

Universitat Politècnica de Catalunya

Doctoral Thesis

---

**Climate networks constructed by using  
information-theoretic measures and  
ordinal time-series analysis**

---

*Author:*

Juan Ignacio Deza

*Supervisors:*

Prof. Cristina Masoller

Prof. Marcelo Barreiro

*A thesis submitted in fulfilment of the requirements  
for the degree of Doctor in Physics  
in the*

Nonlinear Dynamics, Nonlinear Optics and Lasers Group  
Departament de Física i Enginyeria Nuclear





To my parents, Ana María and Roberto,  
to my sister and brother, Agustina and Luis,  
and to every person who has ever helped me to be a better person.



# Contents

<b>Contents</b>	<b>iv</b>
<b>List of Figures</b>	<b>vii</b>
<b>Abbreviations</b>	<b>xi</b>
<b>Summary</b>	<b>xiii</b>
<b>Resumen</b>	<b>xvii</b>
<b>Acknowledgements</b>	<b>xxi</b>
<b>Part 1. Introduction</b>	<b>1</b>
<b>Preface</b>	<b>3</b>
<b>1 Temporal and spatial scales in climate phenomena</b>	<b>5</b>
1.1 Climate data records . . . . .	5
1.1.1 Surface air temperature . . . . .	7
1.2 Climate time-scales . . . . .	8
1.3 Elements of the climate system . . . . .	10
1.4 Patterns of climate variability . . . . .	12
1.4.1 El Niño-Southern Oscillation (ENSO) . . . . .	12
1.4.2 North Atlantic Oscillation (NAO) . . . . .	15
1.5 Overview of the thesis . . . . .	18
<b>2 Climate Networks</b>	<b>19</b>
2.1 Complex Networks . . . . .	19

2.1.1	Properties of complex networks . . . . .	23
2.2	Construction and representation of climate networks . . . . .	26
2.2.1	Data set . . . . .	27
2.2.2	Network construction . . . . .	28
2.2.3	Area Weighted Connectivity . . . . .	30
2.3	Climate phenomena and climate networks . . . . .	32
<b>3</b>	<b>Climate Time Series Analysis</b>	<b>37</b>
3.1	Linear techniques . . . . .	37
3.1.1	Cross Correlation . . . . .	37
3.1.2	Empirical Orthogonal Functions . . . . .	40
3.2	Information theoretic tools . . . . .	41
3.2.1	Entropy . . . . .	41
3.2.2	Mutual Information . . . . .	43
3.2.3	Directionality measure . . . . .	43
3.3	Ordinal analysis . . . . .	45
3.3.1	Ordinal Patterns . . . . .	46
3.4	Statistical Significance . . . . .	49
3.4.1	Surrogate Data and Bootstrap . . . . .	51
	<b>Part 2. Results</b>	<b>55</b>
<b>4</b>	<b>Interdependencies in Climate Networks on Several Time Scales</b>	<b>57</b>
4.1	Data and methodology . . . . .	57
4.2	Results . . . . .	59
4.3	Discussion and Conclusions . . . . .	65
<b>5</b>	<b>Components of Atmospheric Variability and Time Scale Identification</b>	<b>67</b>
5.1	Internal and forced variability . . . . .	68
5.2	Data sets and model used . . . . .	70
5.2.1	Model Validation . . . . .	71
5.3	AWC maps . . . . .	74
5.3.1	Forced variability . . . . .	74

5.3.2	Internal variability . . . . .	77
5.4	Node connectivity maps . . . . .	79
5.4.1	Forced variability . . . . .	81
5.4.2	Internal variability . . . . .	83
5.5	Summary and conclusions . . . . .	85
<b>6</b>	<b>Directionality of Climate Interactions</b>	<b>87</b>
6.1	Statistical significance analysis . . . . .	88
6.2	Analysis of monthly-averaged SAT anomalies . . . . .	90
6.2.1	Influence of the parameter $\tau$ . . . . .	90
6.2.2	Comparison between monthly and daily datasets . . . . .	93
6.3	Analysis of daily-averaged SAT anomalies . . . . .	94
6.3.1	Influence of the parameter $\tau$ . . . . .	94
6.3.2	Influence of $\tau$ in the extra-tropics . . . . .	96
6.3.3	Influence of $\tau$ in the tropics . . . . .	99
6.4	Directionality on the tropical Pacific Ocean . . . . .	99
6.5	Conclusions . . . . .	101
<b>7</b>	<b>Conclusions</b>	<b>103</b>
	<b>Bibliography</b>	<b>109</b>
	<b>Publications</b>	<b>123</b>
	<b>Presentations at International Conferences</b>	<b>125</b>
	<b>Attendance to schools and research stays</b>	<b>129</b>

# List of Figures

1.1	The composite power spectrum of climate variability showing the amount of variance in each frequency range. . . . .	9
1.2	Organization of the climate system. . . . .	11
1.3	Mean “El Niño” and “La Niña” and NINO3.4 index. . . . .	14
1.4	NAO+ phase and NAO- phase. . . . .	16
1.5	Time series of the North Atlantic Oscillation index. . . . .	17
2.1	Classical Königsberg bridges problem. . . . .	20
2.2	Schematic representation of a network. . . . .	21
2.3	Metabolic network. . . . .	22
2.4	Small-world network properties. . . . .	25
2.5	Degree centrality examples . . . . .	26
2.6	How data is gridded in a world map. . . . .	27
2.7	Correlation and adjacency matrices of a Climate network . . . . .	29
2.8	Degree distribution of a climate network . . . . .	30
2.9	Example of representation of CNs. Map of AWC and map of connections to a given point. . . . .	31
3.1	Several sets of (X, Y) points, with the Pearson correlation coefficient for each set. . . . .	39
3.2	Graphical representation of the linear removal procedure. . . . .	40
3.3	An example for ordinal patterns of (a) $n = 3$ and (b) $n = 4$ letters. . . . .	47
3.4	An example of three ordinal patterns in the time series of the NINO 3.4 index (monthly averaged). . . . .	48
3.5	Classical bell-shaped (Gaussian) curve. . . . .	50
3.6	The mechanism of statistical testing used. . . . .	51
3.7	Illustration of the method p-value for (a) the MI and (b) DI. . . . .	53



4.1	AWC and connectivity maps (for a point in the equatorial Pacific) for different methods using 6 bins and the statistical significance threshold. . . . .	60
4.2	AWC and connectivity maps (for a point in the equatorial Pacific) for different methods using 24 bins and the statistical significance threshold. . . . .	61
4.3	AWC and connectivity maps for different methods using 6 bins. The threshold was chosen such that the networks have the same link density (0.03%). . . . .	63
4.4	AWC and connectivity maps for different methods using 24 bins. The threshold was chosen such that the networks have the same link density (0.03%). . . . .	64
5.1	Maps of AWC constructed from reanalysis NCEP/NCAR data. . . . .	72
5.2	Maps of AWC obtained from single model run. . . . .	73
5.3	Maps of AWC computed from averaged time series, only containing information of the forced component of atmospheric variability. . . . .	75
5.4	Maps of AWC of the forced component of the network when the ENSO influence is removed from the time series. . . . .	76
5.5	Maps of averaged AWC, revealing the internal variability network. . . . .	78
5.6	Maps of averaged AWC, revealing the internal variability network when the NAO index is removed from the time series . . . . .	79
5.7	Connectivity map of a node in central Pacific. Forced Variability. . . . .	80
5.8	Connectivity map of a node near Alaska. Forced Variability. . . . .	81
5.9	Connectivity map of a node near New Zealand. Forced Variability. . . . .	82
5.10	Maps of internal variability showing the connectivity of a node in the Labrador Sea. . . . .	84
6.1	Procedure of constructing significant directionality maps from raw DI calculations. . . . .	89
6.2	Plot of DI ( $\tau = 1$ month) <i>vs.</i> MI for all the links of the two nodes in the Pacific and Indian oceans considered in Fig. 6.1. . . . .	90
6.3	Effect of $\tau$ on tropical areas using monthly-averaged data. . . . .	91
6.4	Directionality of the significant links ( $\tau = 1$ month) of two nodes in the extratropics. . . . .	92
6.5	Comparison of results for monthly-averaged and daily-averaged datasets. . . . .	93
6.6	Influence of the parameter $\tau$ when DI is computed from daily-averaged data. . . . .	95
6.7	Effect of $\tau$ in the southern extratropics, when DI is computed from daily data. . . . .	97
6.8	As in Fig 6.7 but for a node in the northern extratropics (the Labrador Sea) . . . . .	98
6.9	Effect of $\tau$ using daily data for a node in the tropics. . . . .	100

6.10 The zonal change of directionality over the equatorial pacific. . . . . 101





# ***Abbreviations***

<b>AGCM</b>	<b>Atmospheric General Circulation Model</b>
<b>AWC</b>	<b>Area Weighted Connectivity</b>
<b>CMI</b>	<b>Conditional Mutual Information</b>
<b>CN</b>	<b>Climate Network</b>
<b>DAS</b>	<b>Data Assimilation System</b>
<b>DI</b>	<b>Directionality Index (based in TE)</b>
<b>ENSO</b>	<b>El Niño Southern Oscillation</b>
<b>EOF</b>	<b>Empirical Orthogonal Functions</b>
<b>GCM</b>	<b>General Circulation Model</b>
<b>GH</b>	<b>Geopotential Height</b>
<b>MI</b>	<b>Mutual Information</b>
<b>MIH</b>	<b>Mutual Information using Histograms of the series</b>
<b>MIOP</b>	<b>Mutual Information using Ordinal Patterns</b>
<b>NAO</b>	<b>North Atlantic Oscillation</b>
<b>OP</b>	<b>Ordinal Pattern</b>
<b>PCA</b>	<b>Principal Component Analysis (another name for EOF)</b>
<b>SAT</b>	<b>Surface Air Temperature</b>
<b>SST</b>	<b>Sea Surface Temperature</b>
<b>TE</b>	<b>Transfer Entropy</b>



# Summary

---

This Thesis is devoted to the construction of global climate networks (CNs) built from time series—surface air temperature anomalies (SAT)—using nonlinear analysis. Several information theory measures have been used including mutual information (MI) and conditional mutual information (CMI).

The ultimate goal of the study is to improve the present understanding of climatic variability by means of networks, focusing on the different spatial- and time-scales of climate phenomena.

An introduction to the main components of this interdisciplinary work are offered in the first three chapters. Climate variability and patterns are introduced Chapter 1, network theory in Chapter 2, and nonlinear time series analysis—especially information theoretic methodology—in Chapter 3.

In Chapter 4, the statistical similarity of SAT anomalies in different regions of the world is assessed using MI. These climate networks are constructed from time series of monthly-averaged SAT anomalies, and from their symbolic ordinal representation, which allows an analysis of these interdependencies on different time scales. This analysis allows identifying topological changes in the networks when using ordinal patterns (OPs) of different time intervals. Intra-seasonal (of a few months), inter-seasonal (covering a year) and inter-annual (several years) timescales are considered. An increase in the ordinal pattern spacing (namely, in the timescale of the ordinal analysis), results in climate networks with increased connectivity in the equatorial Pacific area. Reciprocally, the number of significant links decrease when the ordinal analysis is done with a shorter timescale (i.e. by comparing consecutive months). These effects are interpreted as the El Niño–Southern Oscillation (ENSO) forcing on long scales together with the presence of more stochasticity in the time series at the shorter timescale.

The nature of the interdependencies is then explored in Chapter 5 by using SAT data from an ensemble of atmospheric general circulation model (AGCM) runs, all of them forced by the same historical sea surface temperature (SST). It is possible to separate atmospheric variability into a forced component, and another one intrinsic to the atmosphere. In this way, it is possible to obtain climate networks for both types of variability and characterize them. Furthermore, an analysis using OP allows to construct CNs for several time scales, and evaluate the connectivity of each different network. This selecting both time scale and variability type allows to obtain a further insight into the study of SAT anomalies. The connectivity of the constructed CNs allows to assess the influence of two main climate phenomena: ENSO and the North Atlantic Oscillation (NAO). To do so, these phenomena are linearly removed from the time series and the analysis of these new series can be compared to the analysis of the original ones. The connectivity of the forced variability network is heavily affected by ENSO: removing the NINO3.4 index—that characterizes ENSO—yields a general loss of connectivity. The fact that even connections between regions far away from the equatorial Pacific ocean are lost, suggests that these regions are not directly linked but rather connected via ENSO, particularly at interannual time scales. On the other hand, on the internal variability network—independent of SST forcing—the links connecting the Labrador Sea with the rest of the world are found to be significantly affected by NAO, with a maximum at intra-annual time scales. While the strongest non-local links found are those forced by the ocean, the presence of long range teleconnections associated with internal atmospheric variability is also shown.

In Chapter 6, a natural extension of the network construction methodology is implemented in order to infer the *direction* of the links. A directionality index (DI) is used. DI can be defined as the difference of the CMI between two time series  $x(t)$  and  $y(t)$ , calculated in two ways: i) considering the information about  $x(t)$  contained in  $\tau$  time units in the past of  $y(t)$ , and ii) considering the information about  $y(t)$  contained in  $\tau$  time units in the past of  $x(t)$ . DI is used to quantify the direction of information flow among the series, indicating the direction of the links of the network. Two SAT datasets—one monthly-averaged and another daily-averaged—are used. The links of the obtained networks are interpreted in terms of known atmospheric tropical and extra-tropical variability phenomena. Specific and relevant geographical regions are selected, the net direction of propagation of the atmospheric patterns is analyzed, and the direction of the inferred links is tested using surrogate



data. These patterns are also found to be acting on various time scales, such as synoptic atmospheric waves in the extra-tropics or longer time scale events in the tropics. The dependence of the values of DI with  $\tau$  is investigated. For synoptic time scales ( $\tau < 10$  days), DI is shown to exhibit a dependence with  $\tau$ , with a minimum of connectivity in the tropics, and a maximum—in the form of a train of waves—in the extra-tropics. For larger values of  $\tau$ , links are found to be relatively robust to the choice of this parameter, showing a high connectivity in the tropics and a low connectivity in the extra-tropics. The analysis demonstrates the capability of the DI to infer the net direction of climate interactions, and to improve the present understanding of climate phenomena and climate predictability. The resulting directed network is found to be in full agreement with state-of-the-art knowledge in climate phenomena, validating this methodology for inferring—directly from the data—the net directionality of climate interactions.

The final Chapter 7 presents the main conclusions, and a discussion of future work.



# ***Resumen***

---

El objetivo de esta tesis es la creación de redes climáticas (CN por las siglas en inglés) a partir de un conjunto global de series temporales de temperatura del aire superficial (SAT), utilizando técnicas de análisis no lineal de series temporales. Varias metodologías son aplicadas al estudio de la variabilidad climática, incluyendo la Información mutua (MI) y la información mutua condicional (CMI).

El objetivo principal de esta tesis es estudiar la variabilidad climática a través del análisis de redes haciendo énfasis en los diferentes patrones espaciales y temporales del sistema climático.

Una introducción a los componentes principales de este trabajo interdisciplinario se presenta en los primeros tres capítulos. La variabilidad climática y los patrones atmosféricos se introducen en el Capítulo 1, la teoría de redes en el Capítulo 2, y el análisis no lineal de series temporales, especialmente métodos en teoría de la información, en el Capítulo 3.

En el Capítulo 4, la similitud estadística de las anomalías de SAT en diferentes regiones del mundo es evaluada utilizando MI. Estas redes climáticas globales son construidas a partir de series temporales de SAT promediadas a escalas de tiempo mensuales, y a partir de su representación simbólica, permitiendo un análisis de estas interdependencias en varias escalas temporales. Se identifican cambios topológicos entre las redes, como resultado de variaciones en el intervalo de construcción de los OP. Escalas intra-estacionales (unos meses), inter-estacionales (cubriendo un año) e inter-anuales (varios años), son consideradas. Se encuentra que un incremento en el espaciado de los patrones ordinales (por lo tanto, en la escala de tiempo del análisis ordinal), resulta en redes climáticas con un incremento en la conectividad en el Pacífico ecuatorial. Al contrario, el número de conexiones significativas decrece al realizar el análisis ordinal en una escala de tiempo más corta (es decir, comparando meses consecutivos). Este efecto es interpretado como una consecuencia del efecto de

El Niño-Oscilación Sud (ENSO) actuando en escalas de tiempo más largas y de una mayor estocasticidad en las series temporales en escalas de tiempo más cortas.

La naturaleza de las interdependencias es explorada en el Capítulo 5, utilizando datos de SAT, resultantes de un conjunto de salidas de un modelo atmosférico de circulación global (AGCM), todas forzadas por la misma temperatura de la superficie del mar (SST). Es posible separar la variabilidad atmosférica en una componente forzada y otra intrínseca a la atmósfera. De esta forma, se obtienen redes climáticas para ambos tipos de variabilidad, lo que posibilita caracterizarlas. Un análisis utilizando OP permite crear CNs para diferentes escalas temporales, y encontrar la escala de OP para la cual las diferentes redes presentan mayor conectividad. Este doble proceso de selección permite estudiar la variabilidad de las anomalías de SAT desde un nuevo punto de vista. La conectividad de las redes climáticas así construídas permite evaluar la influencia de dos fenómenos climáticos: ENSO y la Oscilación del Atlántico Norte (NAO). Para esto, se pueden comparar las redes originales, con redes provenientes de series temporales a las que se les quitaron linealmente estos fenómenos. Un resultado clave de este análisis es que la conectividad de la red de variabilidad forzada es muy afectada por ENSO: eliminando el índice NINO3.4 (que caracteriza ENSO), se provoca una pérdida general de la conectividad en la red. El hecho de que incluso conexiones entre áreas muy alejadas del océano Pacífico ecuatorial se hayan perdido al quitar el índice, sugiere que estas regiones no están directamente conectadas sino que ambas son influenciadas por la zona dominada por ENSO, especialmente en escalas de tiempo inter-anales. Por otro lado, en la red de variabilidad interna, independiente del forzado de las SST, las conexiones del Mar del Labrador con el resto del mundo resultan significativamente afectadas por NAO, con un máximo en escalas intra-anales. Aunque las conexiones no locales más fuertes resultan las forzadas por el océano, se muestra la presencia de teleconexiones asociadas con la variabilidad interna.

En el Capítulo 6, una extensión natural de la metodología de construcción de redes es implementada, permitiendo inferir la *dirección* de las conexiones. Un índice de direccionalidad (DI), puede ser definido como la diferencia entre la CMI entre dos series temporales  $x(t)$  e  $y(t)$  calculada de dos formas: i) considerando la información de  $x(t)$  contenida en  $\tau$  unidades de tiempo en el pasado de  $y(t)$  y ii) considerando la información de  $y(t)$  contenida en  $\tau$  unidades de tiempo en el pasado de  $x(t)$ . Este índice DI, se utiliza para cuantificar la dirección del flujo de información entre las series, lo que equivale a la dirección de la conexión

entre los respectivos nodos de la red. Dos conjuntos de series temporales, uno promediado mensualmente y el otro promediado diariamente, son usados. Las conexiones de las redes resultantes son interpretadas en términos de fenómenos de variabilidad tropical y extra-tropical conocidos. Regiones específicas y relevantes son seleccionadas, la dirección neta de propagación de los patrones atmosféricos es analizada y contrastada con un test de inferencia estadística. Se encuentra que diferentes patrones de variabilidad, actúan en varias escalas de tiempo, tales como ondas sinópticas atmosféricas en los extra-trópicos o escalas de tiempo mayores en los trópicos. La dependencia de valores de DI con  $\tau$  es investigada. Para la escala sinóptica ( $\tau < 10$  días), DI presenta una dependencia con  $\tau$ , con un mínimo en los trópicos y máximos (en forma de trenes de ondas) en los extra-trópicos. Para valores mayores de  $\tau$ , los links resultan ser relativamente robustos a la elección del parámetro, mostrando una conectividad alta en los trópicos y baja en los extra trópicos. El análisis demuestra la capacidad de DI de inferir la dirección neta de las interacciones climáticas, y de mejorar la comprensión actual de fenómenos climáticos y de la predictabilidad climática. La red resultante está en total acuerdo con los conocimientos actuales de fenómenos climáticos, validando esta metodología para inferir, directamente de los datos, la dirección neta de las interacciones climáticas.

Finalmente, el Capítulo 7, presenta las conclusiones, y una discusión de trabajo futuro.



# ***Acknowledgements***

---

To say thank you is more than a mere formality, it is a way to admit that no matter how hard I could have worked, I could have never done it without the support, and help of others. I'm very grateful to all of you who directly or indirectly helped me through this research, and, more importantly, through life.

Firstly I want to thank Cristina Masoller. For her patience, for her support, for the freedom she gave me during my research, and the time she invested in me. And overall because she believed in me from the very beginning. I also thank Marcelo Barreiro, who introduced me to an exciting new field, and helped me to understand the difference between studying about climate and studying climate.

I'm grateful to all the People from the LINC project. Especially the seniors: Henk Dijkstra, Emilio Hernandez García, Jürgen Kurths, Avi Golzochiani, Shlomo Havlin, Marcus Abel, Johan Dijkzeul, Karsten Ahnert and Manfred Mudelsee. With some I had very inspiring conversations, others helped me in some specific topics, or gave me wise advise. I have also read your papers and your books, all this contact made me improve and grow.

I also want to thank my fellow LINC colleagues, my friends: especially Giulio, which whom I shared a great deal of things during my time here, but also to the rest of you crazy people! Veronika, Miguel, Verónica, Fernando, Víctor, Hisham, Alexis, Enrico, Liuba, Miguel, Qingyi, Dong and Yang. You all know how, and how much I'm in debt with you (hint: a lot). I hope to be in contact with you for the years to come, no matter how life goes.

From the people from Terrassa I want to thank Ramón Vilaseca, Kestas Staliunas, Carme Torrent, José Trull, Ramon Herrero, Simone Pigolotti, Toni Pons, Crina Cojocar, Laura Carpi, Carles Serrat, Muriel Botey, Josep Lluís Font, José Antonio Diego, and Jordi García Ojalvo. Thank you for your help, for the talks, and simply for being there. Thanks to Montse Gea, to Jaume and to Cristina Fernández for your patience and for your help.

To all the people from work who made my life more enjoyable! Shubham, Andrés, Lina, Ricard, Carlos, Dani, Taciano, John, Bingxia, Yu Chieh, Auro, Sandro, Juan Luis, Safae, Petya, Cristina, Heather, Tomasz, Belén, Marta, Pau, Houssam, Nikhil, Lara, Jose María, Jordi Tiana and Jordi Zamora. Also the people from Uruguay. Pablo, Andrés, Stefanie, Juan, Rodrigo, Nicolás Díaz, Inés, Federico, Franco, Yamil, Florencia, Matilde, Ana, Javier, Nicasio, Santiago, César, Alain, Enrique, Nicolás Rubido, Madeleine and more. Thank you for your support with my studies and the things I learned from you, but mostly for the coffees we shared, for the beers and drinks we had, for the skiing, beach, sightseeing, or lazy days together, the parties, and for all the many random conversations. Those small things that, with time, become the important things to remember.

Thank you to all my family: my parents, my brother and sister, my niece Laureana, my grandmother Chiche, my Aunts Coca, Cristina, Adriana and Tisbe, my uncle Jorge, and all the huge number of cousins (Constanza, Raúl, Marysel, Fernando, Gustavo, Sonia, Federico, Santiago, Nico, Sofía y Carito), second order cousins, the children of most of them, even if I haven't seen them yet, and more and even more people I have seen maybe once or twice in my life. But you are my family. You always show me that you believe in me. You can't imagine how important is this for me, during the dark days—or worse, during those gray, monotonous days—when my self esteem just vanishes and I only want to hide and go away.

I want to thank a good friend I made during my stay in Terrassa, somebody who had to put up with my bad days and my constant whining and lighted my life with all her tropical rhythm and azúcar. Thank you Dennys for being always there.

A big thank you to all my friends (some of them are getting double thank you, you see how life goes). You are too many to name here. You know who you are, and I'll say thank you in private. With time. Thank you for your support, for the good vibes, for the experiences and conversations we shared, for showing me your points of view, especially when I didn't agree, for helping me to grow, to improve, to learn, and to enjoy life. I'm not going to say more here. But believe me, you are in my heart. And will be, forever.

Thank you, the reader of this thesis. No matter why you are reading this, as getting to you is the final purpose of my work. I believe science is altruistic, does not look for praise nor glory nor fame, it is based on small steps, of unknown and forgotten people, just like me. I hope you will find this work interesting, and you will use parts of what I have done here to build something new and beautiful, as good ideas should be, for the good of all.



And thank God.



# PART 1

## INTRODUCTION



# ***Preface***

Earth's climate is a highly complex dynamical system comprising an immensely wide range of time and spatial ranges [1, 2]. *Climate variability* is defined as changes of some variables (e.g. temperature, pressure, or rainfall), over months, seasons and years (for shorter time scales the term *weather variability* is used) with respect to their long term mean. Natural atmospheric variability can be partly explained in terms of patterns—or “oscillations” as first climatologists named them—which constitute recurrent variations of climatic variables with periods in the order of several years, decades or longer, over the long term mean. Some of these variability patterns have been named; for example “El Niño - Southern Oscillation” (ENSO), the “North Atlantic Oscillation” (NAO), the “Pacific Decadal Oscillation” (PDO), the “Atlantic Multidecadal Oscillation” (AMO), or the “Arctic—and Antarctic—Oscillations” (AO and AAO respectively) among others [3].

In the last few decades a great amount of effort has been focused on understanding how climate phenomena in one geographical region affects the climate in other regions. The mechanisms underlying patterns of climate variability are complex and still not fully understood, but it is widely accepted that climate in a region is not the result of local factors, as the atmosphere connects far away regions through waves and advection of mass, heat and momentum. These long-range couplings are called *teleconnections* and have been shown to be an ubiquitous feature of atmospheric dynamics.

For the purpose of modeling and forecasting, many systems lead naturally to the concept of networks of interacting elements, appearing in many fields of science and technology. Some examples of these are: social interactions [4], food webs [5], gene expression networks [6], brain functional networks [7], or the internet [8], to name only a few. In many of these systems, nodes and links can be assigned depending on the (in principle, very complex) features of the system under study. The goal being to map features of the system into the topology of the network, in order to apply network analysis tools (the well developed theory of *graphs*) to the systems under study.

Using this approach is then possible to extract relevant information about a system without yielding an oversimplification, or being forced to handle the full-scale detailed model which can obscure the interpretations or being computationally too expensive. This situation has motivated the use of networks for climate studies and set the birth of the research field of **climate networks** (CNs) [9–12].

Complex networks have indeed revealed themselves as a powerful framework for identifying climate interdependencies. However, in order to further exploit the knowledge of the links uncovered via the network analysis (for, e.g. improvements in prediction), a good understanding of the physical mechanisms represented by these links is required.

The work presented in this thesis is aimed at studying methods of climate network construction, by using nonlinear time series analysis, and information theory measures. Methods based on symbolic analysis, and the mutual information concept have been used. They are particularly capable of detecting and distinguishing time scales and thus are useful to classify variability patterns out of the data. A particular emphasis is made in comparing network construction methods, statistical inference and directionality measures, in order to detect not only the relevant links in the network but their direction as well, and to improve the use of CNs as a tool for modeling the Earth climate system.

## **Chapter**

# **1**

## ***Temporal and spatial scales in climate phenomena***

### **1.1 Climate data records**

Atmospheric patterns occur naturally in the atmosphere in a wide range of time scales. These span from hours—or even seconds—in the case of tornadoes, to millennia and even longer time scales for events like glacial eras or warm periods. Furthermore, what we informally refer to as “normal” climate in an area, changes with time, and probably was not considered normal only some generations ago. There is abundant historical evidence of this fact by means of tales, and paintings from different civilizations, showing how “typical” climate naturally changes. As an example, the tale “A Christmas Carol” of Charles Dickens [13], depicts a many centimeter snow layer in London in late December as something usual. This book dates from 1843, at the end of a period of several hundred years characterized by severe winters now called the “Little Ice Age”; present day London however is usually snowless in december and it only presents a couple of centimeters of snow yearly which seldom last as air temperatures below freezing are unusual [14]. Additionally, human-induced (antropogenic) climate change is superposed to natural variability, making future climate predictions even more uncertain as they are now also subjected to a new forcing with rather unknown consequences [15].

Climate variations can also be known from historical data, which usually is not detailed enough to be used for analysis or forecasting but it gives a rough idea of the conditions met in the past, i.e., how climate looked like before measurements were widespread. The path

taken, for example, by sailing ships or large wind-powered vessels recovered from their log-books during the sail era, shows that patterns of air circulation were different than those in present days. This can be explained due to patterns which occurred during the critical period that marked the transition out from the Little Ice Age in the late seventeenth century [16]. Another study, with historical consequences is reported in [17] where a chronology of El Niño events going back to 1525 is produced. It was compiled from historical reports of conditions from the coastal region and adjacent waters of northwestern South America made during the Spanish colony. As an anecdote, their work suggests, that Francisco Pizarro's conquest of the Incas in 1531–1532 coincided with an El Niño event (a phenomenon to be explained in section 1.4.1). Heavy rains and swollen rivers, which typically occur only during El Niño years in Peru, delayed Francisco Pizarro's advance through the countryside. On the other hand, the same rains made the desert areas bearable to transit, and produced abundant vegetation, providing plentiful fodder for his horses, which were one of the chief tactical advantages (along with swords) that his small contingent of soldiers had over the much more numerous natives.

Another path to understand ancient climates are *paleoclimatic* records [18, 19]. These are indirect measures of a climatic variable of the past through the use of proxies. A proxy is a geological variable which can yield information about an ancient climatic variable. They include ice caps (drilled so an ice core can be extracted and analyzed) [20], old trees and tree-rings (Dendroclimatology) [21], studies in the growing of coral (Sclerochronology), but also marine and land sediments [22, 23], and pollen [24], among many others (see [25] for a review). These methods not only have been used to state geological climate events from the very ancient Huronian glaciation [26]—occurred from 2400 Mya to 2100 Mya before present, but also have yield light on more recent phenomena. Two examples are the reconstruction of European climate during the Holocene—about 11.000 B.P.—using pollen data [27], and studies on the aforestated Little Ice Age using tree-ring data [28]. Paleo-data are immensely useful to understand the nature and extent of present day climate variability patterns, as with present day ENSO [29]. The transcription from proxy data to climatic variables is not without error [25], as e.g. trees have a huge influence on their growth from many unknown factors (humidity, rainfall, the precise species of tree, etc) which can affect the interpretation of the results, yielding sometimes to huge error bars in the data. Although paleo-data is usually of better quality than the historical records as they are not human-biased, the quality



of paleoclimatic records is still very low compared with direct observations. They remain, however, the only known method to characterize the climate of the distant past.

Current climate can be measured directly by using instruments of high precision and resolution, this resulting in the most reliable data source for analysis. Instrumental data can be station-based or gridded with a certain resolution, and this results in spatio-temporal information which allows to find patterns and relationships between far away areas. Phenomena like advection of heat, mass, and momentum, or the propagation of atmospheric waves generate teleconnections which are important properties of climate dynamics. Only in recent decades—after the invention of meteorological satellites [30]—atmospheric data has a full global spatial coverage. Time series dating before the twentieth century can only be trusted at certain locations—as in Europe and regions of north America—even if the most ancient records date to the end of the sixteenth century [31]. In other locations, as regions in Africa or the southern Pacific ocean, there exist almost data-less spots for times as near as the mid-twentieth century. Given the low frequency—decades, even longer—of the signals to be analyzed, this insufficiency of data is a very important problem for the study of climate dynamics, and yields difficult to differentiate naturally occurring multidecadal variability from external factors, as e.g. volcanism or human-caused climate change. Fortunately, data assimilation techniques are in place, creating what is called *reanalysis data*, integrating information from various sources in order to compensate (even partially) for this problem [32]. A reanalysis project involves reprocessing observational data spanning an extended historical period using a consistent modern analysis system, to produce a dataset that can be used for meteorological and climatological studies. Examples of reanalysis datasets include the NCEP/NCAR Reanalysis (USA) [33], the ECMWF re-analysis (Europe) [34], and the JRA-25 [35] reanalysis from the Japan Meteorological Agency.

### 1.1.1 Surface air temperature

During this thesis the climatic variable chosen was SAT (surface air temperature) which is defined as a measure of the average kinetic energy of air molecules in the atmosphere, either over land or over water and expressed in units of degrees Celsius. The variable is often called *near surface air temperature* as the data express the temperature of the air over the surface. In the field, these surface temperatures would be measured by thermometers placed approximately 1 meter above ground. This should not be confused with *surface sea tem-*

perature (SST) which reflects the temperature of water near the surface or the *surface skin temperature* data, which are calculated from measurements of microwave radiation at the land surface taken by satellite.

Temperature is a fundamental variable in climate, affected by all of the absorptive and reflective properties of the atmosphere, oceans, and land surface. The surface temperature is related to the amount of solar radiation a region receives, which is in turn dependent on both astronomical factors (which determine the seasons and longer climate cycles) and weather (clouds, atmospheric air movements, evaporation and precipitation, etc.) The surface temperature is also related to the average global temperature, which is determined by the Earth's radiative balance, as set by the absorptive and reflective properties of the atmosphere modulated by greenhouse gases as  $H_2O$  and  $CO_2$  among others. A changing average global temperature may influence local weather processes in ways that make regional temperature trends more pronounced than average global temperature trends [36,37].

## 1.2 Climate time-scales

Figure 1.1, shows a schematic representation of the “variance”—as called in the figure—or variability of climate in different time scales. No single time-series in terms of length (millions of years) and resolution (a couple of minutes) exists, or has ever been produced as this diagram suggests, making impossible to recreate this figure via, e.g., a Fourier Transform. This diagram has to be regarded as a summary of the time scales in climate dynamics originating from various sources. The leftmost part of the diagram, ranging hundreds of kya (kilo-years or thousands of years) originates from paleoclimatic data. It represents important events including glaciation cycles—alternating warm and freezing temperatures in the extra-tropical areas—as one of the most important features of the Quaternary. These glacial eras have occurred in a cyclic fashion, as reflected in the diagram as sharp peaks in the area between  $10^3$ – $10^6$  years. These peaks have been showed to be of “astronomical origin” as the frequencies of the glaciations are locked to the so called Milankovitch cycles—see [38,39] for the original papers (in german) and [40] for a review—of the Earth. These observations are regarded to be one of the cornerstones of climate science [41]. Moreover, even when these effects alone were too weak to trigger an event as a glaciation, a pioneer paper in stochastic resonance [42] has suggested this mechanism to explain the locking of the glacial periods to the Milankovitch cycles.

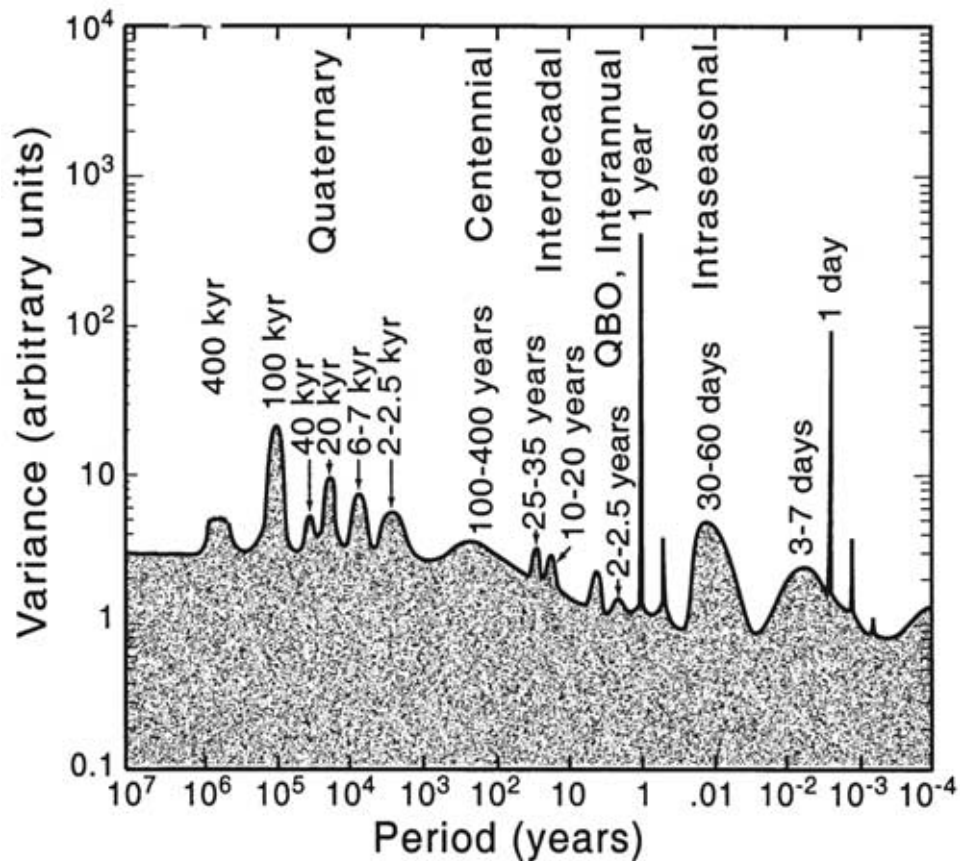


Figure 1.1: An ‘artist’s rendering’ of the composite power spectrum of climate variability showing the amount of variance in each frequency range. (From [2]).

From the higher frequency zone of Fig. 1.1 two peaks instantly catch the eye as they are tall and very narrow. These are the seasonal (one year) and the daily (one day) cycles, of obvious astronomical origin. They are accompanied by their harmonics represented as narrow but shorter peaks. The remaining variability sources are broader and respond to the atmosphere-ocean dynamics: from 2 – 2.5y as the Quasi-biennial oscillation (QBO) which is related to the stratosphere dynamics, through multi-year (ENSO, NAO, among others), decadal and multidecadal—Pacific Decadal Oscillation (PDO), and Atlantic Multidecadal Oscillation (AMO), etc.—and the less known centennial oscillations. For time scales shorter than a year, the intra-seasonal variability takes place (30-60 days). Finally, at 3-7 days, the *synoptic weather* dominates. It has the biggest impact for everyday life, especially in the

extra-tropics, dictating, e.g, the passage between a sunny and a rainy day. It is crucial for weather forecasts but it is usually averaged out for climate studies.

### 1.3 Elements of the climate system

The main reasons for climate to exhibit such a rich behavior are the many degrees of freedom of the system, the nonlinear nature of the many processes involved, and the strong interconnectivity of the multiple factors that affecting climate. This yields a whole spectrum of positive and negative feedbacks and its delicate balance decides the equilibrium state of a particular subsystem and of the system in general. Figure 1.2 shows a simplified scheme of these interconnections. The links show the connections between different sub-systems (atmospheric, soil or marine systems), the role of chemistry, and biochemistry, the role of external factors as the sun, volcanoes, and human activities and the flux of mass and energy between the different actors, following the hydrological, the carbon, the sulphur, or the nitrogen cycles.

This figure doesn't show the time scales associated to these phenomena. However, in many cases the response time to variations is fundamental in order to understand and predict the outcome to any forcing. Time scales for atmospheric phenomena range from *seconds* (e.g. formation of cloud droplets) to about a *week* (e.g. dissipation of midlatitude weather systems). For the ocean, these scales range from *months* (e.g. upper layer ocean mixing), to *thousands of years* (e.g. deep ocean circulation). Time scale ranges in the biosphere are very wide and include respiration and photosynthesis (seconds) in one extreme and changes in biochemistry due to evolutionary processes (millions of years) in the other extreme.

An example which illustrates feedbacks and how time scales are important, is the anthropogenic variations on the carbon cycle. An excess of carbon resulting from human activities as burning fossil fuels is counterbalanced on several time scales. In scales of seconds to minutes, plants usually take carbon out of the atmosphere through photosynthesis and release it back into the atmosphere via respiration.  $CO_2$  is also dissolved into water as wind and waves mix the oceans and the atmosphere. On longer time scales carbon is stored as cellulose in trees and other plants. Dead plant material can be incorporated into soils, where it might reside for years, decades or centuries before being broken down by soil microbes and released back to the atmosphere (short carbon cycle). On even longer time scales, carbon

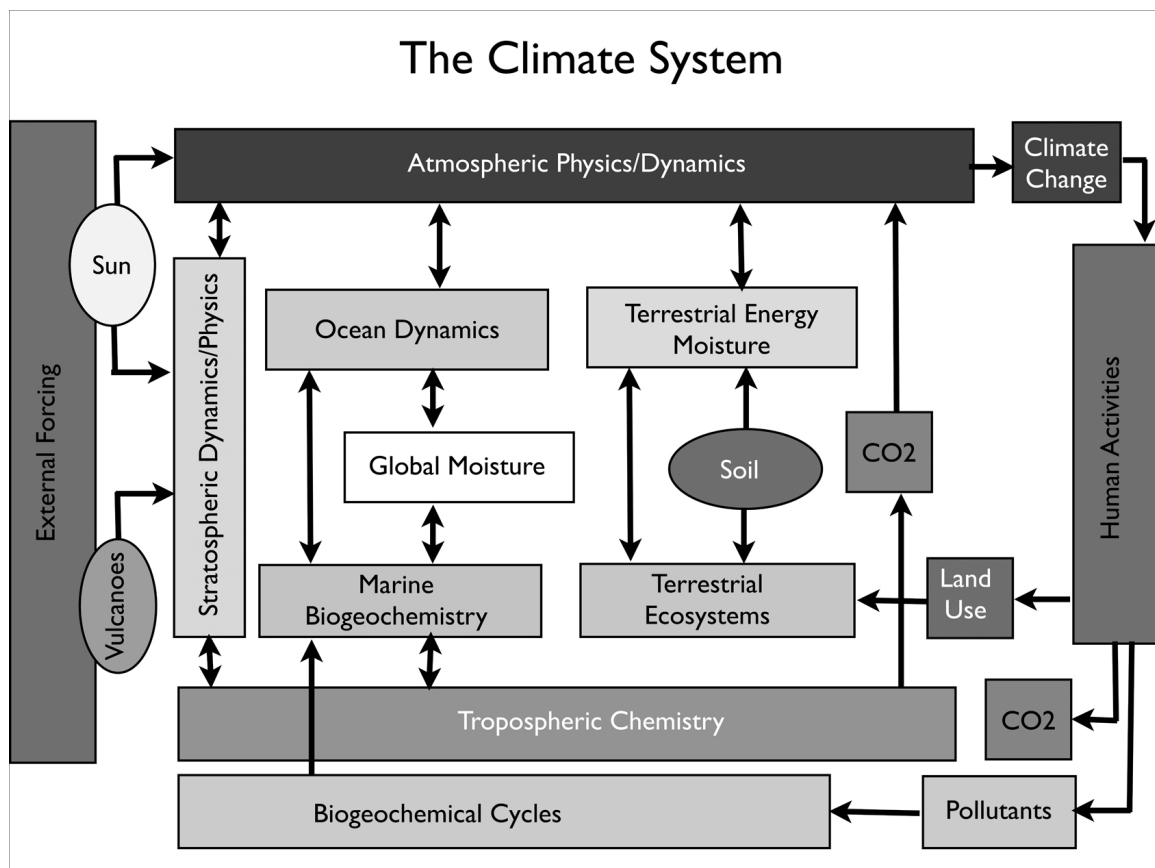


Figure 1.2: Organization of the Climate system. The different components as the atmosphere, the ocean, the land and the ecosystems, together with external forcings and the human activities and how they are interconnected. This is sometimes called the ‘horrendogram’ of the climate system. (From [1]).

containing organic matter that became buried in deep sediments (and protected from decay) is slowly transformed into deposits of coal, oil and natural gas (long carbon cycle). The excess of carbon in the atmosphere—responsible for climate change—is a problem, because the carbon we are releasing is *new*, in the sense that it has been stored for millions of years in sediments and it would take the same time scale to be finally counterbalanced by the long carbon cycle [43]. Note that this would not happen if using renewable fuel based factories only—burning wood, or vegetable coal—as no new carbon would have been added, allowing the short carbon cycle to act in the same time scales of the forcing.

## 1.4 Patterns of climate variability

In this thesis, two natural patterns were studied and analyzed in detail. These are ENSO and NAO. For reference purposes a climatological description of these phenomena will be given below, together with an historical introduction and an assessment of their impact in society. For a more general discussion on patterns of atmospheric variability, several books, as for example chapter 5 of [44] can be consulted.

An useful concept is a *climate index*. It can be used in order to synthesize the time information of a phenomenon in a single time series. A climate index describes the state and changes of a particular region of the ocean or the atmosphere. Indices are standardized and may be downloaded from many internet climate data repositories [45]. They also can be determined directly from time series from monitoring stations or reanalysis data, or identified by means of linear multivariate analysis as Empirical Orthogonal Functions (EOF) analysis (the EOF technique will be explained in section 3.1.2).

### 1.4.1 El Niño-Southern Oscillation (ENSO)

The El Niño-Southern Oscillation phenomenon is a natural part of the global climate system and results from large-scale interactions between the oceans and the atmosphere that occur mainly across the tropical-subtropical Pacific to Indian ocean basins [46–51]. ENSO explains a considerable amount of worldwide climatic variability. As a consequence, most direct climatic shifts, environmental and human impacts are found over, and in countries bordering, the Indo-Pacific sector of the planet [52]. Climate sensitive industries directly impacted by weather such as agriculture, construction, energy distribution, tourism, and outdoor recreation, account for a big share of the GDP of many countries. El Niño can affect commodity prices and the macroeconomy of different countries. It can constrain the supply of rain-driven agricultural commodities; reduce agricultural output, construction, and services activities; create food-price and generalized inflation; and may trigger social unrest in commodity-dependent poor countries that primarily rely on imported food [53].

El Niño—Spanish for a male child, in this case implicitly referring to the Christ Child as the phenomenon historically peaks around Christmas—initially referred to a weak, warm current appearing annually in late December along the coast of Ecuador and Peru and lasting from some weeks to a month only. Nowadays, El Niño refers to a basin warming in the

equatorial Pacific ocean. Every three to seven years, an El Niño event occurs, lasting for many months, having significant economic and atmospheric consequences worldwide [54].

In the eastern tropical Pacific, trade winds generally drive the surface waters westward. They accumulate warm surface water near Indonesia, raising the sea level roughly half a meter higher than in the eastern Pacific. As it moves away, the water is deflected by the Coriolis force, northward in the northern hemisphere and southward in the southern hemisphere, causing water to move away from the equator in both directions. This equatorial divergence induces upwelling in the eastern Pacific that brings colder water up from deeper levels to replace the surface water that has been dragged away. The upwelled water is rich in nutrients and supports an abundance of fish and marine life.

During El Niño years, however, the easterly trade winds are weakened interrupting the upwelling. This allows warmer water from the western Pacific to surge eastward, so the sea level flattens and the eastern Pacific warms. This warming changes the pressure patterns in the atmosphere over the area, producing the so called *Southern Oscillation* (SO), which is defined as the pressure differences between Tahiti and Darwin (Australia) and it is very much correlated to El Niño SSTs, that they are considered two aspects of the same phenomenon: El Niño Southern Oscillation (ENSO) [48].

The warming of the surface water in the eastern Pacific causes heavy rainfall in the area and also limits the amount of nutrient-rich deep water. These nutrients are vital for sustaining the large fish populations normally found in the region as any reduction in the supply of nutrients means a reduction in the fish population. Figure 1.3 (top) shows the mean sea surface temperature (SSTs) variations from the mean (anomalies), during Nov-Mar averaged for several “El Niño” years. Notice the anomalous hot temperature in the equatorial Pacific.

The opposite phase to El Niño is usually called La Niña—female child, because “anti-Niño” would have been too strong an image considering the religious origin of the name—refers to an anomaly of unusually cold sea surface temperatures found in the eastern tropical Pacific. Figure 1.3 (middle) shows the mean sea surface temperature (SSTs) anomalies, during Nov-Mar averaged for several “Niñas”. Notice the anomalous cold equatorial Pacific ocean. The frequency of this Niño-Niña oscillations has been suggested to change due to the increase in global  $CO_2$  concentration because of human activities [56, 57].

The NINO 3.4 index [54] (Fig. 1.3 bottom) is defined as the average of SST anomalies in the equatorial Pacific bounded by latitudes  $5S - 5N$  and by longitudes  $120W - 170W$  using

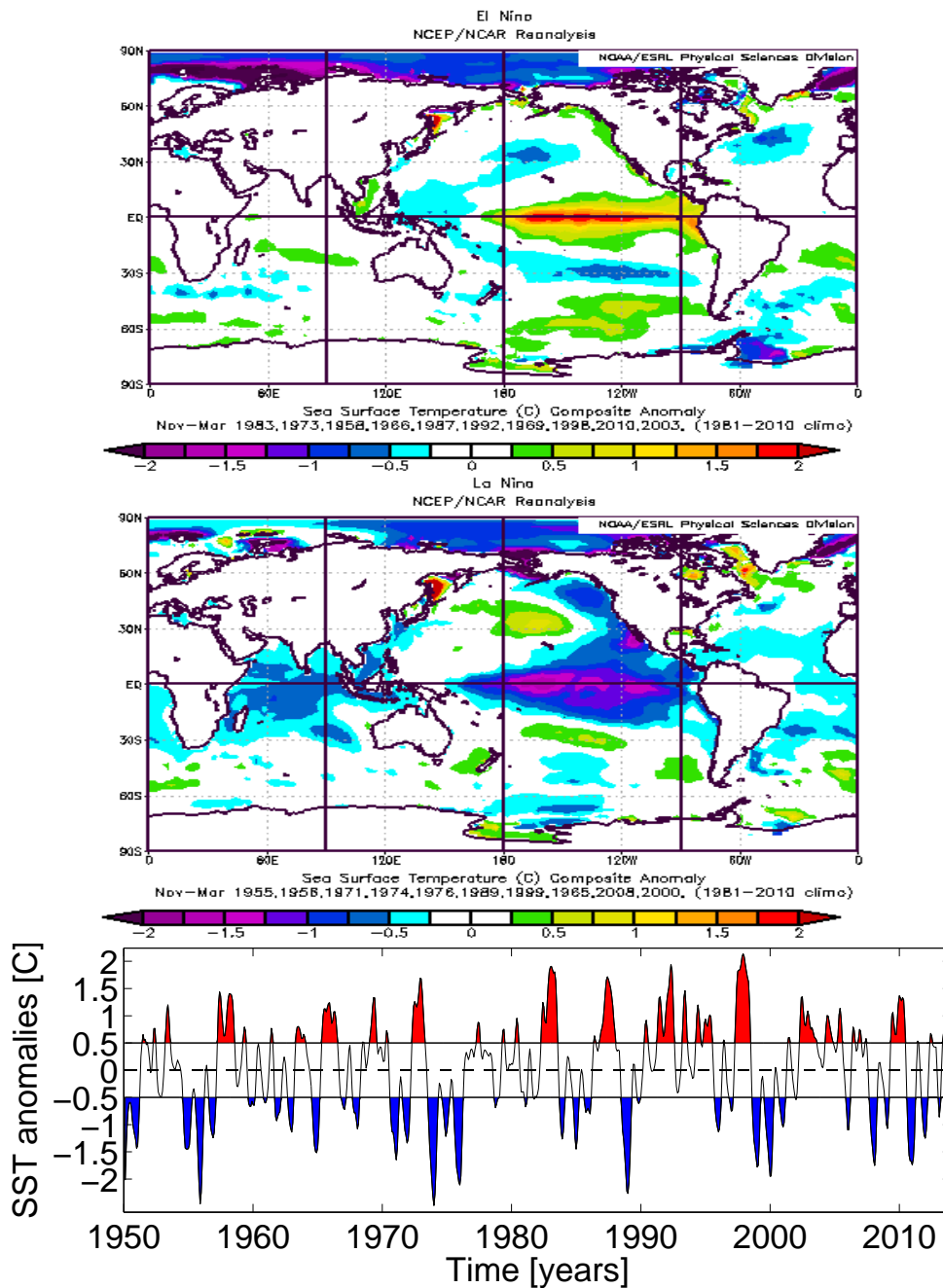


Figure 1.3: (top) SST anomalies mean for “El Niño” years: 1958, 1966, 1968, 1973, 1983, 1987, 1992, 1998, 2003, and 2010. (middle) SST anomalies mean for “la Niña” years: 1955, 1956, 1965, 1971, 1974, 1976, 1989, 1999, 2000, and 2008. Anomalies in both cases are calculated with respect to 1981-2001 climate and are adapted from [55].(Bottom) Plot of the NINO3,4 index filtered with a 3-month running average. Temperature anomalies over 0.5 C indicate “El Niño” conditions and are plotted in red, while when under  $-0.5$  C indicate “La Niña” conditions and are plotted in blue. Data from the National Center for Atmospheric Research (NCAR). USA [45].



SST data. It resumes the information about the temporal behavior of the phenomena. It can be also downloaded from many online repositories (e.g. [45]).

### 1.4.2 North Atlantic Oscillation (NAO)

The North Atlantic Oscillation (NAO) is one of the most prominent and recurrent patterns of atmospheric circulation variability in the northern hemisphere [58]. It affects climate variability from the eastern seaboard of the United States to Siberia and from the Arctic to the subtropical Atlantic, especially during boreal winter, so variations in the NAO are important to society and for the environment in most of the developed world. The NAO refers to a redistribution of atmospheric mass between the Arctic and the subtropical Atlantic, and swings from one phase to another produce large changes in the mean winds, the heat and moisture transport between the Atlantic and the neighboring continents, and the intensity, path and number of storms. Agricultural harvests, water management, energy supply and demand, and yields from fisheries, among many other things, are directly affected by the NAO [59,60].

The NAO is one of the oldest known weather modulating patterns, as some of its earliest descriptions were from seafaring Scandinavians several centuries ago. The history of scientific research on the NAO is rich, an historical review can be found in chapter 2 of [58]. In this text the words of the Danish missionary Hans Egede Saabye (1745), showing the recognition of a wintertime pattern, are recollected:

*“In Greenland, all winters are severe, yet they are not alike. The Danes have noticed that when the winter in Denmark was severe, as we perceive it, the winter in Greenland in its manner was mild, and conversely.”*

Strong positive phases (NAO+) tend to be associated with above-average temperatures in the eastern United States and across northern Europe and below average temperatures in Greenland and oftentimes across southern Europe and the Middle East. They are also associated with above-average precipitation over northern Europe and Scandinavia in winter, and below-average precipitation over southern and central Europe. Opposite patterns of temperature and precipitation anomalies are typically observed during strong negative phases (NAO-). During particularly prolonged periods dominated by one particular phase of the NAO, anomalous height and temperature patterns are also often seen extending well into central Russia and north-central Siberia (Fig. 1.4).

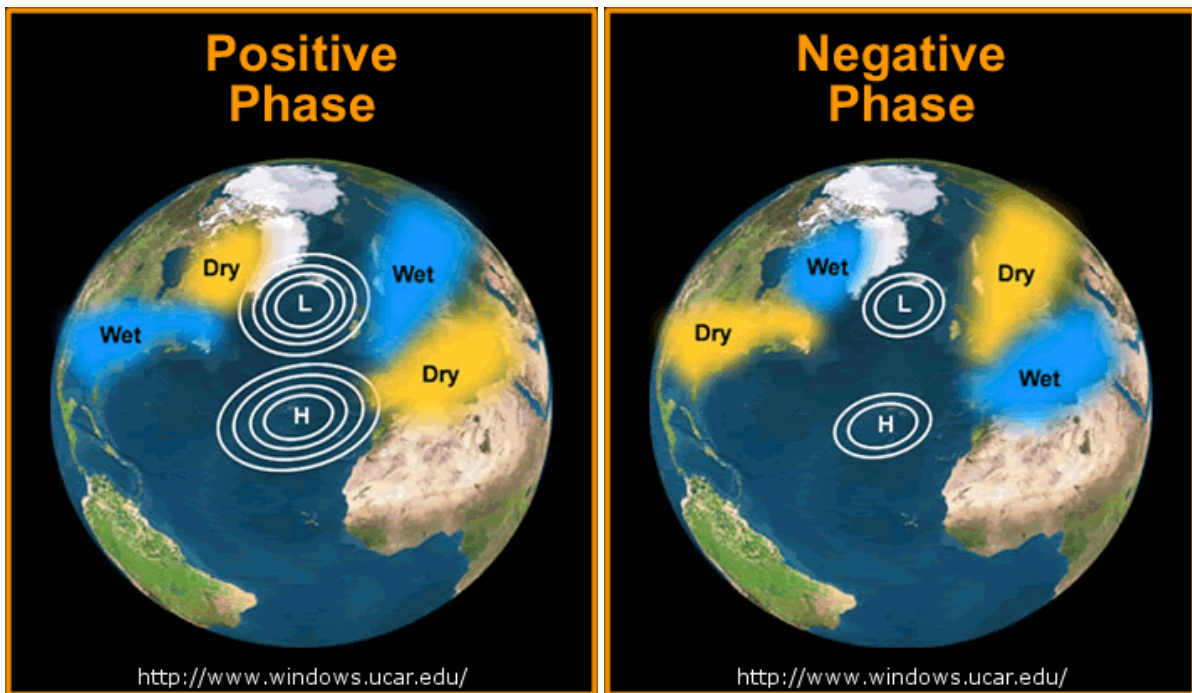


Figure 1.4: (a) NAO+ phase and (b) NAO- phase. White contours depict the seasonal mean sea-level pressure field; regions in red and blue were respectively warmer and colder than normal; regions in green and brown respectively experienced higher and lower rainfall than normal. The presence of stronger winds from the west (westerlies) over northern Europe during a NAO+ phase (a) is marked with a grey arrow (from [61] ).

There is no unique way to define the spatial structure of the NAO, or its temporal evolution. It was historically defined as the difference of sea level pressure in December to February between Ponta Delgada (Azores) and Stykkisholmur (Iceland). However, several other approaches exist, some of them using EOF (see [58] and also section 3.1.2 of this thesis). In this last case the index is defined as the time series—also called principal component (PC)—of the evolution of the spatial pattern which explains the biggest amount of variance in the region (the leading EOF pattern). In this thesis this has been the method preferred to identify the NAO index in chapter 5. Indices can be also downloaded from internet repositories as [45]. Figure 1.5 (top), shows the NAO index. In red the positive phase NAO+ and in blue the negative NAO-. Filtering the time series with a one year running average (Fig 1.5) allows distinguishing considerable variability on interannual and multi decadal time scales [58,62].

For example, as seen in the bottom plot of Fig. 1.5, the negative phase of the NAO dom-

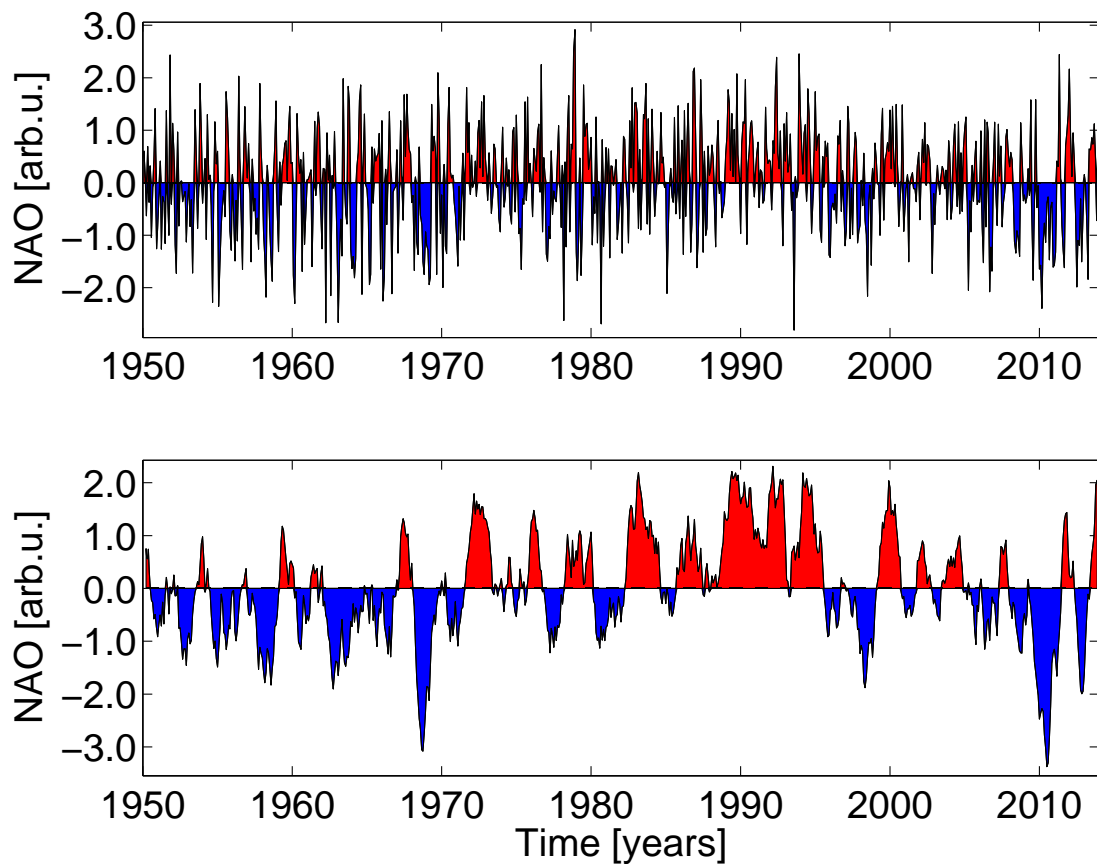


Figure 1.5: (top) Time series of the North Atlantic Oscillation index. Observed (black line) winter NAO index, December to February 1950 – 2014. The NAO+ phases are shown in red while NAO- is shown in blue. The lower graph shows the same time series filtered by a one-year running mean in order to appreciate the decadal variability of the pattern. (Data from [45]).

inated the circulation from the mid-1950's through the end of the 1970's, only to become essentially positive during the 1980's and 1990's. However, the NAO is a mode of variability predominantly *internal to the atmosphere*. The fact that not all of its variability can be attributed to intra-seasonal stochastic atmospheric processes points to a role for external forcing [63] and, perhaps, to a small but useful amount of predictability [59]. In [58] it was shown that the north Atlantic ocean responds to the NAO, because the changes in surface wind patterns associated with NAO have an influence on heat transfer and freshwater exchange on the Atlantic surface. The relationship between the NAO and variations in surface temperature, storms and precipitation, and thus the economy, and how the ocean and

ecosystem respond to NAO variability, are still under study [64–67].

## 1.5 Overview of the thesis

As stated before, the aim of this work is to study methods of climate network construction. Hence, the research covers fields of **climate dynamics**, **climate networks**, and **time series** analysis; this last part focusing in **information theoretic** quantities. Accordingly, after a primer of climate variability and patterns given in this chapter, an introduction to climate networks is given in **Chapter 2** where the methodology of construction of the networks is explained together with some definitions of complex network measures. **Chapter 3** deals with time series analysis, information theory similarity measures such as mutual information and transfer entropy are introduced. This chapter also introduces the method of symbolic analysis used, known as *ordinal patterns*, and the criteria used for assessing the statistical significance of the results.

Chapters 4, 5, and 6 present the main results obtained in the thesis. In **chapter 4** several construction methods are compared. Statistical tests focused on discovering weak but significant links in the network are used and the links are interpreted as the signature of teleconnections patterns. Using Ordinal patterns a time scale analysis of different phenomena is also presented, and the obtained networks are compared. **Chapter 5** deals with the classification of climate network links by *variability type*, separating the—usually entangled—forced and internal atmospheric variability. The role of the strongest variability patterns on the network connectivity on a global scale was assessed. **Chapter 6** deals with the analysis of the *direction* of the obtained links. Using a directionality measure based on the conditional mutual information, the direction of existing links is inferred. This provides an alternative approach to CNs construction, allowing to assess the transfer of information, by means of teleconnection patterns and wave propagation. Finally **Chapter 7** summarizes the conclusions and discusses possible lines of research for future work.

## Chapter

# 2

## ***Climate Networks***

In this chapter an introduction to complex networks and their applications will be given. A description of special types of networks (small world and scale free) and their properties will be given in section 2.1. In section 2.2, methods for construction of networks from climatic data will be discussed. Finally, previous works which have motivated the work presented in this thesis, are reviewed in section 2.3.

### 2.1 Complex Networks

The first paper in network theory is considered to be of Leonhard Euler which solved in 1741 a famous problem about bridges—see [68] for the original paper in latin, and an historical review [69]. The problem stated that there were seven bridges in the East Prussian city of Königsberg (now Kaliningrad - Russia) that spanned the various sections of the river—as the illustration in Euler’s original paper shows in Fig. 2.1. The problem posed was: “Could a person devise a path through Königsberg so that one could cross each of the seven bridges only once and return home?\*

Long thought to be impossible to do, the first mathematical demonstration of this impossibility was presented by Euler as an example of a class of problems belonging to a Leibniz-called “geometry of position”, defining each of the regions A,B,C, and D of Fig. 2.1 as ‘nodes’ and each of the bridges as ‘links’. The topic was named graph theory and was considered in the beginning an obscure and rather unuseful branch of mathematics. Even Euler stated about the type of mathematics he had to develop in order to give his Königsberg bridges solution: “This branch is concerned only with the determination of position and its properties;

---

\*This is: return to the same part of the river, marked with A,B,C or D in Fig. 2.1

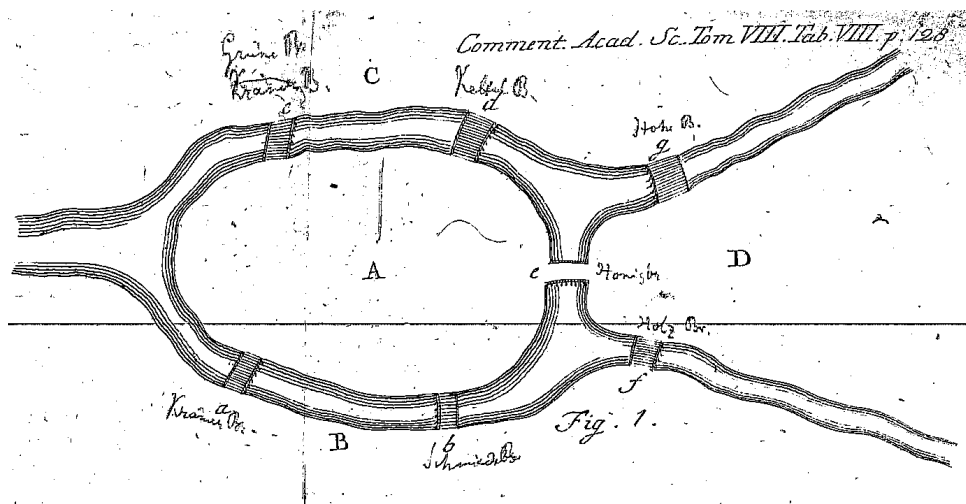


Figure 2.1: Classical Königsberg bridges problem: “Could a person devise a path through Königsberg so that one could cross each of the seven bridges only once and return home?” In 1741 Euler demonstrated that it is impossible, writing in the process the first article in networks. Figure extracted from the original paper of Leonhard Euler [68] which can be downloaded from [70].

it does not involve distances, nor calculations made with them”. However, and especially after the development of fast computers, and the bloom of numerical solutions in the second half of the twentieth century, graph theory has become a well developed topic.

From a mathematical point of view, a network (graph) is a set of vertices (or nodes) connected via edges (or links). Figure 2.2 shows a schematic representation. It is this sheer simplicity which makes it suitable for tackling a wide range of problems in systems composed of many interacting elements. This adaptability to many modeling schemes, have set the basis for the birth of the field of *Complex networks*.

Complex networks are graphs which possess non-trivial topological features, as patterns of connections between their elements that are neither purely regular nor fully random. These patterns do not occur in simple networks such as lattices or random graphs but often occur in graphs that model real systems. Network properties have been intensively studied [71–79] and are reviewed in several books [80, 81]. They include the degree distribution, the clustering coefficient, assortativity among vertices, community and hierarchical structure, among others. Some of these properties will be described in section 2.1.1.

Applications of complex networks are ubiquitous. Social science has been one of the first

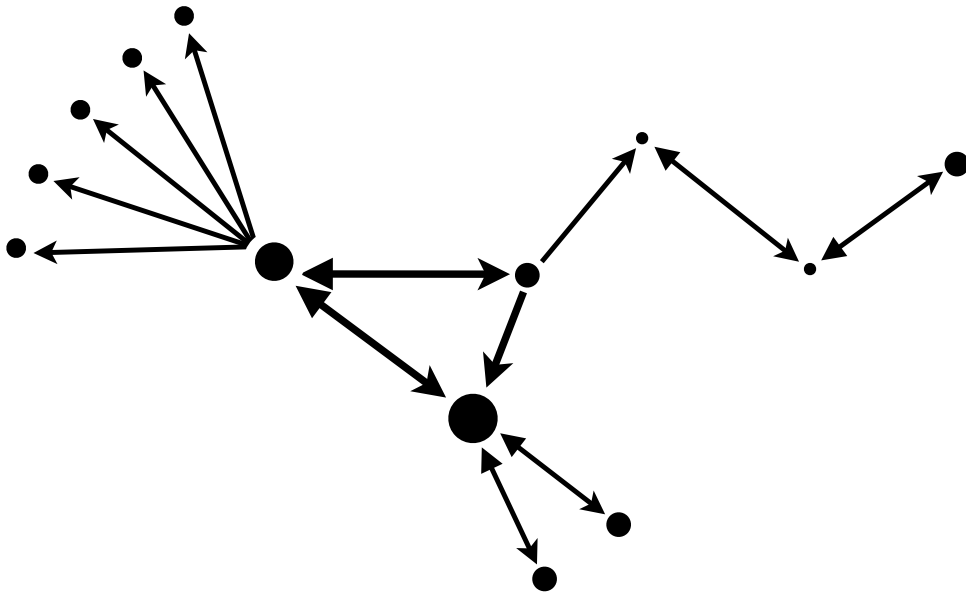


Figure 2.2: Schematic representation of a network, where nodes are indicated as dots and links are shown as lines. The strength of the links is displayed using the line width and the direction of the links with arrows. When a network is used to model a certain physical problem, properties as the number of links of a node or the shortest path between two nodes are magnitudes that can be reinterpreted in terms of the system under study.

research fields that used the complex networks approach [4], and it naturally extended previous works in analytical sociology [82]. Nowadays it is a very active field of study including vote dynamics [83], cultural diversity [84], and behavioral sciences [85, 86], among others.

Some biological problems that are suitable modeled by complex networks include food webs—see earlier [5, 87, 88], and more recent work [89]—where the relationships among organisms in an ecosystem are mapped by a network. A node is defined to be a species and a link, the fact that a species can eat or be eaten by another species. The transference of mass and energy from vegetal life into carnivorous animals and back to unicellular life can be traced and important nodes on which the ecosystem is based are naturally highlighted by the topology of the network.

Gene regulatory networks [6, 90–94] have also become an active research topic in recent years, as the network represents metabolic pathways, gene-gene interactions and self-regulation. In this case the nodes are mRNAs (messenger Ribonucleic acid strains) and proteins that arise from gene expression, while the links are the interactions between them.

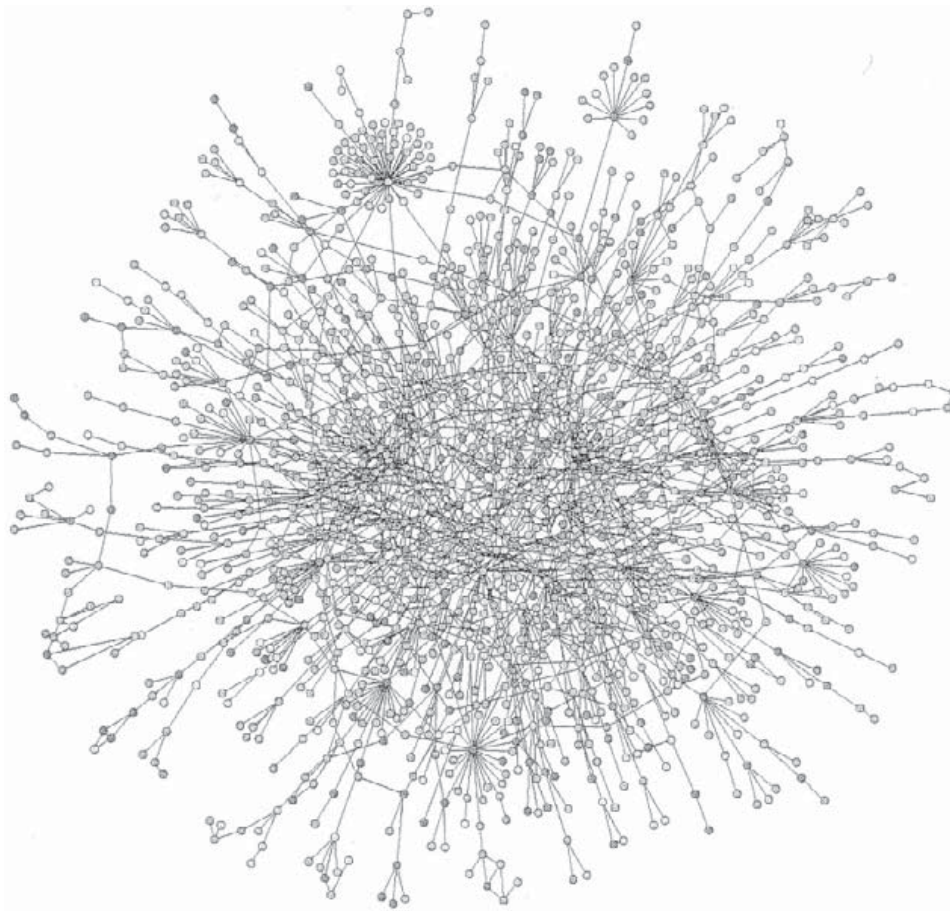


Figure 2.3: The network of interactions between the proteins in *Saccharomyces cerevisiae*, known as baker's yeast. Studying the number of links for each node gives an insight of the importance of each protein the life of the organism [9, 90].

In figure 2.3 the network of interactions between the proteins in *Saccharomyces cerevisiae*, known as baker's yeast, is shown. Notice the complex structure of the topology of the network and the presence of 'hubs' or super connected nodes.

Another branch of biology which has benefited from the networks concept is epidemiology [95–97] where networks are used to model the probability of infection and the pathways of diseases. Nodes are individuals connected by links when an epidemic (or a rumor) spreads through the population. A classical example is the network of human sexual interactions presented in [98] which helped to map the transmission of HIV/AIDS. The network was constructed from one particular person (possibly infected), for whom sexual relations to other persons were mapped by means of questionnaires. Afterwards, the collection of data



was extended to contacts between other persons in the same network. Accordingly, network measures were then used to yield light on the epidemic parameters for better modeling the outbreaks.

Neural networks—computational models inspired by biological networks as the central nervous systems of animals, in particular the brain—are a field by itself, exceeding the limits of biology. They have been studied from a purely theoretic point of view [99], as a source of algorithms for solving problems as voice or written text recognition, time series prediction, or robotics, among many others [100, 101]. They are also used as a tool for a detailed anatomical study of the brain [7, 102, 103]. Nodes are defined as neurons, which are connected physically by axons. The massive parallelism of our brain and the facility humans and other animals can perform many tasks, which are extremely difficult for computers, empowers research in this topic. Indeed many present day algorithms are based in neural networks, e.g., the work presented in [104] where the house numbers taken from Google street view are read and interpreted by a neural network-based algorithm for a better localization of houses in the maps.

Many other applications of complex networks exist, including studies about the internet [8], World Wide Web size and properties [75, 105], traffic (airports [106], or highways [107]) etc. In addition, networks have allowed to do ‘meta-science’, this is science-on-science including scientific citation statistics [108–110]. More complex objects such as networks of networks [111–113] and spatially embedded networks (see [112, 114–116] and chapter 8 of [81]) are areas of intense research. In particular, phenomena as synchronization in complex network topologies has been intensively studied [117, 118].

Climate networks [9, 10, 12, 119, 120] stand as an application of network theory to climate studies. The main idea, as will be explained in section 2.2 is to assume that different geographical regions can also form a network. Nodes are defined from a continuous or gridded field while the links are assigned by using a similarity measure.

### 2.1.1 Properties of complex networks

Real-world networks are complex, and possess many properties different to those of ‘simple’ networks as lattices or randomly connected graphs (as e.g. Erdős-Rényi networks [121]). Some of them have a fat-tailed distribution of links per node (degree distribution), strong

correlations of degrees of connected vertices (highly connected nodes are directly connected among them) and an abundance of loops (paths returning to the original node).

A complete description of a graph of  $N$  vertices is provided by its  $N \times N$  adjacency matrix  $A$ . Each element  $A_{ij}$  can be either 0 or 1, where 1 means a link between nodes  $i$  and  $j$ , and 0 no link. The case of a *directed* network, yields an asymmetrical  $A_{ij} \neq A_{ji}$  adjacency matrix. If the link strength is to be represented, network links can be *weighted*, by setting  $A_{ij} \in \mathfrak{R}$ .

The degree of a node  $v$ ,  $k_v$ , is the number of first neighbors the node has, and can be calculated from  $A$  using

$$k_v = \sum_{i=1}^N A_{vi} \quad (2.1)$$

Vertices with exceptionally high degree are usually referred to as hubs or super-nodes.

Two well-known classes of complex networks are scale-free networks [75], and small-world networks [73]. Scale-free networks were introduced by Barabási and Albert and present a very broad degree distribution. Specifically, if the degree distribution can be fitted by a power-law:

$$P(k) \propto k^{-\gamma}, \quad (2.2)$$

this case it is said to possess a scale-free distribution (as it contains no specific degree scale) and the network is referred to as scale-free. In the case of regular lattices, or random networks, degree distributions are usually narrowly concentrated forming peaks.

The second class of complex networks, was proposed by Watts and Strogatz [4, 71, 73]. They have named them small-world networks (see Fig. 2.4) inspired in the adage “It’s a small world” used when finding e.g. common friends with a stranger in unexpected situations. These networks consist of lattices, where randomly chosen vertices are connected by long-range shortcuts. The proposed mechanism to construct these networks starts from a regular lattice (Fig. 2.4(a)) and rewires iteratively some of the connections with some probability  $p$  (Fig. 2.4(b)) to obtain a small-world network. However, if  $p$  is too large, the small-world effect is lost and the obtained is a random graph (Fig. 2.4(c)).

Indeed, a small-world network can be regarded as a superposition of a lattice and a classical random graph. If one wants to estimate the distance between two chosen nodes in a regular one dimensional lattice connected to first and second neighbors as in Fig. 2.4(d), this will grow linearly as  $\bar{\ell} \propto N$ . This will continue to be so in a small world network if the path

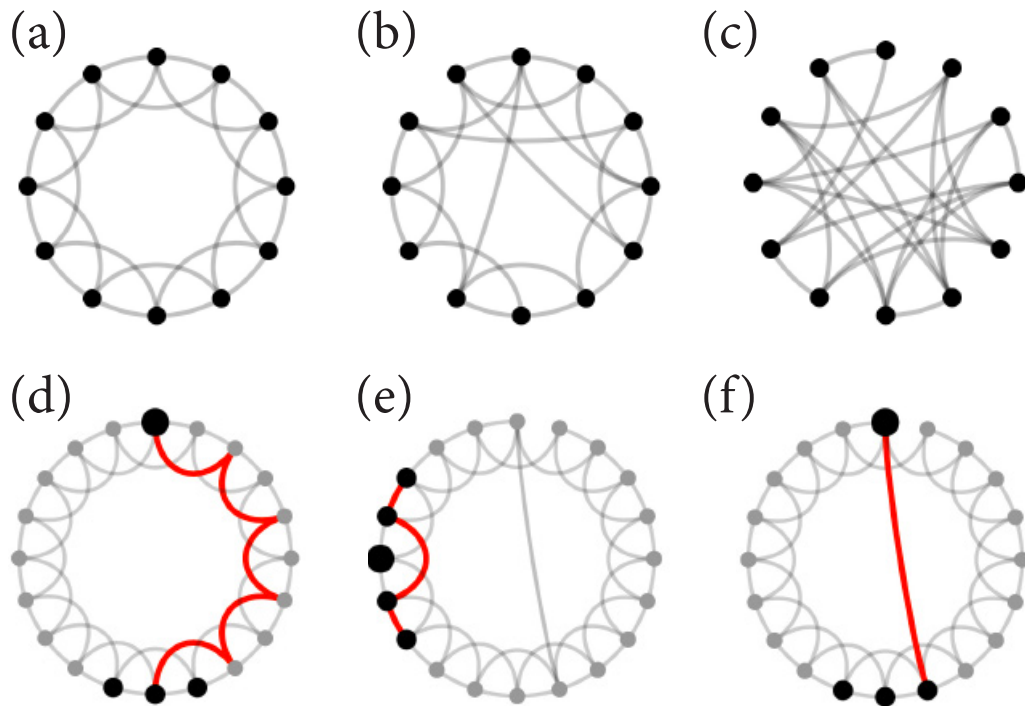


Figure 2.4: An explanation of the ‘small-world’ effect. (a) a regular one-dimensional periodic network with links to first- and second- neighbors. (b) same networks where some links where rewired (small-world network). (c) as the rewiring continues, the network becomes a random network. (d) in a network like (a) there are no shortcuts between two randomly chosen nodes. (e) this is also true in principle for the average links in small-world networks. (f) However some ‘special’ nodes possessing long-range connections act as shortcuts reducing drastically the diameter of the network. Adapted from [4].

does not go through any ‘shortcut’ or long range connection as in Fig. 2.4(e). Finally some special nodes which possess such long range connection act as the focus of connectivity in the network, even if the node possesses a degree similar to the rest of the nodes (Fig. 2.4(f)). It was shown [4] that the mean distance between nodes is proportional to the logarithm of the number of nodes  $N$  in the network, that is:  $\bar{\ell} \propto \log N$ . In [122] it was showed analytically that scale-free networks are ultra-small worlds. In this case, due to the presence of hubs, the shortest paths become significantly smaller and scale as  $\bar{\ell} \propto \log(\log N)$ .

In figure 2.5 the degree centrality for three types of networks is shown. The degree of an Erdős-Rényi Random network 2.5 (a) usually is around a certain value with a disper-

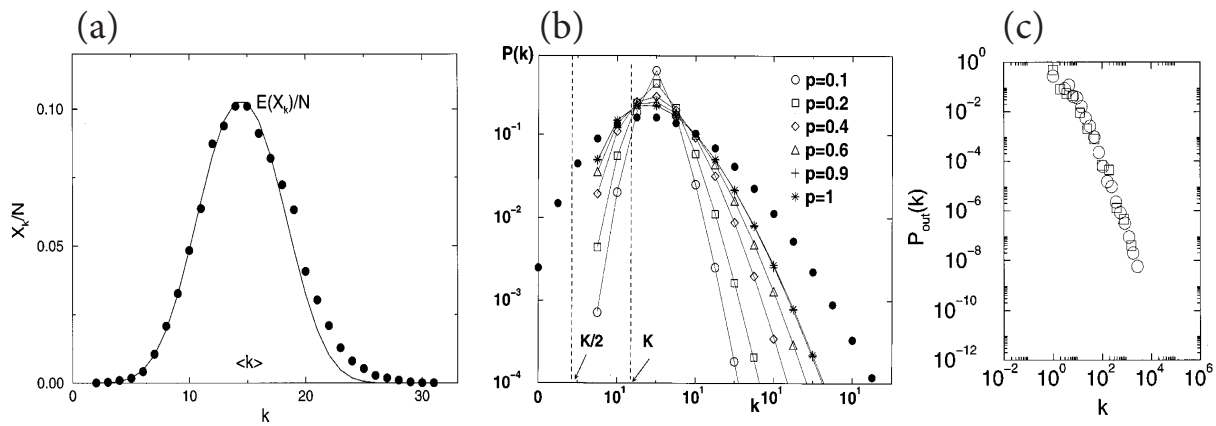


Figure 2.5: Examples of different degree centralities ( $P(k)$ ) (a) a Random network (Erdős-Rényi). Here the number of nodes with probability  $k$  over all the possible nodes  $N$  is shown as  $X_k/N$ . (b) Small world networks for several values of the rewiring probability  $p$ . (c) A scale free network, in this example the number of the in links of the internet as measured in 2001. Adapted from [75].

sion which depends on the number of nodes. A small world network 2.5 (b) shares some of the properties of the underlying lattice—which has a delta-like degree distribution around a fixed value of neighbors—giving to this degree distribution a maximum value. Finally scale free networks (c) possess no typical degree. The distribution is almost linear and show how a small number of nodes possess an enormous quantity of links while the other nodes possess less and less in a seamless manner.

## 2.2 Construction and representation of climate networks

Climate networks allow to investigate the connectivity between different areas over the surface of the Earth. This connectivity is based on the similar behavior of time series; hence it is to be associated to modes of variability or patterns that occur often enough to be recognizable. Such kind of spatial structures occur in the atmosphere on all time scales, examples of which are the North Atlantic Oscillation and the teleconnection patterns associated with El Niño-Southern Oscillation. Network connectivity is usually based on (linear and nonlinear) correlations as the climate network approach intends to embed the properties of climate in the topology of the network.

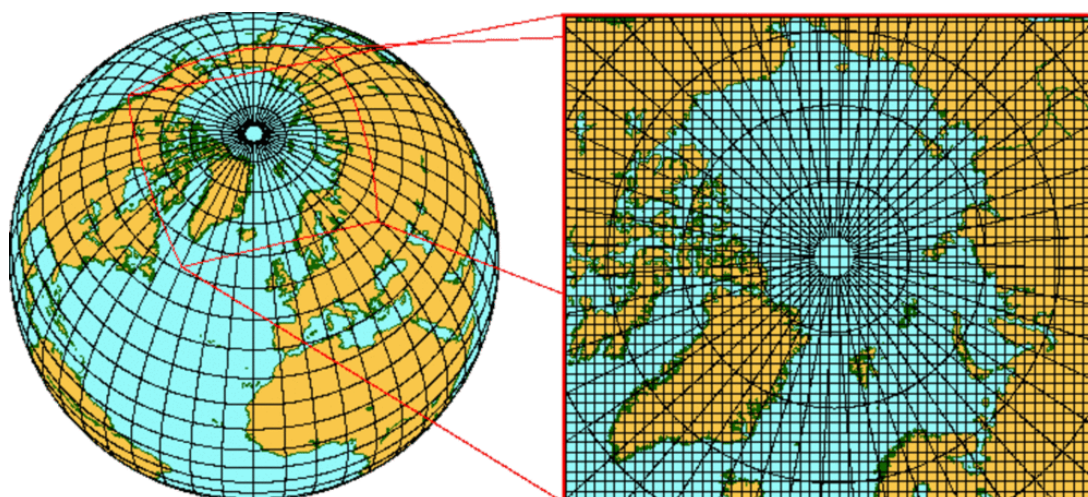


Figure 2.6: Map showing the typical grid of climate data, by latitude and longitude. In the left panel, it can be seen how the density of points increases as near the poles. In order to overcome spurious degree distributions, the area weighted connectivity—Eq. (2.3)—is used. Figure extracted from [123].

The climate network properties will depend on the methodology employed to infer the presence of connections between two nodes, i.e., the similarity measure used to include a particular link in the network and the procedure to filter out those correlations that may have occurred merely by chance. The similarity measures used and the statistical significance tests performed to validate the connections are explained in Chapter 3.

The nodes of the network are defined by the data set, which is distributed in a regular grid (Fig. 2.6 left). This way, links between any pair of nodes are defined depending on their climate interdependency. As the earth is approximately a sphere, a latitude-longitude based nodes scheme will necessary yield an inhomogeneous distribution of nodes in physical space, especially near the poles (Fig. 2.6 right). This effect will be addressed in section 2.2.3.

### 2.2.1 Data set

The variable chosen to construct climate networks during this thesis is the near surface air temperature or SAT (see section 1.1.1 for an explanation of the variable).

SAT data is obtained from a data assimilation system (DAS). A DAS is a general circulation model (GCM) employed to fill in the gaps—in space and in time—in real observed

data. The gaps exist because of missing data or bad data quality. The model must blend seamlessly with the real data and hence being the most accurate representation of past climatic variables. Because of this, the data assimilation system uses the observed data to make the output from the model consistent with that data, which ensures that the model accurately produces environmental variables. Thus, most output variables from a weather DAS are based on data acquired by remote sensing or ground station measurements. Also because of this processing, these products usually are distributed as gridded data, the kind of data used during this work. The DAS used for this work is called “NCEP/NCAR Reanalysis I: 1948-present” from the National Oceanic and Atmospheric Administration (NOAA) of the United States. It uses a global, high resolution, coupled atmosphere-ocean-land surface-sea ice system model to provide the best estimate of the state of these coupled domains over the period [124]. The climate model was initialized with a wide variety of weather observations: ships, planes, station data, satellite observations and many more. The dataset is kept current using near real-time observations and updated daily.

SAT can also come from the output of atmospheric models alone. In this case an atmospheric general circulation model (AGCM) is forced with historical sea surface temperatures and the simulated SAT depends only on the skill of the model to reproduce the behavior of the real atmosphere. Even though all models have biases, the use of an AGCM allows the design of numerical experiments to isolate a particular phenomenon. Chapter 5 of this Thesis uses the output of an ensemble of AGCM runs to separate the variability associated with internal dynamics and the one forced from the boundary.

Another important definition is that of temperature anomaly, which is calculated as the departure from the climatology. A positive anomaly indicates that the observed temperature was warmer than the long term mean, while a negative anomaly indicates that the observed temperature was colder.

Given that the seasonal cycle is strong in many regions of the world, the use of the anomaly field allows to study more easily that part of the atmospheric variability not directly related to astronomical forcing.

### 2.2.2 Network construction

In order to construct a Climate network, let's suppose there is a correlation measure and a method to assess the significance of the links. A correlation matrix  $C_{ij}$  as the one shown

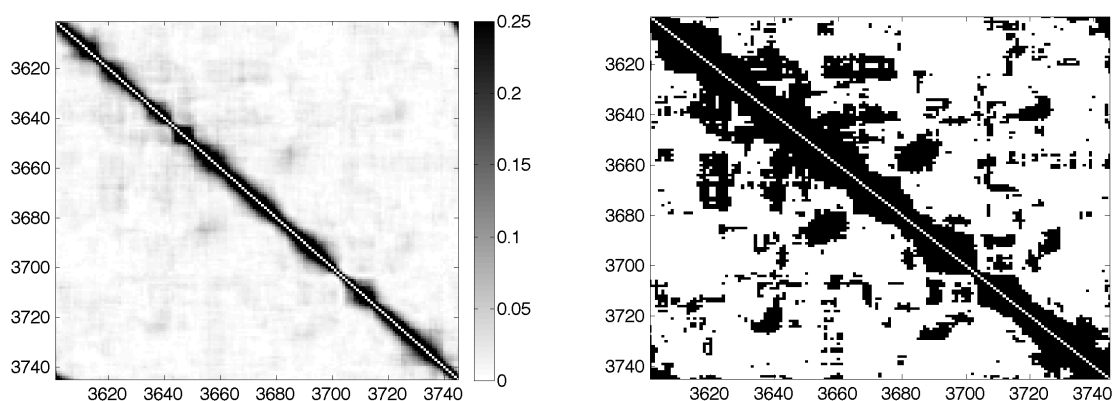


Figure 2.7: Correlation matrix of a climate network (detail). (left) A part—over a single latitude—of the correlation matrix of a climate network constructed during this thesis (see Chapter 4). Notice that the correlation of each link with itself has been set to zero. (right) The adjacency matrix after applying a threshold. Links are shown in black while no links are represented in white.

in figure 2.7 (left) can be calculated. This panel—for simplicity—shows the structure for a single latitude (the equator) and thus only longitudinal correlations are shown. Notice that the diagonal elements are zero as the correlation of a time series with itself was not taken into account. After an appropriate significance test, an adjacency matrix  $A_{ij}$  is created—shown in 2.7 (right)—where the connectivity of the network is summarized. Notice the abundance of short range links with respect to longer range links (which reflect teleconnections).

Because climate networks are embedded in physical space—as nodes have precise geographic coordinates—they are partially lattices and display the already mentioned ‘Small-World’ Effect [119, 125] as they are usually very connected to geographically close nodes and sporadically connected to far away areas. This is an efficient topology for information spreading [97] compared with lattices or complete random networks, thus allowing information spread in several time scales as some of the information will propagate through the lattice-like structure, and other through the shortcuts. This effect can be used to model the different time scales that are observed in climate .

These networks, however, are not ‘scale-free’ in the Barabási-Albert sense—as shown in [119]—as they usually do not display potential law in the degree distribution [126] but show

traces of the underlying spatial embedding (latitude-longitude grid) resulting in link degrees around a maximum as shown in figure 2.8.

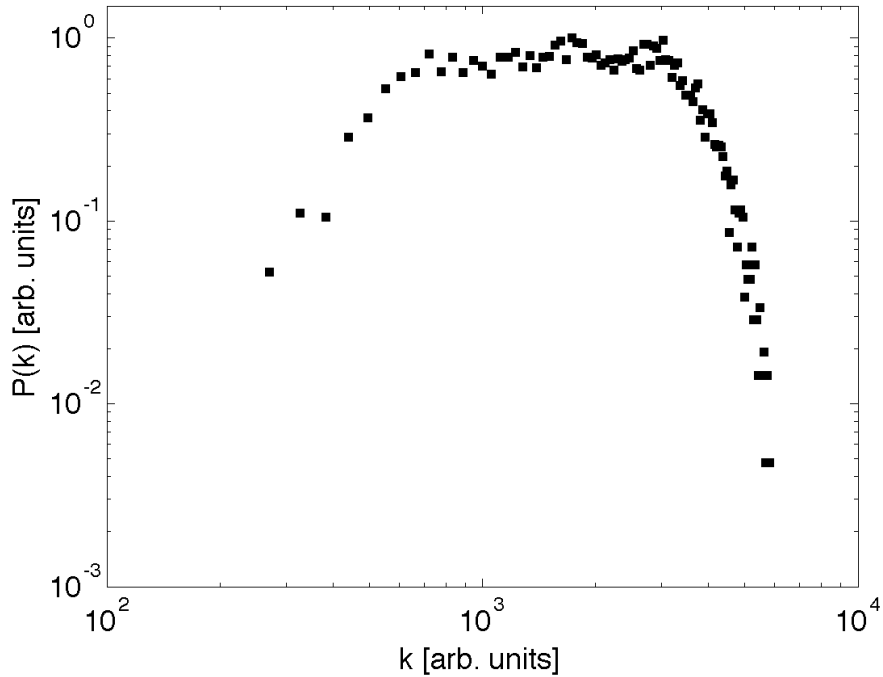


Figure 2.8: Degree distribution of the climate network shown in fig 2.7, it can be seen that it is not a power law distribution.

### 2.2.3 Area Weighted Connectivity

As stated above, a latitude-longitude node based scheme yields an inhomogeneous distribution of distances between nodes. Thus, correlations near the poles could be higher in part because nodes are physically closer, altering the results. In order to overcome this, when representing the connectivity of the network on a map, the degree and connections among nodes should be weighted by the *areas* the nodes represent instead of the nodes themselves.

A commonly used measure to represent the network is the *area-weighted connectivity* (AWC) [120, 125, 127–129] and it is given by:

$$AWC_i = \frac{\sum_j^N A_{ij} \cos(\lambda_j)}{\sum_j^N \cos(\lambda_j)}, \quad (2.3)$$

where  $\lambda_i$  is the latitude of node  $i$  and  $A_{ij}$  is the adjacency matrix.



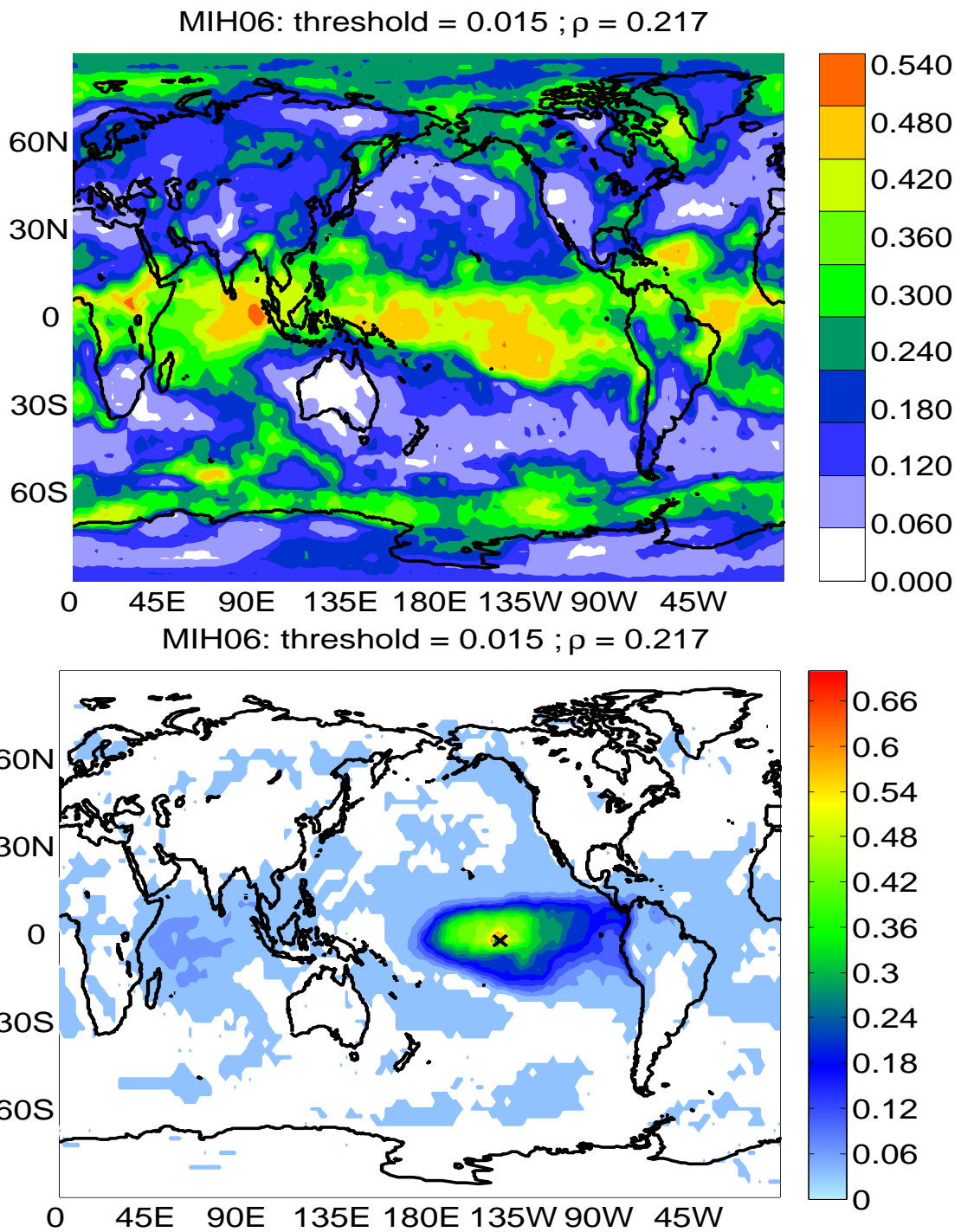


Figure 2.9: (top) Map of AWC from surface air temperature, showing the degree distribution *per area* on a world map. Notice most of the connectivity is in the tropics. The high connectivity near the poles has been related with the presence of sea-ice and will be further explained in chapter 4. (bottom) Connections to a point in the Pacific ocean (marked with an  $\times$ ). A threshold has been applied. Figures adapted from [130].

The AWC plots the number of links every node has, taking into account that the nodes represent geographic regions with different areas. AWC plots provide information about the portion of the Earth to which a node is connected, but do not indicate to which regions this node is connected. In figure 2.9 (top) a map of AWC is shown for the same network of figures 2.7 and 2.8. Figure 2.9 (bottom) represents the connections of a single node (referred as  $\times$  and located in the central Pacific ocean) and the color scale shows the strength of these connections having only discarded those under the significance threshold.

## 2.3 Climate phenomena and climate networks

In the last years, the network approach to climate studies has received a large interest from the community. In this section an overview of the pioneer works and recent articles relevant to the work presented in this thesis will be made.

The field was first developed with a series of papers from several groups which proposed the network theory as means of analysis of climate variables [9–12, 114, 119, 120, 131–135]. Since then, various methods for constructing climate networks have been proposed (computing correlation measures [9, 119] from temperature or pressure fields, analyzing global climate [12, 120] or particular geographic locations [132, 136–140]).

A. Tsonis *et al* [119] in 2004 considered climate as a network of many dynamical systems and studied the coupling architecture of this network as the interaction of two interweaved subnetworks. One subnetwork was in the tropics and the other at higher latitudes with the equatorial one establishing links between the two hemispheres. Both subsystems displayed ‘small-world’ properties, but there were differences between each other. The tropical one was an almost fully connected network, whereas the mid-latitude behaved more like a scale-free network characterized by dominant super nodes.

Two papers appearing in the same issue of “Physical Review Letters” (one by Yamasaki *et al* and the other by A. Tsonis *et al*) [10, 11] pointed in the same direction, constructing a climate network from temperature data for “El Niño”, “La Niña” and neutral years, and showed that during the Niño years many links were broken in comparison to the other two cases. They further suggested that the number of surviving links could be used as a measure for gauging “El Niño” events, and the stability of the system in the various ENSO phases.

In [141] the study of [10] was extended to examine climate networks constructed from other daily data. Daily temperature or geopotential height data usually present high vari-

ability and are hard to predict, but the authors found long lasting links between the nodes yielding to a robust network pattern. Part of this robustness was attributed to the geographical regular embedding of the network; however, physically robust coupling between different locations was also found, with the coupling on the equator being significantly less pronounced than in the extratropics.

A systematic comparison of networks measures and construction on local, mesoscopic and global topological scales was done by J.F. Donges *et al* in [120]. Climate networks were constructed from the same global climatological data set using the linear Pearson correlation coefficient and the mutual information as measures of dynamical similarity between regions. A high degree of similarity was observed on the local and mesoscopic topological scales for surface air temperature fields. This was called the backbone of the climate network.

Multivariate approaches have also been used by comparing several climatic variables or the same variable at different levels. For example in [127] a multilayer network of geopotential height (GH) at different heights was studied. A measure called “cross-betweenness” was defined in order to help to identify regions which are particularly important for mediating vertical wind field interactions as interaction between the GH networks. K. Steinhaeuser *et al* in [142, 143] studied several variables as sea surface temperature (SST), sea level pressure (SLP), and GH, among others. Community maps of one variable and between the variables were obtained. The euclidean distance in this projected space was used as a similarity measure in order to build a network, finding time stability of some communities. Guez *et. al* in [144] studied the structure of climate networks in relation to NAO. A network covering the North Atlantic for wintertime data only was constructed from both temperature and geopotential height. They found that within the different phases of the NAO the correlation values of the links in the climate network was significantly different. They proposed this as a measure to track the NAO pattern.

CNs are currently researched for a wide variety of uses, from characterization of specific regions to prediction of extreme events. The topology of extreme precipitation networks during the pre-monsoon, Indian Summer Monsoon, and post-monsoon seasons have been studied by V. Stolbova *et. al.* [139]. The network of the Indian Summer Monsoon has three essential spatial domains defined by topography: North Pakistan (NP), Eastern Ghats (EG), and the Tibetan Plateau (TP). Precipitation links form in the pre-monsoon season, and dis-

appear during the post-monsoon season. The large number of connections and the long average link length of the NP, TP and EG regions during the ISM season imply that these regions strongly affect extreme rainfall event synchronization all over the Indian subcontinent. Also, the average and maximal link lengths of these regions are significantly increased in comparison to the pre-monsoon and post-monsoon periods.

The reliability and robustness of the networks have also been analyzed in terms of a critical comparison of the networks found with the various methods used [145–148]. A main conclusion of these studies is that it is crucial to analyze the robustness of the method used to quantify climate similarities because trends and serial correlations in the time series, as well as time lags, can significantly affect the topology of the network obtained.

Other papers applied the network approach to climate indices. A. Tsonis *et. al.* in [131] used PDO, NAO, ENSO etc, as a path for investigating climate shifts. The distance between series and their synchronization was measured. Afterwards, using symbolic dynamics they made predictions for the synchronization and desynchronization of the network on certain year through a linear transformation. In [136] interdependence between ENSO and the Indian monsoon was analyzed with the use of Granger causality estimation from data for the period 1871-2006. A similar approach has been used by Tirabassi *et. al.* [149] to study the possibility of disentangling the air-sea interaction in the region of the South Atlantic Convergence Zone (SACZ) also using Granger Causality as a measure of the directional coupling between the ocean and atmosphere.

A. Tantet *et al* [150] studied the interannual to multidecadal timescales variability patterns in sea surface temperature through the community structure of interaction networks constructed from SST. The community structure was interpreted using known dominant patterns of variability, such as the El Niño/Southern Oscillation and the Atlantic Multidecadal Oscillation (AMO). The study of the relationship between the communities and indices of global surface temperature showed that, while ENSO was most dominant on interannual timescales, the Indian West Pacific and North Atlantic may also play a key role on decadal timescales.

Another use of Climate networks has been as predictors of regime change. Van der Mheen *et al* [151] proposed early warning indicators of the collapse of the Atlantic Meridional Overturning Circulation (MOC), a part of the large-scale ocean circulation that depends on global

density gradients created by surface heat and freshwater fluxes . These new indicators are based on a climate network of spatial correlations in the time series of the Atlantic ocean temperature field. A meridional-depth model of the MOC is used for which the critical conditions for collapse can be explicitly computed. This network is used to monitor changes in spatial correlations in the model temperature time series as the critical transition is approached. The indicators were based on changes in topological properties of the network, in particular changes in the distribution functions of the degree and the clustering coefficient. The expectation value of the normalized degree distribution of the network increased steeply and smoothly when the tipping point approached.

M. Barreiro *et. al* [128] analyzed climate networks using techniques of nonlinear time series symbolic analysis. Specifically ordinal patterns and binary representations was used on monthly averaged surface air temperature anomalies. The symbolic analysis results were able to separate the time variability of SAT anomalies in patterns of oscillatory behavior related to intraseasonal variations—especially in the extra-tropics—and to El Niño on seasonal to interannual time scales. By mapping these processes into global networks, using ordinal patterns and binary representations, the structure of the network was found to change drastically at different time scales. This study has set the basis for the work presented in this Thesis.



## Chapter

# 3

## ***Climate Time Series Analysis***

This chapter presents the similarity measures and the significance criterion employed for constructing the climate networks. Section 3.1 shows linear analysis techniques, including Pearson cross correlation and linear regression (section 3.1.1), and Empirical Orthogonal Functions (EOF) also known as principal components analysis (section 3.1.2). Section 3.2 is devoted to *non-linear* information theoretic-based techniques. Firstly entropy is defined (section 3.2.1) as a measure of the information content—or the lack of it—in a time series. Afterwards mutual information is defined (section 3.2.2) as measure of similarity between two time series, which takes in account the information *shared* by the two time series, and finally a directionality index based on the conditional mutual information is introduced (section 3.2.3) which can quantify the rate of information transfer among the time series. Section 3.3 deals with symbolic ordinal analysis using ordinal patterns. This method is used during this thesis to investigate the connectivity of the climate networks, on several time scales. Finally section 3.4 is devoted to a study of statistical significance tests, especially through the use of surrogate data.

### 3.1 Linear techniques

#### 3.1.1 Cross Correlation

The most familiar measure of interdependency between two time series is the Pearson correlation coefficient [152]. It is defined as:

$$\rho_{X,Y} = \text{corr}(X, Y) = \frac{\text{cov}(X, Y)}{\sigma_X \sigma_Y} = \frac{E[(X - \mu_X)(Y - \mu_Y)]}{\sigma_X \sigma_Y} \quad (3.1)$$

where  $X$  and  $Y$  are two random variables with expected values  $\mu_X$  and  $\mu_Y$  and standard deviations  $\sigma_X$  and  $\sigma_Y$  and  $E$  is the expected value. For the discrete case (typical for time series analysis) it is defined as:

$$E[(X - \mu_X)(Y - \mu_Y)] = \frac{1}{(n-1)} \sum_{i=1}^n (x_i - \mu_x)(y_i - \mu_y) \quad (3.2)$$

where the  $x_i$  and  $y_i$  are  $X$  and  $Y$  written as time series.

Karl Pearson's  $\rho$  was the first formal correlation measure, and it is still the most widely used measure of relationship.

From the definition shown in Eq. (3.1) the Pearson correlation cannot exceed 1 in absolute value and it is defined only if both of the standard deviations are finite and nonzero. The correlation coefficient is also symmetric:

$$\text{corr}(X, Y) = \text{corr}(Y, X) \quad (3.3)$$

It is possible to lag one of the time series a time  $j$  in order to calculate the *lagged* response defined as:

$$\rho_{X, Y_j} = \frac{1}{(n-1)\sigma_x\sigma_{y_j}} \sum_{i=1}^n (x_i - \mu_x)(y_{i+j} - \mu_{y_j}) \quad (3.4)$$

If the variables are independent, Pearson's correlation coefficient is 0, but the converse is not true because the correlation coefficient detects only linear dependencies between two variables. In Fig. 3.1 several cases of the correlation of two variables are shown. In the special case (top row) when  $X$  and  $Y$  are jointly normal, Pearson correlation is an exact measure of interdependency and zero-correlation is equivalent to independence. Notice in the middle row of Fig. 3.1 that when the variables are linearly related with minimal noise,  $\rho$  reflects the sign of the slope only, and for a completely horizontal line the correlation is not defined as the variance of  $Y$  is zero. Many shapes which possess an obvious (but nonlinear) relationship between the variables yield to a zero correlation providing the slope of a linear regression on them is zero, as shown in the bottom row.

### Removing linear correlations between time series

A handy technique to analyze—to the first order—the effects of one time series over another is to perform a linear regression at zero lag. In Fig. 3.2, a SAT time series  $Y$  over the Pacific



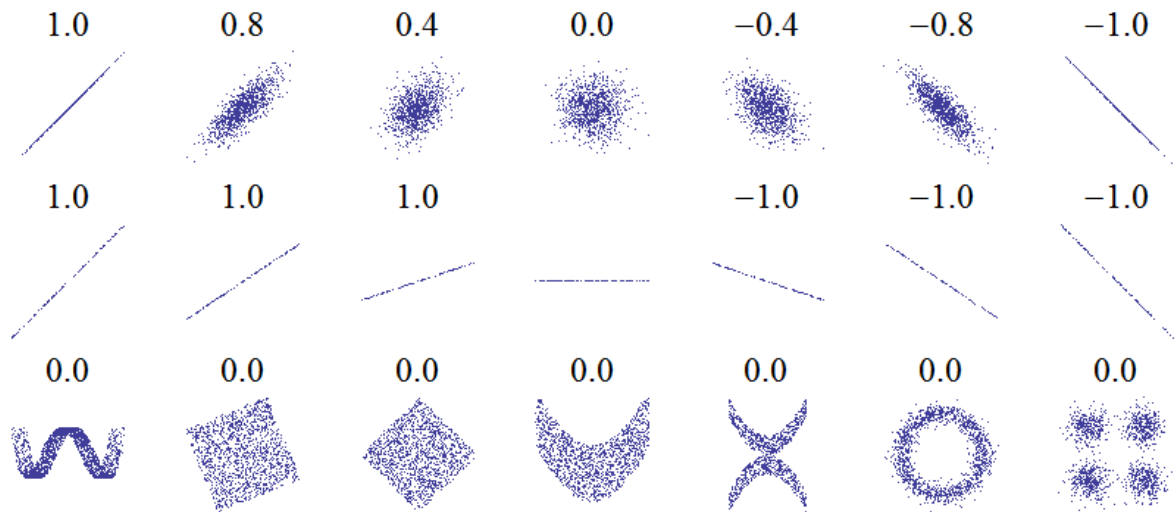


Figure 3.1: Several sets of  $(X, Y)$  points, with the Pearson correlation coefficient for each set. Note that the correlation reflects the noisiness and direction of a linear relationship (top row), but not the slope of that relationship when the variables are linearly correlated (middle)—the figure in the center has an undefined value of  $\rho$  because the variance of  $Y$  is zero—it also does not reflect many aspects of nonlinear relationships providing the linear regression of those relationships give zero slope (bottom). Figure adapted from [153].

ocean is correlated to the SST-based NINO3,4 index. A straight line  $\tilde{Y}$  can be calculated via a simple linear regression (shown in red). Eliminating its slope by subtracting the value of the regressed line from each of the points:

$$R = Y - \tilde{Y} \quad (3.5)$$

effectively eliminates the linear correlation of one series over the other. Nonlinear effects may—and, they usually do—subsist, however, in many practical cases, removing the linear effects is first step for a more detailed non-linear analysis.

This simple approach has been proved extremely useful in climate dynamics to distinguish linear from nonlinear influences [51, 63]. It can be however improved using nonlinear methods for calculating the relationship (see e.g. [155]), or lagged regressions—Eq. (3.4)—could be considered.

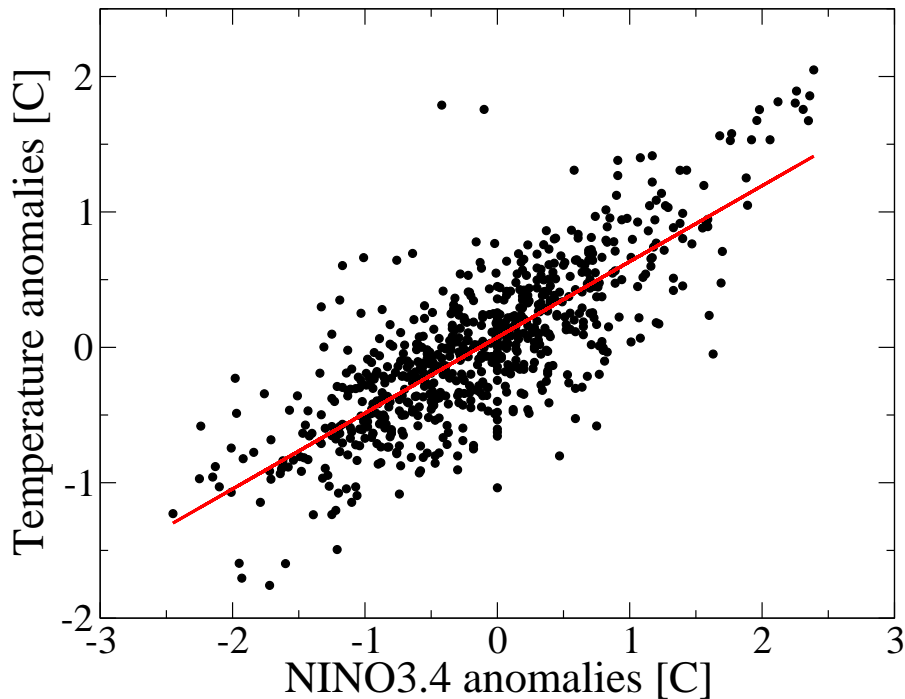


Figure 3.2: Graphical representation of the linear removal procedure. One time series (in this case SAT anomalies) are compared at zero lag with another time series (in this case the NINO3.4 index) and a linear regression is performed (in red). If the slope of this regression is subtracted from the data, as shown in the bottom line of Fig. 3.1 the linear correlation will be zero. Figure taken from [154].

### 3.1.2 Empirical Orthogonal Functions

The technique of Empirical Orthogonal Functions (EOF), also known as Principal component (PCA) analysis, is a linear multivariate analysis tool [152, 156]. It creates a new set of orthogonal variables that contains the same information as the original set, where the axes of variation are rotated while keeping orthogonality, and ordered so that they summarize decreasing proportions of the variance. It is used in climate studies to identify spatial patterns which—by themselves—can explain a sizable amount of the total variance of the dataset, as the ones mentioned in chapter 1. During this research the EOF was constantly used as a complementary linear technique for testing the data, for noise filtering, and for defining most patterns and teleconnections in the usual manner. This by itself motivates the inclusion of this technique in this chapter. Furthermore the Results this technique is explicitly

used for calculating the NAO indices in chapter 5.4.2, which is a standard procedure in climate science.

Let a matrix of time-series  $\mathbf{X}_t^i$ , from model or observational gridded data recorded in  $N$  different geographical locations, not necessarily regular. If the  $\mathbf{X}^i$  are normalized to zero mean the covariance matrix will result

$$\mathbf{C} = \mathbf{X}^T \mathbf{X}; \quad (3.6)$$

$\mathbf{C}$  can be diagonalized through the transformation

$$\mathbf{A}\mathbf{C} = \mathbf{C}\mathbf{D}, \quad (3.7)$$

where  $\mathbf{A}$  is the matrix of eigenvectors and  $\mathbf{D}$  is the diagonal matrix containing the covariance eigenvalues. Since  $\mathbf{C}$  is symmetric, the eigenvectors form an orthogonal basis on which the data matrix  $\mathbf{X}$  can be projected:

$$X_t^i = \sum_{j=1}^N P_t^j A_{ji}, \quad (3.8)$$

here, the  $\vec{P}_j$  are called the principal components (PCs) and the  $\vec{A}_j$  are the EOFs of the field  $\mathbf{X}$ . The EOFs have the same dimension of the spatial dimension of the database, and they can be interpreted as spatial patterns and plot as maps. In the other hand, from Eq. (3.8) the PCs characterize the time evolution of the maps represented by the EOFs.

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

$$\lambda_j = \frac{D_{jj}}{\text{Tr}(\mathbf{D})}, \quad (3.9)$$

represents the variance fraction explained by each empirical mode of variability  $\vec{P}_j * \vec{A}_j$ . The first mode is the only one not constrained by orthogonality—the second mode has to be orthogonal to the first, the third mode to the two before, and so forth. Non-orthogonal EOFs can be derived through opportune algorithms as the already classical varimax rotation algorithm [157] among others, and are widely used in climate studies [158]. As EOF is a linear technique, the original field  $\mathbf{X}_t^i$  can be recovered summing back all the terms in Eq. (3.8).

## 3.2 Information theoretic tools

### 3.2.1 Entropy

In the time series analysis context, “entropy” usually refers to Shannon’s definition [159]. This is generally defined in terms of the probability density function (PDF)  $p(x)$ , of the sys-

tem to be in state  $x$  out of a possible set of states  $\mathcal{A}$ :

$$H = - \int_{x \in \mathcal{A}} p(x) \log_b p(x) dx, \quad (3.10)$$

or—more appropriate for the analysis of time series—in the discrete form [160]:

$$H = - \sum_{i \in \mathcal{A}} p_i \log_b p_i. \quad (3.11)$$

Here,  $p_i$  is the probability of the value number  $i$  to appear in a sequence of characters of a given time series. The base  $b$  with respect to which the logarithm is taken, determines the units in which  $H$  is measured: If  $b = 2$ ,  $H$  is measured in “bits”; for  $b = e$  in “nats” and for  $b = 10$ , in “bans”<sup>\*</sup>.

To understand the notion of entropy in the information theory context, a classical example is to consider a coin toss. When the coin is fair, that is, when the probability of heads is the same as the probability of tails, the entropy of the coin toss is also a maximum. It is impossible to predict the outcome of the coin toss ahead of time—the best prediction will be correct with probability  $1/2$ . Such a coin toss (when using  $b = 2$ ) has one bit of entropy since there are two possible outcomes that occur with equal probability. From Eq. (3.11), learning the actual outcome contains one bit of information as

$$H = - \sum_{i \in (0,1)} p_i \log_2 p_i = -2 \times \frac{1}{2} \log_2 \left(\frac{1}{2}\right) = 1 \quad (3.12)$$

Contrarily, if the coin was not fair, lets say  $p_1 = \frac{1}{4}$  and  $p_2 = \frac{3}{4}$ , and if it was possible to know the probability of the outcomes of the coin tossing event beforehand, it could be possible to do a better prediction, and thus, the entropy would be lower:

$$H = - \sum_{i \in (0,1)} p_i \log_2 p_i = -\frac{1}{4} \log_2 \left(\frac{1}{4}\right) - \frac{3}{4} \log_2 \left(\frac{3}{4}\right) \approx 0.81. \quad (3.13)$$

Now lets assume one has a sequence of events e.g. many sports matches, many coin tosses, or more generally, a deterministic, chaotic or stochastic flow, this is, any sequence of numbers that can be written as a time series. Entropy will act as a measure of the novelty of the information of the time series and, therefore, of its unpredictability.

---

<sup>\*</sup> States with zero probability of occurrence are ignored, as the sum in Eq. (3.11) would be undefined.

### 3.2.2 Mutual Information

The mutual information (MI) is a measure of the mutual dependence of two variables [161, 162]. Contrarily to the correlation coefficient, it is a more general measure of the information—in the Shannon sense—shared by the variables.

Mutual information is computed from the probability density functions (PDFs) that characterize two time series in two nodes,  $p_i$  and  $p_j$ , as well as their joint probability function,  $p_{ij}$  [162–164]:

$$M_{ij} = \sum_{m,n} p_{ij}(m,n) \log \frac{p_{ij}(m,n)}{p_i(m)p_j(n)}. \quad (3.14)$$

$M_{ij}$  is a symmetric measure

$$M_{ij} = M_{ji} \quad (3.15)$$

of the degree of statistical interdependence for the time series  $i(t)$  and  $j(t)$ ; if they are independent:

$$p_{ij}(m,n) = p_i(m)p_j(n) \quad (3.16)$$

and thus  $M_{ij} = 0$ . As with the entropy the most common unit of measurement of mutual information is the bit but it can also be measured in nats and bans.

In this thesis, the PDFs  $p_i$ ,  $p_j$  and  $p_{ij}$  are computed in two ways: by histograms of the original values (this case will be referred to as MIH) and by using a symbolic transformation, in terms of probabilities of ordinal patterns [128, 165–167] (this case will be referred to as MIOP and will be explained in section 3.3).

### 3.2.3 Directionality measure

Both Pearson correlation coefficient and MI give a measure of statistical similarity but not of causality. Generally, one would like to know the *net* effect of one series over the other, and in the case of a net driving, its direction. This can be done for regular and noise free time series, using classical methods based in the phase differences between the series [168]. These results have been extended for the more general case of series embedded in small noise [169–172], but usually these techniques fail for series where the power spectrum is not sharp and phases cannot be unambiguously defined. Using information theoretic techniques allows

us to tackle this problems [173–175], having obtained already successful results in EEG tests and other fields [173].

Time series  $i(t)$  and  $j(t)$  are characterized by probability distribution functions  $p_i$ ,  $p_j$ , and by their joint PDF,  $p_{ij}$ . The directionality between them can be assessed via the *directionality index* (DI) as defined in [176, 177]:

$$DI_{ij}(\tau) = \frac{TE_{ij}(\tau) - TE_{ji}(\tau)}{TE_{ij}(\tau) + TE_{ji}(\tau)}, \quad (3.17)$$

where  $TE_{ij}(\tau)$  and  $TE_{ji}(\tau)$  are referred to as *transfer entropy* [178, 179] and defined as:

$$\begin{aligned} TE_{ij}(\tau) &\equiv M(i; j|i_\tau) \\ &= H(i|i_\tau) + H(j|i_\tau) - H(i, j|i_\tau) \end{aligned} \quad (3.18)$$

$$\begin{aligned} TE_{ji}(\tau) &\equiv M(j; i|j_\tau) \\ &= H(j|j_\tau) + H(i|j_\tau) - H(j, i|j_\tau); \end{aligned} \quad (3.19)$$

here,  $i_\tau = i(t - \tau)$ ,  $j_\tau = j(t - \tau)$ —the nature of  $\tau > 0$  will be explained below—and  $H(i|j)$  is the conditional entropy [176, 177]

$$H(i|j) = \sum_{m,n} p_{ij}(m, n) \log \frac{p_i(m)}{p_{ij}(m, n)}. \quad (3.20)$$

Transfer entropy is a particular case of the *conditional mutual information* (CMI) [168, 178] which is defined as:

$$M(i; j|k) = H(i, k) + H(j, k) - H(i, j, k) - H(k) = H(i|k) - H(i|j, k) \quad (3.21)$$

and can be rewritten as:

$$M(i; j|k) = \sum_{m,n,l} p_{ijk}(m, n, l) \log \frac{p_k(l) p_{ijk}(m, n, l)}{p_{ik}(m, l) p_{jk}(n, l)}; \quad (3.22)$$

it indicates the amount of information shared between  $i(t)$  and  $j(t)$ , given the effect of the series  $k(t)$  over  $j(t)$ .

If  $k(t)$  is replaced by the past of the time series  $i(t)$  to account for the information transfer time  $\tau$ , the transfer entropy case is obtained. This yields  $TE_{ij}(\tau) \equiv M(i; j|i_\tau)$ , Eq. (3.19),

which quantifies the amount of information shared between  $i(t)$  and  $j(t)$ , given the influence of  $i(t - \tau)$  over  $j(t)$ .

Analogously, to assess the information transfer from  $j$  to  $i$ ,  $k(t)$  is replaced by the past of  $j(t)$ , Eq. (3.19). This way, the directionality index,  $DI_{ij}$ , is able to quantify the *net* information flow.

From the definition of  $DI_{ij}$ , Eq.(3.17), it is clear that if  $DI_{ij}(\tau)$  is positive, it will mean a net directionality from  $i$  to  $j$  as  $TE_{ij}(\tau) > TE_{ji}(\tau)$ . This will be represented in red on the maps on chapter 6. Conversely, if  $TE_{ij}(\tau) < TE_{ji}(\tau)$  then  $DI_{ij}(\tau) < 0$  will mean a net direction of information flux from  $j$  to  $i$  and the maps of chapter 6 will show it in blue. Also from the definition:

$$DI_{ij} = -DI_{ji} \quad (3.23)$$

and also,

$$-1 \leq DI_{ij} \leq 1. \quad (3.24)$$

Another properties are:

$$DI_{ij} = 1 \text{ if and only if } TE_{ij} \neq 0, TE_{ji} = 0 \quad (3.25)$$

(i.e., the information flow is  $i \rightarrow j$  and there is no back coupling  $j \rightarrow i$ ), and correspondingly:

$$DI_{ij} = -1 \text{ if and only if } TE_{ji} \neq 0, TE_{ij} = 0 \quad (3.26)$$

(i.e., the information flow is  $j \rightarrow i$  and there is no back coupling  $i \rightarrow j$ ).

Naturally,  $\tau > 0$  is a parameter that has to be tuned appropriately to the time-scales involved in the series. If  $\tau$  is too small  $DI_{ij}(\tau)$  will capture short time scale directionality, and may fail if the time series behave too similarly on those time scales as they do if they are subjected to the same external forcing. On the other hand, if  $\tau$  is too large, larger than the decorrelation time of the time series, the effect of the past  $i$  over  $j$  (and of  $j$  over  $i$ ) will be negligible and  $DI_{ij}(\tau)$  will be a small and in principle random value.

### 3.3 Ordinal analysis

Ordinal symbolic analysis is an approach to the investigation of long and complex time series. It is concerned with the order relations between successive values instead of the values

themselves. It was introduced by Bandt and Pompe [167], and is an approach to complex dynamics aiming to capture the essential aspects of complexity by the use of conceptually simple models.

As mentioned in section 3.2.1, the quantification of the predictability of a system is one of the aims of non-linear time series analysis. The problem becomes more complicated as these time series can be embedded in noise, even in non-linear ways (known as *multiplicative noise*). If there is no recognizable structure in the system, this can be, in principle, considered to be stochastic. However, recurrent intrinsic patterns can be hidden in the dynamics of the system and can be overlooked. In order to overcome this problem, in [167] a different approach to perform this analysis has been proposed. The order relation between the values of a time series is considered, notwithstanding the values themselves. A permutation complexity measure is defined, called *Permutation Entropy*, which is based on the distribution of ordinal patterns. Formal proofs and applications of the analysis of real world time series are provided in [180].

### 3.3.1 Ordinal Patterns

Consider that one wants to compute the probability distributions of a time series, for example, for computing the similarity measures discussed previously. This can be the output of a random process, an orbit of a dynamical system, observed or modeled climate data, and it can be embedded in external and dynamical noise. In order to compute many measures as the entropy—Eq. (3.11)—or the mutual information, among others, it is necessary to obtain a good estimate of the probability distribution  $p_i$  of the time series.

A simple but widespread way is to do so in terms of histograms of values computed with a certain number of bins,  $N_{bin}$ , which is limited by the length of the time series. The selection of the number of bins is a nontrivial problem, as different bin sizes can reveal different features of the data. Using wider bins where the density—the amount of points per bin—is low, reduces noise due to sampling randomness; using narrower bins where the density is high—and therefore the signal is higher than the noise—gives greater precision to the density estimation. The technique of variable bin-width can be beneficial in some cases [182]. On the other hand equal-width bins are widely used for simplicity. Another approach is to construct Ordinal Pattern (OP) time series from the original data. Ordinal Patterns are calculated by noting the value of a data point relative to other values in the series. Using e.g.



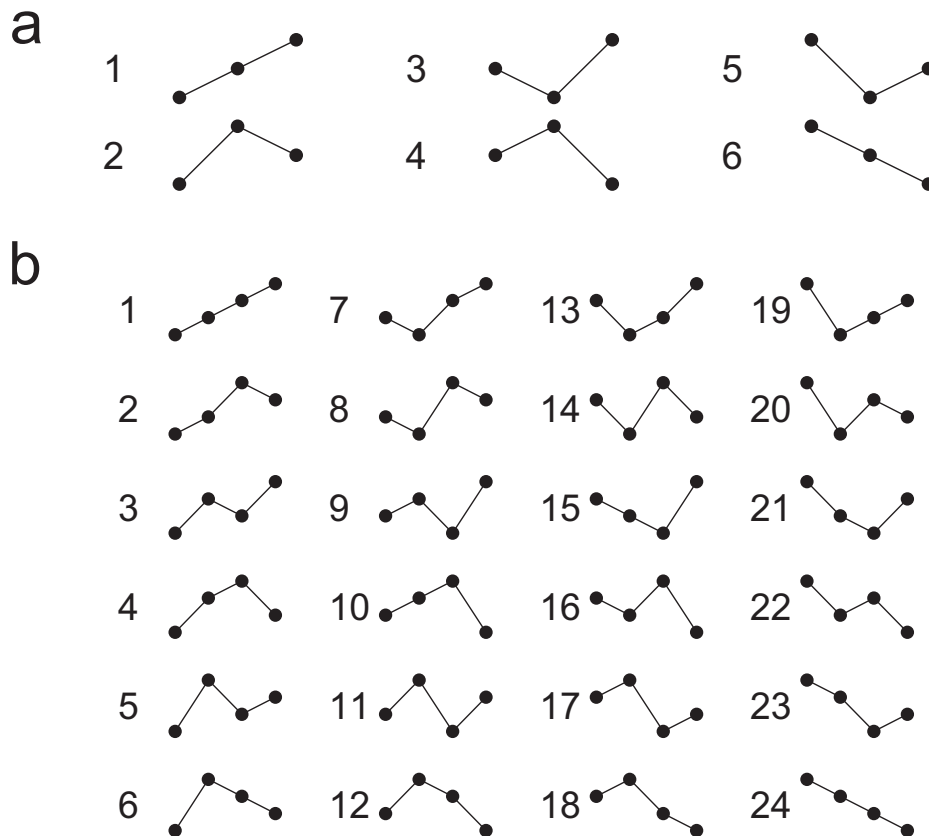


Figure 3.3: An example for ordinal patterns of (a)  $n = 3$  and (b)  $n = 4$  letters. For both cases, the  $n!$  possible patterns for are shown. This can be easily extended for larger values of  $n$  where the possible patterns will rapidly increase. Figure adapted from [181].

three symbols (letters)  $3! = 6$  different patterns exist, for four symbols there will be  $4! = 24$  patterns, and so forth—see Fig. 3.3 (a) and (b) respectively. The possibility of two equal adjacent values is not considered, partly due to the presence of noise in the time series.

As shown in Fig. 3.3 (a), OPs of length 3 are formed by 3 symbols in the following way: if a value ( $x_i(2)$ ) is higher than the previous one ( $x_i(1)$ ) but lower than the next one ( $x_i(3)$ ), it will yield the pattern '123' (Fig. 3.3 (a-1)), while the opposite case ( $x_i(3) < x_i(2) < x_i(1)$ ) will give the pattern '321' (Fig. 3.3 (a-6)), etc. This symbolic transformation allows to detect correlations in the sequence of values which are not taken into account when using histograms of values as they do not consider the order in which the values appear in the time series. As a drawback, this technique does not contain information about the relative magnitudes; this is usually useful as a natural robustness under low to moderate noise embedding.

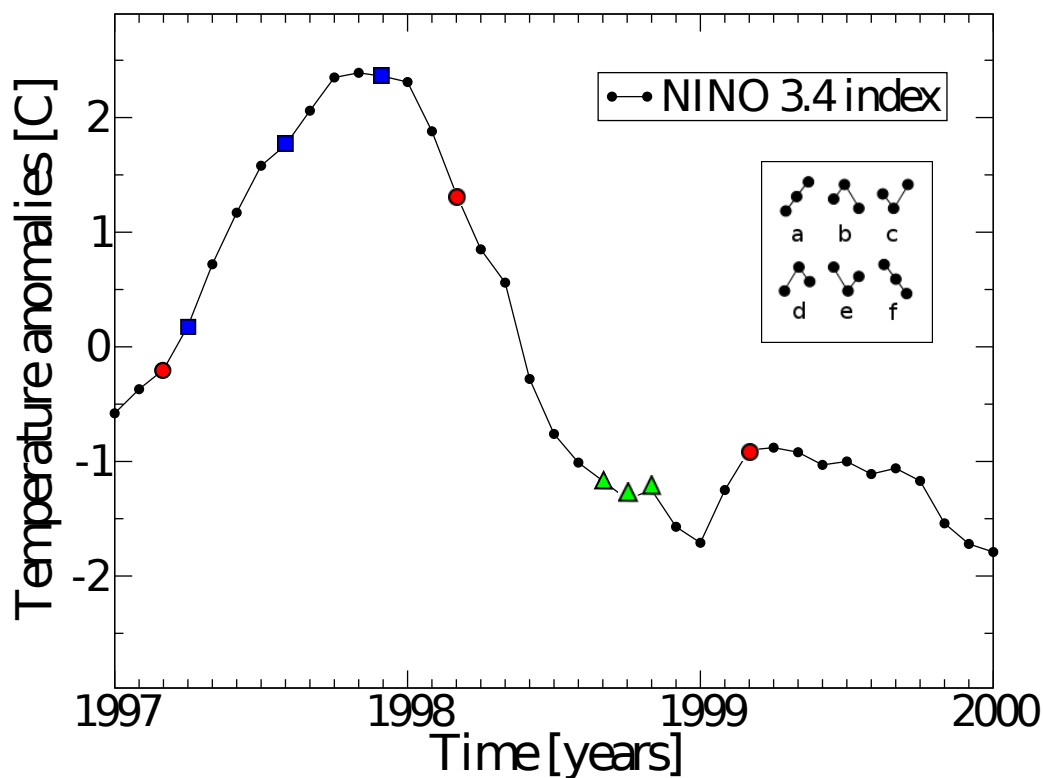


Figure 3.4: An example of three ordinal patterns in the time series of the NINO 3.4 index (monthly averaged). Green triangles: intraseasonal pattern, blue squares: intra-annual pattern and red circles: interannual pattern. The possible patterns for  $D = 3$  are shown in the inset. In this example, the intraseasonal pattern corresponds to an “e”, the intra-annual, to an “a” and the interannual, to a “b”. Figure taken from [154].

After constructing time series of the OPs, histograms can be calculated from them. In this symbolic approach, the number of bins is naturally defined by the number of possible patterns, which in turn is determined by the number of symbols in the ordinal pattern. As explained above, if the OP word is of length  $n$ , there will be  $n!$  possible patterns, and this will be the number of bins used for computing the probabilities associated with the symbolic sequences. This eliminates the binning problems which frequently appear when using histograms.

Ordinal patterns do not need to be constructed with immediately adjacent data points only. It is possible to construct them with data points that are separated in time, and in this way consider different time scales. Lets assume monthly sampled climatic data as e.g. anomalies in the central Pacific ocean temperature (NINO3,4 index) shown in Fig. 3.4: the

time-interval of the ordinal patterns can be varied by considering not only 3 consecutive months—intra-season time scale; e.g. January, February, March; February, March, April; etc shown in green triangles—but also in 3 consecutive years—inter-annual time scale; e.g., January 2010, January 2011, January 2012; February 2010, February 2011, February 2012; etc, shown in red circles—or any other convenient time scale as inter-season time scales as shown in blue squares. In this way OPs allow to characterize timescale-dependent phenomena, which is very difficult to detect when using histograms of the data. For example, a separation of 12 months allows grouping together individual months of the year.

This symbolic transformation keeps the information about correlations present in a time series at the selected time scale, but does not keep information about the absolute values of the data points. Therefore, the mutual information computed from ordinal patterns (MIOP) can be expected to provide complementary information with respect to the standard method of computing the mutual information (MIH).

This feature of OP for climate time series analysis, was first introduced in [128], where several sizes of ordinal patterns were also studied.

## 3.4 Statistical Significance

The present-day concept of statistical significance originated from Ronald Fisher when he developed statistical hypothesis testing in the early 20th century [183]. His methods are used to determine whether the outcome of a study would lead to a rejection of the *null hypothesis* based on a pre-specified low probability threshold called *p-values*, which can help to decide if a result contains sufficient information to cast doubt on the null hypothesis [184–186].

Statistics quantifies the outlier status of an observation  $x$  by the probability of sampling another observation from the *null distribution* that is as far or farther away [184]. The null distribution is the one obtained assuming that the null hypothesis is true. It is an integral part of statistical hypothesis testing where it helps to decide if a null hypothesis can be rejected.

In any experiment or observation that involves drawing a sample from a population there is always the possibility that an observed effect would have occurred due to sampling error or chance alone. However if the probability of obtaining an at least as extreme result—large difference between two or more sample means—given the null hypothesis is true, is less than a pre-determined threshold (e.g. 5% chance), then it is possible to conclude that the

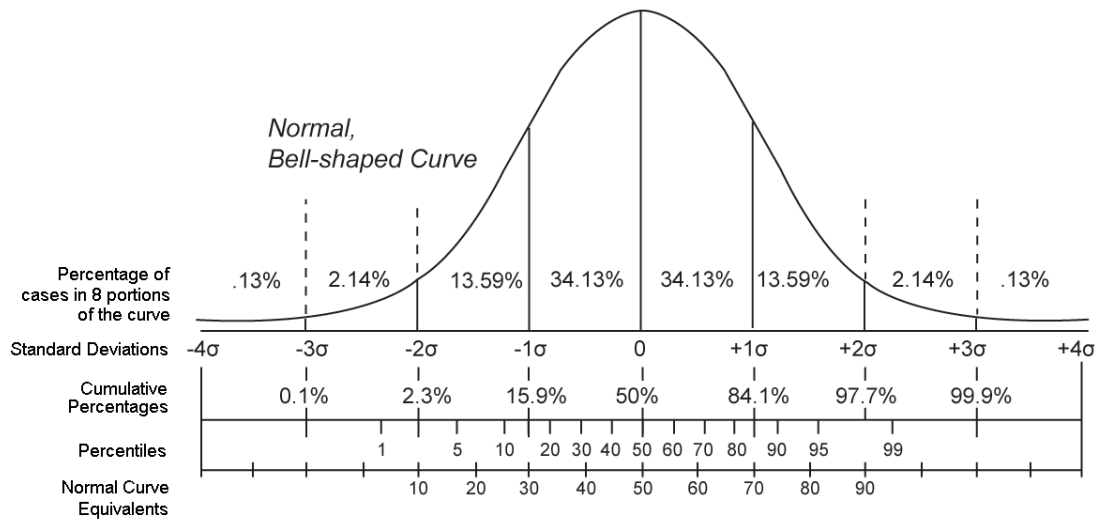


Figure 3.5: Classical bell-shaped (Gaussian) curve where the cumulative Percentages are shown, for several values of standard deviations  $\sigma$ . The percentiles and normal curve equivalents are also shown. Adapted from [187]

observed effect is due to chance [152]. The use of p-values imply a Gaussian null distribution; however other methods exist, notably the use of quantiles or percentiles [188, 189] (see Fig. 3.5).

In Figure 3.6 the mechanism of statistical testing is shown. Suppose a value of  $x = 12$  is observed while there is a reason to believe the reference value is  $\mu = 10$  as seen in figure 3.6 (a). A (Gaussian) distribution of reference—Fig. 3.6 (b)—called the null distribution is used. It embodies the null hypothesis ( $H_0$ ) that the observation is a sample from the pool of all possible instances of measuring the reference, that is that the event is a pure result of chance. The purpose of a statistical test is to locate the observation on this distribution to identify the extent to which it is an outlier. This probability is the p-value, which is the output of common statistical tests. It is calculated from the area under the distribution curve in the shaded regions—Fig. 3.6 (c).

The p-values are often coupled to a significance or  $\alpha$  level or cumulative percentages, which is also set ahead of time, usually at  $\alpha = 0.05$ . Thus, if a p-value was found to be less than 0.05, then the result will be considered statistically significant and the null hypothesis will be rejected. Other significance levels, such as  $\alpha = 0.1$  or  $\alpha = 0.01$ , may also be used,

depending on the accuracy needed. Another method is to use  $\mu \pm n\sigma$ , with  $\mu$  and  $\sigma$  being the mean value and the standard deviation of the null distribution, and  $n$  a number used to quantify the separation from the mean.

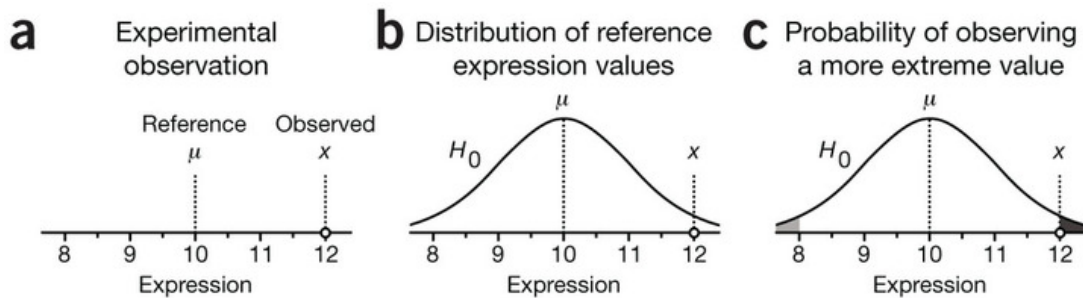


Figure 3.6: The mechanism of statistical testing used. (a-c) the significance of the difference between observed ( $x$ ) and reference ( $\mu$ ) values (a) is calculated by assuming that observations are sampled from a distribution  $H_0$  with mean  $\mu$  (b). the statistical significance of the observation  $x$  is the probability of sampling a value from the distribution that is at least as far from the reference, given by the shaded areas under the distribution curve (c). This is the p-value. Figure adapted from [184].

### 3.4.1 Surrogate Data and Bootstrap

To find an adequate null hypothesis in complex system is a nontrivial problem [190, 191]. Historically, as shown in figure 3.5, the method stated above was used, assigning a Gaussian of a certain mean and standard deviation and subsequently adjusting its parameters with further observations.

Another approach is to generate *surrogate data*. This is data not coming from observations but responding to the null hypothesis.

Many algorithms to generate surrogate data have been proposed. They are usually classified in two groups [160, 192]:

- Typical realizations: data series are generated as outputs of a well-fitted model to the original data (as climate models) [193].

- **Constrained realizations:** data series are created directly from original data, generally by some suitable transformation of it which destroys some of the properties while conserving others.

The constrained realization surrogate data methods do not depend on a particular model, nor on any parameters—and thus are called non-parametric methods. These surrogate data methods are usually based on preserving the linear structure of the original series. Some of the methods used are:

**Random Shuffle (RS)** [128, 190, 191]: New data are created simply by random permutations of the original time series. The permutations guarantee the same amplitude distribution than the original series, but destroy any linear correlation. This method is associated to the null hypothesis of the data being uncorrelated in time.

**Surrogate Fourier Transform (SFT)** also known as RP, for Random Phases [145, 190, 194, 195]: In order to preserve the linear correlation (and the power spectra) of the series, surrogate data are created by the inverse Fourier Transform of the modules of Fourier Transform of the original data with new (uniformly random) phases. If the surrogates must be real, the Fourier phases must be antisymmetric with respect to the central value of data. This method, however, does not conserve the amplitude distribution. The **Amplitude Adjusted Fourier Transform (AAFT)** has approximately the advantages of the two previous ones: it attempts to preserve both the linear structure and the amplitude distribution transforming the PDF in a gaussian before applying the SFT to take advantage of the property of Gaussians under Fourier transforms, and then applying the opposite transformation. As a drawback it is a slow method and due to these nonlinear transformations the linear structure may be changed. Another drawback is that a long time series is in general needed, which is not usually the case in climatology.

**The bootstrap method (BS)** [196, 197]: Randomly resamples with replacement from the original set of data using blocks of data of approximately the size of the autocorrelation of the time series, and re-computes the estimator on the collection of resampled data. This is repeated many (e.g., 100+) times and the value of the estimator for each resample is stored. This way, an empirical distribution of the estimator is constructed from which various statistical tests as the p-value can be performed to compare different values of the estimator.

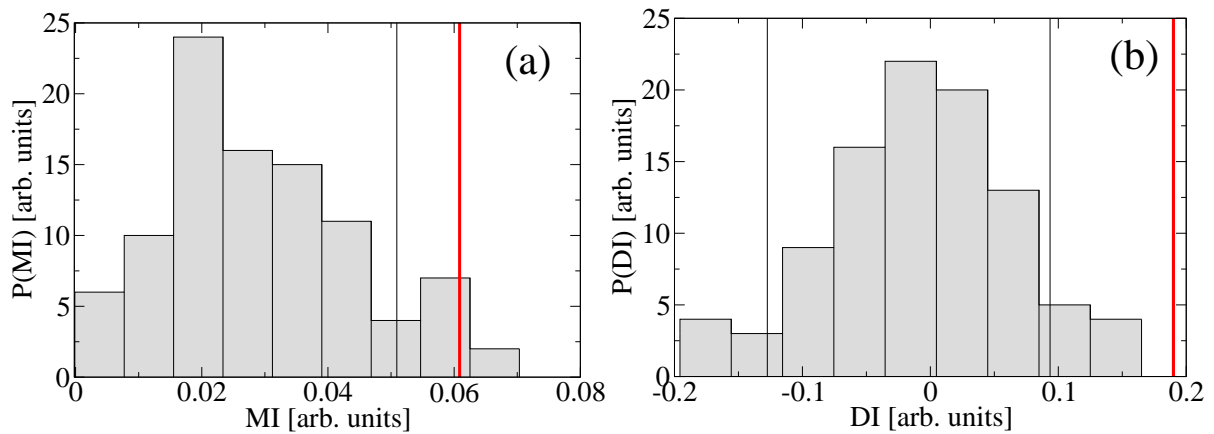


Figure 3.7: Illustration of the method p-value for (a) the MI and (b) DI. In both cases the BS method was used. The observed value is marked with a red vertical line, while the significance marks ( $\mu + 3\sigma$  for MI and  $\mu \pm 3\sigma$  for DI) are shown in black vertical lines. Both values in this illustration are significant.

Figure 3.7, shows the methodology explained in two cases for real data. In panel 3.7(a) 100 BS surrogates were generated and result in the distribution shown in gray. The black vertical line correspond to the chosen p-value ( $\mu + 3\sigma$  of the distribution). The red vertical line is the MI calculated from the real data. As in this case, the value is higher than the black line, this value is significant. Panel 3.7(b) shows a similar analysis but for DI where there are two significance bars, (1) corresponding to the case  $1 \rightarrow 2$  ( $X > \mu + 3\sigma$ ) and (2)  $2 \rightarrow 1$  ( $X < \mu - 3\sigma$ ). The case represented in the figure corresponds to the first series significantly driving the second.

The methodology used during the present work is the explained above, with some variations. In chapters 4 and 5 random shuffle data was used (as the data was monthly sampled) and the threshold was global, this is, a single p-value was obtained from the statistical analysis and used for thresholding of all nodes. In chapter 6, the BS method was preferred because of the great autocorrelation of *daily data* and the more computationally expensive local thresholding method was preferred.





PART 2

RESULTS



## Chapter

# 4

## ***Interdependencies in Climate Networks on Several Time Scales***

This chapter presents results for the construction of climate networks applying the methods described in the previous chapters. Firstly, in section 4.1 the data analyzed is described together with the methodology of ordinal patterns used in this chapter to select time scales. Intra-season time-intervals (e.g., the patterns are formed by anomalies in consecutive months) and inter-annual time-intervals (e.g., the patterns are formed by anomalies in consecutive years) are considered. The criteria used for thresholding are described in section 4.2, which discusses how the network density and topology change with the time scales, and provide evidence of correlations between geographically distant regions that occur on specific time scales. In particular, it shows that an increase in the ordinal pattern spacing (i.e., an increase in the timescale of the ordinal analysis), results in climate networks with increased connectivity in the equatorial Pacific area. On the contrary, the number of significant links decreases when the ordinal analysis is done with a shorter timescale (by comparing consecutive months). Finally, section 4.3 presents a discussion of these results, which were summarized in J.I. Deza *et. al* [130].

### 4.1 Data and methodology

The monthly-averaged surface air temperature (SAT) anomalies, taken from the reanalysis of the National Center for Environmental Prediction/National Center for Atmospheric Research, NCEP/NCAR [33] was analyzed. The anomalies were calculated as the actual tem-

perature values minus the monthly average, and were normalized by the standard deviation. The data are given on a grid over the Earth's surface with latitudinal and longitudinal resolution of  $2.5^\circ$ , resulting in  $N = 10512$  grid points or network nodes. The data cover the period from January 1949 to December 2006, and thus, in each node, there is a time series of 696 data points. This is the same dataset used in [128] and some of the results presented here are compared to those of that article.

Networks were constructed using MI values. These were computed calculating the probabilities as histograms, directly from the time-series, or as probabilities of occurrence of ordinal patterns (OP). Several OPs—of different lengths and spacing—were considered. The sensitivity of the results to the number of  $N_{\text{bin}}$  used for computing the probabilities was also studied.

More specifically, when using histograms (in the following, referred as MIH), 6 and 24 bin histograms were constructed. They were used for comparison with histograms of OPs of length 3 and 4 letters which yield  $3! = 6$  and  $4! = 24$  bins respectively.

The ordinal patterns were formed by:

- consecutive months (i.e., by comparing three or four consecutive values in the time series); the MI computed in this way will be referred to as MIOP3L01 and MIOP4L01 respectively;
- months in consecutive years (i.e., by comparing  $x_i(t)$ ,  $x_i(t + 12)$ , and  $x_i(t + 24)$ ); the MI computed in this way will be referred to as MIOP3L12 (OPs of length 3) and MIOP4L12 (OPs of length 4).
- equally spaced months covering a one-year period. For patterns of length 3, this is done by comparing  $x_i(t)$ ,  $x_i(t + 4)$ , and  $x_i(t + 8)$ ; for patterns of length 4, by comparing  $x_i(t)$ ,  $x_i(t + 3)$ ,  $x_i(t + 6)$ , and  $x_i(t + 9)$ . The MI computed in these ways will be referred to as MIOP3L04 and MIOP4L03 respectively.

A main goal during this study was to find an adequate thresholding methodology, that results in climate networks that i) contain only truly relevant connections—the links that represent random correlations are filtered, and ii) do not disregard the weak links that are significant—i.e., the links representing statistically significant deviations from random correlations are not filtered.

In order to achieve this, firstly constant density networks already reported [120, 128] were considered. A global threshold  $\tau$  was calculated (and indicated in each map) as well as the network density (the number of links divided by the number of possible links). This methodology was compared to the random shuffle surrogate data method which was introduced in section 3.4.1. This yielded another value of  $\tau$  whose maps allowed finding weaker but still significant links. In both cases, links with mutual information values below the threshold were considered not significant and disregarded. The physical interpretation of the accepted links was then discussed from a climatological point of view.

## 4.2 Results

Figures 4.1 and 4.2 present the results of the four methods of analysis, considering 6 bins (Fig. 4.1) and 24 bins (Fig. 4.2). Ordered from top to bottom, the network was constructed by computing the MIH (top row), the MIOP with OPs formed by consecutive months (second row), the MIOP with OPs covering a one-year period (third row) and with the same month in consecutive years (bottom row). Note that in order to cover a year period, the spacing on the third row in Fig. 4.1 is four months (as there are three letters) and it is three months in Fig. 4.2 (which uses four letters). In both figures the left column presents the AWC plots, and the right column, the connectivity maps for a point in the equatorial Pacific.

The AWC obtained here compares well to that calculated in [128] (left panel, Fig. 1), except that the network density is about double. This is explained by a different way to calculate the significance threshold. In [128] the threshold consisted in the largest MI value computed from shuffled data, while in Fig. 4.2 of this study the threshold is defined in terms of the standard deviation of the MI distribution calculated from shuffled data.

Considering the left column of Figs. 4.1 and 4.2, (showing the AWC plots for 6 bins/3 letters and 24 bins/4 letters respectively), one can observe highly connected spots on the first row which are present only in some of the other three maps. See, for example, the highly connected green spot in the Labrador Sea, which is also seen in the second and to a lesser extent in the third row; but is not present in the plot on the fourth row. The Labrador Sea is one of the most important regions of deep water formation in the north Atlantic. The formation of this water occurs in wintertime and depends on the passage of extratropical storms that cool the surface increasing its density. The passage of storms is in turn related

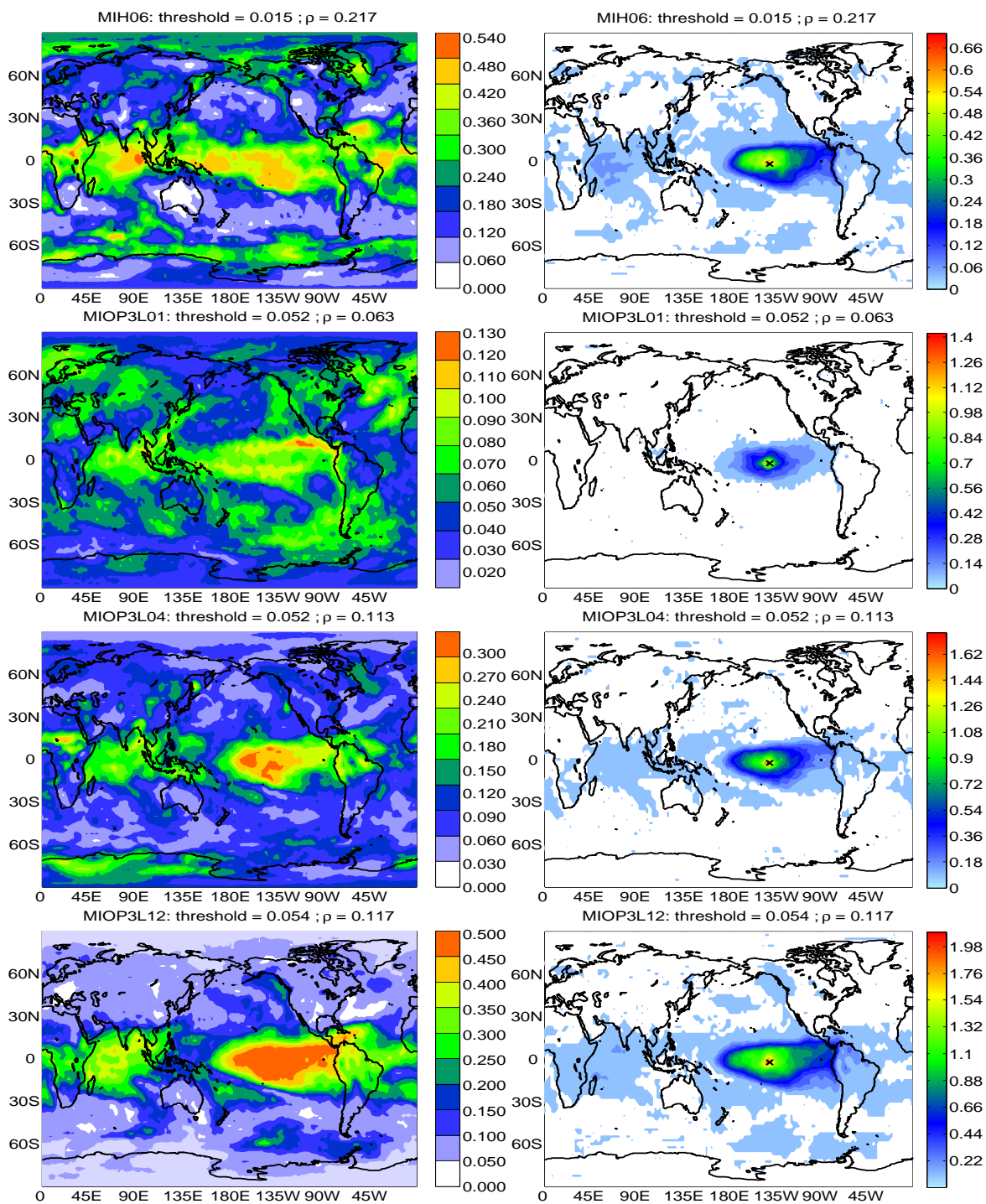


Figure 4.1: AWC (left column) and connectivity maps (right column) using the different methods of network construction described in the text, computing the probabilities with 6 bins and using the significance threshold  $\tau = \mu + 3\sigma$  to define the links. The methods are: histograms of anomaly values MIH (top row); OPs formed with three consecutive months (second row); OPs formed with three equally spaced months covering a one-year period (third row); and OPs formed with three months in consecutive years (bottom row).

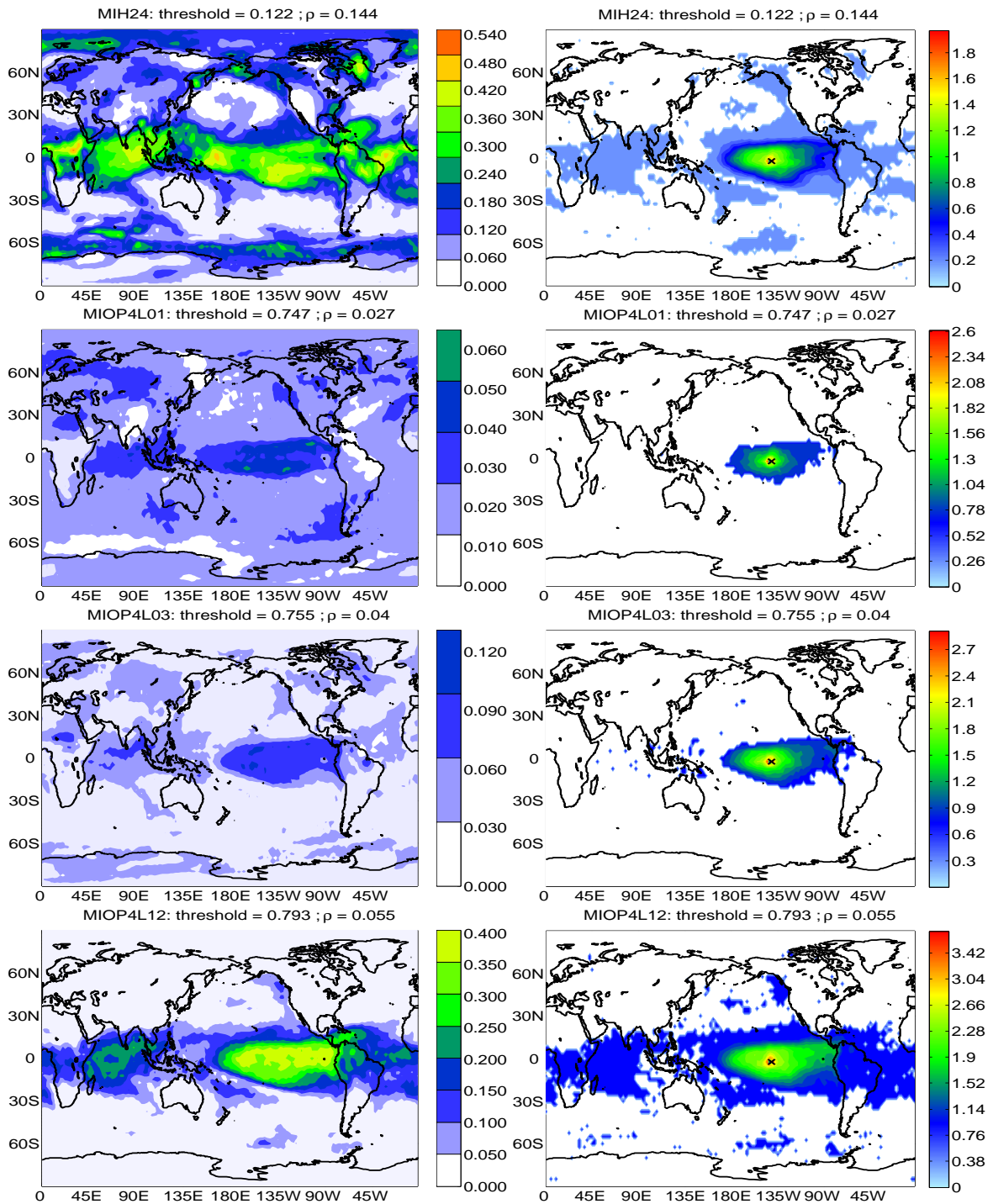


Figure 4.2: As Fig. 4.1 but the probabilities are calculated with 24 bins: histograms of anomaly values MIH (top row); OPs formed with four consecutive months (second row); OPs formed with four equally spaced months covering a one-year period (third row); and OPs formed with four months in consecutive years (bottom row). In order to better compare the AWC plots, each left column map has same color-AWC correspondence of the corresponding map in the left column in Fig. 4.1.

to the state of the North Atlantic Oscillation, a preferred pattern of atmospheric variability in the north Atlantic basin. As result, there is a clear connection of the Labrador Sea with the rest of the north Atlantic mainly on seasonal time scales and is mostly independent on ENSO activity (see second row of Figure 4.1 ). This topic will be further addressed in chapter 5.

In addition, in the first row of the same Fig. 4.1, it can be noticed highly connected areas in Africa, the equatorial Atlantic and western tropical north Atlantic which are not present in the short-time scale networks (second row) but that are seen in the long-time scale networks, (third and fourth rows). Thus, these connections arise because regions are connected on inter-annual, but not on monthly or seasonal time scales.

On interannual time scales El Niño teleconnections include a decrease of the northern trade winds that reduce the heat loss in the western tropical Atlantic, as well as a tropospheric warming over most of the tropical band, thus inducing warming over several regions, including the equatorial Atlantic and the Indian ocean [198]. At the same time, the air-sea interaction in the equatorial Atlantic leads to inter-annual modes of variability which can interact constructively or not with the anomalies induced by El Niño, resulting in less number of links compared to those in the Indian ocean [199].

On interannual time scales ENSO also influences climate over the northeastern Pacific and the south Pacific, including the Antarctic Peninsula, probably through the propagation of Rossby waves (Figure 4.1, last row). The latter Pacific link has been recently suggested as the mechanism responsible for the temperature trend in the Antarctic Peninsula during the last 30 years [200].

According to these results the tropical region becomes interconnected on seasonal time scales, while the extratropics become connected to the equatorial Pacific only when considering interannual time scales. These teleconnection patterns evidence the propagation of Kelvin and Rossby waves from the equatorial Pacific on different time scales. While the overall picture is similar in Figs. 4.1 and 4.2, it is clear that the use of 6 bins represents more adequately the known atmospheric processes, probably due to the shortness of the time series that prevents weak links to be declared significant when using 24 bins to calculate the mutual information.

To analyze the influence of the significance criterion, in Figs. 4.3 and 4.4 the networks are presented when the threshold to define the links is such, that all the networks have the



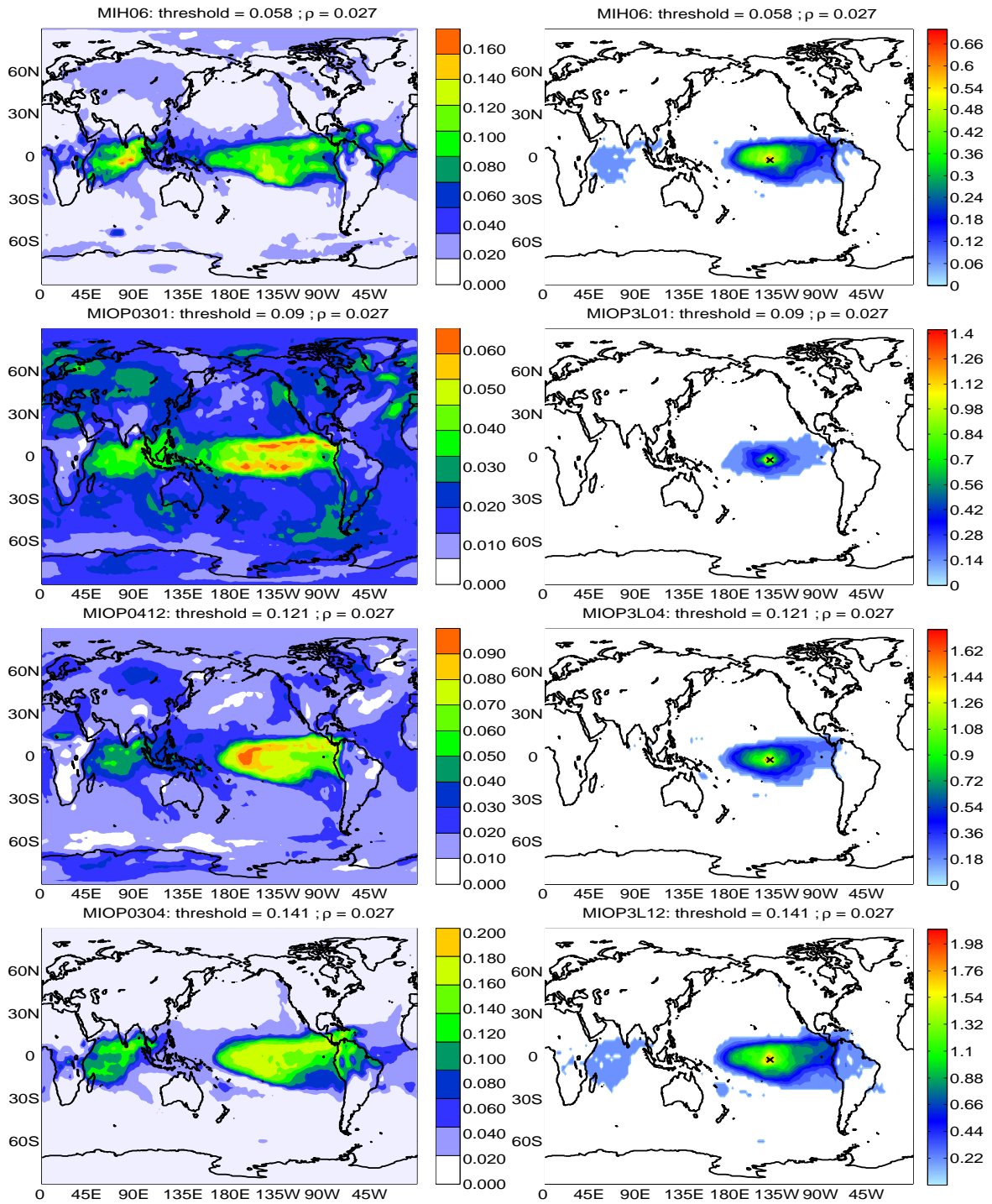


Figure 4.3: As Fig.4.1 but with the threshold  $\tau$  chosen such that the networks have the same link density (0.03%).

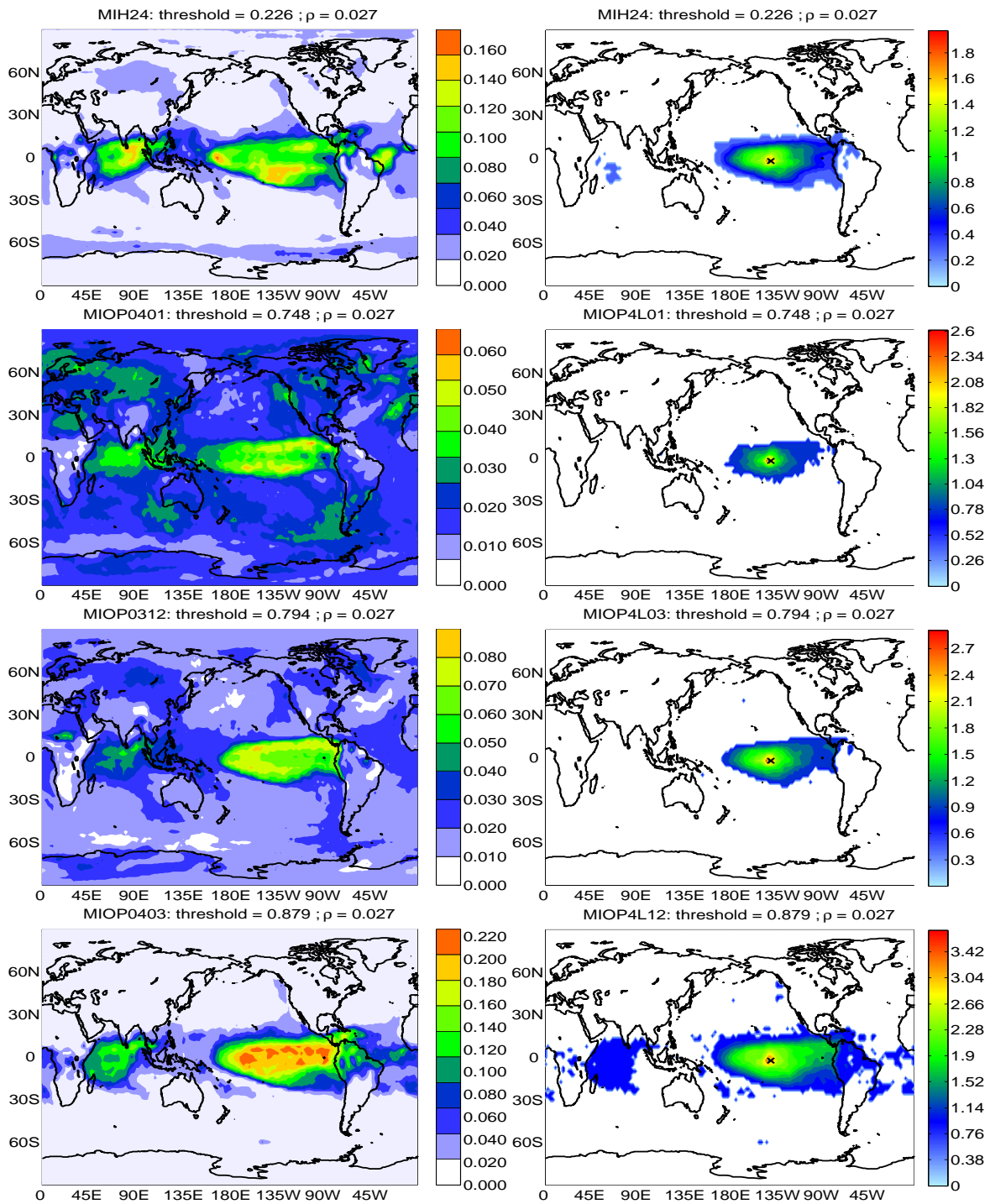


Figure 4.4: As Fig.4.2 but with the threshold  $\tau$  chosen such that the networks have the same link density (0.03%). In order to better compare the AWC plots, each left column map has same color-AWC correspondence of the corresponding map in the left column in Fig. 4.3.

same number of links. A density of links of 0.027% was chosen, because all the networks presented in Fig. 4.1 and 4.2 have a density equal or larger than this value. This is the approach used in [128]. In this previous work, it was found that the short-time scale network MIOP4L01 was more uniformly connected than the year-time scale network MIOP4L12. The present analysis confirms this result and extends it to the case where the density depends on the significance test (Figs 4.1 and 4.2). This suggests that on short time scales there is no dominant phenomenon that interconnects remote regions. Instead, temperature anomalies seem to be governed by regional patterns of atmospheric internal variability.

### 4.3 Discussion and Conclusions

In this chapter different aspects of the global climate were studied, some were associated to short time scales—of a few months—and others to longer time scales—of a few years—using the framework of climate networks and nonlinear time-series analysis. The goal was to examine the influence on the network topology of i) the methodology used for quantifying the degree of statistical similarity between two nodes, and ii) the significance criteria used for thresholding, to define the links.

The mutual information measure was employed, calculated from probabilities that were defined over i) a small number of bins (6, Figs. 4.1 and 4.3), and ii) a large number of bins (24, Figs. 4.2 and 4.4); also, the probabilities were computed directly from histograms of SAT anomaly values, and from the symbolic, ordinal representation of the anomaly time series. This was done in order to compare the dependence of the network topologies on the number of letters of the ordinal representation.

The OP method allowed considering different time scales when transforming the anomalies time series into ordinal patterns. The influence of thresholding was considered, and the threshold was defined i) in terms of surrogate data (significance test) (Figs. 4.1, 4.2) and ii) to obtain a network with a given link density (Figs. 4.3, 4.4).

It was found that an increase in the ordinal pattern spacing (therefore an increase in the timescale), generates a growth of the connectivity in the equatorial Pacific area (figures 4.1–4.4, left column, rows 2–4). It was also found that this increase in connectivity is associated with an increase on the teleconnections from points situated on this area. (right column of the same figures). This result is consistent with previous work by Barreiro *et. al.* [128],

and with current understanding of ENSO influence. Thus our method allows us to obtain networks where the effect of ENSO goes from weak (monthly OP) to strong (yearly OP) independently of the number of letters used for the ordinal patterns and of the thresholding method.

It was also observed that the number of significant links is smaller when the ordinal patterns are constructed with a shorter timescale—by comparing consecutive months—what was interpreted as due a larger stochasticity in the time-series in the short timescale.

It was found that, when the networks are set at a fixed link density (figures 4.3, 4.4) the networks—constructed with 6 and 24 bins respectively—are remarkably similar. It can be argued that 24 bins is a tradeoff of resolution and data length, a measure equivalent to the 6 bins MI, but more flexible in its essence due to its higher resolution, it would certainly give better results for e.g. longer time series. Moreover, these results indicate that the significance of the links in climate networks should be carefully examined in order to avoid disregarding weak but significant links.

While in [128] the significance threshold used was the maximum MI value obtained from shuffled time-series, a more “tolerant” threshold was here considered—the mean value plus 3 standard deviations of the MI distribution—that resulted in networks with a higher number of links, as compared to those in the above mentioned work. The networks obtained are consistent with those in [128] and it was found that, by increasing the number of links, networks that display in more detail the complexity of the atmospheric teleconnections were found. Another conclusion of this chapter is that, since Figs. 4.3 and 4.4 are very similar, for a fixed link density, the main features of the network are independent of the bin number used for computing the mutual information, which confirms the robustness of the climate networks constructed with this thresholding methodology.

## **Chapter**

# **5**

## ***Components of Atmospheric Variability and Time Scale Identification***

To further exploit the knowledge of the links uncovered via the network analysis shown in the previous chapter, a good understanding of the physical mechanisms underlying these links is required. In this chapter the focus is on constructing climate networks representing internal atmospheric variability, and variability forced by the surface ocean conditions using the output of an ensemble of runs from an Atmospheric General Circulation Model (AGCM). In Section 5.1 the methodology used for separating variability types is introduced. The data and the model employed are discussed in Section 5.2 together with the model validation in the context of climate networks. Sections 5.3 and 5.4 present the results. Finally section 5.5 presents the discussion and conclusions.

A main strength of the work presented in this chapter is that the networks are constructed using MIOP, which allows the separation of intraseasonal, intra-annual and interannual time scales. This provides a further insight to the analysis of climatological data.

It was found that the connectivity of the ocean forced variability network is heavily affected by “El Niño”: removing its influence (using the methodology explained in section 3.1.1) yields a general loss of connectivity; even connections between regions far away from the equatorial Pacific ocean are lost, suggesting that these regions are not directly linked, but rather, are indirectly interconnected via El Niño, particularly at interannual time scales.

On the contrary, in the case of the internal variability network – independent of Sea Surface Temperature (SST) forcing it was found that NAO plays a dominant role with a maximum on intra-annual time scales in the North Atlantic region. It was further shown that the strongest non-local links found are those forced by the ocean.

The results presented in this chapter were summarized in [154].

## 5.1 Internal and forced variability

Atmospheric variability can be considered, to first order, as a superposition of an internal part due to intrinsic dynamics, and an external part due to the variations of the boundary conditions, primarily given by the sea surface temperature (SST) forcing. These two components can be separated by using Atmospheric General Circulation Models (AGCMs) forced with prescribed historical SSTs [201–205].

This separation between internal and forced atmospheric variability is a standard procedure to study the impact of the oceans on the atmosphere and has led to important advances on our understanding of the dynamics involved. Two books, [206] and [54] provide a summary of the processes involved—mainly based on the propagation of Rossby waves and the generation of teleconnection patterns. Although there are some nonlinear secondary effects, the theory asserts that to first order the observed propagation and establishment of teleconnection patterns is linear.

Separating forced from internal atmospheric variability is also important as it may allow for improvements seasonal in climate prediction. In many geographical regions, the atmosphere is strongly influenced by local or remote SST variations that force persistent regional anomalies [207]. Because the evolution of the tropical oceans presents some predictability at time scales longer than the atmosphere, prediction of atmospheric variables beyond the chaotic time scale of 7-10 days is possible provided that the atmospheric dynamics is being forced by the ocean [207].

The usual modeling strategy to separate atmospheric components of variability consists in forcing AGCMs with idealized or observed SST anomalies. The experiment involves the generation of an ensemble of runs initialized differently but forced with the same SST as boundary conditions. Then, the simulated time series of anomalies of a climatic field (e.g. SAT anomalies) is considered as a combination of internal and forced variability, e.g.  $x = x_{for} + x_{int}$ . Thus, for each run  $i$  it results

$$x^i = x_{for}^i + x_{int}^i = x_{for} + x_{int}^i \quad (5.1)$$

as  $x_{for}$  does not depend on the initial conditions.

Averaging over  $N$  runs yields

$$\bar{x} = x_{for} + (1/N) \sum_i x_{int}^i \quad (5.2)$$

thus, the ensemble mean is a biased estimate of the forced component. If  $N$  is large enough, the second term will be small as each model run will have a different value. Thus, to the first order  $\bar{x} \approx x_{for}$

In other words, each time series  $x^i$  can be separated into a part that changes from run to run of the model,  $x_{int}^i$ , and a part that does not depend on the initial conditions (is forced by the boundary conditions only, and is the same for all runs),  $x_{for} \approx \bar{x}$ .

Similarly one can define the internal variability as

$$x_{int}^i = x^i - x_{for} \quad (5.3)$$

this yields  $N$  different internal variability data sets, which are processed separately producing  $N$  different climate networks. The Internal variability network is then calculated as an average of these networks.

This method allows to construct two types of networks, one in which the links represent similarities in internal atmospheric variability (referred as *internal variability network*), and another in which the links represent similarities in forced atmospheric variability (the *forced variability network*).

Here, connectivity of these networks will be used to assess the influence of two main phenomena: El Niño-Southern Oscillation – characterized by the NINO3.4 index –, and the North Atlantic Oscillation – characterized by the NAO index. This was done by calculating the networks from time series where either the NINO3.4 index or the NAO index was linearly removed using the methodology explained in section 3.1.1.

The forced variability networks is found to be intimately related to El Niño phenomenon and that linearly removing its evolution yields a breakdown of the long range teleconnections of the climate network, particularly at interannual time scales. A similar result is observed for the internal variability network in the Northern Hemisphere when NAO is removed, with maximum effect at intra-annual time scales.

## 5.2 Data sets and model used

In this study the AGCM from the International Centre for Theoretical Physics (ICTP AGCM) has been used. It consists of a full atmospheric model with simplified physics and an horizontal resolution of T30 ( $3.75^\circ \times 3.75^\circ$ , which results in  $N = 608$  grid points or network nodes) with eight vertical levels [203]. The model is forced with historical global sea surface temperatures (ERSSTv.2) [208]. In order to separate forced from internal atmospheric variability nine runs using the same boundary (SSTs) conditions but slightly different initial conditions were performed.

In our experiment design SST is taken as a boundary condition and it is not changed by the atmospheric flow. In the real world there is a two-way interaction between the ocean and the atmosphere. This limitation is especially important in the extratropics where the SST evolution strongly depends on the atmospheric forcing [209, 210]. However, current understanding indicates that the atmosphere is most sensitive to SST anomalies in the tropics and thus the forced atmospheric variability will be related to the evolution of the tropical oceans [211]. This model setup allows, as explained in the Introduction, to separate the *forced* and *internal* components of the atmospheric variability. While an ensemble of only nine model runs might seem insufficient for a robust estimation of the forced response, as it could be contaminated by noise due to the relatively small ensemble size, it will be shown that the results found here are consistent with well known climate phenomena, indicating that, at least at the “first order” description of the network via AWC, nine model runs are enough to separate forced and internal variability. This is consistent with previous works that show that an ensemble of about 10 runs is enough to separate internal and forced variability in most places (e.g. [51, 212–214]). More sophisticated methods for identifying the forced variability despite the small-ensemble noise contamination are discussed in [199, 202, 215–217].

Monthly averaged SAT in the period January 1948 - December 2006 was analyzed. This results in a total of 708 data points per node. For each node, the time series were linearly detrended and the anomalies of these series were computed by subtracting the long term average to each monthly data point.

To validate the model (see Section 5.2.1) the reanalysis data from NCEP/NCAR [33] in the same time period (1948-2006) was considered. Since NCEP/NCAR reanalysis data is given on



a  $2.5^\circ \times 2.5^\circ$  grid, for easier comparison it was resampled using bilinear interpolation of the gridded data to fit the grid of the ICTP-AGCM data. The detrended and normalized anomalies were computed as stated with the model data.

To construct the network, a link between nodes  $i$  and  $j$  is defined if  $M_{ij}$  is above an appropriate threshold, which is calculated in terms of surrogate shuffled data as explained in previous chapters.

To represent the network the AWC is plotted, which yields the fraction of the total area of Earth to which each node is connected as explained in section 2.2.3.

It is of particular interest to identify significant weak links, as the strongest links are usually of shorter spatial range. In all the AWC maps presented, the color scale has been set from zero to a fixed value (0.4), and any node with stronger connectivity is shown with the color code of 0.4. This allows to visualize more clearly the weakest part of the accepted significant links. It also allows for a direct comparison of all the AWC maps.

The significant connections of some selected geographical regions (represented by individual network nodes) are also explored. In these connectivity maps the value of the interdependency measure (MIH or MIOP) will be displayed using a color scale which is also fixed, from zero up to 0.3; MI values larger than this will be shown using the same color code as 0.3.

### 5.2.1 Model Validation

While the ICTP-AGCM model has been used extensively in the literature (see, e.g. [51, 203, 204, 218] and references therein), the model has not yet been validated in the context of climate networks. Therefore, the first step of our study is to validate the model by comparing the networks obtained from one model run with the networks obtained from reanalysis data [130].

This is done by comparing Fig. 5.1 with Fig. 5.2. Figure 5.1(a) displays the AWC map computed from reanalysis data using MIH as interdependency measure; Fig. 5.2(a) displays the AWC map computed from one model run, also using MIH. Clearly, the model is able to capture the same overall pattern of global connectivity with a maximum in the central tropical Pacific, relative maxima in the tropical Atlantic and Indian oceans and over Alaska, Labrador Sea and the Southern ocean. Differences are mainly in the magnitude of the AWC, with the model underestimating the connectivity in most places. Similar observation applies

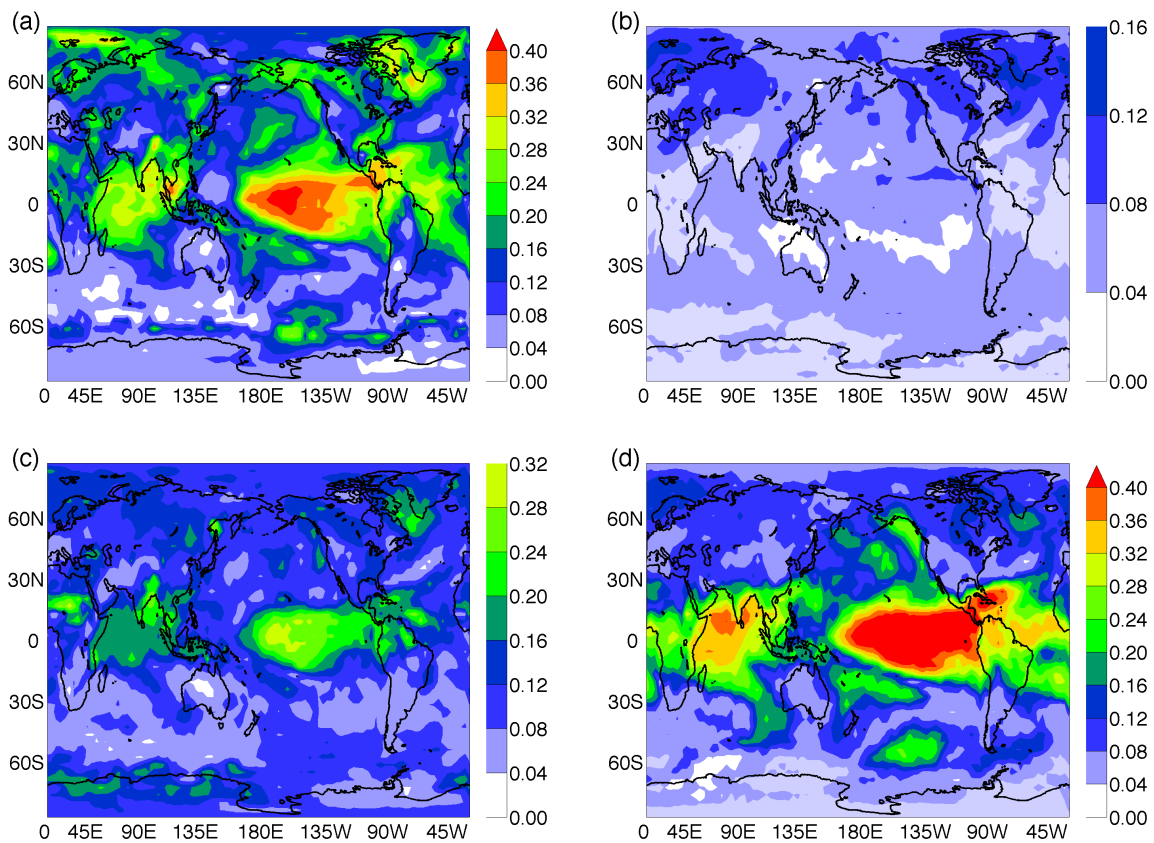


Figure 5.1: Maps of AWC constructed from reanalysis NCEP/NCAR data. The statistical interdependencies are quantified via (a) MIH, (b) MIOP intraseasonal, (c) intra-annual, and (d) interannual time scales (see sections 3.2.2 and 3.3 for details). The color scale is the same for all panels and for all the following AWC maps and it is set to saturate at 0.40 to highlight the weaker values of AWC.

to the comparison between panels b, c and d on Fig. 5.1 and the corresponding panels on Fig. 5.2 where the network was built by using the MIOP as interdependency measure.

Fig. 5.2, panel (a) shows the AWC using MIH and thus, reveals global interdependencies, on all time scales; panels (b)-(d) show the AWC using MIOP in intraseasonal, intra-annual and interannual time scale, respectively. Clearly, the connectivity increases as the time scale increases, in good agreement with the results found in chapter 4 using reanalysis data. Many other features of the AWC maps are also qualitatively well reproduced by the model.

While the networks obtained from AGCM and reanalysis data, when visualized via the AWC, look qualitatively very similar, quantitative differences are seen, for example, with re-

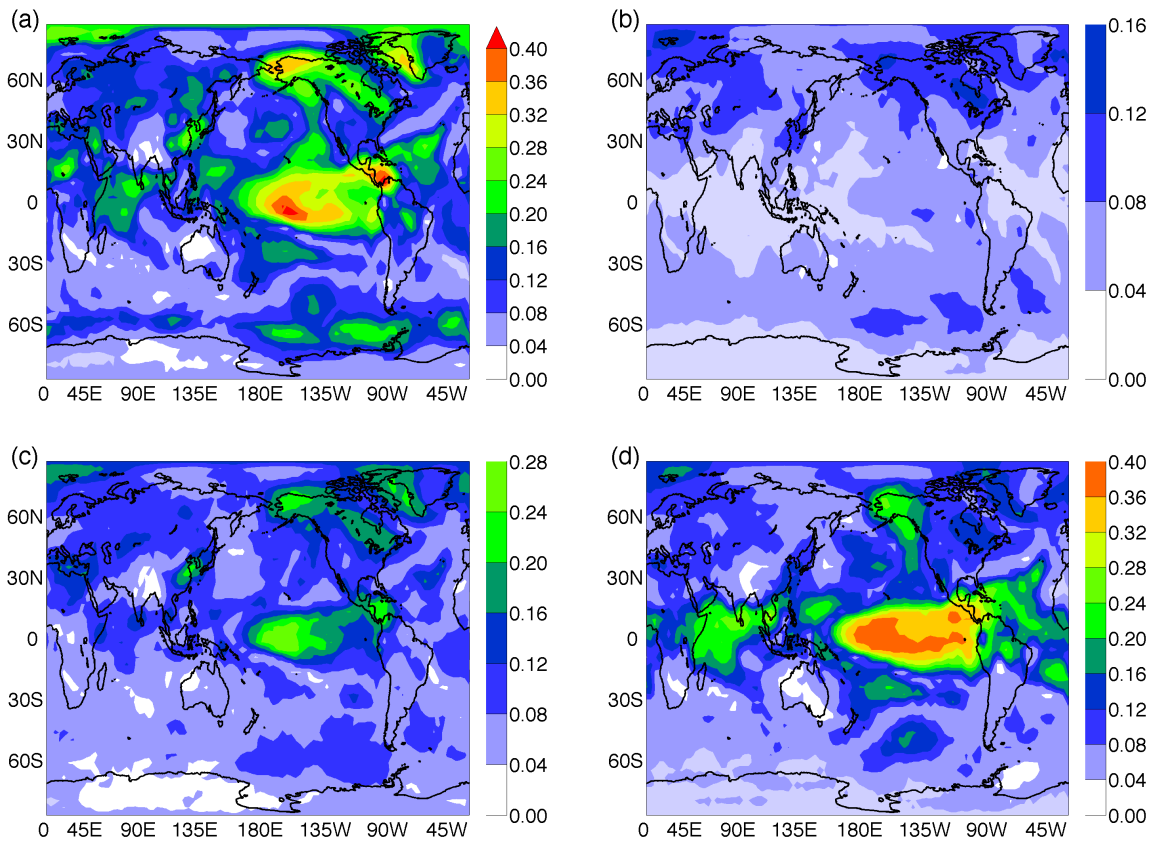


Figure 5.2: Maps of AWC obtained from single model run. The statistical interdependencies are quantified via (a) MIH, (b) MIOP intraseasonal, (c) intra-annual and (d) interannual (see sections 3.2.2 and 3.3 for details). Comparing panel (a) with Fig. 5.1(a) and panel (d) with Fig. 5.1(b) it can be observed that the main features of the maps are the same, providing a visual validation of the model.

spect to the spatial extent of the structures. These differences might be relevant, especially if more sophisticated network measures were to be used. Nevertheless, the good qualitative agreement between networks constructed from model and reanalysis data, lets us focus on using model output to distinguish the networks associated with intrinsic and forced atmospheric variability.

## 5.3 AWC maps

### 5.3.1 Forced variability

The AWC maps presented in Fig. 5.2 for one run of the model, contain information of both forced and internal variability. To analyze *forced variability* only, the network was constructed from averaged time series (over nine model runs),

$$\bar{x} = x_{for} + (1/N) \sum_i x_{int}^i \approx x_{for}$$

where  $(1/N) \sum_i x_{int}^i$  will be small to first order as explained in the Introduction.

The results are presented in Fig. 5.3. Panel (a) displays the AWC map when the MIH is used to quantify statistical interdependencies. Here, connectivity is higher in the tropics and on the Pacific, Indian and Atlantic basins than in the rest of the world. It is worth noting that while tropical connectivity is relatively symmetrical about the equator for Pacific and Indian oceans, the north Atlantic is significantly more connected than the southern Atlantic. Panels 5.3(b-d) show that the connectivity of the forced variability increases with the time scale. At intraseasonal time scales connectivity is very low compared with the connectivity from Fig. 5.3(a). If the time scale is increased to intra-annual – as in panel 5.3(c) – all the tropical area becomes more connected than the extratropics, indicating a better longitudinal energy and momentum exchange. Forced by the tropical Pacific SST anomalies a long range strong teleconnection is found in Alaska [219]. For interannual timescales (three years) which is within the period of the El Niño events (from 2 to 7 years) many very connected areas, especially in the tropics but also in the extratropics are found. The presence of highly connected spots is observed in the extratropics especially in the Pacific basin but also in the Indian and Atlantic oceans. Comparing these three maps with that in panel 5.3(a) which, as explained before, was computed via MIH and thus contains information from all the time series, it can be inferred that most of the connections seen in Fig. 5.3(a) occur at long time scales, because they are clear only in Fig. 5.3(d), and are weak or not seen in Figs. 5.3(b), (c).

Figure 5.4 represents the same maps as Figure 5.3 but after removing the ENSO influence by a linear regression. Panels 5.3(a) and 5.4(a) show large differences. It is clear that the signal of El Niño in the tropical Pacific was successfully removed, and moreover, the connection hotspots in the extratropics were also removed, indicating that they were mainly forced by El

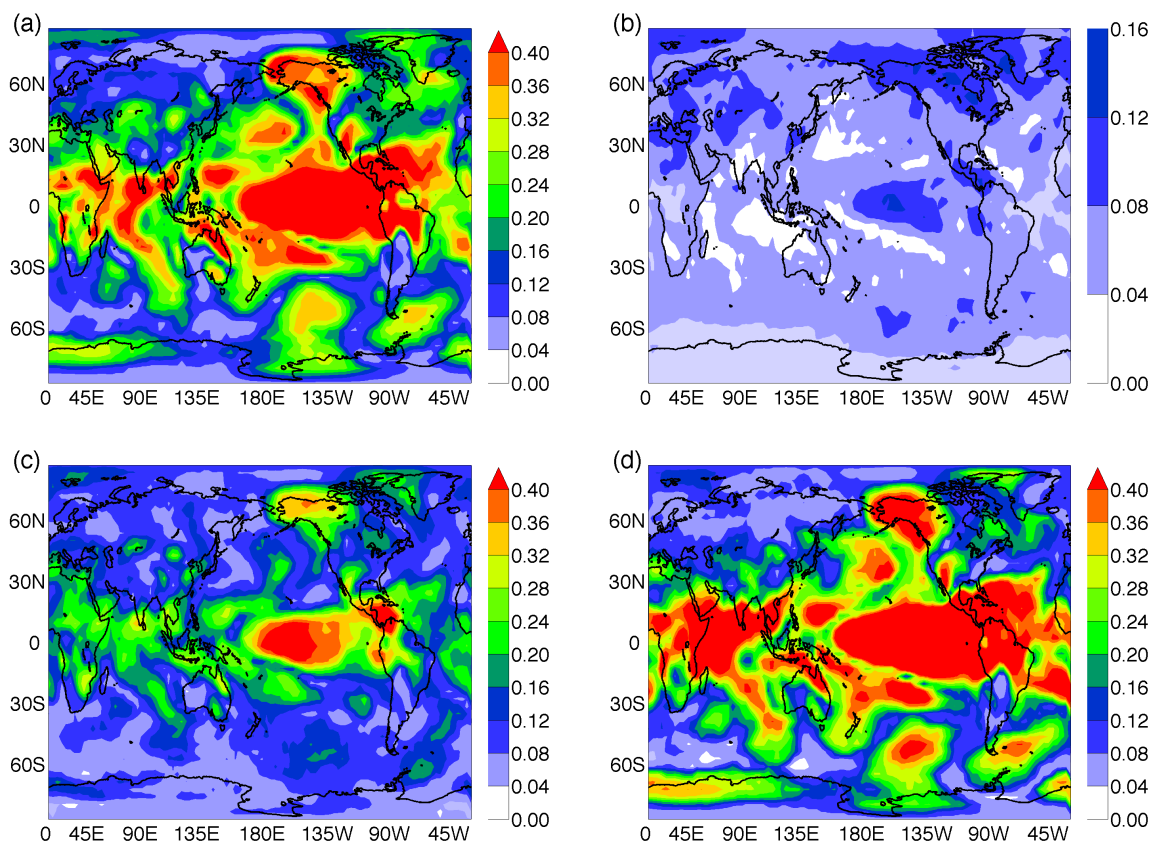


Figure 5.3: Maps of AWC computed from averaged time series, and thus containing information only of the forced component of atmospheric variability. The quantifiers of statistical similarity are as in Fig. 5.2: (a) MIH, (b) MIOP intraseasonal, (c) intra-annual and (d) inter-annual. It can be noticed that in the shorter time scale the tropical area, especially the Pacific ocean has a weak influence, and it grows stronger with increasing time scale. The fact that the maps in panels (a) and (d) are similar suggests that most of the links uncovered by the MIH, panel (a), actually reflect interdependencies in the longer time scale and thus, are seen in panel (d).

Niño. However, a few small well-connected areas remain over the equatorial Pacific, indicating that a linear regression is not sufficient to fully eliminate the ENSO effect on the network connectivity.

The Caribbean and north Atlantic are the largest regions that maintain a similar AWC even after Niño has been removed. Note, however, that the instantaneous regression does not completely remove the ENSO signal if there is a lag in the response. This is so in the

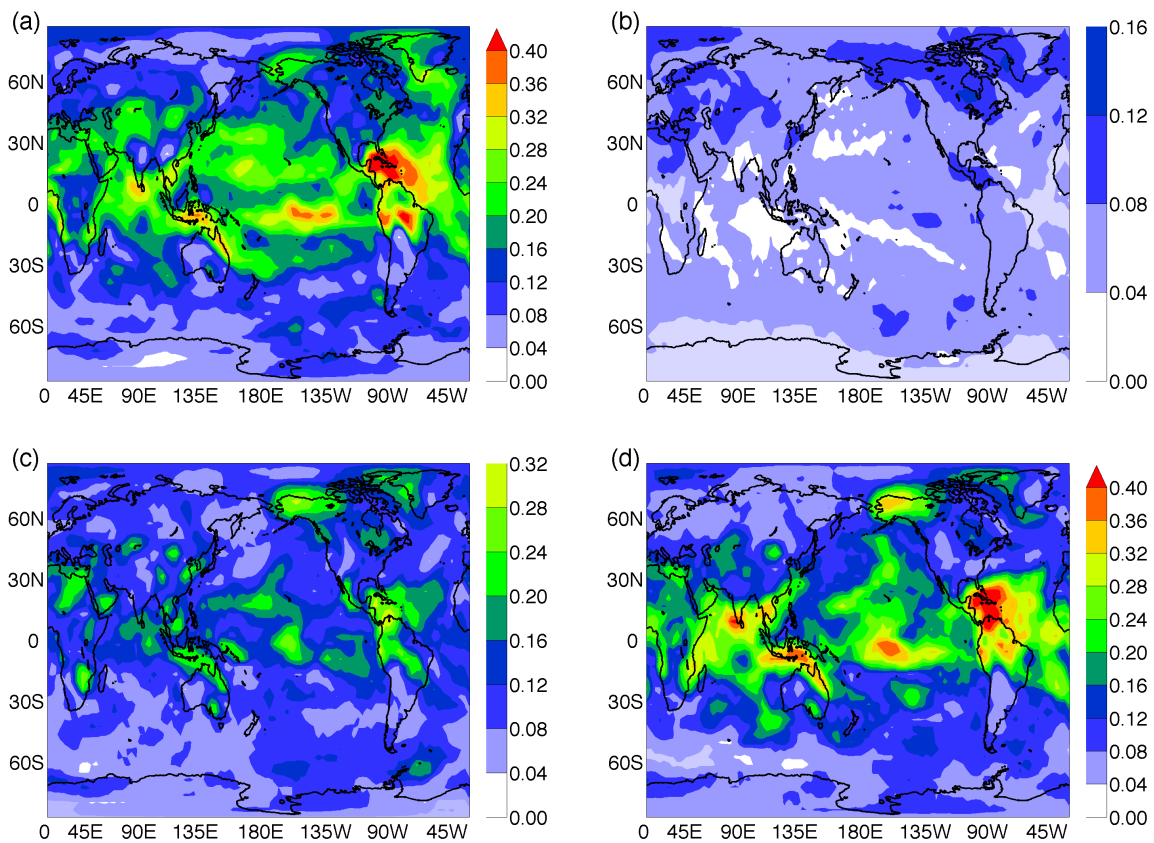


Figure 5.4: Maps of AWC of the forced component of the network when the ENSO influence is removed from the time series (for the description of the index and for the removal procedure, see sections 3.1.1 and 1.4.1). The statistical interdependencies are quantified as in Fig. 5.2: (a) MIH, MIOP (b) intraseasonal, (c) intra-annual and (d) interannual. A comparison with Fig. 5.3 allows assessing the influence of El Niño phenomenon over the network connectivity.

tropical north Atlantic [220], where El Niño affects sea surface temperature through heat flux changes that, given the ocean’s heat capacity, take a few months to induce an anomaly. Thus, this might be a reason for the still large connectivity observed in the Caribbean in Fig. 5.4(a).

Other areas, like over China and central Asia, which are weakly connected to the El Niño phenomenon show the same connectivity in Figs. 5.3 and 5.4. The fact that areas not related to ENSO do not change when removing the index hints that the statistical test used to fix the network density is robust and allows to compare maps with and without a particular phenomenon.

Panel 5.4(b) is very similar to panel 5.3(b) except on the absence of a connected (dark blue) area on the Pacific ocean, suggesting that the influence of El Niño at these time scales is very low and restricted to the tropical Pacific. At intra-annual time scales, panel 5.4(c) shows the disappearance of many links from the corresponding Fig. 5.3(c). This suggests that at this time scale, even if El Niño signal is not as strong as on interannual scales, it is already connecting far away tropical and extratropical areas as Alaska [198]. Thus, removing El Niño signal affects very heavily the connectivity of the network. For longer time scales – shown in panel 5.4(d) – the scenario is similar as for 5.4(a) with only a remnant of connectivity in the tropical region.

### 5.3.2 Internal variability

Figure 5.5 shows AWC maps of internal variability, computed by averaging the nine AWC maps obtained from the individual model runs, where in each time series, the forced signal (the average of the nine runs) was removed. Contrary to the forced variability case presented before, in this case the most connected areas are on the extratropics. This is consistent with results of previous figures and indicates that in the tropics the ocean forces the largest portion of atmospheric variability. As the tropical atmosphere cannot sustain horizontal gradients generated by SST anomalies, it induces vertical movements of air, convection and release of latent heat, thus giving rise to atmospheric circulation anomalies.

In the extratropics internal atmospheric variability is larger leading to stronger connections. The larger connectivity in the northern hemisphere suggests that the large landmasses affect atmospheric variability, which is consistent with our current understanding of storm track dynamics and low frequency transients [206].

The most connected spot in Fig. 5.5(a) is over the Labrador sea. The rest of the highly connected areas (in green) are present mostly in the northern hemisphere. In the southern hemisphere connectivity is largest over the Southern ocean. Investigation over this well connected area near Antarctica –only found using MIH to quantify interdependencies– showed that in this area histograms have a higher skewness than in the rest of the nodes, an effect that has also been reported and discussed in [182]. This effect is found on the internal-plus-forced AWC map of Fig. 5.2(a) and using reanalysis data as shown in panel Fig. 5.1(a). When considering other measures to quantify interdependencies, such as Pearson cross correlation or MIOP, the AWC maps do not show high connectivity in this region [130].

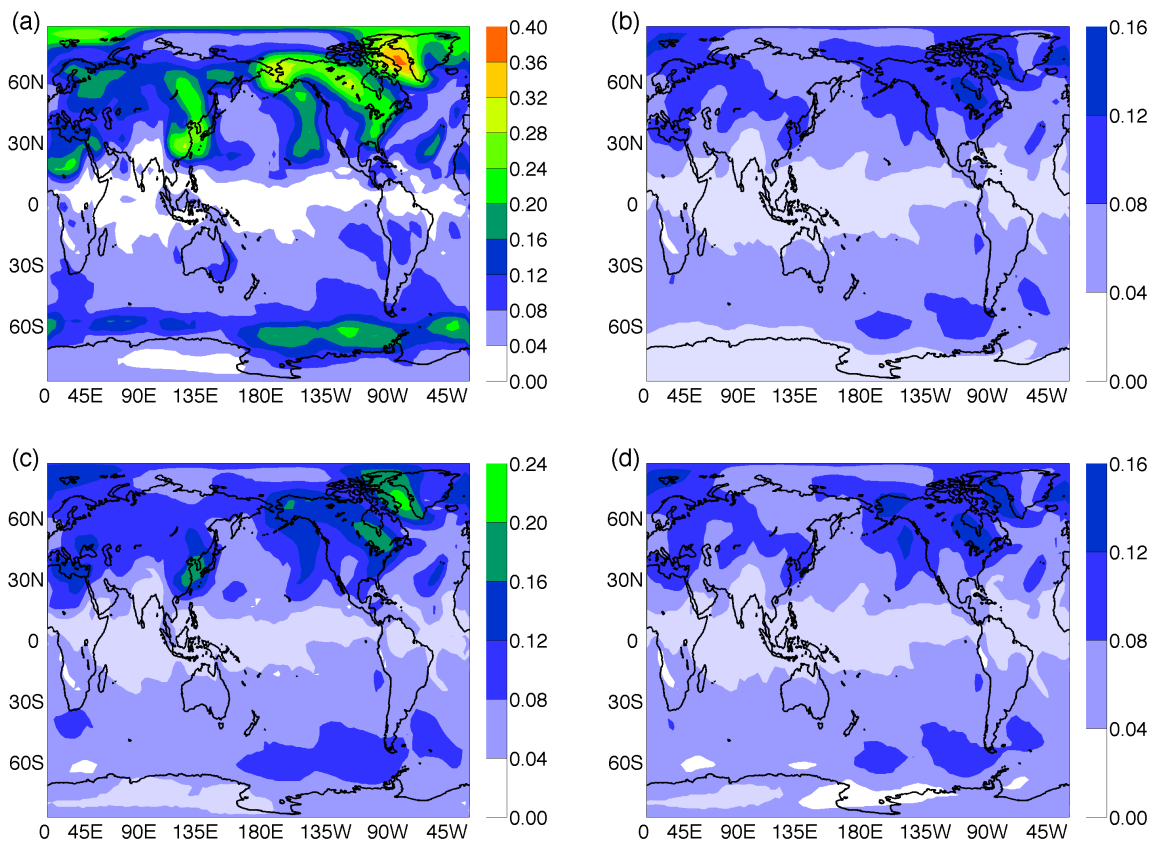


Figure 5.5: Maps of averaged AWC, revealing the internal variability network (see text for details). The statistical interdependencies are quantified as in Fig. 5.2 (a) MIH, MIOP (b) intraseasonal, (c) intra-annual and (d) interannual. It can be noticed that in this network the time scale showing more connectivity is the intra-annual time scale. This is consistent with the shorter memory of the atmosphere compared with the ocean.

With respect to the AWC maps computed using MIOP, in contrast to the forced case, the intraseasonal, intra-annual and interannual maps are very similar to each other. This is a sign of “multiscale variability”. i.e. variability distributed over many time scales. Internal variability cycles are less well defined, with spectra similar to “red” noise. It can be seen that the most connected AWC map is the intra-annual one, stronger than both the intraseasonal and the interannual, consistent with the fact that atmospheric anomalies are less persistent than oceanic ones [210, 221].

The fact that the most connected area in Fig. 5.5(a) is over the Labrador sea, suggests that it is related to NAO. In order to verify this, NAO was removed from the time series using the



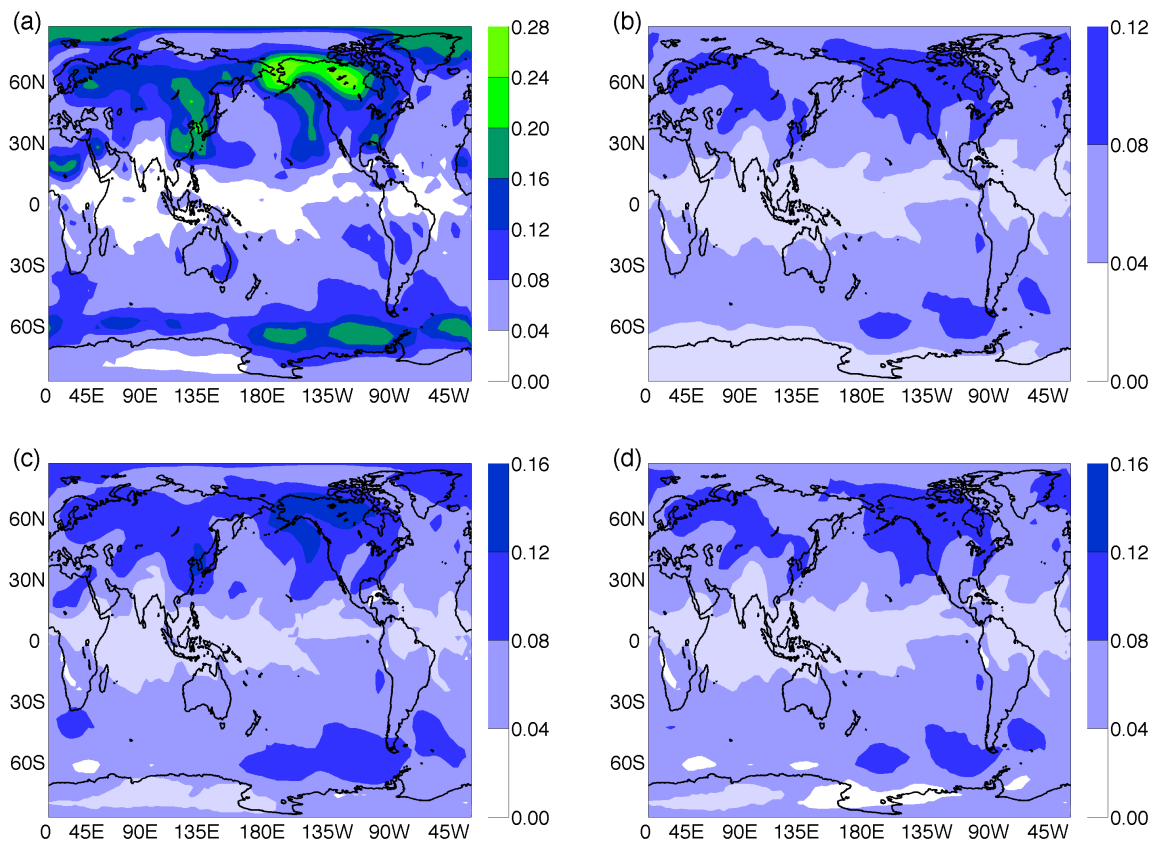


Figure 5.6: Maps of averaged AWC, revealing the internal variability network when the NAO index is removed from the time series (see sections 3.1.1 and 1.4.2 for details). The statistical interdependencies are quantified as in Fig. 5.2 (a) MIH, MIOP (b) intraseasonal, (c) intra-annual and (d) interannual. It can be noticed that in this network the time scale showing more connectivity is the intra-annual time scale. This is consistent with the shorter memory of the atmosphere compared with the ocean.

same procedure as with NINO3,4, explained above. The results are shown in Fig. 5.6. Here, indeed the Labrador connected area disappears in all the panels while the connectivity unrelated to NAO (i.e. over southern hemisphere or China) remains almost unchanged.

## 5.4 Node connectivity maps

AWC maps provide information of the connectivity of the geographical regions, but no information about the nature –spatial range or distribution– of the links. It is expected that nearby points behave similarly and this leads to high values of correlation between nearby

places [12, 140]. The distance over which the climate variables are well connected is related with the Rossby radius of deformation (RRD) [206], which is the distance that a particle or wave travels before being significantly affected by the Earth's rotation. Also, in the tropics, this proximity effect can be greatly enhanced as there the information is propagated very fast longitudinally. Here the goal is to unveil the presence of teleconnections, that is, connections between regions separated more than the RRD.

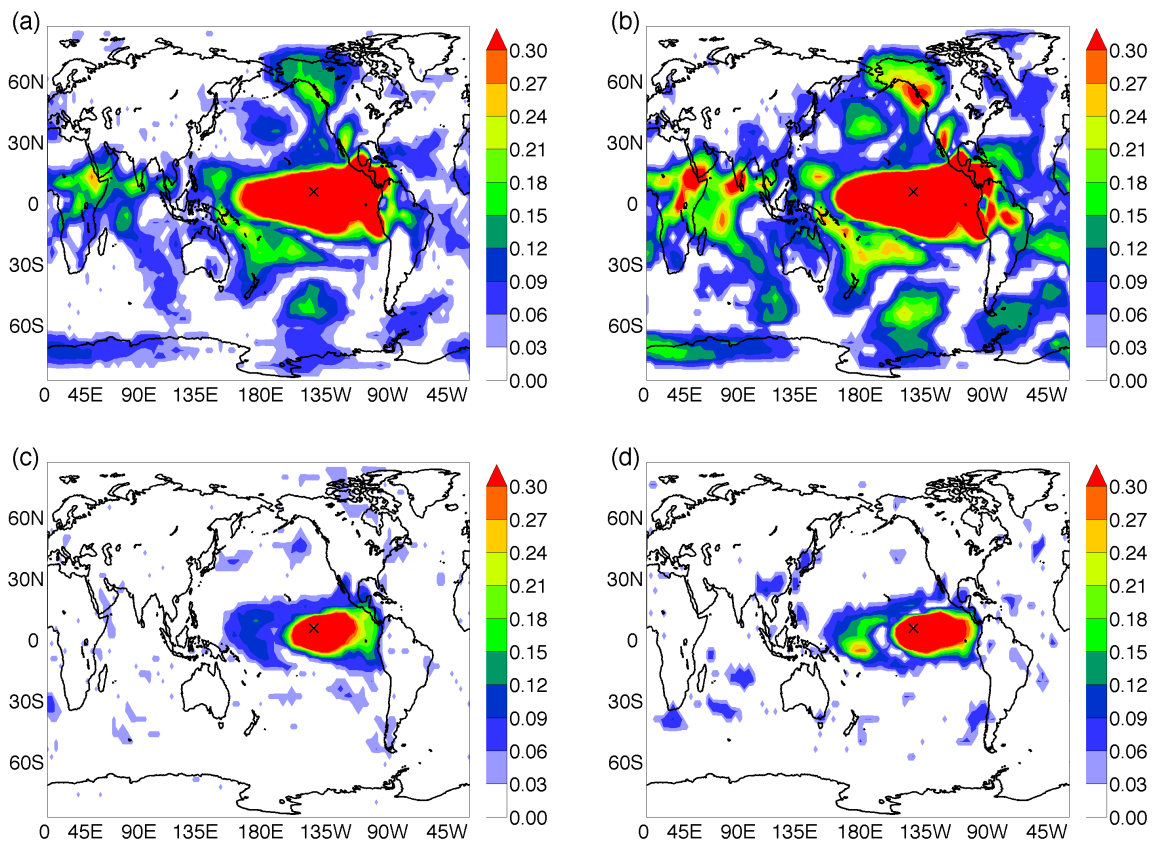


Figure 5.7: Connectivity map of a node in central Pacific (indicated with  $\times$ ). Panels (a) and (b) are computed from forced time series (averaging over nine model realizations); panels (c) and (d) are computed also from forced time series, but with ENSO3.4 linearly removed and thus not containing –to the first order– contributions due to El Niño . In (a), (c) interdependencies are quantified via MIH; in (b), (d) via MIOP interannual time scale.

Plots 5.7, 5.8, 5.9, and 5.10 show the connections of a node, indicated with “ $\times$ ”. Figures display MIH in the left column and MIOP in the right column. The time scale of the MIOP maps is interannual for the forced variability network and intra-annual for the internal vari-

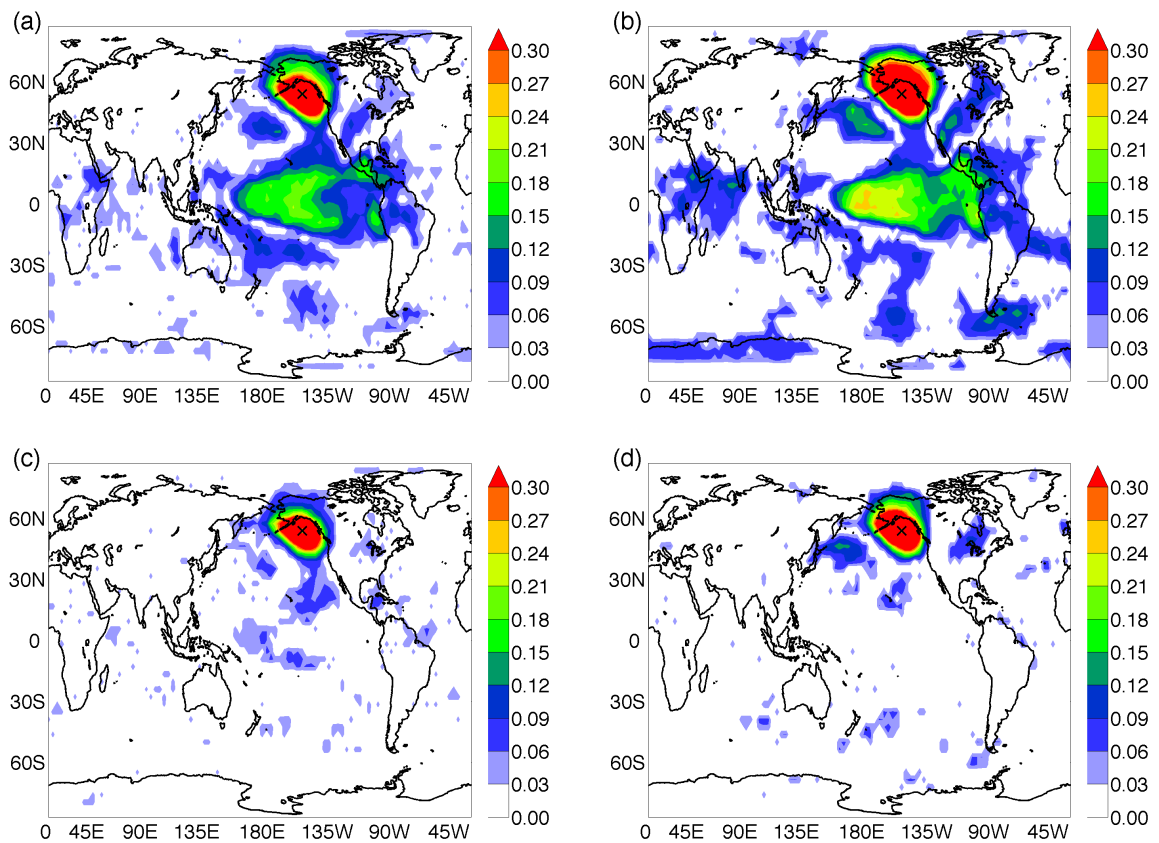


Figure 5.8: As Fig. 5.7 but considering a node near Alaska (indicated with  $\times$ ). Comparing with Fig. 5.7 one can notice that the teleconnection between this region and the Pacific is due mainly to El Niño.

ability network, following above results. Since there is a focus on unveiling weak but long range significant links, the color scale for nearby links was saturated. In this way it is possible to see the weak links with good resolution, losing information for the stronger links (stronger than 0.3 on the arbitrary scale of MI, where the highest links have values of 1 or 2 on the same scale, as shown in the previous chapter) which will be all represented with the same color.

#### 5.4.1 Forced variability

Figure 5.7 shows the connections of a point in the central Pacific ocean in the forced variability network. given the similarity of panels 5.7(a) (which contains all time scales) and 5.7(b) (which contain information for the interannual time scale only) it can be inferred that most

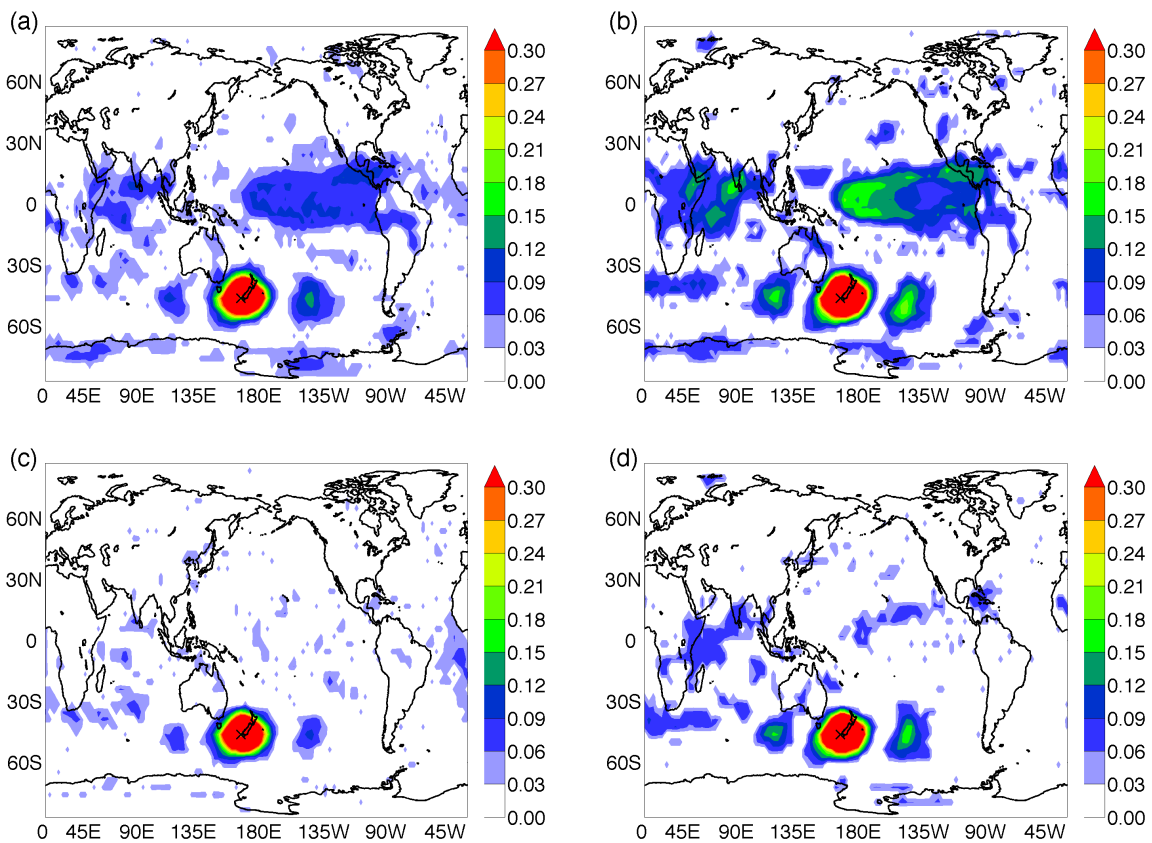


Figure 5.9: As Fig. 5.7 but of a node near New Zealand (indicated with  $\times$ ). In panels (b) and (d) the MIOP is adjusted to interannual time scale. Compare with Figs 5.7 and 5.8.

of the links are links reflect behavior in the interannual time scale.

Panels 5.7(c) and 5.7(d) display the same node connectivity maps as on 5.7(a) and 5.7(b) respectively, however, in this case the NINO3.4 index has been removed from the time series and thus (to first order) they do not contain links due to El Niño phenomenon. The differences between panels 5.7(a) and 5.7(c) and between 5.7(b) and 5.7(d) are evident. First, after eliminating the effects of El Niño the tropical and extratropical teleconnection patterns associated to the spot in the Pacific disappear independently of the methodology used to quantify interdependencies (MIH or MIOP): the connectivity becomes restricted to the tropical Pacific basin. Even inside this region the connectivity is greatly decreased as seen by a much smaller red spot of links over 0.3, although the remaining connections indicate that, either a linear regression is not enough to fully remove the influence of El Niño, or the ENSO dynamics is not fully represented by the NINO3.4 index.

According to panels (a) and (b) of Fig. 5.7, Alaska is an area well connected to the equatorial Pacific ocean. To further investigate, Fig. 5.8 shows global connections to a point nearby Alaska. It can be seen in panels 5.8(a) and 5.8(b) that it indeed presents connections to the equatorial Pacific ocean with a maximum close to the dateline.

Furthermore, connections to the southern Pacific ocean, Central Africa, Indian ocean and even the Drake passage are found. These connections are stronger in panel (b) especially those linking Alaska with the Indian and southern Atlantic ocean and Drake Passage. If NINO3.4 is removed, a dramatic change in the maps is found. Connections become almost local and all the north - south teleconnections are lost; only connections probably associated with an imperfect removal of the El Niño signal remain. This indicates that there are no direct teleconnections between Alaska and (for example) the Drake Passage, but both are strongly connected to El Niño. As these networks are constructed using symmetrical measures of dependency, calculated directly from the data, they are unable to distinguish between a direct connection and an indirect one. In the next chapter a methodology to construct directed networks is explored.

Figure 5.9 is as Figs. 5.7 and 5.8, but for a node in the southern hemisphere extratropics. We chose southern New Zealand because it shows a relatively high forced density [seen in Fig. 5.3 (a,d)] and it is connected to the selected point over the tropical Pacific of Fig. 5.7 (a,b). Panel 5.9(a) shows connectivity between the chosen point and the Pacific and Indian oceans, as well as wave patterns (probably a Rossby wavetrain) along the extratropics. Figure 5.9(b) adds information to 5.9(a) showing that these teleconnections are of interannual type. If NINO3.4 is removed (panels 5.9(c) and 5.9(d)) not surprisingly the links to the tropical Pacific disappear, but also some of the connectivity to the Indian ocean suggesting that part of the links with the Indian ocean are indirect. Nevertheless, the extra-tropical wavetrain remains, and Fig. 5.9(d) suggests that the wave train may be forced by the Indian ocean at interannual time scales. As in the previous figure, some weak north-south teleconnections are found, but they disappear if NINO3.4 index is removed, indicating again an indirect connection between the extratropics through the Pacific ocean.

## 5.4.2 Internal variability

Figure 5.10 displays the *internal* variability connections of a node over the most connected area of Fig. 5.5. The average of the resulting nine different connectivity maps is shown. In

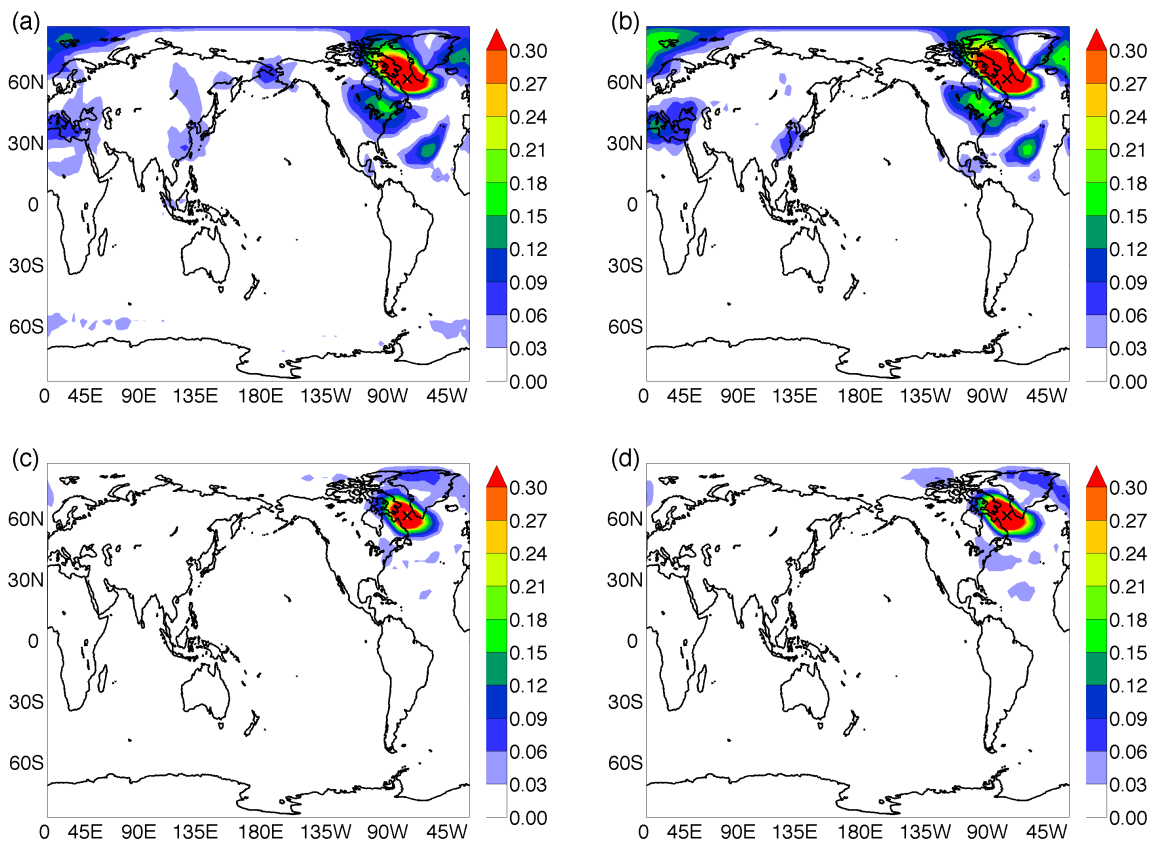


Figure 5.10: Maps of internal variability showing the connectivity of a node in the Labrador Sea (indicated with  $\times$ ). Panel (a), (b) correspond to the original internal time series as in Fig. 5.5; in panels (c), (d) the NAO was linearly removed and thus the links do not contain –to the first order– contributions due to the North Atlantic Oscillation. In (a), (c) interdependencies are quantified via MIH; in (b), (d) via MIOP intra-annual time scale.

the left column the connectivity computed using MIH is displayed, while in the right column, the intra-annual scale is shown, using MIOP. This time scale shows the strongest response for internal variability. In Fig. 5.10(a) the original internal variability connections are shown, revealing teleconnections extending over the northern hemisphere, especially over Scandinavia, Mediterranean Europe, east coast of North America and tropical north Atlantic. Figure 5.10(a) also shows connections to eastern China and the Aleutian islands. The pattern shown in Fig. 5.10(b) mainly corresponds to the known influence of the North Atlantic Oscillation. This is further substantiated in panels (c) and (d) of the same figure, where the NAO influence is removed and the connections of the Labrador sea, particularly in the northern

Atlantic basin, are strongly weakened. As the NAO indices are mostly internal and do not depend of the ocean forcing, they were calculated as the Principal component (PC) related to the leading EOF mode for each different model run, using geopotential height (GH) data. Afterwards, each index was removed from the corresponding SAT data and the internal climate networks were constructed as before.

## 5.5 Summary and conclusions

The monthly variability of the surface air temperature field has been decomposed into a part forced by the ocean temperature, and another due to internal atmospheric variability. This has been performed using an ensemble of nine AGCM runs forced with the same SST data, and starting from slightly different initial conditions. The model data was firstly validated by observing a qualitative agreement between the networks constructed from one model run and those constructed from reanalysis data. Afterwards, climate networks were constructed from model data, for the forced and for the internal variability components, using Mutual Information to assess the interdependencies between the time series. Ordinal patterns have been used in order to separate and determine the strength of the links at different time scales.

While the main conclusions of this analysis (the connectivity of the forced variability network is heavily affected by El Niño, whereas that of the internal variability network is significantly affected by the NAO) are not new, new information has been uncovered as ordinal analysis allows to study these phenomena on different time scales. This has revealed that most of the links detected in the forced variability proceed from long time scales, while the contributions of intra-annual time scales to the internal variability are the most important. The results of this chapter also open the possibility of studying how various network measures, such as the average path length, assortativity, clustering coefficient, betweenness, etc. depend on the time scale considered for quantifying statistical interdependencies.

Another conclusion of this chapter is that forced and internal atmospheric variability are characterized by very different networks. Because the separation of internal and forced variability done here requires averaging over several model runs, the networks obtained here could not have been obtained from observational/reanalysis data only. It is shown that the forced variability is stronger in the tropics, while the internal variability peaks in the mid latitudes. The network of forced variability has the strongest connections at interannual time

scales. *Long range teleconnections* from the tropics to the extratropics and even from different hemispheres in the forced network were observed and explained by the influence of El Niño. On the other hand, the network of internal atmospheric variability has the strongest connections in the extratropics, and it was found that connections to the Labrador sea are heavily affected by the North Atlantic Oscillation. To the best of our knowledge this is the first time this effect is observed in terms of network connectivity.

This study is focused on the lowest levels of the atmosphere. A complementary analysis was performed by Arizmendi *et. al.* [205], devoted to the study of the evolution of the upper atmosphere during the 20th century and aiming at distinguishing the oceanically forced component from the atmospheric internal variability on different time scales. The methodology proposed here for distinguishing links in spatial range (short and long), time scale (intra-seasonal, intra-annual and interannual) and type of variability (forced vs. internal) is a novel approach for the study of climate networks that provides new insight into the climatological meaning of the links found and their connection to physical phenomena.



## Chapter

# 6

## ***Directionality of Climate Interactions***

This chapter explores a natural extension of the previous research. Information-theoretic tools are used to construct *directed* climate networks from time-series of observed climatological data. Specifically, surface air temperature anomalies are considered. Two datasets—one monthly-averaged and another daily-averaged—are used. Directed links between network nodes are defined via an analysis of the net direction of information transfer. A predictability measure—based on conditional mutual information and explained in section 3.2.3—that quantifies the amount of information in a time-series  $x(t)$ , contained in  $\tau$  time units in the past of another time series,  $y(t)$ , is used. The resulting analysis of the directed network yields full agreement with state-of-the-art knowledge in climate phenomena, validating this methodology for inferring the net directionality of climate interactions, directly from the data. No assumptions are made, except for the appropriate setting of the parameter  $\tau$  which is sensitive to the shorter or longer auto-correlation of the time series.

As a first stage, we calculate link directionality, using monthly-averaged SAT data in order to compare with the undirected networks reported in the previous chapters. Afterwards, the results of the analysis of the daily-averaged datasets are presented. They are consistent with those obtained from monthly-mean data, furthermore uncovering additional patterns of atmospheric variability, not observed in monthly-mean data as the averaging procedure filters out high frequency and variability.

## 6.1 Statistical significance analysis

To address the significance of the values of the directionality index (DI) employed, 100 surrogates were generated using the bootstrap (BS) algorithm [197]. The null hypothesis for DI is that there is no preferential direction of the interaction. Also, as before, the null hypothesis considered for MI is that the processes are independent from each other.

The BS algorithm randomly resamples with replacement from the original datasets using blocks of data of approximately the size of the autocorrelation time of the time series, and then computes the estimators (MI and DI) from the resampled data. In this way, for each link, two different empirical distributions are obtained (for MI and for DI), from which significance thresholds are extracted.

Afterwards, for each link, the MI value was calculated from the original datasets and a significance test was applied. A MI value is considered significant if:  $MI > \Theta_{MI}$ , where  $\Theta_{MI}$  is the threshold derived from the bootstrap MI distribution. In addition—for the MI significant links only—the DI value computed from the original dataset was compared to the DI upper threshold  $\Theta_{DI}^+$  and the lower threshold  $\Theta_{DI}^-$ . The DI value was considered significant if  $DI > \Theta_{DI}^+$  or if  $DI < \Theta_{DI}^-$ . This two-step significance test was performed with the scope of assessing firstly the presence of a link, and afterwards its directionality.

The significance thresholds,  $\Theta_{MI}$ ,  $\Theta_{DI}^+$  and  $\Theta_{DI}^-$ , extracted from the BS surrogates, were computed as  $\Theta_{MI} = \mu_{MI} + 3 \times \sigma_{MI}$ ,  $\Theta_{DI}^+ = \mu_{DI} + 3 \times \sigma_{DI}$ , and  $\Theta_{DI}^- = \mu_{DI} - 3 \times \sigma_{DI}$ , with  $\mu$  and  $\sigma$  being the mean and standard deviation of the corresponding MI and DI bootstrap distributions. See section 3.4.1 for a full explanation of the surrogate procedure.

A graphical explanation of the full procedure is shown in Fig. 6.1 where, for two nodes (one over the Pacific and one over the Indian ocean) the unfiltered DI maps are displayed in panels (a,d), the significant MI values are displayed in panels (b,e) and finally the DI significant values for the MI significant links are displayed in panels (c,f). The DI maps show positive DI values in red, which mean outgoing links, while the incoming links are shown in blue.

In Fig. 6.2, panels (a) and (b) show that high MI values do not imply high DI values. In Fig. 6.2 (b) one can notice that most of the blue dots are located in a narrow range of MI values while they are distributed in terms of DI values. An inspection of panel (f) in Fig. 6.1 shows

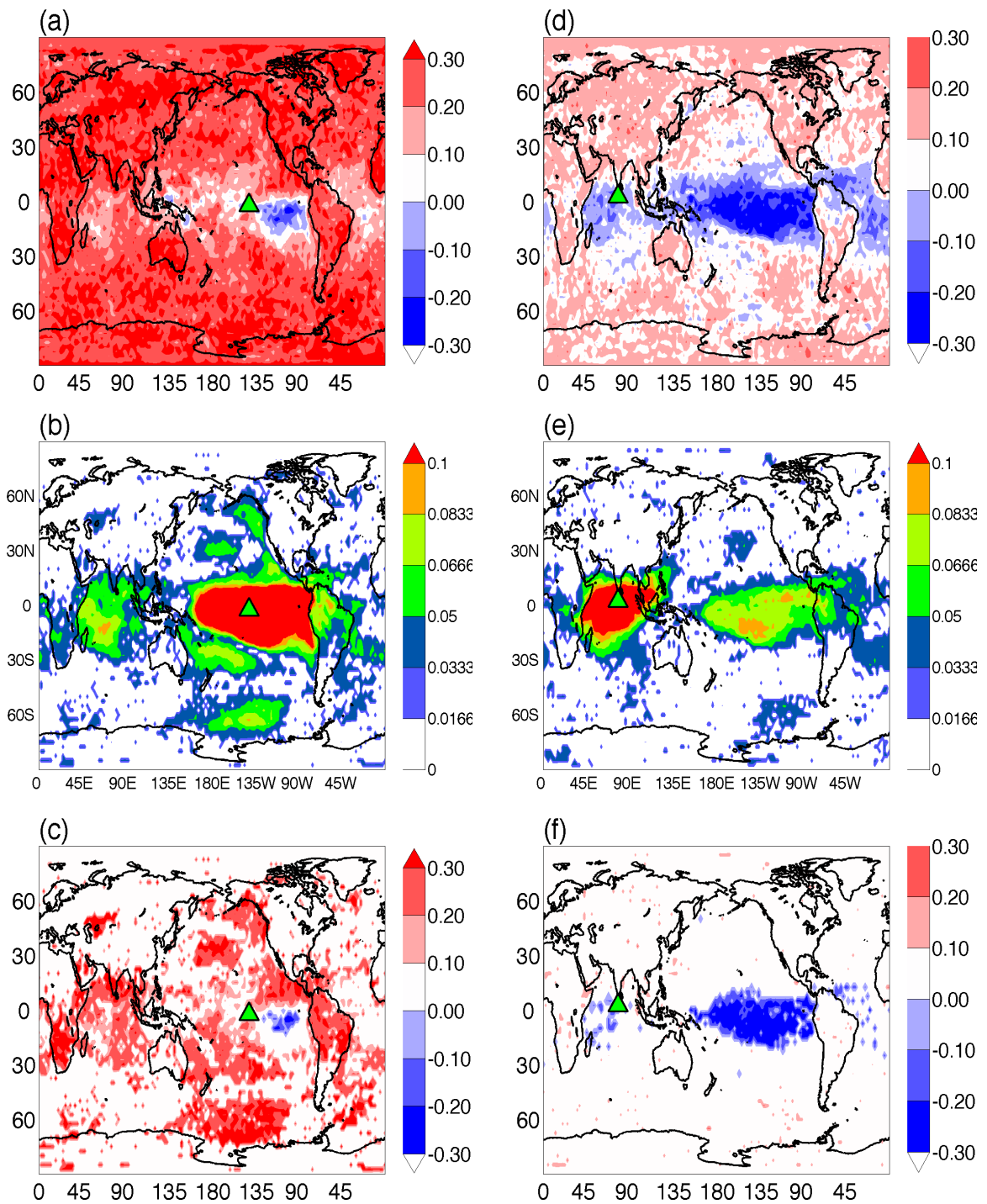


Figure 6.1: Procedure of constructing significant directionality maps from raw DI calculations. In panels (a) and (d) the unfiltered DI maps are shown for two nodes, one in the central Pacific ocean and one in the Indian ocean, indicated with triangles. Incoming links are shown in blue while outgoing links are in red. In panels (b) and (e) only the statistically significant MI values are shown. These results are combined in (c) and (f) where only the links that have both, statistically significant MI and DI values are shown using the same color code as in panels (a) and (d).

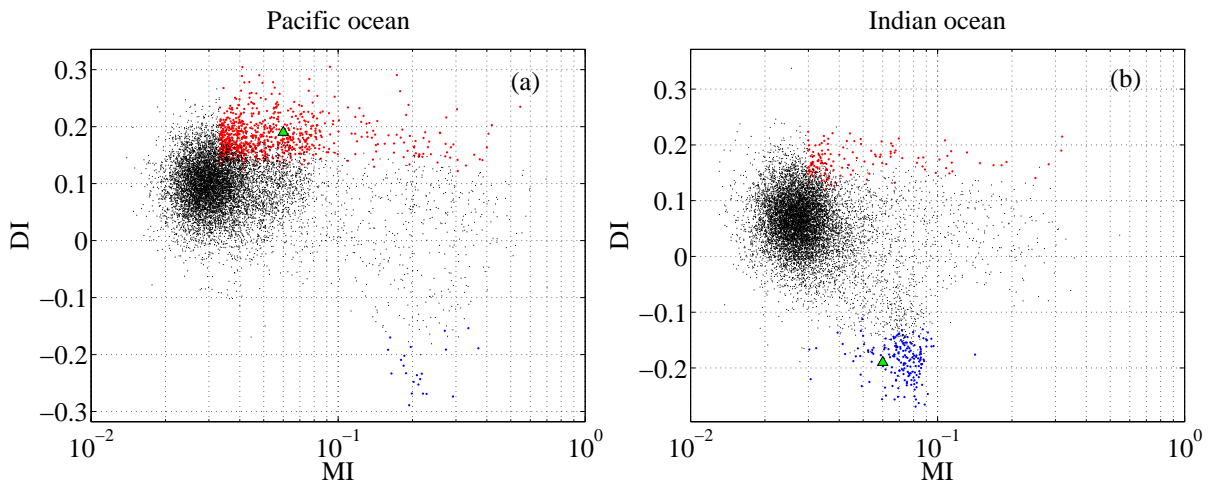


Figure 6.2: (a), (b) Plot of DI ( $\tau = 1$  month) *vs.* MI for all the links of the two nodes in the Pacific and Indian oceans considered in Fig. 6.1. Incoming links are indicated in blue and outgoing links, in red. Significant links are plotted in red or blue dots; the black dots indicate the disregarded links (either because MI is not statistically significant, or because MI is significant but DI is not). The triangles indicate the (MI, DI) values of the particular link between the two nodes.

that the blue links related to the node in the Indian Ocean are from a well-defined region in the central Pacific Ocean. On the other hand, one can notice in Fig. 6.2 (b) that the few red dots are more distributed in the MI, DI plane, consistent with Fig. 6.1 (f) the red outgoing links connect the node in the Indian Ocean to various regions on Earth.

## 6.2 Analysis of monthly-averaged SAT anomalies

### 6.2.1 Influence of the parameter $\tau$

As stated in the methodology the correct choice of the value of  $\tau$  is necessary for obtaining consistent results. As  $\tau$  can be only an integer, for monthly averaged data its minimum value is one month. In the tropical areas the influence of the ocean on the surface air temperature is a dominant characteristic. Moreover, because of the large heat capacity of water and the ocean's dynamics, the sea surface temperature anomalies vary in the scale of months. Calculating DI for a point in the central pacific (NINO3 area) for different values of  $\tau$  yields the results shown in Fig. 6.3. The point considered is the same as in Fig. 6.1 (a-c) and moreover,

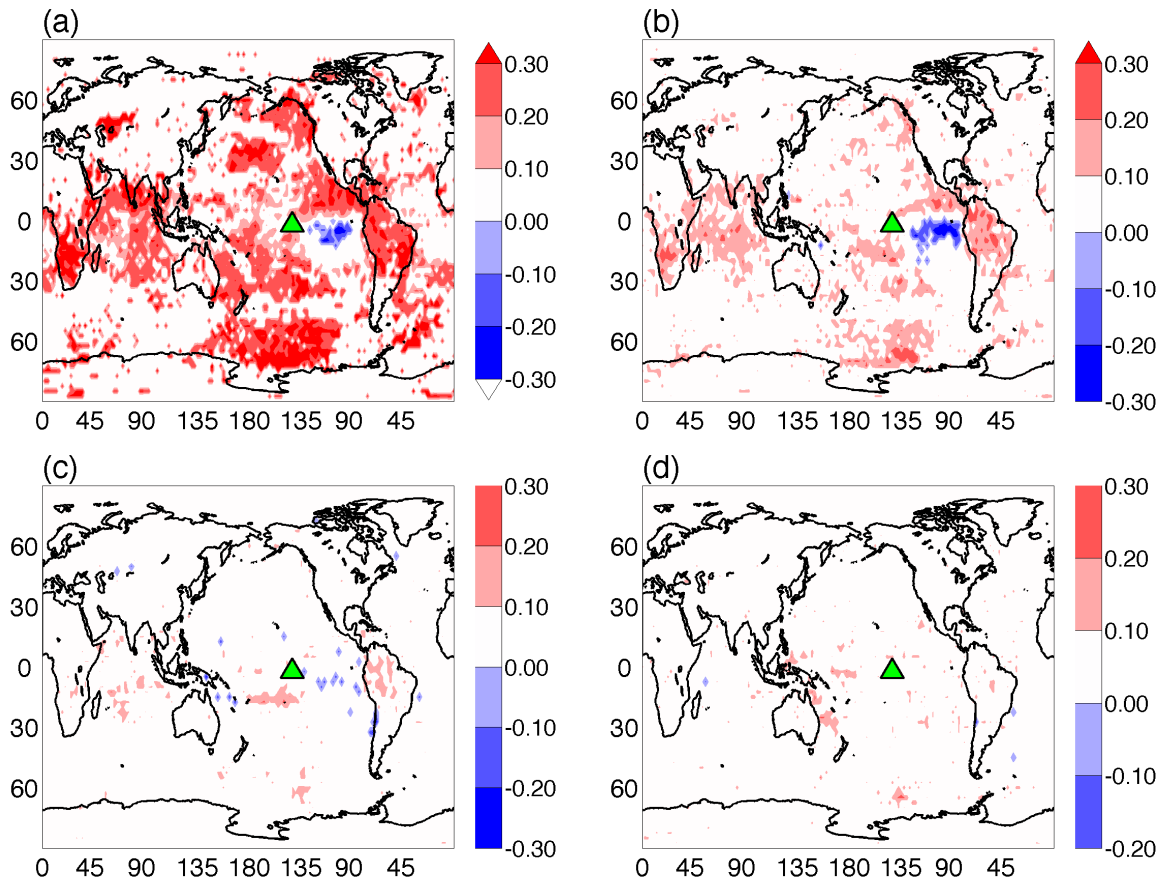


Figure 6.3: Effect of  $\tau$  on tropical areas using monthly-averaged data. In this case a point in central pacific (the same point used in the top row of Fig. 6.1 and in [130] is considered. The values of  $\tau$  are: (a) 1 month, (b) 3 months, (c) 6 months, (d) 12 months. Notice the decorrelation of the time series for large  $\tau$ . Incoming links are in blue while outgoing links are in red.

the panel 6.1 (c) is the same as panel 6.3 (a).

For  $\tau = 1$  Fig.6.3 (a) shows the central Pacific influenced by (in blue) the eastern equatorial Pacific and influencing (in red), presumably through atmospheric teleconnections, the global tropics and the extratropical Pacific ocean. However, as  $\tau$  grows the number of significant connections decreases, suggesting that the time-scale of decorrelation of the SAT is less than 6 months. This is consistent with the persistence time scale of 3 to 6 months of observed sea surface temperature.

As shown in chapter 5, the extratropical atmosphere shows larger internal variability than the tropics and the impact of the extratropical SST on the atmosphere is much more limited

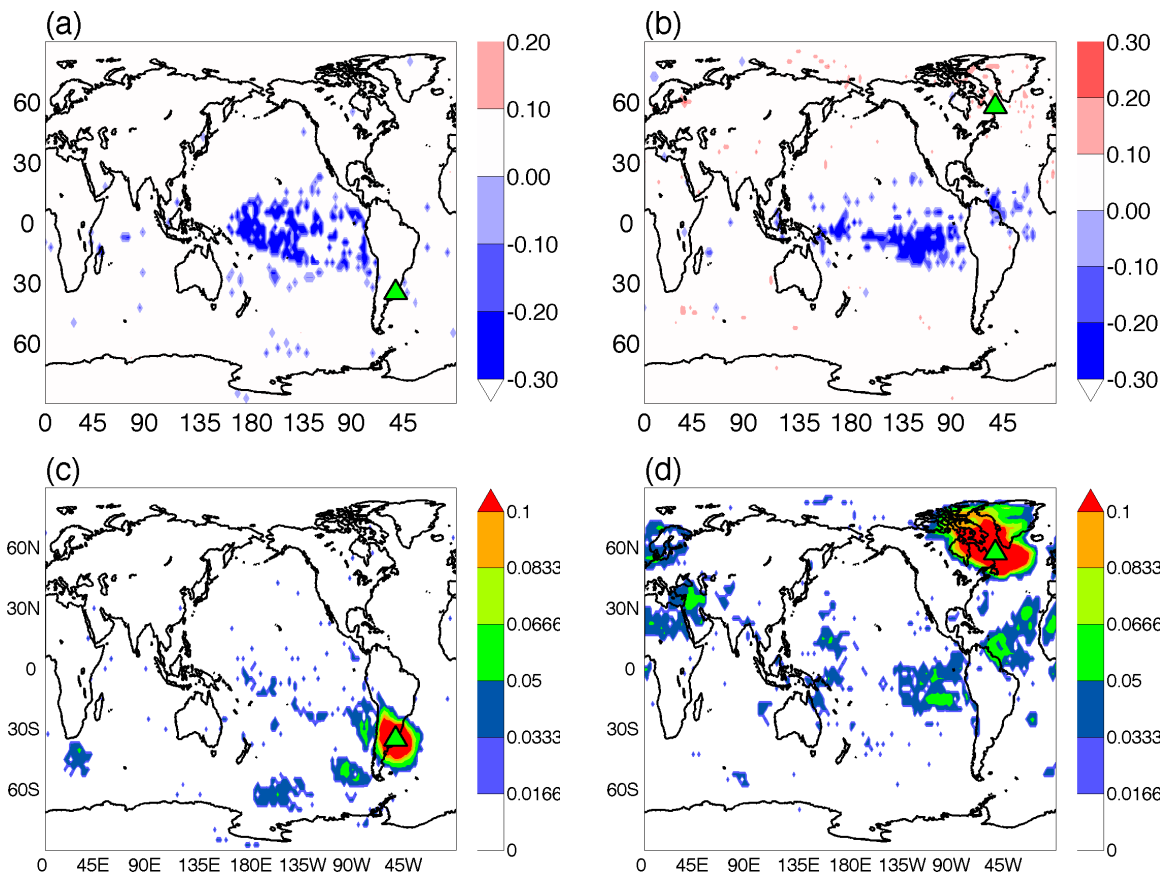


Figure 6.4: (a), (b) Directionality of the significant links ( $\tau = 1$  month) of two nodes in the extratropics indicated with triangles: (a) in southern South America (de la Plata basin) and (b) in Labrador Sea. As the decorrelation time on the extratropics is very fast, no clear structures are seen. In order to show that it is not a problem of statistics, panels (c) and (d) display the significant MI values. Incoming links are in blue while outgoing links are in red.

than in the tropics. Thus, the variability of extratropical SAT is dominated by synoptic atmospheric dynamics and has time scales of a few days. Longer persistence time scales might appear in the extratropics if the region is influenced by tropical SST. This motivates the use of a small value of  $\tau$  when considering extra-tropical variability.

In Fig. 6.4 the DI and MI maps for two points in the extratropics are shown. Panels 6.4 (a,c) show links related to a point in southeastern South America, while panels 6.4 (b,d) show links related to a point in the Labrador sea, whose characteristics are linked to the North Atlantic Oscillation (NAO) as shown in Chapter 5. The top panels show DI for  $\tau = 1$  month while the bottom panels show MI.

Consistent with the previous description, the extratropical SAT show only some incoming links from the tropical region for  $\tau = 1$  month. The point over the Labrador Sea seems to show also some outgoing links to the northeast, although there is no clear structure. As will be shown in Sec. 6.3, a better identification of link directionality is obtained by using daily data.

### 6.2.2 Comparison between monthly and daily datasets

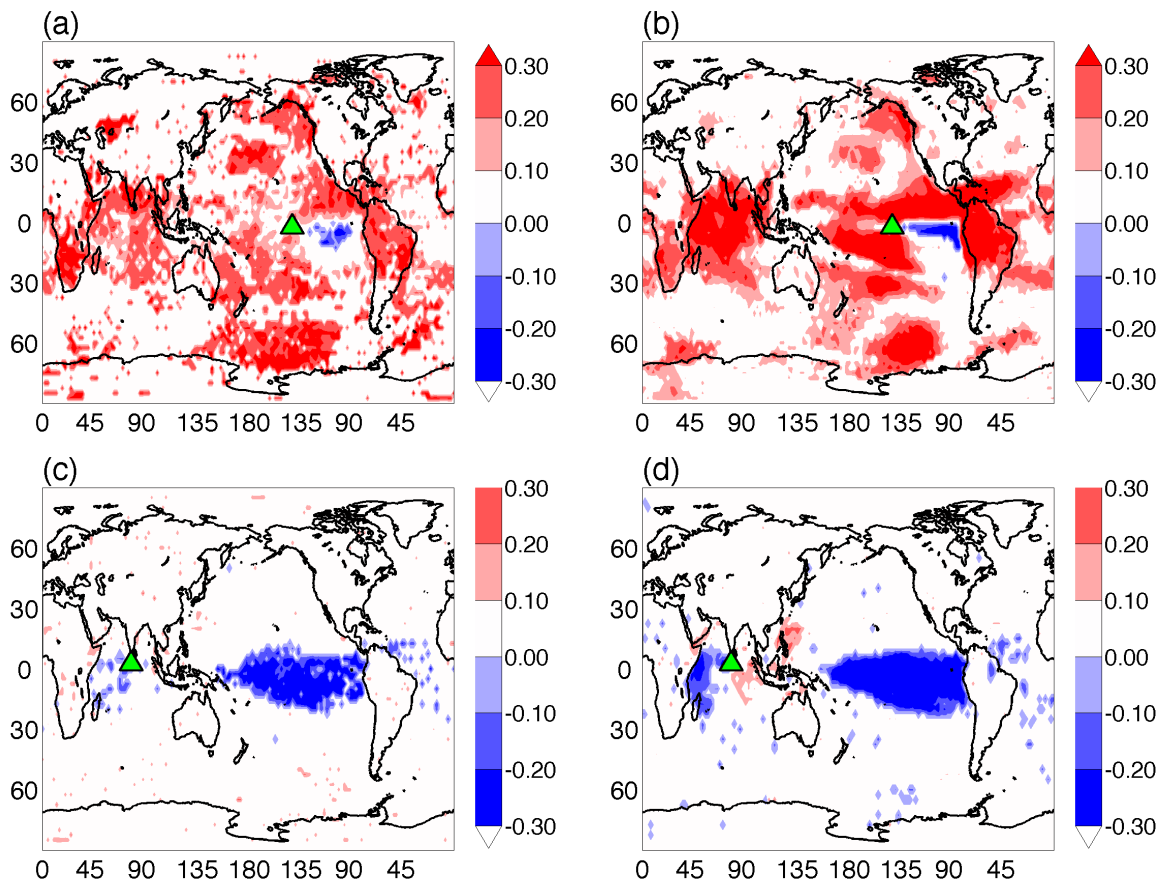


Figure 6.5: Comparison of results for monthly-averaged and daily-averaged datasets. The nodes considered are as in Fig. 6.1. Panels (a) and (c) display the DI measure ( $\tau = 1$  month) calculated from monthly data—same as Figs. 6.1 (c) and (f). Panels (b) and (d) display the DI measure ( $\tau = 30$  days) calculated from daily data. Results are consistent and the resolution using daily data is better.

In order to obtain more temporal resolution daily averaged data has been used. Figure 6.5 shows a comparison between DI calculated for monthly-mean data—panels 6.5 (a) and

6.5 (c)—and using daily-mean data—panels 6.5 (b) and 6.5 (d), corresponding to the same points considered in Fig. 6.1. In order to adequately compare the datasets,  $\tau$  was chosen equal to one month in the monthly-mean data and 30 days in the daily-mean data.

The maps using monthly- and daily-mean data show similar features and no inconsistencies are found. Nevertheless, the map constructed using daily-mean data captures much better the local and remote dependencies and directionality of the links. Areas with significant links are better defined and some regions that are known to be influenced by equatorial Pacific SST, like the tropical north Atlantic [220], clearly appear using daily-mean data, but only very roughly using monthly-mean data. Thus, the increase in temporal resolution improves the representation of the links related to tropical regions.

## 6.3 Analysis of daily-averaged SAT anomalies

### 6.3.1 Influence of the parameter $\tau$

To analyze the influence of the parameter  $\tau$  in the DI measure when it is computed from daily data, four links connecting the node in the Pacific ocean are considered in Fig. 6.6 with:

(a) a region in Labrador Sea (60.0 N, 52.5 W), (b) a region in tropical north Atlantic (25.0 N, 40.0 W), (c) a region in southern Pacific (5.0 N, 80.0 E) and, (d) a region in the Indian ocean (−22.5 N, 132.5 E). Figure 6.6 displays, for each link, the DI value vs  $\tau$ . In order to filter the noise, the DI value of nearby points—over a box of  $\pm 5^\circ$  latitude and longitude centered in the region of interest (second neighbors in the grid)—was calculated and averaged.

In panel (a), the directionality index that characterizes the influence of the equatorial Pacific to the Labrador Sea shows a small increase for small  $\tau$  that persists up to about  $\tau = 10$  days. This is the typical time scale associated with the setup of the atmospheric anomalies forced by anomalies in tropical convection. Afterwards, the DI decreases exponentially-like, becoming non-significant at  $\tau = 30$  days—as the DI value enters the shaded area. In the tropical north Atlantic—Fig. 6.6 (b)—the DI related to the equatorial Pacific shows similar values as for the Labrador Sea case, for up to  $\tau = 10$  days. For larger values of  $\tau$ , however, the DI has significant values (approximately constant) up to about  $\tau = 60$  days. Afterwards, it decreases becoming non-significant for  $\tau$  larger than 4 months. The difference in behavior between the tropical north Atlantic and the Labrador Sea might be because the remote forcing from ENSO



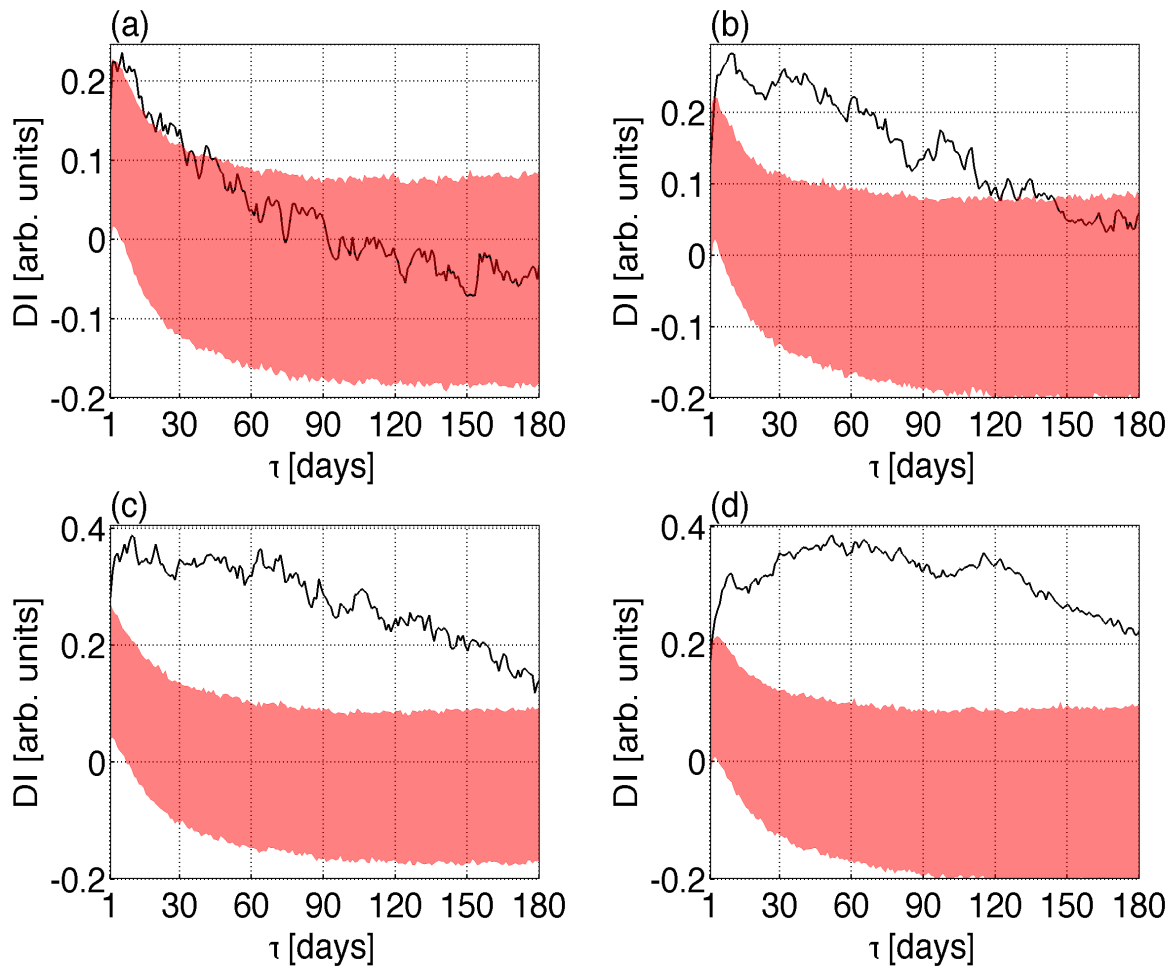


Figure 6.6: Influence of the parameter  $\tau$  when DI is computed from daily-averaged data. The DI value is plotted *vs.*  $\tau$  for the four links that connect a node in the Pacific ocean with (a) a node in Labrador sea (60.0 N, 52.5 W), (b) a node in tropical north Atlantic (25.0 N, 40.0 W), (c) a node in southern Pacific (5.0 N, 80.0 E) and (d) a node in the Indian ocean (-22.5 N, 132.5 E). In order to filter the noise the DI value was averaged over second neighbors in the grid. The shaded area (red online) indicates the DI value computed from 100 BS surrogates, as explained in section 6.1.

induces a clear regional response in the surface temperatures of the tropical Atlantic [220], which will add persistence to the remote signal. On the other hand, in the Labrador Sea, the ocean does not respond strongly and the large atmospheric variability obscures the signal from the equatorial Pacific.

A similar behavior to the one seen in the tropical Atlantic is also found in the Southern

ocean—Fig. 6.6 (c). There is a fast time scale for small values of  $\tau$ , but in this case the influence of the equatorial Pacific persists for  $\tau$  up to 80 days, where DI starts to decrease, with DI values significantly larger than in the Atlantic. The ENSO influence over the south Pacific is one of the most robust signals in the extratropics, consequence of atmospheric teleconnections associated with the Pacific South American pattern [222]. The time scale of about 3 months seen in the DI is likely associated with the time it takes to the surface ocean to respond to anomalous atmospheric fluxes and to the seasonal dependence of the atmospheric teleconnection pattern on the mean state of the extratropical atmosphere.

The behavior of the influence of the equatorial Pacific onto the Indian ocean—Fig. 6.6 (d)—also shows a fast time scale of a few days, but in this case the largest value of the DI is seen for a  $\tau$  of about 60 days. Also, there are large values of DI for  $\tau$  values of more than 4 months. This suggests that the Indian ocean responds to the incoming ENSO signal in a time scale of about 2 months, through thermodynamic and dynamic coupling [223, 224].

Thus, the DI plots as a function of  $\tau$  for different ocean regions suggest that there is one fast time scale of about 10 days associated to the setup of the atmospheric teleconnections associated with changes in tropical convection. Moreover, in some regions the DI shows a second longer time scale of about 2 or 3 months associated with the response of the local ocean to the circulation anomalies forced from the tropics. Finally, longer time scales related to oceanic inertia also affect the DI value but longer datasets are needed in order to perform a robust estimation of their effects.

### 6.3.2 Influence of $\tau$ in the extra-tropics

The improvement in characterizing directionality using daily-mean data is larger in the extratropics. As mentioned above, the extratropical SAT is strongly dependent on synoptic scale perturbations (a time scale of a few days). Thus, the use of daily-mean data should allow to uncover these relationships and investigate the direction of the links as the lag increases. In order to do so, the point in southeastern South America shown in Fig. 6.4 is considered and the directionality network for several values of  $\tau$  ranging from a day to one month (Fig. 6.7) is constructed. For synoptic time scales of a few days the methodology uncovers the existence of a wave train connected to southeastern South America propagating with a southwest-northeast direction. Moreover, there is a clear separation line between

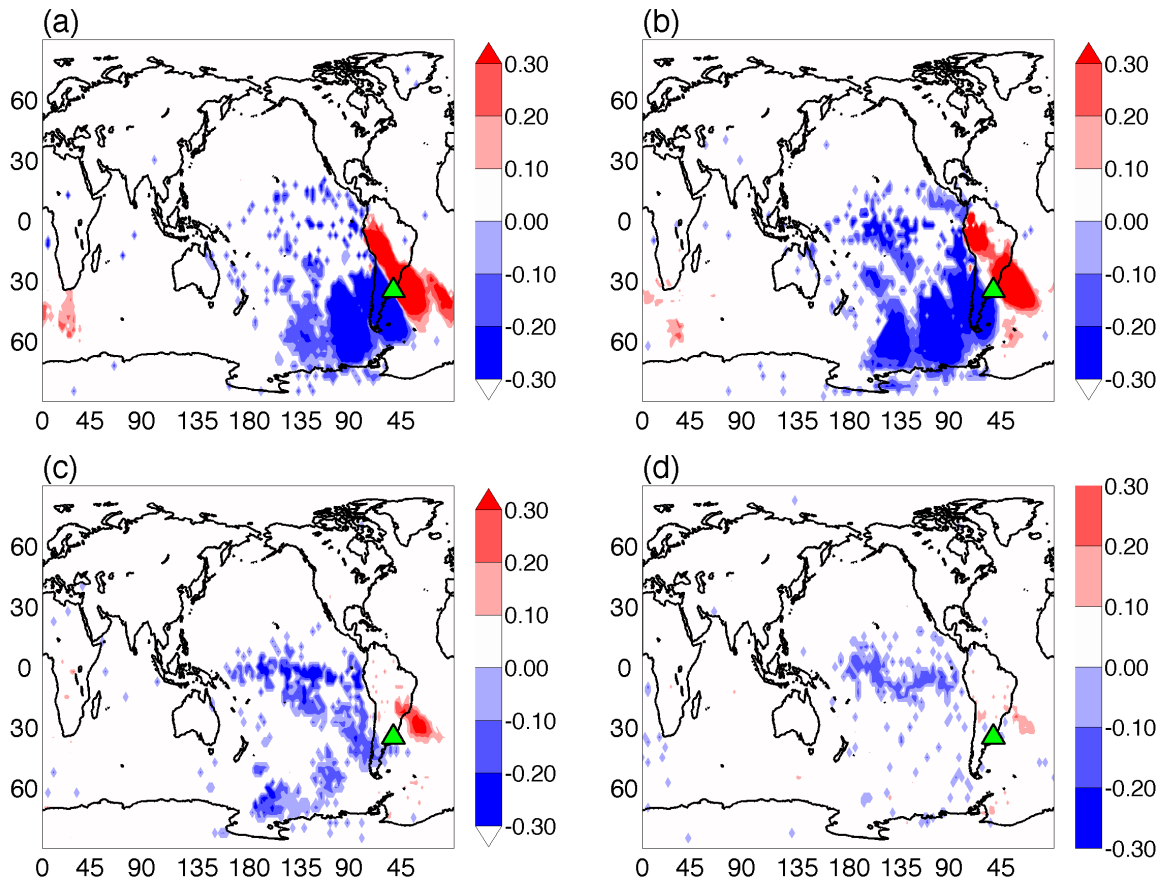


Figure 6.7: Effect of  $\tau$  in the southern extratropics, when DI is computed from daily data. The same node in Fig. 6.4 (a) is considered, and the values of  $\tau$  are: (a) 1 day, (b) 3 days, (c) 7 days, and (d) 30 days. Small  $\tau$  capture wave trains propagating from west to east, while for  $\tau$  values longer than a week the decorrelation is higher and only the influence from the Pacific ocean persists.

regions with incoming and outgoing links. This configuration is characteristic of the progression of a front through the reference point and does not imply that the SAT over the reference point influences the region to the northeast but it only happens to be in the path of the perturbation. As the lag time increases, the extratropical wave train associated with synoptic time scales fades and only the points in the tropics remain, consistent with an influence of the equatorial Pacific on the region on longer time scales, perhaps related to ENSO.

A similar behavior is seen taking as reference point the SAT over the Labrador Sea (Fig. 6.8). For small values of  $\tau$  the progression of a front is clearly detected using this procedure: given the mean westerly winds at these latitudes, the front moves from west to east and is

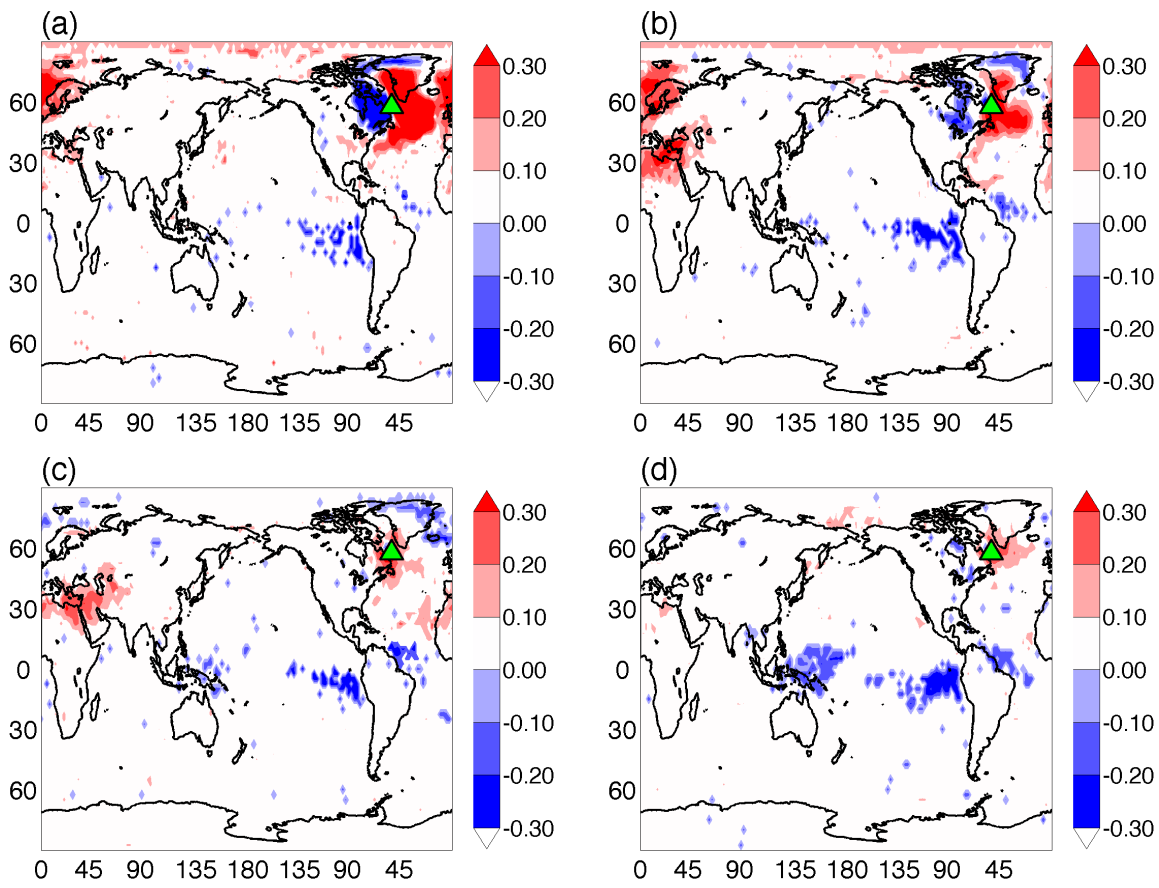


Figure 6.8: As in Fig 6.7 but for a node in the northern extratropics —the Labrador Sea, as in Fig. 6.4 (b). The values of  $\tau$  are as in Fig 6.7. The Labrador sea area is related to a source of atmospheric variability of the north Atlantic ocean that affects Europe. As in Fig. 6.7, smaller values of  $\tau$  can capture wave trains propagating from west to east, over Europe, while values of  $\tau$  longer than a week loose this effect and only the influence from the Pacific ocean persists.

clearly marked as the boundary between the incoming and outgoing links. It is also seen for  $\tau = 3$  suggesting that in about three days the front reaches the Mediterranean region affecting temperatures there. Again, as  $\tau$  increases mainly the tropical links remain. However, even for  $\tau = 30$  there is a well defined region of outgoing links that remain over the Labrador Sea, suggesting that the SAT in the region may have relatively long time scales of variability, perhaps related to the North Atlantic Oscillation. Another possibility is the influence of the local ocean that increases the persistence of atmospheric temperature anomalies through

thermodynamic coupling, as shown by Barsugli *et. al.* [210].

### 6.3.3 Influence of $\tau$ in the tropics

Figure 6.9 presents the effect of  $\tau$  for a point in the tropics. Panel 6.9 (a) shows the point considered in the Pacific ocean for  $\tau = 1$  day. Contrarily to e.g. figure 6.7, where there is a clear front propagation due to the existence of wave trains in the extratropics, in this case the connections are weak. The opposite case, for  $\tau > 30$  days—Panels 6.9 (b,c)—are consistent with figure 6.6 as the structures are robust for a wide range of  $\tau$ . Notice in Panel 6.9 (c) the presence of a second blue area near the coast of Chile for high values of  $\tau$ . In order to complement the information provided in panels 6.9 (a,b,c) and also in Fig. 6.6. Panels 6.9 (d,e,f) show maps for a point in the Indian ocean, for values of  $\tau = 1, 45,$  and  $90$  days respectively where similar results are obtained with weak connections for low values of  $\tau$  and robustly strong for  $\tau > 30$  days.

## 6.4 Directionality on the tropical Pacific Ocean

In Fig. 6.10 the DI for  $\tau = 30$  days has been plotted for points covering all the equatorial Pacific. They begin at  $95.0^\circ$  W—panel 6.10 (a)—and end at  $125.0^\circ$  E—panel 6.10 (d). Clearly, the influence the Pacific ocean exerts is almost global, over tropics and extratropics, in agreement with previous studies (e.g. in [211]). The DI allows to show that—even if there are feedbacks and the Pacific is affected by extratropical perturbations and other ocean basins (e.g. the tropical Atlantic)—the influence is *effectively* from the Pacific to the rest of the world. Moreover, the maps show that the largest influence is done by the equatorial Pacific close to the dateline. This is clear in the extratropical atmosphere, as well as in the tropical north Atlantic. On the other hand, the connection to the Indian ocean and south Atlantic is not so sensitive of the point considered over the equatorial Pacific. As the reference point moves further west from the dateline the influence decreases substantially, only remaining weak connections to the tropical north Atlantic and Indian oceans. The methodology can thus be applied to find the best region to construct an index that describes the Pacific influence over the area where climate anomalies are studied.

Notice that all maps show a blue tongue of incoming links to the east of the point considered; it is seen first in panel 6.10 (a) and extends westward until covering the whole Pacific

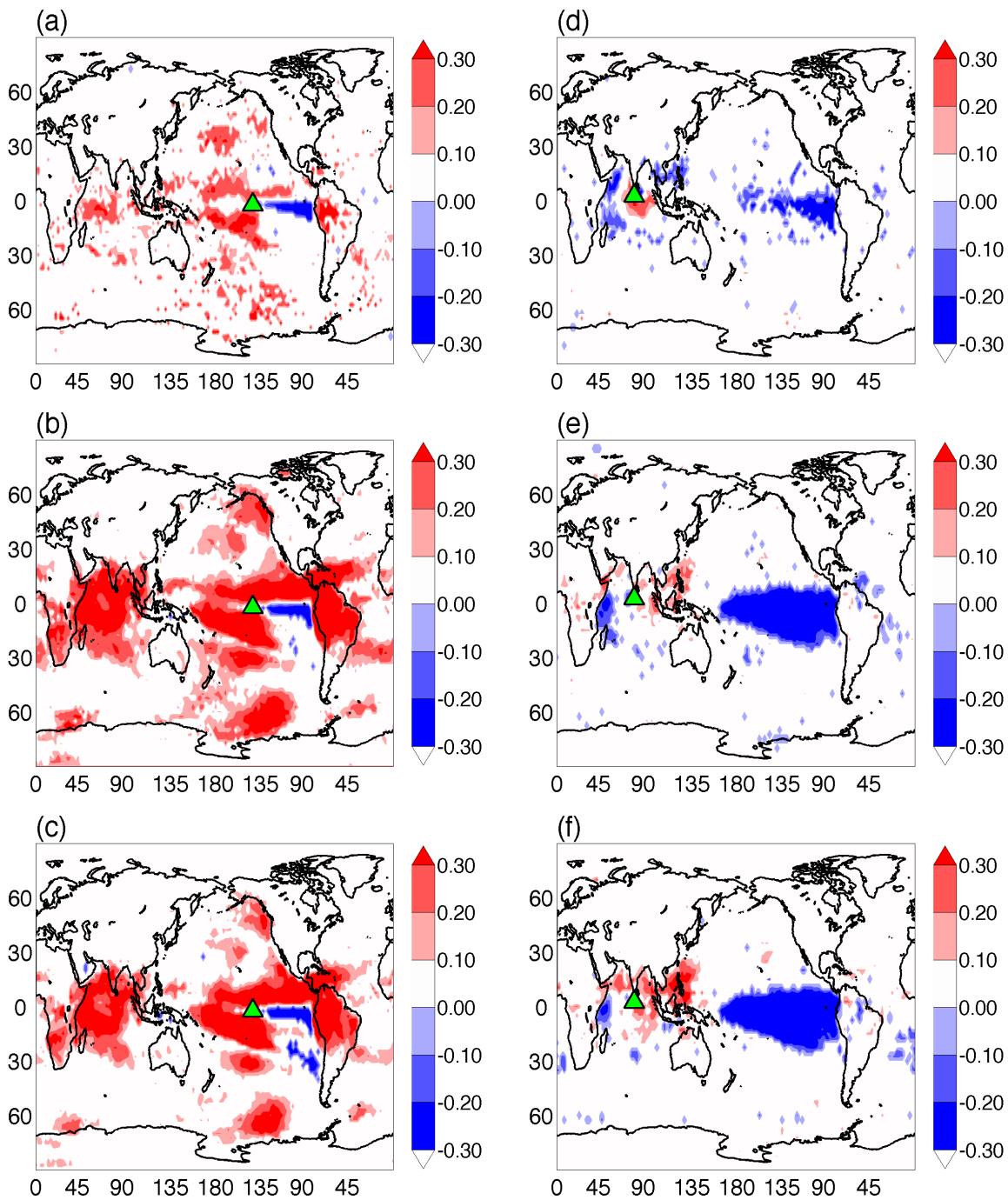


Figure 6.9: Effect of  $\tau$  using daily data for a node in the tropics. In panels (a-c) the node is in the Pacific ocean, and (d-f), in the Indian ocean. The values of  $\tau$  are: (a,d) 1 day, (b,e) 45 days, and (c,f) 90 days. The maps for  $\tau = 30$  were shown in Fig. 6.3. In contrast to the extra-tropics, where the propagation of waves in the scale of days is dominant, and are clearly seen in the maps for time scales of a few days, in the tropics the variability is longer as the ocean adds persistence to the nearly barotropic atmosphere. For higher values of  $\tau$  only long lasting phenomena are observed, related to the strongest variability pattern in this area, which is ENSO.

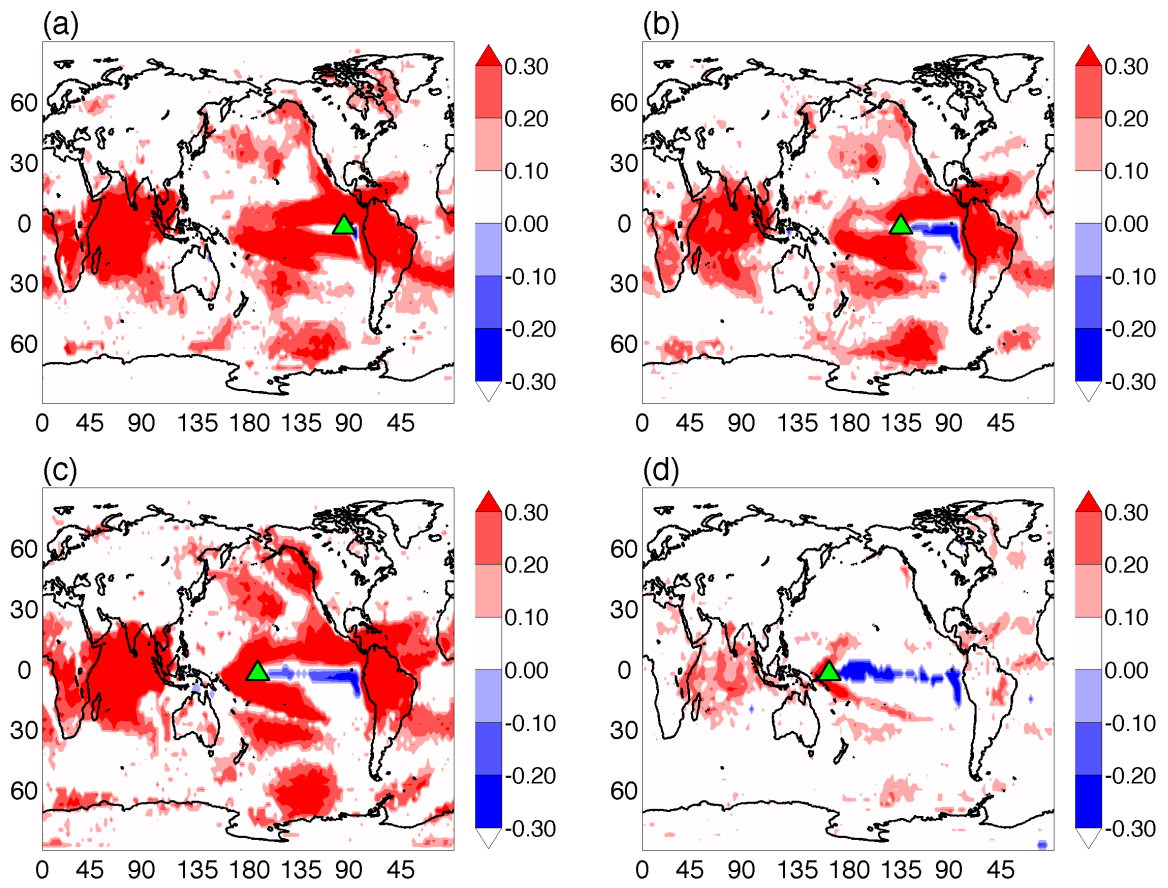


Figure 6.10: The zonal change of directionality over the equatorial Pacific is shown. In all cases  $\tau = 30$  days. From (a) near the south American coast, to (d) in the western Pacific ocean. As seen in the maps, most of the points over central and eastern Pacific have an important effect over a large part of the world; especially over the tropical areas and the rest of the Pacific ocean—notice that in (a)-(c) the teleconnections remain basically the same, although the intensity varies. Incoming links are in blue while outgoing are in red.

ocean in 6.10 (d). This feature is related to the existence of the equatorial cold tongue and the fact that easterly trades blow over the equator thus advecting air from the east to the west of the point.

## 6.5 Conclusions

An information theoretic measure to calculate the direction of propagation of information directly from the data was implemented and used to construct directed climate networks of surface air temperature. The applicability of the method was shown for monthly- and daily-

mean data. Using monthly-averaged SAT data the results from previous chapters have been recovered, and the system correctly found the direction of propagation of the patterns.

The measure included a parameter  $\tau$ , that represents the time required for the information to travel from one node to another. It was found that by adequately tuning  $\tau$  the network connectivity varied revealing the various time-scales of atmospheric processes; too short values of  $\tau$  failed to capture several long-range links, while for too large values of  $\tau$  (above the length of decorrelation in the data) the connectivity of the network decreased drastically. A further study of the dependence of some links with  $\tau$  was made and it was found to be consistent with the propagation of information in the tropics and extra tropics.

For monthly-mean data, it was shown that the method recovers the known direction of propagation, and it was shown to be more suited for the tropics where variability is slower due to the influence of the SSTs. In addition, when considering daily-averaged SAT data, the analysis revealed variability patterns consistent with known features of the global climate dynamics. For example, in the extratropics the long time average synoptic weather was correctly inferred: as specific examples two geographical regions in different hemispheres—one node in the La Plata basin and another in the Labrador Sea—were considered, and the link direction revealing wave trains propagating from west to east, in both hemispheres was shown.



## **Chapter**

# 7

## **Conclusions**

This Thesis was centered in the construction of climate networks based on the surface air temperature field, and the use of these networks to study several aspects of the global climate. Built out by means of information theoretic measures and ordinal time series analysis, these networks allowed to study statistical similarity among geographical regions, time scales of variability and the direction of their interactions. Statistical similarity was evaluated by means of information theoretic measures and special efforts were made in assessing the statistical significance of the links of the networks by means of classical statistics, surrogate data and bootstrap.

In Chapter 4, two thresholding methodologies were employed. The first one was based on surrogate data to define a global threshold, chosen as the mean value plus 3 standard deviations of the surrogate MI distribution (Figs. 4.1, 4.2). In order to compare with previous work [12, 120, 128], the second methodology fixed the link density (Figs. 4.3, 4.4). The networks obtained using both thresholding schemes were consistent. Furthermore, the first method was found to display the complexity of the atmospheric teleconnections in more detail due to the increasing of the number of links. For the second methodology, the main features of the network for fixed link density were found to be independent of the bin number used for computing the mutual information, so confirming the robustness of the climate networks constructed with this thresholding methodology.

The use of ordinal patterns allowed to study the properties of networks built at different time scales. An increase in the ordinal pattern spacing—therefore, an increase in the timescale—was found to generate a richer connectivity pattern in the equatorial Pacific area

(figures 4.1–4.4, left column, rows 2–4). This increase in connectivity is associated with an increase in the teleconnections from points situated in this region. This result is consistent with climatological interpretations in terms of atmospheric teleconnection patterns. As the equatorial Pacific is known to be dominated by ENSO on scales longer than several months, this method allowed to obtain networks where the effect of ENSO went from weak (monthly OP) to strong (yearly OP), independently of the number of letters used for the ordinal patterns and of the thresholding methodology.

\*\*\*

In Chapter 5 the variability of the monthly averaged surface air temperature field was decomposed—using an ensemble of nine AGCM runs forced with the same SST data, and starting from slightly different initial conditions—into a part forced by the ocean temperature, and another due to internal atmospheric variability. As the separation of internal and forced variability required averaging over several model runs, the networks obtained could not have been calculated from observational/reanalysis data only.

Separate climate networks were constructed for the forced and for the internal variability components, using Mutual Information to assess the interdependencies between the time series. Ordinal patterns were also used in order to separate and determine the strength of the links on different time scales. The NINO3.4 and NAO indices were linearly removed from the original data and the CNs were recalculated to assess the effect of these phenomena on the connectivity of the network.

The main conclusions were that the global connectivity of the *forced* variability network is heavily influenced by ENSO (Figs. 5.3 and 5.4), and that of the *internal* variability network is significantly influenced—in the north Atlantic—by the NAO (Figs. 5.5 and 5.6). This is in excellent agreement with well known climatological evidence [1]. Moreover, this has revealed that most of the links detected in the forced variability proceed from long time scales, while the contributions of intra-annual time scales to the internal variability are the most important.

Another conclusion is that forced and internal atmospheric variabilities are characterized by *very different* networks. The connectivity of the *forced* variability was shown to be strong in the tropics, while that of the *internal* variability has a maximum in mid latitudes.

---

*Long-range* teleconnections from the tropics to the extratropics—and even between different hemispheres—were observed in the *forced* network, and explained by the influence of ENSO. On the other hand, the network of *internal* atmospheric variability has the strongest connections in the extratropics, and connections to the Labrador Sea were found to be heavily affected by the North Atlantic Oscillation.

\*\*\*

In the last part of the thesis, Chapter 6, *directed* climate networks were constructed by using a directionality index, which has not been previously employed for the analysis of climate data. The presence of significant links was inferred using mutual information, while the direction of those links was inferred using the directionality index (Fig. 6.1). To assess the statistical significance of these results the bootstrap method was employed.

For monthly-averaged SAT, the analysis extended previous works [120, 128, 130], to inferring the net direction of propagation of the information. The inference method was tested against the value of the parameter  $\tau$ , that represents the time required for the information to be transferred from one node to another. The network connectivity was found to vary as a function of  $\tau$  (figure 6.6), revealing the various time-scales of atmospheric processes; too short values of  $\tau$  failed to capture several long-range links, while for too large values of  $\tau$  (above the decorrelation length of the data) the connectivity of the network decreased drastically. This effect was also dependent of the latitude and the variability patterns present on the regions considered.

In addition, when considering daily-averaged SAT data, the analysis revealed the influence of variability patterns (as ENSO, NAO and ) consistent with known features of the global climate dynamics. In the extra-tropics, the averaged synoptic weather patterns were correctly inferred. For example, considering nodes in La Plata basin and the Labrador Sea, the link direction revealed wave trains propagating from west to east in both hemispheres, consistent with the synoptic weather patterns that dominate atmospheric variability in mid-latitudes (Figs. 6.7 and 6.8). On the other hand, in the tropical Pacific and Indian oceans, ENSO was found to be the main component on long time-scale variability (Fig. 6.9). These are not only fully consistent with previous results presented in this thesis (e.g. see top right panel of Fig. 4.1), but also complemented these patterns establishing the direction of transfer of the information.

To summarize, the work presented in this thesis has advanced the state of the art in a number of ways:

- The networks constructed in Chapter 4 were found to be fully consistent with those previously reported in [12, 128]; furthermore, a careful significance analysis allowed to construct non-sparse, higher density networks—by identifying weak but significant links—which displayed features, as teleconnection patterns, only hinted in the previous analysis but expected by the present knowledge of climate science.
- In Chapters 4 and 5, ordinal analysis allowed for a clear identification of the different time scales involved patterns in both internal and forced atmospheric variability types.
- In Chapter 5, the linear removal of ENSO and NAO indices allowed to understand the effect well known climate patterns on the climate network topology. This effect has found to be considerable, as teleconnection patterns resulted to be major components of the long-range network connectivity on both internal and forced variability networks.
- The directionality index of Chapter 6—not used before on climate studies—permitted a direct analysis of the transference of information among nodes, from the data. The patterns revealed are consistent with those of chapters 4 and 5, and with the state of the art knowledge of variability patterns.
- The analysis of  $\tau$  performed in Chapter 6 allowed to identify the time-scales of net link directionality; for example, the analysis revealed that the Indian ocean responds to the incoming ENSO signal in a time scale of about 2 months, this time being shorter for other areas in the globe (figure 6.6).

As future work, it would be interesting to apply this methodology to analyze how the climate network changes for different climatic variables—as vertical velocity (related with rainfall in the tropics) or geopotential height—in different seasons, or in different phases of some phenomenon as ENSO. Also to focus in some specific areas of the globe, studying their dynamics and time-scales, together with the linear removal of leading patterns of variability is a direct application of the methods devised during this thesis.

---

This work also opens the possibility of studying how various network measures—such as average path length, assortativity, clustering coefficient, betweenness, etc, calculated from the adjacency matrices obtained—depend on the time scale considered for quantifying statistical interdependencies, together with the physical interpretation of these measures in terms of climate dynamics. Also an analysis of synchronization of the networks at different time scales could be of interest. As the focus of this thesis was on the construction of the networks, much of this analysis has not been yet done.

Another interesting issue is to analyze the effect of directionality on several time scales, by computing the directionality index with the ordinal patterns approach. This would allow to construct networks that characterize short-scale (days to months) or long-scale (few years) atmospheric processes.

Some of these works are in progress as a natural continuation of this research.



# Bibliography

- [1] Henk A. Dijkstra. *Nonlinear Climate Dynamics*. Cambridge University Press, 2013.
- [2] Henk A. Dijkstra and Michael Ghil. Low-frequency variability of the large-scale ocean circulation: A dynamical systems approach. *Rev. Geophys.*, 43(3):RG3002, 2005.
- [3] Howard A. Bridgman and John E. Oliver. *The Global Climate System*. Cambridge University Press, 2006.
- [4] Duncan J. Watts and Steven H. Strogatz. Collective dynamics of 'small-world' networks. *Nature*, 393(6684):440–2, June 1998.
- [5] Gary A. Polis and Donald R. Strong. Food web complexity and community dynamics. *Am. Nat.*, 147(5):813–846, 1996.
- [6] Serban Nacu, Rebecca Critchley-Thorne, Peter Lee, and Susan Holmes. Gene expression network analysis and applications to immunology. *Bioinformatics*, 23(7):850–8, April 2007.
- [7] Edward T. Bullmore and Olaf Sporns. Complex brain networks: graph theoretical analysis of structural and functional systems. *Nat. Rev. Neurosci.*, 10(3):186–98, March 2009.
- [8] Reuven Cohen, Keren Erez, Daniel Ben-Avraham, and Shlomo Havlin. Resilience of the Internet to Random Breakdowns. *Phys. Rev. Lett.*, 85(21):4626–4628, November 2000.
- [9] Anastasios A. Tsonis, Kyle L. Swanson, and Paul J. Roebber. What Do Networks Have to Do with Climate? *Bull. Am. Meteorol. Soc.*, 87(5):585–595, May 2006.
- [10] Kazuko Yamasaki, Avi Gozolchiani, and Shlomo Havlin. Climate Networks around the Globe are Significantly Affected by El Niño. *Phys. Rev. Lett.*, 100(22):228501, June 2008.
- [11] Anastasios A. Tsonis and Kyle L. Swanson. Topology and Predictability of El Niño and La Niña Networks. *Phys. Rev. Lett.*, 100(22):228502, June 2008.
- [12] Jonathan F. Donges, Yong Zou, Norbert Marwan, and Jürgen Kurths. Complex networks in climate dynamics. *Eur. Phys. J. Spec. Top.*, 174(1):157–179, July 2009.
- [13] Charles Dickens. *A Christmas Carol*. Chapman and Hall, 1843.
- [14] <http://www.metoffice.gov.uk/public/weather/climate/gcpsvf37b> (October 2014).
- [15] Intergovernmental Panel on Climate Change. *Climate Change 2014: Impacts, Adaptation and Vulnerability*. Cambridge University Press, 2014.
- [16] Dennis Wheeler, Ricardo Garcia-Herrera, Clive W. Wilkinson, and Catharine Ward. Atmospheric circulation and storminess derived from Royal Navy logbooks: 1685 to 1750. *Clim. Change*, 101(1-2):257–280, October 2009.
- [17] William H. Quinn, Victor T. Neal, and Santiago E. Antunez De Mayolo. El Niño occurrences over the past four and a half centuries. *J. Geophys. Res.*, 92(C13):14449, 1987.

## BIBLIOGRAPHY

---

- [18] Michael E Mann, Raymond S Bradley, and Malcolm K Hughes. Global-scale temperature patterns and climate forcing over the past six centuries. *Nature*, 392:779–787, 1998.
- [19] Michael Ghil, Michael C Maccracken, John S Perry, and Ted Munn. Natural Climate Variability. In Ted Munn, Michael C MacCracken, and John S Perry, editors, *Encycl. Glob. Environ. Chang. Vol. 1*, volume 1, pages 544–549. Wiley, 2002.
- [20] Douglas G. Martinson, Kirk Bryan, Michael Ghil, Melinda M. Hall, Thomas R. Karl, Edward S Sarachik, Soroosh Sorooshian, and Lynne D. Talley. *Natural climate variability on decade-to-century time scales*. National Academy Press, 1995.
- [21] Dana L. Royer and Peter Wilf. Why do toothed leaves correlate with cold climates? Gas exchange at leaf margins provides new insights into a classic paleotemperature proxy. *Int. J. Plant Sci.*, 167(September 2005):11–18, 2006.
- [22] Claire L. Smith, Andy Baker, Ian J. Fairchild, Silvia Frisia, and Andrea Borsato. Reconstructing hemispheric-scale climates from multiple stalagmite records. *Int. J. Climatol.*, 26(10):1417–1424, August 2006.
- [23] Maria Bianca Cita, Colette Vergnaud-Grazzini, Christian Robert, Hervé Chamley, Neri Ciaranfi, and Sara D’Onofrio. Paleoclimatic record of a long deep sea core from the eastern Mediterranean. *Quat. Res.*, 8(2):205–235, September 1977.
- [24] Brian Huntley and Iain Colin Prentice. July temperatures in europe from pollen data, 6000 years before present. *Science (80-. )*, 241(4866):687–90, August 1988.
- [25] Raymond S Bradley. *Paleoclimatology: reconstructing climates of the Quaternary*. Academic Press, 1999.
- [26] Victor A. Melezhik. Multiple causes of Earth’s earliest global glaciation. *Terra Nov.*, 18(2):130–137, March 2006.
- [27] Basil A.S. Davis, Simon Brewer, Anthony C. Stevenson, and Joël Guiot. The temperature of Europe during the Holocene reconstructed from pollen data. *Quat. Sci. Rev.*, 22(15-17):1701–1716, July 2003.
- [28] Ricardo Villalba. Tree-ring and glacial evidence for the medieval warm epoch and the little ice age in southern South America. *Clim. Change*, 26(2-3):183–197, March 1994.
- [29] Bert Rein, Andreas Lückge, Lutz Reinhardt, Frank Sirocko, Anja Wolf, and Wolf-Christian Dullo. El Niño variability off Peru during the last 20,000 years. *Paleoceanography*, 20(4):n/a–n/a, December 2005.
- [30] Janice Hill. *Weather from Above: America’s Meteorological Satellites*. Smithsonian, 1991.
- [31] David E. Parker, Tim P. Legg, and Chris K Folland. A new daily central England temperature series, 1772–1991. *Int. J. Climatol.*, 12(4):317–342, May 1992.
- [32] Michael Ghil and Paola Malanotte-Rizzoli. Advances in Geophysics Volume 33. In *Adv. Geophys.*, volume 33 of *Advances in Geophysics*, pages 141–266. Elsevier, 1991.
- [33] Eugenia Kalnay and Others. The NCEP/NCAR 40-Year Reanalysis Project. *Bull. Am. Meteorol. Soc.*, 77(3):437–471, March 1996.
- [34] Sakari M. Uppala and Others. The ERA-40 re-analysis. *Q. J. R. Meteorol. Soc.*, 131(612):2961–3012, October 2005.
- [35] [http://jra.kishou.go.jp/JRA-25/index\\_en.html](http://jra.kishou.go.jp/JRA-25/index_en.html) (October 2014).



- [36] Richard A Wood Uk. Climate Models and Their Evaluation. *Evaluation*, 323:589–662, 2007.
- [37] IPCC. *Climate Change 2014: Impacts, Adaptation, and Vulnerability. Part A: Global and Sectoral Aspects. Contribution of Working Group II to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* [Field, C.B., V.R. Barros, D.J. Dokken, K.J. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 2014.
- [38] Milan Milankovitch. *Mathematische Klimalehre und astronomische Theorie der Klimaschwankungen*. Gebruder Borntraeger, 1930.
- [39] Milan Milankovitch. Ausbau und gegenwärtiger Stand der astronomischen Theorie der erdgeschichtlichen Klimate. *Experientia*, 4(11):413–418, November 1948.
- [40] Peter Huybers. Combined obliquity and precession pacing of late Pleistocene deglaciations. *Nature*, 480(7376):229–32, December 2011.
- [41] André Berger. Milankovitch Theory and climate. *Rev. Geophys.*, 26(4):624, 1988.
- [42] Roberto Benzi, Giorgio Parisi, Alfonso Sutera, and Angelo Vulpiani. Stochastic resonance in climatic change. *Tellus*, 34(1):10–16, February 1982.
- [43] Gavin C. Cawley. On the Atmospheric Residence Time of Anthropogenically Sourced Carbon Dioxide. *Energy & Fuels*, 25(11):5503–5513, November 2011.
- [44] Dennis L. Hartmann and Dennis L. Hartmann. *Global Physical Climatology*. Academic Press, San Diego, USA, 1994.
- [45] <http://www.esrl.noaa.gov/psd/data/climateindices/list/> (October 2014).
- [46] Chester F Ropelewski and Michael S. Halpert. Global and Regional Scale Precipitation Patterns Associated with the El Niño/Southern Oscillation. *Mon. Weather Rev.*, 115:1606–1626, 1987.
- [47] George N. Kiladis and Henry F. Diaz. Global Climatic Anomalies Associated with Extremes in the Southern Oscillation. *J. Clim.*, 2(9):1069–1090, September 1989.
- [48] S. George Philander. *El Niño, La Niña, and the Southern Oscillation*. Academic Press, 1990.
- [49] Michael S. Halpert and Chester F. Ropelewski. Surface Temperature Patterns Associated with the Southern Oscillation. *J. Clim.*, 5(6):577–593, June 1992.
- [50] S. Brönnimann. Impact of El Niño-Southern Oscillation on European climate. *Rev. Geophys.*, 45(3):RG3003, September 2007.
- [51] Marcelo Barreiro. Influence of ENSO and the South Atlantic Ocean on climate predictability over Southeastern South America. *Clim. Dyn.*, 35(7-8):1493–1508, September 2009.
- [52] [http://www.srh.noaa.gov/jetstream/tropics/enso\\_impacts.htm](http://www.srh.noaa.gov/jetstream/tropics/enso_impacts.htm) (October 2014).
- [53] Paul Cashin, Kamiar Mohaddes, and Mehdi Raissi. Fair Weather or Foul? The Macroeconomic Effects of El Niño. *Cambridge Work. Pap. Econ.*, cwpe:1418, June 2014.
- [54] Kevin E. Trenberth. The Definition of El Niño. *Bull. Am. Meteorol. Soc.*, 78(12):2771–2777, December 1997.
- [55] <http://www.ncdc.noaa.gov/teleconnections/enso/indicators/sst.php> (October 2014).

## BIBLIOGRAPHY

---

- [56] Axel Timmermann, Josef M. Oberhuber, Andrea Bacher, Marjolein Esch, Mojib Latif, and Erich Roeckner. Increased El Niño frequency in a climate model forced by future greenhouse warming. *Nature*, 398(June 1982):1996–1999, 1999.
- [57] Matthew Collins. El Niño- or La Niña-like climate change? *Clim. Dyn.*, 24(1):89–104, December 2004.
- [58] James W. Hurrell, Yochanan Kushnir, Geir Ottersen, and Martin Visbeck, editors. *The North Atlantic Oscillation: Climatic Significance and Environmental Impact*, volume 134 of *Geophysical Monograph Series*. American Geophysical Union, Washington, D. C., 2003.
- [59] James W. Hurrell. Decadal trends in the north atlantic oscillation: regional temperatures and precipitation. *Science (80-. )*, 269(5224):676–9, August 1995.
- [60] Yochanan Kushnir. Climatology: Europe’s winter prospects. *Nature*, 398(6725):289–291, March 1999.
- [61] <http://www.nc-climate.ncsu.edu/climate/patterns/NAO.html> (October 2014).
- [62] James W. Hurrell. Influence of variations in extratropical wintertime teleconnections on northern hemisphere temperature. *Geophys. Res. Lett.*, 23(6):665–668, March 1996.
- [63] Mark J. Rodwell, David P. Rowell, and Chris K. Folland. Oceanic forcing of the wintertime North Atlantic Oscillation and European climate. *Nature*, 398(6725):320–323, March 1999.
- [64] Martin Visbeck, James W. Hurrell, Lorenzo Polvani, and Heidi M Cullen. The North Atlantic Oscillation: past, present, and future. *Proc. Natl. Acad. Sci.*, 98(23):12876–7, November 2001.
- [65] John Marshall and Others. North Atlantic climate variability: phenomena, impacts and mechanisms. *Int. J. Climatol.*, 21(15):1863–1898, December 2001.
- [66] Richard C. Cornes, Philip D. Jones, Keith R. Briffa, and Timothy J. Osborn. Estimates of the North Atlantic Oscillation back to 1692 using a Paris-London westerly index. *Int. J. Climatol.*, 33(1):228–248, January 2013.
- [67] Cécile Penland and Leslie M. Hartten. Stochastic forcing of north tropical Atlantic sea surface temperatures by the North Atlantic Oscillation. *Geophys. Res. Lett.*, 41(6):2126–2132, March 2014.
- [68] Leonhard Euler. Solutio problematis ad geometriam situs pertinentis (The solution of a problem relating to the geometry of position). *Comment. Acad. Sci. Petropolitanae*, 8:128–140, 1741.
- [69] Robin J. Wilson. An Eulerian trail through Königsberg. *J. Graph Theory*, 10(3):265–275, January 1986.
- [70] <http://eulerarchive.maa.org/pages/E053.html> (November 2014).
- [71] Mark E. J. Newman, Cristopher Moore, and Duncan J. Watts. Mean-Field Solution of the Small-World Network Model. *Phys. Rev. Lett.*, 84(14):3201–3204, April 2000.
- [72] Luis A. Nunes Amaral, Antonio Scala, Marc Barthelemy, and H. Eugene Stanley. Classes of small-world networks. *Proc. Natl. Acad. Sci.*, 97(21):11149–52, October 2000.
- [73] Steven H. Strogatz. Exploring complex networks. *Nature*, 410(6825):268–76, March 2001.
- [74] Alejandro Sánchez, Juan López, and Miguel Rodríguez. Nonequilibrium Phase Transitions in Directed Small-World Networks. *Phys. Rev. Lett.*, 88(4):048701, January 2002.

- 
- [75] Réka Albert and Albert-László Barabási. Statistical mechanics of complex networks. *Rev. Mod. Phys.*, 74(1):47–97, January 2002.
- [76] Chaoming Song, Shlomo Havlin, and Hernán A Makse. Self-similarity of complex networks. *Nature*, 433(7024):392–395, January 2005.
- [77] Stefano Boccaletti, Vito Latora, Yamir Moreno, Martín Chavez, and Dong-Uk Hwang. Complex networks: Structure and dynamics. *Phys. Rep.*, 424(4-5):175–308, February 2006.
- [78] Santo Fortunato. Community detection in graphs. *Phys. Rep.*, 486(3-5):75–174, February 2010.
- [79] AL Do, Stefano Boccaletti, and Thilo Gross. Graphical notation reveals topological stability criteria for collective dynamics in complex networks. *Phys. Rev. Lett.*, 108(19):194102, 2012.
- [80] Sergey N. Dorogovtsev and Jose F.F. Mendes. *Evolution of Networks*. Oxford University Press, 2003.
- [81] Reuven Cohen and Shlomo Havlin. *Complex Networks: Structure, Robustness and Function*. Cambridge University Press, 2010.
- [82] Linton C. Freeman. Centrality in social networks conceptual clarification. *Soc. Networks*, 1(3):215–239, January 1978.
- [83] Gesa A. Böhme and Thilo Gross. Fragmentation transitions in multistate voter models. *Phys. Rev. E*, 85(6):066117, June 2012.
- [84] Luca Valori and Francesco Picciolo. Reconciling long-term cultural diversity and short-term collective social behavior. *Proc. Natl. Acad. Sci.*, 109(4):1068–73, January 2012.
- [85] Michael Szell, Renaud Lambiotte, and Stefan Thurner. Multirelational organization of large-scale social networks in an online world. *Proc. Natl. Acad. Sci.*, 107(31):13636–41, August 2010.
- [86] Nicholas A. Christakis and James H. Fowler. Friendship and natural selection. *Proc. Natl. Acad. Sci.*, 10.1073:1–6, July 2014.
- [87] Robert T Paine. Food web complexity and species diversity. *Am. Nat.*, 100(910):65–75, 1966.
- [88] Joel E. Cohen, Frédéric Briand, and Charles M. Newman. *Community food webs: Data and theory*. Springer Berlin Heidelberg, 1990.
- [89] Jennifer A Dunne, Richard J Williams, and Neo D Martínez. Food-web structure and network theory: The role of connectance and size. *Proc. Natl. Acad. Sci.*, 99(20):12917–22, October 2002.
- [90] Hawoong Jeong, Bálint Tombor, Réka Albert, Zoltan N. Oltvai, and Albert-László Barabási. The large-scale organization of metabolic networks. *Nature*, 407(6804):651–4, October 2000.
- [91] Chung-Kang Peng, Sergey V. Buldyrev, Shlomo Havlin, Mark Simons, H. Eugene Stanley, and Ary L. Goldberger. Mosaic organization of DNA nucleotides. *Phys. Rev. E*, 49(2):1685–1689, February 1994.
- [92] Trey Ideker, Vesteinn Thorsson, Jeffrey A. Ranish, Rowan Christmas, Jeremy Buhler, Jimmy K. Eng, Roger Bumgarner, David R. Goodlett, Ruedi Aebersold, and Leroy Hood. Integrated genomic and proteomic analyses of a systematically perturbed metabolic network. *Science*, 292(5518):929–34, May 2001.

## BIBLIOGRAPHY

---

- [93] Melissa S. Cline and Others. Integration of biological networks and gene expression data using Cytoscape. *Nat. Protoc.*, 2(10):2366–82, January 2007.
- [94] Sabrina Hempel, Aneta Koseska, Zoran Nikoloski, and Jürgen Kurths. Unraveling gene regulatory networks from time-resolved gene expression data - a measures comparison study. *BMC Bioinformatics*, 12(1):292, January 2011.
- [95] Romualdo Pastor-Satorras and Alessandro Vespignani. Epidemic Spreading in Scale-Free Networks. *Phys. Rev. Lett.*, 86(14):3200–3203, April 2001.
- [96] Marc Barthélemy and Alain Barrat. Velocity and hierarchical spread of epidemic outbreaks in scale-free networks. *Phys. Rev. Lett.*, 2004.
- [97] Linyuan Lü, Duan-Bing Chen, and Tao Zhou. The small world yields the most effective information spreading. *New J. Phys.*, 13(12):123005, December 2011.
- [98] Fredrik Liljeros, Christofer R. Edling, and Luis A. Nunes Amaral. Sexual networks: implications for the transmission of sexually transmitted infections. *Microbes Infect.*, 5(2):189–196, February 2003.
- [99] Simon Haykin. *Neural Networks: A comprehensive foundation*. Prentice Hall, 2004.
- [100] Olaf Sporns, Dante R Chialvo, Marcus Kaiser, and Claus C Hilgetag. Organization, development and function of complex brain networks. *Trends Cogn. Sci.*, 8(9):418–25, September 2004.
- [101] Patrick McGraw and Michael Menzinger. Self-sustaining oscillations in complex networks of excitable elements. *Phys. Rev. E*, 83(3):037102, March 2011.
- [102] Danielle S. Bassett and Edward T. Bullmore. Small-world brain networks. *Neuroscientist*, 12(6):512–23, December 2006.
- [103] Olaf Sporns. Structure and function of complex brain networks. *Dialogues Clin. Neurosci.*, 15(3):247–62, September 2013.
- [104] Ian J. Goodfellow, Yaroslav Bulatov, Julian Ibarz, Sacha Arnoud, and Vinay Shet. Multi-digit Number Recognition from Street View Imagery using Deep Convolutional Neural Networks, Internal google publication, also available from. <http://arxiv.org/abs/1312.6082>, December 2013.
- [105] Réka Albert, Hawoong Jeong, and Albert-László Barabási. Internet: Diameter of the World-Wide Web. *Nature*, 401(September):398–399, 1999.
- [106] Roger Guimerà and Luis A. Nunes Amaral. Modeling the world-wide airport network. *Eur. Phys. J. B*, 38(2):381–385, March 2004.
- [107] M. Argollo de Menezes and Albert-László Barabási. Fluctuations in Network Dynamics. *Phys. Rev. Lett.*, 92(2):028701, January 2004.
- [108] Eugene Garfield, Irving H Sher, and Richard J. Torpie. The use of citation data in writing the history of science. Technical report, Institute for scientific information inc. Philadelphia, 1964.
- [109] Albert-László Barabási, Hawoong Jeong, Z. Néda, E. Ravasz, Albert Schubert, and T. Vicsek. Evolution of the social network of scientific collaborations. *Physica A*, 311(3-4):590–614, August 2002.
- [110] Tanmoy Chakraborty, Sandipan Sikdar, Niloy Ganguly, and Animesh Mukherjee. Citation interactions among computer science fields: a quantitative route to the rise and fall of scientific research. *Soc. Netw. Anal. Min.*, 4(1):187, April 2014.

- 
- [111] Jianxi Gao, Sergey V. Buldyrev, Shlomo Havlin, and H. Eugene Stanley. Robustness of a network of networks. *Phys. Rev. Lett.*, 107(19):195701, 2011.
- [112] Li Daqing, Kosmas Kosmidis, Armin Bunde, and Shlomo Havlin. Dimension of spatially embedded networks. *Nat. Phys.*, 7(6):481–484, February 2011.
- [113] Elizabeth Quill. When Networks Network. *Sci. News*, 182(6):1–17, 2012.
- [114] Kosmas Kosmidis, Shlomo Havlin, and Armin Bunde. Structural properties of spatially embedded networks. *Europhys. Lett.*, 82(4):48005, 2008.
- [115] Seth Bullock, Lionel Barnett, and Ezequiel A. di Paolo. Spatial embedding and the structure of complex networks. *Complexity*, 16(2):20–28, 2010.
- [116] Thorsten Emmerich, Armin Bunde, Shlomo Havlin, Guanliang Li, and Daqing Li. Complex networks embedded in space: Dimension and scaling relations between mass, topological distance, and Euclidean distance. *Phys. Rev. E*, 87(3):032802, March 2013.
- [117] Alex Arenas, Albert Díaz-Guilera, Jürgen Kurths, Yamir Moreno, and Changsong Zhou. Synchronization in complex networks. *Phys. Rep.*, 469(3):93–153, December 2008.
- [118] Jesús Gómez-Gardeñes, Sergio Gómez, Alex Arenas, and Yamir Moreno. Explosive Synchronization Transitions in Scale-Free Networks. *Phys. Rev. Lett.*, 106(12):128701, March 2011.
- [119] Anastasios A. Tsonis and Paul J. Roebber. The architecture of the climate network. *Physica A*, 333:497–504, February 2004.
- [120] Jonathan F. Donges, Yong Zou, Norbert Marwan, and Jürgen Kurths. The backbone of the climate network. *Europhys. Lett.*, 87(4):48007, August 2009.
- [121] Paul Erdős and Alfréd Rényi. On Random Graphs. *Publ. Math.*, 6:290–297, 1959.
- [122] Reuven Cohen and Shlomo Havlin. Scale-Free Networks Are Ultrasmall. *Phys. Rev. Lett.*, 90(5):058701, February 2003.
- [123] <http://www.awi.de/> (October 2014).
- [124] <https://reanalyses.org/atmosphere/overview-current-reanalyses> (November 2014).
- [125] Anastasios A. Tsonis, Kyle L. Swanson, and Geli Wang. On the Role of Atmospheric Teleconnections in Climate. *J. Clim.*, 21(12):2990–3001, June 2008.
- [126] Stephan Bialonski, Marie-Therese Horstmann, and Klaus Lehnertz. From brain to earth and climate systems: small-world interaction networks or not? *Chaos*, 20(1):013134, March 2010.
- [127] Jonathan F. Donges, Hanna C H Schultz, Norbert Marwan, Yong Zou, and Jürgen Kurths. Investigating the topology of interacting networks. *Eur. Phys. J. B*, 84(4):635–651, April 2011.
- [128] Marcelo Barreiro, Arturo C. Martí, and Cristina Masoller. Inferring long memory processes in the climate network via ordinal pattern analysis. *Chaos*, 21(1):013101, March 2011.
- [129] Jobst Heitzig, Jonathan F. Donges, Yong Zou, Norbert Marwan, and Jürgen Kurths. Node-weighted measures for complex networks with spatially embedded, sampled, or differently sized nodes. *Eur. Phys. J. B*, 85(1):38, January 2012.

## BIBLIOGRAPHY

---

- [130] J. Ignacio Deza, Marcelo Barreiro, and Cristina Masoller. Inferring interdependencies in climate networks constructed at inter-annual, intra-season and longer time scales. *Eur. Phys. J. Spec. Top.*, 222(2):511–523, June 2013.
- [131] Anastasios a. Tsonis, Kyle L. Swanson, and Sergey Kravtsov. A new dynamical mechanism for major climate shifts. *Geophys. Res. Lett.*, 34(13):1–5, July 2007.
- [132] Kazuko Yamasaki, Avi Gozolchiani, Shlomo Havlin, and Ramat Gan. Climate Networks Based on Phase Synchronization Analysis Track El-Niño. *Prog. Theor. Phys. Suppl.*, 179:178–188, 2009.
- [133] Avi Gozolchiani, Kazuko Yamasaki, Oz M. Gazit, and Shlomo Havlin. Pattern of climate network blinking links follows El Niño events. *Europhys. Lett.*, 83(2):28005, 2008.
- [134] Avi Gozolchiani and Oz M. Gazit. Dynamical Patterns of Climate Networks: Blinking Links and Stable Structures. *EGU Gen. Assem. . . .*, 11(2008):228501, 2009.
- [135] Jonathan F. Donges. *Complex networks in the climate system*. PhD thesis, University of Potsdam, 2009.
- [136] Igor I. Mokhov, Dmitry A Smirnov, Pavel I. Nakonechny, Sergey S. Kozlenko, Evgeny P. Seleznev, and Jürgen Kurths. Alternating mutual influence of El-Niño/Southern Oscillation and Indian monsoon. *Geophys. Res. Lett.*, 38(8):L00F04, April 2011.
- [137] Nishant Malik, Bodo Bookhagen, Norbert Marwan, and Jürgen Kurths. Analysis of spatial and temporal extreme monsoonal rainfall over South Asia using complex networks. *Clim. Dyn.*, 39(3-4):971–987, August 2011.
- [138] Jan H. Feldhoff, Reik V. Donner, Jonathan F. Donges, Norbert Marwan, and Jürgen Kurths. Geometric detection of coupling directions by means of inter-system recurrence networks. *Phys. Lett. A*, 376(46):3504–3513, October 2012.
- [139] Veronika Stolbova, Paul Martin, Bodo Bookhagen, Norbert Marwan, and Jürgen Kurths. Topology and seasonal evolution of the network of extreme precipitation over the Indian subcontinent and Sri Lanka. *Nonlinear Process. Geophys.*, 21(4):901–917, August 2014.
- [140] Alexander Radebach, Reik V. Donner, Jakob Runge, Jonathan F. Donges, and Jürgen Kurths. Disentangling different types of El Niño episodes by evolving climate network analysis. *Phys. Rev. E*, 88(5):052807, November 2013.
- [141] Yehiel Berezin, Avi Gozolchiani, Oded Guez, and Shlomo Havlin. Stability of climate networks with time. *Sci. Rep.*, 2:666, January 2012.
- [142] Karsten Steinhaeuser, Nitesh V. Chawla, and Auroop R. Ganguly. An Exploration of Climate Data Using Complex Networks. *ACM SIGKDD Explor. Newsl.*, 12(1):25–32, 2009.
- [143] Karsten Steinhaeuser, Auroop R. Ganguly, and Nitesh V. Chawla. Multivariate and multiscale dependence in the global climate system revealed through complex networks. *Clim. Dyn.*, 39(3-4):889–895, June 2011.
- [144] Oded Guez, Avi Gozolchiani, Yehiel Berezin, Steve Brenner, and Shlomo Havlin. Climate network structure evolves with North Atlantic Oscillation phases. *Europhys. Lett.*, 98(3):38006, May 2012.
- [145] Milan Paluš, Dennis L. Hartmann, Jaroslav Hlinka, and M. Vejmelka. Discerning connectivity from dynamics in climate networks. *Nonlinear Process. Geophys.*, 18(5):751–763, October 2011.

- 
- [146] Jaroslav Hlinka, David Hartman, Martin Vejmelka, Jakob Runge, Norbert Marwan, Jürgen Kurths, and Milan Paluš. Reliability of Inference of Directed Climate Networks Using Conditional Mutual Information. *Entropy*, 15(6):2023–2045, May 2013.
- [147] Elliot A. Martin, Maya Paczuski, and Jörn Davidsen. Interpretation of link fluctuations in climate networks during El Niño periods. *Europhys. Lett.*, 102(4):48003, May 2013.
- [148] Giulio Tirabassi and Cristina Masoller. On the effects of lag-times in networks constructed from similarities of monthly fluctuations of climate fields. *Europhys. Lett.*, 102(5):59003, June 2013.
- [149] Giulio Tirabassi, Cristina Masoller, and Marcelo Barreiro. Air-sea interaction in the South Atlantic Convergence Zone. *Int. J. Climatol.*, 2014.
- [150] Alexis Tantet and Henk A. Dijkstra. An interaction network perspective on the relation between patterns of sea surface temperature variability and global mean surface temperature. *Earth Syst. Dyn.*, 5(1):1–14, January 2014.
- [151] Mirjam van der Mheen, Henk a. Dijkstra, Avi Gozolchiani, Matthijs den Toom, Qingyi Feng, Jürgen Kurths, and Emilio Hernandez-Garcia. Interaction network based early warning indicators for the Atlantic MOC collapse. *Geophys. Res. Lett.*, 40(11):2714–2719, June 2013.
- [152] Daniel S. Wilks. *Statistical Methods in the Atmospheric Sciences*. Elsevier, 2011.
- [153] [http://commons.wikimedia.org/wiki/File:Correlation\\_examples.png](http://commons.wikimedia.org/wiki/File:Correlation_examples.png).
- [154] J. Ignacio Deza, Cristina Masoller, and Marcelo Barreiro. Distinguishing the effects of internal and forced atmospheric variability in climate networks. *Nonlinear Process. Geophys.*, 21(3):617–631, May 2014.
- [155] Arellano J. Gámez, Changsong Zhou, Axel Timmermann, and Jürgen Kurths. Nonlinear dimensionality reduction in climate data. *Nonlinear Process. Geophys.*, 11(3):393–398, September 2004.
- [156] Henk A. Dijkstra. *Data Analysis: A short Introduction*. Springer, 2011.
- [157] Michael B. Richman. Rotation of principal components. *J. Climatol.*, 6(3):293–335, January 1986.
- [158] Tao Lian and Dake Chen. An Evaluation of Rotated EOF Analysis and Its Application to Tropical Pacific SST Variability. *J. Clim.*, 25(15):5361–5373, August 2012.
- [159] Claude E. Shannon. A Mathematical Theory of Communication. *Bell Syst. Tech. J.*, 27(3):379–423, July 1948.
- [160] Michael Small. *Applied Nonlinear Time Series Analysis*. World Scientific, 2005.
- [161] Thomas M. Cover and Joy A. Thomas. *Elements of information theory*. John Wiley & Sons, 2012.
- [162] Alexander Kraskov, Harald Stögbauer, and Peter Grassberger. Estimating mutual information. *Phys. Rev. E*, 69(6):066138, June 2004.
- [163] Christopher J. Cellucci, Alfonso M. Albano, and Paul E. Rapp. Statistical validation of mutual information calculations: Comparison of alternative numerical algorithms. *Phys. Rev. E*, 71(6):066208, June 2005.

## BIBLIOGRAPHY

---

- [164] Shiraj Khan, Sharba Bandyopadhyay, Auroop R. Ganguly, Sunil Saigal, David Erickson, Vladimir Protopopescu, and George Ostrouchov. Relative performance of mutual information estimation methods for quantifying the dependence among short and noisy data. *Phys. Rev. E*, 76(2):026209, August 2007.
- [165] José M. Amigó and Matthew B. Kennel. Topological permutation entropy. *Phys. D*, 231(2):137–142, July 2007.
- [166] Bernd Pompe and Jakob Runge. Momentary information transfer as a coupling measure of time series. *Phys. Rev. E*, 83(5):051122, May 2011.
- [167] Christoph Bandt and Bernd Pompe. Permutation entropy: a natural complexity measure for time series. *Phys. Rev. Lett.*, 88(17):174102, April 2002.
- [168] Holger Kantz and Thomas Schreiber. *Nonlinear Time Series Analysis*. Cambridge University Press, 2004.
- [169] Jacques Martinerie, Claude Adam, Michel Le Van Quyen, Michel Baulac, Stéphane Clemenceau, Bernard Renault, and Francisco J. Varela. Epileptic seizures can be anticipated by non-linear analysis. *Nat. Med.*, 4(10):1173–6, October 1998.
- [170] Michel Le Van Quyen, Jacques Martinerie, Claude Adam, and Francisco J. Varela. Non-linear analyses of interictal EEG map the brain interdependences in human focal epilepsy. *Phys. D*, 127(3-4):250–266, March 1999.
- [171] Michael Rosenblum and Arkady Pikovsky. Detecting direction of coupling in interacting oscillators. *Phys. Rev. E*, 64(4):045202, September 2001.
- [172] M. Carmen Romano, Marco Thiel, Jürgen Kurths, and Celso Grebogi. Estimation of the direction of the coupling by conditional probabilities of recurrence. *Phys. Rev. E*, 76(3):036211, September 2007.
- [173] Milan Paluš, Vladimír Komárek, Zbyněk Hrnčíř, and Katalin Štěrbová. Synchronization as adjustment of information rates: Detection from bivariate time series. *Phys. Rev. E*, 63(4):046211, March 2001.
- [174] Martin Vejmelka. Inferring the directionality of coupling with conditional mutual information. *Phys. Rev. E*, 77(2):026214, February 2008.
- [175] Janez Jamšek, Milan Paluš, and Aneta Stefanovska. Detecting couplings between interacting oscillators with time-varying basic frequencies: Instantaneous wavelet bispectrum and information theoretic approach. *Phys. Rev. E*, 81(3):036207, March 2010.
- [176] Alireza Bahraminasab, Fatemeh Ghasemi, Aneta Stefanovska, Peter VE McClintock, and Holger Kantz. Direction of Coupling from Phases of Interacting Oscillators: A Permutation Information Approach. *Phys. Rev. Lett.*, 100(8):084101, February 2008.
- [177] Milan Paluš. From nonlinearity to causality: statistical testing and inference of physical mechanisms underlying complex dynamics. *Contemp. Phys.*, 48(6):307–348, 2007.
- [178] Aaron D. Wyner. A definition of conditional mutual information for arbitrary ensembles. *Inf. Control*, 38(1):51–59, July 1978.
- [179] Thomas Schreiber. Measuring Information Transfer. *Phys. Rev. Lett.*, 85(2):461–464, July 2000.
- [180] José M. Amigó. *Permutation Complexity in Dynamical Systems*. Springer Series in Synergetics. Springer Berlin Heidelberg, Berlin, Heidelberg, 2010.



- 
- [181] Ulrich Parlitz, Sebastian Berg, Stefan Luther, Alexander Schirdewan, Jürgen Kurths, and Niels Wessel. Classifying cardiac biosignals using ordinal pattern statistics and symbolic dynamics. *Comput. Biol. Med.*, 42(3):319–27, March 2012.
- [182] Jaroslav Hlinka, David Hartman, Martin Vejmelka, Dagmar Novotná, and Milan Paluš. Non-linear dependence and teleconnections in climate data: sources, relevance, non-stationarity. *Clim. Dyn.*, 42(7-8):1873–1886, May 2013.
- [183] Ronald A. Fisher. *Statistical methods for research workers*. Oliver and Boyd, 1925.
- [184] Martin Krzywinski and Naomi Altman. Points of significance: Significance, P values and t-tests. *Nat. Methods*, 10(11):1041–1042, October 2013.
- [185] Martin Krzywinski and Naomi Altman. Points of significance: error bars. *Nat. Methods*, 10(10):921–2, October 2013.
- [186] Martin Krzywinski and Naomi Altman. Points of significance: Power and sample size. *Nat. Methods*, 10(12):1139–1140, November 2013.
- [187] Sheldon M. Ross. *Introduction to probability and statistics for engineers and scientists*. Academic Press, 2009.
- [188] Rob J. Hyndman and Yanan Fan. Sample Quantiles in Statistical Packages. *Am. Stat.*, 50(4):361, November 1996.
- [189] Rand R. Wilcox. *Introduction to robust estimation and hypothesis testing*. Academic Press, 2012.
- [190] James Theiler, Stephen Eubank, André Longtin, Bryan Galdrikian, and J. Doayne Farmer. Testing for nonlinearity in time series: the method of surrogate data. *Phys. D*, 58(1-4):77–94, September 1992.
- [191] Thomas Schreiber and Andreas Schmitz. Improved Surrogate Data for Nonlinearity Tests. *Phys. Rev. Lett.*, 77(4):635–638, July 1996.
- [192] James Theiler, Dean Prichard, Radiation Measurements Group, and Los Alamos National. Constrained-realization Monte-Carlo method for hypothesis testing. *Phys. D*, 94:221–235, 1996.
- [193] Hans von Storch. Misuses of Statistical Analysis in Climate Research. In Hans von Storch and Antonio Navarra, editors, *Anal. Clim. Var.*, pages 11–26. Springer Berlin Heidelberg, Berlin, Heidelberg, 1999.
- [194] Milan Paluš and Dagmar Novotná. Testing for nonlinearity in weather records. *Phys. Lett. A*, 193(September):67–74, 1994.
- [195] Reeda Kunhimangalam, Paul K. Joseph, and Sujith Owallath. Nonlinear analysis of EEG signals: Surrogate data analysis. *Irbm*, 29(4):239–244, September 2008.
- [196] Bradley Efron and Robert J. Tibshirani. *An introduction to the bootstrap*. Chapman and Hall, 1994.
- [197] Manfred Mudelsee. *Climate Time Series Analysis*, volume 42 of *Atmospheric and Oceanographic Sciences Library*. Springer Netherlands, Dordrecht, 2014.
- [198] John C. H. Chiang and Adam H. Sobel. Tropical Tropospheric Temperature Variations Caused by ENSO and Their Influence on the Remote Tropical Climate\*. *J. Clim.*, 15(18):2616–2631, September 2002.

## BIBLIOGRAPHY

---

- [199] Marcelo Barreiro, Ping Chang, Link Ji, Ramalingam Saravanan, and Alessandra Giannini. Dynamical elements of predicting boreal spring tropical Atlantic sea-surface temperatures. *Dyn. Atmos. Ocean.*, 39(1-2):61–85, April 2005.
- [200] Qinghua Ding, Eric J. Steig, David S. Battisti, and Marcel Küttel. Winter warming in West Antarctica caused by central tropical Pacific warming. *Nat. Geosci.*, 4(6):398–403, April 2011.
- [201] David M. Straus and Jagadish Shukla. Distinguishing between the SST-forced variability and internal variability in mid latitudes: Analysis of observations and GCM simulations. *Q. J. R. Meteorol. Soc.*, 126(567):2323–2350, August 2010.
- [202] Marcelo Barreiro, Ping Chang, and Ramalingam Saravanan. Variability of the South Atlantic Convergence Zone Simulated by an Atmospheric General Circulation Model. *J. Clim.*, 15(7):745–763, April 2002.
- [203] Franco Molteni. Atmospheric simulations using a GCM with simplified physical parametrizations. I: model climatology and variability in multi-decadal experiments. *Clim. Dyn.*, 20(2-3):175–191, 2003.
- [204] Annalisa Bracco, Fred Kucharski, Rameshan Kallummal, and Franco Molteni. Internal variability, external forcing and climate trends in multi-decadal AGCM ensembles. *Clim. Dyn.*, 23(6):659–678, July 2004.
- [205] Fernando Arizmendi, Arturo C. Martí, and Marcelo Barreiro. Evolution of atmospheric connectivity in the 20th century. *Nonlinear Process. Geophys.*, 21:825–839, 2014.
- [206] Ian N. James. *Introduction to circulating atmospheres*. Cambridge University Press, 1994.
- [207] Jagadish Shukla. Predictability in the Midst of Chaos: A Scientific Basis for Climate Forecasting. *Science (80-. )*, 282(5389):728–731, October 1998.
- [208] Thomas M. Smith and Richard W. Reynolds. Improved extended reconstruction of SST (1854-1997). *J. Clim.*, 2004.
- [209] Claude Frankignoul and Klaus Hasselmann. Stochastic climate models, Part II Application to sea-surface temperature anomalies and thermocline variability. *Tellus*, 29(4):289–305, August 1977.
- [210] Joseph J. Barsugli and David S. Battisti. The Basic Effects of Atmosphere-Ocean Thermal Coupling on Midlatitude Variability. *J. Atmos. Sci.*, 55(4):477–493, 1998.
- [211] Kevin E. Trenberth, Grant W. Branstator, David J. Karoly, Arun Kumar, Ngar-Cheung Lau, and Chester F. Ropelewski. Progress during TOGA in understanding and modelling global teleconnections associated with tropical sea surface temperatures. *J. Geophys. Res.*, 103(C/):14291–14324, 1998.
- [212] Marcelo Barreiro and Nicolas Díaz. Land-atmosphere coupling in El Niño influence over South America. *Atmos. Sci. Lett.*, 12(4):351–355, October 2011.
- [213] Holger Pohlmann. Atlantic versus Indo-Pacific influence on Atlantic-European climate. *Geophys. Res. Lett.*, 32(5):L05707, 2005.
- [214] Richard Seager, Naomi Naik, Walter Baethgen, Andrew W. Robertson, Yochanan Kushnir, Jennifer Nakamura, and Stephanie Jurburg. Tropical Oceanic Causes of Interannual to Multidecadal Precipitation Variability in Southeast South America over the Past Century\*. *J. Clim.*, 23(20):5517–5539, October 2010.

- [215] Myles R. Allen and Leonard A. Smith. Optimal filtering in singular spectrum analysis. *Phys. Lett. A*, 234(6):419–428, 1997.
- [216] Stephan Venzke, Myles R. Allen, Rowan T. Sutton, and David P. Rowell. The atmospheric response over the North Atlantic to decadal changes in sea surface temperature. *J. Clim.*, 12:2562–2584, 1999.
- [217] Mingfang Ting, Yochanan Kushnir, Richard Seager, and Cuihua Li. Forced and Internal Twentieth-Century SST Trends in the North Atlantic\*. *J. Clim.*, 22(6):1469–1481, March 2009.
- [218] Fred Kucharski, Franco Molteni, and Annalisa Bracco. Decadal interactions between the western tropical Pacific and the North Atlantic Oscillation. *Clim. Dyn.*, 26(1):79–91, November 2005.
- [219] Chester F. Ropelewski and Michael S. Halpert. North American Precipitation and Temperature Patterns Associated with the El Niño/Southern Oscillation (ENSO). *Mon. Weather Rev.*, 114(12):2352–2362, December 1986.
- [220] Ping Chang, Ramalingam Saravanan, Link Ji, and Gabriele C Hegerl. The effect of local sea surface temperatures on atmospheric circulation over the tropical Atlantic sector. *J. Clim.*, 13(1985):2195–2216, 2000.
- [221] Klaus Hasselmann. Stochastic climate models Part I. Theory. *Tellus*, 28(6):473–485, December 1976.
- [222] Kingtse C. Mo and Wayne Higgins. The Pacific–South American Modes and Tropical Convection during the Southern Hemisphere Winter. *Mon. Weather Rev.*, 126(6):1581–1596, June 1998.
- [223] Hanna Annamalai, Ragu Murtugudde, James T Potemra, Shang-Ping Xie, Philip L-F Liu, and Bin Wang. Coupled dynamics over the Indian Ocean: spring initiation of the Zonal Mode. *Deep Sea Res. Part II Top. Stud. Oceanogr.*, 50(12-13):2305–2330, July 2003.
- [224] Xin Wang and Chunzai Wang. Different impacts of various El Niño events on the Indian Ocean Dipole. *Clim. Dyn.*, 42(3-4):991–1005, February 2014.



# ***Publications***

---

Publications related to the work presented in this thesis.

- Deza, J. Ignacio, Marcelo Barreiro, and Cristina Masoller. “Inferring Interdependencies in Climate Networks Constructed at Inter-Annual, Intra-Season and Longer Time Scales.” *The European Physical Journal Special Topics* 222.2 (2013): 511–523.
- Deza, J. Ignacio, Cristina Masoller, and Marcelo Barreiro. “Distinguishing the Effects of Internal and Forced Atmospheric Variability in Climate Networks.” *Nonlinear Processes in Geophysics* 21.3 (2014): 617–631.
- Deza, J. Ignacio, Marcelo Barreiro, and Cristina Masoller. “Assessing the direction of climate interactions by means of complex networks and information theoretic tools.” Submitted to *Chaos* (2014).

Additional work not related to this thesis

- Deza, J. Ignacio et al. “Wide-Spectrum Energy Harvesting out of Colored Levy-like Fluctuations, by Monostable Piezoelectric Transducers.” *Europhys. Lett.* 100.3 (2012): 38001.



# ***Presentations at International Conferences***

---

- IBERSINC network conference (Zaragoza, 20-21 mar 2012), poster presentation.  
Poster title: “Detecting long-range teleconnections in the climate network via ordinal pattern time-series analysis”  
<http://www.ibersinc.org>
- 76th annual meeting of the Deutsche Physikalische Gesellschaft (DPG) (25-30 mar 2012), poster presentation  
Poster title: “Detecting long-range teleconnections in the climate network via ordinal pattern time-series analysis”  
[http://www.dpg-physik.de/veranstaltungen/tagungen/tagung\\_2012.html](http://www.dpg-physik.de/veranstaltungen/tagungen/tagung_2012.html)
- European Geosciences Union (EGU) general assembly (22-27 apr 2012), poster presentation.  
Poster title: “Detecting long-range teleconnections in the climate network via ordinal pattern time-series analysis”  
<http://www.egu2012.eu>
- NOLINEAL 2012 conference (Zaragoza 4-6 jun 2012), poster presentation.  
Poster title: “Climate networks at inter-annual and intra-season time scales constructed via ordinal time-series analysis”.  
<http://neptuno.unizar.es/jgg/nolineal2012/>
- Data Analysis and Modeling in Earth Sciences (DAMES - 2012) (Potsdam, 8-10 oct 2012), oral presentation.

- Presentation title: “Detecting long-range teleconnections in the climate network via ordinal pattern time-series analysis”.
- <http://dames.pik-potsdam.de>
- Spanish Statistical physics conference “FisEs” (UIB Mallorca, 18-20 oct 2012), poster presentation.  
Poster title: “Detecting long-range teleconnections in the climate network via ordinal pattern time-series analysis”.  
<http://www.gefenol.es/FisEs/12/>
  - Nonlinear Theory and its Applications (NOLTA - 2012) (Palma de Mallorca, 22-26 Oct 2012), oral presentation at the “Networks of Networks and their Applications II”, session.  
Presentation title: “Inferring Interdependencies in Climate Networks Constructed at Inter-Annual, Intra-Season and Longer Time Scales”.  
<http://nolta2012.org>
  - European Conference on Complex Systems (ECCS’13), (Barcelona, Spain, 16 - 20 September 2013), poster presentation.  
Poster title: “Internal and forced atmospheric variability by means on climate networks”  
<http://www.eccs13.eu/>
  - WCRP Conference for Latin America and the Caribbean: Developing, linking and applying climate knowledge , (Montevideo, Uruguay, March 17-21, 2014), poster presentation.  
Poster title: “Constructing climate networks: variability types, time scales and directionality”  
Winner of the best poster award.  
<http://www.cima.fcen.uba.ar/WCRP/>
  - European Geosciences Union (EGU) General Assembly 2014 (Vienna, Austria, 27 April-02 May 2014), poster presentation.  
Poster title: “Constructing climate networks: variability types, time scales and direc-



tionality”

<http://www.egu2014.eu/>

- The 10th AIMS Conference on Dynamical Systems, Differential Equations and Applications (Madrid, Spain 07-11 July, 2014), oral presentation.

Presentation title: “Interdependencies in Climate Networks Constructed using Information Measures at different Time Scales”.

<http://www.aims.org/conferences/2014/>

- European conference in complex systems 2014 (ECCS’14) (Lucca, Italy 22-26 Sept, 2014), poster presentation.

Poster title: “Directionality in climate networks: an information-theoretic approach”

The conference included a satellite event organized by LINC on Thursday, 25 Sept 2014, where I gave a 2 minute presentation.

<http://www.eccs14.eu/index.php?lang=en>



# ***Attendance to schools and research stays***

---

- Attendance to several talks from the Universitat Politècnica de Catalunya (Terrassa, 2012):
  - “Per qué cal patentar? La importància de les patents en la recerca universitària” (Why do you need to patent? The importance of patents in university research). (UPC, Terrassa)
  - “UPC. Universitat emprenedora” (UPC. enterprising univesity) cycle of talks. (UPC, Terrassa)
  - “Tècniques d’escriptura científica (Scientific Writing Skills ) (EET, UPC, Terrassa)”
- 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 (Barcelona 4-6 set 2012).  
<http://www.ibersinc.org>
- 1st LINC School and Workshop (10 - 13 sep 2012).  
<http://www.climatelinc.eu/events/1st-linc-school-10-12-sept-2012-mallorca/>
- 2nd LINC School (Soesterberg, The Netherlands 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>
- Max Plank “Scales and patterns in the earth system” Workshop (Dresden, 05-09 Nov 2012).  
<http://www.mpipks-dresden.mpg.de/earth12/>
- Research stay in Universidad de la República, (Montevideo, Uruguay, 1 May-31 July 2013).  
<http://meteo.fisica.edu.uy/>
- 21 - 30 October 2013 "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.  
<http://indico.ictp.it/event/a12220/>
- 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/>
- 4th LINC Workshop, (Montevideo, Uruguay, 24-26 March, 2014).  
<http://www.climatelinc.eu/events/linc-workshop-4-march-2014-montevideo-uruguay/>
- Company internship in Ambrosys (Potsdam, Germany, 27 March-27 April 2014)  
<http://www.ambrosys.de/>
- 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](http://www.to.isac.cnr.it/aosta_old/aosta2014/index.htm)
- 5th LINC Workshop, (Lucca, Italy 27-29 Sept, 2014).  
<http://www.climatelinc.eu/events/eccs-14-and-linc-workshop-5-sept-2014-/lucca-italy/>