CAUSAL INFERENCE AND FORECASTING METHODS FOR CLIMATE DATA ANALYSIS

RICCARDO SILINI

Ph.D. Thesis

Nonlinear Dynamics, Nonlinear Optics and Lasers group (DONLL)
Department of Physics
Universitat Politecnica de Catalunya

Supervisor: Prof. Cristina Masoller
Co-supervisor: Prof. Marcelo Barreiro

June 2022

Riccardo Silini: *Causal inference and forecasting methods for climate data analysis,* Ph.D. Thesis, © June 2022

Well, maybe it started that way. As a dream, but doesn’t everything? Those buildings. These lights. This whole city. Somebody had to dream about it first. And maybe that is what I did. I dreamed about coming here, but then I did it.

— James and the Giant Peach

Dedicated to my loved ones, living by my side, living in me.

And above all, watch with glittering eyes the whole world around you because
the greatest secrets are always hidden in the most unlikely places.
Those who don't believe in magic will never find it.

– Billy and The Minpins.

ABSTRACT

To advance time series forecasting we need to progress on multiple fronts. In this thesis, we develop algorithms to identify causal relations which allow to identify the driving processes containing useful information for the prediction of the process of interest. Complementing this, machine learning algorithms allow to exploit such information to build data-driven forecast models, and to correct dynamical models.

The identification from time series analysis of reliable indicators of causal relationships, is essential for many disciplines. Main challenges are distinguishing correlation from causality and discriminating between direct and indirect interactions. Over the years, many methods for data-driven causal inference have been proposed; however, their success largely depends on the characteristics of the system under investigation. Often, their data requirements, computational cost or number of parameters, limit their applicability. In this thesis, we propose a computationally efficient measure for causality testing, with the goal of overcoming the applicability limitations of information-theoretic measures, due their high computational cost. The proposed metric is very valuable when causality networks need to be inferred from the analysis of a large number of relatively short time series. It can also be very useful for the inputs selection of machine learning algorithms; in fact, it allows to identify those processes which contain useful information for the prediction of a given process. This feature is particularly useful for systems composed of a large number of processes, whose interactions are poorly understood.

The socioeconomic impact of weather extremes draws the attention of researchers to the development of novel methodologies to make more accurate weather predictions. The Madden-Julian Oscillation (MJO) is the dominant mode of variability in the tropical atmosphere on sub-seasonal time scales, and can promote or enhance extreme events in both, the tropics and the extratropics. Currently, the estimated MJO predictability is far from being reached, leaving a large room for the improvement of forecast models. To improve its prediction skill, in this thesis we take two different machine learning approaches; first we use machine learning as a stand-alone technique, showing that two artificial neural networks, a feed-forward neural network and a recurrent neural network, allow a competitive prediction, yet not exceeding the skill of the state-of-art dynamical models. Then, we combine dynamical models with machine learning, which allows to improve the predictions of the best dynamical model. In particular, machine learning allows to improve the prediction of the events intensity and geographical localization.

RESUMEN

Para avanzar en el pronóstico de series temporales, es necesario avanzar en múltiples frentes. En esta tesis, desarrollamos algoritmos para descubrir relaciones causales que identifican los procesos que actúan como fuentes potenciales de información y pueden ayudar a mejorar la predicción del proceso de interés. En complementación a esto, los algoritmos de aprendizaje automático, permiten explotar dicha información para construir modelos de pronóstico basados en la observación de datos para, de esta forma, corregir los modelos dinámicos. La identificación de indicadores fiables de relaciones de causalidad a partir de series temporales es esencial en muchas disciplinas. Los principales desafíos en este ámbito se encuentran en distinguir la correlación de la causalidad, así como diferenciar entre las interacciones directas e indirectas. A lo largo de los años, se han propuesto numerosos métodos de inferencia causal basados en la observación de datos. No obstante, su éxito depende enormemente de las características del sistema a investigar. A menudo, los requisitos de sus datos, el coste computacional o el número de parámetros limitan su aplicabilidad. En esta tesis, se propone una medida computacionalmente eficiente para el testeo de causalidad, con el fin de solucionar las limitaciones de aplicabilidad de las medidas teóricas de la información, debido a su alto coste computacional. La métrica que se propone resulta ser muy valiosa cuando las redes neuronales de causalidad necesitan inferirse a partir de análisis de un gran número de series temporales relativamente cortas. También puede resultar muy útil en la selección de entradas en los algoritmos de machine learning. De hecho, permite identificar aquellos procesos que contengan información útil en la predicción de cierto proceso dado. Esta característica es particularmente útil para sistemas compuestos por un gran número de procesos, cuyas interacciones son escasamente conocidas. El impacto socioeconómico de los fenómenos meteorológicos extremos llama la atención a los investigadores en el desarrollo de nuevas metodologías con el objetivo de obtener predicciones meteorológicas más precisas. La Oscilación de Madden-Julian (MJO) es el modo dominante de variabilidad en la atmósfera tropical en escalas temporales subestacionales, y puede promover o aumentar eventos extremos tanto en el trópico como el extratrópico. Actualmente, la prediccion de la MJO está lejos de alcanzarse, lo que deja un gran margen de mejora en los modelos de pronóstico. Para mejorar su habilidad de predicción, en esta tesis, se escogerán dos aproximaciones diferentes de aprendizaje automático. Primero, se usará el machine learning como una técnica independiente, mostrando que dos redes neuronales artificiales, una red neuronal feed-forward y una red neuronal recurrente, permiten una predicción competitiva, pero sin superar la habilidad de los modelos dinámicos actuales. Posteriormente, se combinarán modelos dinámicos con machine learning, que permitirán mejorar las predicciones del mejor modelo dinámico. En particular, el aprendizaje automático permite mejorar la predicción de la intensidad de los eventos y, así como su localización geográfica.

One of the secrets of life is that
all that is really worth the doing
is what we do for others.

— Lewis Carroll.

ACKNOWLEDGEMENTS

It has been a long journey, but I walked alongside with awesome people.

The first person that I want to thank here is Cristina. Cristina, my supervisor, not "only" allowed me to achieve my professional goals, she did much more. She guided me, yet giving me the freedom to choose what to work on based on my preference and motivation; she trusted me and my intuitions, yet making me ponder my ideas; she always did what is best for me, allowing me to adapt my secondments for my interests, and for my future; she strongly improved the quality of my work, and the quality of this thesis; sometimes, I feel like a river, I want to explore, branch out, and Cristina played the role of the riverbank, letting me branch out, yet making me focus on properly defined paths, to not make me overflow, and stop me at the right moment to give me the time to look behind, and properly frame my work. I am very grateful for all of this, and I wish all Ph.D. students to have the luck I had.

Talking about supervisors, I also wish to thank Marcelo, my co-supervisor. Although far, he has been close. He enriched me with his expertise on climate, his kindness and helpfulness. In all CAFE workshops we met, I had a great time talking with him. I would like to thank Holger Kantz, who hosted me during my secondment at the Max Planck Institute in Dresden and has always been very kind and helpful. A big thank also to Daniel San Martin, CEO of Predictia, who hosted me in Santander for my secondment, and gave me the opportunity to experience the work in a company outside the academic world. I had a great time during my secondments.

I wish to show my appreciation to the ITN MSCA CAFE project for giving me the opportunity to conduct my Ph.D, and all the CAFE family. In particular, I wish to acknowledge all the ESRs: Iago Perez, Monica Minjares, Noemie Ehstand, Pedro Herrera Lormendez, Xinjia Hu, Emmanuel Rouges, Amal John, Shraddha Gupta, Niclas Rieger, Nikos Mastrantonas, and Meriem Krouma, which have been like a family, always supporting and helping each other, collaborating, and having a great time together. I also acknowledge all the supervisors and collaborators that made this project possible, and helped us becoming better researchers with their expertise.

I would like to thank the seniors of the DONLL group of UPC in Terrassa: Cristina, Jordi, Crina, Ramon Herrero, Ramon Vilaseca, Muriel, Antonio, Carles, Kestutis, José.

A special thank to my past and present colleagues, Pablo, Donatus, Hossam, Sandeep, Judith, Salim, Tetsu, Shroddha, Nayeem, Alex, Guillermo. In particular, a special thank goes to Maria Masoliver, Maria Duque Gijón (Mexican Team), Laura, and Giulio, which are the colleagues with whom I spent most of the time, and supported me the most. Everything has been better thanks to you!

A big thank to my close friends for making me feel home even far from it. Thanks to Jordi, Carluz, Blasco, Jo, Luca, Vale, Mili.

Now, I would like to spend some words for my family. Grazie mamma e papà. Grazie Andy. È sempre difficile scrivere a parole sentimenti cosí forti. Un'unica parola come 'grazie', può portare dentro di sè un grande spettro di sentimenti. Io vi devo tutto, e questo 'grazie' è il più profondo. Grazie alla mia ragazza Stefania. Voi mi avete sempre lasciato la libertà di decidere per la mia vita, di perseguire i miei sogni, e fare ciò che più mi rende felice. Tutto ciò restando al mio fianco, per consigliarmi, supportarmi e sopportarmi. Non mi avete mai fatto sentire solo, perché il vostro amore non conosce confini. Grazie Rina per tutto quello che fai, e grazie Paola. Grazie a tutti i nonni, agli zii e ai cugini, questa tesi è anche per i piccoli della nuova generazione, Robin, Alessandro e Matteo.

Finally, thanks to COVID-19 with all its variants, which made me spend more time with my family.

CONTENTS

I INTRODUCTION 1

1 INTRODUCTION 3

1.1 Motivation . . . 3

1.2 Objectives . . . 5

1.3 Outline . . . 6

2 CORRELATION AND CAUSALITY 9

2.1 Correlation . . . 9

2.1.1 Pearson correlation . . . 9

2.1.2 Spearman correlation . . . 10

2.1.3 Cross-correlation . . . 10

2.1.4 Mutual information . . . 11

2.2 Causality . . . 12

2.2.1 Granger causality . . . 14

2.2.2 Transfer Entropy . . . 14

2.2.3 Other approaches . . . 15

3 THE MADDEN-JULIAN OSCILLATION 17

3.1 The phenomenon . . . 17

3.2 MJO impact . . . 17

3.3 MJO indices . . . 20

3.4 MJO forecast . . . 20

4 MACHINE LEARNING 25

4.1 Supervised, unsupervised, and reinforcement learning . . 25

4.2 Artificial Neural Networks . . . 26

4.3 Applications in climate . . . 27

4.3.1 Prediction of the MJO . . . 28

II PSEUDO TRANSFER ENTROPY 29

5 PSEUDO TRANSFER ENTROPY 31

5.1 Definition . . . 31

6 APPLICATION TO SYNTHETIC DATA 37
6.1 Models 37
6.2 Statistical significance 38
6.3 Implementation 39
6.4 Results 40
6.5 Discussion 51
7 APPLICATION TO CLIMATE INDICES 53
7.1 Monthly atmospheric and Ocean time series 53
7.2 Methodology 57
7.2.1 Statistical significance analysis 58
7.2.2 Long-term causality variation 59
7.2.3 Results 60
7.3 Discussion 62
7.4 Conclusions 67
III PREDICTION OF THE MADDEN-JULIAN OSCILLATION 69
8 PREDICTION OF THE MJO FROM OBSERVATIONS 71
8.1 Data set 71
8.2 Prediction skill metrics 71
8.3 Results 73
8.3.1 Prediction skill 73
8.3.2 Seasonally resolved prediction skill 75
8.4 Discussion 77
9 IMPROV. THE PRED. OF THE MJO BY POST-PROCESSING 83
9.1 RMM Data 83
9.2 ECMWF RMM reforecasts 83
9.3 Results 84
9.4 Discussion 89
IV CONCLUSIONS 93
10 CONCLUSIONS 95
V APPENDIX 99
A AUTOREGRESSIVE MODELS 101
A.1 Autoregressive model AR 101

A.2 Moving average model MA 102
A.3 Autoregressive moving average model ARMA 102
A.4 Autoregressive integrated moving average model ARIMA 102
A.5 Akaike and Bayesian information criteria 103
B SIGNIFICANCE TESTS 105
B.1 F test 105
B.2 Surrogates 106
B.2.1 Independent Gaussian processes 107
B.2.2 Iterative amplitude adjusted Fourier transform (IAAFT) surrogates 107
B.2.3 Time-shifted surrogates 109
C PTE ON SYNTHETIC DATA: SUPPLEMENTARY RESULTS 111
D ARTIFICIAL NEURAL NETWORKS 115
D.1 MJO forecast 115
D.2 MJO forecast post-processing 117
BIBLIOGRAPHY 121
RESEARCH ACTIVITIES 141

ACRONYMS

AAFT	Amplitude adjusted Fourier transform
AIC	Akaike information criterion
AMO	Atlantic multidecadal oscillation
ANN	Artificial neural network
AR	Autoregressive
AR-RNN	Autoregressive recurrent neural network
ARIMA	Autoregressive integrated moving average
ARMA	Autoregressive moving average
BIC	Bayesian information criterion
CAFE	Climate advanced forecasting of sub-seasonal extremes
CCM	Convergent cross mapping
CMI	Conditional mutual information
COR	Bivariate correlation coefficient
DJF	December-January-February
DL	Deep learning
DGP	Data generating process
ECMWF	European centre for medium-range weather forecasts
ENSO	El Niño-Southern oscillation
EOFs	Empirical orthogonal functions
FFNN	Feedforward neural network
FFT	Fast Fourier transform

GC	Granger causality
GMT	Global mean temperature
GRU	Gated recurrent unit
HURR	Atlantic hurricanes index
IAAFT	Iterative amplitude adjusted Fourier transform
IC	Information criterion
IPCC	Intergovernmental panel on climate change
ISM	Indian summer monsoon
JJA	June-July-August
JJASON	June-July-August-September-October-November
LSTM	Long-short term memory
MAE	Mean absolute error
MAM	March-April-May
MC	Maritime continent
MI	Mutual information
MJO	Madden-Julian oscillation
ML	Machine learning
MSE	Mean squared error
MSLE	Mean squared logarithmic error
NAO	North Atlantic oscillation
NCAR	National center for atmospheric research
NCEP	National centers for environmental prediction
NOAA	National oceanic and atmospheric administration
NP	North Pacific pattern
NTA	North Tropical Atlantic index

OLR	Outgoing longwave radiation
PDO	Pacific decadal oscillation
pTE	pseudo transfer entropy
QBO	Quasi-biennial oscillation
ReLU	Rectified linear unit
RMM	Real-time multivariate MJO
RMSE	Bivariate root-mean-square error
RNN	Recurrent neural network
Sahel	Sahel standardized rainfall
SOI	Southern oscillation index
SON	September-October-November
SST	Sea-surface temperature
TE	Transfer entropy
T-S	Time-shifted
TSA	Tropical Southern Atlantic index

Part I

INTRODUCTION

1 INTRODUCTION

1.1 MOTIVATION

Humans started to physically store information through prehistoric wall paintings between 43000 to 65000 years ago. As we evolved, emerging languages and the invention of writing, culminating in the invention of paper in China about 2000 years ago, gave humankind extremely powerful tools to store information. Since the appearance of the first printed book in 1377, and the Gutenberg movable type in 1453-1455, books have been the most used support for information storage for centuries. The invention of the transistor in 1947 has been a crucial turning point for humankind also for storing information, and since 1996, digital storage of data became more cost-effective than the paper one, making it the largely preferred support. Over the last 30 years, technology allowed to reduce the average cost per gigabyte, from hundreds of thousand dollars of the first hard drives, to just fractions of a cent with the Cloud. From occupying an entire room, to fit on a fingertip.

At present, thanks to such technical evolution and progress, humans are collecting and storing an amount of data as never before, and the collection rate is exponentially increasing over the years. According to Social Media Today, it is estimated that every day 2.5 exabytes (10^{18}) are collected, which are estimated to grow up to 463 exabytes in 2025 (Raconteur, 2019). The total amount of data collected up to the beginning of 2020, was estimated to be 44 zettabytes (10^{21}), which is 40 times larger than the number of stars in the observable universe.

With huge amounts of data to analyze, computationally efficient tools are needed.

A field largely benefiting from this evolution is climate. Every day across the globe we are collecting data on temperature, pressure, wind, rain, snow, and much more. These data allow us to improve our understanding of climate, its phenomena and the interactions between them, to reconstruct the past, and predict the future. From these data, we define climate indices that characterize large-scale climate phenomena, validate models, and forecast the weather.

Extreme weather events such as cyclones, hurricanes, droughts, floods, wild fires, cold and warm spells, have huge socio-economic impact. The World Health Organisation has estimated that worldwide, climate extremes cost around 150'000 lives every year (Patz et al., 2005), and hundreds of billion dollars (Kramer and Ware, 2021). According to the lattest IPCC report (IPCC, 2022), climate change is expected to increase the rate and severity of extremes, suggesting a future increment of those, already frightening, numbers. It is crucial then to progress in the forecasting of such events, to be able to provide reliable early warnings that can save lives.

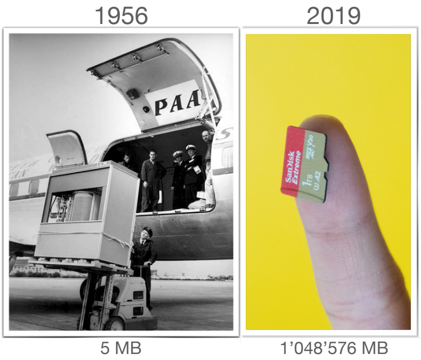

Figure 1: Left: 5MB IBM 305 RAMAC (Photo: IBM). Right: SanDisk Extreme 1TB microSD card (Credit: Raymond Wong / Mashable)

1.2 OBJECTIVES

In this thesis, part of the *Climate Advanced Forecasting of sub-seasonal Extremes (CAFE)* project, we aim at improving the sub-seasonal predictability (from about 10 days to 3 months) of extreme weather events. Predicting such events is very challenging due to the poor understanding of the phenomena acting at this time scale, such as the *Madden-Julian oscillation (MJO)* (Madden and Julian, 1994; Madden and Julian, 1971, 1972), planetary waves and atmospheric blockings, to cite a few.

To improve sub-seasonal predictability, in this dissertation we have two main goals: first, to develop a generic approach to infer causality from data that is computationally cost-effective; second, to improve the prediction of the MJO, which is a main source of predictability at the sub-seasonal scale.

Identifying, from time series analysis, reliable indicators of causal relationships is important in all fields of science and technology. Main challenges are distinguishing correlation from causality and discriminating between direct and indirect interactions. Over the years, many methods for data-driven causal inference have been proposed (Granger, 1969; Paluš and Vejmelka, 2007; Schreiber, 2000; Sugihara et al., 2012); however, their success largely depends on the characteristics of the system under investigation. Often, their data requirements, computational costs, or number of parameters, limit their applicability. In this thesis we propose a computationally efficient measure for causality testing, which we refer to as pseudo transfer entropy (pTE). We apply this metric on both synthetic and real data, showing its strengths and weaknesses.

Regarding the second objective, the MJO is the dominant mode of variability in the tropical atmosphere on sub-seasonal time scales, and can promote or enhance extreme events in both, the tropics and the extratropics (Ferranti et al., 2018; Lau and Waliser, 2011; Vitart, 2009; Zhang et al., 2013). In this thesis, we show the MJO prediction skill achieved using different machine learning models. Moreover, we build artificial neural networks to improve the best state-of-the-art climate model's predictions through post-processing.

1.3 OUTLINE

In the following chapters of Part I, we introduce the main concepts used in the studies presented in Part II and III. In particular, in Chapter 2 we present metrics to compute correlation and causality, which will be the starting point for the results obtained in Part II. Afterwards, we will introduce the notions needed for Part III; in Chapter 3 we introduce the MJO, its impact, its characterization, and the state-of-the-art of its forecast; in Chapter 4 we introduce the basic concepts of machine learning, some of its algorithms, and its applications on climate, with a particular focus in the prediction of the MJO.

In Part II, we present our contribution to causal data analysis, with the presentation of the pseudo transfer entropy and its applications.

In Chapter 5, we present the mathematical derivation of the pTE from the TE, a concept that we presented in Chapter 2.

In Chapter 6, we apply pTE to synthetic data. We first introduce then the data generating processes, and the inferred causality with statistical significance. We compare our results with the literature, we showcase an application on real data and we conclude the chapter with a general discussion of the results. The results presented in this chapter were published in Silini and Masoller (2021).

In Chapter 7, we adopt the pTE to unveil interactions between a selection of climate indices, to build causality networks. We also explore how the interactions between those indices changed across decades.

In Part III, we present our contribution to the prediction of the MJO.

In Chapter 8, we show the machine learning prediction of the MJO, its prediction skill, the phase and amplitude errors, and how the seasons and initial MJO phases influence the predictions. The results presented in this chapter were published in Silini, Barreiro, and Masoller (2021).

In Chapter 9, we correct the predictions of MJO obtained using the current best weather model, using artificial neural networks. In particular, we show how machine learning manages to improve the MJO amplitude and phase error, better than a linear post-processing. We also show how the improvement depends on the initial MJO phase, and that machine learning is helpful to overcome the Maritime continent barrier (presented in Chapter 3).

In Part IV we present the final conclusions of the thesis, and the future perspectives.

Finally, Part V is devoted to the Appendix. In this last part we present the autoregressive models, which are used to determine a parameter of the pTE in Part II, and as baseline in Part III. Then, we cover the significance testing, presenting the surrogates that are used to assess significance of the causality, and the supplementary results of Silini and Masoller (2021), where we apply the pTE on several data generating processes. Finally, we include the architecture details of the artificial neural networks employed in Part III.

CORRELATION AND CAUSALITY 2

2.1 CORRELATION

Unveiling and quantifying the strength of interactions from the analysis of observed data is a problem of capital importance for real-world complex systems. Typically, the details of the system are not known, but only observed time series are available, often short and noisy.

One way to evaluate the degree of association between time series is to compute the correlation. There are different types of correlation measures, and in this chapter we will present the most commonly found in literature: *Pearson correlation*, *Spearman correlation*, *cross-correlation*, and *mutual information*.

2.1.1 *Pearson correlation*

To find a linear relationship between two data sets, a very common measure of correlation is the *Pearson product moment correlation (PPMC)*, usually called Pearson correlation in short. This measure requires particular attention when applied, since it does not distinguish between dependent and independent variables. The Pearson correlation coefficient r between two variables X and Y is given by

$$r = \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_i (x_i - \bar{x})^2 \sum_i (y_i - \bar{y})^2}}, \tag{1}$$

where x_i and y_i are the values in the samples, while $\bar{x}$ and $\bar{y}$ are the mean of the values of the X and Y variable, respectively. The coefficient

r will be equal to 0 for independent variables, and takes a value of 1 for a perfect linear relationship. In some case, the relationship between two variables does not change at a constant rate, although being monotonic. This lead to a low Pearson coefficient, while the two variables are actually correlated. To solve this issue, one may use the Spearman correlation, described in the following section.

2.1.2 *Spearman correlation*

To find a monotonic relationship between two data sets, we can use the Spearman correlation. Differently from the Pearson correlation, the relationship between the two variables can also be nonlinear (but monotonic), and it is a nonparametric statistic. To use the Spearman correlation ρ we need first to rank the data, and then compute it using the following formula:

$$\rho = 1 - \frac{\sum_i d_i^2}{n(n^2-1)}, \tag{2}$$

where n is the size of the samples, and d_i is the difference between the ranks of each observation.

2.1.3 *Cross-correlation*

Let's consider now X and Y as two time series. The relationship between the processes generating these two time series, could not be instantaneous. One of the two process could be correlated with the other with a lag. Here is where the cross-correlation comes in handy. The cross-correlation measures the correlation between the two series X and Y as a function of the displacement of one with respect to the other. By normalizing the cross correlation we obtain a time-dependent Pearson correlation coefficient $\hat{r}(\tau)$, where τ here is the time lag, which can be written as

$$\hat{r}(\tau) = \frac{\sum_i (x_i - \bar{x})(y_{i-\tau} - \bar{y})}{\sqrt{\sum_i (x_i - \bar{x})^2 \sum_i (y_{i-\tau} - \bar{y})^2}}, \tag{3}$$

where x_i and y_i are the values in the samples at time i, τ is the time lag, $\bar{x}$ and $\bar{y}$ are the mean of the values of the X and Y variable, respectively.

2.1.4 *Mutual information*

The *mutual information (MI)* is a measure that captures nonlinear correlations, and which has been used in many studies to quantify the overlap of information contained in two processes.

Let's suppose to have a process X following a probability distribution $p_X(i)$ where i is a given process state, then the *Shannon entropy* (Shannon and Weaver, 1949) is given by

$$H_X = -\sum_i p_X(i) \log p_X(i), \tag{4}$$

where the sum extend over all possible states. The relative Shannon entropy, also called *Kullback-Leibler divergence* (Kullback, 1959) is a measure of how one probability distribution diverges from another one. Let's call this second distribution $q(i)$, then the Kullback-Leibler divergence K_X, is given by

$$K_X = \sum_i p_X(i) \log \left[\frac{p_X(i)}{q_X(i)} \right]. \tag{5}$$

Given two processes X and Y, to quantify the impact of assuming the independence between them, we would pose $q_{XY}(i,j) = p_X(i)p_Y(j)$ obtaining as Kullback-Leibler divergence

$$M_{XY} = \sum_{i,j} p_{XY}(i,j) \log \left[\frac{p_{XY}(i,j)}{p_X(i)p_Y(j)} \right], \tag{6}$$

which takes the name of mutual information. From this formula we can notice the symmetry of M_{XY} under the exchange of X and Y, thus not containing any directional sense nor dynamics. In the case of independence of processes X and Y the probabilities would be independent as well, turning the argument of the logarithm to 1 and the mutual information to 0. Practically, although the two processes X and Y can be fully independent, due to the limited number of samples the joint probability $p_{XY}(i, j)$ will not be equal to the product of the two independent probabilities $p_X(i)p_Y(j)$. Therefore M_{XY} will give a positive value (bias) also in case of independence of the two processes.

2.2 CAUSALITY

Correlation does not imply causation. Although two time series could be very well correlated, that doesn't mean that it exists causality between them. A well known example is the correlation between the number of storks and the number of births with a p-value of 0.008 (Matthews, 2000), meaning that there is only one chance out of 125 that it is a spurious correlation, therefore concluding with some certainty that storks deliver babies. Other examples are the lack of pirates that causes global warming, and that the most firefighters are sent to a fire, the more damage is done. For the interested reader, other hilarious spurious correlations can be found in Vigen (2015b), which are part of an extended collection (Vigen, 2015a). Clearly just by intuition we can understand that those are not actual causalities, but when we cannot make use of our intuition, things get more complicated. Moreover, while we can have a negative correlation, causality measures are positively defined: if positive peaks of a process are followed by negative dips of another, we have negative correlation, but positive causality.

A first attempt to try to quantify causality from observations was done in 1956 by Wiener (1956) and formalized in 1969 by Granger (1969). According to the *Granger causality (GC)*, given two processes X and Y, it is said that "Y G-causes X" if the information about the past of Y improves, in conjunction with the past of X, the prediction of the future of X, than the latter's past alone. Since then, several variations have been proposed (Amblard and Michel, 2013; Baccala and Sameshima, 2001; Barnett and Seth, 2014; Chen et al., 2004; Dhamala, Rangarajan, and Ding, 2008; Marinazzo, Pellicoro, and Stramaglia, 2008), and have been applied to a broad variety of fields, such as econometrics (Chiou-Wei, Chen, and Zhu, 2008; Hiemstra and Jones, 1994; Salahuddin and Gow, 2016), neurosciences (Seth, Barrett, and Barnett, 2015), physiology (Porta and Faes, 2016) and Earth sciences (McGraw and Barnes, 2018; Mosedale et al., 2006; Runge and al., 2019; Tirabassi, Masoller, and Barreiro, 2014; Tirabassi, Sommerlade, and Masoller, 2017) to cite a few.

An information-theoretic measure, known as *transfer entropy (TE)*, a form of conditional mutual information (CMI) (Paluš and Vejmelka, 2007), approaches this problem from another point of view: instead of predicting the future of X, it tests whether the information about the past of Y is able to reduce the uncertainty on the future of X. Since its introduction (Schreiber, 2000), TE has found applications in different fields such as neurosciences (Bielczyk and al., 2019; Lizier et al., 2011; Pereda, Quiroga, and Bhattacharya, 2005; Staniek and Lehnertz, 2008; Ursino, Ricci, and Magosso, 2020; Vicente et al., 2011; Wibral et al., 2013), physiology (Faes, Nollo, and Porta, 2011, 2013; Mueller et al., 2016), climatology (Bhaskar et al., 2017; Delgado-Bonal et al., 2020; Deza, Barreiro, and Masoller, 2015; Pompe and Runge, 2011; Pothapakula, Primo, and Ahrens, 2019), financial (He and Shang, 2017; Korbel, Jiang, and Zheng, 2019; Sandoval, 2014; Yao and Li, 2020) and social sciences (Porfiri and al., 2019).

For Gaussian processes the equivalence between GC and TE is well established (Barnett, Barrett, and Seth, 2009). There are no clear links though between GC and TE for non Gaussian processes. In practical terms, while TE provides a model-free approach, the need of estimating several probability distributions (see Eq. 8) makes TE substantially more computationally demanding than GC.

In the following sections we present the GC, the TE, and other approaches.

2.2.1 *Granger causality*

The first measure of causality quantification that we present in this work is the GC.

The mathematical formulation of the GC is based on linear regression modeling of stochastic processes. We write then process X and Y as autoregressive linear models of order p:

$$\begin{aligned} X(t) &= \sum_{i=1}^{p} a_i X(t-i) + \sum_{i=1}^{p} b_i Y(t-i) + \epsilon_X(t), \\ Y(t) &= \sum_{i=1}^{p} c_i X(t-i) + \sum_{i=1}^{p} d_i Y(t-i) + \epsilon_Y(t), \end{aligned} \tag{7}$$

where a_i, b_i, c_i and d_i are the coefficients of the model, $\epsilon_X(t)$ and $\epsilon_Y(t)$ are the residuals for each time series. If in the equation of $Y(t)$ the residual $\epsilon_Y(t)$ is reduced by including the values of X, then it is said that X G-causes Y. The limitations of the Granger causality are two assumptions on the data: the processes must be stationary, and they can be properly described by autoregressive linear models.

2.2.2 *Transfer Entropy*

The TE is an information-theoretic measure that quantifies the information transfer between two stochastic processes. Suppose to have two processes, namely X and Y: the TE $T_{Y \to X}$ is the amount of uncertainty reduction in future values of Y knowing the past values of X and Y. This means that TE is a measure to quantify how well we can predict the future values of a time series Y, given its past and the values of another time series X. A non negligible value of the TE indicates that process X *drives* process Y.

We can write the TE $T_{Y \to X}$ as (Schreiber, 2000)

$$T_{Y \to X} = \sum_{i,j} p\left(i_{n+1}, i_n^{(k)}, j_n^{(l)}\right) \log\left[\frac{p\left(i_{n+1} \mid i_n^{(k)}, j_n^{(l)}\right)}{p\left(i_{n+1} \mid i_n^{(k)}\right)}\right], \tag{8}$$

where $p\left(i_{n+1}, i_n^{(k)}, j_n^{(l)}\right)$ is the probability of process X to be in state i_{n+1} at time step $n+1$ and in states $i_n^{(k)}$ in the previous k time steps, and process Y to be in states $j_n^{(l)}$ in the previous l time steps. The conditional probabilities $p\left(i_{n+1} \mid i_n^{(k)}\right)$ and $p\left(i_{n+1} \mid i_n^{(k)}, j_n^{(l)}\right)$, are the probabilities for process X to be in state i_{n+1} at time step $n+1$, given the past states of X alone, and the past states of both X and Y, respectively.

2.2.3 *Other approaches*

For the sake of completeness, we mention here the *convergent cross mapping (CCM)* (Sugihara et al., 2012), that enriches the field of causality analysis in pairwise dynamical systems, and it is based on the nonlinear state space reconstruction. While GC and TE are suited for purely stochastic systems where causal influences are independent, the CCM can be applied in systems where causal influences have synergistic effects (Ye et al., 2015).

The success of the mentioned approaches strongly depends on the characteristics of the system under study (its dimensionality, the strength of the coupling, the length and the temporal resolution of the data, the level of noise contamination, etc.). Those approaches can fail in distinguishing genuine causal interactions from correlations that arise due to similar governing equations, or correlations that are induced by the presence of common external forcings. In addition, when the system under study is composed by more than two interacting processes, the mentioned metrics can return fake causalities, i.e., fail to discriminate between direct and indirect causal interactions. Many methods have been proposed to address these problems (Harnack et al., 2017; Hirata et al., 2016; Jiang et al., 2016; Korenek and Hlinka, 2020; Kugiumtzis, 2013; Leng et al., 2020; Ma, Aihara, and Chen, 2014; Ma et al., 2017; Nowack et al., 2020; Runge et al., 2019; Sun, Taylor, and Bollt, 2015; Vannitsem and Ekelmans, 2018; Zhao et al., 2016); however, their performance depends on the characteristics of the data, and their data requirements, computational cost, and number of parameters that need to be estimated may limit their applicability. Moreover, in order to discriminate between real and fake causality, it is required a complete knowledge of the system and the processes involved, which is often not the case in real complex systems.

3 THE MADDEN-JULIAN OSCILLATION

3.1 THE PHENOMENON

The MJO, Fig 2, is the major fluctuation in tropical weather on subseasonal time scale (Ferranti et al., 2018; Lau and Waliser, 2011; Vitart, 2009; Zhang et al., 2013), with a typical 30- to 60-days oscillation. Discovered in 1971 by Dr. Madden and Dr. Julian, the MJO is characterized by an eastward progression along the equator of large regions of enhanced and suppressed rainfall, from Western Africa to the Pacific Ocean, as shown in Fig 3.

3.2 MJO IMPACT

The MJO has a considerable worldwide socioeconomic impact. The reason lies in its influence on the tropical and extratropical climate. The MJO has a strong influence on the tropical weather, for example modulating cyclogenesis (Camargo, Wheeler, and Sobel, 2009; Fowler and Pritchard, 2020; Klotzbach, 2010). It is also a main source of intraseasonal variability for the different monsoon systems (Díaz, Barreiro, and Rubido, 2020; Taraphdar et al., 2018; Wheeler et al., 2009), and interacts with *El Niño-Southern Oscillation (ENSO)* (Bergman, Hendon, and Weickmann, 2001). Moreover, it impacts rainfall and temperature in the extratropics through atmospheric teleconnections (Alvarez, Vera, and Kiladis, 2017; Fauchereau, Pohl, and Lorrey, 2016; Ungerovich, Barreiro, and Masoller, 2021; Vecchi and Bond, 2004), affects the boreal winter extratropical circulation (Garfinkel, Benedict, and Maloney, 2014), and modulates the extratropical cyclone activity (Kunkel et al., 2012; Ma et al., 2017).

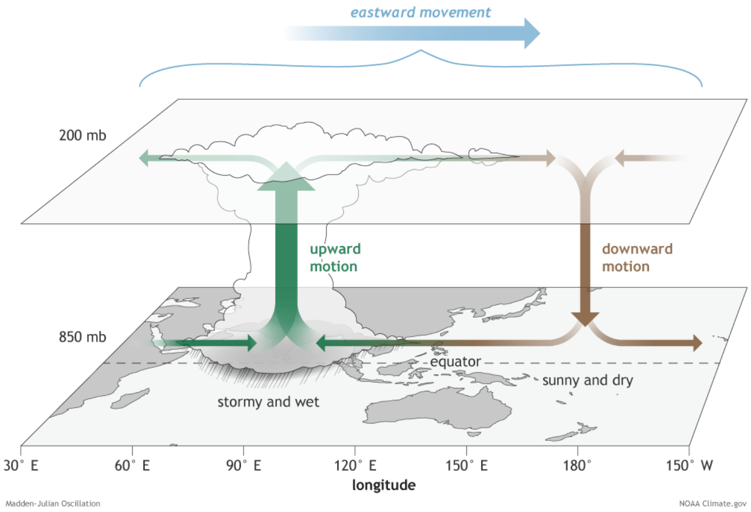

Figure 2: Structure of the MJO when the enhanced rainfall region is above the Indian ocean, and the dry region is over the Pacific Ocean. The green and brown arrows indicates air flows, while the blue one the whole system's movement. Climate.gov drawing by Fiona Martin.

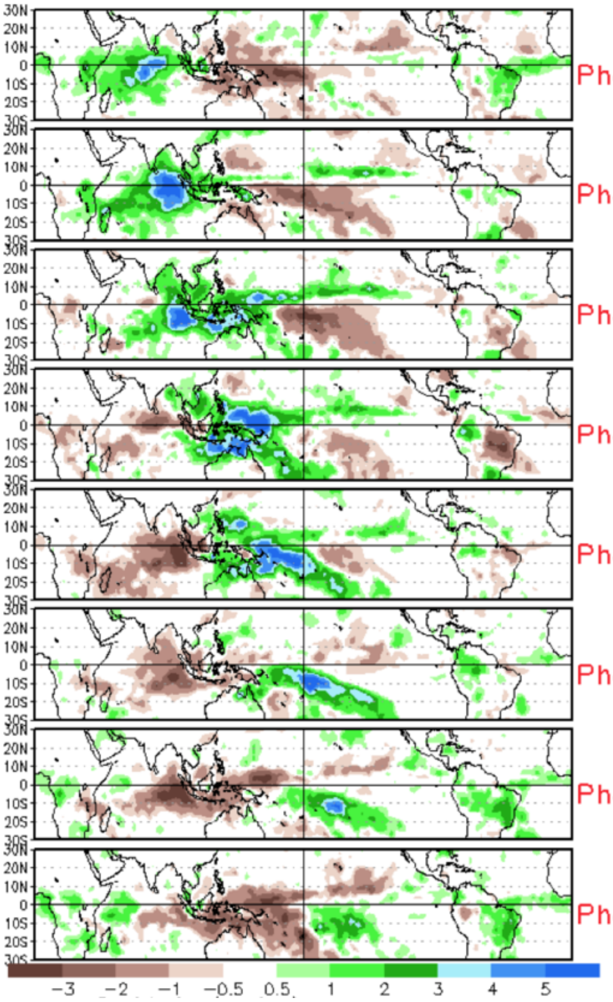

Figure 3: Difference from average rainfall for all MJO events from 1979-2012 for November-March for the eight phases described in the text. The green shading denotes above-average rainfall, and the brown shading shows below-average rainfall. To first order, the green shading areas correspond to the extent of the enhanced convective phase of the MJO and the brown shading areas correspond to the extent of the suppressed convective phase of the MJO. Note eastward shifting of shaded areas with each successive numbered phase as you view the figure from top to bottom. Image taken from https://www.climate.gov.

For example (Yoo, Feldstein, and Lee, 2011), MJO phases 4–6 are followed by Arctic warming with a lag of 1–2 weeks, and similarly, MJO phases 1–2 are followed by Arctic cooling. Thus, prediction of the MJO provides a source of climate predictability to many regions of the world on intraseasonal time scales.

3.3 MJO INDICES

In 2004, Wheeler and Hendon developed an index to characterize the MJO: the daily Real-time Multivariate MJO (RMM) index (Wheeler and Hendon, 2004). The RMM index is calculated as the first two principal components (RMM1 and RMM2) of the combined empirical orthogonal functions (EOFs) of outgoing longwave radiation (OLR), zonal wind at 200 and 850 hPa averaged between 15°N and 15°S. Applying a polar transformation to these two variables, it is possible to obtain the MJO phase and amplitude. The phase is classified in one of eight sectors of the phase diagram, Fig. 4, defining the observed MJO life cycle, shown in Figure 3, while the amplitude, characterizing the events' intensity, when smaller than 1 it corresponds to a non-active MJO. While other MJO indices exist, such as the OLR MJO index OMI (Liebmann and Smith, 1996), the real-time OLR MJO index ROMI (Kikuchi, Wang, and Kajikawa, 2012), the filtered OLR MJO index FMO (Kiladis et al., 2014), and the velocity potential MJO index VPM (Ventrice et al., 2013), the RMM index is the most frequently consideed.

3.4 MJO FORECAST

Until about a decade ago, empirical techniques exhibited a higher prediction skill compared to the numerical models, reaching up to 2 weeks (Kim, Vitart, and Waliser, 2018; Lau and Waliser, 2011; Neena et al., 2014). Significant advances in the understanding of the physics involved in the MJO and better dynamical forecasting systems, have allowed to improve the skill of MJO prediction. For the climate models the prediction

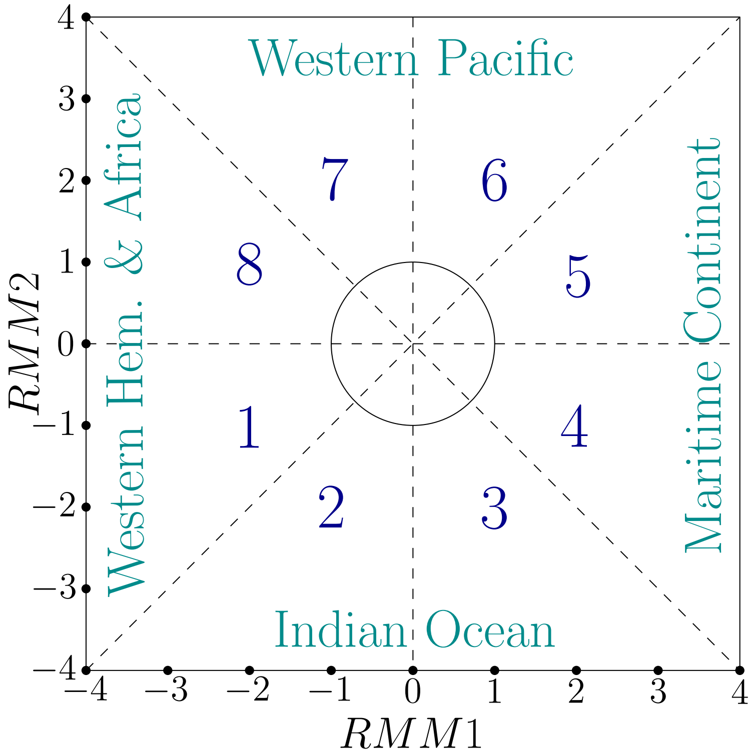

Figure 4: Wheeler-Hendon phase diagram. The two axes correspond to RMM1 and RMM2, which take values between -4 and 4. The dashed lines divide the phase space into eight sectors defining the MJO phases (MJO geographical location). The unitary circle corresponds to a MJO amplitude of 1, dividing the space into active (> 1) and non-active (< 1) MJO.

skill of MJO is sensitive to the physics of the model and the quality of the initial conditions. Of the dynamical models considered in 2014 by Neena and coworkers (Neena et al., 2014), shown in Fig. 5, the ensemble-mean prediction skill is highest for the model of the European Centre for Medium-Range Weather Forecast (ECMWF, 28 days) and for the model of the Australian Bureau of Meteorology (ABOM2, 24 days), and it is in the range of 15–20 days for most other models. More recently, the prediction skill of ECMWF has improved to exceeded 4 weeks, while most models have improved their skill to the range of 20–25 days (Kim, Vitart, and Waliser, 2018). The MJO prediction skill has also been shown to depend on the initial amplitude and phase, the season of the year, the background mean state and the extratropical influence (Kim, Vitart, and Waliser, 2018). Boreal winter leads, for most models, to a higher prediction skill that reaches up to 25–26 days, except for the ECMWF model that approaches 5 weeks (Jiang et al., 2020). Recently, also machine learning approaches have been used to predict the MJO (Martin, Barnes, and Maloney, 2021; Silini, Barreiro, and Masoller, 2021), yet not exceeding the prediction skill of the best dynamical models. Nevertheless, a combination of machine learning and dynamical models exceeded the prediction skill of the latter alone (Kim et al., 2021; Silini et al., 2022a), suggesting a promising route to follow in the future to improve the prediction of the MJO.

To have a phenomenological interpretation of the prediction error, it can be helpful to focus on the amplitude and phase. As shown in Figure 6, most models tend to underestimate the amplitude for all lead times, suggesting a faster decay of the predicted MJO with respect to the observations. Most models also predict a slower propagation of the MJO, accumulating about 2 days of delay in a 30 days prediction.

The Maritime Continent (MC) is the Southeast Asia archipelago comprising, among others, Philippines, Indonesia, and Papua New Guinea. It is the largest archipelago on Earth, and its orography, strong diurnal convection, as well as other factors, make it a complex land-atmosphere system. The MJO travelling from the Indian Ocean often weakens, or is disrupted, across the MC (Hendon and Salby, 1994; Rui and Wang, 1990).

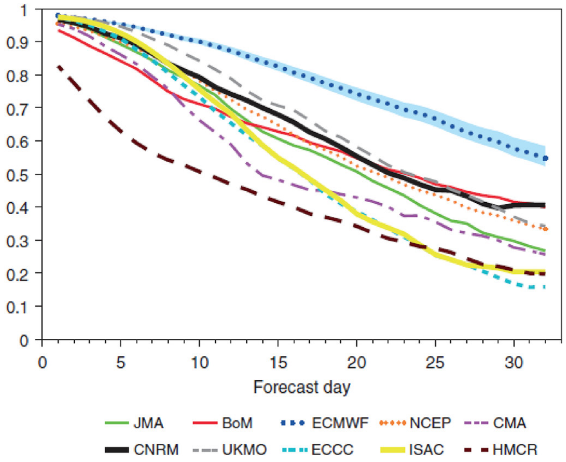

Figure 5: RMM bivariate correlation between the model ensemble means and ERA-Interim for 10 S2S models [Fig. 1 from Vitart, 2017].

Nevertheless, climate models tend to exaggerate this MC barrier due to poor modelling, and not due to a barrier on the predictability (Kim et al., 2014; Neena et al., 2014), leaving a margin for improvement.

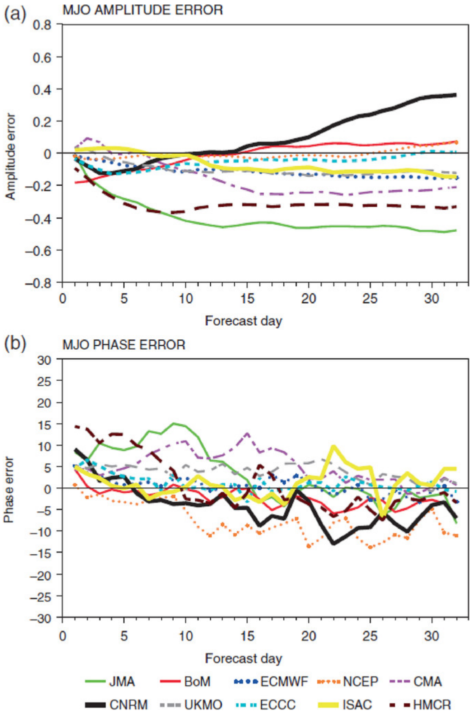

Figure 6: Evolution of the MJO (a) amplitude error and (b) phase error relative to ERA-Interim as a function of lead time. In (a), a positive (negative) value indicates a too strong (weak) MJO. In (b), a positive (negative) value indicates a too fast (slow) MJO propagation [Fig. 3 from Vitart, 2017].

4 MACHINE LEARNING

Machine learning (ML) algorithms aim to make a computer *learn* to perform particular tasks. ML algorithms are nowadays widely used in science and technology. There are many different ML algorithms that can be employed for different problems. We can find algorithms to extract information from texts, to predict values, to find anomalies, to discover structures, generate recommendations, or to perform classifications. We can make a computer recognize objects in an image, transcript texts from audios, mimic artists, do translations, recognize people emotions from their expressions, and much more. In this chapter we introduce the main types of ML algorithms and explain their use in this Thesis.

4.1 SUPERVISED, UNSUPERVISED, AND REINFORCEMENT LEARNING

Currently, there are three main categories of machine learning: supervised learning, unsupervised learning, and reinforcement learning.

In supervised learning we have access to the target (label), and the computer learns to perform a task trying to reduce the error between its guess and the target. Let's consider the example of a program that tries to attribute a sentiment to a given text input provided by an user. We present to the computer a list of users' texts with the associated sentiment provided by humans. We train then the computer to learn which features of the text are associated to a given sentiment. After the training, the test step consists of presenting to the computer unlabeled text (not seen before), and evaluate if its output (sentiment) is correct or not.

In unsupervised learning, we don't have access to the labels, and the computer tries to find patterns and information from the data without human intervention. The aim is to somehow organize the data or to describe its structure. For example, we want the computer to regroup users' texts that are similar between each other, according to some internal rules.

In reinforcement learning, we adopt a reward system, were we let the computer perform some actions and reward or punish it depending on the action taken. It is used to learn an optimal policy maximizing the reward. Like our pets, the computer will understand which actions to perform in order to get a reward with minimal effort. It is often used in robotics (Gullapalli, Franklin, and Benbrahim, 1994; Ibarz et al., 2021; Mahadevan and Connell, 1991), where each data point is given by a set of information retrieved by the sensors, and the robot has to choose its next action, and is also often used to create gaming bots (OpenAI et al., 2019; Silver et al., 2016).

In this thesis, all ML algorithms used are part of supervised learning, since the labels are known.

4.2 ARTIFICIAL NEURAL NETWORKS

A widely used family of ML algorithms are known as the Artificial Neural Networks (ANN).

ANNs are biologically-inspired algorithms, which are built to mimic the neural networks in the brain. They are composed of artificial neurons, a simplification of their biological counterparts, which are interconnected through links, that model axons, dendrites and synapses. Artificial neurons lay in layers that can be linked through different connective structures. ANNs are characterized by an input layer, one or more hidden layers, and an output layer, as shown in Fig. 7.

ANNs that contain more than one hidden layer, or that keep the memory of previous network states are known as Deep Learning (DL) algorithms.

The details of the ANNs used in this thesis and their training procedures are presented in the Appendix.

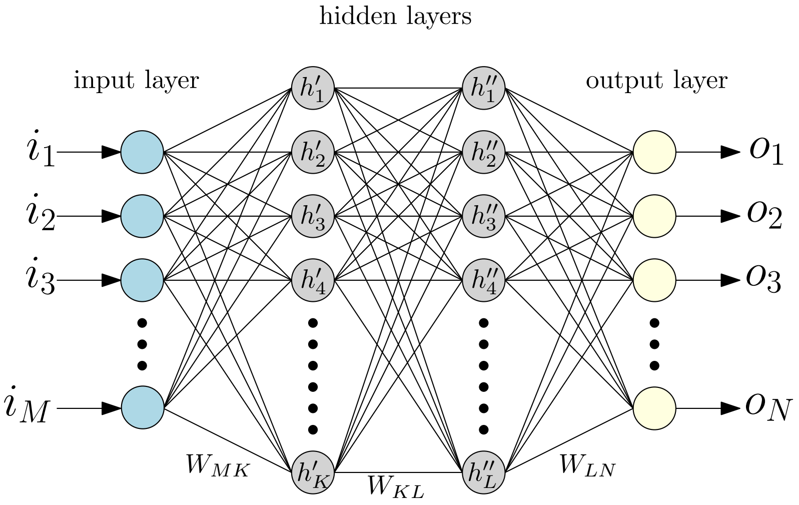

Figure 7: Example of an ANN, composed of an input layer with M neurons, two hidden layers with K and L neurons respectively, and an output layer with N neurons. The layers are interconnected through the matrices W_{MK}, W_{KL}, and W_{LN}.

4.3 APPLICATIONS IN CLIMATE

In climate science, ML algorithms are increasingly being employed due to their success.

In the case of ENSO, ML has allowed to improve the forecast lead time by several months with respect to the previous state-of-the-art dynamical forecast systems networks (Ham, Kim, and Luo, 2019). ML techniques have also been used to reconstruct the historical MJO index (Tseng, Barnes, and Maloney, 2020), to compete or outperform dynamical models in forecasting large/scale spatial patterns of precipitation (Gibson et al., 2021), and to reduce the costs of climate change scenarios computations (Mansfield et al., 2020), to cite a few.

A complete review of applications is out of the scope of this thesis, and we steer the interested reader to an extensive analysis of ML applications in weather prediction and climate analysis, that can be found in Bochenek and Ustrnul (2022).

4.3.1 *Prediction of the MJO*

Many efforts have been made for the prediction of the MJO in the last decades (Jiang et al., 2020), with dynamical models leading to the current best forecasts, but despite the continuous progress of the dynamical models, there is still room for improvement in the prediction of the MJO (Jiang et al., 2020; Zhang et al., 2013). In particular, an improvement of the prediction skill when MJO crosses the Maritime Continent (MC) barrier (Barrett et al., 2021; Kim et al., 2016; Wu and Hsu, 2009) will be of practical importance due to the influence of MJO on ENSO, as an improved MJO prediction may contribute to improving the prediction of ENSO.

However, to the best of our knowledge, ML algorithms were not yet been used to predict MJO before Silini and Masoller (2021), except for DL algorithms that had been used in post-processing to correct the bias of MJO dynamical multi-models means (Kim et al., 2021).

Although MJO predictions obtained using ML models do not outperform dynamical models (Martin, Barnes, and Maloney, 2021; Silini and Masoller, 2021) yet, a hybrid approach (Silini et al., 2022a), combining dynamical models and ML techniques, manages to improve the dynamical models results. In this way, it is possible to use dynamical models that have been developed across decades, based on physical phenomena, in combination with data-driven ML techniques, an approach that has shown its ability to reduce the gap between observations and dynamical models' forecasts (Haupt et al., 2021; McGovern et al., 2019; Rasp and Lerch, 2018; Scheuerer et al., 2020; Vannitsem et al., 2021).

Part II

PSEUDO TRANSFER ENTROPY

In this second part, we introduce our implementation of a fast and effective metric, to compute the G-causality called pseudo transfer entropy (pTE), we test it using synthetic data, and apply it to real data. Afterwards, we employ pTE to unveil information flow between climate indices, which functions as feature selection tool for the prediction of the considered climate indices.

The results we present in Chapter 6, have been published in Silini and Masoller (2021), and those presented in Chapter 7 have been submitted for publication (Silini et al., 2022b).

5 PSEUDO TRANSFER ENTROPY

5.1 DEFINITION

In this chapter we present the derivation of the pTE from the TE, which in its turn is derived from the MI.

With the TE, Schreiber purpose was to find a statistic measure, sharing some of the desired properties of the MI, albeit keeping into account the dynamics and directionality of information.

It is possible to give a directional sense to MI *ad hoc* by introducing a time lag τ like follows

$$M_{XY}(\tau) = \sum_{i,j} p_{XY}(i_n, j_{n-\tau}) \log \left[\frac{p_{XY}(i_n, j_{n-\tau})}{p_X(i_n) p_Y(j_{n-\tau})} \right]. \tag{9}$$

If we consider a system approximated by a Markov process of order k, then the probability to be in the state i_{n+1} of the process X at time $n+1$ is independent of the state i_{n-k}. For what follows we will use the notation $i_n^{(k)} = (i_n, \dots, i_{n-k+1})$.

The entropy of an additional state knowing all the previous state can be written as

$$h_X = -\sum_i p_X \left(i_{n+1}, i_n^{(k)} \right) \log \left[p_X \left(i_{n+1} \mid i_n^{(k)} \right) \right]. \tag{10}$$

Since by the definition of conditional probability

$$p_X \left(i_{n+1} \mid i_n^{(k)} \right) = \frac{p_X \left(i_{n+1}^{(k+1)} \right)}{p_X \left(i_n^{(k)} \right)}, \tag{11}$$

then

$$h_X = -\sum_i p_X\left(i_{n+1}, i_n^{(k)}\right)\left\{\log\left[p_X\left(i_{n+1}^{(k+1)}\right)\right] - \log\left[p_X\left(i_{n+1}^{(k)}\right)\right]\right\}, \tag{12}$$

also called *entropy rate*. In a Markov process of order k, the conditional probability to find X in state i_{n+1} at time $n+1$ is independent of the state i_{n-k}, therefore we can write the equality

$$p_X\left(i_{n+1} \mid i_n^{(k)}\right) = p_X\left(i_{n+1} \mid i_n^{(k+1)}\right), \tag{13}$$

which leads to

$$h_X = H_{X^{(k+1)}} - H_{X^{(k)}}. \tag{14}$$

In order to find the MI rate by generalizing h_X to two processes, the best way to go is to measure the deviation from the generalized Markov property. This choice is taken since by using the Kullback-Leibler divergence we would end up with a symmetric quantity under exchange of the two processes. We can write this independence as

$$p_X\left(i_{n+1} \mid i_n^{(k)}\right) = p_{XY}\left(i_{n+1} \mid i_n^{(k)}, j_n^{(l)}\right), \tag{15}$$

which implies the absence of information flow from Y to X. The divergence from this independence assumption is yet again quantified by the Kullback-Leibler divergence, obtaining

$$TE_{Y\to X} = \sum_{i,j} p\left(i_{n+1}, i_n^{(k)}, j_n^{(l)}\right)\log\left[\frac{p\left(i_{n+1} \mid i_n^{(k)}, j_n^{(l)}\right)}{p\left(i_{n+1} \mid i_n^{(k)}\right)}\right], \tag{16}$$

which was named *transfer entropy* by Schreiber (2000).

The TE from process Y to process X can be re-written as

$$\begin{aligned} TE_{Y\to X} = \sum_{i,j} p\left(i_{n+1}, i_n^{(k)}, j_n^{(l)}\right)\Bigg\{ &\log\left[p\left(i_{n+1} \mid i_n^{(k)}, j_n^{(l)}\right)\right] \\ &- \log\left[p\left(i_{n+1} \mid i_n^{(k)}\right)\right]\Bigg\}, \end{aligned} \tag{17}$$

where, just like in Eq. 8, $p\left(i_{n+1}, i_n^{(k)}, j_n^{(l)}\right)$ is the probability of process X to be in state i_{n+1} at time step $n+1$ and in states $i_n^{(k)}$ in the previous k time steps, and process Y to be in states $j_n^{(l)}$ in the previous l time steps. The conditional probabilities $p\left(i_{n+1} \mid i_n^{(k)}\right)$ and $p\left(i_{n+1} \mid i_n^{(k)}, j_n^{(l)}\right)$, are the probabilities for process X to be in state i_{n+1} at time step $n+1$, given the k past states of X alone, and the k past states of X combined with the l past states of Y, respectively.

By using the definition of conditional probabilities, Eq. 17 can be rewritten as sum of four Shannon entropies (Shannon and Weaver, 1949) as

$$\mathrm{TE}_{Y\to X} = H\left(i_n^{(k)}, j_n^{(l)}\right) - H\left(i_{n+1}, i_n^{(k)}, j_n^{(l)}\right) + H\left(i_{n+1}, i_n^{(k)}\right) - H\left(i_n^{(k)}\right), \tag{18}$$

where H is given by

$$H = -\sum_i p(i) \log p(i), \tag{19}$$

and the sum extends over all possible states i.

The computation of the TE with Eq. 8 is challenging because a good estimation of the probability distributions is often not available. Considering the processes X and Y to follow normal distributions i.e. $X \sim \mathcal{N}(x \mid \mu_x, \Sigma_x)$ and $Y \sim \mathcal{N}(y \mid \mu_y, \Sigma_y)$, where $\mu_{x,y}$ are the mean values, and $\Sigma_{x,y}$ are the covariances, it substantially simplifies the computation using in fact that the entropy of a p-variate normal variable x, is given by

$$\begin{aligned} H_p(x) &= \int_{-\infty}^{+\infty} \mathcal{N}(x \mid \mu_x, \Sigma_x) \log\left[\mathcal{N}(x \mid \mu_x, \Sigma_x)\right] dx \\ &= -\mathbb{E}\left[\log\left(\mathcal{N}(x \mid \mu_x, \Sigma_x)\right)\right], \end{aligned} \tag{20}$$

where $\mathbb{E}[\cdot]$ is the expected value.

By definition of the multivariate Gaussian, we can rewrite Eq. 20 as

$$H_p\left(x\right) = -\mathbb{E}\left[\log\left((2\pi)^{-\frac{p}{2}} \mid \Sigma \mid^{-\frac{1}{2}} e^{-\frac{1}{2}(x-\mu_x)^T \Sigma_x^{-1}(x-\mu_x)}\right)\right], \qquad (21)$$

which, by the property of the logarithm of products becomes

$$H_p\left(x\right) = \frac{p}{2}\log(2\pi) + \frac{1}{2}\log(\mid \Sigma_x \mid) + \frac{1}{2}\mathbb{E}\left[(x-\mu_x)^T \Sigma^{-1}(x-\mu_x)\right]. \qquad (22)$$

By noticing that $\mathbb{E}\left[(x-\mu_x)^T \Sigma_x^{-1}(x-\mu_x)\right] = \mathrm{tr}(\Sigma_x^{-1}\Sigma_x) = p$, we obtain

$$H_p(x) = \frac{1}{2}\left(p + p\log(2\pi) + \log|\Sigma_x|\right), \qquad (23)$$

where $|\Sigma|$ is the determinant of the $p \times p$ positive definite covariance matrix. By substituting Eq. 23 in Eq. 18, we can estimate the TE as follows:

$$\begin{aligned} TE_{Y\to X} = &\frac{1}{2}\left[k + l + (k+l)\log(2\pi) + \log\left(\left|\Sigma\left(\mathbf{I}_n^{(k)} \oplus \mathbf{J}_n^{(l)}\right)\right|\right)\right] \\ &- \frac{1}{2}\left[k + l + 1 + (k+l+1)\log(2\pi) + \log\left(\left|\Sigma\left(\mathbf{i}_{n+1} \oplus \mathbf{I}_n^{(k)} \oplus \mathbf{J}_n^{(l)}\right)\right|\right)\right] \\ &+ \frac{1}{2}\left[k + 1 + (k+1)\log(2\pi) + \log\left(\left|\Sigma\left(\mathbf{i}_{n+1} \oplus \mathbf{I}_n^{(k)}\right)\right|\right)\right] \\ &- \frac{1}{2}\left[k + k\log(2\pi) + \log\left(\left|\Sigma\left(\mathbf{I}_n^{(k)}\right)\right|\right)\right], \end{aligned} \qquad (24)$$

which finally can be written as

$$TE_{Y\to X} = \frac{1}{2}\log\left(\frac{\left|\Sigma\left(\mathbf{I}_n^{(k)} \oplus \mathbf{J}_n^{(l)}\right)\right| \cdot \left|\Sigma\left(\mathbf{i}_{n+1} \oplus \mathbf{I}_n^{(k)}\right)\right|}{\left|\Sigma\left(\mathbf{i}_{n+1} \oplus \mathbf{I}_n^{(k)} \oplus \mathbf{J}_n^{(l)}\right)\right| \cdot \left|\Sigma\left(\mathbf{I}_n^{(k)}\right)\right|}\right), \qquad (25)$$

where $\Sigma(A \oplus B)$ is the covariance of the concatenation of matrices A and B, $\mathbf{i}_{n+1}$ is the vector of the future values of X, $\mathbf{I}_n^{(k)}$ and $\mathbf{J}_n^{(l)}$ are the matrices containing the previous k and l values of processes X and Y respectively.

Whenever X and Y are not Gaussian processes, we call the quantity in Eq. 25 *pseudo Transfer entropy (pTE)*. For Gaussian variables pTE coincides with the TE and is equivalent to GC (Barnett, Barrett, and Seth, 2009). The Gaussian form for CMI/TE for causality inference was also used by Cliff et al. (2021), Molini, Katul, and Porporato (2010), Paluš (2014a), and Paluš (2014b).

In the following chapter we evaluate the performance of pTE with several known models, comparing it with the implementation of GC and TE of well-known Python libraries. We will cover its strengths and weaknesses, and show an example of application for real climatological data.

6 APPLICATION TO SYNTHETIC DATA

6.1 MODELS

Three data generating processes (DGPs) were analyzed. For these DGPs the null hypothesis of non-causality is not satisfied for process Y to process X. Results obtained with other DGPs are presented in the Appendix.

The first DGP is a linear model (Diks and DeGoede, 2001) given by:

$$X_t = 0.6X_{t-1} + C \cdot Y_{t-1} + \epsilon_{1t}, \qquad Y_t = 0.6Y_{t-1} + \epsilon_{2t}, \tag{26}$$

where ϵ_{1t} and ϵ_{2t} are white noises with zero mean and unit variance, and C is the coupling strength.

The second DGP is a nonlinear model (Taamouti, Bouezmarni, and Ghouch, 2014) that reads:

$$X_t = 0.5X_{t-1} + C \cdot Y_{t-1}^2 + \epsilon_{1t}, \qquad Y_t = 0.5Y_{t-1} + \epsilon_{2t}. \tag{27}$$

The third DGP consists of two Lorenz chaotic systems, coupled on the first variable:

$$\begin{aligned} \dot{X}_1 &= 10(-X_1 + X_2) + C \cdot (Y_1 - X_1) & \dot{Y}_1 &= 10(-Y_1 + Y_2) \\ \dot{X}_2 &= 21.5X_1 - X_2 - X_1X_3 & \dot{Y}_2 &= 20.5Y_1 - Y_2 - Y_1Y_3 \\ \dot{X}_3 &= X_1X_2 - \tfrac{8}{3}Y_3 & \dot{Y}_3 &= Y_1Y_2 - \tfrac{8}{3}Y_3 \end{aligned} \tag{28}$$

Examples of time series of these three DGPs, normalized to zero mean and unit variance, are displayed in Fig. 8.

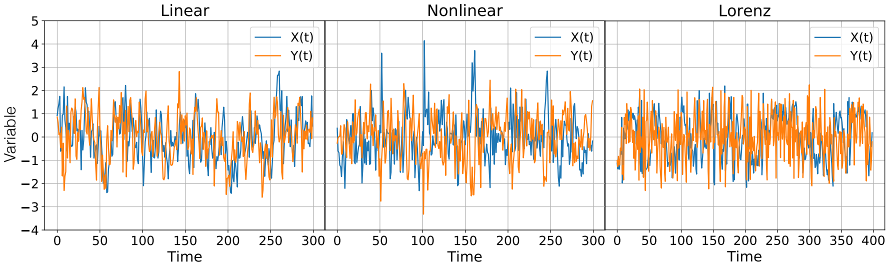

Figure 8: Examples of time series of the three data generating processes (DGPs) analyzed in the main text. In the three cases there is causality from Y to X; the coupling strength is (a), (b) $C = 0.5$, (c) $C = 8$.

6.2 STATISTICAL SIGNIFICANCE

We used surrogate data to test the significance of the pTE, TE and GC values. The number of surrogates needed depends on the characteristics of the data, the available computational resources and time limitations: given enough resources and time, one should use a large number of surrogates and select a confidence interval (Paluš and Vejmelka, 2007); however, with limited time or computational resources, when the spread of surrogates data is not too large one can use an alternative strategy: analyze a small number of surrogates and, in the case of a one sided test, select as significance threshold the maximum or minimum value obtained with the surrogates. In this case, $M = K/\alpha - 1$ surrogates should be generated, where K is a positive integer number and α is the probability of false rejection (Lancaster et al., 2018). Therefore, a minimum of 19 surrogates ($K = 1$) are required for a significance level of 95%. For this study we used the *iterative amplitude adjusted Fourier transform (IAAFT)* algorithm the time-shifted (T-S) surrogates.

6.3 implementation

To calculate pTE we developed an algorithm in *python* (available on *GitHub* (Silini, 2020)), while we used the *statsmodels* implementation of GC (Fulton, 2020) and the *pyunicorn* implementation of TE (Donges et al., 2015). The code has been thought to be as user friendly as possible to be used to build networks. It takes as arguments all the time series of the studied system, the embedding parameter and the statistical significance test that the user decides to apply. As result it returns the matrix of pTE values computed from the original data, and the matrix of the maximum values obtained from the surrogates (i.e., the statistically significant thresholds).

In the analysis of synthetic data generated with the DGPs the causality measures were run over 1000 realizations with different initial conditions and noise seeds. For each realization the first 100 data points were discarded. For the computation of GC and pTE we chose a lag equal to 1, which implies considering the models as auto-regressive processes of order 1, AR(1), since by the considered models construction, the dependent variable is influenced by the previous step of the independent one; for the computation of TE the k-nearest neighbors method is used, and we chose $k = \sqrt{N}$, where N is the number of data points in the time series (Lall and Sharma, 1996).

To calculate the causality between two time series, the time series were first linearly detrended and L2-normalized. The significance of the pTE, GC and TE values obtained were then tested against the values obtained from 19 couples of surrogates (as explained in the previous section, 19 surrogates is the minimum for achieving a significance level of 95%). Unless otherwise specifically stated, the results presented in the text were obtained by using IAAFT surrogates.

6.4 RESULTS

First, we use the three DGPs described in section 6.1 to compare the performance of pTE, GC and TE in terms of the power and size. If by construction there is no causality from X to Y, the percentage of times the causality is higher than the significance threshold returned by the surrogate analysis will be called "size" of the test, i.e., is the probability that a causality is detected when there is no causality by construction. On the other hand, if by construction X causes Y, the percentage of times the method finds causality from X to Y is called "power" of the test. With the surrogate analysis adopted, the causality between the original data will be compared to the maximum one found within 19 surrogates (Lancaster et al., 2018), and the probability that the original data displays by chance the highest causality is 5%.

We analyze the power and size for the two possible causal directions ($X \to Y$ and $Y \to X$), as a function of the coupling strength and of the length of the time series. Fig. 9 displays the power and size of the three methods, pTE, GC and TE, for the linear model, when the coupling is such that there is causality from Y to X (the size is shown in the top row, and the power, in the bottom row). The similarity between pTE and GC in finding the true causality is evident. With a coupling strength $C < 0.1$ the three methods fail to detect causality, while for $C > 0.4$, for both pTE and GC, the number of data points in the time series needed to find causality is quite small, in fact 100 data points are sufficient to achieve a power of 100. In Fig. 10 we plot the cross sections of the highest values of the time series length and coupling strength of Fig. 9. From the left panel it is possible to notice that for a coupling strength of 0.5, a time series of 200 data points is needed to retrieve the correct causality for all three methods with a power above 95. From the information contained in the right panel, for time series composed of 500 data points, a coupling strength of about 0.25 is necessary to find a power larger than 95 for all three methods.

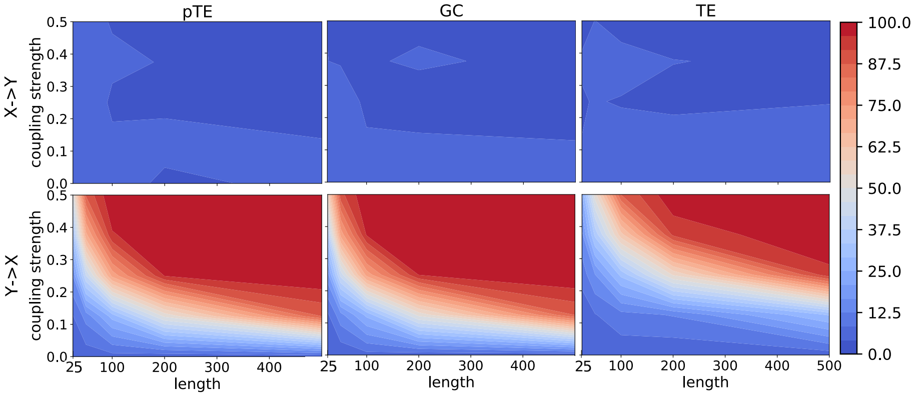

Figure 9: Power and size [the percentage of times that causality is detected when there is causality (power) and when there is no causality (size)] obtained using pTE (first column), GC (second column) and TE (third column) on the linear model, as a function of the length of the time series and of the strength of the coupling, for the two possible causality directions (top row: X → Y, bottom row: Y → X). By construction the model has causality from Y to X; therefore, the top row displays the size, and the bottom row, the power. The performance of pTE and GC is very similar, as both find the correct causality with moderate coupling strength even for short time series. TE finds the correct causality, but for stronger coupling.

Fig. 11 displays the results obtained for the nonlinear model, and we notice that they are very similar to the ones obtained with the linear model, probably due to the weak nonlinearity considered. We note that, in comparison with the linear model, in this model, with short time series the power and size returned by the three methods are more similar.

Regarding the two chaotic Lorenz oscillators, which are coupled in the first variable, the situation is very different, as shown in Fig. 12. When looking at the causality between the coupled variables, for both pTE and GC the causality is detected for a moderate coupling strength and a rather long time series. Causality X → Y is not detected for any (coupling strength, time series length), which is correct by construction. TE instead finds causality also for X → Y, which is wrong by construction.

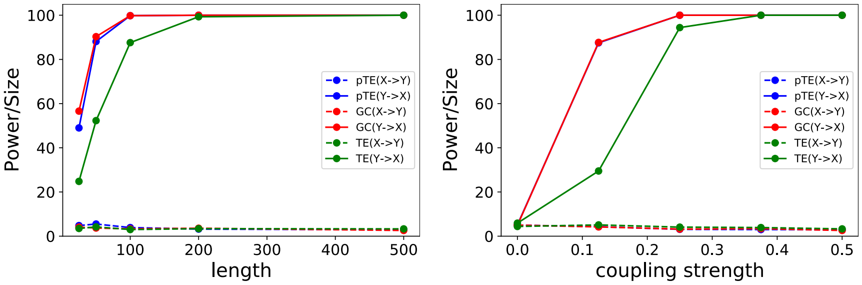

Figure 10: Cross sections of Fig. 9. In the left panel we fix the coupling strength to 0.5 and we plot the power and size of the linear model as a function of the time series length for pTE, GC and TE. In the right panel we fix the number of data points to 500 and plot the power and size as a function of the coupling strength.

This observation for TE can be attributed to insufficient conditioning treated by Paluš (Palus et al., 2001; Paluš and Vejmelka, 2007), in fact the directionality of the coupling cannot be inferred when the systems are fully synchronized.

Next, we compare the computational cost of using pTE, GC and TE. Fig. 13 displays the time required to calculate X → Y and Y → X causalities, as a function of the length, N, of the time series. The figure shows the time required when the codes are run on Google colab CPUs (Intel® Xeon® CPU @ 2.20GHz), and includes preprocessing the time-series (detrending and normalizing) and performing the statistical significance test.

For short time series we see a large advantage of using pTE instead of GC. TE sits back as the slowest of the three methods. The reason is attributed to the scaling of parameter k in the k-nearest neighbors method used to compute TE, which scales as $\sqrt{N}$.

Table 1 displays the computational time required to calculate X → Y and Y → X causalities, and the corresponding power and size obtained using the linear model. While in Fig. 13 we showed the total computational

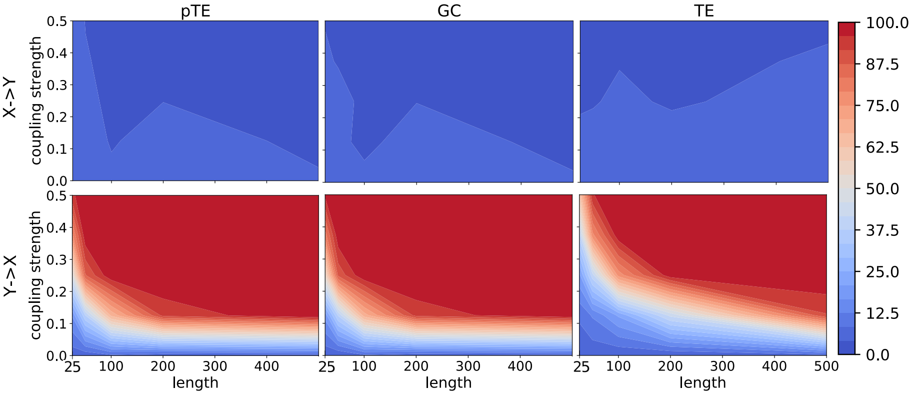

Figure 11: As Fig. 9, but using the nonlinear model. We again see that pTE and GC both find the correct causality, and their performance is very similar. TE finds the correct causality, but for stronger coupling.

time, in Table 1 we show only the time required for the calculation of the pTE, GC and TE values (without signal preprocessing and without performing statistical significance analysis). We see that, for time series of 25 data points, the time required for pTE calculation (averaged over 1000 runs) is 200% faster than GC; however, this percentage reduces to 12% for time series of 500 data points. From these results, we argue on the value of using pTE to analyze a large number of short time series, which is often the case when causality methods are used to build complex networks from observed data. We remark that all the codes used to generate the results shown in this article are publicly available at *GitHub* (Silini, 2020).

The use of T-S surrogates (Lancaster et al., 2018; Quian Quiroga et al., 2002) results in a substantial reduction of the computational time, in comparison to the widely used IAAFT surrogates, as seen in Fig. 13 and Table 2. The computational cost is reduced by approximately 98%, albeit displaying very similar results in terms of power and size. Clearly, T-S surrogates give a major boost in causality testing. As an example, for time series of length $N = 100$, using pTE with T-S surrogates will reduce the computational cost by approximately 82% with respect to

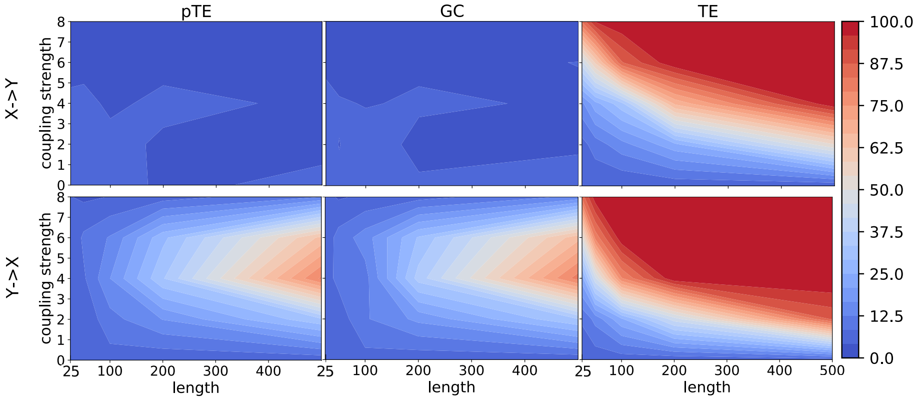

Figure 12: As Fig. 9, but using the chaotic model composed by two coupled Lorenz systems. The performance of pTE and GC is very similar, as both find the correct causality when the time series is long enough, and the coupling strength is moderate. TE finds Y → X causality, but it also finds X → Y causality, which is wrong by construction.

GC with IAAFT surrogates, while a reduction of approximately 77% is found with respect to GC with T-S surrogates. However, for causal inference T-S surrogates should be used with caution, because when there are time-delayed interactions, it can lead to fake conclusions.

To study the resilience to observational noise, we add, to the time series generated with the DGPs, X and Y, a Gaussian noise $\xi_{1,2}$ of zero mean and unit variance, tuning its contribution with a parameter $D \in [0,1]$. In this way we generate and analyze the signals X' and Y' given by $X'_t = (1-D)X_t + D\xi_{1t}$, $Y'_t = (1-D)Y_t + D\xi_{2t}$.

Fig. 14 shows that pTE and GC perform very similarly (they are almost indistinguishable) and are quite resilient to noise. For the linear DGP, up to 40% of noise contribution can be present without a significant effect on the results, while for the nonlinear DGP, the methods start failing for a lower noise level. For the chaotic DGP the three methods are very resilient to noise. As previously noticed in Fig. 12, TE detects causality in both directions.

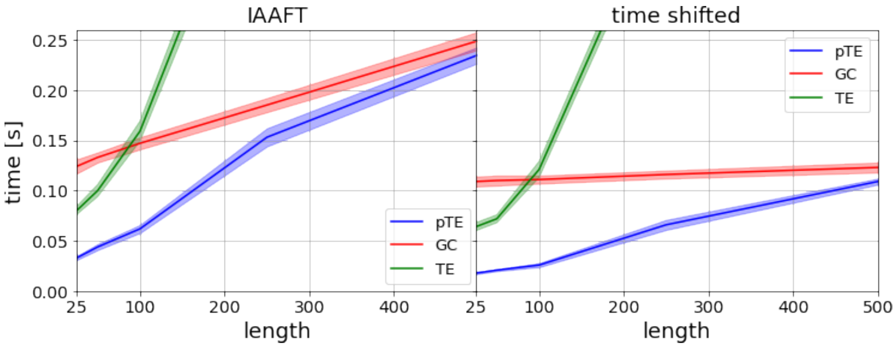

Figure 13: Computational times required to infer causal interactions in the two directions, X → Y and Y → X, using pTE, GC or TE, as a function of the length of the time series, N. The times, calculated with the linear model after averaging over 1000 realizations, include preprocessing the time series and performing the statistical significance analysis. In the left panel IAAFT surrogates are used, while in the right panel time shifted surrogates are used.

Finally, moving beyond synthetic data, we apply the pTE measure to two well-known climatic indices, and compare the results with GC and TE. The time series analysed, the NINO3.4 index and All India Rainfall (AIR) index, shown in Fig. 15, represent the dynamics of two large-scale climatic phenomena, the El Niño–Southern Oscillation (ENSO) and the Indian Summer Monsoon (ISM), whose causal inter-relationship is represented by long-range links (teleconnections) between the Central Pacific and the Indian Ocean (Dijkstra et al., 2019). The time series were downloaded from Explorer (2020) and Tropical Meteorology (2020). The NINO3.4 index begins in 1854 while AIR index begins in 1813. Monthly-mean values are available, and their shared period is from 1854 to 2006 (153 years, 1836 months),

Table 3 displays the results of the analysis of monthly-sampled data, and of yearly-sampled data. In the latter case we used the average of December, January and February (DJF) values, where the ENSO phenomenon peaks, and the average of June, July and August (JJA), where the mon-

Table 1: Average computational time of pTE, GC and TE per realization for four time series lengths, N. The mean and standard deviation are computed over 1000 realizations and the values in the table are expressed in milliseconds. pTE is the fastest up to $N = 500$ data points, with the difference with GC diminishing as N increases. TE time increases exponentially as the k parameter of the k-nearest neighbors scales with $\sqrt{N}$. The power and size are computed for the linear model. We note that for $N = 100$, pTE and GC give very similar results, even though pTE takes half the time. The last column displays the average computational cost reduction of pTE with respect to GC.

Data points	Time [ms]			Power/Size		
N	pTE	GC	TE	pTE	GC	TE
25	1.3 ± 0.3	3.7 ± 0.6	3.0 ± 0.5	49.0/4.8	56.6/4.1	24.8/3.5
100	1.9 ± 0.4	4.0 ± 0.6	8.6 ± 0.8	99.8/3.9	99.8/3.3	87.6/3.0
250	3.0 ± 0.6	4.6 ± 0.8	34 ± 2	100/3.2	100/3.6	99.3/3.4
500	4.1 ± 0.3	4.6 ± 0.3	112 ± 2	100/2.9	100/2.6	100/3.3

Data points N	Computational time reduction (%)
25	64.9
100	52.5
250	34.8
500	10.9

soon peaks. Therefore, the length of the yearly-sample time series is 152 data points because for the last year the last data point, DJF, is not available. We used, for the yearly-sampled data, an autoregressive integrated moving average (ARIMA) model of order 4 (consistent with Tirabassi, Sommerlade, and Masoller (2017)) and, for the monthly-sampled data, of order 3. The order of the model was selected by using the Akaike information criterion (AIC).

Table 2: Average computational time to generate time shifted (T-S) and IAAFT surrogates, for four time series lengths, N. The mean and standard deviation are computed over 1000 realizations and the values in the table are expressed in milliseconds. T-S surrogates are substantially faster than IAAFT, allowing to reduce the average computational time required to create surrogates by approximately 98%. The causality testing using pTE with the two surrogate methods gives very similar results in terms of power and size for the linear model.

Data points	Time [ms]		Power/Size		Computational time
N	T-S	IAAFT	T-S	IAAFT	reduction (%)
25	0.020 ± 0.004	0.6 ± 0.1	48.9/3.2	51.5/4.9	96.7
100	0.035 ± 0.005	1.5 ± 0.3	97.5/0.0	99.2/3.0	97.7
250	0.07 ± 0.01	3.2 ± 0.6	100/0.0	100/2.9	97.8
500	0.13 ± 0.02	6 ± 1	100/0.0	100/2.6	97.8

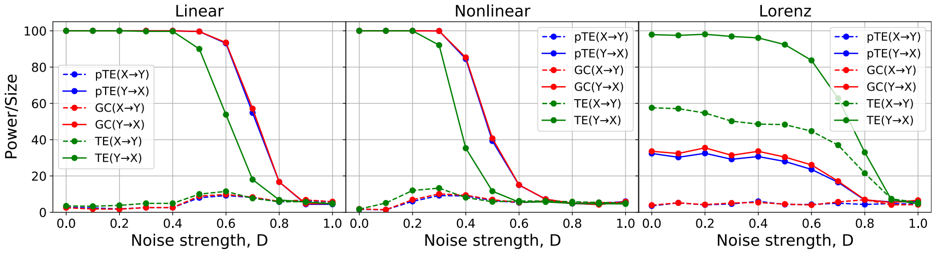

Figure 14: Resilience to noise of pTE, GC and TE, using the linear, nonlinear and chaotic models. pTE and GC perform very similarly (they are almost indistinguishable). The three measures are quite resilient to noise: for the linear model, up to 40% of noise can be present without significantly affecting the results, while for the nonlinear model, the three methods start failing at a lower noise strength. For the chaotic model, as previously noticed in Fig. 12, TE detects causality in both directions. The length of the time series is N = 300 and the coupling strength is C = 0.5 for the linear and nonlinear models, C = 4 for the chaotic model (for this value of the coupling TE has the largest difference between the two directions).

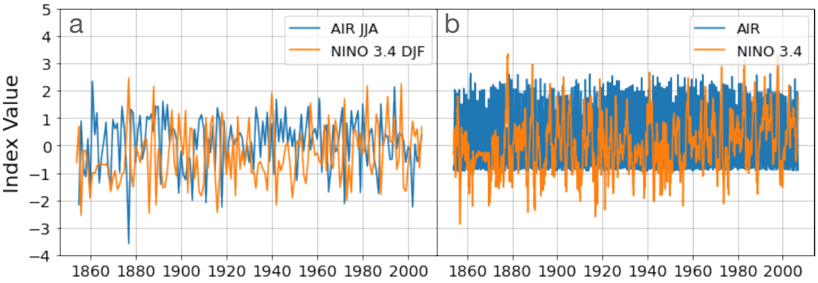

Figure 15: L2-normalized and linearly detrended time series of NINO3.4 and All India Rainfall (AIR) indices from 1854 to 2006. In panel (a) it is shown the average value of DJF for NINO3.4 index and JJA for AIR index, while in panel (b), the monthly sampled time series.

Table 3: Results of the analysis of the NINO3.4 and AIR indices yearly- and monthly-sampled using T-S and IAAFT surrogates. The table indicates the the number of datapoints in the time series, the pTE, GC and TE values obtained, the significance threshold, and the computational time required to calculate the causality including statistical significance analysis.

Direction	N	pTE / th / sig.	Time(s)	GC / th / sig.	Time(s)	TE / th / sig.	Time(s)
			Time shifted				
ENSO→AIR	152	0.028 / 0.020 / YES	0.02	1.9 / 1.8 / YES	0.18	0.045 / 0.018 / YES	0.35
AIR→ENSO	152	0.003 / 0.029 / NO	0.02	0.49 / 1.8 / NO	0.18	0.06 / 0.02 / YES	0.35
ENSO→AIR	1836	0.002 / 0.006 / NO	0.18	1.6 / 8.8 / NO	0.38	0.03 / 0.02 / YES	34
AIR→ENSO	1836	0.0059 / 0.0058 / YES	0.18	5.4 / 5.2 / YES	0.38	0.02 / 0.04 / NO	34
			IAAFT				
ENSO→AIR	152	0.028 / 0.020 / YES	0.02	1.9 / 1.7 / YES	0.20	0.07 / 0.04 / YES	0.51
AIR→ENSO	152	0.003 / 0.029 / NO	0.02	0.49 / 1.8 / NO	0.20	0.06 / 0.02 / YES	0.51
ENSO→AIR	1836	0.002 / 0.006 / NO	0.25	1.6 / 8.8 / NO	0.44	0.03 / 0.02 / YES	34
AIR→ENSO	1836	0.0059 / 0.0058 / YES	0.25	5.4 / 5.2 / YES	0.44	0.02 / 0.04 / YES	34

In Table 3 we see that for the yearly-sampled data, pTE and GC only detect the dominant causality (ENSO→AIR), while TE detects both (in good agreement with Tirabassi, Sommerlade, and Masoller (2017)). We note similarities with the results presented in Fig. 12: while unidirectional causality is found with pTE and GC, TE causality is found in both directions. The computational times clearly show that pTE is faster than GC (and of course also faster than TE, which is the slowest method). In the monthly-sampled data we see an opposite direction of causality, a result that we interpret as due to different time scales in the mutual influence between ENSO and ISM: while ENSO effects on the Indian monsoon precipitations are pronounced on an annual time scale, the influence of the Indian monsoon on ENSO acts on a shorter, monthly time scale. To exclude the fact that this change in directionality is an artifact due to the different time series lengths, we analyzed the monthly-sampled time series using segments of 152 consecutive data points (which is the length of the annually-sampled data). In this case we did not find any significant causality, which suggests that the change in directionality when considering annually-sampled or monthly-sampled data is not an artifact but has a physical origin, that we interpret as due to different time scales in the mutual interaction and that 152 data points are not sufficient to find causality in the monthly-sampled data.

Finally, we note that the computational times shown in Table 3 are higher than those that can be estimated from Fig. 13. In Fig. 13 we see that, for 150 datapoints, the time required for the GC calculation with T-S surrogate analysis is about 0.11 s while in Table 3 we see that the time required for GC and T-S calculation (two directions) is 0.36 s. The difference is due to the fact that in Fig. 13 a model of order 1 was used, while in Table 3, for the yearly-sampled data, a model of order 4 is used. The computational time increases with the order of the model, especially for GC, because the algorithm used (`statsmodels grangercausalitytest`) computes causality for all model orders up to the chosen one. For the NINO3.4 and AIR indices we also analysed the effect of varying the order of the model (from 1 to 10) and found either the same significant causal directionality (with stronger or weaker values), or we did not find any significant causality.

The linear DGP was used by Diks and DeGoede (Diks and DeGoede, 2001) to test nonlinear Granger causality. With a coupling strength of $C = 0.5$ and a time series length of 100 points with a lag of 1, they obtained a power of 95.6 and a size of 3.0. Using pTE under the same conditions, we obtain a power of 99.8 and a size of 3.9.

The nonlinear DGP was used by Taamouti, Bouezmarni, and Ghouch (2014) to quantify linear and nonlinear Granger causalities. With a coupling strength of $C = 0.5$, 200 data points, a pvalue of 5% and a resampling bandwidth k for the bootstrap as the integer part of $2 \cdot 200^{1/2}$, they obtained a power of 100 and a size of 4.4. Using pTE we obtained a power of 100 and a size of 3.3.

The Krakovská, Jakubík, and Chvostekova (2018) coupled Lorenz systems, are very similar to those studied here. By using three state-space based methods, including cross-mapping, they noticed that the highest directionality in the causality is for a coupling $C \approx 4$. From $C > 4$ synchronization is obtained, finding causality in both directions, using time series of 50000 data points. This observation is very similar to our results with TE, while for pTE and GC, once synchronization has been achieved, no causality is found. This supports their conclusion, warning the reader that the blind application of causality test can easily lead to incorrect conclusions. While GC and pTE can successfully be used to analyze AR processes and weakly nonlinear Gaussian-like processes, for more complex processes (high dimensional and/or highly nonlinear) advanced information-theoretic methods such as TE are needed.

6.5 DISCUSSION

We have proposed a new measure, *pseudo transfer entropy* (pTE), to infer causality in systems composed by two interacting processes. Using synthetic time series generated with processes where the underlying causality is known, and also, a real-world example of two well-known climatic indices, we have found a remarkable similarity between the results of pTE and Granger causality (GC), in terms of the power and size, and the robustness to noise, but pTE can be significantly faster, particularly for short time series. For example, for time series of 100 datapoints, while giving extremely similar results, pTE with time-shifted (T-S) surrogate testing reduces the computational time by approximatelly 92% with respect to GC with IAAFT surrogate testing, and by 48% with respect to GC with T-S surrogate testing (on Google colab CPU, the total computational time for pTE and T-S is 2.5 ms, while for GC and IAAFT is 32.5 ms, and for GC and T-S, 4.7 ms).

Since the computational cost is of capital importance for the analysis of large datasets, the causality testing methodology proposed here will be extremely valuable for the analysis of short and noisy time series whose probability distributions are approximately Gaussian. We remark that many real-world signals follow distributions that are nearly normal. Although we do not claim that our method can be applied to any pair of signals, the information presented in the *Appendix* supports the method's generic applicability. The algorithms are freely downloadable from *GitHub* (Silini, 2020).

In the next chapter we apply the pTE on thirteen well-known climate indices to disentangle the interactions among the phenomena that they represent, which we represent as a causality network.

7 APPLICATION TO CLIMATE INDICES

7.1 MONTHLY ATMOSPHERIC AND OCEAN TIME SERIES

We focus this study in the climatic indices described below. They are monthly sampled timeseries and are freely accessible at the NOAA website (https://psl.noaa.gov/data/climateindices/list/), with the exception of the All Indian Rainfall index, which is available via the Indian Institute of Tropical Meteorology (https://www.tropmet.res.in/).

AIR: All Indian Rainfall. The area-weighted integral of the rainfall measured by the Indian national network of rain gauges.

AMO: Atlantic Multidecadal Oscillation. The detrended area-weighted average over the North Atlantic, from the equator up to 70N, of the sea surface temperature (SST) anomalies from the Kaplan SST dataset (Kaplan et al., 1998; Reynolds and Smith, 1994).

GMT: The Global Mean Temperature anomaly as computed by NASA/GISS. The anomaly is computed with respect to the period 1951-1980.

HURR: The total number of hurricanes or named tropical storms in a given month in the Atlantic region.

NAO: The North Atlantic Oscillation. The north-south dipole of pressure anomalies over the North Atlantic, with one center over Greenland and the other center of opposite sign between 35N and 40N.

NINO34: The East Central Tropical Pacific SST anomaly. It integrates the NOAA ERSST V5 anomalies in the region (5N-5S)×(170W-120W).

NP: North Pacific pattern. The area-weighted sea level pressure over the region (30N-65N)×(160E-140W).

NTA: North Tropical Atlantic index. The SST anomalies averaged over the two regions (60W-20W)×(6N-18N) and (20W-10W)×(6N-10N) map. Anomalies are obtained from the ERSST V3b dataset relative to the 1981-2010 climatology, smoothed by three months running mean and projected onto 20 leading EOFs.

PDO: Pacific Decadal Oscillation. The leading principal component of monthly SST anomalies in the North Pacific Ocean.

QBO: Quasi-Biennial Oscillation. The zonal average of the 30mb zonal wind at the equator as computed from the NCEP/NCAR Reanalysis.

Sahel: Sahel Standardized Rainfall. Average rainfall recorded by 14 weather stations in the region (8N-20N)×(20W-10E).

SOI: Southern Oscillation Index. The standardized difference in surface air pressure between Tahiti and Darwin. The SOI is a proxy of the strength of the Walker circulation and it's strictly related to ENSO.

TSA: Tropical Southern Atlantic Index. The SST anomaly with respect to the 1971-2000 period in the region (0-20S)×(10E-30W). HadISST and NOAA OI $1^\circ \times 1^\circ$ datasets are used to create this index.

The various indices span different regions and focus on different variables. variability of the ocean and the atmosphere on different spatio-temporal scales, with particular attention to the tropical belt.

The majority of the timeseries span six decades, overlapping in the period 1951-2016 (i.e., in 792 data points). In this period, the timeseries are depicted in Fig. 16.

The various indices display different spectral properties. Several have a defined periodic component, either seasonal, as in the case of rainfall and storm indices, or longer, like the case of the QBO. Some display trends (GMT), others slow non-linear oscillations (NINO34). The high frequencies as well are very heterogeneous.

All these complex spectral properties influence the indices' distributions. Generally, we observe skewed distributions and hints of multimodality.

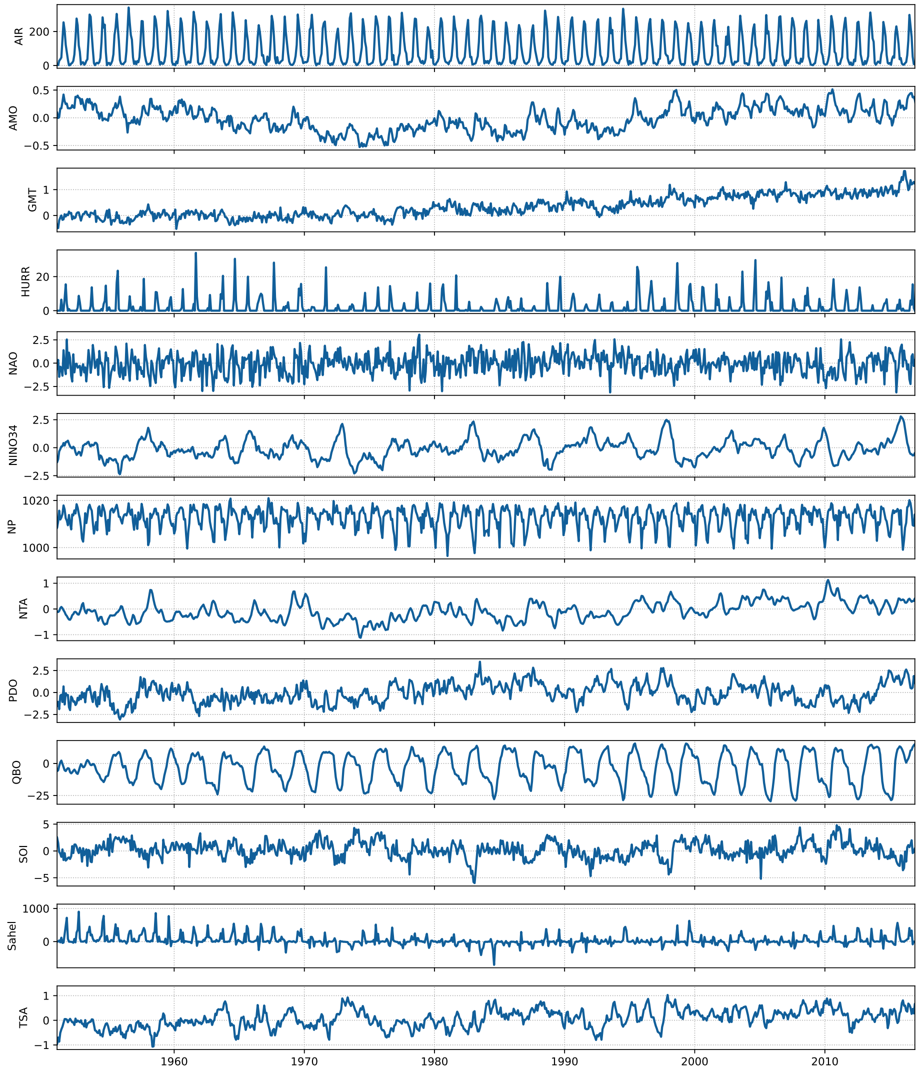

Figure 16: Raw time series of the thirteen climatic indices under study.

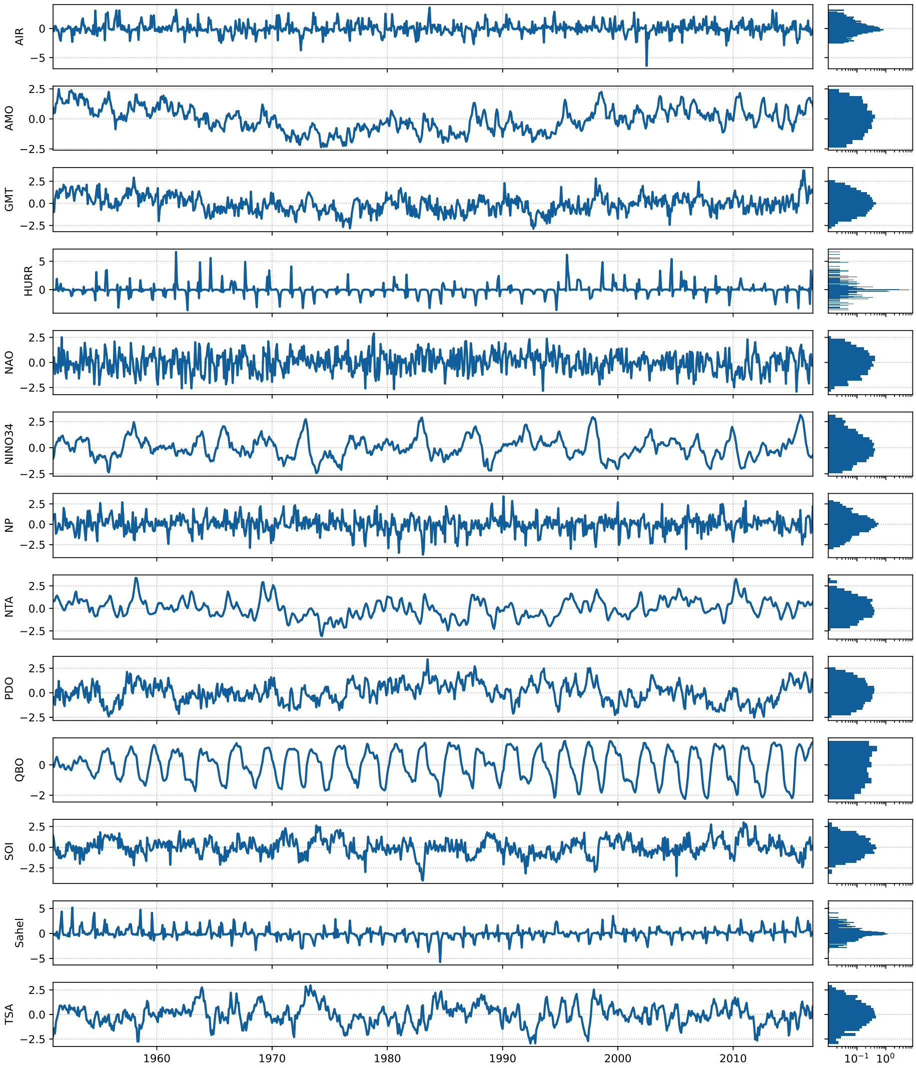

Figure 17: Post-processed time series and histograms of values in log scale.

AIR, HURR, Sahel and NP indices display a strong seasonal component (Fig. 16). To avoid spurious signals, we removed the seasonality subtracting the mean of every different month from these four indices. The other indices have already been constructed after removing the seasonal cycle of the variables used.

Since the indices under consideration span a broad range of values, we standardized their distribution by removing a linear trend and rescaling the series to have unitary variance. Even if information-related quantities such as pTE, transfer entropy, or mutual information in principle do not depend on the variables gauges, the rescaling to unitary variance is considered good practice for their numerical computation. The resulting post-processed time series and their values distributions are depicted in Fig. 17.

7.2 METHODOLOGY

We analyze the causality structure of the set of indices indices using bivariate analysis, i.e., we address the problem of determining whether one index influences another without considering the possible influence of a third index that could mediate the interaction.

For consistency, we focused on the period 1950–2016, which is the time span for which all the time series have continuous records.

To measure the bivariate causality between two indices, we use the pseudo transfer entropy (pTE) (Silini and Masoller, 2021), a method that has recently proven to be a computationally fast alternative to traditional transfer (TE) entropy (Schreiber, 2000). Analogously to the TE, the pTE measures the transfer of information from Y at time t to X at time $t+\tau$ conditional to the information flowing from X at time t to X at time $t+\tau$, providing a mean to assess causal relationships between two processes.

The pTE from series Y to X is calculated from the analysis of a vector containing the future elements of X at $t \geqslant \tau$, and the matrices containing the k past values of X and Y (see Silini and Masoller (2021) for details).

Index	k
AIR	1
AMO	2
GMT	2
HURR	1
NAO	1
NINO34	2
NP	1
NTA	8
PDO	1
QBO	2
Sahel	1
SOI	3
TSA	1

Table 4: Summary of index model orders, k, used in this study.

The embedding dimension of time series X, hereafter k, has to be selected before carrying on the calculation (Silini and Masoller, 2021). There are different possibilities to determine its optimal value. Here we model X as an autoregressive process and fix the model order, k, minimizing the Bayesian information criterion (BIC) score.

7.2.1 *Statistical significance analysis*

Once the pTE between two indices is computed, we have to address whether it is significant or not. Unlike the case of cross-correlation, the null model of X being independent of the past of Y doesn't allow the analytical calculation of p-values for the pTE distribution. For this reason, we have to rely on surrogate analysis to understand if a pTE value is significantly different from zero.

Surrogate timeseries can be obtained from real-world data through different kinds of manipulation (Lancaster et al., 2018). Surrogate time series should retain all the properties of the original timeseries with the exception of the one we are interested in. In the case of causal relationships, we want surrogates that are independent from each other while preserving the autocorrelation function of the original time series. In fact, preserving the autocorrelation ensures that we preserve the dependence of the timeseries on itself. To achieve this we employed an algorithm known as *iterative amplitude adjusted Fourier transform* (IAAFT) (Schreiber and Schmitz, 1996, 2000), which preserves both the amplitude distribution and the power spectrum of the original series.

From the original dataset, we generated $N = 1000$ independent surrogate datasets using IAAFT. From the surrogate datasets we obtain N surrogate measures of pTE between each pair of timeseries. Thus, the quantiles of the surrogate pTEs can be viewed as significance threshold for the measured pTE values. In the following, we considered a pTE value significant if it falls within the highest 1% of its surrogates distribution.

7.2.2 *Long-term causality variation*

Measuring the variation of pTE between the first and second half of the dataset allows us to explore possible long-term variation in the index interactions. To address the significance of such variations we rely again on surrogate analysis. Once we split the dataset into two halves, ranging respectively from 1950 to 1983 and from 1984 to 2016, we create $N = 1000$ surrogates for each half, and from the surrogates, we compute N pTE values for each pair of variables on each half. For each surrogate pTE value of the second half, we randomly sample 100 surrogate pTEs from the first half and we compute the average difference. This way, we end up with N surrogate differences for each pair of indices. A difference is considered significant if the empirical p-value is either below 1% for a negative difference or above 99% for a positive difference.

7.2.3 *Results*

In Fig. 18(a) we report an example of pTE calculation focusing on NINO34 as "forcing" node, considering significant pTEs for different values of τ. Most of the results match with the current knowledge regarding ENSO dynamics. We can observe a 5-months cutoff in the NINO34 → SOI interaction, which is in line with the ENSO build-up time scale when the ocean and the atmosphere are coupled. Moreover, a maximum of around 4 months in the NINO34 → NTA is expected too, given that their interaction is mediated by heat fluxes: because of its thermal inertia, the SST of the ocean boundary layer changes in a time scale of roughly three months, producing the pTE delayed maximum. From this perspective, the behavior of the AMO is analogous. The AIR and HURR have a 1-month impact, which is reasonable given that the interaction is mediated directly by the atmosphere. It is interesting to note that the HURR index has a small but significant pTE tail up to $\tau = 3$, which may result from indirect interactions mediated by the NTA.

Results of some indices are, however, unexpected to a degree. For the PDO and the NP, we would have expected a behavior more similar to the NTA one. Instead, we observe in Fig. 18(a) relatively high pTEs up to 4 months. We interpret this as due to the fact that the local air-sea interaction increases the persistence of the remotely forced ENSO signal. In contrast, the pTE values for NAO (Fig. 18(b)) show a very rapid decrease with τ, indicating interactions on a much shorter time scale.

In the following, for every pair of indices, we calculated the pTE and its significance for $\tau = 1$, 3 and 6, that is one month, one season and half a year into the future. In Fig. 19, we report the significant connections between the various indices. For better clarity, results are displayed both as a network and an adjacency matrix. In the network representation, directed connections are represented by arrows. In particular, we draw a link only if the value of the pTE between two indices is significant for at least one value of τ.

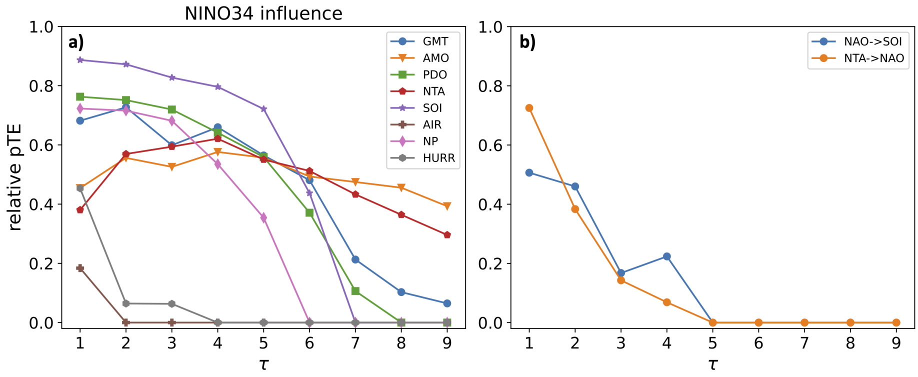

Figure 18: Influence of NINO34 (a) and NAO (b) to a subset of indices for various lags τ. The vertical axis shows the relative pTE value, which is the relative difference between the measured pTE and the significance threshold determined by using the surrogate analysis (see Sec. 7.2 for details). For clarity, pTE values that are not significant are not shown.

We investigate the role of τ in the pTE values in Fig. 20, again representing connections by directed arrows. The arrows' colors represent the value of τ for which the pTE is the highest, while the arrows' width represents this maximum value. We can observe that, in general, connections have relatively low τ. Also, the values of pTE for $\tau = 6$ seems to be lower than for other lags.

The full network picture allows to visualize the global structure; however, the different links can not be clearly distinguished. For the sake of clarity, we select a subset of key indices (AMO, NTA, PDO, and the NINO34/SOI pair), and report in Fig. 21 the forward and inward links separately. The pivotal role of ENSO in the climate network is evident, with numerous forward connections tying NINO34 and SOI to the vast majority of the studied indices.

Finally, we report in Fig. 22 the significant variations of pTE values between pre-1984 and post-1984. We observe that the variations are heterogeneous, with main changes being an increase in the strength of the

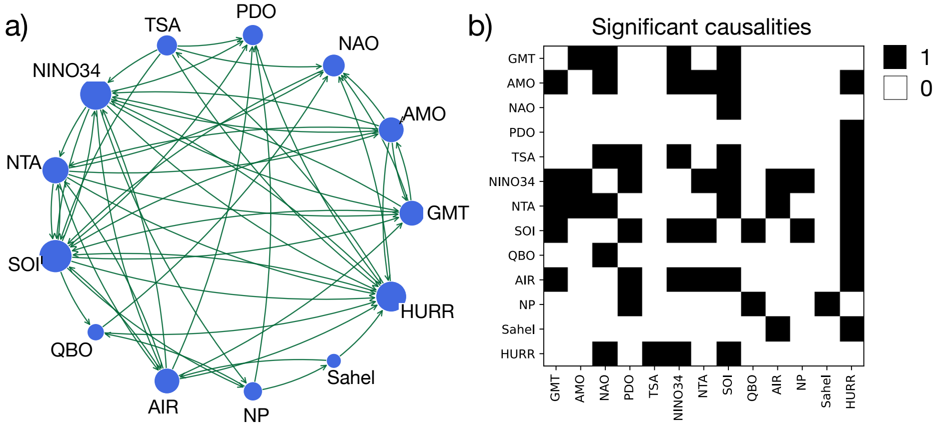

Figure 19: a) Network representation of causal dependencies. The arrows go from the driving to the driven indices. The size of the nodes is proportional to their degree. b) Binary matrix representation of the causal dependencies. The nodes (the columns and rows) represent climate indices, while the links (the black squares) indicate significant causality between indices. Row labels represent the forcing indices, while the columns stand for the forced ones. All significant connection are shown, regardless of the lag, τ.

NP-PDO link and the decrease of the strength of the PDO-GMT link, and of several links that affect AMO. The discussion of the uncovered links, and their variations, is presented the next section.

7.3 DISCUSSION

While some of the uncovered connections are well-known, others are either previously unknown or are possible false positives. In the latter case, a common driver of two indices could produce a significant amount of shared information, inducing an apparent connection.

Let's start with the connection between ENSO and the north Pacific indices (NP and PDO). It is well known that ENSO induces an atmospheric teleconnection that modulates the Aleutian low over the North Pacific

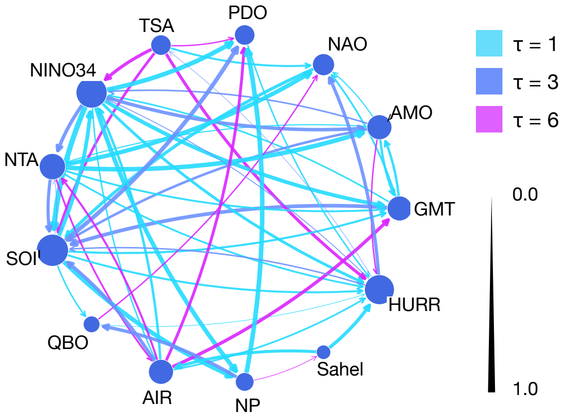

Figure 20: As Fig. 19, here the width of the represents the strength of the causal dependency (the relative value of pTE, with respect to the significance threshold found with a p-value of 0.01) and the color represents the lag, τ, for which maximum causality is found. The size of the nodes is proportional to their degree.

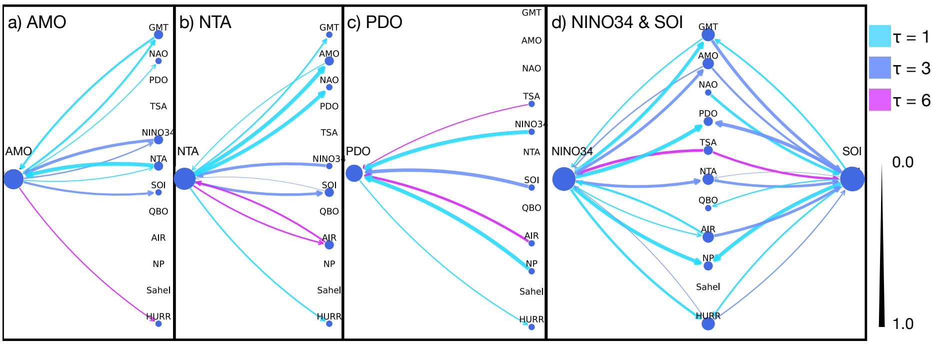

Figure 21: Connections between selected indices and the rest of the network. a) AMO, b) NTA, c) PDO, d) NINO34 and SOI.

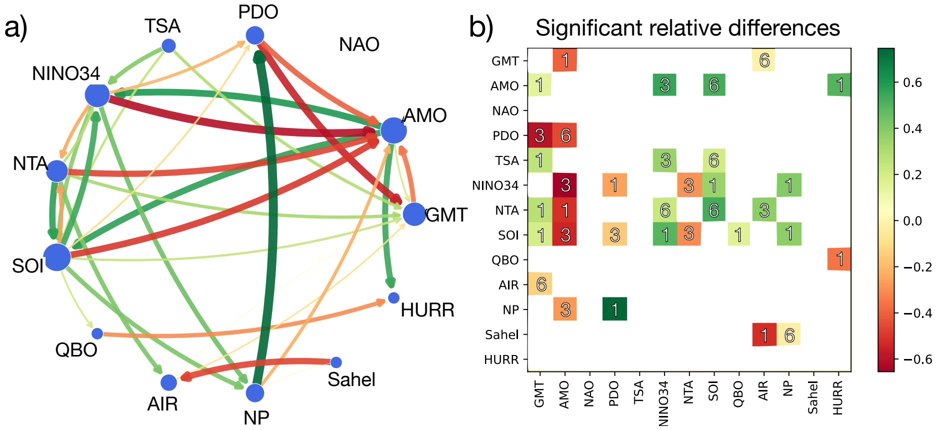

Figure 22: Significant differences in the causality networks between two time windows, represented both as network and adjacency matrix. The first time window contains the years 1950-1983, while the second correspond to the window 1984-2016. Green (red) links correspond to an increment (reduction) in the causality from the first to the second window. For each connection, we report the largest significant difference across the three studied values of τ. The value of tau for which the difference is the largest is reported in adjacency matrix. The width and color of the links are proportional to the relative increments (reductions), with respect to the significance threshold found with IAAFT surrogates (p-value = 0.01).

Wang et al., 2012. As the Aleutian low is characterized by the NP index, the ENSO → NP causality is well understood. Changes in the surface winds associated with the Aleutian low in turn induce SST anomalies in the north Pacific thus affecting the PDO (NP → PDO). Finally, air-sea interaction in the north Pacific generates the causality PDO → NP Newman et al., 2016.

On the other hand, it is unlikely that the number of tropical storms in the north Atlantic (HURR) can drive NAO, TSA, ENSO and NP, as suggested by the results. Instead, these causalities likely result as consequence of complex interactions among the different atmospheric and oceanic phenomena characterized by these indices. For example, it is well known that during El Niño the number of hurricanes in the north Atlantic decreases because of enhanced vertical shear Klotzbach, 2011 as found by pTE. At the same time a warm NTA, which is also influenced by ENSO Chang, Saravanan, and Ji, 2003, favours the development of tropical storms Pérez-Alarcón et al., 2021. Similar complex interactions explain the links of the AIR index.

The lag is a crucial parameter that has a large impact on the analysis. The global surface air temperature warms up by about 0.1°C during an El Niño event, with a lag of about 6 months Trenberth et al., 2002. This causality is detected in the analysis, although ENSO → GMT is maximum at lag 1. At longer lags we find the opposite causality (GMT → SOI at lag 3 and NINO34 at lag 1), which may be understood as consequence of the persistence of the ENSO events that last between 6 and 9 months.

The causality identified QBO→ NAO in our analysis has been reported in the literature to occur during boreal winter Andrews et al., 2019; Marshall and Scaife, 2009, although the link had been assessed as relatively weak. We found the largest causality for a lag of 6 months, suggesting that the mechanism through which the QBO affects the NAO may last more than one season.

It is well known that the NAO is the main driver of the sea surface temperature anomalies over the tropical north Atlantic mainly through changes in surface heat fluxes Visbeck et al., 2001. However, we don't find this causality, probably due to the index used to describe the Atlantic SST. As mentioned above the NTA index uses SST anomalies that have been smoothed thus filtering out the response to NAO. On the other hand, the analysis detected the NTA → NAO connection at lag 1, which is consistent with the literature that shows that SSTa in the tropical north Atlantic can induce atmospheric teleconnections that project onto NAO Okumura et al., 2001.

Another index that presents links with several other phenomena is the TSA. As in the case of the tropical north Atlantic, it is likely that the TSA → NAO link is direct, as the SSTa in the south Atlantic can control the position of the Intertropical Convergence Zone which could promote the development of a teleconnection to the north Atlantic Okumura et al., 2001. Interestingly, connections with other indices occur with lags of 3 or 6 months, suggesting that some of these links are indirect. The tropical south Atlantic is known to influence the equatorial Pacific through changes in the Walker circulation with a lag of several months Rodríguez-Fonseca et al., 2009, consistent with our results. Thus, we hypothesize that the connections of the TSA with the PDO and HURR indices occurr via the ENSO influence.

Our analysis also shows that the impact of TSA on ENSO has grown in recent decades (see Fig. 22), in agreement with the literature Rodríguez-Fonseca et al., 2009.

Looking at Fig. 22, the strongest and most consistent signal is the change in causality between SOI (and Nino34) and AMO. That AMO variability influences the ENSO variability is already documented Levine, McPhaden, and Frierson, 2017. While the literature on the link between AMO and ENSO is extensive, to our knowledge, there is no report that this influence is getting stronger, which could have implications for the ENSO predictions. As already mentioned, El Niño warms up the NTA, which is part of the AMO index, therefore it's not surprising that ENSO appears driving AMO. On the other hand, the Atlantic can influence ENSO variability by changing the mean state in the equatorial Pacific, altering the

Walker circulation and trades, as pointed out above. From Fig. 22, we infer that the AMO impact on ENSO is increasing while the ENSO impact on AMO is becoming weaker. For longer lags (e.g., $\tau = 9$) only the link AMO → ENSO remains (not shown). The link ENSO → AMO is weak for long leads, while AMO → ENSO can still be strong, due to the different time scales of the phenomena.

Figure 22 also shows an increase in the NP → PDO link in the last decades, which may be related to the fact that the ENSO teleconnection to the north Pacific has also increased (NINO34 → NP, SOI → NP). On the other hand the link PDO → NP does not seem to have changed. Combined these results suggest that the SST anomalies in the north Pacific have become more dependent on the equatorial Pacific conditions compared to local air-sea interactions.

7.4 CONCLUSIONS

We have used the pseudo transfer entropy (pTE), which is a simplified expression of the transfer entropy, to evaluate causal dependencies between thirteen indices that represent large-scale climate pattern. Taken together, our results have unveiled the well-known complexity of the network of interactions and feedback loops, and their interdecadal variations. The majority of the links recovered by our analysis have been documented in the literature and can be explained through known physical mechanisms; however, we have also found undocumented or likely spurious interactions. While it is important for advancing the understanding of our climate to identify the links that represent genuine connections, from a practical standpoint, to improve the forecast of an index variability, the pTE analysis yields useful knowledge because it tells us which signals contain information relevant for the future of another signal. In this way, the pTE represents a useful tool of time series analysis, to identify features that can potentially improve the forecast of the evolution of a climate index. As an example, we have found that the link between

AMO and ENSO is becoming stronger, which may be important for ENSO predictability.

Another application of the pTE algorithm is for performing model intercomparisons, e.g., for contrasting the causal links found in model data with those found in observed data, in order to determine the skill of different climate models in representing the interactions and lags in our climate.

Part III

PREDICTION OF THE MADDEN-JULIAN OSCILLATION

In the following part, we present our contribution to the prediction of the MJO. We show the machine learning prediction of the MJO, its prediction skill, the phase and amplitude errors, and how the seasons and initial MJO phases influence the predictions. We apply then machine learning as a post-processing technique to improve the current best numerical model's predictions.
The results we present in Chapter 8 have been published in Silini, Barreiro, and Masoller (2021), and those presented in Chapter 9 have been submitted for publication (Silini et al., 2022a).

8 PREDICTION OF THE MJO FROM OBSERVATIONS

8.1 DATA SET

In this study, we use the daily RMM indices. RMM1 and RMM2, as well as the phase and amplitude since June 1, 1974 were downloaded from *RMM data* (2021). The same tools used in this study could also be applied to other MJO indices, such as the OLR MJO index (OMI), the original OLR MJO index (OOMI), the real-time OLR MJO index (ROMI) and the filtered OLR MJO index (FMO), which can be downloaded from *MJO indices data* (2021).

Due to missing data in the first years we limit the study to the period between January 1, 1979 and December 31, 2020, which is L2-normalized.

8.2 PREDICTION SKILL METRICS

In order to assess the quality of the MJO predictions for a given model, we focus on its prediction skill, and its MJO amplitude and phase errors. To do so, we adopt the same quantifiers as in Kim, Vitart, and Waliser (2018), which are adapted from Lin, Brunet, and Derome (2008) and Rashid et al. (2011).

The bivariate correlation coefficient (COR) and the root-mean-squared error (RMSE) are defined as:

$$\mathrm{COR}(\tau) = \frac{\sum_{t=1}^{N}[a_1(t)b_1(t,\tau) + a_2(t)b_2(t,\tau)]}{\sqrt{\sum_{t=1}^{N}[a_1^2(t) + a_2^2(t)]}\sqrt{\sum_{t=1}^{N}[b_1^2(t,\tau) + b_2^2(t,\tau)]}}, \quad (29)$$

$$\mathrm{RMSE}(\tau) = \sqrt{\frac{1}{N}\sum_{t=1}^{N}[|a_1(t) - b_1(t,\tau)|^2 + |a_2(t) - b_2(t,\tau)|^2]}, \qquad (30)$$

where $a_1(t)$ and $a_2(t)$ are the observed RMM1 and RMM2 at time t, and $b_1(t,\tau)$ and $b_2(t,\tau)$ are the respective forecasts for time t with a lead time of τ days, and N is the number of predictions. COR expresses the strength of co-occurrence between the forecast and the observations, while RMSE does a term-by-term comparison of the actual difference between the forecast and the observations. The values COR=0.5 and RMSE=1.4 are usually used as skill thresholds (Rashid et al., 2011): the prediction skill refers to the time when the COR falls below 0.5 and RMSE grows above 1.4.

Through a change of coordinates from Cartesian to polar, we calculate the amplitude and phase, (RMM1, RMM2)→(A, φ) (Rashid et al., 2011) as follows

$$A(t) = \sqrt{\mathrm{RMM1}^2(t) + \mathrm{RMM2}^2(t)}, \qquad (31)$$

and

$$\varphi(t) = \tan^{-1}\left(\frac{\mathrm{RMM2}(t)}{\mathrm{RMM1}(t)}\right), \qquad (32)$$

and we define their errors as

$$\begin{aligned} E_A(\tau) &= \frac{1}{N}\sum_{t=1}^{N}[A_{pred}(t,\tau) - A_{obs}(t)], &(33)\\ E_\varphi(t,\tau) &= \frac{1}{N}\sum_{t=1}^{N}\tan^{-1}\left(\frac{a_1(t)b_2(t,\tau) - a_2(t)b_1(t,\tau)}{a_1(t)b_1(t,\tau)}\right), &(34)\end{aligned}$$

where $A_{obs}(t)$ is the observed amplitude at time t and $A_{pred}(t, \tau)$ is the predicted amplitude at time t with a lead time of τ days. a_1, a_2, b_1 and b_2 are the same used for Eqs. 29, 30.

8.3 RESULTS

8.3.1 *Prediction skill*

We begin by computing RMM COR and RMSE as a function of the forecast lead time, τ, for the two ANNs (see *Methods*). Averaging over all seasons we obtain the results shown in Fig. 23, where we display COR and RMSE as a function of $\tau = 5, 10, \ldots, 60$ days, for an initial RMM amplitude larger than 1. In this figure we see that both ANNs perform very similarly. The AR-RNN seems to perform slightly better than FFNN up to 10 days prediction, after which, the two curves overlap up to 50 days, when the latter starts providing a better prediction. Using the standard value COR=0.5 to define the prediction skill, we find a prediction skill of about 26–27 days for both ANNs, which is comparable to the best known prediction skills obtained from most models (Kim, Vitart, and Waliser, 2018), except ECMWF. Regarding the RMSE, using the standard value RMSE=1.4 to define the prediction skill, we see that the prediction skill is longer than 60 days, as for both ANNs, RMSE never crosses this value for τ values up to 60 days. A video (Silini, 2021b) showing the real and the predicted MJO evolution in the Wheeler-Hendon phase diagram clearly visualizes the very good prediction ability.

We then compute the error of the predictions for the MJO amplitude and phase (see *Methods*). The results are presented in Fig. 24, where we notice that, for both ANNs, the phase is well predicted but the amplitude is underestimated, and its absolute error grows as the lead time increases.

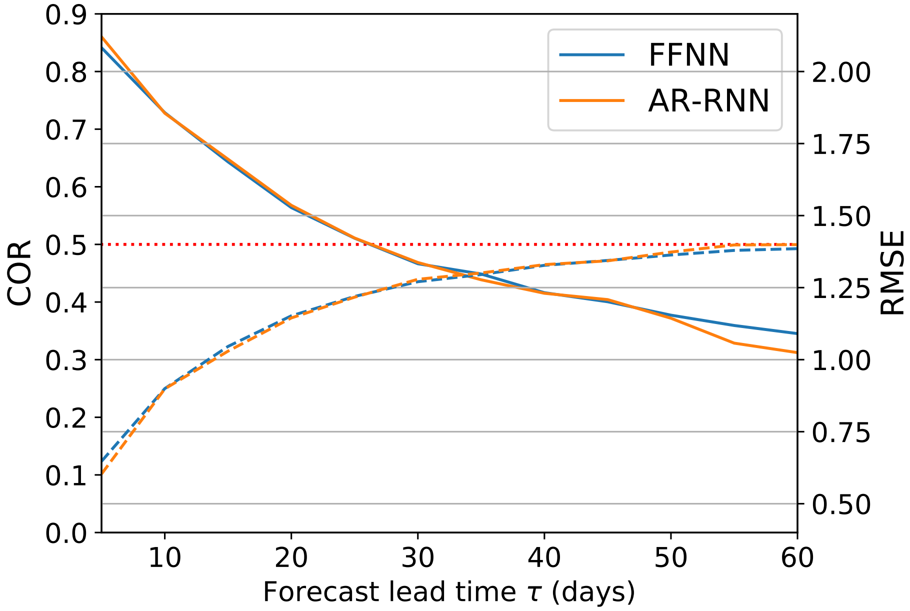

Figure 23: Bivariate correlation coefficient (COR) (solid) and root-mean-squared error RMSE (dashed) averaged over all seasons in the test set, as a function of the forecast lead time τ. The color indicates the artificial neural network (FFNN: feed-forward neural network; AR-RNN: autoregressive recurrent neural network) and the dotted line indicates the threshold that defines the prediction skill. While the RMSE threshold (1.4) is never crossed, the COR value falls below 0.5 around 26–27 days.

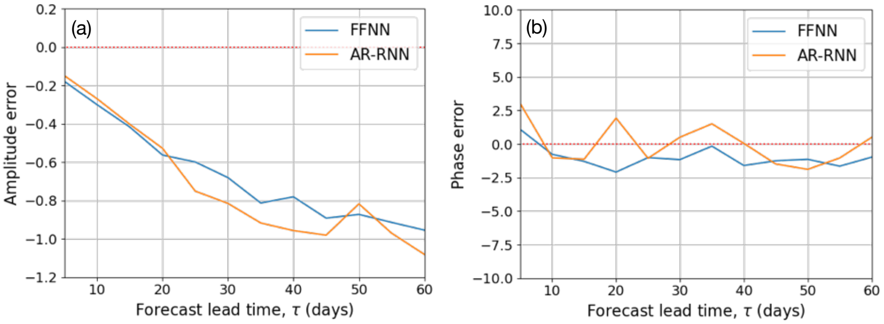

Figure 24: MJO amplitude (a) and phase (b) error averaged over all seasons in the test set, as a function of the lead time.

8.3.2 *Seasonally resolved prediction skill*

We now perform the same analysis for the dataset restricted to each season using the FFNN, which is the fastest and simplest of the two ANNs. The results are presented in Figs. 25 and 26.

In Fig. 25, we see a large difference in the prediction skill in different seasons. Boreal spring (March–May, MAM) and fall (September–November, SON), the transition seasons, are the least predictable with COR prediction skills of 23–24 days and 16–17 days, respectively. In boreal summer (June–August, JJA) the prediction skill is around 31 days, while in boreal winter December-February (DJF) it is around 45 days. We also note that DJF has the largest RMSE, which means that the prediction correlates well with the observations, but the predicted and actual values are quite different. On the contrary, JJA has a very low RMSE, which means that even if JJA has a lower COR than DJF, the prediction is more accurate. The transition seasons are in the middle, with SON showing larger RMSE than MAM, as found for COR. The highest COR and RMSE are for DJF, which is likely due to the fact that MJO is most active during the extended boreal winter (DJFM), which would also partially explain the large (yet smaller than DJF), RMSE of MAM.

Fig. 26 displays the amplitude and phase errors as a function of the lead time (as in Fig. 24, but here for the individual seasons). We notice that boreal winter (DJF) has the largest amplitude error, while boreal summer (JJA) has the lowest one. Regarding the phase error, we note that in JJA the predicted MJO propagation is faster than the real one, while in the other three seasons, the predicted propagation is slower.

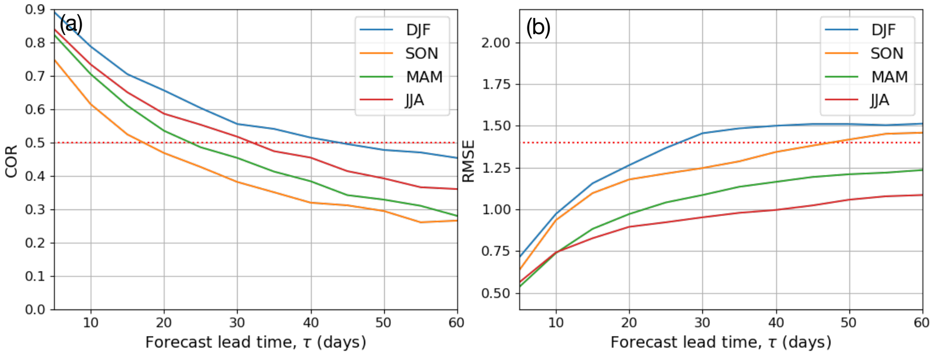

Figure 25: COR (a) and RMSE (b) as a function of the leading time in days, obtained with the feed-forward neural network (FFNN). The different colors represent different seasons.

Finally, we study the dependence of the COR and RMSE prediction skill as a function of the MJO initial phase and the season. The results are presented in Figs. 27 (COR) and 28 (RMSE). In boreal winter (DJF in blue), we can notice that starting from phase 1, 2, 5 and 8 the prediction skill using COR is very high, in fact, it has skill for up to 60 days or longer, while it falls below 20 days for phase 7. Nevertheless, Fig. 28 shows that for phases 5 and 8 the threshold is crossed below 30 days. By combining the information presented in the two figures, we can infer a prediction skill of about 60 days for phases 1 and 2.

For boreal fall (SON, orange) we also see a strong dependence of the skill on the initial phase: it is around 50 days for phases 4 and 7, while all other initial phases lead to prediction skills lower than 20 days. The skill in boreal spring (MAM, green) and summer (JJA, red) is more uniform

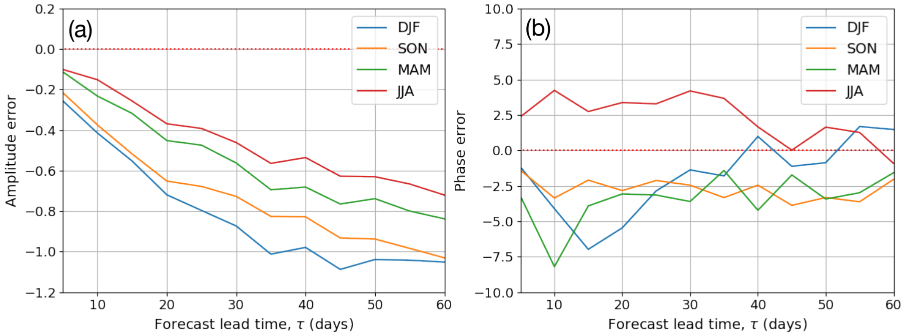

Figure 26: Amplitude (a) and phase (b) errors for the different seasons (represented with different colors), obtained with the FFNN.

across different initial phases, but the highest prediction skill achieved (given by COR) is around 40 days, and the lowest (below 20 days) are in phases 1, 3, 8 and 1, 5, 8, respectively. Overall, we can notice that the initial phase 1 provides a very high prediction skill in boreal winter, while it is low in all other seasons. Starting from phase 2, the prediction skill is larger than 35 days from December to May, while for initial phase 3 the highest prediction skill (around 40 days) is found in winter and summer. The initial phase 4 provides high skill (more than 40 days) in the transition seasons. Starting from initial phase 6, provides high skill from March to August, while starting from phase 7 gives a prediction skill above 40 days from June to November. Lastly, starting from phase 8 the prediction skill is always below 20 days.

In Fig. 28 we also notice that the RMSE for MAM and JJA never crosses the 1.4 threshold, for up to 100 days.

8.4 DISCUSSION

We have used two types of ANNs to predict the MJO. We have used a feed-forward neural network (FFNN) and an autoregressive recurrent

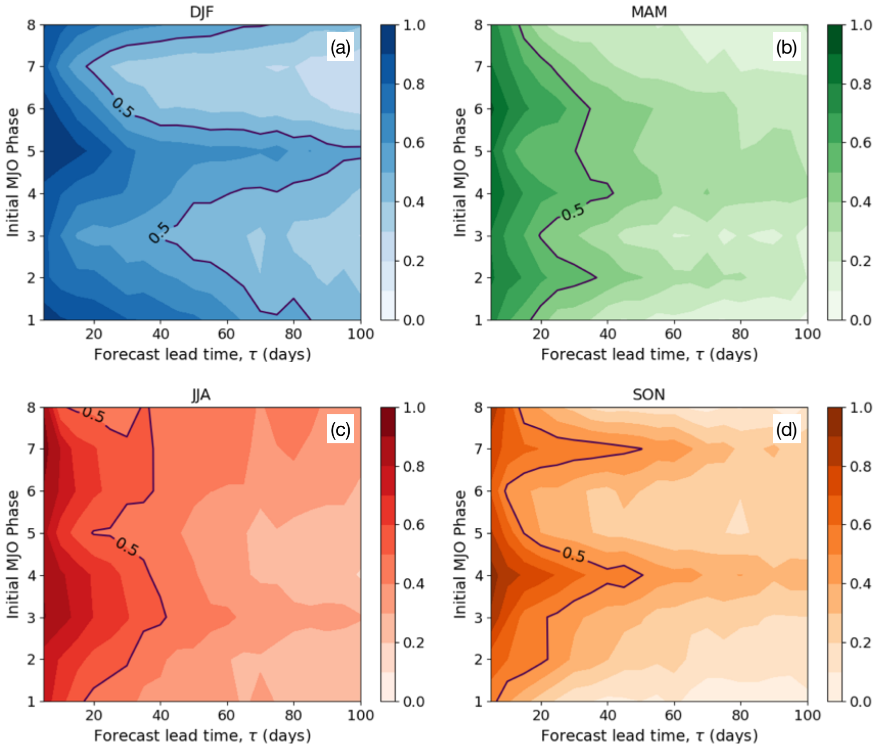

Figure 27: COR as a function of the initial MJO phase and forecast lead time τ. Each plot corresponds to a different season: boreal winter (a; blue), spring (b; green), summer (c; red) and fall (d; orange).

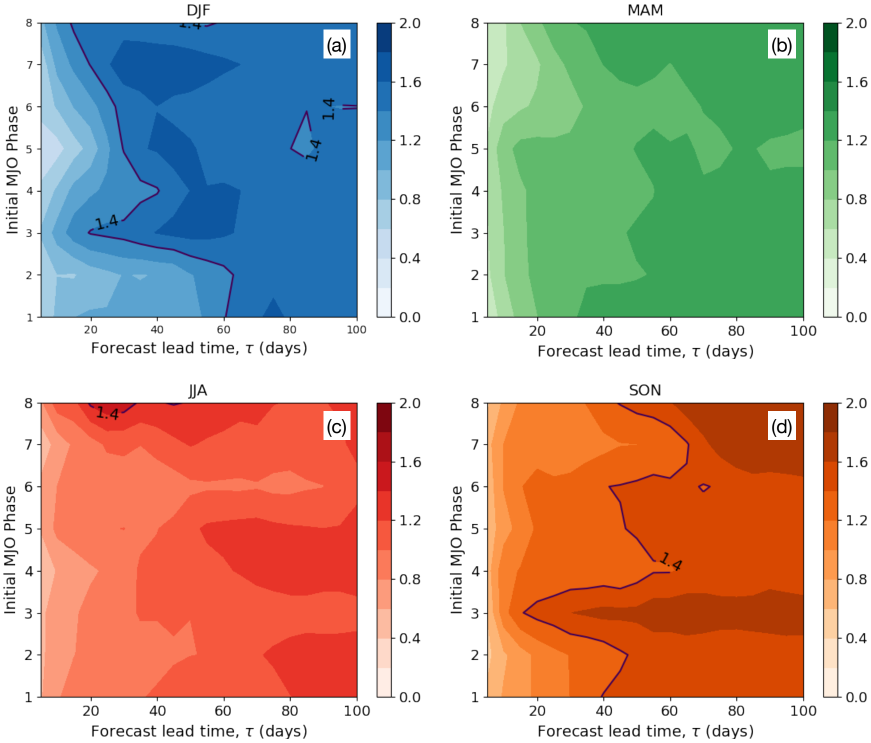

Figure 28: RMSE as a function of the initial MJO phase and forecast lead time τ. Each plot corresponds to a different season: boreal winter (a; blue), spring (b; green), summer (c; red) and fall (d; orange).

neural network (AR-RNN) to predict the daily Real-time Multivariate MJO indices, RMM1 and RMM2, analyzing the period between January 1, 1979 and December 31, 2020. First we considered the whole dataset, and in a second step, we considered individual seasons (boreal winter, DJF, spring, MAM, summer, JJA and fall, SON). We have quantified the prediction skill as a function of the leading time, τ, using standard magnitudes and thresholds (COR and RMSE with thresholds 0.5 and 1.4, respectively (Rashid et al., 2011)).

For the full dataset, using COR we have found a prediction skill of 26–27 days, which is comparable to most dynamical models. Using the RMSE, the prediction skill we have obtained is up to 60 days.

We have obtained a very good prediction of the RMM phase, but a poorer prediction of the RMM amplitude, which was systematically underestimated. Comparing these results with those reported in Vitart (2017), we notice that the two ANNs used here lead to a worse prediction of the amplitude, but to a better prediction of the phase, in comparison with the predictions obtained from most dynamical models. The larger amplitude error is due to the systematic underestimation, as the error adds up. In contrast, dynamical models sometimes overestimate and sometimes underestimate, which leads to a lower amplitude error, due to a partial compensation of positive and negative errors.

Consistent with previous studies (Lin et al., 2006; Rashid et al., 2011; Seo, 2009; Wheeler and Weickmann, 2001; Wu et al., 2016) we have found significant differences among seasons.

We found that boreal fall and spring have the lowest prediction skill, being 16–17 and 23–24 days, respectively. In accordance with Lin et al. (2006), Rashid et al. (2011), Seo (2009), and Wheeler and Weickmann (2001), we found the highest prediction skill in boreal winter, which in our case is of around 45 days. Another study (Wu et al., 2016) found the highest prediction skill in boreal fall. In boreal summer we have found a prediction skill of about 31 days.

We have also studied the dependence of the prediction skill as a function of the initial MJO phase. We have found a large variability in prediction skill in boreal winter and fall. In the best conditions, in boreal winter with an initial MJO phase of 1 and 2, the ANN has a prediction skill for up to 60 days or more. Our results indicate that the most difficult conditions to predict MJO is in boreal fall when the initial MJO phase is phase 1.

A major advantage of the ANNs considered is that they are computationally low-cost, and they do not have the limitations of dynamical models, where the MJO prediction skill depends strongly of the model's physics, initialization and ocean-atmosphere coupling processes. On the other hand, the very own nature of ANNs preclude understanding of the physical processes involved and thus they represent a complementary approach that, according to our results, is worth pursuing.

For future work, the MJO prediction skill could potentially be improved by training the ANNs independently for each season (for simplicity, here we have trained them on all seasons and test them on individual seasons). A study of the predictability barrier of the RMM index from different seasons and phases could also shed light on the results obtained with machine learning methods (Liu, Jin, and Rong, 2019).

Summarizing, in this chapter we presented a data-driven method to predict the MJO which, although computationally efficient, does not improve the prediction skill of (computationally demanding) state-of-the-art dynamical models. In the next chapter, we present an alternative approach which allows to exceed the current MJO best prediction skill.

9 IMPROVING THE PREDICTION OF MJO FROM MODEL DATA BY POST-PROCESSING

9.1 RMM DATA

For this study, we use the Real-time Multivariate MJO (RMM) index (Wheeler and Hendon, 2004) as labels for the supervised learning method, which is used to characterize the MJO geographical position and intensity. The first two principal components of the combined empirical orthogonal functions (EOFs) of outgoing longwave radiation (OLR), zonal wind at 200 and 850 hPa averaged between 15°N and 15°S are labeled RMM1 and RMM2. With a polar transformation, it is possible to define the MJO phase and amplitude. The phase is divided into 8 classes, each corresponding to a different sector of the phase diagram defining the observed MJO life cycle. The amplitude, describing the MJO intensity, when smaller than 1 defines a non-active MJO. The ERA5 RMM1 and RMM2 from 13th June 1999 to 29th June 2019 were downloaded from *ECMWF RMM reforecasts data* 2021. This time window is selected to match the ECMWF reforecasts, presented in the previous section.

9.2 ECMWF RMM REFORECASTS

The samples used as input for the ANN and to assess the model performance, are the ECMWF reforecasts with Cyrcle 46r1 freely available from *ECMWF RMM reforecasts data* 2021. This dataset is composed of 110 initial dates per year for 20 years, between the 13th June 1999 and the 29th June 2019. In total there are 2200 starting dates, from which a 46-lead-days prediction is available. The dataset provides the prediction of four variables: the first two principal components of the RMM index, and their polar transformation. For each starting day and variable there are 12 time series of 46 points. One is the controlled forecast (cf) corresponding to a forecast without any perturbations, then there are 10 perturbed forecasts members (pf) which have slightly different initial conditions from the cf to take into consideration errors in observations and the chaotic nature of weather. Finally there is the ensemble mean (em), which corresponds to the mean of the 11 members (cf + 10 pf). In this particular study, we made use solely of the em data.

9.3 RESULTS

The first part of this section will be devoted to the results obtained for the MJO amplitude and phase. In the second part we present the prediction skill assessment using the COR 0.5 level, and RMSE 1.4 level as metrics, while in the last part of the section we show how the different forecast methods perform for different MJO initial phases.

The results are obtained training the ANN from 13th of June 1999 using a walk-forward validation, and averaging the error obtained by testing over different unseen time windows from 5th December 2014 to 29th June 2019. The size of the windows is defined by the selected number of initial days from which the ECMWF forecast starts. Due to the bi-weekly acquisition of ECMWF, this means that each window of 200 points corresponds to 2 years approximately. Each member of the ensemble over which the average is performed, corresponds to a test set used for the walk-forward validation. Different sizes of the test set between 100 and 500 samples have been tested, leading to prediction skills that vary sensibly. For this reason, it is important to take into account that results may vary depending on the test set and its size, albeit preserving the same general result: the post-processing corrections improve the ECMWF forecasts.

In Fig. 29, we show the error on the MJO amplitude for events starting with an amplitude larger than 1. We can notice an underestimation of the amplitude as expected (Jiang et al., 2020). Nevertheless, the post-processed amplitudes are closer to the observed ones, with respect to the raw ECMWF forecast. The maximum improvement occurs for a lead time of 28 days when the ECMWF-ANN model has a RMSE similar to the RMSE of the uncorrected ECMWF at a lead time of 20 days.

By the definition of the amplitude error, errors of opposite sign could potentially cancel out resulting in misleading conclusions. For this reason in Fig. 29 we also provide the RMSE of the amplitude error, which shows a similar behavior as before. Both post-processing techniques improve the results, with the ANN bringing the highest benefits in terms of the magnitude of error reduction, and the forecasting horizon of the improvement.

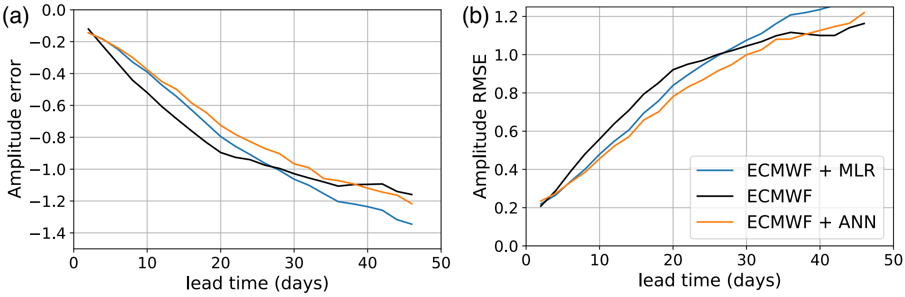

Figure 29: **(a)** MJO amplitude error and **(b)** amplitude RMSE (b) as a function of the lead time for events starting with an amplitude larger than 1. The color indicates the forecast model, the black line corresponds to the ECMWF forecast, the blue line corresponds to the MLR correction of the ECMWF forecast, while the orange line corresponds to the post-processed ECMWF forecast with an ANN.

In Fig. 30, we present the MJO phase error. The post-processing techniques provide an improved prediction, during which all three models predict a negative phase. A positive phase error indicates a faster propagating MJO, while a negative error represents a slower propagation. The ECMWF forecast shows an overall slower propagation of the MJO with respect to the observations, and both post-processing corrections provides an increment of the MJO speed prediction. In particular, at the 18 days lead time we can notice an increment of the ECMWF phase error, which MLR and ML tend to correct.

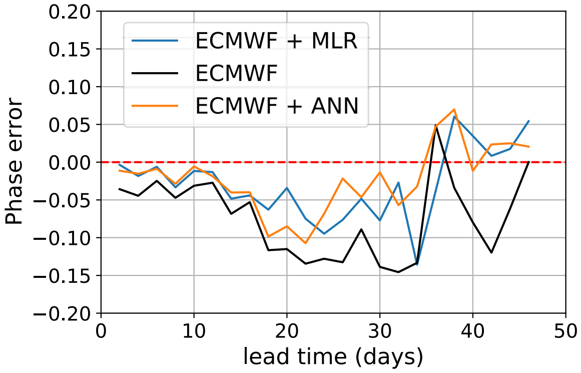

Figure 30: MJO phase error for events starting with an amplitude larger than 1. The color indicates the forecast model, the black line corresponds to the ECMWF forecast, the blue line corresponds to the MLR correction of the ECMWF forecast, while the orange line corresponds to the post-processed ECMWF forecast with an ANN.

Figure 31, shows the COR and RMSE of the ECMWF ensemble mean forecasts, the MLR, and ANN post-processing. A COR of 0.5 is taken here as baseline for useful prediction skill. We see an improvement of the a prediction skill at the COR=0.5 level of about 1 day. However, in terms of RMSE, up to a lead time of 4 weeks, neither post-processing technique crosses the RMSE-threshold of 1.4, and therefore, they both improve the prediction skill with respect to the raw, unprocessed output of the ECMWF model.

In Fig. 32, we display the comparison between the observations, the ECMWF forecast, and its corrections, in a Wheeler-Hendon phase diagram for two different starting dates of the same MJO event. The dots are marked every 7 days to identify the weeks. In the left panel, the 3 weeks prediction starts on the 21st November 2018 and displays its progression from the Western Hemisphere over the Indian Ocean. It is

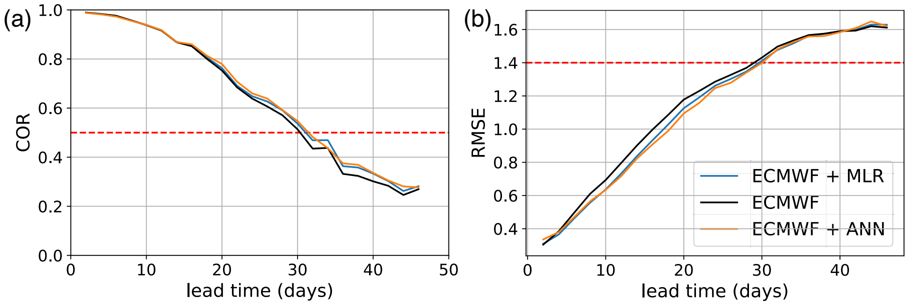

Figure 31: **(a)** COR and **(b)** RMSE as a function of the forecast lead time for events starting with an amplitude larger than 1. The color indicates the forecast model and the red dashed line indicates the prediction skill threshold of COR=0.5 and RMSE=1.4. The black line corresponds to the ECMWF forecast, the blue line corresponds to the post-processed ECMWF forecast with MLR, while in orange it is shown the post-processed ECMWF forecast with an ANN.

possible to notice that both post-processing techniques display very similar prediction, with a slightly larger amplitude than ECMWF, closer to the observations for all lead times. In the right panel, the 3 weeks prediction starts on the 5th December 2018 in the Indian Ocean. We can see a drop of accuracy in the ECMWF prediction, and the MLR post-processing, approaching the MC. The ML correction instead preserves a larger amplitude, closer to the observations.

It is also possible to notice that while the speed of the MJO event is well predicted in the left panel, in the right one there is a drop of the MJO speed forecast over the Indian Ocean and MC.

Here we presented an example of a strongly active MJO event, where the corrections clearly improve the ECMWF prediction and it is among the best found. All predictions from the 12th of December 2014 to the 18th of June 2019, can be found in Silini, 2021a. Looking at these results it is possible to appreciate the general improvement provided by the post-processing corrections.

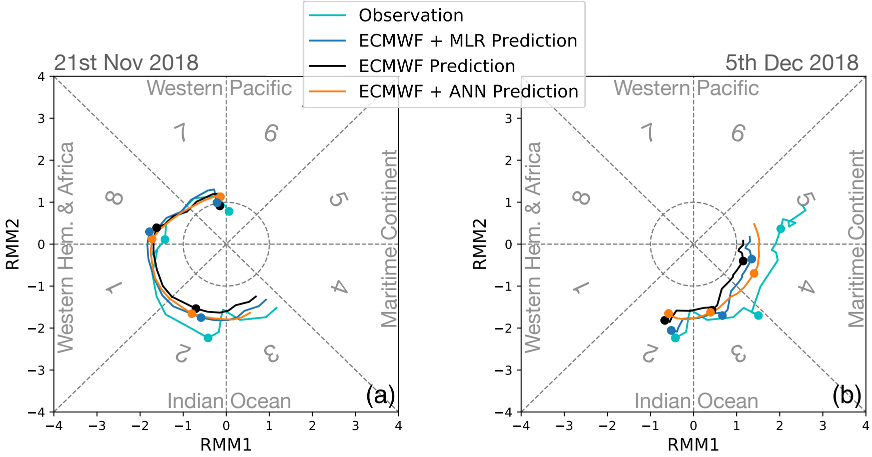

Figure 32: Wheeler-Hendon phase diagram for two different starting dates of the same MJO event, and a 3 weeks prediction. Panel **(a)** starting date is the 21st November 2018. The MJO enhanced rainfall region travels across the western Hemisphere and Indian Ocean. Panel **(b)** starting date is the 5th of December 2018, and represents a 3 weeks prediction approaching and traveling over the MC. The rotation of the event in the phase diagram is counter-clockwise, and the dots are included every 7 days, marking the different weeks.

Finally we study the amplitude error, the phase error, the COR, and RMSE, as a function of the different initial phases of MJO. As displayed in Fig. 33, applying post-processing methods improves the amplitude error for all initial phases. The MLR provides an improvement with respect to the ECMWF model, but the ML correction leads to the lowest error. Concerning the initial phases, we find the lowest amplitude error when an MJO event starts over the MC, while the largest is found in phase 2, over the Indian Ocean. With the MJO propagating at an average speed of 5 ms^{-1}, events starting in phase 2 will cross the MC in 2-3 weeks time (Kim et al., 2014). The phase error displays a large worsening of the MJO localization prediction, when the forecast starts between the MC and Western Pacific (phase 6-8). This observation is consistent with Fig. 32, where we noticed a drop in the accuracy of the MJO speed prediction over the Indian Ocean and MC. The COR finds its maximum when starting over the MC continent, consistently with the amplitude error. The ML correction has the highest COR except for phase 8, where MLR leads to the highest one. The RMSE is very consistent with the COR, in which we find the the minimum in phase 4, and the ML correction having the lowest error, except for phase 8. Overall, we can conclude that the ML post-processing is worth applying especially to reduce the error on the amplitude prediction, while MLR could be useful for a better prediction of the MJO location.

9.4 DISCUSSION

It is interesting to compare the results presented in Fig. 31 with those reported in Fig. 7 of the Supplementary Information in Kim et al., 2021, keeping in mind that Kim et al. show the mean BCOR for the 8 dynamical models considered. While it can be seen in Fig. 7 that for short lead times (up to 2 weeks) a clear improvement with DL-post-processing is obtained, the average BCOR for short lead times is quite low compared to the ECMWF prediction. We can also notice that the improvement obtained by Kim et al. (2021) fades away by the 4th week. In contrast, in our case, for short lead times there is no significant improvement (as it

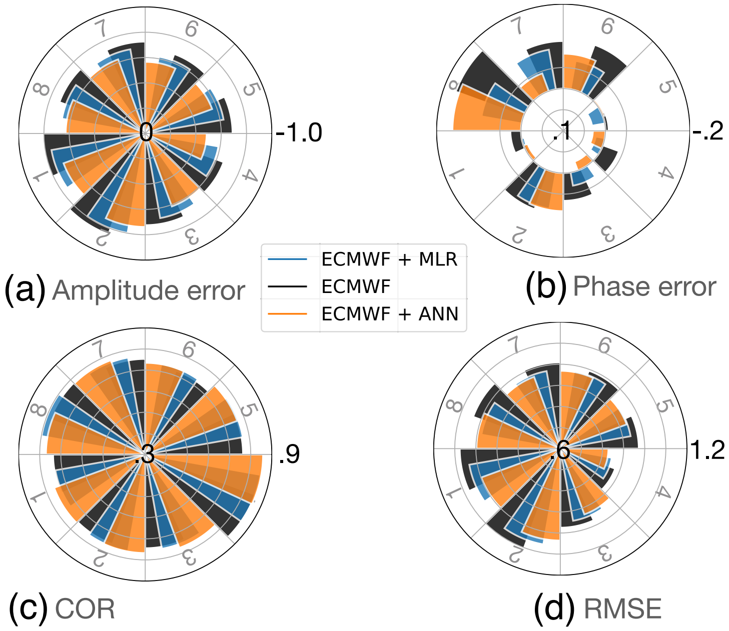

Figure 33: **(a)** Amplitude error, **(b)** phase error, **(c)** COR, and **(d)** RMSE, for the different MJO initial phases, for events starting with an amplitude larger than 1. The plots show the mean for lead times up to 5 weeks. The different colors represent the different prediction methods.

could be expected, due to the fact that ECMWF model provides the best MJO forecast), but our improvement lasts for longer lead times.

It is also interesting to compare the different post-processing approaches used. While we use a feedforward neural network (FFNN) architecture, Kim et al., 2021 used a Deep Learning (DL) network, specifically, a Long Short-Term Memory (LSTM) network. Having a simpler architecture, FFNNs are usually faster to train and to use than LSTMs. While LSTMs have been proven to be powerful for time sequence modeling, as shown in Kim et al., 2021, in our case we are not trying to predict the future of a time series using its past, but we are trying to improve the predictions.

There are other differences in the architecture of the networks used: we found that to improve the prediction of the RMMs for a day t, the information in the past and future predictions can both help the correction, while in Kim et al., 2021, the future model's predictions (which are available) are not used for the correction. Another difference is that while the algorithm used by Kim et al., 2021 performs an expansion of the system dimensionality (hidden nodes $>$ input nodes), we found good results performing a compression (hidden nodes $<$ input nodes).

Comparing our results to those of post-processing ensemble weather predictions on medium-range time-scales (Vannitsem et al., 2021), we find that the general magnitude of improvements over the predictions of the dynamical model is lower. This indicates the increasingly difficult challenge to obtain accurate MJO predictions for longer lead times, likely due to a generally lower predictability and a lower level of useful information that can be learned from the raw ensemble predictions compared to those of many other weather variables on shorter time scales. That said, our results indicating the potential of modern DL methods to improve over classical statistical approaches are well in line with the findings for medium-range post-processing (Haupt et al., 2021; Rasp and Lerch, 2018; Vannitsem et al., 2021).

We employed a MLR and a ML algorithm to perform a post-processing correction of the prediction of the dynamical model that currently holds the highest MJO prediction skill (Jiang et al., 2020), developed by ECMWF.

The largest improvement is found in the MJO amplitude and phase individually, which decreases the underestimation of the amplitude, providing a more accurate predicted geographical location of the MJO. The amplitude and phase estimation are improved for all lead times up to 5 weeks.

We obtained an improved prediction skill of about 1 day for a COR of 0.5.

Plotting the forecasts in a Wheeler-Hendon phase diagram we found an improvement predicting the MJO propagation, notably across the MC, which helps overcome the MC barrier.

Considering the results obtained for each initial MJO phase, we found that both post-processing tools improve the prediction, with the ML correction being the best.

The ML technique provides an improvement over MLR for all initial phases except phase 8. In the case of phase forecast it might be also sufficient to use MLR instead of ML. This suggests a predominance of linear corrections to improve the MJO phase forecast.

This study confirms the potential of post-processing techniques to reduce the knowledge and bias gap between dynamical models forecasts and observations, providing advancement in MJO prediction.

Although the improvement provided by the MLR and ML techniques, a post-processing method will always strongly rely on the accuracy of the dynamical model's forecasts. For this reason, it is crucial to work on both dynamical models and machine learning methods to progress.

Part IV

CONCLUSIONS

10 CONCLUSIONS

The work presented in this thesis was aimed at two main goals (as explained in Chapter 1). As previously mentioned, technological advances allow to have an increasingly finer resolution for data acquisition in many fields. While this improvement drives new discoveries and finer analysis, the amount of data to analyze increases exponentially, and sometimes, supercomputing is not the most cost-efficient solution. For this reason, the development of computationally cost-effective metrics is crucial.

We are often facing complex systems composed of a very large number of subsystems interacting between each other. Finding and quantifying the interactions among the subsystems from the observed data, is of capital importance to unveil the dynamics and structure of the system. Once the driving and driven processes are established, it is possible to predict the future of a driven process, knowing the past, and current state, of itself and its driving counterparts.

In Part II, we proposed a cost-effective metric for this purpose. We showed in Chapter 6 that with the pTE we significantly reduce the computational time with respect to conventional causality metrics, such as GC and TE, of well reputed Python libraries. For time series of 100 points, pTE reduces the computational time of `statsmodels` GC by ~50%, and by ~80% the time to compute the `pyunicorn` TE, albeit leading to very similar results. The proposed metric can be very valuable in a large variety of fields, where dynamical systems are composed of relatively short time series (<500 data points).

In Chapter 7, we showed an application of pTE to evaluate causal dependencies between thirteen indices that represent large-scale climate pattern. Aside from the conclusions we can infer on the interpretation of the interactions between specific climate indices, pTE showed its potential to identify those indices containing useful information to improve the prediction of a specific index. Without the complete knowledge of the system under study and the processes involved, it is not possible to identify which are actual causalities that are not due to indirect connections. For example, if two processes appear to be connected only because of a common driver, and the latter is not known or not considered for the analysis, there is no way to assess from the observed data whether there is a direct or indirect causality. For real complex systems that are not completely understood, it is quite common to miss information about all the processes that are actually involved. Nevertheless, while the complete knowledge of the system is crucial to build dynamical models, it is not the case for data-driven models, where both causal and indirect links provide helpful information to improve the forecast of a given index.

The pTE has also been successfully applied on the analysis of the influence of the major fire danger indices to the observed burned area, for each ecoregion in the world (Perez et al., 2022). The pTE is a metric that unveils the information transfer among processes. Since it computes the bivariate causality, it is very well suited for time series forecasting as preliminary tool for inputs selection of artificial neural networks, due to the latter suitability for problems with nonindependent inputs. In particular, it is very useful when dealing with complex systems which connections are poorly understood. It is important to notice that it is a metric that can be used on time series generated by a wide range of processes and fields. From climate to physiology, from finance to neurosciences, whenever we want to unveil causal interactions among variables evolving in time, pTE could provide a solution. Moreover, pTE could also be applied for performing model inter-comparisons and model validation, to check that the models under study preserve the same interactions as the observations. For this reason, we are positive about its applicability in future. On GitHub the metric has been forked and extended to include multiple user-friendly features, leading to the `AdapTE` function, whose core is the pTE, and it is in continuous development. Moreover, we implemented a generalization of the pTE on trivariate problems, which removes indirect links, and as future direction, it would be interesting to generalize the metric for multivariate systems.

In Part III, we put our effort in improving the state-of-the-art prediction of the MJO. In Chapter 8, we presented a pure machine learning approach for the forecast of the MJO, which back then was not yet explored in the literature. We showed its strengths and weaknesses, how it is resilient to the MC barrier, albeit not reaching the prediction skill of dynamical models. Counter to expectations, we found out that feed-forward neural networks performs very similarly to recurrent neural networks in this problem. Without any prior knowledge of the system, nor the physics of the MJO, a very small and simple ANN allows to outperform some dynamical models developed about a decade ago, in a matter of minutes. For this reason, and due to the success of ML in similar problems, we are confident that in the coming years ANNs will close the gap with the dynamical models. Moreover, the success of interpretable and explainable ANNs, brings forward another quality of ML, that gives insights, and improves the understanding of the underlying physical processes.

In Chapter 9, we explored another possible application of ML to the prediction of the MJO. This time, we don't stress the ANNs making it learn all MJO underlying structures from the observed data, and in particular, to learn what we already know thanks to years of research. In this chapter, we showed how ANNs can be used as post-processing technique, to correct and improve the predictions of the best dynamical model. We managed to improve both MJO intensity, and localization prediction; in particular, we did a step forward in overcoming the dynamical models' issue with the MC barrier. A future work that would be very interesting to explore, is to apply ML as post-processing for probabilistic forecasts, following the idea of Rasp and Lerch (2018). We are positive about the fact that the combination of two continuously progressing worlds will outperform models of both worlds, if taken individually.

A natural follow up to this work, is to combine the two Parts presented, by using ML as post-processing not only using the dynamical model's predictions, but also include further inputs selected with the pTE. This could potentially not only improve the prediction of the MJO, but also suggest which variables could be considered to refine the dynamical model.

Part V

APPENDIX

A AUTOREGRESSIVE MODELS

In order to compute the pTE, it is needed an estimation of the embedding parameters k and l, which correspond to the order of the Markov processes of the two time series that are considered to compute the pTE. This can be done by building autoregressive models of different degrees and by using a model selection criterion to estimate the best model degree that unveils the structure of the real data.

A.1 AUTOREGRESSIVE MODEL AR

Stationary stochastic time series can be modeled by the use of an autoregressive AR(p) process Z of a given order p as

$$z(t) = a_0 + \epsilon(t) + \sum_{i=1}^{p} a_i z(t-i), \tag{35}$$

where $\epsilon(t)$ is a white noise process, $a_1, \ldots, a_p$ are the autoregressive parameters and a_0 a constant. Once the AR estimates are built for a range of degrees, it is possible to apply a model selection criterion, from which it is extracted the degree of the AR model that fits the real data the best. When applied to the processes X and Y, the best models will be given by

$$\begin{aligned} x(t) &= a_{x,0} + \sum_{i=1}^{k} a_{x,i} x(t-i) + \epsilon_x(t), \quad \text{and} \\ y(t) &= a_{y,0} + \sum_{i=1}^{l} a_{y,i} y(t-i) + \epsilon_y(t). \end{aligned} \tag{36}$$

Due to the independence of pTE from the chosen model, the parameters that are used are k and l, which will play the role of the embedding size.

A.2 MOVING AVERAGE MODEL MA

With the AR model, we have seen how the value at time t of stationary stochastic time series can be computed using the values at the previous time steps. With the moving average MA(q) model we consider unexpected external factors, known as Errors or Residuals, that affect the time series. The effect of these residuals ϵ_i, is modulated by a set of parameters α_i, and we can write the value of process Z at time t, $z(t)$, as:

$$z(t) = \sum_{i=1}^{q} \alpha_i \varepsilon(t-i) + \varepsilon(t). \tag{37}$$

A.3 AUTOREGRESSIVE MOVING AVERAGE MODEL ARMA

In order to consider both the past values and the associated error, we can build a combination of the AR and MA models, obtaining an autoregressive moving average ARMA(p, q) model. In this case, for a process Z, we can write the model as

$$z(t) = a_0 + \varepsilon(t) + \sum_{i=1}^{p} a_i z(t-i) + \sum_{i=1}^{q} \alpha_i \varepsilon(t-i). \tag{38}$$

A.4 AUTOREGRESSIVE INTEGRATED MOVING AVERAGE MODEL ARIMA

Until now we considered process Z to be stationary. In case of non-stationarity, we can differentiate d times the time series, until reaching stationarity. In this case, we are modeling process Z as an autoregressive integrated moving average ARIMA(p, d, q), and we can write $z(t)$ as

$$z^{(d)}(t) = a_0 + \epsilon(t) + \sum_{i=1}^{p} a_i z^{(d)}(t-i) + \sum_{i=1}^{q} \alpha_i \epsilon(t-i), \tag{39}$$

where we denoted the d-th differentiation of $z(t)$ as $z^{(d)}(t)$.

A.5 AKAIKE AND BAYESIAN INFORMATION CRITERIA

In supervised learning problems there are many model selections criteria. From the very accurate but very expensive cross validation (Mosteller and Tukey, 1968) to global fit likelihood based criteria like Akaike information criterion (AIC) (Akaike, Petrov, and Csáki, 1973), Bayesian information criterion (BIC) (Schwarz, 1978), Deviance information criterion (DIC) (Spiegelhalter et al., 2002) and their shades, to criteria using particularly selected model parameters like the Focused information criterion (FIC) (Claeskens and Hjort, 2003).

The penalized-likelihood criteria like AIC and BIC find a balance between a good fit and a low computational cost. The precision on the choice of the embedding value for the pTE is not crucial, in fact it is sufficient to have a large enough embedding to not lose a causality coming from an higher order, and small enough to keep the computational cost low. For this reason, even if both the AIC and BIC have been criticized for having unrealistic asymptotic assumptions, they will provide a good estimate for the pTE embedding. Nevertheless, using both criteria ensure to not pick too big models, which is a tendency of AIC, nor too small models, that can happen using BIC.

The two criteria differs in how they penalize the number of parameters. The general information criterion (IC) can be written as

$$IC = P(p) - 2\ln\left(\hat{L}\right), \tag{40}$$

where $\hat{L}$ is the maximum value of the likelihood function for the model and $P(p)$ is the penalty function which corresponds to $2p$ for AIC and $\ln(n)p$ for BIC, where p is the number of parameters of the model and n the number of data points.

SIGNIFICANCE TESTS B

The causality tests' results could lead to inaccurate, or even erroneous conclusions, if not analyzed correctly. Let's take an example: consider two time series representing how many people are awake at a given time. The first one for people living in New York, and the second one in San Francisco. Due to the different time zones, we will find a strong causality from the first to the second one, which would suggest that the people awaking in New York would cause people in San Francisco to wake up. This is clearly a spurious causality, caused by a lagged external common forcing on the two processes. A meticulous preprocessing of the time series can avoid incurring in this kind of problems, but sometimes it is difficult to do so, due to lack of a complete knowledge of the system.

B.1 F TEST

Similar to a Z and T statistic for testing the statistical significance of a single variable, the F statistics gives a value to test if the variance between two populations means are significantly different. The F test will tell if a group of variables are jointly significant. The F value for each correlation is given by

$$F = \frac{s_1^2}{s_2^2}, \tag{41}$$

where s_1 and s_2 are the samples variances of the two populations. Using the size of the samples, it is possible to obtain the degrees of freedom (DoF) of the numerator and denominator of Eq. 41 by subtracting 1 to the populations sample sizes n_1 and n_2 respectively. The DoFs will define the shape of the F distribution, which is positive and asymmetric righttailed. Once the F distribution is obtained, using a confidence interval it is possible to discriminate the significant correlations arising by defining a critical value of F, i.e. when the F value is larger than the critical value it is possible to reject the null hypothesis. In the framework of this study, an F value will be calculated for every value of the pTE. Using as sample sizes the embedding size k used to compute the pTE, and the difference between the length of the time series N and two times the embedding size, i.e.

$$n_1 = k \qquad \text{and} \qquad n_2 = N - 2k, \tag{42}$$

the DoFs are easily obtained, defining the F distribution's shape The critical value is then computed using a p value of 0.05 with a Bonferroni corrected inverse cumulative distribution function.

B.2 SURROGATES

Another way to account for significance in the values of pTE is to create surrogates. There are multiple ways to create surrogates that allows to test against different null hypothesis. It is possible to test against noise, for nonlinearity or for independence. The testing against noise is done by creating white noise surrogates, or with random permutation (RP) surrogates (Theiler et al., 1992) to destroy a possible temporal structure in the real data. The surrogates for nonlinearity testing spans from autoregressive methods like the autoregressive moving average (ARMA) and the autoregressive integrated moving average (ARIMA), to Fourier transform surrogates which preserves the power spectrum such as the amplitude adjusted Fourier transform (AAFT), the iterative amplitude adjusted Fourier transform (IAAFT) and the iterative digitally filtered shuffled (IDFS) surrogates. They allow to avoid fitting of model parameters and it is not necessary to assume any model equation, by generating instead resamples that have a certain set of properties in common with the real data set. Finally for what concerns independence testing, the most used surrogates are the intersubject surrogates, cyclic phase permutations (CPP), the twin surrogates (TS) and the time-shifted surrogates.

B.2.1 *Independent Gaussian processes*

Arguably the simplest surrogate method is the white Gaussian noise surrogates to test against noise. To do that, two independent Gaussian processes with same length, mean and standard deviation as the real data are created.

B.2.2 *Iterative amplitude adjusted Fourier transform (IAAFT) surrogates*

Among the most commonly used surrogates to test for nonlinearity in data we find the amplitude adjusted Fourier transform (AAFT) and the iterative amplitude adjusted Fourier transform (IAAFT) surrogates. They have been extensively studied as well as their limits, and been used in a wide range of applications. Both AAFT and IAAFT are based on Fourier transform (FT) surrogates, which null hypothesis is to consider the data as generated by a stationary linear Gaussian process. The algorithmic procedure of FT surrogates is known as phase randomization, which preserves the power spectrum/autocorrelation, but destroys any nonlinear behavior. The considered data used here don't follow Gaussian distributions of values, and they could lead to false rejections of the null hypothesis of FT surrogates purely based on differences in amplitude distribution. To overcome this issue, the AAFT surrogates allows to preserve both power spectrum and amplitude distribution. The null hypothesis is that the data represent a rescaled linear Gaussian process. For finite data sets, the power spectrum of the AAFT surrogates appears flattened, and that's where IAAFT surrogates come into play. In fact IAAFT surrogates manage to reduce the whitening effect using an iterative approach which asymptotically leads to the same estimate of the spectral density and amplitude distribution as the original data. The null hypothesis in this case is that the data represent a stationary linear Gaussian process, measured through an invertible, time-independent instantaneous measurement function. It must be noted that the iteration involved in the IAAFT algorithm can't preserve both the power spectrum and amplitude distribution perfectly. The problem with FT and IAAFT surrogates is their strength as well, they can exactly preserve the power spectrum of the original time series. Therefore, the digitally filtered shuffled surrogates (DFS), as well as their iterative version (IDFS), introduces a controlled variation into the power spectrum, to account for possible real systems where the power spectrum of repeated processes varies.

In this work, it is used the IAAFT-1, which preserves the amplitude distribution exactly (while IAAFT-2 preserves the power spectrum exactly), since it has been observed that for mutual information the IAAFT-1 is the most successful between IAAFT-1, IAAFT-2 and DFS (Nichols and Murphy, 2016).

In Fig. 34, it is shown one example of a time series IAAFT surrogate, the Fast Fourier Transform (FFT) which represent the power spectrum, and the amplitude distribution. It is possible to notice the preservation of both power spectrum and amplitude distribution of the IAAFT algorithm, and the shape similarity between the actual time series with their surrogates.

The required number of surrogates heavily depends on the characteristics of the time series. In order to have an idea of how many are needed for the considered time series, 100 surrogates are created and tested to see how much the spread of values of the pTE varies as the number of surrogates increases. Since the computational cost is a limiting factor, and since the spread of the surrogate data is not large, for a one-sided test, $M = K/\alpha - 1$ surrogates can be generated, where K is a positive integer and α is the probability of false rejection, or p-value. Therefore, if $K = 1$ and for a pvalue α of 0.05, the number of needed surrogates would be 19 (Lancaster et al., 2018).

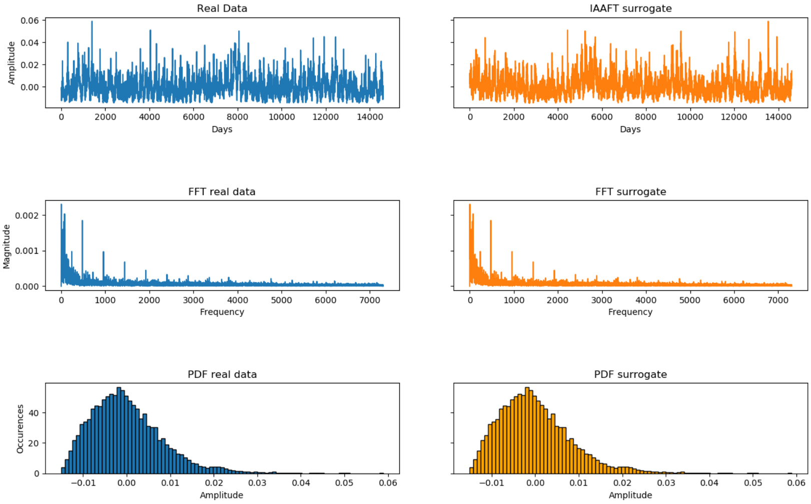

Figure 34: Comparison of a time series, its power spectrum, and its amplitude distribution, with an IAAFT surrogate.

B.2.3 *Time-shifted surrogates*

First proposed in Quian Quiroga et al. (2002) and further developed in Andrzejak et al. (2003), the time-shifted surrogates with periodic boundary conditions fully preserve all of the properties of the original time series. The algorithm to create time-shifted surrogates is computationally efficient: to generate a realization, it only requires the selection of a time shift and to wrap the time series end to its beginning. This analysis tests the hypothesis of having two linear or nonlinear processes without any significant cross-correlation and nonlinear interdependence (Lancaster et al., 2018). Due to the fact that the surrogates cannot be fully randomized, realizations consistent with the null hypothesis may occur with higher probability than for surrogates like IAAFT. For this reason, a larger number of surrogates should be used for testing.

C PTE ON SYNTHETIC DATA: SUPPLEMENTARY RESULTS

Table 5: List of DGPs studied for the comparison between pTE, GC and TE (the results are reported in Table 6). Models M0-M2 have no causality by construction. Models M3-M11 have causality from Y to X, while M12-M14 have bidirectional causality. M0 is Gaussian white noise, M1 is a bivariate process with a linear dependence, M2 corresponds to spurious causality and M3 corresponds to a nonlinear model (Taamouti, Bouezmarni, and Ghouch, 2014). M4 is a nonlinear model where the t-th point of process X is built using the an autoregressive model of order 2, and it's influenced by the $t-3$ value of process Y (Péguin-Feissolle and Teräsvirta, 2001). M5 is a heteroskedasticity mean causality, M6 a heteroskedasticity variance, while M7 is an homoskedasticity (Vilasuso, 2001). M8 and M9 have instantaneous causalities (Tjostheim, 1981), and M10 is a nonlinear ARX model (He et al., 2014). M11 are two Rössler systems (Rössler, 1976) coupled by the first variable. M12 and M13 are the circle map (Aragoneses et al., 2014) with unidirectional and bidirectional causality respectively. M14 has bidirectional causality (Taamouti, Bouezmarni, and Ghouch, 2014).

Model	X	Y	Causality
M0	ϵ_{1t} white noise	ϵ_{2t} white noise	$Y \not\rightarrow X$
M1	$x_t \sim \mathcal{N}(0, 1, \gamma_{xy}), \gamma_{xy} = 0.5$	$y_t \sim \mathcal{N}(0, 1, \gamma_{xy}), \gamma_{xy} = 0.5$	$Y \not\rightarrow X$
M2	$x_t = (0.01 + 0.5x_{t-1}^2)^{0.5}\epsilon_{1t}$	$y_t = 0.5y_{t-1} + \epsilon_{2t}$	$Y \not\rightarrow X$
M3	$x_t = 0.5x_{t-1}y_{t-1} + \epsilon_{1t}$	$y_t = 0.5y_{t-1} + \epsilon_{2t}$	$Y \rightarrow X$
M4	$x_t = 0.1 + 0.4x_{t-2} + \frac{2.4 - 0.9y_{t-3}}{1 + e^{-4y_{t-3}}} + \epsilon_{1t}$	$y_t = 0.7y_{t-1} + \epsilon_{2t}$	$Y \rightarrow X$

Model	X	Y	Causality
M5-M7	$x_t = 0.25x_{t-1} + 0.5y_{t-1} + \sigma_{1t}$	$y_t = 0.2 + 0.1y_{t-1} + \sigma_{2t}$	Y → X
	$\sigma_{it} = \eta_{it}\sqrt{H_{iit}}, \quad \eta_{it} \sim \mathcal{N}(0,1)$		
	$H = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} + A \begin{pmatrix} \sigma_{1t} \\ \sigma_{2t} \end{pmatrix} \begin{pmatrix} \sigma_{1t} \\ \sigma_{2t} \end{pmatrix}^T A^T$		
	$M5 : A = \begin{pmatrix} 0.2 & 0.0 \\ 0.0 & 0.9 \end{pmatrix} \quad M6 : A = \begin{pmatrix} 0.2 & 0.7 \\ 0.0 & 0.9 \end{pmatrix} \quad M7 : A = \begin{pmatrix} 0.0 & 0.0 \\ 0.0 & 0.0 \end{pmatrix}$		
M8	$x_t = 0.65x_{t-1} + 0.38y_{t-1} + 0.01x_{t-2}$ $- 0.21y_{t-2} + \epsilon_{1t}$	$y_t = 1.29y_{t-1} + 0.18x_{t-1} - 0.35y_{t-2}$ $- 0.16x_{t-2} + \epsilon_{2t}$	Y → X
M9	$x_t = 0.06x_{t-1} - 1.14y_{t-1} + 0.48x_{t-2}$ $+ 0.51y_{t-2} - 0.23x_{t-3} - 0.51y_{t-3} + \epsilon_{1t}$	$y_t = 1.1y_{t-1} - 0.09x_{t-1} - 0.36y_{t-2}$ $- 0.29x_{t-2} + 0.09y_{t-3} - 0.15x_{t-3}\epsilon_{2t}$	Y → X
M10	$x_t = 0.5x_{t-1} - 0.3x_{t-2} + 0.1y_{t-2} + 0.1x_{t-2}^2 +$ $+ 0.4y_{t-1}y_{t-2} + \epsilon_{1t}$	$y_t = \sin(4\pi t) + \sin(6\pi t) + \epsilon_{2t}$	Y → X
M11	$\dot{x}_1 = -(1 + 0.015)x_2 - x_3 + 0.1(y_1 - x_1)$ $\dot{x}_2 = (1 + 0.015)x_1 + 0.15x_2$ $\dot{x}_3 = 0.2 + x_3(x_1 - 10)$	$\dot{y}_1 = -(1 - 0.015)y_2 - y_3$ $\dot{y}_2 = (1 - 0.015)y_1 + 0.15y_2$ $\dot{y}_3 = 0.2 + y_3(y_1 - 10)$	Y → X
M12-M13	$x_t = \Big(x_{t-1} + \rho + \frac{K}{2\pi}\sin(2\pi x_{t-1}) +$ $+ \beta\epsilon_{1t}\Big) \mod 1 + C_1(x_{t-1} - y_{t-1})$	$y_t = \Big(y_{t-1} + \rho + \frac{K}{2\pi}\sin(2\pi y_{t-1}) +$ $+ \beta\epsilon_{2t}\Big) \mod 1 + C_2(y_{t-1} - x_{t-1})$	Y ↔ X
	$\rho = 0.23, \quad K = 0.04, \quad \beta = 0.002$ M12: $C_1 = 0.5, C_2 = 0,$ M13: $C_1 = C_2 = 0.5$		
M14	$x_t = 0.3 + 0.15x_{t-1} + 0.7y_{t-1} + \epsilon_{1t}$	$y_t = 0.2 + 0.1y_{t-1} + 0.8x_{t-1} + \epsilon_{2t}$	Y ↔ X
	$\begin{pmatrix} \epsilon_{1t} \\ \epsilon_{2t} \end{pmatrix} \sim \mathcal{N}\left[\begin{pmatrix} 0.0 \\ 0.0 \end{pmatrix}, \begin{pmatrix} 1 & 0.2 \\ 0.2 & 1 \end{pmatrix}\right]$		

Table 6: Power and size obtained with the DGPs listed in Table 5 using pTE, GC and TE. We can notice that there are no significant differences between pTE and GC. The results were obtained using time series of length 1000, where the first 100 are discarded and they are averaged over 1000 realizations. The last three columns correspond to the directionality index DI, eg. $(\mathrm{pTE}_{Y\to X}-\mathrm{pTE}_{X\to Y})/(\mathrm{pTE}_{Y\to X}+\mathrm{pTE}_{X\to Y})$, which shows that pTE performs better in most of the models in assessing the directionality. The pTE has been calculated with an embedding parameter of 1 for all models except for M10, where an embedding parameter of 2 has been used to match the causality lag imposed by construction.

Model	pTE		GC		TE		DI		
	Y → X	X → Y	Y → X	X → Y	Y → X	X → Y	pTE	GC	TE
M0	3.8	3.9	5.1	5.0	4.4	4.4	−0.01	0.01	0.00
M1	2.3	2.6	3.3	3.1	100	100	−0.06	0.03	0.00
M2	4.2	4.7	5.5	5.9	4.7	4.9	−0.06	−0.04	−0.02
M3	100	4.5	100	4.8	70.2	5.6	0.91	0.91	0.85
M4	80.7	3.8	84.2	4.9	96.0	4.7	0.91	0.89	0.91
M5	100	2.2	100	3.1	100	3.8	0.96	0.94	0.93
M6	100	1.8	100	2.8	100	4.3	0.96	0.95	0.92
M7	100	2.8	100	3.4	100	4.0	0.95	0.93	0.92
M8	100	4.5	100	5.6	100	100	0.91	0.89	0.00
M9	100	0.1	100	0.1	100	100	1.00	1.00	0.00
M10	62.6	3.1	67.3	4.3	12.2	4.5	0.91	0.88	0.46
M11	46.1	43.1	53.1	49.8	37.8	45.0	0.03	0.03	−0.09
M12	99.9	1.0	100	0.9	100	0	1.0	1.0	1.0
M13	100	100	100	100	100	100	0.00	0.00	0.00
M14	100	100	100	100	100	100	0.00	0.00	0.00

ARTIFICIAL NEURAL NETWORKS D

D.1 MJO FORECAST

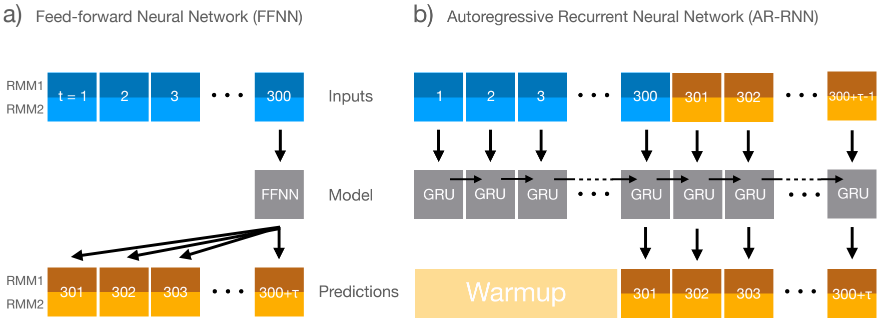

Figure 35: Diagram of the ANNs employed in this study. Panel (a) represents the FFNN, while panel (b) the AR-RNN.

In this study we use two well-known ANNs, schematically shown in Fig. 35 a feed-forward neural network (FFNN) and an autoregressive recurrent neural network (AR-RNN), both having an input layer of 300 units.

The FFNN uses the last point of the input layer and links it to one hidden layer composed of 64 units, itself linked to an output layer of τ units fully connected, where $\tau = 5, 10, \dots, 100$ is the forecast lead time. Each input and output is composed by two values, corresponding to RMM1 and RMM2, as shown in Fig. 7 panel a.

The AR-RNN is a single Gated Recurrent Unit (GRU) (Cho et al., 2014) layer composed of 64 units, displayed in Fig. 35 panel b. Instead of predicting the entire output sequence in a single step, with this recurrent neural network we decompose the prediction into individual time steps that are fed back into the network after a warm-up, which updates the internal state of the network and discards the outputs considering them poor predictions. GRU is chosen over a classical RNN to prevent the vanishing gradient problem, which corresponds to the potential tendency of the loss function gradients to approach zero, making the backpropagation of the error to not affect the first layers neurons of a multi-layer network. It is also preferred over a long short-term memory ANN due to the lower computational time required. Since we don't have several hidden layers, the vanishing gradient problem is not an issue, and in this way we leave open the possibility of increasing the number of layers for achieving a better prediction skill.

For the FFNN the activation function is a rectified linear activation function (ReLU), which is responsible for transforming the summed weighted input from the node into the activation of the node or output for that input. Sigmoid functions generally work better in the case of classifiers, and just like `tanh` functions might be avoided due to the vanishing gradient problem. If by increasing the number of hidden neurons one might encounter multiple dead neurons, i.e. non active neurons, we suggest using the leaky version of ReLU, or its parameterized version.

The Mean Squared Error (MSE) is used as loss function, which is the default loss used for regression problems and the RMM values are not widely spread and do not have outliers, which motivates this choice instead of using Mean Squared Logarithmic Error (MSLE) or Mean Absolute Error (MAE).

Finally, the Adam optimizer is used for training, with a maximum of 10 epochs. We selected a patience of 1, used for the early stopping of the training to avoid overtraining, which corresponds to the delay in stopping. Adam optimizer is chosen being the best common method among adaptive optimizers, which doesn't require a tuning of the learning rate value. The maximum number of epochs is never reached as the learning is stopped if the validation error starts growing. We could increase the patience to account for possible local minima of the validation error, but that would require more computational time, and we preferred to use fast and simple ANNs for a demonstration of their ability for MJO prediction.

To perform the backtesting, or hindcast, we selected a train-validation-test splitting that preserves the temporal order of observations. Other methods like multiple train-test splits or the walk-forward validation could be applied and would result in a more robust estimation of the model performance on out of sample data. The drawback of such methods is the cost of creating multiple models, which would sensibly slow down the training.

The dataset is divided in three sets: the *train set* contains data from 1.1.1979 to 30.11.2006, the *validation set*, from 1.12.2006 to 30.11.2015, and the *test set*, from 1.12.2015 to 31.12.2020.

The ANNs are trained on the *train set*, and the model's internal parameters are updated every 16 (batch size) exposure of different training samples. After the training, the ANN is evaluated using the *validation set* to fine-tune the hyperparameters. This training and validation process is repeated a maximum of 10 times. Then, a single evaluation is performed using the *test set*, which was not previously seen by the ANNs.

D.2 MJO FORECAST POST-PROCESSING

The post-processing machine learning tool built for this study is a fully connected feedforward neural network (FFNN) composed of an input layer containing N_{in} neurons, a single hidden layer with N_h neurons, and an output layer with N_{out} neurons, as shown in Fig 36. The activation function used is the Rectified Linear Unit (ReLU), which transforms the weighted sum of the input values by returning 0 in case of a negative-sum, and the result of the sum otherwise. Dealing with a supervised regression problem, the mean-squared error (MSE) is extensively employed as loss function, and it is used in the framework of this study

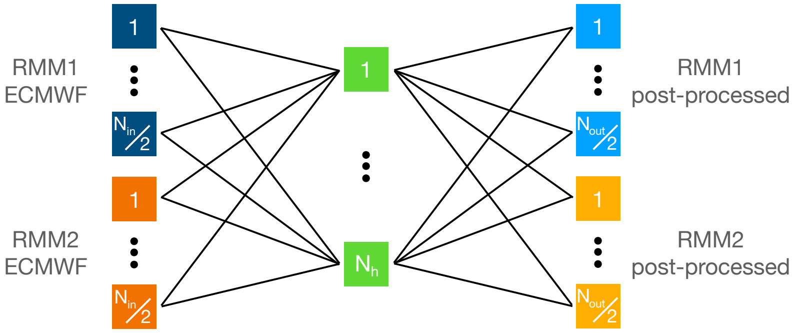

Figure 36: ANN architecture employed for this study.

to compare the neural network output with the observations (labels). An adaptive optimizer (Adam) is selected to automatically manage the learning rate during the training phase.

We use an adaptive number of neurons depending on the number of days we want to forecast. The ECMWF reforecasts provide predictions up to a lead time of 46 days for both RMM1 and RMM2, and we build a different network for each lead time. This means that the number of output neurons N_{out} can fall between 2 and 92 because we use both RMM1 and RMM2.

After selecting the number of output neurons (which is even and in fact defines our lead time, $\tau = N_{out}/2$), we adapt the number of input N_{in} and hidden neurons N_h as follows. As input, the networks receive the ECMWF reforecasts, which also limit the number of input neurons N_{in} in the range between 2 and 92. After training the networks with different N_{in}, we found the best result is obtained with $N_{in} = N_{out} + 6$ with an upper limit of 92. This means that for all lead times $\tau > 44$, $N_{in} = N_{out} = 92$. For lead times larger than 30-35 days, the prediction skill of the models falls below the thresholds of 0.5 and 1.4 imposed by the COR and RMSE, respectively, and thus, the lead times $\tau > 44$ are not crucial. Using all 92 inputs, the prediction skill for short lead times slightly decreases. For simplicity, a fixed number of 92 inputs could also be used. An interpretation of the reason behind this result is that to correct the prediction values for a given day, the future predicted values can help up to some extent. To correct the prediction of a given day, for each RMM we use the predicted values of up to 3 days after that particular day. To avoid overfitting, we want the number of hidden neurons to be relatively small, for this reason after some tests, we select $N_h = N_{in}/2$. The training has been performed over 100 epochs which allows to not overfit the model. The model performance is tested using a walk-forward validation. The procedure is as follows. First, we train the network on an expanding train set, and then test its performance on a validation set that contains the N samples that follow the train set. In our case, we found the best minimum number of samples for the train set, out of 2200 available, to be 1700. Then, the train set is extended by 100 samples (~1 year) for each run, and validated on the subsequent 200 samples (~2 years). This method of walk-forward validation ensures that no information coming from the future of the test set is used to train the model. Other methods to avoid overfitting could also be used, such as early stopping or drop-out.

MLR in the ordinary least squares (OLS) linear regression where the observed RMMs are a linear combination of the ECMWF-predicted RMMs. To compute the MLR we adopt the Python library `scikit-learn` (Pedregosa et al., 2011). With MLR we correct the RMMs separately, and apply the same walk-forward validation used for the ANNs.

BIBLIOGRAPHY

Akaike, H., B. N. Petrov, and F. Csáki (1973). "Information theory and an extension of the maximum likelihood principle." In: *2nd International Symposium on Information Theory, Tsahkadsor, Armenia, USSR*, pp. 267–281.

Alvarez, Mariano S., Carolina S. Vera, and George N. Kiladis (2017). "MJO Modulating the Activity of the Leading Mode of Intraseasonal Variability in South America." In: *Atmosphere* 8.12. ISSN: 2073-4433. DOI: https://doi.org/10.3390/atmos8120232.

Amblard, P. and O. Michel (2013). "The relation between granger causality and directed information theory: A review." In: *Entropy* 15.1, pp. 113–143. DOI: http://dx.doi.org/10.3390/e15010113.

Andrews, Martin B, Jeff R Knight, Adam A Scaife, Yixiong Lu, Tongwen Wu, Lesley J Gray, and Verena Schenzinger (2019). "Observed and simulated teleconnections between the stratospheric quasi-biennial oscillation and Northern Hemisphere winter atmospheric circulation." In: *Journal of Geophysical Research: Atmospheres* 124.3, pp. 1219–1232.

Andrzejak, Ralph G., Alexander Kraskov, Harald Stögbauer, Florian Mormann, and Thomas Kreuz (2003). "Bivariate surrogate techniques: Necessity, strengths, and caveats." In: *Physical Review E* 68 (6), p. 066202. DOI: 10.1103/PhysRevE.68.066202.

Aragoneses, A., S. Perrone, T. Sorrentino, M. C. Torrent, and C. Masoller (2014). "Unveiling the complex organization of recurrent patterns in spiking dynamical systems." In: *Scientific Reports* 4, pp. 1–6. DOI: http://dx.doi.org/10.1038/srep04696.

Baccala, L. A. and K. Sameshima (2001). "Partial directed coherence: a new concept in neural structure determination." In: *Biological Cybernetics* 84, p. 463. DOI: http://dx.doi.org/10.1007/PL00007990.

Barnett, L., A. B. Barrett, and A. K. Seth (2009). "Granger causality and transfer entropy are equivalent for Gaussian variables." In: *Physical Review Letters* 103.23, p. 238701. DOI: http://dx.doi.org/10.1103/PhysRevLett.103.238701.

Barnett, L. and A. K. Seth (2014). "The MVGC multivariate Granger causality toolbox: A new approach to Granger-causal inference." In: *Journal of Neuroscience Methods* 223, pp. 50–68. DOI: http://dx.doi.org/10.1016/j.jneumeth.2013.10.018.

Barrett, Bradford S., Casey R. Densmore, Pallav Ray, and Elizabeth R. Sanabia (2021). "Active and weakening MJO events in the Maritime Continent." In: *Climate Dynamics*. DOI: https://doi.org/10.1007/s00382-021-05699-8.

Bergman, John W., Harry H. Hendon, and Klaus M. Weickmann (2001). "Intraseasonal Air–Sea Interactions at the Onset of El Niño." In: *Journal of Climate* 14.8, pp. 1702–1719. ISSN: 08948755, 15200442.

Bhaskar, Ankush, Durbha Sai Ramesh, Geeta Vichare, Triven Koganti, and S Gurubaran (2017). "Quantitative assessment of drivers of recent global temperature variability: an information theoretic approach." In: *Climate Dynamics* 49. DOI: https://doi.org/10.1007/s00382-017-3549-5.

Bielczyk, N. Z. and et al. (2019). "Disentangling causal webs in the brain using functional magnetic resonance imaging: A review of current approaches." In: *Network Neuroscience* 3, p. 1. DOI: http://dx.doi.org/10.1162/netn_a_00062.

Bochenek, Bogdan and Zbigniew Ustrnul (2022). "Machine Learning in Weather Prediction and Climate Analyses–Applications and Perspectives." In: *Atmosphere* 13.2. DOI: https://doi.org/10.3390/atmos13020180.

Camargo, Suzana J., Matthew C. Wheeler, and Adam H. Sobel (2009). "Diagnosis of the MJO Modulation of Tropical Cyclogenesis Using an Empirical Index." In: *Journal of the Atmospheric Sciences* 66.10, pp. 3061 –3074. DOI: https://doi.org/10.1175/2009JAS3101.1.

Chang, Ping, R Saravanan, and Link Ji (2003). "Tropical Atlantic seasonal predictability: The roles of El Niño remote influence and thermodynamic air-sea feedback." In: *Geophysical Research Letters* 30.10.

Chen, Y., G. Rangarajan, J. Feng, and M. Ding (2004). “Analyzing multiple nonlinear time series with extended Granger causality.” In: *Physical Letters A* 324, p. 26. DOI: http://dx.doi.org/10.1016/j.physleta.2004.02.032.

Chiou-Wei, S. Z., C-F. Chen, and Z. Zhu (2008). “Economic growth and energy consumption revisited: evidence from linear and nonlinear Granger causality.” In: *Energy Economics* 30.6, pp. 3063–3076. DOI: http://dx.doi.org/10.1016/j.eneco.2008.02.002.

Cho, Kyunghyun, Bart van Merriënboer, Caglar Gulcehre, Dzmitry Bahdanau, Fethi Bougares, Holger Schwenk, and Yoshua Bengio (2014). “Learning Phrase Representations using RNN Encoder–Decoder for Statistical Machine Translation.” In: *arXiv:1406.1078*. URL: https://arxiv.org/abs/1406.1078.

Claeskens, G. and N. L. Hjort (2003). “The focused information criterion.” In: *Journal of the American Statistical Association* 98, pp. 879–899. DOI: http://dx.doi.org/10.1198/016214503000000819.

Cliff, Oliver M., Leonardo Novelli, Ben D. Fulcher, James M. Shine, and Joseph T. Lizier (2021). “Assessing the significance of directed and multivariate measures of linear dependence between time series.” In: *Physical Review Research* 3 (1), p. 013145. DOI: http://dx.doi.org/10.1103/PhysRevResearch.3.013145.

Delgado-Bonal, Alfonso, Alexander Marshak, Yuekui Yang, and Daniel Holdaway (2020). “Analyzing changes in the complexity of climate in the last four decades using MERRA-2 radiation data.” In: *Scientific Reports* 10. DOI: https://doi.org/10.1038/s41598-020-57917-8.

Deza, J. I., M. Barreiro, and C. Masoller (2015). “Assessing the direction of climate interactions by means of complex networks and information theoretic tools.” In: *Chaos* 25, p. 033105. DOI: http://dx.doi.org/10.1063/1.4914101.

Dhamala, M., G. Rangarajan, and M. Ding (2008). “Estimating Granger causality from Fourier and wavelet transforms of time series data.” In: *Physical Review Letters* 100.1. DOI: http://dx.doi.org/10.1103/PhysRevLett.100.018701.

Díaz, Nicolás, Marcelo Barreiro, and Nicolás Rubido (2020). “Intraseasonal Predictions for the South American Rainfall Dipole.” In: *Geo-*

physical Research Letters 47.21. DOI: https://doi.org/10.1029/2020GL089985.

Dijkstra, Henk A., Emilio Hernandez-Garcia, Cristina Masoller, and Marcelo Barreiro (2019). *Networks in Climate*. Cambridge, U. K.: Cambridge University Press.

Diks, C. G. H. and J. DeGoede (2001). *A general nonparametric bootstrap test for Granger causality*. London: Institute of Physics.

Donges, J.F. et al. (2015). "Unified functional networkand nonlinear time series analysis for complex systems science: The pyunicorn package." In: *Chaos* 25, p. 113101. DOI: http://dx.doi.org/10.1063/1.4934554.

ECMWF RMM reforecasts data (2021). https://acquisition.ecmwf.int/ecpds/data/list/RMMS/ecmwf/reforecasts/. Accessed: 2021-02.

Explorer, Climate (2020). *NINO3.4 from = "https://climexp.knmi.nl/start.cgi"*.

Faes, L., G. Nollo, and A. Porta (2011). "Information-based detection of nonlinear granger causality in multivariate processes via a nonuniform embedding technique." In: *Physical Review E* 83.5, p. 051112. DOI: http://dx.doi.org/10.1103/PhysRevE.83.051112.

Faes, L., G. Nollo, and A. Porta (2013). "Non-uniform multivariate embedding to assess the information transfer in cardiovascular and cardiorespiratory variability series." In: *Computers in Biology and Medicine* 42.3, pp. 290–297. DOI: http://dx.doi.org/10.1016/j.compbiomed.2011.02.007.

Fauchereau, Nicolas, Benjamin Pohl, and A. Lorrey (2016). "Extratropical impacts of the Madden-Julian oscillation over New Zealand from a weather regime perspective." In: *Journal of Climate* 29.6, pp. 2161–2175. DOI: https://doi.org/10.1175/JCLI-D-15-0152.1.

Ferranti, Laura, Linus Magnusson, Frédéric Vitart, and David S. Richardson (2018). "How far in advance can we predict changes in large-scale flow leading to severe cold conditions over Europe?" In: *Quarterly Journal of the Royal Meteorological Society* 144.715, pp. 1788–1802. DOI: https://doi.org/10.1002/qj.3341.

Fowler, M. D. and M. S. Pritchard (2020). "Regional MJO Modulation of Northwest Pacific Tropical Cyclones Driven by Multiple Transient

Controls." In: *Geophysical Research Letters* 47.11, e2020GL087148. DOI: https://doi.org/10.1029/2020GL087148.

Fulton, C. (2020). *Statsmodels = https://github.com/statsmodels/statsmodels*.

Garfinkel, Chaim I., James J. Benedict, and Eric D. Maloney (2014). "Impact of the MJO on the boreal winter extratropical circulation." In: *Geophysical Research Letters* 41.16, pp. 6055–6062. DOI: https://doi.org/10.1002/2014GL061094.

Gibson, Peter B, William E Chapman, Alphan Altinok, Luca Delle Monache, Michael J DeFlorio, and Duane E Waliser (2021). "Training machine learning models on climate model output yields skillful interpretable seasonal precipitation forecasts." In: *Communications Earth Environment* 2 (1), p. 159. DOI: https://doi.org/10.1038/s43247-021-00225-4.

Granger, C. W. J. (1969). "Investigating Causal Relations by Econometric Models and Cross-spectral Methods." In: *Econometrica* 37, pp. 424–459.

Gullapalli, V., J. Franklin, and H. Benbrahim (1994). "Acquiring robot skills via reinforcement learning." In: *Control Systems Magazine, IEEE* 14.1, pp. 13–24.

Ham, Yoo-Geun, Jeong-Hwan Kim, and Jing-Jia Luo (2019). "Deep learning for multi-year ENSO forecasts." In: *Nature* 573 (7775), pp. 568–572. DOI: https://doi.org/10.1038/s41586-019-1559-7.

Harnack, D., E. Laminski, M. Schünemann, and K. R. Pawelzik (2017). "Topological Causality in Dynamical Systems." In: *Physical Review Letters* 119.9, pp. 1–5. DOI: http://dx.doi.org/10.1103/PhysRevLett.119.098301.

Haupt, Sue Ellen, William Chapman, Samantha V. Adams, Charlie Kirkwood, J. Scott Hosking, Niall H. Robinson, Sebastian Lerch, and Aneesh C. Subramanian (2021). "Towards implementing artificial intelligence post-processing in weather and climate: proposed actions from the Oxford 2019 workshop." In: *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 379.2194, p. 20200091. DOI: https://doi.org/10.1098/rsta.2020.0091.

He, F., S. A. Billings, H-. L. Wei, and P. G. Sarrigiannis (2014). "A nonlinear causality measure in the frequency domain: Nonlinear partial directed coherence with applications to EEG." In: *Journal of Neuroscience Methods* 225, pp. 71–80.

He, Jiayi and Pengjian Shang (2017). "Comparison of transfer entropy methods for financial time series." In: *Physica A: Statistical Mechanics and its Applications* 482, pp. 772–785. DOI: https://doi.org/10.1016/j.physa.2017.04.089.

Hendon, H. H. and M. L. Salby (1994). "The life cycle of the Madden–Julian oscillation." In: *Journal of Atmospheric Science* 51, pp. 2225–2237. DOI: https://doi.org/10.1175/1520-0469(1994)051<2225:TLCOTM>2.0.CO;2.

Hiemstra, C. and J. D. Jones (1994). "Testing for linear and nonlinear Granger causality in the stock price-volume relation." In: *The Journal of Finance* 49.5, pp. 1639–1664.

Hirata, Y., J. M. Amigo, Y. Matsuzaka, R. Yokota, H. Mushiake, and K. Aihara (2016). "Detecting Causality by Combined Use of Multiple Methods: Climate and Brain Examples." In: *PLOS ONE* 11, e0158572. DOI: http://dx.doi.org/10.1371/journal.pone.0158572.

IPCC (2022). *Climate change 2022 = "https://report.ipcc.ch/ar6wg3/pdf/IPCC_AR6_WGIII_FinalDraft_FullReport.pdf"*. [Online; accessed 22-April-2022].

Ibarz, Julian, Jie Tan, Chelsea Finn, Mrinal Kalakrishnan, Peter Pastor, and Sergey Levine (2021). "How to train your robot with deep reinforcement learning: lessons we have learned." In: *The International Journal of Robotics Research* 40.4-5, pp. 698–721. DOI: https://doi.org/10.1177/0278364920987859.

Jiang, J. J., Z. G. Huang, L. Huang, H. Liu, and Y. C. Lai (2016). "Directed dynamical influence is more detectable with noise." In: *Scientific Reports* 6, pp. 1–9. DOI: http://dx.doi.org/10.1038/srep24088.

Jiang, Xianan, Ángel F. Adames, Daehyun Kim, Eric D. Maloney, Hai Lin, Hyemi Kim, Chidong Zhang, Charlotte A. DeMott, and Nicholas P. Klingaman (2020). "Fifty Years of Research on the Madden–Julian Oscillation: Recent Progress, Challenges, and Perspectives." In: *Jour-*

nal of Geophysical Research: Atmospheres 125.17, e2019JD030911. DOI: https://doi.org/10.1029/2019JD030911.

Kaplan, Alexey, Mark A Cane, Yochanan Kushnir, Amy C Clement, M Benno Blumenthal, and Balaji Rajagopalan (1998). "Analyses of global sea surface temperature 1856–1991." In: *Journal of Geophysical Research: Oceans* 103.C9, pp. 18567–18589.

Kikuchi, K., B. Wang, and Y. Kajikawa (2012). "Bimodal representation of the tropical intraseasonal oscillation." In: *Climate Dynamics* 38, pp. 1989–2000. DOI: https://doi.org/10.1007/s00382-011-1159-1.

Kiladis, G. N., J. Dias, K. H. Straub, M. C. Wheeler, S. N. Tulich, K. Kikuchi, K. M. Weickmann, and M. J. Ventrice (2014). "A comparison of OLR and circulation-based indices for tracking the MJO." In: *Monthly Weather Review* 142, pp. 1697–1715. DOI: https://doi.org/10.1175/MWR-D-13-00301.1.

Kim, H., Y. G. Ham, Y. S. Joo, and S. W. Son (2021). "Deep learning for bias correction of MJO prediction." In: *Nature Communications* 12.3087. DOI: https://doi.org/10.1038/s41467-021-23406-3.

Kim, H., F. Vitart, and D. E. Waliser (2018). "Prediction of the Madden–Julian Oscillation: A Review." In: *Journal of Climate* 31.23, pp. 9425 –9443. DOI: https://doi.org/10.1175/JCLI-D-18-0210.1.

Kim, Hye-Mi, Daehyun Kim, Frederic Vitart, Violeta E. Toma, Jong-Seong Kug, and Peter J. Webster (2016). "MJO Propagation across the Maritime Continent in the ECMWF Ensemble Prediction System." In: *Journal of Climate* 29.11, pp. 3973–3988. DOI: https://doi.org/10.1175/JCLI-D-15-0862.1.

Kim, Hye-Mi, Peter J. Webster, Violeta E. Toma, and Daehyun Kim (2014). "Predictability and Prediction Skill of the MJO in Two Operational Forecasting Systems." In: *Journal of Climate* 27.14, pp. 5364 –5378. DOI: 10.1175/JCLI-D-13-00480.1.

Klotzbach, Philip J. (2010). "On the Madden–Julian Oscillation–Atlantic Hurricane Relationship." In: *Journal of Climate* 23.2, pp. 282–293. DOI: https://doi.org/10.1175/2009JCLI2978.1.

Klotzbach, Philip J (2011). "El Niño–Southern Oscillation's impact on Atlantic basin hurricanes and US landfalls." In: *Journal of Climate* 24.4, pp. 1252–1263.

Korbel, Jan, Xiongfei Jiang, and Bo Zheng (2019). "Transfer Entropy between Communities in Complex Financial Networks." In: *Entropy* 21.11. DOI: https://doi.org/10.3390/e21111124.

Korenek, J. and J. Hlinka (2020). "Causal network discovery by iterative conditioning: Comparison of algorithms." In: *Chaos* 30, p. 013117. DOI: http://dx.doi.org/10.1063/1.5115267.

Krakovská, A., J. Jakubík, and M. Chvostekova (2018). "Comparison of six methods for the detection of causality in a bivariate time series." In: *PRE* 97, p. 042207.

Kramer, K. and J. Ware (2021). *Counting the cost 2021 – A year of climate breakdown = "https://www.christianaid.org.uk/sites/default/files/2021-12/Counting%20the%20cost%202021%20-%A%20year%20of%20climate%20breakdown.pdf"*. [Online; accessed 22-April-2022].

Kugiumtzis, D. (2013). "Direct-coupling information measure from nonuniform embedding." In: *Physical Review E* 87, p. 062918. DOI: http://dx.doi.org/10.1103/PhysRevE.87.062918.

Kullback, S. (1959). *Information Theory and Statistics*. New York: Wiley.

Kunkel, Kenneth E., David R. Easterling, David A. R. Kristovich, Byron Gleason, Leslie Stoecker, and Rebecca Smith (2012). "Meteorological Causes of the Secular Variations in Observed Extreme Precipitation Events for the Conterminous United States." In: *Journal of Hydrometeorology* 13.3, pp. 1131 –1141. DOI: https://doi.org/10.1175/JHM-D-11-0108.1.

Lall, U. and A. Sharma (1996). "A nearest nighbor bootstrap for resampling hydrologic time seriess." In: *Water Resources Research* 32.3, pp. 679–693. DOI: http://dx.doi.org/10.1029/95WR02966.

Lancaster, G., D. Iatsenko, A. Pidde, V. Ticcinelli, and A. Stefanovska (2018). "Surrogate data for hypothesis testing of physical systems." In: *Physics Reports* 748, pp. 1–60. DOI: http://dx.doi.org/10.1016/j.physrep.2018.06.001.

Lau, W. K. M. and D. E. Waliser (2011). "Predictability and forecasting." In: *Intraseasonal Variability in the Atmosphere-Ocean Climate System*. Berlin, Heidelberg: Springer Berlin Heidelberg. DOI: https://doi.org/10.1007/978-3-642-13914-7_12.

Leng, S., H. Ma, J. Kurths, Y. C. Lai, W. Lin, K. Aihara, and L. Chen (2020). "Partial cross mapping eliminates indirect causal influences." In: *Nature Communications* 11.1, p. 2632. DOI: http://dx.doi.org/10.1038/s41467-020-16238-0.

Levine, Aaron FZ, Michael J McPhaden, and Dargan MW Frierson (2017). "The impact of the AMO on multidecadal ENSO variability." In: *Geophysical Research Letters* 44.8, pp. 3877–3886.

Liebmann, B. and C. A. Smith (1996). "Description of a complete (interpolated) outgoing long-wave radiation dataset." In: *Bulletin of the American Meteorological Society* 77.6.

Lin, Hai, Gilbert Brunet, and Jacques Derome (2008). "Forecast Skill of the Madden–Julian Oscillation in Two Canadian Atmospheric Models." In: *Monthly Weather Review* 136.11, pp. 4130 –4149. DOI: https://doi.org/10.1175/2008MWR2459.1.

Lin, Jia-Lin et al. (2006). "Tropical Intraseasonal Variability in 14 IPCC AR4 Climate Models. Part I: Convective Signals." In: *Journal of Climate* 19.12, pp. 2665 –2690. DOI: https://doi.org/10.1175/JCLI3735.1.

Liu, Z., Y. Jin, and X. A Rong (2019). "A theory for the seasonal predictability barrier: threshold, timing, and intensity." In: *Journal of Climate* 32, pp. 423–443. DOI: https://doi.org/10.1175/JCLI-D-18-0383.1.

Lizier, J. T., J. Heinzle, A. Horstmann, J.-D. Haynes, and M. Prokopenko (2011). "Multivariate information-theoretic measures reveal directed information structure and task relevant changes in fMRI connectivity." In: *Journal of Computational Neuroscience* 30.85.

MJO indices data (2021). https://psl.noaa.gov/mjo/mjoindex/. Accessed: 2021-02.

Ma, H., K. Aihara, and L. Chen (2014). "Detecting causality from nonlinear dynamics with short-term time series." In: *Scientific Reports* 4, pp. 1–10. DOI: http://dx.doi.org/10.1038/srep07464.

Ma, H., S. Leng, C. Tao, X. Ying, J. Kurths, Y. C. Lai, and W. Lin (2017). "Detection of time delays and directional interactions based on time series from complex dynamical systems." In: *Physical Review E* 96.1, pp. 1–8. DOI: http://dx.doi.org/10.1103/PhysRevE.96.012221.

Madden, R. and P. Julian (1994). “Observations of the 40–50-Day Tropical Oscillation–A Review.” In: *Monthly Weather Review* 122, pp. 814–837. DOI: https://doi.org/10.1175/1520-0493(1994)122<0814:OOTDTO>2.0.CO;2.

Madden, Roland A. and Paul R. Julian (1971). “Detection of a 40–50 Day Oscillation in the Zonal Wind in the Tropical Pacific.” In: *Journal of Atmospheric Sciences* 28.5, pp. 702 –708. DOI: https://doi.org/10.1175/1520-0469(1971)028%3C0702:DOADOI%3E2.0.CO;2.

Madden, Roland A. and Paul R. Julian (1972). “Description of Global-Scale Circulation Cells in the Tropics with a 40–50 Day Period.” In: *Journal of Atmospheric Sciences* 29.6, pp. 1109 –1123. DOI: https://doi.org/10.1175/1520-0469(1972)029%3C1109:DOGSCC%3E2.0.CO;2.

Mahadevan, Sridhar and Jonathan Connell (1991). “Scaling Reinforcement Learning to Robotics by Exploiting the Subsumption Architecture.” In: *Machine Learning Proceedings 1991*. Ed. by Lawrence A. Birnbaum and Gregg C. Collins. San Francisco (CA): Morgan Kaufmann, pp. 328–332. ISBN: 978-1-55860-200-7. DOI: https://doi.org/10.1016/B978-1-55860-200-7.50068-4.

Mansfield, L A, P J Nowack, M Kasoar, R G Everitt, W J Collins, and A Voulgarakis (2020). “Predicting global patterns of long-term climate change from short-term simulations using machine learning.” In: *npj Climate and Atmospheric Science* 3 (1), p. 44. DOI: https://doi.org/10.1038/s41612-020-00148-5.

Marinazzo, D., M. Pellicoro, and S. Stramaglia (2008). “Kernel method for nonlinear Granger causality.” In: *Physical Review Letters* 100.14. DOI: http://dx.doi.org/10.1103/PhysRevLett.100.144103.

Marshall, Andrew G and Adam A Scaife (2009). “Impact of the QBO on surface winter climate.” In: *Journal of Geophysical Research: Atmospheres* 114.D18.

Martin, Zane K., Elizabeth A. Barnes, and Eric D. Maloney (2021). “Using simple, explainable neural networks to predict the Madden-Julian oscillation.” In: *Earth and Space Science Open Archive*. DOI: https://doi.org/10.1002/essoar.10507439.1.

Matthews, Robert (2000). “Storks Deliver Babies (p= 0.008).” In: *Teaching Statistics* 22.2, pp. 36–38. DOI: https://doi.org/10.1111/1467-9639.00013.

McGovern, A., R. Lagerquist, D. John Gagne II, G. E. Jergensen, K. L. Elmore, C. R. Homeyer, and T. Smith (2019). “Making the Black Box More Transparent: Understanding the Physical Implications of Machine Learning.” In: *Bulletin of the American Meteorological Society* 100.11, pp. 2175–2199. DOI: https://doi.org/10.1175/BAMS-D-18-0195.1.

McGraw, M. C. and E. A. Barnes (2018). “Memory Matters: A Case for Granger Causality in Climate Variability Studies.” In: *Journal of Climate* 31, pp. 3289–3300. DOI: http://dx.doi.org/10.1175/JCLI-D-17-0334.1.

Molini, Annalisa, Gabriel G. Katul, and Amilcare Porporato (2010). “Causality across rainfall time scales revealed by continuous wavelet transforms.” In: *Journal of Geophysical Research: Atmospheres* 115.D14. DOI: https://doi.org/10.1029/2009JD013016.

Mosedale, T. J., D. B. Stephenson, M. Collins, and T. C. Mills (2006). “Granger causality of coupled climate processes: Ocean feedback on the North Atlantic Oscillation.” In: *Journal of Climate* 19.7, pp. 1182–1194. DOI: http://dx.doi.org/10.1175/JCLI3653.1.

Mosteller, F. and J. W. Tukey (1968). “Data analysis, including statistics.” In: *Handbook of Social Psychology* 2.

Mueller, A., J. F. Kraemer, T. Penzel, H. Bonnemeier, J. Kurths, and N. Wessel (2016). “Causality in physiological signals.” In: *Physiological Measurements* 37.5, R46–R72. DOI: http://dx.doi.org/10.1088/0967-3334/37/5/R46.

Neena, J. M., June Yi Lee, Duane Waliser, Bin Wang, and Xianan Jiang (2014). “Predictability of the Madden—Julian Oscillation in the Intraseasonal Variability Hindcast Experiment (ISVHE).” In: *Journal of Climate* 27.12, pp. 4531 –4543. DOI: https://doi.org/10.1175/JCLI-D-13-00624.1.

Newman, Matthew, Michael A Alexander, Toby R Ault, Kim M Cobb, Clara Deser, Emanuele Di Lorenzo, Nathan J Mantua, Arthur J Miller, Shoshiro Minobe, Hisashi Nakamura, et al. (2016). "The Pacific decadal oscillation, revisited." In: *Journal of Climate* 29.12, pp. 4399–4427.

Nichols, J. M. and K. D. Murphy (2016). *Modeling and Estimation of Structural Damage*. Wiley.

Nowack, P., J. Runge, V. Eyring, and J. D. Haigh (2020). "Causal networks for climate model evaluation and constrained projections." In: *Nature Communications* 11, p. 1415. doi: http://dx.doi.org/10.1038/s41467-020-15195-y.

Okumura, Yuko, Shang-Ping Xie, Atusi Numaguti, and Youichi Tanimoto (2001). "Tropical Atlantic air-sea interaction and its influence on the NAO." In: *Geophysical Research Letters* 28.8, pp. 1507–1510.

OpenAI et al. (2019). *Dota 2 with Large Scale Deep Reinforcement Learning*. doi: https://doi.org/10.48550/ARXIV.1912.06680. url: https://arxiv.org/abs/1912.06680.

Palus, M., V. Komárek, Z. Hrncír, and K. Sterbová (2001). "Synchronization as adjustment of information rates: detection from bivariate time series." In: *Physical Review E* 63, p. 046211. doi: http://dx.doi.org/10.1103/PhysRevE.63.046211.

Paluš, Milan and Martin Vejmelka (2007). "Directionality of coupling from bivariate time series: How to avoid false causalities and missed connections." In: *Physical Review E* 75, p. 056211. doi: http://dx.doi.org/10.1103/PhysRevE.75.056211.

Paluš, M. (2014a). "Multiscale Atmospheric Dynamics: Cross-Frequency Phase-Amplitude Coupling in the Air Temperature." In: *Physical Review Letters* 112. doi: http://dx.doi.org/10.1103/PhysRevLett.112.078702.

Paluš, Milan (2014b). "Cross-Scale Interactions and Information Transfer." In: *Entropy* 16.10, pp. 5263–5289. doi: http://dx.doi.org/10.3390/e16105263.

Patz, J., D. Campbell-Lendrum, T. Holloway, and J. Foley (2005). "Impact of regional climate change on human health." In: *Nature* 438, pp. 310–317. doi: https://doi.org/10.1038/nature04188.

Pedregosa, F. et al. (2011). "Scikit-learn: Machine Learning in Python." In: *Journal of Machine Learning Research* 12, pp. 2825–2830.

Pereda, E., R. Quian Quiroga, and J. Bhattacharya (2005). "Nonlinear multivariate analysis of neurophysiological signals." In: *Progress in Neurobiology* 77, p. 1. DOI: http://dx.doi.org/10.1016/j.pneurobio.2005.10.003.

Pérez-Alarcón, Albenis, José C Fernández-Alvarez, Rogert Sorí, Raquel Nieto, and Luis Gimeno (2021). "The combined effects of SST and the North Atlantic subtropical high-pressure system on the Atlantic basin tropical cyclone interannual variability." In: *Atmosphere* 12.3, p. 329.

Perez, Antonio, Riccardo Silini, Ivan Sanchez, and Joaquin Bedia (2022). "Ecoregion based attribution analysis of the influence of several fire danger indices on the amount of burned area at a global scale by means of pseudo transfer entropy." In: *Advances in Forest Fire Research*.

Pompe, B. and J. Runge (2011). "Momentary information transfer as a coupling measure of time series." In: *Physical Review E* 83.5, p. 051122. DOI: http://dx.doi.org/10.1103/PhysRevE.83.051122.

Porfiri, M. and et al. (2019). "Media coverage and firearm acquisition in the aftermath of a mass shooting." In: *Nature Human Behaviour* 3, p. 913. DOI: http://dx.doi.org/10.1038/s41562-019-0636-0.

Porta, A. and L. Faes (2016). "Wiener-Granger causality in the network physiology with applications to cardiovascular control and neuroscience." In: *Proceedings of the IEEE* 104.2, pp. 282–309. DOI: http://dx.doi.org/10.1109/JPROC.2015.2476824.

Pothapakula, Praveen Kumar, Cristina Primo, and Bodo Ahrens (2019). "Quantification of Information Exchange in Idealized and Climate System Applications." In: *Entropy* 21.11. DOI: https://doi.org/10.3390/e21111094.

Péguin-Feissolle, A. and T. Teräsvirta (2001). "Causality tests in a nonlinear framework." In: *Working paper, Stockholm School of Economics, Stockholm*.

Quian Quiroga, R., A. Kraskov, T. Kreuz, and P. Grassberger (2002). "Performance of different synchronization measures in real data: A case study on electroencephalographic signals." In: *Physical Review E* 65 (4), p. 041903. DOI: 10.1103/PhysRevE.65.041903.

RMM data (2021). https://iridl.ldeo.columbia.edu/SOURCES/.BoM/.MJO/.RMM/index.html?Set-Language=en. Accessed: 2021-02.

Raconteur (2019). *A Day in Data = "http://res.cloudinary.com/yumyoshojin/image/upload/v1/pdf/future-data-2019.pdf"*. [Online; accessed 22-April-2022].

Rashid, H. A., H. H. Hendon, M. C. Wheeler, and O. Alves (2011). "Prediction of the Madden–Julian oscillation with the POAMA dynamical prediction system." In: *Climate Dynamics* 36, pp. 649–661. DOI: https://doi.org/10.1007/s00382-010-0754-x.

Rasp, S. and S. Lerch (2018). "Neural networks for postprocessing ensemble weather forecasts." In: *Monthly Weather Review* 146.11, pp. 3885–3900. DOI: https://doi.org/10.1175/MWR-D-18-0187.1.

Reynolds, Richard W and Thomas M Smith (1994). "Improved global sea surface temperature analyses using optimum interpolation." In: *Journal of climate* 7.6, pp. 929–948.

Rodríguez-Fonseca, Belén, Irene Polo, Javier García-Serrano, Teresa Losada, Elsa Mohino, Carlos Roberto Mechoso, and Fred Kucharski (2009). "Are Atlantic Niños enhancing Pacific ENSO events in recent decades?" In: *Geophysical Research Letters* 36.20.

Rui, H. and B. Wang (1990). "Development characteristics and dynamic structure of tropical intraseasonal convection anomalies." In: *Journal of Atmospheric Science* 47, pp. 357–379. DOI: https://doi.org/10.1175/1520-0469(1990)047<0357:DCADSO>2.0.CO;2.

Runge, J., P. Nowack, M. Kretschmer, S. Flaxman, and D. Sejdinovic (2019). "Detecting and quantifying causal associations in large nonlinear time series datasets." In: *Science Advances* 5, eaau4996. DOI: http://dx.doi.org/10.1126/sciadv.aau4996.

Runge, J. and et al. (2019). "Inferring causation from time series in Earth system sciences." In: *Nature Communications* 10, p. 2553. DOI: http://dx.doi.org/10.1038/s41467-019-10105-3.

Rössler, O. E. (1976). "An equation for continuous chaos." In: *Physics Letters* 57A.5, pp. 397–398. DOI: http://dx.doi.org/10.1016/0375-9601(76)90101-8.

Salahuddin, M. and J. Gow (2016). "The effects of Internet usage, financial development and trade openness on economic growth in South

Africa: A time series analysis." In: *Telematics and Informatics* 33.4, pp. 1141–1154. DOI: http://dx.doi.org/10.1016/j.tele.2015.11.006.

Sandoval, L. J. (2014). "Structure of a Global Network of Financial Companies Based on Transfer Entropy." In: *Entropy* 16, p. 4443. DOI: http://dx.doi.org/10.3390/e16084443.

Scheuerer, M., M. B. Switanek, R. P. Worsnop, and T. M. Hamill (2020). "Using Artificial Neural Networks for Generating Probabilistic Subseasonal Precipitation Forecasts over California." In: *Monthly Weather Review* 148.8, pp. 3489–3506. DOI: https://doi.org/10.1175/MWR-D-20-0096.1.

Schreiber, T. (2000). "Measuring Information Transfer." In: *Physical Review Letters* 85.2, pp. 461–464.

Schreiber, Thomas and Andreas Schmitz (1996). "Improved surrogate data for nonlinearity tests." In: *Physical review letters* 77.4, p. 635.

Schreiber, Thomas and Andreas Schmitz (2000). "Surrogate time series." In: *Physica D: Nonlinear Phenomena* 142.3-4, pp. 346–382.

Schwarz, G. E. (1978). "Estimating the dimension of a model." In: *Annals of Statistics* 6.2, pp. 461–464.

Seo, K.-H. (2009). "Statistical-dynamical prediction of the Madden-–Julian oscillation using NCEP Climate Forecast System (CFS)." In: *International Journal of Climatology* 29, pp. 2146 –2155. DOI: https://doi.org/10.1002/joc.1845.

Seth, A. K., A. B. Barrett, and L. Barnett (2015). "Granger causality analysis in neuroscience and neuroimaging." In: *Journal of Neuroscience* 35.8, pp. 3293–3297. DOI: http://dx.doi.org/10.1523/JNEUROSCI.4399-14.2015..

Shannon, C. E. and W. Weaver (1949). *The Mathematical theory of Information*. Urbana IL: University of Illinois Press.

Silini, R. (2020). *GitHub: https://github.com/riccardosilini/pTE*. DOI: http://dx.doi.org/10.5281/zenodo.4271219.

Silini, R. (2021b). *Wheeler-Hendon phase diagrams videos: https://doi.org/10.5281/zenodo.4733942*. DOI: https://doi.org/10.5281/zenodo.4733942.

Silini, R. (2021a). *Wheeler-Hendon phase diagrams*. DOI: https://doi.org/10.5281/zenodo.5801415. URL: \url{https://doi.org/10.5281/zenodo.5801415}.

Silini, R., M. Barreiro, and C. Masoller (2021). "Machine learning prediction of the Madden-Julian Oscillation." In: *npj Climate and Atmospheric Science* 4.57. DOI: https://doi.org/10.1038/s41612-021-00214-6.

Silini, R., S. Lerch, N. Mastrantonas, H. Kantz, M. Barreiro, and C. Masoller (2022a). "Improving the prediction of the Madden-Julian Oscillation of the ECMWF model by post-processing." In: *Unpublished*.

Silini, R. and C. Masoller (2021). "Fast and effective pseudo transfer entropy for bivariate data-driven causal inference." In: *Scientific Reports* 11, p. 8423. DOI: https://doi.org/10.1038/s41598-021-87818-3.

Silini, R., G. Tirabassi, M. Barreiro, L. Ferranti, and C. Masoller (2022b). "Assessing causal dependencies in climatic indices." In: *Unpublished*.

Silver, David et al. (2016). "Mastering the game of Go with deep neural networks and tree search." In: *Nature* 529 (7587), pp. 484–489. DOI: https://doi.org/10.1038/nature16961.

Spiegelhalter, D. J., N. G. Best, B. P. Carlin, and A. van der Linde (2002). "Bayesian measures of model complexity and fit." In: *Journal of the Royal Statistical Society B* 64.4, pp. 583–639.

Staniek, M. and K. Lehnertz (2008). "Symbolic transfer entropy." In: *Physical Review Letters* 100, p. 158101. DOI: http://dx.doi.org/10.1103/PhysRevLett.100.158101.

Sugihara, G., R. May, H. Ye, C. H. Hsieh, E. Deyle, M. Fogarty, and S. Munch (2012). "Detecting causality in complex ecosystems." In: *Science* 338.6106, pp. 496–500. DOI: http://dx.doi.org/10.1126/science.1227079.

Sun, J., D. Taylor, and E. M. Bollt (2015). "Causal Network Inference by Optimal Causation Entropy." In: *SIAM Journal on Applied Dynamical Systems* 14, pp. 73–106. DOI: http://dx.doi.org/10.1137/140956166.

Taamouti, A., T. Bouezmarni, and A. El Ghouch (2014). "Nonparametric estimation and inference for conditional density based Granger causality measures." In: *Journal of Econometrics* 180, pp. 251–264. DOI: http://dx.doi.org/10.1016/j.jeconom.2014.03.001.

Taraphdar, S., F. Zhang, L. R. Leung, X. Chen, and O. M. Pauluis (2018). "MJO affects the Monsoon Onset Timing Over the Indian Region." In: *Geophysical Research Letters* 45.18. DOI: https://doi.org/10.1029/2018GL078804.

Theiler, J., S. Eubannk, A. Longtin, B. Galdrikian, and J. Doyne Farmer (1992). "Testing for nonlinearity in time series: the method of surrogate data." In: *Physica D* 58.1-4, pp. 77–94. DOI: http://dx.doi.org/10.1016/0167-2789(92)90102-S.

Tirabassi, G., C. Masoller, and M. Barreiro (2014). "A study of the air–sea interaction in the South Atlantic Convergence Zone through Granger causality." In: *International Journal of Climatology* 35.12, pp. 3440–3453. DOI: http://dx.doi.org/10.1002/joc.4218.

Tirabassi, Giulio, Linda Sommerlade, and Cristina Masoller (2017). "Inferring directed climatic interactions with renormalized partial directed coherence and directed partial correlation." In: *Chaos* 27, p. 035815. DOI: http://dx.doi.org/10.1063/1.4978548.

Tjostheim, T. (1981). "Granger-causality in multiple time series." In: *Journal of Econometrics* 17, pp. 157–176.

Trenberth, Kevin E, Julie M Caron, David P Stepaniak, and Steve Worley (2002). "Evolution of El Niño–Southern Oscillation and global atmospheric surface temperatures." In: *Journal of Geophysical Research: Atmospheres* 107.D8, AAC–5.

Tropical Meteorology, Indian Institute of (2020). *All-India Rainfall data from = "https://www.tropmet.res.in/"*.

Tseng, Kai-Chih, Elizabeth A. Barnes, and Eric Maloney (2020). "The Importance of Past MJO Activity in Determining the Future State of the Midlatitude Circulation." In: *Journal of Climate* 33.6, pp. 2131 –2147. DOI: https://doi.org/10.1175/JCLI-D-19-0512.1.

Ungerovich, Matilde, Marcelo Barreiro, and Cristina Masoller (2021). "Influence of Madden–Julian Oscillation on extreme rainfall events in Spring in southern Uruguay." In: *International Journal of Climatology*, pp. 1–13. DOI: https://doi.org/10.1002/joc.7022.

Ursino, Mauro, Giulia Ricci, and Elisa Magosso (2020). "Transfer Entropy as a Measure of Brain Connectivity: A Critical Analysis With the Help of Neural Mass Models." In: *Frontiers in Computational Neuroscience* 14. DOI: 10.3389/fncom.2020.00045.

Vannitsem, S. et al. (2021). "Statistical Postprocessing for Weather Forecasts: Review, Challenges, and Avenues in a Big Data World." In: *Bulletin of the American Meteorological Society* 102.3, E681–E699. DOI: https://doi.org/10.1175/BAMS-D-19-0308.1.

Vannitsem, Stephane and Pierre Ekelmans (2018). "Causal dependences between the coupled ocean–atmosphere dynamics over the tropical Pacific, the North Pacific and the North Atlantic." In: *Earth System Dynamics* 9, 1063–1083. DOI: http://dx.doi.org/10.5194/esd-9-1063-2018.

Vecchi, G. A. and N. A. Bond (2004). "The Madden-Julian Oscillation (MJO) and northern high latitude wintertime surface air temperatures." In: *Geophysical Research Letters* 31.L04104. DOI: https://doi.org/10.1029/2003GL018645.

Ventrice, Michael J., Matthew C. Wheeler, Harry H. Hendon, Carl J. Schreck, Chris D. Thorncroft, and George N. Kiladis (2013). "A Modified Multivariate Madden–Julian Oscillation Index Using Velocity Potential." In: *Monthly Weather Review* 141.12, pp. 4197 –4210. DOI: https://doi.org/10.1175/MWR-D-12-00327.1.

Vicente, R., M. Wibral, M. Lindner, and G. Pipa (2011). "Transfer entropy–a model-free measure of effective connectivity for the neurosciences." In: *Journal of Computational Neuroscience* 30.45.

Vigen, Tyler (2015a). *Spurious correlations*. New York: Hachette Books.

Vigen, tyler (2015b). *Spurious Correlations "http://tylervigen.com/spurious-correlations"*.

Vilasuso, J. (2001). "Causality tests and conditional heteroskedasticity: Monte Carlo evidence." In: *Journal of Econometrics* 101, pp. 25–35.

Visbeck, Martin H, James W Hurrell, Lorenzo Polvani, and Heidi M Cullen (2001). "The North Atlantic Oscillation: past, present, and future." In: *Proceedings of the National Academy of Sciences* 98.23, pp. 12876–12877.

Vitart, F. (2009). “Impact of the Madden Julian Oscillation on tropical storms and risk of landfall in the ECMWF forecast system.” In: *Geophysical Research Letters* 36. DOI: https://doi.org/10.1029/2009GL039089.

Vitart, Frédéric (2017). “Madden—Julian Oscillation prediction and teleconnections in the S2S database.” In: *Quarterly Journal of the Royal Meteorological Society* 143.706, pp. 2210–2220. DOI: https://doi.org/10.1002/qj.3079.

Wang, Hui, Arun Kumar, Wanqiu Wang, and Yan Xue (2012). “Influence of ENSO on Pacific decadal variability: An analysis based on the NCEP Climate Forecast System.” In: *Journal of climate* 25.18, pp. 6136–6151.

Wheeler, Matthew C. and Harry H. Hendon (2004). “An All-Season Real-Time Multivariate MJO Index: Development of an Index for Monitoring and Prediction.” In: *Monthly Weather Review* 132.8, pp. 1917 –1932. DOI: https://doi.org/10.1175/1520-0493(2004)132<1917:AARMMI>2.0.CO;2.

Wheeler, Matthew C., Harry H. Hendon, Sam Cleland, Holger Meinke, and Alexis Donald (2009). “Impacts of the Madden-Julian Oscillation on Australian Rainfall and Circulation.” In: *Journal of Climate* 22.6, pp. 1482–1498. DOI: https://doi.org/10.1175/2008JCLI2595.1.

Wheeler, Matthew and Klaus M. Weickmann (2001). “Real-Time Monitoring and Prediction of Modes of Coherent Synoptic to Intraseasonal Tropical Variability.” In: *Monthly Weather Review* 129.11, pp. 2677 – 2694. DOI: https://doi.org/10.1175/1520-0493(2001)129<2677:RTMAPO>2.0.CO;2.

Wibral, M., N. Pampu, V. Priesemann, F. Siebenhhner, H. Seiwert, M. Lindner, J. T. Lizier, and R. Vicente (2013). “Measuring information-transfer delays.” In: *PloS One* 8.2, e55809. DOI: http://dx.doi.org/10.1371/journal.pone.0055809.

Wiener, N. (1956). “Nonlinear Prediction and Dynamics.” In: *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability* 3, pp. 247–252.

Wu, Cheng-Han and Huang-Hsiung Hsu (2009). “Topographic Influence on the MJO in hte Maritime Continent.” In: *Journal of Climate* 22.20, pp. 5433–5448. DOI: https://doi.org/10.1175/2009JCLI2825.1.

Wu, Jie, Hong-Li Ren, Jinqing Zuo, Chongbo Zhao, Lijuan Chen, and Qiaoping Li (2016). “MJO prediction skill, predictability, and teleconnection impacts in the Beijing Climate Center Atmospheric General Circulation Model.” In: *Dynamics of Atmospheres and Oceans* 75, pp. 78–90. DOI: https://doi.org/10.1016/j.dynatmoce.2016.06.001.

Yao, Can-Zhong and Hong-Yu Li (2020). “Effective Transfer Entropy Approach to Information Flow Among EPU, Investor Sentiment and Stock Market.” In: *Frontiers in Physics* 8. DOI: https://doi.org/10.3389/fphy.2020.00206.

Ye, Hao, Ethan R Deyle, Luis J Gilarranz, and George Sugihara (2015). “Distinguishing time-delayed causal interactions using convergent cross mapping.” In: *Scientific Reports* 5 (1), p. 14750. DOI: https://doi.org/10.1038/srep14750.

Yoo, Changhyun, Steven Feldstein, and Sukyoung Lee (2011). “The impact of the Madden-Julian Oscillation trend on the Arctic amplification of surface air temperature during the 1979–2008 boreal winter.” In: *Geophysical Research Letters* 38 (24). DOI: https://doi.org/10.1029/2011GL049881.

Zhang, Chidong, Jon Gottschalck, Eric D. Maloney, Mitchell W. Moncrieff, Frederic Vitart, Duane E. Waliser, Bin Wang, and Matthew C. Wheeler (2013). “Cracking the MJO nut.” In: *Geophysical Research Letters* 40.6, pp. 1223–1230. DOI: https://doi.org/10.1002/grl.50244.

Zhao, J., Y. Zhou, X. Zhang, and L. Chen (2016). “Part mutual information for quantifying direct associations in networks.” In: *Proceedings of the National Academy of Sciences of the United States of America* 113.18, pp. 5130–5135.

RESEARCH ACTIVITIES

PUBLICATIONS

Covered in this thesis

Fast and effective pseudo transfer entropy for bivariate data-driven causal inference.

👥 Silini R & Masoller C.

📅 2021 📖 Scientific Reports 11(1):8423

Machine learning prediction of the Madden-Julian oscillation.

👥 Silini R, Barreiro M & Masoller C.

📅 2021 📖 npj Climate and Atmospheric Sciences 4, 57

Improving the Madden-Julian oscillation prediction of weather models by post-processing.

👥 Silini R, Lerch S, Mastrantonas N, Kantz H, Barreiro M & Masoller C.

📅 2022 📖 Earth System Dynamics, under revision

Causality analysis of atmospheric and ocean climate indices.

👥 Silini R, Tirabassi G, Barreiro M, Ferranti L & Masoller C.

📅 2022 📖 Climate Dynamics, under revision

Not in this thesis

Ensemble forecast of the Madden Julian Oscillation using analogs of the geopotential at 500 hPa and stochastic weather generator.

👥 Krouma M, Silini R, Yiou P.

📅 2022 📕 Earth System Dynamics, under revision

Ecoregion based attribution analysis of the influence of several fire danger indices on the amount of burned area at a global scale by means of pseudo transfer entropy.

👥 Perez A, Silini R, Sanchez I, Bedia J.

📅 2022 📕 Advances in Forest Fire Research 2022, Coimbra Univ. Press

Outlier mining in high-dimensional datasets based on Jensen-Shannon distances and graph structure analysis.

👥 Toledo A, Silini R, Carpi L, Masoller C.

📅 2022 📕 Complex Networks 2022, 56, under revision

SECONDMENTS

📍 Universidad de la Republica, Montevideo, Uruguay

📅 July-August 2020

👤 Prof. Marcelo Barreiro

Online secondment focused on the prediction of the Madden-Julian Oscillation, which led to the collaboration with Prof. Barreiro for all subsequent publications.

📍 Max Planck Institute for the Physics of Complex Systems, Dresden, Germany

📅 June-July 2021

👤 Prof. Holger Kantz

Secondment focused on the post-processing of Madden-Julian Oscillation prediction, which led to the collaboration with Prof. Kantz for a publication.

📍 Predictia Intelligent Data Solutions, Santander, Spain

📅 January-February 2022

👤 Daniel San Martin Segura

Secondment in the private sector. The work conducted at Predictia, led to a publication in the 9th International Conference on Forest Fire Research, and to the accomplishment of a client request, which involved a data-driven identification of climate extremes.

CONFERENCES TALKS

European Geosciences Union, EGU21

💬 *On the predictability of the Madden-Julian Oscillation phase*

👥 **Silini R**, Barreiro M & Masoller C.

📅 19-30 Apr 2021

Congreso de Física Estadística Joven, FisEs’21

💬 *Fast and effective measure for bivariate data-driven causal inference*

👥 **Silini R** & Masoller C.

📅 5-6 May 2021

ECMWF Machine Learning Workshop

💬 *Improving the Madden-Julian Oscillation prediction of weather models by post-processing*

👥 **Silini R**, Lerch S, Mastrantonas N, Kantz H, Barreiro M & Masoller C.

📅 28-30 Mar 2022

IX Jornada de Complexitat

🗋 *Machine learning algorithms for the prediction of the Madden-Julian Oscillation*

👥 **Silini R**, Lerch S, Mastrantonas N, Kantz H, Barreiro M & Masoller C.

📅 15 Jun 2022

CAFE WORKSHOPS

1st CAFE Workshop

📍 Sitges, Spain

📅 November 2019

2nd CAFE Workshop

📍 Freiberg, Germany

📅 March 2020

💬 Talk: *Assessment of seasonal and sub-seasonal atmospheric interaction and extreme events*

🗋 Poster: *A new approach for inferring causality and its directionality from time series analysis*

3rd CAFE Workshop

📍 Toulouse, France

📅 November 2021

💬 Talk: *Machine learning and causal data analysis tools for the assessment of atmospheric interaction and extreme events in seasonal and sub-seasonal time series*

🗋 Poster: *Machine learning prediction of the Madden-Julian Oscillation*

4th CAFE Workshop

📍 Reading, UK

📅 March 2022

💬 Talk: *Improving the Madden-Julian Oscillation prediction of weather models by post-processing*

Outreach activities

CAFE ESRs presentation, *Get to know: Riccardo Silini (Universitat Politècnica de Catalunya, Barcelona, Spain)*
https://www.youtube.com/watch?v=AzxGpdKOOMo
31 May 2021

Nit de la recerca, *Predicció de fenòmens meteorològics utilitzant la intel·ligència artificial*
https://www.youtube.com/watch?v=plGJWSOPwHk&t
24 Sep 2021

Predictia Blog post, *Hablamos con: Riccardo Silini, doctorando de CAFE*
https://predictia.es/es/news/CAFE-Riccardo-Silini-Madden-Julian-causalidad
11 Feb 2022

Institut Nou Barris Barcelona, *A Machine learning introduction*
6 Apr 2022

Now this is not the end. It is not even the beginning of the end.
But it is, perhaps, the end of the beginning.

– Winston Churchill