also not present in the model. On the other hand, using yearly averaged data eliminates this problem, but it reduces the length of the database in a significant way, thus making the method not robust. It would be also interesting to analyse the developed indicator more in detail, to test how much they are robust against false alarms.

In Chapter 4 we shown that derived variables could lead to a better identification of the functional connectivity with the structural one. Thus, it will be interesting to analyse climate networks built from time series of frequencies computed through the first derivative of series obtained through a Hilbert transformation of the original one, checking then the difference with the known results from direct time-series and the known pattern of interaction and circulation.

In Chapter 5 we proposed two new methods to identify communities that we provided results in agreement with climate knowledge. It would be interesting, now that we have been able to validate the method, to apply it to other systems in which the community decomposition is not obvious and thus obtain a new insight in the system properties. Also, it could be interesting to analyse the basins' eastern boundaries in the tropics and subtropics to understand why they display a greater thermal inertia than the surrounding water masses.

Another interesting possibility is offered by the ordinal patterns. Changing the interval with which the words are constructed it is possible to investigate different time-scales. Once obtained different networks for different scales it could be possible to study the interaction among scales as well as the directionality of these interactions.

Many of these topics are already under research.

A Appendix

A.1 Analysis of the distribution of Cross Correlation values for two uncoupled Kuramoto oscillators.

In Fig. 4.2 of Chapter 4 we reported that uncoupled Kuramoto oscillators can display arbitrary large values of cross correlation among them, independently of the length of the time-series. This is doubtlessly a counterintuitive feature, although we can prove that it has a solid mathematical reason.

Namely, what we are plotting in Fig. 4.2 is the maximum value over an ensemble of 120 possible correlations among the time series of uncoupled Kuramoto oscillators. The equations for this system can be written in the form

$$d\theta_i = \omega_i \, dt + D \, dW_i^t, \tag{A.1}$$

and, since we processes the series removing a linear trend and normalizing respect to the variance, it can be reformulated as an autoregressive process of order 1 (AR(1)) and unitary memory:

$$\theta_i^{t+1} = \theta_i^t + \epsilon_i^t \tag{A.2}$$

where ϵ is a gaussian innovation with mean 0 and variance 1.

The correlation among two independent AR(1) processes can be proven to be equal to the correlation among the first stochastic elements of the two time-series, namely $r_W = \langle \epsilon_i^1 \epsilon_j^1 \rangle$ [55]. This correlation, of course, cannot be dependent of the length of the time-series since it is relative to the first step. (This explains why we detect an independence of the maximum value respect to the integration time.

The reason why the maximum can reach very high values can also be understood: the distribution of the AR(1) correlation can be transformed into a normal one with the so-called Fisher's transformation: $Z = \tanh^{-1}(r_W)$. This transformation is monotonous, so the maximum values of Z correspond univocally to the maximum values of r_W. Now, as stated in the caption of the figure, we are considering the maximum over $k = 120$ independent realizations, that is the maximum over 120 independent values of r_W i.e. of Z. We call this variable $x = \max(Z)$. The cumulative distribution of such a maximum, $F(x)_k$, can be proven to be given by the following relation: $F_k = [F_1]^k$ that, in our case, is the k-th power of the error function plus a constant. In fact, the probability that the maximum over k independent values is less than a certain x is equivalent to the probability that all the k values are below x, for the definition of maximum; this simply means $F_k = [F_1]^k$.

It has to be noted that for k sufficiently large the distribution F_k converges to a limit function referred as General Extreme Value (GEV) distribution, that is the analogous of the normal distribution of the central limit theorem for extreme values, i. e., the convergence of the x PDF to this function is guaranteed independently of the form of the Z PDF.

Now that we have the cumulative distribution of the variable $x = \max(Z)$ it is possible to compute its expectation as

$$m = \int_{-\infty}^{+\infty} x \, \frac{\mathrm{d}F_k}{\mathrm{d}x} \, \mathrm{d}x \tag{A.3}$$

This integral can be numerically estimated as $m \sim 2.57$ for $k = 120$ that lead, with an inverse Fisher's transformation, to the expected maximum correlation of $r_m \sim 0.988$. Estimating also the variance give use the one sigma interval [0.973; 0.995] for the expected value r_m, corresponding to a ~ 70% confidence interval.

Regarding the Y-projection case, instead, we cannot say much, since the nonlinear transformation to the AR(1) processes impel to compute neither the correlation explicitly nor the moments of the distribution via the generatrix function in a simple way.

For the frequencies-CC maximum values, instead, we have a decrease of the maximum with increasing length of the time-series, T, as $1/T$. This can be easily explained assuming that the maximum values scale with the length of the time-series together with the variance, since the distribution moments of the correlation are computable using the associated moment-generating function, and it can be seen that the variance indeed scales as $1/T$. In fact, from equation A.2 it is possible to see that the cross correlation for two frequencies time-series can be written as

$$CC_{ij} = \frac{1}{T} \sum_{t}^{T} \epsilon_i^t \epsilon_j^t. \tag{A.4}$$

We call the product of two normal independent variable $z^t = \epsilon_i^t \epsilon_j^t$. In can be proven that this variable follows a certain distribution with an associated characteristic function $\phi_z(u) = (1+u^2)^{-1/2}$. Thus, being z^t independent variables, the characteristic function of CC_{ij} can be written as

$$\phi_{CC}(u) = \prod_{t}^{T} \phi_{z^t}\left(\frac{u}{T}\right) = \left[\phi_z\left(\frac{u}{T}\right)\right]^T. \tag{A.5}$$

Deriving twice and evaluating for $u = 0$ we obtain the variance of CC, namely

$$\mathrm{Var}[CC] = \frac{1}{T}. \tag{A.6}$$

Bibliography

[1] M. J. Suarez and P. S. Schopf, “A delayed action oscillator for enso,” *Journal of the atmospheric Sciences*, vol. 45, no. 21, pp. 3283–3287, 1988.

[2] G. K. Vallis, *Atmospheric and Oceanic Fluid Dynamics.* Cambridge University Press, 2006.

[3] M. Steyvers and J. B. Tenenbaum, “The large-scale structure of semantic networks: Statistical analyses and a model of semantic growth,” *Cognitive science*, vol. 29, no. 1, pp. 41–78, 2005.

[4] H. Jeong, B. Tombor, R. Albert, Z. N. Oltvai, and A.-L. Barabási, “The large-scale organization of metabolic networks,” *Nature*, vol. 407, no. 6804, pp. 651–654, 2000.

[5] F. Bretherton *et al.*, “Earth system sciences: A closer view,” *Earth System Sciences Committee, NASA*, 1988.

[6] H. A. Dijkstra and M. Ghil, “Low-frequency variability of the large-scale ocean circulation: A dynamical systems approach,” *Reviews of Geophysics*, vol. 43, no. 3, p. RG3002, 2005.

[7] C. Wang and J. Picaut, *Understanding ENSO physics—A review.* Wiley Online Library, 2004.

[8] N. J. Mantua and S. R. Hare, “The Pacific Decadal Oscillation,” *Journal of Oceanography*, vol. 58, no. 1, pp. 35–44, 2002.

[9] M. E. Schlesinger and N. Ramankutty, “An oscillation in the global climate system of period 65-70 years,” *Nature*, vol. 367, no. 6465, pp. 723–726, 1994.

[10] C. Zhang, “Madden-Julian Oscillation,” *Rev. Geopyhsics*, vol. 43, no. 2004, pp. 1–36, 2005.

[11] J. Marshall and R. A. Plumb, *Atmosphere, Ocean and Climate Dynamics: An Introductory Text.* Academic Press, 2008.

[12] P. Imkeller and J.-S. Von Storch, *Stochastic climate models*, vol. 49. Springer Science & Business Media, 2001.

[13] K. Hasselmann, "Stochastic climate models part I: Theory," *Tellus*, vol. 28, no. 6, pp. 473–485, 1976.

[14] D. B. Stephenson, H. Wanner, S. Brönnimann, and J. Luterbacher, *The history of scientific research on the North Atlantic Oscillation.* Wiley Online Library, 2003.

[15] J. Lifland, "The North Atlantic Oscillation: climatic significance and environmental impact," *Eos, Transactions American Geophysical Union*, vol. 84, no. 8, pp. 73–73, 2003.

[16] J. W. Hurrell, H. V. Loon, and H. Van Loon, "Decadal variations in climate associated with the North Atlantic Oscillation," *Clim. Chang. High Elev. Sites*, vol. 36, pp. 301–326, 1997.

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

[18] C. Wang, "A Unified Oscillator Model for the El Nino-Southern Oscillation," *J. Climate*, vol. 14, no. 1991, pp. 98–115, 2001.

[19] M. H. Glantz, R. W. Katz, and N. Nicholls, *Teleconnections linking worldwide climate anomalies.* Information Systems Division, National Agricultural Library, 1991.

[20] I. M. Held, M. Ting, and H. Wang, "Northern winter stationary waves: Theory and modeling," *J. Climate*, vol. 15, pp. 2125–2144, 2002.

[21] V. Barros, M. Doyle, M. González, I. Camilloni, R. Bejarán, and R. M. Caffera, "Climate variability over subtropical South America and the South American monsoon: a review," *Meteorologica*, vol. 27, no. 1-2, pp. 33–57, 2002.

[22] L. M. V. Carvalho, C. Jones, and B. Liebmann, "Extreme precipitation events in southeastern South America and large-scale convective patterns in the South Atlantic convergence zone," *J. Climate*, vol. 15, pp. 2377–2394, 2002.

[23] L. M. V. Carvalho, C. Jones, and B. Liebmann, "The South Atlantic Convergence Zone: Intensity, form, persistence, and relationships with intraseasonal to interannual activity and extreme rainfall," *J. Climate*, vol. 17, pp. 88–107, 2004.

[24] R. J. Bombardi, L. M. V. Carvalho, C. Jones, and M. S. Reibota, "Precipitation over eastern South America and the South Atlantic Sea surface temperature during Neutral ENSO periods," *Clim. Dyn*, pp. 1553–1568, 2013.

[25] Y. M. Kodama, "Large-scale common features of subtropical precipitation zones (the Baiu frontal zone, the SPCZ, and the SACZ) Part I: Characteristics of subtropical frontal zones.," *J. Meteor. Soc. Japan*, vol. 70, pp. 813–836, 1992.

[26] J. Nogués-Paegle and K. C. Mo, "Alternating wet and dry conditions over South America during summer," *Monthly Weather Review*, vol. 125, pp. 279–291, 1997.

[27] M. Barreiro, P. Chang, and R. Saravanan, “Variability of the South Atlantic Convergence Zone as simulated by an atmospheric general circulation model,” *J. Climate*, vol. 15, p. 745, 2002.

[28] C. Junquas, C. Vera, L. Li, and H. L. Treut, “Summer precipitation variability over Southeastern South America in a global warming scenario,” *Climate dynamics*, vol. 38, pp. 1867–1883, 2012.

[29] A. W. Robertson and C. R. Mechoso, “Interannual and Interdecadal Variability of the South Atlantic Convergence Zone,” *Monthly Weather Review*, vol. 128, no. 8, pp. 2947–2957, 2000.

[30] M. Barreiro, P. Chang, and R. Saravanan, “Simulated precipitation response to SST forcing and potential predictability in the region of the South Atlantic Convergence Zone,” *Clim. Dyn*, vol. 24, pp. 105–114, 2005.

[31] R. R. Chaves and P. Nobre, “Interactions between sea surface temperature over the South Atlantic Ocean and the South Atlantic Convergence Zone,” *Geophys. Res. Lett*, vol. 31, pp. L03204–1, 2004.

[32] D. Almeida, R. A. F., P. Nobre, R. J. Haarsma, and E. J. D. Campos, “Negative ocean-atmosphere feedback in the South Atlantic Convergence Zone,” *Geophys. Res. Lett*, vol. 34, no. 18, p. L18809, 2007.

[33] T. M. Lenton, “Early warning of climate tipping points,” *Nat. Clim. Chang.*, vol. 1, no. 4, pp. 201–209, 2011.

[34] N. M. Shnerb, P. Sarah, H. Lavee, and S. Solomon, “Reactive grass and vegetation patterns,” *Phys. Rev. Lett.*, vol. 90, p. 038101, 2003.

[35] M. Scheffer, S. Carpenter, J. A. Foley, C. Folke, and B. Walker, “Catastrophic shifts in ecosystems,” *Nature*, vol. 413, no. 6856, pp. 591–596, 2001.

[36] M. Rietkerk, S. C. Dekker, P. C. de Ruiter, and J. van de Koppel, “Self-organized patchiness and catastrophic shifts in ecosystems,” *Science*, vol. 305, no. 5692, pp. 1926–1929, 2004.

[37] S. Kéfi, M. Rietkerk, C. L. Alados, Y. Pueyo, V. P. Papanastasis, A. ElAich, and P. C. De Ruiter, “Spatial vegetation patterns and imminent desertification in Mediterranean arid ecosystems,” *Nature*, vol. 449, no. 7159, pp. 213–217, 2007.

[38] S. Kéfi, V. Dakos, M. Scheffer, E. H. Van Nes, and M. Rietkerk, “Early warning signals also precede non-catastrophic transitions,” *Oikos*, vol. 122, pp. 641–648, 2013.

[39] M. Hirota, M. Holmgren, E. H. Van Nes, and M. Scheffer, “Global resilience of tropical forest and savanna to critical transitions,” *Science*, vol. 334, no. 6053, pp. 232–235, 2011.

[40] Z. Yin, S. C. Dekker, B. van den Hurk, and H. A. Dijkstra, “Bimodality of woody cover and biomass caused by vegetation structures in West Africa,” *Earth Sys. Dynam. Discuss.*, vol. 5, no. 1, pp. 83–120, 2014.

[41] V. Dakos, S. Kéfi, M. Rietkerk, E. H. van Nes, and M. Scheffer, “Slowing down in spatially patterned ecosystems at the brink of collapse,” *Am. Nat.*, vol. 177, no. 6, pp. 154–166, 2011.

[42] M. Scheffer, J. Bascompte, W. A. Brock, V. Brovkin, S. R. Carpenter, V. Dakos, H. Held, E. H. V. Nes, M. Rietkerk, and G. Sugihara, “Early-warning signals for critical transitions,” *Nature*, vol. 461, no. 7260, pp. 53–59, 2009.

[43] C. Wissel, “A universal law of the characteristic return time near thresholds,” *Oecologia*, vol. 65, no. 1, pp. 101–107, 1984.

[44] V. Guttal and C. Jayaprakash, “Spatial variance and spatial skewness: leading indicators of regime shifts in spatial ecological systems,” *Theor. Ecol.*, vol. 2, no. 1, pp. 3–12, 2009.

[45] V. Dakos, E. H. van Nes, R. Donangelo, H. Fort, and M. Scheffer, “Spatial correlation as leading indicator of catastrophic shifts,” *Theor. Ecol.*, vol. 3, no. 3, pp. 163–174, 2010.

[46] J. M. Drake and B. D. Griffen, “Early warning signals of extinction in deteriorating environments,” *Nature*, vol. 467, no. 7314, pp. 456–459, 2010.

[47] S. R. Carpenter, J. J. Cole, M. L. Pace, R. Batt, W. A. Brock, T. Cline, J. Coloso, J. R. Hodgson, J. F. Kitchell, D. A. Seekell, *et al.*, “Early warnings of regime shifts: a whole-ecosystem experiment,” *Science*, vol. 332, no. 6033, pp. 1079–1082, 2011.

[48] L. Dai, D. Vorselen, K. S. Korolev, and J. Gore, “Generic indicators for loss of resilience before a tipping point leading to population collapse,” *Science*, vol. 336, no. 6085, pp. 1175–1177, 2012.

[49] É. Blayo, M. Bocquet, E. Cosme, and L. F. Cugliandolo, *Advanced Data Assimilation for Geosciences: Lecture Notes of the Les Houches School of Physics: Special Issue, June 2012*. Oxford University Press, 2014.

[50] R. Kistler, W. Collins, S. Saha, G. White, J. Woollen, E. Kalnay, M. Chelliah, W. Ebisuzaki, M. Kanamitsu, V. Kousky, *et al.*, “The NCEP-NCAR 50-year reanalysis: Monthly means CD-ROM and documentation,” *Bull. Am. Meteor. Soc.*, vol. 82, no. 2, pp. 247–267, 2001.

[51] D. Dee, S. Uppala, A. Simmons, P. Berrisford, P. Poli, S. Kobayashi, U. Andrae, M. Balmaseda, G. Balsamo, P. Bauer, *et al.*, “The ERA-Interim reanalysis: Configuration and performance of the data assimilation system,” *Quarterly Journal of the Royal Meteorological Society*, vol. 137, no. 656, pp. 553–597, 2011.

[52] Y. Berezin, A. Gozolchiani, O. Guez, and S. Havlin, “Stability of climate networks with time,” *Sci. Rep.*, vol. 2, p. 666, 2012.

[53] A. Gozolchiani, S. Havlin, and K. Yamasaki, “Emergence of El Niño as an autonomous component in the climate network,” *Phys. Rev. Lett.*, vol. 107, no. 14, p. 148501, 2011.

[54] E. A. Martin and J. Davidsen, “Estimating time delays for constructing dynamical networks,” *Nonlin. Process. Geophys.*, vol. 21, no. 5, pp. 929–937, 2014.

[55] M. Mudelsee, *Climate Time Series Analysis: Classical Statistical and Bootstrap Methods*, vol. 397. Springer, 2010.

[56] T. M. Cover and J. A. Thomas, *Elements of information theory.* John Wiley & Sons, 2012.

[57] C. Bandt and B. Pompe, “Permutation entropy: A natural complexity measure for time series,” *Phys. Rev. Lett.*, vol. 88, no. 17, p. 174102, 2002.

[58] A. D. Wyner, “A definition of conditional mutual information for arbitrary ensembles,” *Information and Control*, vol. 38, no. 1, pp. 51–59, 1978.

[59] T. Schreiber, “Measuring information transfer,” *Phys. Rev. Lett.*, vol. 85, no. 2, p. 461, 2000.

[60] A. Bahraminasab, F. Ghasemi, A. Stefanovska, P. V. McClintock, and H. Kantz, “Direction of coupling from phases of interacting oscillators: a permutation information approach,” *Phys. Rev. Lett.*, vol. 100, no. 8, p. 084101, 2008.

[61] M. Paluš, “From nonlinearity to causality: statistical testing and inference of physical mechanisms underlying complex dynamics,” *Contemporary Physics*, vol. 48, no. 6, pp. 307–348, 2007.

[62] I. I. Mokhov, D. A. Smirnov, P. I. Nakonechny, S. S. Kozlenko, E. P. Seleznev, and J. Kurths, “Alternating mutual influence of El-Niño Southern Oscillation and Indian monsoon,” *Geophys. Res. Lett.*, vol. 38, pp. 2–6, 2011.

[63] G. A. Seber and A. J. Lee, *Linear regression analysis*, vol. 936. John Wiley & Sons, 2012.

[64] G. Schwarz, “Estimating the dimension of a model,” *Ann. Stat*, vol. 6, no. 2, pp. 461–464, 1978.

[65] C. Hiemstra and J. D. Jones, “Testing for linear and nonlinear Granger causality in the stock price-volume relation,” *The Journal of Finance*, vol. 49, no. 5, pp. 1639–1664, 1994.

[66] A. Brovelli, M. Ding, A. Ledberg, Y. Chen, R. Nakamura, and S. L. Bressler, “Beta oscillations in a large-scale sensorimotor cortical network: directional influences revealed by Granger causality,” *Proc. Natl. Acad. Sci.*, vol. 101, no. 26, pp. 9849–9854, 2004.

[67] L. Sommerlade, K. Henschel, J. Wohlmuth, M. Jachan, F. Amtage, B. Hellwig, C. H. Lücking, J. Timmer, and B. Schelter, “Time-variant estimation of directed influences during parkinsonian tremor,” *Journal of Physiology-Paris*, vol. 103, no. 6, pp. 348–352, 2009.

[68] G. D. Salvucci, J. A. Saleem, and R. Kaufmann, "Investigating soil moisture feedbacks on precipitation with tests of Granger causality," *Advances in Water Resources*, vol. 25, no. 8, pp. 1305–1312, 2002.

[69] T. J. Mosedale, D. B. Stephenson, M. Collins, and T. C. Mills, "Granger causality of coupled climate processes: Ocean feedback on the north atlantic oscillation," *J. Climate*, vol. 19, p. 1182, 2006.

[70] G. Caldarelli, *Scale-Free Networks: complex webs in nature and technology.* Oxford University Press, 2007.

[71] http://www.coolantarctica.com/Antarctica

[72] J. F. Donges, Y. Zou, N. Marwan, and J. Kurths, "The backbone of the climate network," *EPL*, vol. 87, no. 4, p. 48007, 2009.

[73] K. Steinhaeuser, N. V. Chawla, and A. R. Ganguly, "Complex Networks as a Unified Framework for Descriptive Analysis and Predictive Modeling in Climate Science," *Statistical Analysis and Data Mining: The ASA Data Science Journal*, vol. 4, no. 5, pp. 497–511, 2011.

[74] 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. Discuss.*, vol. 4, pp. 743–783, 2014.

[75] Z. Levnajić and A. Pikovsky, "Network reconstruction from random phase resetting.," *Phys. Rev. Lett.*, vol. 107, 2011.

[76] M. Timme and J. Casadiego, "Revealing networks from dynamics: an introduction," *J. Phys. A*, vol. 47, 2014.

[77] M. Timme, "Revealing Network Connectivity from Response Dynamics," *Phys. Rev. Lett.*, vol. 98, no. 22, p. 224101, 2007.

[78] S. G. Shandilya and M. Timme, "Inferring network topology from complex dynamics," *New J. Phys.*, vol. 13, no. 1, p. 13004, 2011.

[79] J. Ren, W.-X. Wang, B. Li, and Y.-C. Lai, "Noise bridges dynamical correlation and topology in coupled oscillator networks," *Phys. Rev. Lett.*, vol. 104, no. 5, p. 58701, 2010.

[80] D. Yu and U. Parlitz, "Inferring network connectivity by delayed feedback control," *PLoS One*, vol. 6, no. 9, p. e24333, 2011.

[81] N. Rubido, A. C. Martí, E. Bianco-Martínez, C. Grebogi, M. S. Baptista, and C. Masoller *New Journal of Physics*, vol. 16, no. 9, p. 093010, 2014.

[82] E. S. C. Ching, P.-Y. Lai, and C. Y. Leung, "Extracting connectivity from dynamics of networks with uniform bidirectional coupling," *Phys. Rev. E*, vol. 88, no. 4, p. 42817, 2013.

[83] S. Bialonski, M. Horstmann, and K. Lehnertz, "From brain to earth and climate systems: small-world interaction networks or not ?," *Chaos*, vol. 20, p. 013134, 2010.

[84] J. Hlinka *et al.*, "Reliability of inference of directed climate networks using conditional mutual information.," *Entropy*, vol. 15, no. 6, pp. 2023–2045, 2013.

[85] 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, no. 5, p. 59003, 2013.

[86] E. Bullmore and O. Sporns, "Complex brain networks: graph theoretical analysis of structural and functional systems," *Nat. Rev. Neurosci.*, vol. 10, no. 3, pp. 186–198, 2009.

[87] A. A. Tsonis and P. J. Roebber, "The architecture of the climate network," *Physica A*, vol. 333, pp. 497–504, 2004.

[88] T. L. *et al.*, "Characterizing the evolution of climate networks.," *Nonlin Processes Geophys.*, vol. 21, pp. 705–711, 2014.

[89] J. I. Deza, M. Barreiro, and C. Masoller, "Inferring interdependencies in climate networks constructed at inter-annual, intra-season and longer time scales," *Eur. Phys. J. Spec. Top*, vol. 222, no. 2, pp. 511–523, 2013.

[90] S. Scarsoglio, F. Laio, and L. Ridolfi, "Climate Dynamics: A Network-Based Approach for the Analysis of Global Precipitation," *PLoS One*, vol. 8, no. 8, p. e71129, 2013.

[91] V. M. Eguiluz, D. R. Chialvo, G. A. Cecchi, M. Baliki, and A. V. Apkarian, "Scale-free brain functional networks," *Phys. Rev. Lett.*, vol. 94, no. 1, p. 18102, 2005.

[92] A. P. Masucci, S. Arnaud-Haond, V. M. Eguíluz, E. Hernández-García, and E. A. Serr ao, "Genetic flow directionality and geographical segregation in a cymodocea nodosa genetic diversity network.," *EPJ Data Science*, vol. 1, pp. 1–11, 2012.

[93] W. X. Wang, Y. C. Lai, C. Grebogi, and J. Ye, "Network reconstruction based on evolutionary-game data via compressive sensing.," *Phys. Rev. X*, vol. 1, 2011.

[94] J. A. Acebrón, L. L. Bonilla, C. J. P. Vicente, F. Ritort, and R. Spigler, "The Kuramoto model: A simple paradigm for synchronization phenomena," *Rev. Mod. Phys.*, vol. 77, no. 1, p. 137, 2005.

[95] J. F. Donges, Y. Zou, N. Marwan, and J. Kurths, "Complex networks in climate dynamics: Comparing linear and nonlinear network construction methods," *Eur. Phys. J. Spec. Top.*, vol. 174, no. 1, pp. 157–179, 2009.

[96] M. Barreiro, A. C. Marti, and C. Masoller, "Inferring long memory processes in the climate network via ordinal pattern analysis," *Chaos*, vol. 21, no. 1, p. 13101, 2011.

[97] K. Steinhaeuser, A. R. Ganguly, and N. V. Chawla, "Multivariate and multiscale dependence in the global climate system revealed through complex networks," *Clim. Dyn.*, vol. 39, no. 3-4, pp. 889–895, 2011.

[98] K. Yamasaki, a. Gozolchiani, and S. Havlin, “Climate networks around the globe are significantly affected by El Niño,” *Phys. Rev. Lett.*, vol. 100, no. 22, p. 228501, 2008.

[99] J. F. Donges, H. C. H. Schultz, N. Marwan, Y. Zou, and J. Kurths, “Investigating the topology of interacting networks,” *Eur. Phys. J. B*, vol. 84, no. 4, pp. 635–651, 2011.

[100] A. Gozolchiani, O. Gazit, K. Yamasaki, and S. Havlin, “Dynamical Patterns of Climate Networks : Blinking Links and Stable Structures,” vol. 11, no. 2008, p. 228501, 2009.

[101] O. Guez, a. Gozolchiani, Y. Berezin, S. Brenner, and S. Havlin, “Climate network structure evolves with North Atlantic Oscillation phases,” *EPL*, vol. 98, no. 3, p. 38006, 2012.

[102] V. Stolbova, P. Martin, B. Bookhagen, N. Marwan, and J. Kurths, “Topology and seasonal evolution of the network of extreme precipitation over the Indian subcontinent and Sri Lanka,” *Nonlin. Proc. Geoph.*, vol. 21, no. 4, pp. 901–917, 2014.

[103] V. Rossi, E. Ser-Giacomi, C. López, and E. Hernández-García, “Hydrodynamic provinces and oceanic connectivity from a transport network help designing marine reserves,” *Geophys. Res. Lett.*, vol. 41, no. 8, pp. 2883–2891, 2014.

[104] A. A. Tsonis, K. L. Swanson, and P. J. Roebber, “What do networks have to do with climate?,” *B. Am. Meteorol. Soc.*, vol. 87, no. 5, pp. 585–596, 2006.

[105] G. Tirabassi, R. Sevilla-Escoboza, J. M. Buldú, and C. Masoller, “Inferring the topology of small random networks with a statistical similarity analysis,” *to be published in Sci. Rep.*, 2015.

[106] Y. Kuramoto, *Chemical oscillations, waves, and turbulence.* Courier Dover Publications, 2003.

[107] T. L. Carroll and L. M. Pecora, *Nonlinear dynamics in circuits.* World Scientific, 1995.

[108] A. N. Pisarchik, R. Jaimes-Reátegui, and J. H. García-López, “Synchronization of coupled bistable chaotic systems: experimental study,” *Philos. Trans. R. Soc. A Math. Phys. Eng. Sci.*, vol. 366, no. 1864, pp. 459–473, 2008.

[109] A. N. Pisarchik, R. Jaimes-Reátegui, J. R. Villalobos-Salazar, J. H. Garcia-Lopez, and S. Boccaletti, “Synchronization of chaotic systems with coexisting attractors.,” *Phys. Rev. Lett.*, vol. 96, no. 24, p. 244102, 2006.

[110] E. A. Martin, M. Paczuski, and J. Davidsen, “Interpretation of link fluctuations in climate networks during El Nino periods,” *EPL*, vol. 102, no. 4, p. 48003, 2013.

[111] A. Arenas, A. Díaz-Guilera, J. Kurths, Y. Moreno, and C. Zhou, “Synchronization in complex networks,” *Phys. Rep.*, vol. 469, no. 3, pp. 93–153, 2008.

[112] G. Tirabassi and C. Masoller, “Unravelling the community structure of the climate system by using lags and symbolic analysis,” *under submission.*

[113] C. Masoller, Y. Hong, S. Ayad, F. Gustave, S. Barland, A. J. Pons, S. Gómez, and A. Arenas, "Quantifying sudden changes in dynamical systems using symbolic networks," *New Journal of Physics*, vol. 17, no. 2, p. 023068, 2015.

[114] M. Rosvall and C. T. Bergstrom, "An information-theoretic framework for resolving community structure in complex networks," *Proc. Natl. Acad. Sci.*, vol. 104, no. 18, pp. 7327–7331, 2007.

[115] G. Tirabassi, C. Masoller, and M. Barreiro, "A study of the air–sea interaction in the South Atlantic Convergence Zone through Granger causality," *Int. J. Clim.*, p. 59003, 2014.

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

[117] C. S. Bretherton, C. Smith, and J. M. Wallace, "An intercomparison of methods for finding coupled patterns in climate data," *J. Climate*, vol. 5, pp. 541–560, 1992.

[118] M. T. Kayano, R. V. Andreoli, and R. A. F. de Souza, "Relations between ENSO and the South Atlantic SST modes and their effects on the South American rainfall," *Int. J. Clim.*, vol. 33, no. 8, pp. 2008–2023.

[119] G. N. Kiladis and K. M. Weickmann, "Horizontal structure and seasonality of large-scale circulations associated with submonthly tropical convection," *Monthly Weather Review*, vol. 125, pp. 1997–2013, 1997.

[120] I. F. A. Cavalcanti and M. T. Kayano, "High-frequency patterns of the atmospheric circulation over the Southern Hemisphere and South America," *Meteorology and Atmospheric Physics*, vol. 69, no. 3-4, pp. 179–193, 1999.

[121] M. Alexander and J. Scott, "The influence of ENSO on air-sea interaction in the atlantic," *Geophys. Res. Lett*, vol. 29, no. 14, pp. 46–51, 2002.

[122] A. M. Grimm and T. Ambrizzi, *Teleconnections into South America from the tropics and extratropics on interannual and intraseasonal timescales.* Past Climate Variability in South America and Surrounding Regions pp 159–191, Springer, 2009.

[123] M. Barreiro, A. Giannini, P. Chang, and R. Saravanan, *On the Role of the South Atlantic Atmospheric Circulation in Tropical Atlantic Variability in Earth's Climate.* (eds C. Wang S.P. Xie and J.A. Carton), American Geophysical Union, Washington, D. C, 2004.

[124] F. N. M. Oliveira, L. Carvalho, and T. Ambrizzi, "A new climatology for Southern Hemisphere blockings in the winter and the combined effect of ENSO and SAM phases," *Int. J. Climatol*, vol. 34, no. 5, pp. 1676–1692, 2014.

[125] G. Tirabassi, J. Viebahn, V. Dakos, H. A. Dijkstra, C. Masoller, and M. Rietkerk, "Interaction network based early-warning indicators of vegetation transitions," *Ecol. Complex.*, vol. 19, pp. 148–157, 2014.

[126] V. Guttal and C. Jayaprakash, “Impact of noise in bistable ecological systems,” *Ecol. Modell.*, vol. 201, pp. 420–428, 2007.

[127] M. Mheen, H. A. Dijkstra, A. Gozolchiani, M. den Toom, J. Feng Q. Kurths, and E. Hernandez-Garcia, “Interaction network based early warning indicators for the Atlantic MOC collapse,” *Geophys. Res. Lett.*, vol. 40, pp. 2714–2719, 2013.

[128] E. H. Van Nes and M. Scheffer, “Slow recovery from perturbations as a generic indicator of a nearby catastrophic shift,” *Am. Nat.*, vol. 169, no. 6, pp. 738–747, 2007.

[129] J. Viebahn and H. A. Dijkstra, “Critical transition analysis of the deterministic wind-driven ocean circulation – a flux-based network approach,” *Int. J. Bifurc. Chaos*, vol. 24, no. 2, p. 1430007, 2014.

[130] C. Boettiger and A. Hastings, “Early warning signals and the prosecutor’s fallacy,” *Proc. R. Soc. B Biol. Sci.*, vol. 279, no. 1748, pp. 4734–4739, 2012.

[131] V. Dakos, S. R. Carpenter, W. A. Brock, A. M. Ellison, V. Guttal, A. R. Ives, S. Kefi, V. Livina, D. A. Seekell, E. H. Van Nes, *et al.*, “Methods for detecting early warnings of critical transitions in time series illustrated using simulated ecological data,” *PLoS One*, vol. 7, no. 7, p. e41010, 2012.

[132] H. A. Dijkstra, “Vegetation Pattern Formation in a Semi-Arid Climate,” *Int. J. Bifurc. Chaos*, vol. 21, no. 12, pp. 3497–3509, 2011.

[133] S. C. Dekker, M. Rietkerk, and M. F. P. Bierkens, “Coupling microscale vegetation–soil water and macroscale vegetation–precipitation feedbacks in semiarid ecosystems,” *Glob. Chang. Biol.*, vol. 13, no. 3, pp. 671–678, 2007.

[134] 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, no. 3, p. 033105, 2015.

Publications

1. Tirabassi, G. and Masoller, C. On the effects of lag-times in networks constructed from similarities of monthly fluctuations of climate fields. *EPL* 102, 59003 (2013).

2. Tirabassi, G., Viebahn, J., Dakos, V., Dijkstra, H.A., Masoller, C., Rietkerk, M. and Dekker, S.C. Interaction network based early-warning indicators of vegetation transitions. *Ecol. Complex.* 19, 148–157 (2014).

3. Tirabassi, G., Masoller, C. and Barreiro, M. A study of the air-sea interaction in the South Atlantic Convergence Zone through Granger causality. *Int. J. Climatol.* (2014). DOI: 10.1002/joc.4218

4. Tirabassi, G., Sevilla-Escoboza, R., Buldú, J.M. and Masoller, C. Inferring the topology of small random networks with a statistical similarity analysis. *To be published in Sci. Rep.* (2015).

5. Tirabassi, G. and Masoller, C. Unravelling the community structure of the climate system by using lags and symbolic analysis. *Under revision in Geophys. Res. Lett.*

Additional work not related to the thesis:

6. Tirabassi, G., Montani, G., and Carlevaro, N. Self-collimated axial jet seeds from thin accretion disks. *Phys. Rev. E* 88, 043101 (2013).

Conferences and Workshops

- Data Analysis and Modeling in Earth Sciences (DAMES - 2012) (Potsdam, 8-10 oct 2012), poster presentation.
  Poster title: "On the effects of lag-times in networks constructed from similarities of monthly fluctuations of climate fields".
  http://dames.pik-potsdam.de
- XVIII Congreso de Física Estadística, FISES 12, (Palma de Mallorca, 18-20 oct 2012), poster presentation
  Poster title: "On the effects of lag-times in networks constructed from similarities of monthly fluctuations of climate fields".
  http://www.gefenol.es/FisEs/12/
- Max Plank "Scales and patterns in the earth system" Workshop (Dresden,05-09 Nov 2012).
  http://www.mpipks-dresden.mpg.de/earth12/
- Dynamic Days Europe 2013 (Madrid, Spain 3-7 June, 2013), oral presentation.
  Presentation Title: "No signs of lag-time effects in the connectivity of climate networks constructed with surface temperature anomalies".
  http://dynamics-days-europe-2013.org
- 3rd LINC Workshop and Mid-Term Review, (Potsdam, Germany, 17-20 Nov. 2013).
  http://www.climatelinc.eu/events/linc-mid-term-review-17-20-nov-potsdam/
- WCRP Conference for Latin America and the Caribbean (Montevideo, Uruguay, March 17-21, 2014), oral presentation.
  Presentation title: "A study of the air-sea interaction in the South Atlantic Convergence Zone through Granger causality"
  http://www.cima.fcen.uba.ar/WCRP/
- 4th LINC Workshop, (Montevideo, Uruguay, 24-26 Mar, 2014).
  http://www.climatelinc.eu/events/linc-workshop-4-march-2014-montevideo-uruguay/

- European Geosciences Union (EGU) General Assembly 2014 (Vienna, Austria, 27 April-02 May 2014), poster presentation.

Poster title: "A study of the air-sea interaction in the South Atlantic Convergence Zone through Granger causality"
http://www.egu2014.eu/

- European conference in complex systems 2014 (ECCS'14) (Lucca, Italy 22-26 Sept, 2014), poster presentation.
  Poster title: "Disentangling the air-sea interaction in the South Atlantic Convergence Zone"
  http://www.eccs14.eu
- 5th LINC Workshop, (Lucca, Italy, 27-29 Sept 2014).
  http://www.climatelinc.eu/events/eccs-14-and-linc-workshop-5-sept-2014-/lucca- italy/
- 6th LINC Workshop, (Wien, Austria, 13-17 Apr 2015).
  http://www.egu2015.eu
- European Geosciences Union (EGU) General Assembly 2015 (Vienna, Austria, 13 April-17 April 2015), poster presentation.
  Poster title: "New Climate Network Communities from Anomalies and Seasonal Cycle Time-Series"
  http://www.egu2015.eu/

Schools, Courses and Research Stays

- Euroscience open forum (ESOF - 2012) and Marie Curie Actions Conference (Dublin, 11-16 jul 2012). On Marie Curie Actions Conference I attended to a public speaking workshop.
  http://esof2012.org
- School IBERSINC Workshop (Barcelona4-6set2012).
  http://www.ibersinc.org
- 1^{st} LINC Schooland Workshop (10-13sep2012).
  http://www.climatelinc.eu/events/1st-linc-school-10-12-sept-2012-mallorca/
- 2^{nd} LINC School (Soesterberg,TheNetherlands, 21-30 April 2013).
  http://www.climatelinc.eu/events/2nd-linc-school-21-30-april-2013-the-netherlands/
- SICC international School “Complex networks in action” (Catania, Italy, 26-28 Sep 2012).
  http://www.dees.unict.it/mfrasca/PhDSchool2012.html
- Research stay in Universidad de la República, (Montevideo, Uruguay, 1 Sep-31 Oct 2013).
  http://meteo.fisica.edu.uy/
- "School and Workshop on Weather Regimes and Weather Types in the Tropics and Extra-tropics: Theory and Application to Prediction of Weather and Climate", (Trieste, Italy, 21 - 30 Oct 2013).
  http://indico.ictp.it/event/a12220/
- Research stay in Utrecht University (Utrecht, Netherlands, 1 Nov 2013-28 Feb 2014).
  http://imau.nl
- Alpine summer school: Dynamics, Stochasticity and Predictability of the Climate System (Valsavarenche, Valle d’Aosta, Italy, 9-18 June, 2014).
  http://www.to.isac.cnr.it/aosta_old/aosta2014/index.htm
- Advanced Course on Climate Time-Series Analysis at Mudelsee Climate Risk Analysis, (Hannover, Germany, (21-23 Oct 2014)
  http://www.manfredmudelsee.com