

An information-theoretical approach to identify seismic precursors and earthquake-causing variables

Arthur Lopes da Silva Valencio

BSc Physics, Universidade Estadual de Campinas (Unicamp), Brazil, 2015

Supervisors:

Dr Murilo da Silva Baptista, Prof Celso Grebogi

A thesis presented for the degree of

Doctor of Philosophy

of the

University of Aberdeen.



School of Natural and Computing Sciences – Department of Physics
and Institute for Complex Systems and Mathematical Biology

24th May 2018

Contents

Declaration	v
Publications	vi
Abstract	vii
Acknowledgements	viii
Foreword	xi
Glossary	xii
1 Introduction	1
1.1 Motivations	1
1.2 Objective	4
1.3 Structure of this thesis	4
1.4 Summary of results	6
1.4.1 Development of a tool for analysis of causality from time-series	6
1.4.2 Application to test-bench systems	6
1.4.3 Application to precursors and earthquake-causing variables	7
2 Overview of observational seismology	9
2.1 What is a seismic event?	10
2.1.1 Types of seismic source: faulting	10
2.1.2 Stress, strain and seismic waves	11

2.1.3	Trace, magnitude and intensity	15
2.1.4	Earthquake occurrence: location and magnitude distribution	19
2.1.5	Foreshocks, aftershocks and swarms	22
2.1.6	Gravity and displacement changes	24
2.1.7	Exotic phenomena	25
2.2	Earthquake prediction	27
2.2.1	The field of earthquake prediction	27
2.2.2	IASPEI list of precursors	27
2.2.3	The case for b-value anomalies	31
2.2.4	The case for tidal triggering of earthquakes	31
2.2.5	The case for pre-seismic gravity anomalies	32
2.3	Instrumentation in seismology	32
2.3.1	Seismometer	32
2.3.2	Gravimeter	35
2.3.3	Other devices	37
	2.3.3.1 Displacement sensors	38
	2.3.3.2 Strainmeters	39
	2.3.3.3 Tiltmeters	40
3	Time-series considered	41
3.1	Geophysical time-series	41
	3.1.1 Seismic data	41
	3.1.2 Gravity data	45
3.2	Data pre-processing	47
	3.2.1 Tidal analysis and removal	47
	3.2.2 Symbolic encoding and partitioning	55
3.3	Test-bench dynamical system: Logistic map	57
	3.3.1 Coupled logistic maps	59
	3.3.2 Logistic networks	65

4	Information theoretical quantities and causality	67
4.1	Entropy and Mutual Information	67
4.2	Transfer Entropy	70
4.3	Causal Mutual Information (CaMI)	71
4.4	Local and delayed Causal Mutual Information	74
4.5	Rate of information measures	75
4.6	Pointwise information measures	76
4.7	Algorithm of the method and usage	78
4.8	Computational demands	83
5	Assessing causality to test-bench systems	87
5.1	Pseudo-random noise	88
5.2	Application to coupled logistic maps	91
5.2.1	Error levels: uncoupled system	91
5.2.2	Diffusive coupling	94
5.2.3	CML coupling	101
5.2.4	Including dynamical noise	106
5.2.5	Connecting a chaotic and an intermittent system	107
5.3	Application to logistic networks	111
5.3.1	Serial network	111
5.3.2	Parallel network	115
5.3.3	Directed tree network with maximum distance 3	120
5.3.3.1	A variation: periodic channel	122
5.4	Discussion of results	124
6	Causal analysis of earthquake precursor candidates	126
6.1	Tidal triggering of earthquakes	129
6.2	Gutenberg-Richter's b-value temporal variations	136
6.3	Pre-seismic gravity variations	142
6.4	Preceding seismicity leading to new events	147

6.5 Summary	150
7 Conclusions	151
7.1 Review of context, objective and findings	151
7.2 Open questions	152
7.3 Applications to other fields of knowledge	153
Bibliography	155
Appendix A Comment on definitions of disasters and media representation	166
Appendix B Steps in the development of a Cold Atom Gravimeter	173
Appendix C Pointwise information measures applied to precursor candidates	176

Declaration

This work was produced by me. No portion of the work contained in this document has been submitted in support of an application for a degree or qualification of this or any other university or other institution of learning. All verbatim extracts have been distinguished by quotation marks, and all sources of information have been specifically acknowledged. Some of the work in this thesis has been published in peer-reviewed journals or peer-reviewed book chapters, or is being prepared for submission. The details of these publications are given on the next page.

Signed:

Arthur Lopes da Silva Valencio

Date: 24th May 2018

Publications

On earthquakes and geophysical time-series analysis

Published

- A. Valencio, C. Grebogi and M.S. Baptista. Methods for removal of unwanted signals from gravity time-series: comparison using linear techniques complemented with analysis of system dynamics. *Chaos*, 27(10):103126, 2017. doi: 10.1063/1.4996452.
- A. Valencio, N. Valencio and F.Y. Hanai. Keeping the feet on the ground: reflecting, monitoring and taking action in front of seismic events (*Mantendo os pés no chão: refletir, monitorar e agir diante eventos sísmicos*). *Revista Eletrônica de Tecnologia e Cultura*, 20:21–40, 2017. Original document in Portuguese.

In preparation

- A. Valencio, C. Grebogi and M.S. Baptista. The topology of causality. (Provisional title).
- A. Valencio, C. Grebogi and M.S. Baptista. Investigation of earthquake precursor candidates with causality. (Provisional title).

Additional publications (in the theme of disasters)

- N. Valencio and A. Valencio. Media coverage of the UK flooding crisis: a social panorama. *Disasters*, (early view), *in press*. doi: 10.1111/disa.12255.
- N. Valencio and A. Valencio. Cities plagued by dangerous waters (*Cidades atormentadas pelas águas perigosas*). *In preparation*. Original document in Portuguese.
- N. Valencio and A. Valencio. News coverage of disasters in Brazil: sociopolitical dimensions put at the margins of public debate (*Cobertura jornalística sobre desastres no Brasil: dimensões sociopolíticas marginalizadas no debate público*). *Disertaciones*, 12(2):165–185, 2017. doi: 10.12804/revistas.urosario.edu.co/disertaciones/a.4791. Original document in Portuguese.
- N. Valencio and A. Valencio. Vulnerability as social oppression: the traps of risk-prevention actions. In V. Marchezini, B. Wisner, S. Saito, and L. Londe, editors, *Reduction of Vulnerability to Disasters: from Knowledge to Action*, pages 111–137. RiMa, São Carlos, 2017.

Abstract

Several seismic precursors and earthquake-causing variables have been proposed in the last decades based on physical considerations and case observations, however none has been confirmed on long datasets using linear analysis. This work adopts an information-theoretical approach to investigate the occurrence of causal flow between these precursors and causing variables and seismicity. It starts by introducing the key concepts in seismology and presenting the current main precursor candidates. Four variables will be considered as possible precursors or anomalies leading to earthquakes: large tidal amplitudes, temporal fluctuations in the Gutenberg-Richter's b-value, surface gravity changes, and preceding anomalous seismicity patterns. To perform the causality test between these variables and their effects, it is developed a method which allows the fast calculation of Transfer Entropy for any two time-series, detecting the direction of the flow of information between the variables of interest. The method is tested to coupled logistic maps and networks with different topologies before application to geophysical events. The analysis shows mutual information relating to coupling strength and also allows inference of the causal direction from data using the Transfer Entropy, both in bivariate systems and in networks. The method was then applied to the earthquake analysis for an interval of 4018 days on an area comprising the Japan trench. Within a conservative margin of confidence, the results could not at this point confirm any of the four precursor options considered, but future work can clarify initial suggestions regarding tidal amplitudes link to seismicity, and pre-seismic gravity changes and cumulative daily magnitude anomalies. The Matlab/Octave codes for our method are open-source and available at <https://github.com/artvalencio/causality-toolbox>. We hope the method is able to support the quest for other precursor candidates, and to assist other fields of knowledge.

Acknowledgements

First I must thank immensely to Murilo and Celso. Murilo, a restless mind, found quite a demanding student to match, so I must thank him infinitely for the incredible patience – and for holding a big smile at all times! Looking back, it was very fun indeed all the moments of long discussions, checking results (and some bug-fixing), and all the opportunities I've received for presenting the research, applying for funding, publishing, and teaching, thanks to him. Thanks also for all the great companionship along the way. Celso, more than a supervisor, was a mentor, someone who I could not only discuss about the state-of-art (and beyond) of the topics of this research, but also to talk about simpler things of life, seek advice, ask for help when needed, share good moments (of the research or good news in general), enjoy a simple good cup of coffee/tea, and be inspired by his life experience. Thanks to both Murilo and Celso for welcoming me into their families, making me feel like home. Not enough words can express my gratitude.

I extend my thanks to the other members of Institute for Complex Systems and Mathematical Biology (ICSMB, University of Aberdeen), who were great friends and colleagues. In particular, thanks to the lecturers, for their openness, good moments and lunch break invitations. I also acknowledge and thanks the use of the ICSMB and the University of Aberdeen Maxwell High-Performance Computing Clusters, together with the technical support from Naveed Khan and Michael Chung. Special thanks to Marco Thiel and Alessandro Moura for the most valuable inputs in the previous oral examinations and openness for discussions. Thanks as well to the final examination board for the thorough reading and valuable questions, corrections and remarks.

A massive thanks to the many friends I had the privilege to have in Aberdeen: Daniel Rodrigues, Maria Angélica Araújo, Srikanth Phanendra, Anna Wiedemann, Meng Yu,

Helari Ramler, Teodora Oniga, Amanda Catharina, Julia Gezatt, Paulo Abelha, Juliana Milheiro, Rodrigo de Oliveira, Daniela Madureira, Juan Afanador, Lucas Fernandes, Fábio Tamaki, Gabriela Mol, Nicolás Rubido, Helen Shiells, Marco Faggian, Akura Johnson, Daniel Vogel, Claudia Schnabel, Pau Clusella, Ricardo Contreras, Suzannah Morrison, Gloria Cecchini, Mallory Karlu, Alessandro Gillies, Olivier Veraguth, Grace Ihongo, James Reid, Pierre Bonin, Thisiane Sant and Luke Colburn. Also to the friends in Brazil who kept in contact: Taiguara Sugawara, William Depetri, Artur Pinto, Shamaila Manzoor, Thiago Rosales, Hugo Freitas, Olivia Miura, Fabrício Machado, Rodrigo Dias de Oliveira, Ítalo Aventurato and Otávio Gomide.

I must also thank all the teachers who helped me along the way and continue to be a source of inspiration. I especially dedicate this to the teachers of my primary school, Oca dos Curumins, who supported my earliest dreams of becoming a physicist and stimulated me with science and mathematics projects and participation at competitions. Your driving force means the world to me. Special thanks to my science teacher, Tarso Pinto. Also, thanks to the lecturers who taught me at higher education in undergraduate and post-graduate levels in Unicamp, University of Aberdeen and other institutions of the Scottish Universities Physics Alliance. Finally, thanks to the students at the University of Aberdeen who I had the privilege of giving tutorials to. I am looking forward to see their great discoveries.

This research extends preceding theoretical work in Dynamical Systems and Information Theory and applies novel techniques to Seismology. Hence, thanks to the contributors from these areas. Special thanks to the operators of the seismic and gravity networks which I have collected data from (IRIS, USGS and IGETS). Without their continuous effort of maintenance and on allowing public research use of data used in this work, this research would not have been possible.

Thanks to the Brazilian people, who financed this and other research on the Science without Borders Programme. Thanks to the Brazilian National Council for Science and Technology Development – CNPq, for the PhD scholarship [206246/2014-5].

Above all thanks to my family, especially my mother, who supported me at all times.

To mom

Foreword

On 5 September 2016 the European Seismological Commission gathered in Trieste, Italy, for the opening of its biennial General Assembly. Coincidentally, a major earthquake occurred only two weeks before, on the evening of the 24 August, with epicentre in the Central Apennines. The town of Amatrice was severely hit, with over 240 deaths, including members of a seismology team. The seismic event echoed the one occurred seven years before in the neighbouring town of L'Aquila, which made over 65,000 people homeless and was considered a national catastrophe. In the case of L'Aquila event, the limitations in forecasting and alerting the population about the earthquake led to controversial judicial disputes. The possibility of history repeating itself triggered a rush, of rescue actions from the part of Civil Protection and related personnel, reconstruction and retrofitting from the part of civil engineers, and scenario analysis and testing of earthquake prediction hypothesis from the part of seismologists. The 35th General Assembly of European Seismological Commission became a key stage to these discussions, with special sessions dedicated exclusively to the analysis of the Amatrice event. A significant fraction of the presentations focused on the physical description and case observations of earthquake precursors, especially for L'Aquila and Amatrice. However, there was no consensual agreement on any of the options displayed (which is part of the healthy scientific discussion). At the occasion, I was only presenting a preliminary, but surprisingly tricky, stage of the data analysis (tide removal of a geophysical time-series). Yet, it was clear that there was a big gap of knowledge on the identification of precursors, for which information-theory could help solving. This led to this work, an initial effort in fulfilling this gap. I hope it can support the identification of a seismic precursor in future.

Glossary

Seismology and Geodesy

- **F**: Force vector
- **T**: Traction vector
- σ : Stress tensor
- ε : Strain tensor
- δ_{ij} : Kronecker delta
- λ : First Lamé parameter
- μ : Shear Young modulus (second Lamé parameter)
- U : Potential energy
- M_w : Moment magnitude
- M_L : Local (Richter) magnitude
- M_s : Surface wave magnitude
- N : Number of earthquakes
- b -value: Gutenberg-Richter seismicity-magnitude rate
- t_s : Time-series sampling rate
- v_p : P-wave velocity
- v_s : S-wave velocity
- D : Epicentral distance
- δt : time-difference of arrival of P and S waves
- A : Peak amplitude of P-wave, as measured from a Wood-Anderson seismograph
- $n(t)$: Frequency of aftershocks
- Λ : Seismicity rate function
- $\log L$: Log-likelihood of set of seismic events given an historic
- δg : Relative gravity, gravity change
- θ_1 : Colatitude
- θ_2 : Longitude
- m_1, m_2 : Geolocation of polar cap

Time-series Analysis, Dynamical Systems and Information Theory

- t_s : Sampling time
- t_d : Delay time defining a window size for local analysis
- n_b : Number of initial bins (or alphabet length)
- τ : Reconstruction delay for application of Takens' method
- M : System embedding dimension
- S : Symbol attributed on the first partition
- ϕ : Symbol in the higher-order partitioned space
- p : Probability of an event
- L : Length of the symbolic sequence
- H : Entropy
- I : Mutual information
- $CaMI$: Causal mutual information
- TE : Transfer entropy
- MIR : Mutual information rate
- $CaMIR$: Causal mutual information rate
- TER : Transfer entropy rate
- PMI : Pointwise mutual information
- $PCaMI$: Pointwise causal mutual information
- PTE : Pointwise transfer entropy
- $DirIdx$: Directionality index (net flow of causal information)
- d : Time-delay for analysis
- N_T : Total number of points
- σ : Coupling strength
- λ : Amplitude of dynamical noise
- r : Logistic map free parameter
- A_{ij} : Adjacency matrix terms
- k_i : Degree of node i

Chapter 1

Introduction

1.1 Motivations

The Earth is dynamic. Its parts are constantly moving, settling, building up or releasing stress, ejecting mass, reshaping convective fluxes, and so on [1–3]. The changes on the planet surface can occur in the time-scale of millennia, such as the creation of mountains or the drifts of continents, but also on time-scales as short as days or minutes. The latter is the case of earthquakes, which are sudden ground shakes produced from energy release on the Earth’s crust[3–7], lasting for about a few minutes. Small earthquakes happen on a daily basis, but they are generally unnoticed or pass by producing no harmful effects [3]. However, when a large earthquake happens, it can have dire consequences: collapse of buildings, trigger landslides or avalanches, liquefying the ground making structures sink or float, trigger volcanic eruptions, and producing tsunamis. Although earthquake shaking do not typically constitute a direct cause for death or injuries, all these consequences pose a significant threat to vulnerable communities. Indeed, earthquakes are ranked by the United Nations Office for Disaster Risk Reduction and the Centre for Research on the Epidemiology of Disasters (UNISDR and CRED) as the natural hazard linked to most deaths, accounting for 55.6% of the total fatalities catalogued in the period 1996-2015 on the database EM-DAT [8]. Hence, it is important to distinguish the earthquake as an event of nature, which can be understood as a *hazard* for its potential to cause disruption, and the vulnerability of the citizens in the space they occupy. When we refer to the occurrence of a *disaster*, the underlying condition for it is the social vulnerability to a hazard. It

means that a seismic event in isolation, however strong it might be, cannot be considered a disaster¹. An example is given by comparing events of the similar magnitude and place observing the disparity in number of affected people (Table 1.1). In the cases exemplified, better building practices and civil protection policies were implemented as consequence of major events in Mexico, Japan and Italy, which helped reducing the impacts when similar events happened to strike again in the same regions. However, there is still significant room for improvement, as the death toll and the number of homeless remain high.

Table 1.1: Comparison of death toll and number of homeless for similar seismic events (same magnitude and region) in Italy (2009 and 2016), Mexico (1985 and 2017) and Japan (1995 and 2016)

Seismic event	Magnitude (M_w)	Death toll	Number of homeless
2009 L' Aquila (Central Italy)	6.9	309	65000+
2016 Amatrice (Central Italy)	5.7	299	4500
1985 Mexico City	8.0	up to 45000	250000
2017 Chiapas (Mexico)	7.1	370	40000
1995 Kobe (Southern Japan)	6.9	6400	250000–300000
2016 Kumamoto (Southern Japan)	7.0	50	7000

There is no single recipe to reduce the social vulnerability to the seismic hazard. In terms of prevention and preparedness strategies, the two main approaches are:

1. to adapt the built environment to resist to the ground movements and their consequences, which is the object of research and practice of *Earthquake Engineering*;
2. to monitor and analyse the seismic phenomena, including understanding the past events, definition of geologically vulnerable areas, and assessment of the likelihood of future events, which are the primary objects of research and practice of *Seismology*.

The ultimate goal of seismology historically has been to be able to predict earthquakes. However, this has not yet been achieved, and it is an on-going discussion whether it is even achievable [9]. Instead, the primary preparedness practice in the field focuses

¹We expand the discussion of these concepts on Appendix A

in detecting events near the source and broadcasting a warning to vulnerable populations with a time advance of seconds to minutes before the arrival of the tremor. This procedure is often referred as EEW – Earthquake Early-Warning (Fig. 1.1). Unfortunately, such strategy restricts the possible civil preparedness actions possible in heavily populated areas, which become limited to ducking under a table or stepping away from windows so to reduce the likelihood of physical injuries in case of partial building collapse. These measures might have a positive effect, but cannot prevent fatalities from total building collapse. If more time is allowed, in the scale of hours to days before a large event, then other options such as preventive evacuation and sheltering are available, with the potential to save significantly more lives.

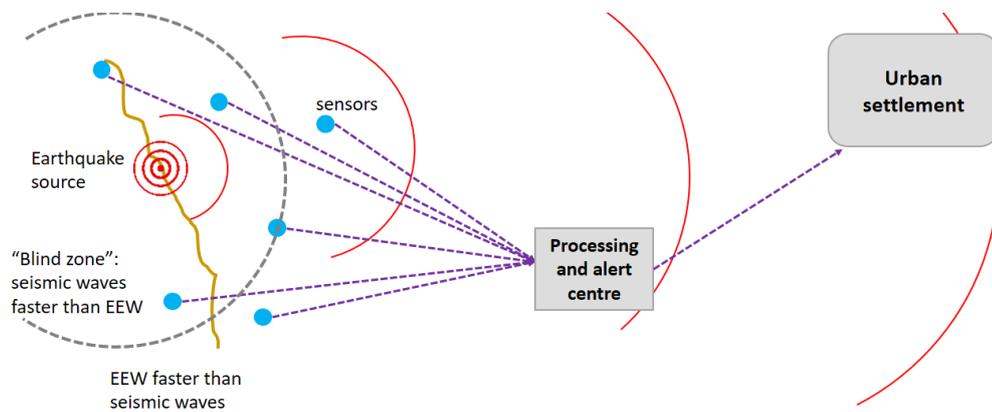


Figure 1.1: Simplified diagram of Earthquake Early-Warning (EEW) system. Adapted from our publication on [10].

Intending to identify earthquakes with more time in advance, two currents in seismology have developed: *earthquake prediction* and *earthquake forecasting*. The first refers to the identification of events associated to earthquake occurrence, the *precursors*, and consequently their monitoring. The second is about probabilistic assessments of future earthquakes considering past events. In principle, prediction and forecast can act together, as the monitoring of a precursor could be incorporated into a probabilistic model for future events. However, no reliable seismic precursor has yet been found, mostly due to disagreements on whether the observed relation with a seismic event is indeed causal or just a spurious correlation. This has not, though, curbed the efforts to find a precursor.

In this thesis we will mention a few phenomena being investigated in the field, such

as foreshocks, strain and tilt changes, groundwater level variations, and crustal displacement. Our particular focus will be in four cases: tidal triggering of earthquakes, fluctuations in Gutenberg-Richter's seismicity-magnitude rate (b -value), pre-seismic gravity anomalies, and anomalous seismicity patterns. The purpose is not to unequivocally validate or discard any of these precursor candidates, but to use them as a testing ground for a method able to identify a precursor in future.

1.2 Objective

In this work we will present a method based on information theory to analyse the causal relation between two variables from their time-series. Approaches of causal analysis from time-series gained notoriety with Granger's studies on econometrics [11, 12], but the original method was only applicable to linearisable models. The historical difficulties encountered in identifying a precursor are indications that linear models are insufficient for solving seismology problems, particularly earthquake prediction.

This work has the objective of offering an alternative or complementary method to analyse causal relations between variables from the time-series even when the system exhibits strong nonlinear behaviour. In particular, we are interested in performing this analysis in a computationally efficient way, compatible with the necessities for earthquake prediction.

The method that will be presented in the following chapters, Causal Mutual Information (CaMI), is tested for a number of systems: fully stochastic, logistic map systems, and real-world seismic and gravity data.

1.3 Structure of this thesis

Chapter 2 begins with a review in seismology, aiming to understand the specificities of the type of physical phenomena and data to be analysed. The goals of this chapter are: to describe what is an earthquake, how their occurrences are distributed, the key concepts in observational seismology, the practices in *earthquake forecasting* and *earthquake prediction*, and to provide a brief description of the instrumentation and systems that enabled

the data collection. Aspects of this chapter were presented by the author on [10].

Chapter 3 presents the time-series considered, seismic and gravitational, followed by the necessary pre-processing techniques for filtering and analysis. It proceeds by describing the dynamical systems used as testbench: the logistic map, and networks built by coupling logistic maps. An aspect of pre-processing procedures in geophysical time-series analysis (tidal removal) was presented by the author on [13].

In chapter 4 it is defined the quantities from Information Theory used in this work. This include the Entropy and the Mutual Information over symbolic sequences, measurements of the causal flow, such as the Transfer Entropy and the Causal Mutual Information (CaMI), and the rate of the causal flow. The algorithm of the implementation and the computational demands are also presented. The codes developed for this chapter are available publicly as an open-source toolbox in <https://github.com/artvalencio/causality-toolbox>.

Chapter 5 shows the application of the method for the analysis of causality for the testbench systems. The method is first tested against a null-hypothesis from two uniformly distributed random time-series, so to give a parameter of expected error levels. Then, it is applied to a coupled logistic maps and networks. The latter exploits the case where many events are simultaneously associated. For coupled maps, we apply pointwise information measures to see how each region of the phase-space of the system contributes differently to the obtained overall information measure. A manuscript containing results from this chapter is currently in preparation.

In chapter 6 the method is applied on a 11 year period of the Japanese trench area. The hypothesis investigated are: tidal triggering of earthquakes, temporal Gutenberg-Richter's b -value anomalies observed before earthquakes, pre-seismic gravity anomalies observed before earthquakes, and anomalous seismicity patterns before large events. The effects, *i.e.* the earthquake occurrence, was considered in three ways: occurrence of large seismic event on a day (maximum daily magnitude exceeding a threshold), large cumulative daily seismic magnitude (exceeding a threshold), and high daily seismicity rate (number of earthquakes exceeding a threshold). The threshold values defining anomalous events are obtained from data. Time-delayed effects are also considered. A manuscript

containing results from this chapter is currently in preparation.

Finally, chapter 7 presents the conclusions from this research, including new questions in dynamical systems and seismology arisen from the development and application of our approach, as well as the description of other disciplinary fields that could benefit from the method.

Three appendices are included. Appendix A briefly discuss the definitions of disasters, risk and hazards and how the media portrays these events. It relates to our work on refs. [14–17]. Appendix B briefly describes the main stages of development of a geophysical instrumentation (cold atom gravimeter) that has the potential of contributing to the field. Appendix C complements chapter 6, by showing the phase-space plots of the data points of seismic precursors and seismic occurrence colour-coded by the pointwise information quantities. These are tools for investigating the contribution of each point to the overall mutual information or flow of information.

1.4 Summary of results

1.4.1 Development of a tool for analysis of causality from time-series

It is implemented a toolbox for estimation of Causal Mutual Information, Transfer Entropy and net flow of information between two variables from their time-series. In addition, the toolbox calculates the Mutual Information and the rates of the mentioned information-theoretical quantities. It is considered that obtaining the Transfer Entropy from the Causal Mutual Information is able to reduce the processing time, compared to a standalone implementation for the Transfer Entropy. The processing times and memory requirements of the current version for Matlab/Octave are presented, and the package is available for use at <https://github.com/artvalencio/causality-toolbox>.

1.4.2 Application to test-bench systems

The test-bench system consists of coupled logistic maps and networks. The coupling types considered are linear diffusive or coupled map lattice (normalising the coupling intensities to contain network dynamics within the domain of validity of the maps). Often

different behaviours were observed in the high and the low coupling regimes, such as for the decrease of the information measures with the increase of the amplitude of an applied dynamic noise, or for the observation of the node with highest transfer entropy in a network. The mutual information revealed a reliable indicator of the coupling strength between two variables, whereas the directionality index, *i.e.* the net flow of causal information, is found to be an effective indicator of the true direction of causality.

1.4.3 Application to precursors and earthquake-causing variables

The method is applied to investigate four candidates of precursors and earthquake-causing variables for a period of 11 years around the Japanese trench area, one of the most active seismic zones in the planet. Information-theoretical values obtained are computed over a range of possible delays between the precursor/cause and the seismicity/effect, up to a maximum delay of 30 days. The threshold defining what is an anomalous behaviour in seismicity, *b*-value and gravity residuals is allowed to vary, and we select the case leading to the largest causal mutual information.

In all cases it is not possible to conclusively state a causal influence of a candidate to the occurrence of seismicity, as the values were close to the a conservative margin of confidence of the method, drawn from comparison of the information quantities with those obtained when applied to a uncoupled uniform random distribution with the same amount of data points. However, in three cases there are initial indications of possible precursory behaviour from the curve of the information-theoretical values with the time-delay applied between the variables. The first is regarding tidal amplitudes as seismicity triggers, as there seems to be a correspondence in the mutual information between the two time-series with a delay of 5–13 days. The second is regarding pre-seismic gravity changes, as anomalies in the gravity residuals have higher mutual information with seismicity occurrence in a time lag of 8–15 days. The last is regarding to a preceding high cumulative daily magnitude, which has a larger mutual information with large earthquake occurrence (measured by maximum daily magnitude) on a delay of 11-15 days. The obtained values are still comparable to our conservative margins of confidence, but future reanalysis, with more data (reducing the error bars), might be able to confirm or discard such options.

The method is openly available for the analysis of other seismic precursors and applications to other geographic areas and periods, according to availability of the datasets. Additionally, it has been included in the computational tool the alternative of calculating the information-theoretical quantities over a sliding window, enabling the real-time monitoring of changing causal behaviour in a location, perhaps associated with changing geomorphological conditions. Furthermore, the pointwise information measures presented in this work also define the regions of the phase-space most contributing for the overall information quantity. It means that it enables to inform which interval of a precursor/cause is most connected to which region of the effect, by means of sharing most information or by transferring most causal information. Unfortunately, due to insufficient data points, the resolution of the pointwise information measures to the geophysical time-series is still low, hence we leave the preliminary results to Appendix C. These approaches are also applicable to other fields aiming to infer causal relations from observed time-series, such as Economics and Neurology.

Chapter 2

Overview of observational seismology

This chapter presents a landscape overview of the main features of natural seismic phenomena and the current proposals for precursors or hypothesis of earthquake causes. Theories for the occurrence of earthquake events have been proposed since the ancient Greek and Chinese civilizations [18, 19]. Nonetheless, the field of Seismology is considered to have started only in the mid-19th century with the seismometers and first systematic experiments with artificially generated earthquakes (using explosives as sources), which enabled comparison with naturally occurring phenomena [20]. The first networks of seismic devices permitted a more detailed understanding of the Earth interior: the global layered structure, the local underground geomorphology, the global patterns of earthquake occurrence, the existence and dynamics of tectonic plates, and so on [3–6]. However, earthquakes also threaten vulnerable communities in the path of seismic waves, once, depending on the intensity, they cause damage or destruction to the built environment [3, 6]. Consequently, one of the key goals of Seismology is to discover when and where seismic events will occur, so communities can be better prepared. Two currents have developed, *earthquake forecasting*, consisting of statistical models for predicting future events, and *earthquake prediction*, consisting of the re-analysis of past events, looking for phenomena that indicated earthquake occurrence in advance. For example, modern earthquake forecasting indicates, from present state of the Coulomb failure function, seismicity historic, or other variables, what is the likelihood of an event tomorrow. Earthquake prediction, on

the other hand, deals with analysis of earthquakes that have already occurred and investigate if for those events (or set of events) it can be found a preceding anomaly or pattern in another geophysical variable, from gravity changes to Radon variations and Gutenberg-Richter's b-value. This work focus on earthquake prediction, as the primary objective is to present an approach to identify the causes or precursors for earthquakes, and not, at this stage, to extrapolate to probabilities of future events. The chapter ends describing the devices associated with the data used. For more details on observational seismology and earthquake prediction, we recommend refs. [3–7, 21–24].

2.1 What is a seismic event?

A seismic event is a sudden movement in the interior and surface of the Earth. Typically, it occurs when a geological fault under stress releases energy in form of slip movements (displacements) and waves that propagate in the solid Earth, which is the main focus of Seismology, and the main threat to communities. Seismic events that differ from this mechanism, such as artificially generated earthquakes (no geological fault necessarily), intraplate events (small fault or construction of a new fault), and 'silent' earthquakes (no clear seismic waves), tend to pose a smaller risk to people.

2.1.1 Types of seismic source: faulting

Rock faulting is classified in strike-slip, dip-slip or a mix of them (Fig. 2.1), and the way the faulting occurs define how the energy is released. In a strike-slip, the movement is horizontal. Examples are the San Andreas fault in California and the Enriquillo-Plaintain Garden fault which crosses Haiti capital, Port-au-Prince. These faults can produce large earthquakes, and are the types associated with largest stress drop following a seismic event [25]. In a dip-slip fault the movement is vertical. The dip-slip category is further divided into two types: normal and reverse faulting. Normal is when the hanging-wall block subsides respective to the foot-wall block. Globally, normal faulting produces a lower proportion of large earthquakes compared to small earthquakes, relative to the other mechanisms [26]. In reverse faulting, however, the hanging-wall moves further upwards. In particular, if the angle of a dip-slip reverse faulting is small ($< 45^\circ$) it is referred as

a thrust fault, and it is the configuration typically associated with the largest earthquakes and production of large tsunamis, such as 2004 Sumatra-Andaman and 2011 Tohoku. Most seismic events are a mix of vertical and horizontal components, referred as oblique faulting. Figure 2.2 shows, for a strike-slip event, the pattern observed for the first seismic motion, whether it is compression or dilatation of the medium, depending on the station location relative to the epicentre. From these patterns of first motions it is possible to infer the faulting mechanism of an earthquake from a set of seismographs.

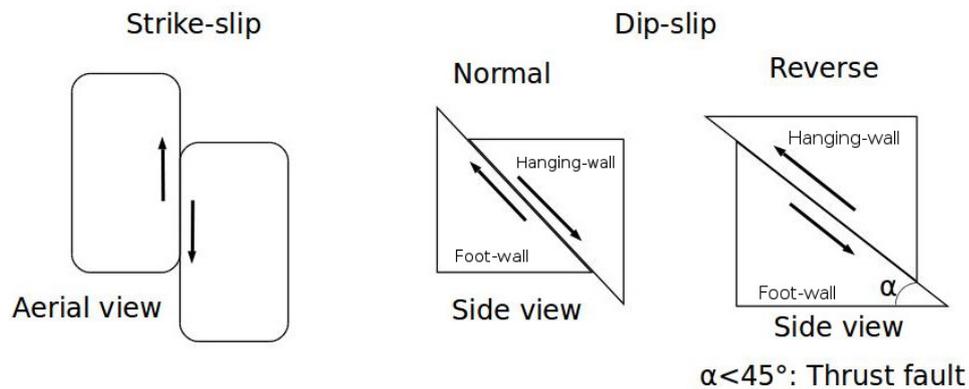


Figure 2.1: Different types of faulting associated with earthquakes.

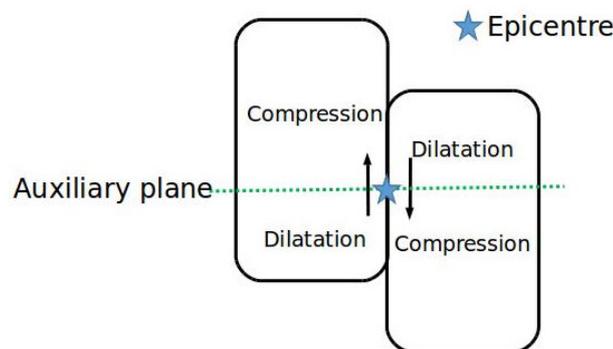


Figure 2.2: Expected behaviour of the first motion near source for a strike-slip event. The auxiliary plane is the plane perpendicular to the motion, containing the epicentre

2.1.2 Stress, strain and seismic waves

Consider a solid under the effect of a force \mathbf{F} representing a portion of the solid Earth. We can analyse how this force leads to the reshaping of the solid. In the infinitesimally small cubes of face area dS constituting the object, this force is translated into tractions ($\mathbf{T} = \lim_{dS \rightarrow 0} \mathbf{F}/dS$) for each cubic face, which then can be further decomposed in a

preferred system of coordinates, creating a stress tensor σ (Fig. 2.3). Considering that the material is not perfectly rigid, the infinitesimal elements will deform slightly, giving origin to a strain tensor ε (Fig. 2.4 and Eq. 2.1). Perfect elasticity is modelled by Hooke's law for continuous media, $\sigma_{ij} = c_{ijkl}\varepsilon_{kl}$, relates the stress (σ) and the strain (ε) tensors. The mechanical properties of the medium are summarised by the the elastic moduli tensor c , which is the equivalent of a spring constant. Hence, at this point we have only modelled a small rock element as a (3-D) spring. Strain is adimensional and stress and elastic moduli have dimension equivalent to pressure

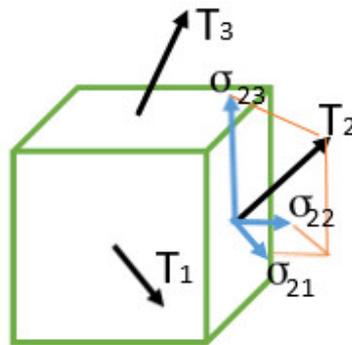


Figure 2.3: Traction vectors T_i and decomposition into stress tensor σ for an infinitesimally small cubic element

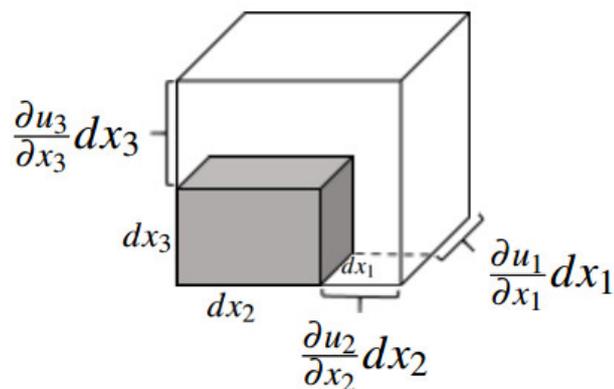


Figure 2.4: Deformation of a solid cubic element (grey), acquiring a new shape (white) after the effect of the displacement field \mathbf{u} . Strain is understood as the ratio of the new extension relative to the original, so each component of a strain tensor is given in Eq.

2.1.

$$\boldsymbol{\varepsilon} = \begin{bmatrix} \frac{\partial u_1}{\partial x_1} & \frac{1}{2} \left(\frac{\partial u_1}{\partial x_2} + \frac{\partial u_2}{\partial x_1} \right) & \frac{1}{2} \left(\frac{\partial u_1}{\partial x_3} + \frac{\partial u_3}{\partial x_1} \right) \\ \frac{1}{2} \left(\frac{\partial u_2}{\partial x_1} + \frac{\partial u_1}{\partial x_2} \right) & \frac{\partial u_2}{\partial x_2} & \frac{1}{2} \left(\frac{\partial u_2}{\partial x_3} + \frac{\partial u_3}{\partial x_2} \right) \\ \frac{1}{2} \left(\frac{\partial u_3}{\partial x_1} + \frac{\partial u_1}{\partial x_3} \right) & \frac{1}{2} \left(\frac{\partial u_3}{\partial x_2} + \frac{\partial u_2}{\partial x_3} \right) & \frac{\partial u_3}{\partial x_3} \end{bmatrix}. \quad (2.1)$$

If this medium has no preferential direction (isotropy), a common mathematical trick is to decompose c in two constants, the shear Young modulus μ and a complementary Lamé constant λ , in the form $c_{ijkl} = \lambda \delta_{ij} \delta_{kl} + \mu (\delta_{ik} \delta_{jl} + \delta_{il} \delta_{jk})$. In this way, the stress-strain relation is considerably simplified to $\boldsymbol{\sigma} = 2\mu \boldsymbol{\varepsilon} + \lambda \text{tr}(\boldsymbol{\varepsilon}) \mathbf{I}$, where \mathbf{I} is the identity matrix. Common rock formations have μ parameter varying in the range 10 – 140 GPa and λ in the range 10 – 160 GPa, with small (less than 5 GPa) dependence on temperature and external pressure (if above 250MPa) [27]. The potential energy accumulated in a rock under stress is $U = 1/2 \int \boldsymbol{\sigma}_{ij} \boldsymbol{\varepsilon}_{ij} dV = 1/2 \int c_{ijkl} \boldsymbol{\varepsilon}_{ij} \boldsymbol{\varepsilon}_{kl} dV$, and a seismic event consists in releasing part of this stored energy, mainly in form of seismic waves.

The total body force on the cubic element is $F_i = \partial_j \boldsymbol{\sigma}_{ij} dV$, but the mass of the cube is $m = \rho dV$, so Newton's second law ($\mathbf{F} = m\mathbf{a}$) reads $\partial_j \boldsymbol{\sigma}_{ij} + f_{ext} = \rho \ddot{u}_i$. We can assume that the external normalised force element f_{ext} is zero for the cases of interest (it only becomes relevant in specific very low frequency modes, where gravity plays a role, and when very close to epicentre, where a source force f_s is present). Replacing $\boldsymbol{\sigma}$ by Hooke's law, after a few algebraic manipulations we reach $(\lambda + 2\mu) \nabla(\nabla \cdot \mathbf{u}) - \mu \nabla \times (\nabla \times \mathbf{u}) = \rho \ddot{\mathbf{u}}$ (see [3–5] for more details). Although this equation is complicated, it simplifies by expressing the displacement vector \mathbf{u} in terms of scalar and vector potentials ϕ and $\boldsymbol{\psi}$ (in analogy to electromagnetism): $\mathbf{u} = \nabla \phi + \nabla \times \boldsymbol{\psi}$. Substituting and using the identity $\nabla^2 \mathbf{v} = \nabla(\nabla \cdot \mathbf{v}) - \nabla \times (\nabla \times \mathbf{v})$ it immediately obtains the seismic wave equation:

$$\nabla \left((\lambda + 2\mu) \nabla^2 \phi(\mathbf{x}, t) - \rho \phi(\ddot{\mathbf{x}}, t) \right) = -\nabla \times \left(\mu \nabla^2 \boldsymbol{\psi}(\mathbf{x}, t) - \rho \boldsymbol{\psi}(\ddot{\mathbf{x}}, t) \right). \quad (2.2)$$

A simple solution is to equal both sides to zero, leading to two wave equations: $\nabla^2 \phi = (1/v_p^2) \ddot{\phi}$ and $\nabla^2 \boldsymbol{\psi} = (1/v_s^2) \ddot{\boldsymbol{\psi}}$, where $v_p = ((\lambda + 2\mu)/\rho)^{1/2}$ and $v_s = (\mu/\rho)^{1/2}$

are the respective velocities in the medium. As expected, the plane wave solution is of the form $\phi(\mathbf{x}, t) = A \exp(i(\omega t \pm \mathbf{k} \cdot \mathbf{x}))$ for the first case, and analogous, but with A being a vector, for the second. For a wave propagating in z -direction, the first case leads to a displacement contribution of $\mathbf{u}(z, t) = \nabla \phi(z, t) = (0, 0, -ikA \exp(i(\omega t \pm kz)))$. This is a compressive wave, like sound. Comparatively, the second case leads to a contribution $\mathbf{u}(z, t) = \nabla \times \boldsymbol{\psi}(z, t) = (ikA_y, -ikA_x, 0) \exp(i(\omega t \pm kz))$, which is a transverse wave solution, like the upwards and downwards motion of a string. The first solution is referred as a P-wave (primary, due to higher velocity, the longitudinal/compressive wave solution), and the second a S-wave (for secondary or shear, the transverse wave solution). The P-wave and each component of displacement of the S-wave have kinetic energy of $K = 1/2 \int \rho (\dot{u}_i)^2 dV = A_i^2 \omega^2 k^2 \rho / 4$ and potential energy due to strain in the wave-front of the same amount, values averaged over the wavelength. These constitute the primary non-frictional components of energy release in a seismic event. The P and S waves are depicted in Fig 2.5.

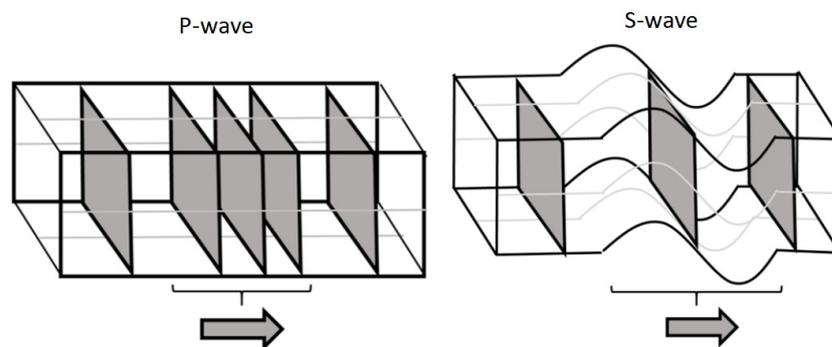


Figure 2.5: Simplified diagram of P (left) and S (right) wavefronts. Adapted from our publication on [10].

It should be noted that the expressions, in particular the seismic velocities v_p and v_s , depend on the density of the medium ρ . It implies that the seismic waves path are subject to refraction (and reflection) wherever there is a boundary, similar to Snell's problem in ray optics. Furthermore, after meeting the boundary, the refracted/reflected wave may have components of the other mode. As the Earth is not homogeneous, the seismic waves sensed by a seismometer are not only the pure direct path P and S waves, but also all the reflections and refractions reaching the device location.

2.1.3 Trace, magnitude and intensity

When the seismic waves arrive at a location, the shaking is recorded by a device that might capture the surface displacement, velocity or acceleration against time, producing a time-series. This time-series is historically referred as seismic trace in Seismology, due to the way it was recorded on paper in old devices. An example is presented on Fig. 2.6 for the 2011 Tohoku-oki thrust event, as observed on the IU-MAJO seismometer (located in the underground Matsushiro Seismological Observatory, near Nagano, Japan), vertical broadband component.

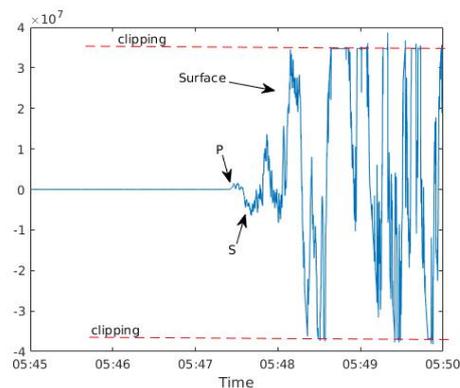


Figure 2.6: Seismic trace of the 2011 Tohoku-oki earthquake from the vertical broadband component (BHZ) of the IU-MAJO seismometer, Japan. The arrival of the P, S and surface waves are indicated. This seismic event was so strong that the motion was above the instrumental limit and the signal was clipped.

In this case, the first peak above the background is the direct P-wave, which we expected from the higher speed. The amplitude increases in this case when we observe the arrival of the S-wave. Around this stage there is already a mix with the refractions and reflections of P and S waves on the Earth's layers. In addition, when the P and/or S waves meet the ground surface, they might satisfy a condition for the generation of a third type, surface waves, which are even more damaging. As the name suggests, they travel along the surface of the Earth, which partially explains why they produce higher ground shaking – the energy is not spread to the planetary interior. Two types of surface waves occur: Rayleigh waves, which is a non-trivial solution originating when a P and a S wave meet at the Earth surface, and Love waves, which is another solution when two S waves constructively interfere in a location where the bottom layer has lower density than the

top, acting like a mirror. A thorough detailing of their mathematical description can be seen in refs. [3, 4, 7].

By measuring the time-delay between the arrival of the P and the S wave, it can be determined the epicentre, by simple triangulation of the result from each sensor in a network. The frequency distribution of the P and S waves peaks around the range 0.1-2 Hz, which is why most seismometers are tailored for such frequency. From the identification of the modes from reflection and the refraction as they arrive at different stations, it is also possible to infer the layered structure of inner Earth. Following global observations, Fig. 2.7 shows the travel-time speeds of the P and S seismic waves depending on the distance to the centre of the Earth, and the inferred variation of the density profile, meaning radial layering. The IASP91 Earth model [28] is based on extrapolation from 57655 travel-time data of 104 events, whereas the Preliminary Earth Reference Model (PREM) [29] is result from the analysis of 26000 events (2000000 detected P-waves and 250000 S-waves). The principal layered structure of the Earth (crust, mantle, outer core, inner core) is defined by the three large discontinuities, respectively at radius 6371 km (25 km depth), 3480 km (2891 km depth) and 1222 km (5149 km depth).

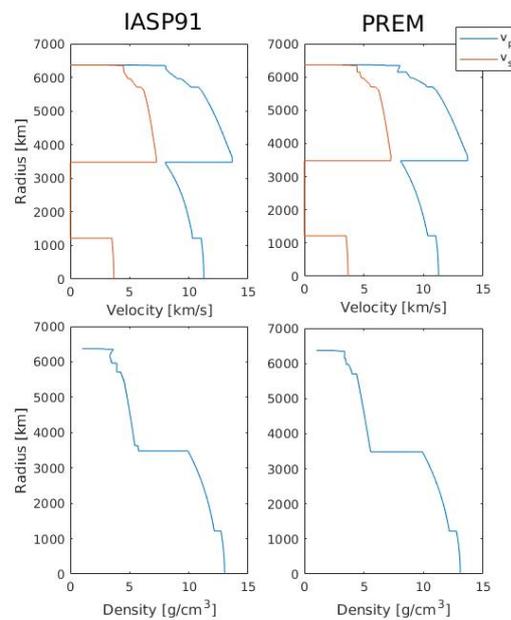


Figure 2.7: Two models of seismic waves velocity profiles, and the inferred density layering of the Earth. IASP91 model definition/data from [28, 30] and PREM model definition/data from [29].

It is convenient to have a simple quantity to assess the energy release of an earthquake. The faster procedure is to assess a magnitude using the Richter scale (M_L) [31]. The origin of this scale is empiric, due to an observed logarithm relation of the amplitude A of the first P-wave as detected on a specific type of seismograph and the distance D of the device to the epicentre (Eq. 2.3). The amplitude A is given in mm of an Wood-Anderson seismograph, which will be further presented in Sec. 2.3.1 (Fig. 2.16), and D is given in km

$$M_L = \log_{10}(A) + 2.76 \log_{10}(D) - 2.48. \quad (2.3)$$

Unfortunately the Wood-Anderson torsion seismometers are no longer a standard (modern devices sometimes implement sophisticated scale conversions), and also the result might vary slightly depending on the location. Other magnitude-scale quantities inspired by empirical observations are also commonly used: body-wave magnitude (m_b) and surface-wave magnitude M_s (Eqs. 2.4 and 2.5)

$$m_b = \log_{10}(A/T) + Q(h,D), \quad (2.4)$$

$$M_s = \log_{10}(A) + 1.66 \log_{10}(D) - 2. \quad (2.5)$$

The body-wave is also calculated over the first P-wave, for amplitude A in microns, period $T < 3s$, and only considering stations at less than 100° from epicentre. $Q(h,D)$ is an empirical function designed to produce a consistent output across all stations in a network, depending also on the earthquake depth h . This standard was adopted by the World Wide Standardized Seismograph Network to monitor if countries were following the 1963 Nuclear Test Ban Treaty. The surface-wave magnitude M_s is measured by taking the amplitude A as the amplitude in microns of the Rayleigh waves with 20s period specifically. However, these three formulations are purely empirical and bear little relation with the physical mechanism producing the earthquake. Furthermore, the three values diverge for strong earthquakes ($M > 6$).

The Moment Magnitude scale (M_w), given by Eq. 2.6, overcomes these issues, as it

is based on the seismic moment M_0 of the earthquake. The seismic moment is defined on terms of the slip parameters of the fault (shear modulus μ at fault depth, mean dislocation of the fault $\langle \Delta U \rangle$ and surface area S), but these relate to the stress drop on the fault and consequently to the radiated energy E_0 in form of seismic waves (see refs. [3, 4] for further details). The seismic moment also connects with the stress drop $\Delta\sigma$ according to Eq. 2.8, where c is a proportionality constant depending on the shape of the fault. Hence, the Moment Magnitude scale is a more robust measure, leading to a unique result for a seismic event. However, to be able to determine the fault rupture parameters precisely, it takes careful analysis from several stations. It means that a precise Moment Magnitude value is typically available only days after an event has passed. For smaller events, might not be calculated, once the different magnitude scales converge for the smaller earthquakes, and M_w

$$M_w = \frac{\log_{10}(M_0)}{1.5} - 10.7, \quad (2.6)$$

$$M_0 = \mu \langle \Delta U \rangle S = 2 \cdot 10^4 E_0, \quad (2.7)$$

$$\Delta\sigma = c \frac{M_0}{S^{3/2}}. \quad (2.8)$$

The magnitude of an earthquake evidences the strength of the *hazard* earthquake. There is another quantity, *seismic intensity*, which shifts the focus to how the seismic event effectively impacted a community, therefore attempting to provide a comparative measure for the *disaster* scale. It is expressed in the modified Mercalli scale, given on Table 2.1. After a strong earthquake, national operators produce maps showing the intensity in each district, or a contour map showing the regions of same intensity (isoseismal map). The data used to produce these are a mix of inputs from seismometers with monitoring of messages sent across social media about a seismic event. The intensity maps support decision-making in rescue and recovery actions and may assist the zoning of areas with higher risk of damage. There are locations, such as Mexico City in the 1985 and 2017 events, where the seismic intensities can be higher than surrounding neighbouring areas, even those closer to epicentre. This is due to a peculiar geomorphology of the region,

Table 2.1: Modified Mercalli seismic intensity scale

Intensity	Definition
I	Not felt (but detected by seismogram)
II	Felt only in special circumstances (upper floor of buildings)
III	Felt indoors, similar to the passage of vehicles
IV	Light shaking, no damage
V	Moderate shaking, very small damage (plaster, dishes)
VI	Strong shaking, few damages (moving furniture)
VII	Very strong shaking, medium damage (affect poorly designed structures)
VIII	Severe shaking, medium-high damage (partial collapse of buildings)
IX	Violent shaking, heavy damage (buildings shift off foundations, ground cracks)
X	Extreme shaking, very heavy damage (structures destroyed with foundations)
XI	Only few structures remain in place, large ground fissures
XII	Total destruction of cities. Objects attached to ground thrown to air.

Based on [3].

consisting of a layering of soft clay with hard rock and stiff soil deposits, which amplifies the effect of the tremors [32].

2.1.4 Earthquake occurrence: location and magnitude distribution

The occurrence of earthquakes empirically follows a frequency-magnitude rule, the Gutenberg-Richter scaling law (Eq. 2.9) [33–35]. This relation can be observed by means of a semi-log plot, where the logarithm cumulative number N of earthquakes with at least magnitude M decreases linearly with the value of this magnitude. The linearity constants a and b may be simply determined by regression. The value of a is an indicator of the local seismicity rate and depends on the number of events in the dataset. The b -value, on the other hand, is the coupling constant between earthquake frequency and magnitude, which is about $b = 1$ in active seismic zones, oscillating only slightly along time. Figure 2.8 shows the distribution of earthquakes with the magnitude for the events around the Japanese trench area for the year of 2017. The discrepancy observed at small magnitudes is due to lack of data – instruments might not be able to detect the smallest events, unless they are located close to the epicentre. For this, a starting value – Magnitude of Completeness – is selected. The moment magnitude of an earthquake is connected to the rupture area S of a fault (as seen in Eq. 2.8), so that the Gutenberg-Richter law can alternatively be expressed as a power-law relation with the rupture area assuming constant stress drop

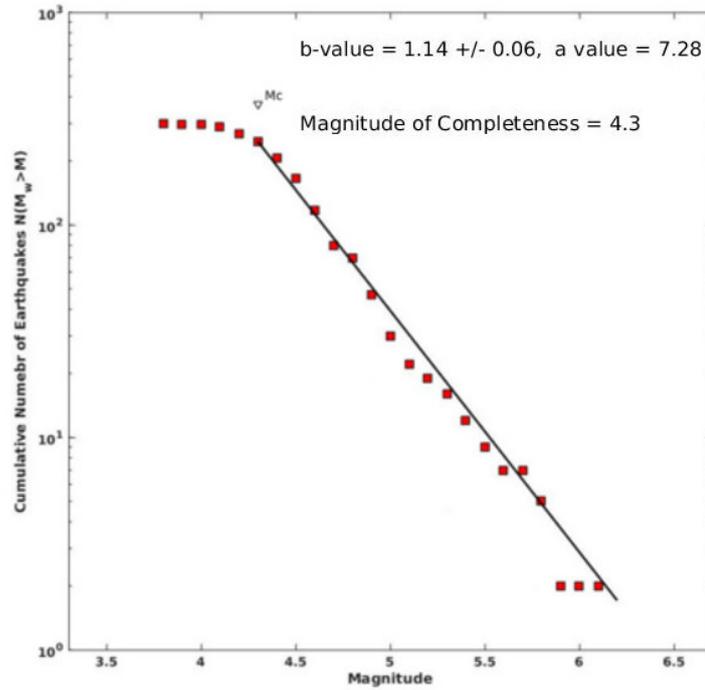


Figure 2.8: Observation of Gutenberg-Richter law for the 299 recorded events on the year of 2017 on the area around the Japanese trench ([latitude range,longitude range]=[34.98–42.48, 138.75–147.02]). Note that b -value is close to 1. Plot constructed by a modification of Gismo toolbox [37]. Data from IRIS consortium, downloaded using Gismo for Matlab.

(Eq. 2.10) [3, 36]. For very high magnitudes (>7.0) it will also be observed a cut-off on the Gutenberg-Richter law, due to the physical constraints of the rupture area of a geological fault. Table 2.2 details the global rate of occurrence of seismic events in relation with the magnitude

$$N = 10^{a+bM}, \quad (2.9)$$

$$\log_{10} N = c - (2/3) \log_{10} M_0, \quad (2.10)$$

$$= c - \log_{10} S. \quad (2.11)$$

The epicentre of large earthquakes follows a global spatial distribution shown in Fig. 2.9 for $M_w > 5.5$ in the period 01 January 2008 to 31 December 2017. These large earthquakes are due to the presence of large rock discontinuities and possibility of movement between the blocks. Such is the case of the subduction (vertical) or transform (horizontal)

Table 2.2: Estimated global earthquake occurrence for each magnitude band.

Magnitude (M_w)	Estimated number of events
4.0-5.0	10000 per year (~ 30 per day)
5.0-6.0	1500 per year (~ 4 per day)
6.0-7.0	150 per year
7.0-8.0	15 per year
>8.0	1 per year - 1 per decade

Note: Estimates based on average retrievals of the IRIS catalogue for the period 2008-2017, except for the latter band where it was considered the period 1968-2017.

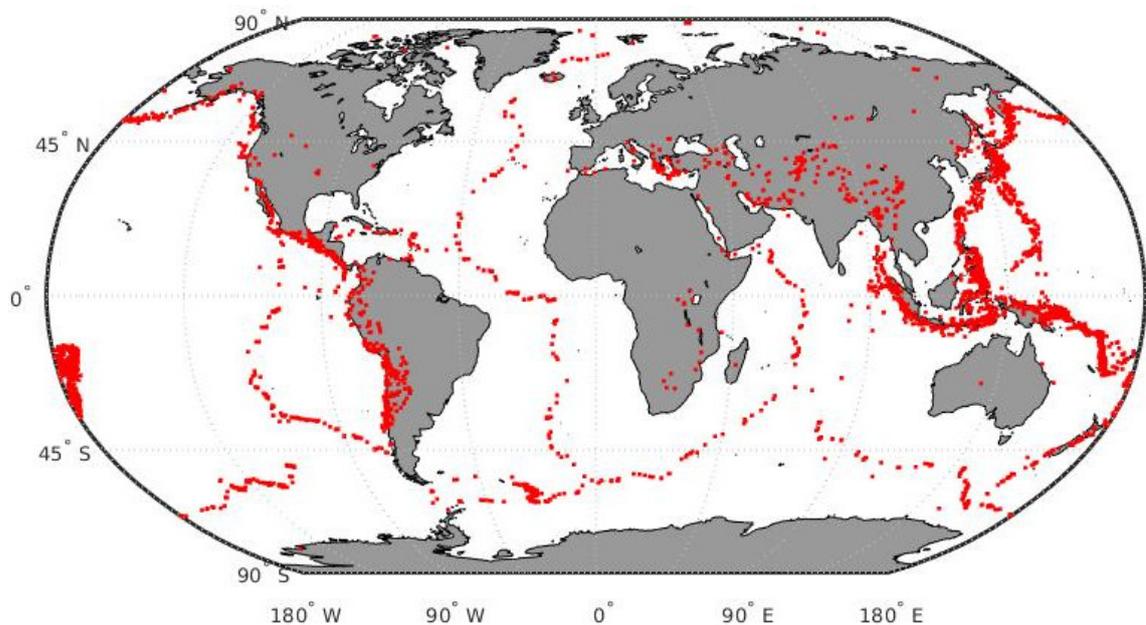


Figure 2.9: Spatial distribution of $M_w > 5.5$ earthquakes around the globe, 2008-2017. Data retrieved from IRIS catalogue.

boundaries at the edge of the tectonic plates that split Earth's crust into large sections able to move independently. The motion of each plate is maintained due to mantle convection, differential gravity force in higher and lower density sections leading to a sliding effect, tidal forces and global deformations. These plate boundaries constitute faults of several kilometres, able to release higher energies when a sudden movement occurs after a period of accumulation of stress.

2.1.5 Foreshocks, aftershocks and swarms

Large earthquake events rarely occur alone. They typically occur together with smaller events either preceding (foreshocks) or succeeding (aftershocks) the largest event (mainshock). There is a third possibility, a *swarm*, when a sequence occur without clear distinction of a driving mainshock and the aftershocks. The identification of a sequence as a swarm or regular, and the individual of single events foreshock, mainshock or aftershock, can only be determined after the sequence has finished. For large events the sequence usually takes 1-2 weeks, but it can extend for months. Figs. 2.10 and 2.11 illustrate the cases.

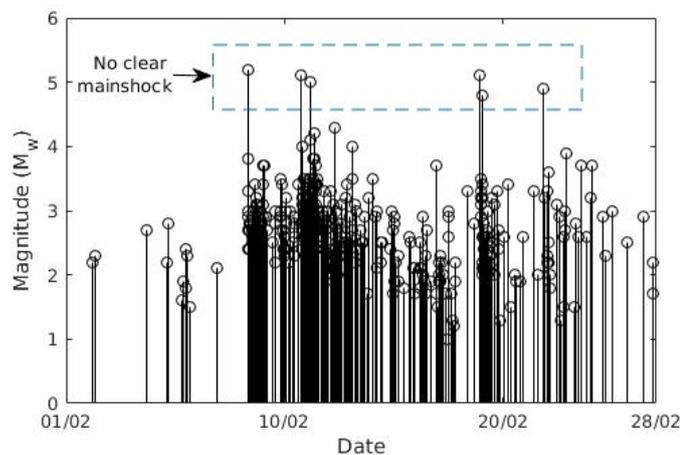


Figure 2.10: Swarm earthquake sequence in Mexicali, US-Mexico border in California, during February 2008

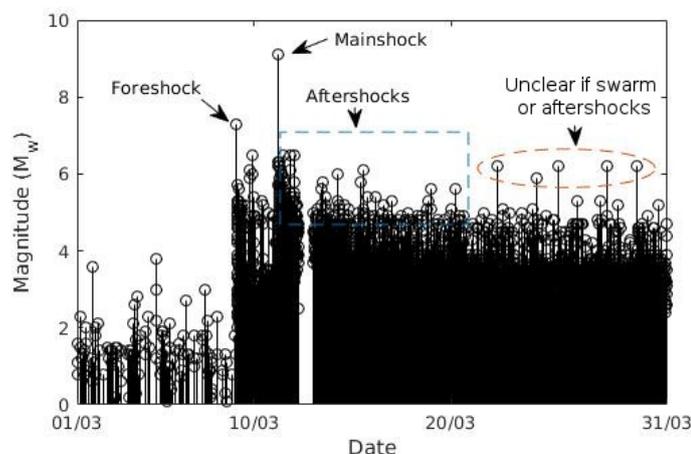


Figure 2.11: Tohoku-oki earthquake sequence, Japan, March 2011. Foreshock, mainshock and aftershocks are displayed. Also a series of events are shown which can be interpreted as either part of the aftershock or as a swarm sequence.

Measuring the frequency of aftershocks $n(t)$ in function of the time t elapsed from the mainshock leads to the observation of a decay given by the (generalised) Omori's law in Eq. 2.12. The constants k and C are also empirically defined varying on the earthquake sequence being studied. The rate of decay p is typically fixed for an area, but also empirically defined. The first description of the law [38] fixed $p = 1$, but latter accounts for other areas led to the generalised version [39], which allows p to vary, typically in the range 0.7–1.5. Most aftershock sequences also observe Bath's law, which is a statement that the magnitude difference of the largest aftershock and mainshock in an area is about 1.2 [40, 41]. This has consequences in civil protection, as implies that the largest earthquakes would be followed by other significant events, with potential of disruption while the response to the mainshock is still taking place

$$n(t) = \frac{C}{(k+t)^p}. \quad (2.12)$$

Foreshock sequences can vary significantly in behaviour. It can be observed single events of high intensity, or a cascade of small events (presumably triggering the motion of the mainshock). There are also cases where no foreshock is observed. Whenever a foreshock is clear, the period of its occurrence is typically within a week of the mainshock. The study of foreshocks is an open debate both in earthquake forecasting and prediction paradigms. Ogata [42] proposes that the statistics that support the aftershock distribution can also be applied for the foreshock, in particular to determine the likelihood of a foreshock sequence leading to the mainshock. This model, described as the Epidemic Type Aftershock-Sequences (ETAS), is given by

$$\Lambda(t|\text{Hist}) = C_0 + \sum_{t_i < t} \frac{C_i}{(k + (t - t_i))^p} e^{\alpha(M_i - M_{\text{cutoff}})}. \quad (2.13)$$

In this expression, Λ is the seismicity rate function depending on the history of events: for an interval $(t, t + \delta t)$, $\Lambda(t|\text{Hist}) = \lim_{\delta t \rightarrow 0} \frac{p_{\delta t}(t|\text{Hist})}{\delta t}$, where p is the probability of an event. The constant C_0 is the background seismicity rate, α is the efficiency of an earthquake generating another, and M_{cutoff} is the cutoff magnitude of the dataset. The

log-likelihood of a set of events $(t_1, M_1), \dots, (t_n, M_n)$ occurring is given by Eq. 2.14. This can be interpreted as the chance of a set of seismic events occurring in future given the history

$$\log L((t_1, M_1), \dots, (t_n, M_n)) = - \sum_{i=1}^n \Lambda(t_i, M_i) - \int_0^T \int_{\delta M} \Lambda((\delta t), (\delta M)) d(\delta t) d(\delta M). \quad (2.14)$$

The ETAS model constitute the backbone of modern short-mid term earthquake forecast.

2.1.6 Gravity and displacement changes

Following an earthquake, there are permanent local gravity and surface displacement changes occurring in two time-scales: co-seismic and post-seismic. Co-seismic effects refer to the comparison of measurements previous to the mainshock with those immediately afterwards (up to a few weeks). Post-seismic effects refer to a similar comparison in the scale of months following the principal seismic event. Hence, the co-seismic gravity and displacement changes are directly associated to the earthquake movement, and are used to assess the characteristics of seismic event, and the current and likely future faulting behaviour in that specific area. They reveal the short-term land topography changes (with possible geographic implications, including the definition of borders between properties or even nations) or the modifications in the underground structure (which affect how future earthquakes will occur). The post-seismic changes, comparatively, are associated to the resettling of the fault zone after the event, and are highly location dependent.

Theoretical co-seismic displacement and gravity change effects due to the dislocation of a finite rectangular fault were respectively defined by Okada and Okubo in [43, 44]. The numerical solution of the application of these to the 2016 M_w 6.2 Amatrice (Italy) earthquake is presented on Fig. 2.12. Close to the epicentre the changes can be considerable, in this case reaching up to $1.5 \mu\text{m/s}^2$ gravity variation (the order of tidal effects) and about 1m maximum absolute displacement. However, the decay is fast as the distance to the epicentre increases. If a sensing network is located at a distance of 10-50km from the epicentre, the observed co-seismic displacements would be not significantly greater than

10cm, and the surface gravity change would be in the order of 10nm/s^2 for a similar event. This is the scale of reported co-seismic changes [45–48].

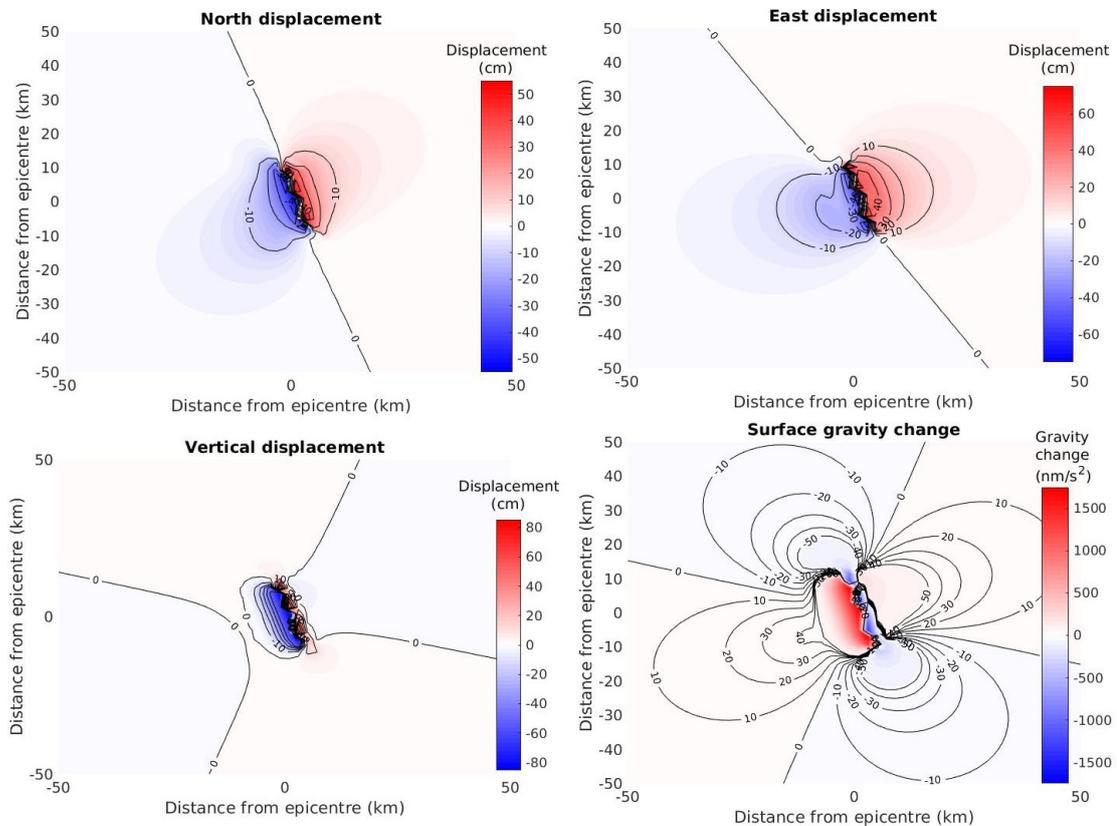


Figure 2.12: Expected co-seismic displacement and gravity change due to the 2016 M_w 6.2 Amatrice earthquake from Okada [43] and Okubo [44] solutions. Computed using okada85 and okubo92 tools for Matlab [49, 50], and adopting the earthquake source parameters for single fault approximation of this event presented on [51].

2.1.7 Exotic phenomena

Recently discovered or exotic phenomena being currently investigated in observational seismology include:

- Prompt/transient gravity changes: small fluctuations in the surface gravity caused by the P-wave. As P-wave are compressional, the mass is redistributed in their passage, with a gravity field effect calculated to depend with the distance to the wavefront as r^{-4} [52, 53]. Modelling and reported attempt observations are of the order of 1nm/s^2 transient gravity changes lasting for seconds to few minutes before P-wave arrival [54, 55].

- Intraplate seismicity: despite being rare, significant intraplate events can happen. This is the case of the events affecting the Northeastern region of Brazil [56–58], including the 1986 João Câmara earthquake sequence which damaged or destroyed 4000 homes. Not always a geological fault system can be attributed to intraplate events. Hypothesis for these events include the reactivation of ancient, deep buried, geological faults [59, 60] which are difficult to observe. Alternatively, it has recently been proposed that the long-term development of conditions for concentrated crustal deformation spots within a plate could lead to these events [61].
- ‘Silent’ earthquakes, or slow earthquakes: these are seismic events which the energy is released in the time-scale of days, instead of seconds-minutes. Because they are slow, the signal might be so weak as to not being able to discern P and S-wave modes, with overall form resembling instead a long-period background noise. These seismic events can have an equivalent high magnitude ($M_w 6 - 7$) and yet no sensible effects be perceived on the surface structures (hence ‘silent’). The most active areas are alongside the Japanese East coast, the North-East USA – South-East Canadian coast, and New Zealand. A review of these events can be found on [62]. Only recently the conditions of stress and frictional properties allowing these events could be reproduced in laboratory conditions [63].
- Episodic tremor and slip (ETS): this event consists of a constant slip between the plates occurring for a defined time (ranging from months to years depending on the location) which periodically stops giving rise typically to a slow earthquake, and the slip starts again in sequence. They are observed in subduction zones, with the epicentre positioned near the boundary of the crust with the mantle (Mohorovic discontinuity). It has been proposed that the presence of a water deposit in the interface of the plates could support this phenomenon [64, 65].

2.2 Earthquake prediction

2.2.1 The field of earthquake prediction

Earthquake prediction consists of determining physical events that precedes earthquake occurrence. Bullen and Bolt [5] argue that there are different kinds of prediction theories: descriptive, when a pattern is simply found from direct observation; inductive, when current theory require additional hypothesis, these with predictive power; and deductive, when current theory leads to consequences which have predictive power. Different than what the name suggests, the practice of earthquake prediction is not on identifying new single earthquake events that will occur, but to investigate past seismic records and determine the other events associated to earthquake occurrence, so that a model could be proposed which, in a future moment, might be applied for forecasting scenarios and development of warning systems with greater time margins. Events associated to earthquake occurrence could be direct or indirect causes, or *precursors*: anomalies occurring in other physical events before the mainshock. Precursors might be a cause for earthquake occurrence, but also could be other events which share a common cause to the seismic activity (mediated causality). As it works with past observations, the methods for analysis in the earthquake prediction tends to be deterministic in essence, *i.e.*, to unambiguously associate a set of physical events to the specific seismic phenomena. Derived models for future events may have stochastic basis, hence linking to *forecasting*, but this is rather an application.

2.2.2 IASPEI list of precursors

The International Association of Seismology and Physics of Earth Interior (IASPEI), which is one of the institutions composing the International Union of Geodesy and Geophysics (IUGG), established in 1991 a set of requirements for candidates of earthquake precursors [24, 66, 67]:

- There should be a explanatory model for the physical mechanism of the precursor-mainshock relation, including predicted distance/magnitude variations.

- The data of observed events must be available, including all the calibration and treatments made.
- It should be presented the definition of the anomaly, enabling cross-validation with other datasets.
- The reasons for associating an anomaly with a event must be clarified.
- It must be included a discussion of possible false alarms and the possibility of the anomaly being co-seismic rather than pre-seismic.

The current precursor candidates on the latest IASPEI list [67] are:

1. Seismicity patterns

- Foreshocks (seismic events hours to weeks before the mainshock) [68]
- Pre-shocks (seismic events months to years before the mainshock) [69]
- Seismic quietness (quiescence) before large events [70]

2. Groundwater chemical properties

- Fluctuation in Radon gas levels dissolved in groundwater [71]

3. Crustal deformation

- Groundwater level increase [72]

In addition, there are nominations for which no final agreement has yet been reached (all of which referring to crustal deformation):

- Strain changes [73]
- Tilt changes [74]
- Crustal displacement [75]

Many of these events share common physical mechanisms. For example, the accumulation of strain in a seismically active zone leads to observable strain changes at the surface measured with a strainmeter. If a confined groundwater reservoir is present at the location, it can act as a strainmeter, as it will be detailed in next section. These crustal changes also link to variations in the tilt angle between a surface and the gravity direction, either because the surface topography is slightly modified, or because the distribution of mass below has steadily changed, leading to a small variation in the direction of the gravity acceleration vector \mathbf{g} .

Foreshocks and pre-shocks are two similar cases, referring to past earthquakes creating the conditions for triggering a earthquake by rearranging the fault in unstable conditions. They differ essentially in the time-scale of the processes considered. Seismic quietness, on the other hand, is the opposite geomorphological case, when the fault becomes locked to the point that smaller earthquakes are not allowed to occur, and all the stress is stored for the mainshock. Crustal displacement changes complement these candidates, by suggesting that these changes can be indirectly monitored by the surface displacement, which is an indicator of the plate dynamics.

The radon fluctuation is a location specific precursor candidate. When some rocks are submitted to stress they release trapped gases. This is more directly observed in the cases of materials that suffer liquefaction or susceptible to microfracturing. Among the typical gases released is Radon ^{222}Rn , a radioactive element product of ^{226}Ra decay on the ^{238}U series, that is continuously monitored in underground water wells in Europe following an EURATOM agreement on natural radioactivity hazard assessment [76]. For example, granite is a type of material rich in ^{238}U , able to release significant quantities of trapped Radon under the appropriate stress conditions. Currently other gases are also monitored in water wells and atmosphere as potential seismic precursors, particularly Ozone [77], but, as a convention, the title ‘Radon emission’ also applies as a common nomenclature for all the other gases. This precursor candidate is among the most controversial, once it appeared present in emblematic cases such as 2009 M_w 6.3 L’Aquila event, but closer consideration of the event shed doubt on the Radon data obtained and attempted

forecast made [78].

Current open research include:

- Changes in the ratio of speeds of P and S waves in small local earthquakes (V_p/V_s ratio)
- Variations in the Gutenberg-Richter b-value
- Electromagnetic emissions at ELF-VLF bands
- Temperature variations
- Changes in Earth resistivity
- Variations in the amplitude of some tidal modes
- Changes in the local surface gravity

A complete review can be found on [21]. On the following section we will expand on the variations of the b -value and gravity changes, as well as presenting a case for tidal triggering of earthquakes.

The precursor candidates are connected according to the blue arrows of Fig. 2.13. The orange arrows of the figure indicate significant co-seismic effect that should be considered.

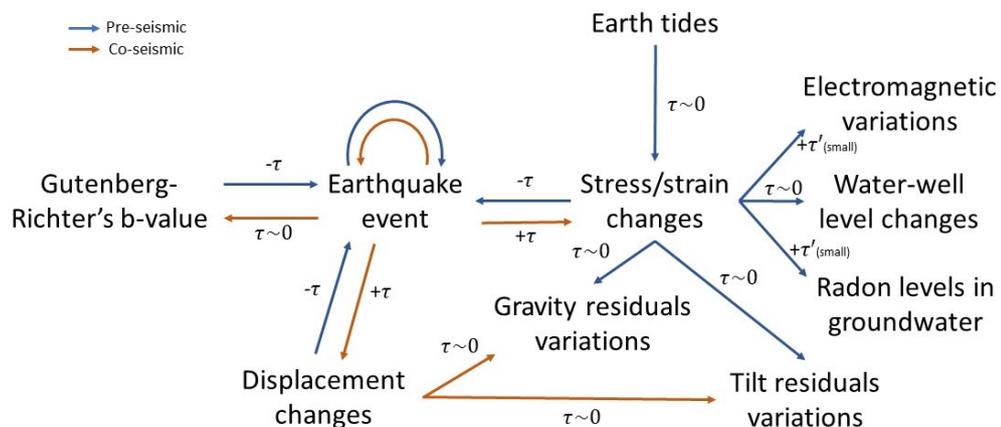


Figure 2.13: Diagram showing the expected physical links between precursor candidates and the occurrence of a seismic event. τ refers to a time-delay between the variables, and the colours clarify the pre-seismic links (blue arrows) and the co-seismic links (orange).

The label τ above the arrows is an indicative of the timing of the process: $-\tau$ means a precursory behaviour, $+\tau$ a seismic effect, $\tau \sim 0$ a reference that the referred process has small response time.

2.2.3 The case for b-value anomalies

Laboratory studies based on application of stress in rocks find that b-value varies for higher stress, presenting particularly high anomalies before rupture [79, 80]. For higher stress it is observed a larger fracturing, leading to an increase of the larger earthquakes compared to the smaller, reducing the b-value. Main *et al* [80] describe two model predictions: for a elastic failure (stress increasing linearly until the rupture point), it should be observed immediately before the rupture a drop in b-value up to 0.5, while for an anelastic failure (stress-strain is nonlinear as it approaches the failure point), it should be observed a small increase in b-value (about 0.2) on a first moment, followed by a sudden drop to b-value close to 0.5 before the rupture. This supports observations by Imoto [81] of unusually small b-values in the interval of weeks up to 10 months before large earthquake events in Japan, while high b-values were sometimes observed in the period 16-24 months before the mainshock.

2.2.4 The case for tidal triggering of earthquakes

The mechanism of earthquake triggering is based on the accumulation of strain/stress at the fault zone, as classically presented by the elastic rebound theory (earthquake occur when the deformation in the fault area surpasses the critical elastic point) . The tidal effects, mainly due to the Moon and the Sun, generates a significant extra stress field and a deformation in the Earth's crust, so that it could provide the extra push in faults near critical loading. Furthermore, in faults located in ocean zones, the drag from water movement above the sea bottom can provide additional stress. In this line of reasoning, it is expected that most earthquakes occur during spring tides (14-day period) – when the largest tidal variations are observed, due to full moon or new moon (syzygy). However, high correlation could be empirically observed only in special cases, particularly consisting of shallow earthquake sources in locations where ocean tides are large, due to the

local topography [82, 83]. In other cases, there is low correlation between tidal events and earthquakes [84–86]. As will be detailed in Chapter 6, we have obtained an initial indication towards shared information between tidal amplitudes and seismicity, however more data is required for a robust conclusion.

2.2.5 The case for pre-seismic gravity anomalies

As the stress accumulates in the fault zone and deformation occurs, the variation in the mass distribution is expected to reflect in potentially observable gravity changes (up to the scale of $\mu\text{m/s}^2$ over a long period). Zhan *et al* [87] conducted a seven year study with a network of 385 gravity stations (25 absolute FG-5 instruments, 360 relative LCR-G instruments) across China, extending further three years in the region of Sichuan. The authors consider that nine large earthquakes in the period 1998-2008 had pre-seismic gravity anomalies up to $1.3\mu\text{m/s}^2$ peak-to-valley, observed both before the 2001 M_w 7.8 Kunlun earthquake of magnitude M_w 7.8 and before the 2008 M_w 7.9 Wenchuan earthquake. The satellite mission Gravity Recovery and Climate Experiment (GRACE, 2002-2017), allowed for observations of gravity anomalies across the globe (even in locations without precise gravity stations), as long as the gravity change is sufficiently large. This has also been employed for the observation of possible gravity changes before large events, such as 2010 M_w 8.8 Chile, 2011 M_w 9.0 Tohoku-Oki and 2012 M_w 8.6 Indian Ocean earthquakes [88]. In these cases anomalies are observed 2-5 weeks prior to the mainshock.

2.3 Instrumentation in seismology

2.3.1 Seismometer

Seismometer is the sensor designed to measure a component of the ground motion during earthquakes. A seismic station typically includes a few seismometers (at least 3, one for each ‘channel’ or cartesian direction) and an accurate clock for time-keeping (often a portable Rubidium atomic clock), besides instruments for recording and communication of the acquired data and possibly geolocation instruments. A simple vertical seismometer is depicted in Fig. 2.14, while a horizontal seismometer is shown in Fig. 2.15. Fig.

2.16 shows a third type, a torsion seismometer, more specifically the Wood-Anderson seismometer used by Richter to define his local magnitude scale [31].

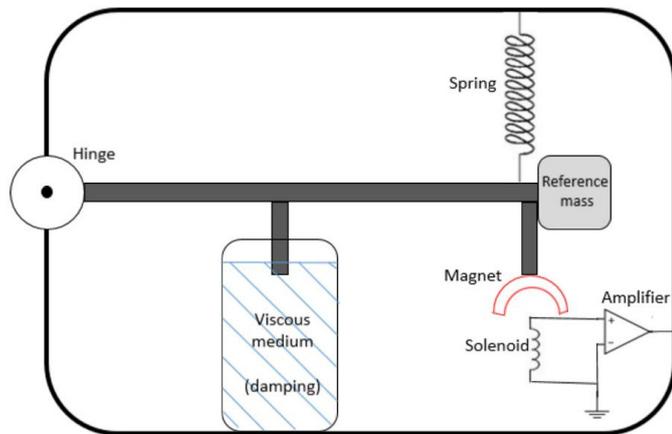


Figure 2.14: Schematic diagram of a simple vertical seismometer.

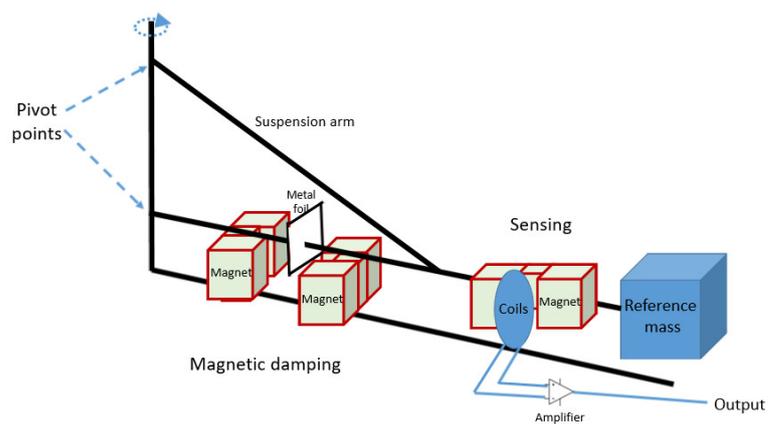


Figure 2.15: Schematic diagram of a garden-gate horizontal seismometer.

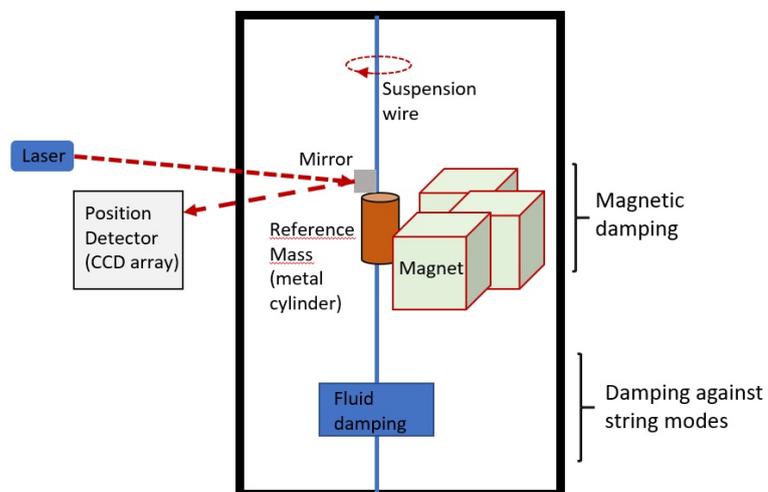


Figure 2.16: Schematic diagram of a Wood-Anderson torsion seismometer.

The appropriate selection of the damping not only provides stability to system, but allows to tailor the desired optimal frequency response. The example systems shown are generally designed for constant amplitude response for frequencies above 1Hz. This is sufficient to retrieve the main behaviour such as the arrival of the P and S waves, the amplitude of the P-wave (to infer the local (Richter) magnitude M_L) and the amplitude of the 20s-period surface wave (to infer the surface magnitude M_s). However, the lower part of the spectrum is not observed. Modern designs, such as the broadband STS-1 and STS-2 seismometers or the borehole KS5400, based on measuring the acceleration relative to the inertial reference, are able to maintain a constant response amplitude starting at 0.01Hz, and a operating range starting at 1mHz. This allows for a more complete picture of the seismic signal. A comprehensive review of seismometry instrumentation principles and response spectra can be found on [3–5, 89].

Data from 1433 networks of seismometers is available from the consortium International Research Institutions for Seismology (IRIS)¹. This is a joint effort from governmental, educational/research and private seismic network operators, which deposit the acquisition from their seismic sensors on real-time or as a historic record. Some operators allow public access and enable real-time acquisition/monitoring. The most relevant to this work is the Global Seismograph Network (codes ‘IU’, ‘GT’, ‘CU’, ‘II’ and ‘IC’, operated by IRIS together with the United States Geological Survey (USGS) and local operators) for the large number of broadband seismic stations (150) almost uniformly spread across the globe².

As a operator detects a seismic event, it reports the observed parameter to one a research partner aiming to discover the source parameters, producing a catalogue entry. Catalogues are created by a number of institutions, such as the International Seismological Centre (ISC), the USGS National Earthquake Information Center (NEIC), the Japanese National Research Institute for Earth Science and Disaster Resilience (NIED) and others. Many of these independent catalogues are also collaborators of IRIS, which is then able

¹IRIS Data Management Center: <https://ds.iris.edu/ds/nodes/dmc/>, 428 of which are permanent installations. List of federated networks: <http://www.fdsn.org/networks/>.

²We have prepared a hands-on activity available on [90] on how to obtain and work with seismic data.

to produce a unified global database of earthquake occurrences.

2.3.2 Gravimeter

Gravimeters are devices designed to measure the local earth gravity acceleration. They are classified in two groups: absolute gravimeters and relative gravimeters. The relative gravimeters measure small variations of the gravity value in time, while the absolute gravimeters provide the reference to which the relative devices are calibrated.

Absolute gravimeters are based on the falling of a reference mass on the effect of the gravity force. For example, the commercial absolute gravimeter FG-5 principle of operation is a falling mirror of defined mass in the movable arm of a Mach-Zehnder interferometer, in a vacuum chamber (Fig. 2.17) [91]. Associating the precise dislocation measurements from the moving fringes of Mach-Zehnder optical interference with a time-keeping provided by a Rubidium atomic clock, a curve of time versus dislocation is obtained, from which the accelerating force due to local gravity is extracted. The accuracy of this instrument (agreement between instruments) is of 20nm/s^2 and precision (least significant bit) of 10nm/s^2 for a total time of 3.75 minutes. A modern alternative is the use of falling cold atom clouds, which are split and recombined during the fall trajectory, producing matter-wave interference pattern linked to the value of absolute local gravity acceleration. This is a technology still under development, and further details of a possible design are explored in Appendix B.

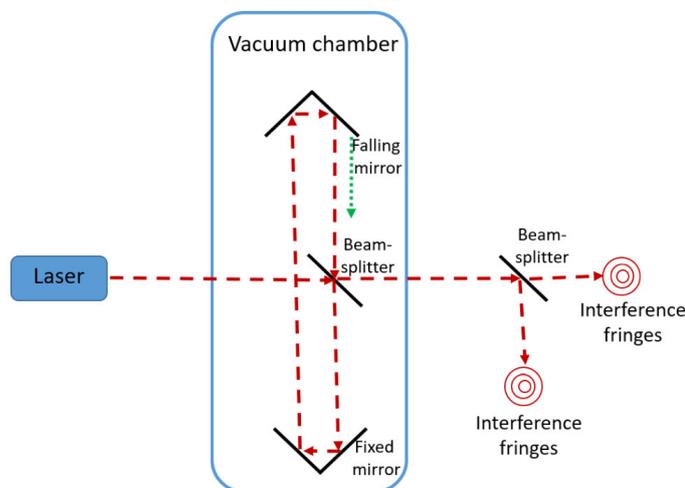


Figure 2.17: Schematic diagram of an absolute gravimeter with Mach-Zehnder optical design.

Relative gravimeters are based on investigating the change in the equilibrium point of a reference mass. The simplest designs, based on a reference mass suspended by springs, allows the vertical position of the mass to change with gravity, and the equivalent effect in distension of the spring or of a lever system supporting the spring is measured. This simpler system allows for great portability, but it has a intrinsic limitation on the frequency range and of being sensible to drifts and noise effects due to the moving mechanical parts which suffer thermal variations and wear with time. An alternative approach is given by the superconducting gravimeters, which also measure relative gravity.

Superconducting gravimeters are the most precise gravity sensing devices, with time-domain accuracy of 1nm/s^2 and precision and long-term frequency-domain accuracy up to 1pm/s^2 . Analogously to the mass-spring counterpart, the principle of the superconducting gravimeter is to investigate the equilibrium point of a reference mass, but this time by determining the conditions to keep it in inertial position when the local gravity acceleration change. The suspension springs in this case are replaced by an equivalent "perfect" spring consisting of magnetic fields repelling a superconducting sphere due to the Meissner effect, able to maintain it levitating in a fixed position. If the gravity changes, capacitive plates encircling the sphere senses the shift in position and a feedback is applied to restore the initial equilibrium point. From the different magnetic field needed to be applied to maintain the sphere in the same position along time, it is obtained the fluctuations in local gravity. Furthermore, the absence of moving mechanical parts reduces the major factors of drift and noise of spring-mass gravimeters. A simplified design of a superconducting gravimeter is shown in Fig. 2.18. More details of these instruments can be found on [92].

A global network of superconducting gravimeters, the Global Geodynamics Project (GGP), was implemented in 1997, containing 25 stations and enabling public use of record data for research with non-commercial purposes. This was replaced in 2015 by the International Geodynamics and Earth Tides Service (IGETS)³, currently with data from 35 superconducting gravity stations and 50 instruments (a station can have more than

³<http://igets.u-strasbg.fr/>

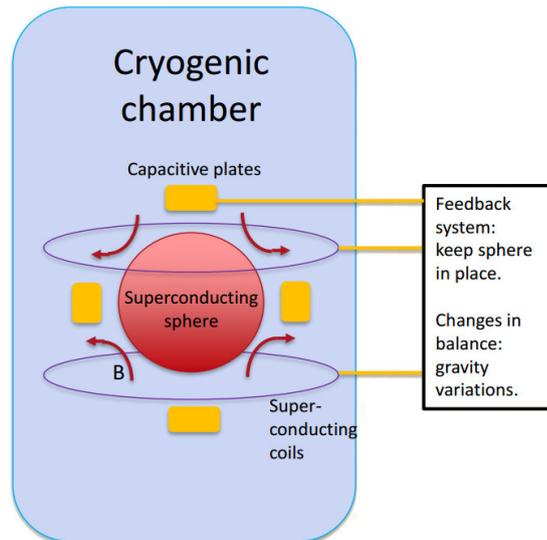


Figure 2.18: Simple schematic diagram of a superconducting gravimeter.

one gravimeter). The IGETS system currently does not enable real-time data acquisition, but 3 stations, Strasbourg (France), Djougou (Benin) and Membach (Belgium), enable it independently, through their websites⁴ or from IRIS Data Management Center.

2.3.3 Other devices

Additional devices used to support the analysis of seismic phenomena include displacement sensors, strainmeters and tiltmeters, which will be briefly described next. To our knowledge there is no unified, cross-institutional, infrastructure for global networks of these instruments, as opposed to IRIS for seismometers and IGETS for superconducting gravimeters. This makes difficult the long-term or global analysis of displacement, strain and tilt patterns, and studies are often restricted to cases. For this reason, here we briefly present these devices and applications for completeness, but only the more complete catalogue from seismometer and gravimeter data has been used for the analysis in the subsequent chapters. However, the method of analysis to be presented on the next chapters can be equally applied for the data analysis from these instruments, once available.

⁴Strasbourg and Djougou: <http://cdg.u-strasbg.fr/PortailEOST/Gravi/v1/>.
Membach: <http://www.seismologie.be/en/gravimetry/observations/real-time-g>

2.3.3.1 Displacement sensors

Currently the most usual displacement sensor is the Global Navigation Satellite System (GNSS) receiver. This type of sensor can pick up the signal from multiple geopositioning satellite systems such as GPS, GLONASS and Galileo, and determine displacements on the order of 1cm precision/accuracy level. Comparatively, the largest tidal component on the solid Earth surface (semidiurnal tide, mode M2) has a vertical amplitude of approximately 38cm and horizontal amplitude of 5cm, so they have to be filtered for precision measurements. Constant movements from the plate tectonics produce displacements of the order of about 0.05 – 10cm/year. Large seismic events such as 2009 M_w 6.3 L’Aquila seismic event produced co-seismic displacements up to 10cm in the horizontal and 30cm in the vertical and post-seismic displacements up to 1cm in the horizontal and 5cm in the vertical [93]. Notice that a GNSS receiver can only provide information about the displacement at a single point, and not always a dense array of instruments is available. A complementary alternative is to use satellite imaging or airborne radar imaging directly to determine the topography changes before and after an event. For higher precision, this is made through techniques such as Interferometric Synthetic Aperture Radar (InSAR, using microwave) or Light Detection and Ranging (LiDAR, using laser), *e.g.* Fig. 2.19. These produce topographic maps with about 1cm level accuracy on any area of the planet.

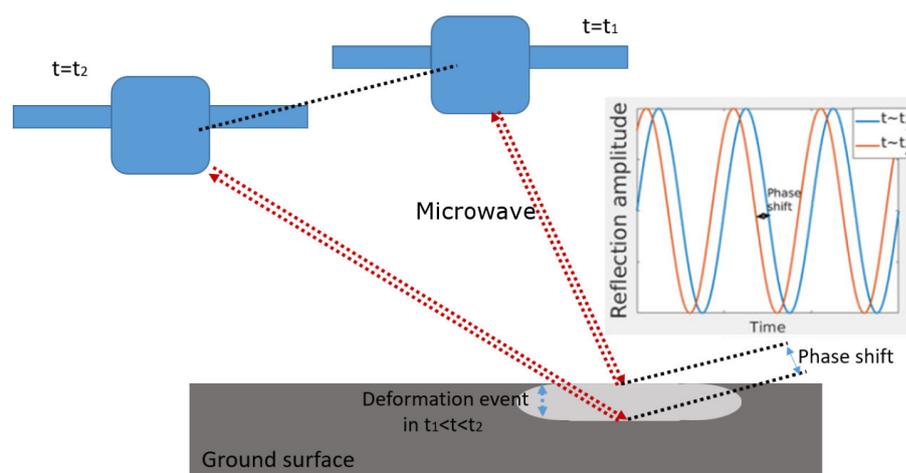


Figure 2.19: Simple diagram of the acquisition of InSAR image. Microwave frequencies are preferred for active remote sensing techniques due to a atmospheric absorption window. The LiDAR is analogous, using as source a 1064nm (infra-red) laser for surface and 532nm (green) laser for sea mapping instead. The point phase difference is calculated over a large area of interest, producing an interferometric map.

2.3.3.2 Strainmeters

Strainmeters are devices designed to measure the differential strain in a rock section as a function of time. From the stress-strain relation ($\sigma = 2\mu\varepsilon + \lambda\text{tr}(\varepsilon)I$), the stress accumulated at the fault can be inferred. A uniaxial strainmeter can be as simple as a spring attached to the ground, or a rod inside a tube (like a piston) with a mechanism to determine its position (*e.g.* by measuring varying capacitance defined the piston walls). Precise optical alternatives are also possible such as the long-baseline strainmeter shown in Fig. 2.20.

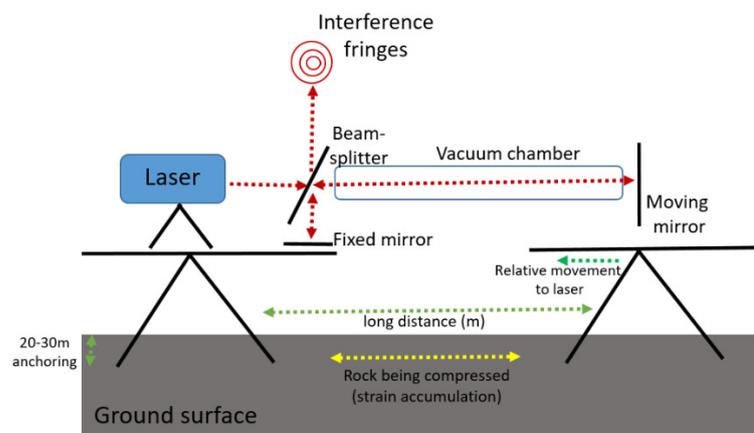


Figure 2.20: Schematic diagram of a long-baseline strainmeter with Michelson interferometer design.

Volumetric strain can be measured by borehole strainmeters or by measuring the level of confined water-well systems. In the case of confined water-wells, as strain accumulates in the system, the pressure in the reservoir increases and the water levels rise proportionally (Fig. 2.21). To identify if a water-well is confined or not it is necessary to observe a long time-series record, separate effects from rainfall/drought and possible consumption, assess if the tides are present and the amplitudes of each tidal frequency, confirm that the observe effect is not atmospheric pressure response, and then determine the aquifer type (fully confined, semi-confined, semi-unconfined, unconfined). Only fully confined aquifers can confidently respond as volume strainmeters.

A review of strainmeter theory and design is presented on [94]. The application of water wells as a volumetric strainmeter is detailed on [95].

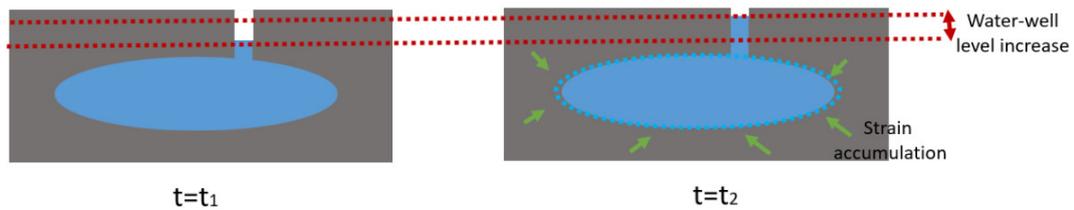


Figure 2.21: Schematic diagram of a water well from a confined aquifer serving as a volume strainmeter.

2.3.3.3 Tiltmeters

Tiltmeters are devices that measure the inclination of the surface they are attached to in comparison to the direction of the gravity acceleration vector. Examples of basic tiltmeters are a bubble/spirit level or a long simple pendulum. High-precision devices are improvements on the readout and stability of these mechanical systems, but are based on the same operating principles. Tilt variations may occur due to displacements in the attached surface or due to small changes in the direction of \mathbf{g} . The first, associated with strain data, is able to inform rotational movements in the fault system, while the latter reveals variations in the local mass distribution, complementing the information provided by gravimeters. An example of a simple tiltmeter in operation is given on Fig. 2.22. The review on [94] provides examples of commercial designs and expected response.

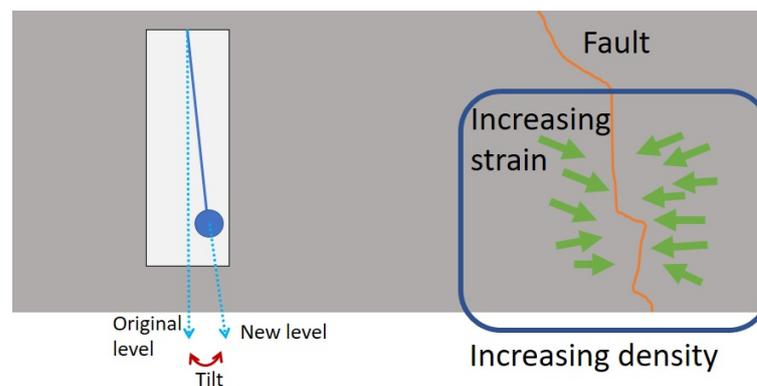


Figure 2.22: Diagram showing the variation of the level of a borehole pendulum tiltmeter due to mass redistribution in a fault system.

Chapter 3

Time-series considered

3.1 Geophysical time-series

We will here consider two types of geophysical time-series: seismic, which might be the trace or the event catalogue, and gravity, obtained from IGETS network of superconducting gravimeters.

3.1.1 Seismic data

As detailed in Chapter 2, when an earthquake occurs the seismic waves might be sensed by seismometers, which will record a trace, such as on Fig. 3.1. A local magnitude of the earthquake can be directly determined, as well as an estimative of the epicentral distance of the station. The epicentral distance is simply estimated by $D = \delta t(v_p v_s)/(v_p - v_s)$ for δt the time-difference of arrival of S and P waves, and v_s and v_p their respective average velocities. By knowing the epicentral distance of a few stations, one determines the location of the epicentre. For the magnitude, one take a closer observation at the P-wave and obtain its maximum amplitude A in mm of an equivalent response from a Wood-Anderson seismometer. The local magnitude, then, is estimated by $M_L = \log_{10} A + 2.76 \log_{10} D - 2.48$, as seen on Sec. 2.1.3 (Eq. 2.3).

If the event is identified by a sparse network of seismometers (between tens and thousand devices), it is possible to precisely determine the earthquake epicentre and depth, the area of rupture, the released energy, and, finally, the moment magnitude. For small events ($M \ll 5.5$) the Richter, surface and moment magnitude scales converge, so the

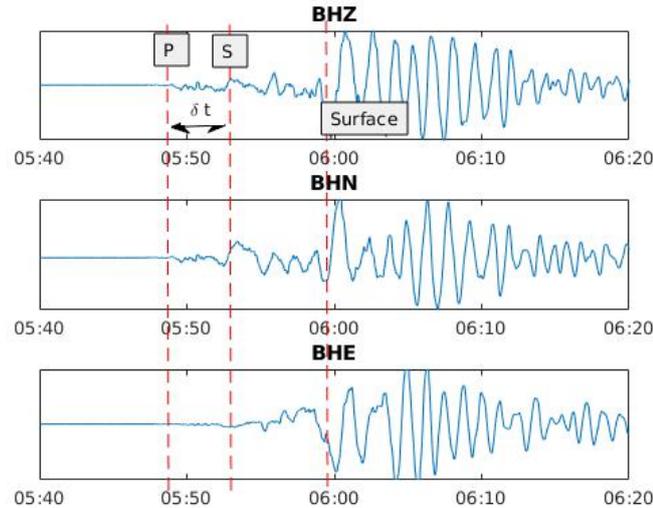


Figure 3.1: Trace from the 11 March 2011 Tohoku-Oki earthquake, as observed by the YSS station on Eastern Russia, close to Northern Japan. The three broadband channels are displayed (vertical: BHZ, horizontal North: BHN, horizontal East: BHE), as well as the start time of the P, S and surface waves, and the difference time δt which enables to infer the epicentral distance. Each channel expresses the signal from a seismometer – either the vertical or horizontal type discussed on the previous chapter – all of which located at the same place.

magnitude information is available immediately. However, for larger events it may be necessary a higher computational effort of inversion of seismic source parameters, not least because of the complexity of the event, which can involve a large number of fractures occurring simultaneously. Hence, it is not unusual that the initial moment magnitude estimate is updated in the first months following an events, only becoming available on international catalogues after an agreement has been reached.

By selecting a region of interest, and organizing the entries of a catalogue in chronological order, we can build a time-series of earthquake occurrences for this area. For this, we define a starting and an end time and a sampling rate t_s . If no event is recorded at a specific time, we assign a null marker, which could be the zero or a *not-a-number* (NaN) value. When an event does happen within the time-window $[t, t + t_s)$ we assign one of the earthquake parameters of interest. In our case, we consider the maximum magnitude (largest earthquake), the cumulative magnitude (sum of magnitudes above a threshold, so to consider the alternative that a large energy release was spread in many smaller events), and the number of earthquakes within the time-window.

Seismic datasets exhibit clustering of earthquake occurrences in time, *i.e.*, there are short time intervals with many events and long periods of relative seismic quietness. The mainshocks are the peaks of highest magnitude in each cluster, such as discussed in Sec. 2.1.5. It is expected for the largest events to be chronologically followed by a number of smaller events, in accordance to the aftershock distribution from Omori's law (Eq. 2.12). However, it is also possible that events happen before a mainshock, possibly explaining its occurrence (foreshock). In the case of a single or few single events leading to the mainshock, it follows the foreshock hypothesis, whereas, in the case of a Omori-like distribution of many small events, it verifies the consistency of the Epidemic-Type Aftershock Sequence (ETAS) model. Another possibility is that before a mainshock there are fewer entries than the normal, the seismic quiescence hypothesis.

Figure 3.2 exemplifies the case with data from the Central Mexico earthquake of 2017. After the mainshock we can observe some features. There is a window with relatively few events, as intuitively expected given that a large fraction of the accumulated stress on the fault was released in the mainshock, but then a few days later there is a new cluster of events, apparently triggered by smaller events.

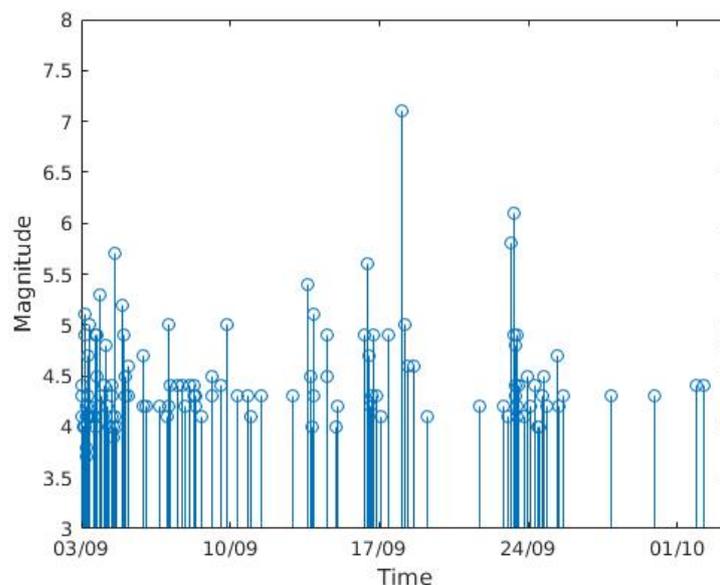


Figure 3.2: Sequence of seismic events around the 2017 M_w 7.1 Central Mexico earthquake. Dataset obtained from IRIS catalogue contains 132 events from 09 to 30 September 2017 on a circular region of 5° radius centred at 18.584° N, -98.399° W.

This clustering behaviour is consistent with Omori's law and derived ETAS model, as the occurrence of a smaller event enhances the probability of occurrence of a novel one. Bath's law is also observed, as the largest aftershock, occurred on 23 September, is $M_w = 6.1$, meaning that the magnitude difference with the mainshock is precisely $\delta M_w = 1$. There are hints at possible phenomena occurring before the mainshock as well. For example, the $M_w = 5.6$ seismic event on 18 September, the day before the mainshock, might be a foreshock candidate. However it raises questions on why the $M_w = 5.7$ event on 10 September did not lead to the same effect. Also we can observe fewer events in the period 13-17 September compared to the base rate on the period preceding the 13 of September, but the difference is not sufficiently clear to confirm seismic quiescence. It could, however, satisfy a precursory ETAS model, by considering the same type of clustering behaviour observed in the aftershock as also occurring in the foreshock region.

A time-series of Gutenberg-Richter's b -value can be built by creating a time-window $[t, t + t_d)$, with t_d in the order of days/weeks (sufficient for an average greater than 30 events within the window), and monitoring the time-series of seismic event magnitudes. For each point, the total number of events N within the window is recorded along with the magnitude M of each. Then the slope of $\log(N(M_w > M))$ as a function of M is assigned as the b -value for the time $t + t_d/2$ (the centre of the window). By moving the time-window by t_s , the calculation can be redone and a b -value is assigned for $t + t_d/2 + t_s$, and so on (Fig. 3.3). Therefore, a time-series of the Gutenberg-Richter's b -values is generated, enabling the verification of anomalies in this variable and whether they are related to mainshock occurrence. In our case, we have adapted the open-source Matlab toolbox GISMO [37], which facilitates the computation of a single b -value for a whole set, to: recursively request seismic occurrence data from IRIS catalogue within $t_d = 14$ -day intervals, calculate the b -value for the intervals, store the result together with a timestamp, and shift the time by $t_s = 1$ day, restarting the process until $t = t_{end}$. The sliding window width of 14-days was chosen as it was the highest resolution allowing for a ± 0.1 error bar in the b -value calculation for the area of interest.

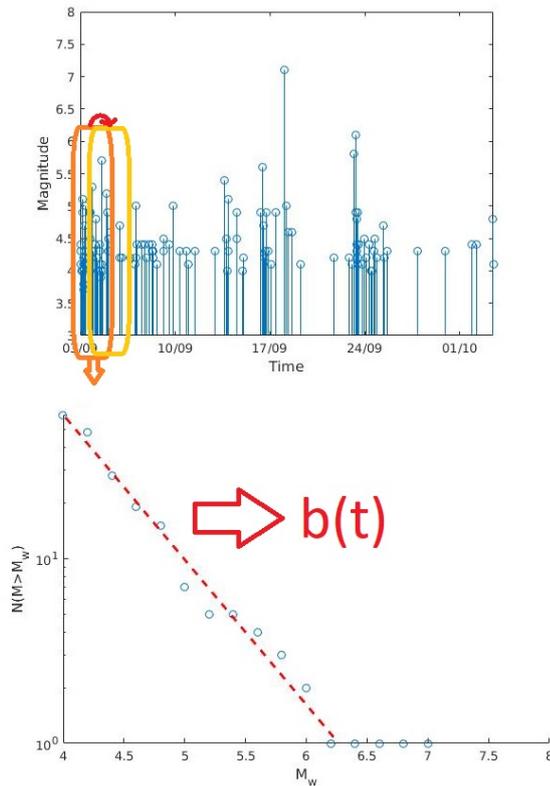


Figure 3.3: Diagram of the process for generating a point of the time-series of the Gutenberg-Richter's b-value. By moving the sliding window to the yellow square interval the next point of b-value is generated.

3.1.2 Gravity data

The surface gravity data is obtained from the International Geodynamics and Earth Tide Service (IGETS) database¹. The data we have used are time-series with a 1-min sampling and manually corrected by the operators to account for unintended changes, such as those caused by equipment maintenance (CORMIN data option). IGETS makes the data available in monthly text files starting with the calibration parameters and instrument location in the header and proceeding with the minute-sampled timestamp, gravity value and local atmospheric pressure in the body. As gaps are common, the original formatting divides the data into blocks, starting with the mark 77777777 and terminating with the mark 88888888. The end of the file is defined by a mark 99999999. In order to work with a longer time-series, not only the files need to be imported and joined together to obtain a longer time-series, but the gaps have to be correctly identified and filled with a null mark,

¹Freely available for research use at <https://isd.c.gfz-potsdam.de/igets-data-base/>

so to later perform calculations only on one single time-series block containing the full desired timespan.

On any point on Earth it can be observed an amplitude variation of surface gravity on the the order of $\pm 1000 \text{ nm s}^{-2}$. Most of the amplitude is due to the tides on Earth, particularly the solid Earth tides. Although the ocean tides typically reach meters height compared to centimetres height in the ground surface case, the solid Earth tides account for greater mass displacement, leading to higher gravity variations. These tides occur at a composition of frequencies clustered in four groups: long-period, diurnal, semidiurnal and terdiurnal.

Fig. 3.4 (a) reveals a gravity signal collected from a superconducting gravimeter in South Africa on August 2010. Analysing the frequency spectrum of this series, we observe the main tidal modes described previously (Fig. 3.4 (b)). It is necessary to filter out these tides if we intend to observe other variations, such as co-seismic gravity changes, which are in the order of $1\text{-}100 \text{ nm s}^{-2}$, considering a distance of $30\text{-}100\text{ km}$ of the nearest station to the epicentre. We also expect any pre-seismic gravity changes to be in the order of nm s^{-2} for a station distant from source, and possibly $\mu\text{m s}^{-2}$ for stations near the epicentre. When comparing the gravity time-series with the time-series of seismic events, we sample the time-series of seismic events in minutes.

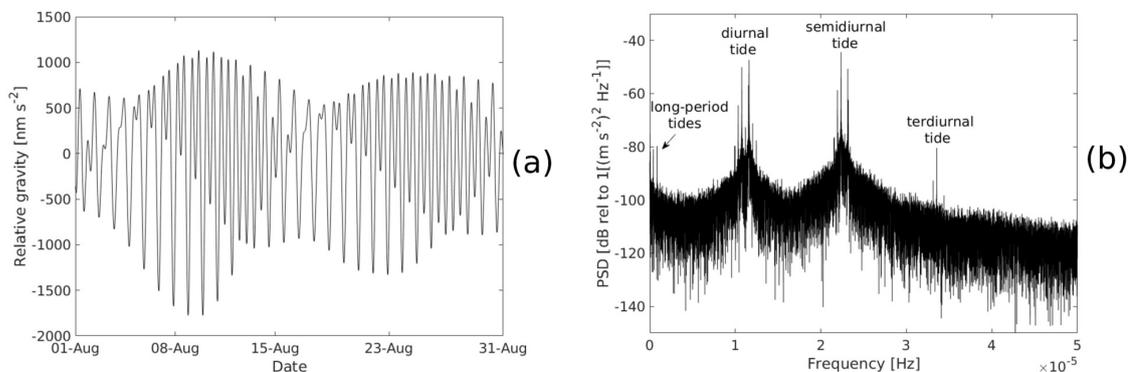


Figure 3.4: Example of gravity signal from the SU3 instrument, in South Africa. Left panel (a): 1 month sample (August 2010) of the relative gravity measurements. Right panel (b): Power spectrum density of the entire time-series. Long-period, diurnal, semi-diurnal, and terdiurnal tides are shown. Published by the author on [13].

3.2 Data pre-processing

3.2.1 Tidal analysis and removal

The Earth deforms as the Moon and the Sun exert slightly different magnitudes of gravitational attraction to different parts of the planet. The total acceleration felt on a point of Earth is also added by consideration on the Earth rotational movement (centrifugal force), the ellipsoidal orbit of the Moon, the planet ellipsoidal orbit around the Sun, and even the precession of the equinoxes (generating an apparent wobble on the polar caps), and this leading to variations in the measured gravitational acceleration on a Earth surface point at different time-scales. This is reflected in Fig. 3.4, where not only can be seen the semi-diurnal and diurnal gravity variations from the Moon and Sun pull at the time interval related to Earth's rotation, but also an envelope of around 14 days period due to the lunar orbit. The maximal amplitudes coincide with the perigee and the minimal with the apogee of the lunar trajectory. When analysing the time-series from the gravimeters or from strainmeters or tiltmeters, signals will also be dominated by tides. These are predicted by calculations based on the astronomical positions (*ephemerides*, provided by NASA-JPL² among others), defining tables of tidal frequencies, amplitude and phases, such as Cartwright and Eden [96], with 505 frequencies, Tamura [97], with 1200, Hartmann and Wenzel [98], with 12935 frequencies (including effect of attraction from other planets until Saturn), and others. Although these tables provide a guidance on the expected tidal frequencies, the observed tides depend on a number of regional considerations, from the geomorphology of the area to weather and climate factors. As the highest tides might be potential earthquake triggers, location-specific solutions for predictions of a next high tide are of interest also to seismology. On the other hand, if the intention is to investigate small gravity precursors, such as pre-seismic gravity changes, the tides have to be removed from the original signal, as these gravity changes are expected to have amplitudes up to three orders of magnitude smaller than the gravity tides. Three different methods are typically adopted for the purpose of filtering the measured signal, eliminating the tidal contribution and obtaining a residual reflecting the response

²https://ssd.jpl.nasa.gov/?planet_eph_export

of the Earth to other geological phenomena:

1. Deleting or damping the tidal frequencies
2. Modelling each physical contribution leading to the gravity signal
3. Fitting the data itself to a tidal model and subtracting the observation from the model

The deletion or dampening is typically done by FFT (Fast Fourier Transform) filtering or FIR (Finite Impulse Response) filtering, respectively. FFT filtering consists in: (1) applying the Fast Fourier Transform (method of fast calculation of Eq. 3.1) to the time-series, which converts it to the equivalent (frequency) series $X(n)$; (2) setting to zero the frequencies near the ones from the tidal tables; and (3) recovering the filtered time-series by applying the Inverse Fast Fourier Transform

$$X_n = \sum_{k=0}^{N-1} x_k e^{-i(2\pi kn/N)}, \quad n=0,1,\dots,N-1. \quad (3.1)$$

Any gaps must be interpolated or filled from theoretical estimates or the FFT cannot be computed. For small gaps (minutes-hours), linear interpolation is applied. Otherwise, the theoretical tides are calculated from the astronomical predictions and fitted to the time-series. After tidal removal, the interpolated sections are removed from residuals too, eliminating the artificial data. FIR filtering consists of a similar principle, based on deleting the undesired frequency bands, but instead do so by the development of a multiband notch filter, which, once implemented, is applied to the signal. This filter design is performed by Matlab in-built routine *fir2*, where the inputs are the desired frequency bands for dampening (regions centred at tidal table frequencies) and the filter order. A filter order too low leads to dampening being insufficient to remove the tides, and order too high creates distortions, especially at higher frequencies³. In practice, the choice of the appropriate order for a particular station is a recursive process of building a filter and analysing the output spectra.

³Equivalent to the misfitting that occurs when a polynomial of higher order than the true value is used to fit a curve

Another approach is to model each physical contribution to the gravity time-series, from theory, based on astronomical positions and observation of Earth-bound processes (global and regional weather, ocean changes, level of moisture in the soil, relative movement of polar ice caps). On Fig. 3.5 we show the modelling of each component until reducing the signal to a residual with tidal amplitude about 15 times smaller than the original series, for a 1-month gravity data sample in New Mexico. For the simulation of the solid Earth tide and ocean tidal loading it was adopted the software Atlantida3.1 [99] assuming theoretical tidal components from the Tamura (1200-frequencies) tables [97], layered Earth model IASP91 [28, 30], and ocean model FES2012 [100]. The gravity potential for the solid Earth solution consists in computing

$$V_{\text{Solid Earth}}(\mathbf{r}, t) = g_{\text{equator}} \operatorname{Re} \left[\sum_{n=2}^{\infty} \sum_{m=1}^n c_{nm}^*(t) Y_{lm}(\theta_1, \theta_2) \right], \quad (3.2)$$

where c_{nm}^* is the set of complex amplitudes tabled and $Y_{lm}(\theta_1, \theta_2)$ the spherical harmonics evaluated at the co-latitude θ_1 and longitude θ_2 of the station.

For the ocean loading the response of the ocean model to the tides is obtained and the gravity attraction to the ocean is defined by the potential

$$V_{\text{loading}}(\mathbf{r}, t) = \rho \iint_{\text{loading surface}} G(\mathbf{r} - \mathbf{r}') H(\mathbf{r}') dS, \quad (3.3)$$

where $G(\mathbf{x})$ is the Green's function and $H(\mathbf{r}')$ is the distribution of the loading element (*e.g.* ocean).

The gravity acceleration is given by the gradient of these potentials. For the atmospheric, non-tidal ocean and hydrology loading, it was used the numerical gravity solutions provided by EOST/University of Strasbourg⁴ for the IGETS stations. These solutions are computed essentially from the same procedure as described by Eq. 3.3, but the loading elements are the atmosphere (local (radius < 10°) and regional (10° < radius < 30°)), the ocean non-homogeneities (such as currents) and the contribution from the soil moisture level. We have selected the results using the atmospheric observations from

⁴<http://loading.u-strasbourg.fr/>

ERA-INTERIM program⁵ sampled at 6h and ocean bottom pressure from ECCO2 model⁶, interpolating the results as necessary.

Finally, we have calculated the contribution from the polar wobble from $\delta g_{polar} = -39 \cdot 10^6 (\sin 2\theta_1 (m_1 \cos \theta_2 + m_2 \cos \theta_2))$, using EOPC04 data⁷ for geolocation coordinate parameters m_1 and m_2 of the polar cap. The residuals of this sample are not sufficiently reduced to observe seismic precursors, but other effects in the same scale could potentially be verified, such as co-seismic gravity changes close to epicentre.

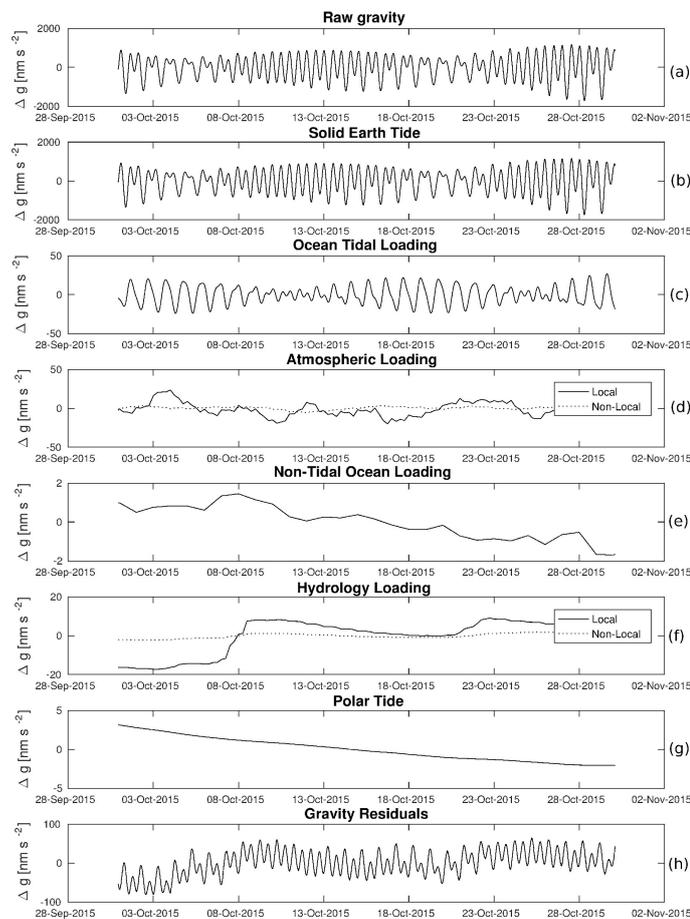


Figure 3.5: Example of physical modelling applied to 1-month data sample (October 2015) of AP instrument (New Mexico). The gravity residuals (h) are obtained by subtracting the measured relative gravity (a) by all the following simulated contributions: solid Earth tide (b), ocean tidal loading (c), atmospheric loading (d), non-tidal ocean loading (e), hydrology loading (f), and polar tide (g). In this sample, there was a misfitting of the amplitudes of the theoretical semidiurnal tides with the observations, so oscillations are still present in the gravity residuals (h), albeit with smaller amplitude than in the original time-series (a). Published by the author on [13].

⁵<http://www.ecmwf.int/en/research/climate-reanalysis/era-interim>

⁶<http://ecco2.jpl.nasa.gov/>

⁷<ftp://hpiers.obspr.fr/iers/eop/eopc04>

Finally, the last tidal removal option is the data-based modelling, which is the preferred method in the geodesy and gravimetry community. It consists in adjusting a tidal model creating a solution specific to the station location, by analysis of the past record of observations. Due to the tides being highly periodic, the most efficient model remains a classical harmonics solution, meaning the expansion of the tidal signal in well-defined frequencies in expressions of general form $V(t_j) = A_0 + \sum_{i=1}^n [A_{1,i} \cos(b_i t_j + \phi_i) + A_{2,i} \sin(b_i t_j + \phi_i)]$. There are different practical software implementations specialised for gravity tides, such as the open source programs ETERNA [101], VAV [102], BAYTAP [103] and UTide [104]. Each implementation re-writes the classical harmonics expression in a different form either to better visualize each tidal group (diurnal, semi-diurnal, long-period), or to perform the calculation in a more efficient way, but the task is common: find the set of parameters that adjusts the theoretical tidal table (such as from Tamura) to the observed dataset. UTide for Matlab is adopted in this study, due to its simplicity of use and equally efficient performance compared to the other options. After adjusting the ideal tidal parameters for the location of the station, we reconstruct, from the model, the tidal contribution from the same period of the dataset. The difference of the reconstructed time-series and the observation is the gravity residuals. In principle, the gravity residuals should present the non-tidal gravity events, i.e., non-periodic signals such as pre-seismic gravity changes should be observable.

However, it is not *a priori* clear which method performs best. All methods will still present some degree of tidal oscillations. The data-based method from ordinary least squares generates gravity residuals with amplitudes of about half the amplitudes from the physical modelling method. On the other hand, from the data-based method has the drawback of losing the connection with the physical process guiding each component, making hard to have control over which variables to preserve or discard and how to determine if the signal is under or overfitted. The frequency filtering methods may provide the smallest amplitude, but they completely remove information contained the selected frequencies, meaning that if a non-tidal event has a contribution in this region, this is lost from the residuals. We discuss this in greater detail in [13].

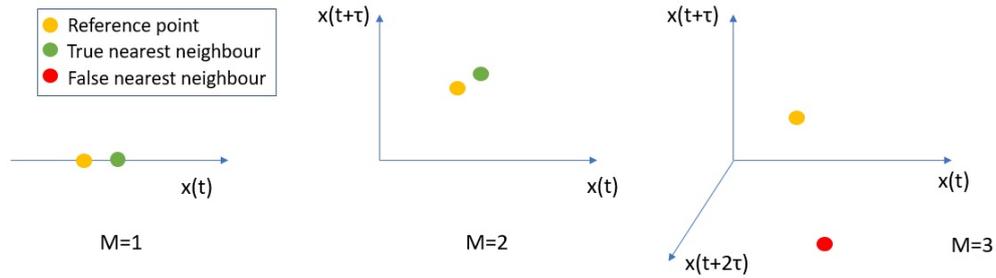


Figure 3.6: Simplified diagram of the false nearest neighbours method. The identification of the false neighbours is done for all points and the fraction of false neighbours relative to true neighbours is used to determine the true dimensionality M of the system.

The attractor from each time-series can be reconstructed using the Takens' method of delays [105], consisting in creating a state-space plot where one axis is the time-series values and each other axis is the same time-series points delayed by $k\tau$, $k=1,2,\dots,M$, with M the system embedding dimension. The delay τ adopted for the reconstruction was the first minima of the auto-mutual information, which is 3.6h for the gravity signals. Also, the embedding dimension, which is the minimum number of axis reconstructing the whole dynamics, is 4 for this systems. This is obtained from the *false nearest neighbours* method [106]. It consists in verifying for a given time the distance to the nearest neighbour of a point. If in the next iterate this distance is too high, the neighbour is false. Too large fraction of false neighbours means the state-space constructed for the M of the hypothesis is actually of lower dimension with respect to the true system – a projection. As long as the system is an attractor with relatively low noise, increasing M would eventually lead to a fraction of false nearest neighbours near zero, which is the true embedding dimension. Figure 3.6 presents a diagram clarifying the method.

The values obtained for τ and M are compatible with previous results from ocean tidal heights [107]. Although $M=4$ can be regarded a low-dimensional system, We still cannot observe the state-space directly. As an alternative, we define an hyperplane given by average point of the fourth component of the attractor, and observed the intersection the attractor of each gravity time-series (original and residuals) with this surface (Fig. 3.7). We also include for comparison a model from 1200 frequencies corresponding to the tides and a correlated (f^{-2}) noise profile, regarded as a common background behaviour in geophysical signals.

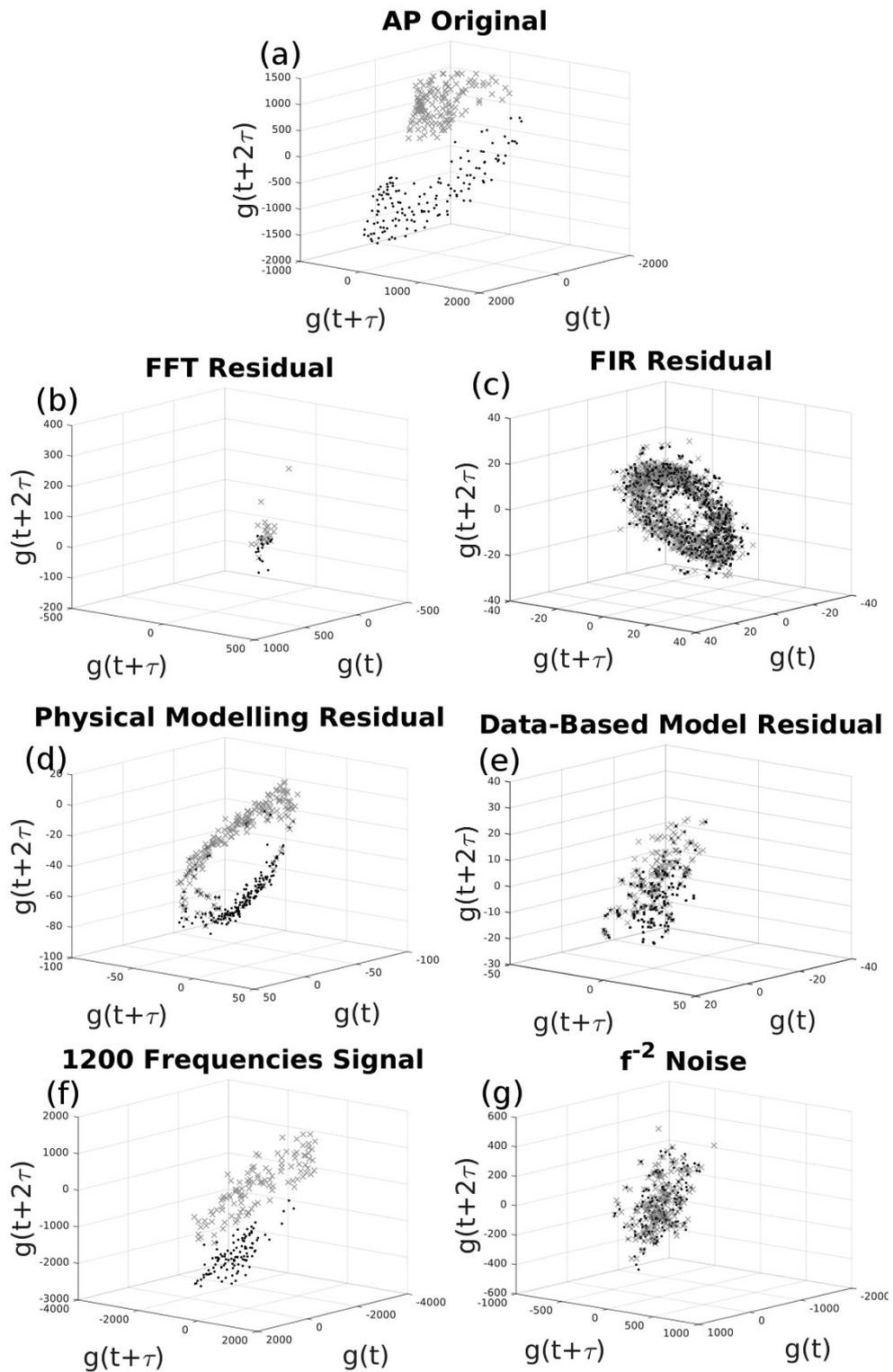


Figure 3.7: Map of the intersection of the trajectory with the surface $x_4 = \langle x_4 \rangle$ for the original gravity signal of AP station (a), gravity residuals obtained from the different methods (b)–(e), artificial signal generated by 1200 sinusoids corresponding to the tidal frequencies (f), and artificial red-noise (g). Similar patterns are observed for the other stations. The selected section is the one with fourth dimension equal to the mean fourth coordinate of the reconstructed attractor. Black dots are the intersections of the attractor trajectory with the surface in the direction of increasing x_4 , and gray crosses are intersections in the direction of decreasing x_4 . Units: nm s^{-2} . Published by the author on [13].

Qualitatively, it can be observed that the data-based residual is similar to the correlated noise. The physical modelling residual seems to retain the characteristic helicoidal shape of the original signal, but also has a spatial separation of the two directions of intersection of the hyperplane similar to the combination of periodic signals, further confirming tides are still present and relevant to this signal. The FFT and FIR residual completely modified the attractor setup, transforming it to a pattern similar from observed on Levy flights and the torus solution.

Observing more carefully the direction which the points cross the hyperplane defined by the average value in the fourth dimension (*i.e.* the plane $x_4 = \langle g(t + 3\tau) \rangle$), by assigning a black dots in direction of increasing x_4 and grey crosses in direction of decreasing x_4 , different patterns can be observed in each case. The original attractor (Fig. 3.7 (a)) exhibits a process where the trajectories from the upper part of the helicoidal structure are mapped on the lower band. These do not happen in the frequency filtered (either by FFT or FIR) signals in Fig. 3.7 (b) and (c). The residuals from physical modelling, though, preserves this structure. The residuals from data-based modelling, although exhibiting a tendency towards maintaining this division, is less evident than in the physical modelling case.

Finally, we have calculated the largest Lyapunov exponent for each gravity residual and observation, comparing also across different time-series from other IGETS stations. The largest Lyapunov exponent is the exponential rate of separation of initial neighbouring trajectories in the embedded space, and as a general rule, it is positive for cases sensitive to initial conditions (suggesting chaos), negative for periodic stable orbits and zero for limit-cycle conditions. For this analysis we have build a Matlab routine based on Rosenstein algorithm for calculation of the largest Lyapunov exponent [108]. We had to reselect the embedding time delay to 30 minutes, as 3.6h was empirically observed to be too large to observe a exponential behaviour, but the method is robust to variations in τ . For improved statistics we have averaged the results over 2000 initial points, with the largest Lyapunov exponents on gravity observations and gravity residuals shown on Table 3.1.

Table 3.1: Largest Lyapunov exponent [bits h⁻¹] of original gravity time-series and residuals after filtering for all stations.

Station	Original series	FFT res.	FIR res.	Phys. Model. res.	Data-based res.
AP	1.412 ± 0.007	0.349 ± 0.007	0.77 ± 0.03	0.93 ± 0.02	0.63 ± 0.02
BF	1.208 ± 0.005	0.29 ± 0.01	0.53 ± 0.03	0.84 ± 0.01	0.26 ± 0.01
MA	1.317 ± 0.006	0.80 ± 0.03	0.76 ± 0.02	0.73 ± 0.01	0.42 ± 0.02
NY	0.860 ± 0.009	0.35 ± 0.04	0.41 ± 0.03	0.81 ± 0.02	0.61 ± 0.02
SU	1.326 ± 0.007	0.68 ± 0.04	0.58 ± 0.03	0.90 ± 0.02	0.84 ± 0.02

Note: Published by the authors on [13].

Again, this is in accordance to previous observation in shallow water ocean levels of 0.57-4.54 bits h⁻¹. The values suggest that the Earth responds to the tides by inheriting a small sensitivity to the initial conditions, thus enhancing the oscillations promoted by tides instead of damping them. The sensitivity to the initial conditions behaviour is maintained in the residuals, but there is a reduction of the value. As the Lyapunov exponent relates to the entropy of the signal, this reduction indicates that the residuals are less entropic than the original signal, as expected. Physical modelling residuals presented Lyapunov exponents closer to the original time-series, while other methods presented larger reductions but also greater differences between the stations.

3.2.2 Symbolic encoding and partitioning

The information-theoretical procedures we are going to apply involve the estimation of probabilities of measurements in a given variable. For sufficiently long time-series these probabilities can be obtained by the procedure of box-binning, which consists of splitting the domain of the variable into smaller ranges and counting the number of elements that fall in each range. For example, for a variable $X \in \mathcal{X} = [x_{min}, x_{max}]$, we can split the domain in n_b bins or subdomains ($\mathcal{X} = \mathcal{X}_1 \cup \mathcal{X}_2 \cup \dots \cup \mathcal{X}_n$, where $\mathcal{X}_i \cap \mathcal{X}_j = \emptyset \forall i, j$) and count the number of points in each (N_i). By dividing this count by the total number of points on the time-series, we obtain the probability of a point x being in the subdomain i , *i.e.* $p_x(i) = N_i / \sum N_i$. The resolution in the probability distribution obtained increases in accordance to the number of bins selected. However, it must be safeguarded that the average number of points in the bins remain sufficiently high, otherwise the frequentist

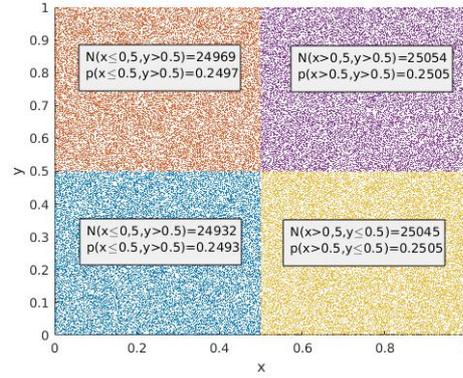


Figure 3.8: Example of calculation of joint probabilities via the binning method for a time-series of 10000 points.

definition of probability is invalid. For the case of joint probabilities $p(x \in i, y \in j)$, the procedure is analogous, but calculated over a grid where each axis is one of the variables (Fig. 3.8). With this, it is already possible to calculate informational quantities, such the Entropy of variable and the Mutual Information of two variables, which will be discussed in subsection 4.1. In this theses we will, however, analyse these quantities over symbolic sequences, which also reveal the dynamical processes of each point.

We monitor the dynamics of the points of the variable as they move from one bin to another with time by assigning to each data point a marker $\mathcal{S}_x(i)$ indicating the subdomain it belongs to. For convenience, if the number of bins (marginal partitions) n_b is 2, \mathcal{S} is binary (0,1), if $n_b = 3$, \mathcal{S} is ternary (0,1,2) and so on. In this thesis we will only work with binary partitions for simplicity, but our code allows for other partitioning. Increasing the number of bins increases the resolution of the analysis, however the computational requirements also increase. We observe, then, for each point in the time-series, what is the initial marker and what is the symbol at the next iterate, and so on. A new symbol can be created to monitor the evolution of each point, which could be given by $\phi_x(t = t') = \mathcal{S}(t = t_0)\mathcal{S}(t = t_0 + 1) \dots \mathcal{S}(t = t_0 + L) = '00 \dots 0', '00 \dots 1', \dots, \text{or } '11 \dots 1'$. For convenience, we transform the symbolic sequence into a natural number by

$$\phi_x(t = t') = \sum_{i=t_0}^{t_0+L} \mathcal{S}_x(t = i) n_b^{t'-i}. \quad (3.4)$$

This new symbol for monitoring the trajectory effectively means the creation of a

new partitioned space for the variable. This is exemplified in Fig. 3.9, showing the construction of partitions with different orders. Observe that the order of a partition is equivalent to the length of the symbolic sequences being used, *i.e.*, the number of points in the original space necessary to create one point in the partitioned space. The calculation of the information theoretical quantities described on chapter 4 are performed in the partitioned space.

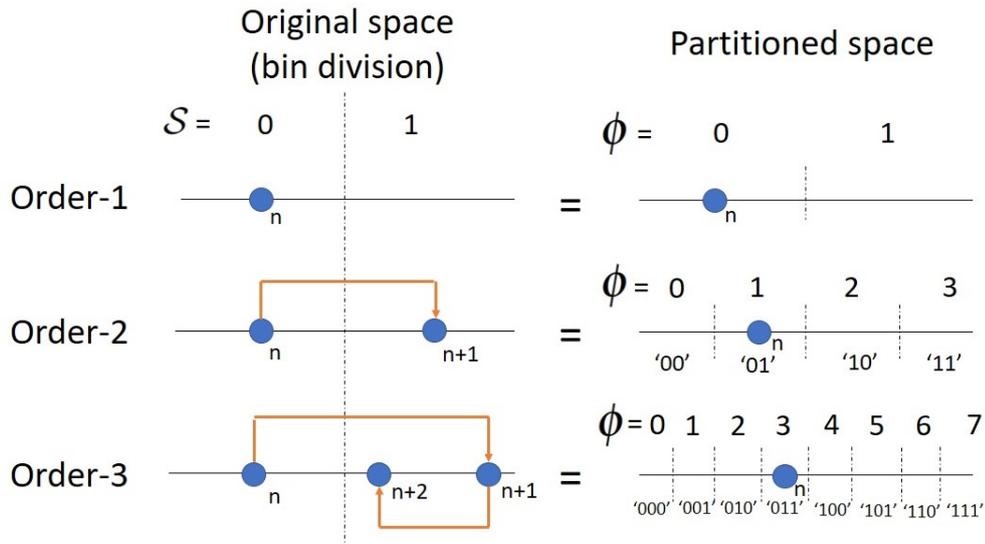


Figure 3.9: Example showing the symbolic encoding and the construction of the partitioned space up to the third order, for $n_b = 2$.

3.3 Test-bench dynamical system: Logistic map

A map is a discrete time system which evolves as $x_{n+1} = f(x_n)$, *i.e.* the next point of the iteration depends on the previous point. The logistic map is one of the classical examples, defined [109] by:

$$x_{n+1} = rx_n(1 - x_n). \tag{3.5}$$

This simple expression, originally used to describe evolution of insect population, reveals very distinct behaviour for varying free parameter r . A few examples of the time-series are displayed on Fig. 3.10. For small values of r , x quickly decays to zero. For r values between 1 and 3, x will stabilize at some value between 1 and 0.67. Above this, we might observe x oscillating between different values, or following complicated

trajectories. For $r > 4$, x diverges to $-\infty$, so we do not consider this option.

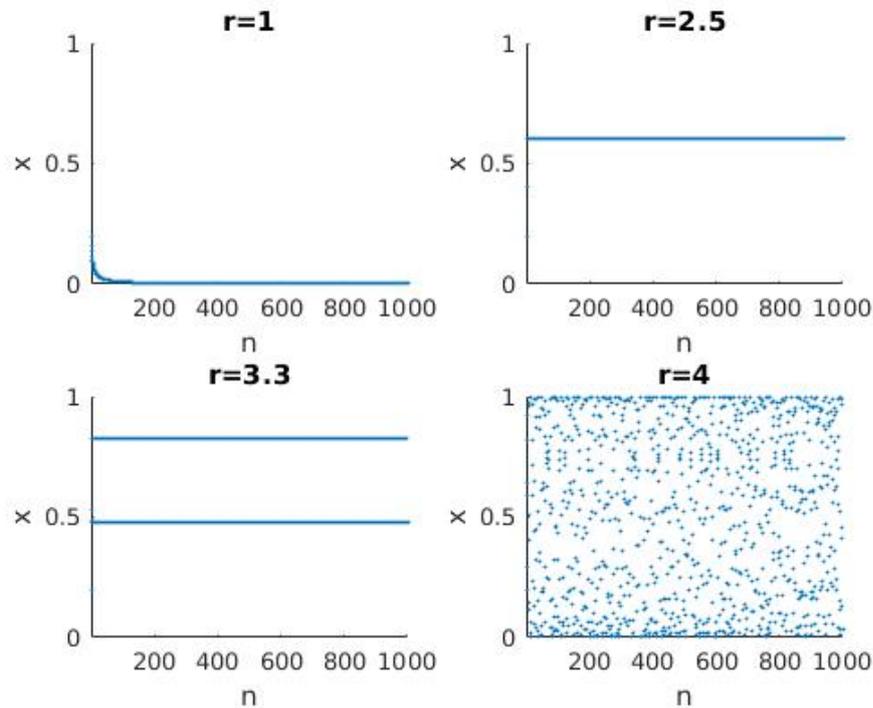
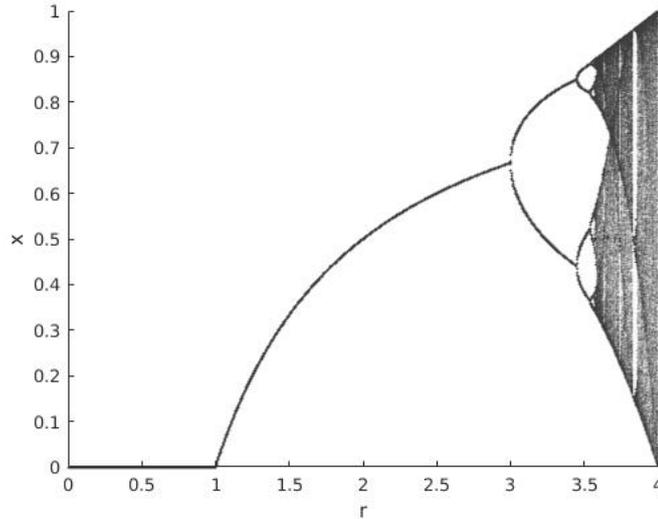


Figure 3.10: Examples of the time-series from the logistic map for different values of the free parameter r . We consider a initial condition $x_0 = 0.2$.

Plotting the possible values of x in function of the free parameter leads to a bifurcation diagram such as Fig. 3.11. We will split this in different regions of interest:

1. zero-value, for $r \in [0, 1]$
2. fixed-point or period-1, for $r \in (1, 3]$
3. period-doubling cascade, for $r \in [3, 3.57]$
4. chaotic region with embedded periodic windows, for $r \in (3.57, 4)$
5. full chaos, for $r = 4$

Some remarkable periodic windows are the period-3 in $r \in (3.83, 3.86)$ or, zooming in, the period-9 in $r \in (3.6871968, 3.6873708)$, among others. These windows occur after the attractor collides with an unstable fixed point or a periodic orbit, a condition referred as *crisis* [110]. This interval in particular is for the attractor meeting a solution of period-3. Such periodic solution is within the attraction basin, and additionally we can observe



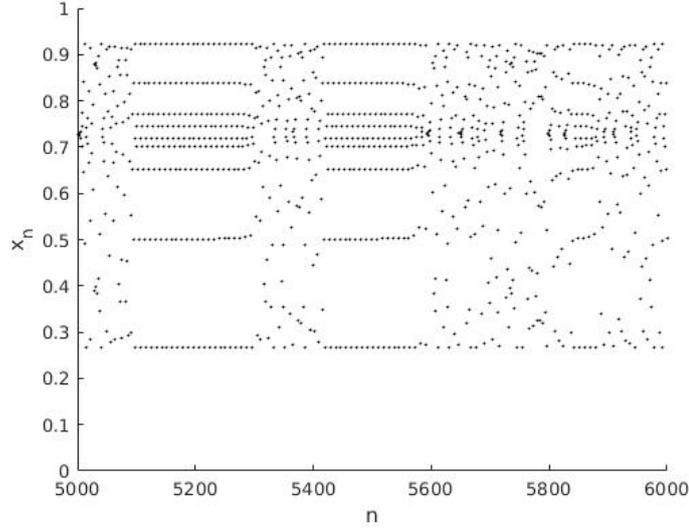


Figure 3.12: Time-series of the logistic map showing intermittency, for $r = 3.6871960$ and initial condition $x_0 = 0.2$. Note the time-series shifting between chaotic and periodic behaviour, despite the system not yet residing in the period-9 window.

but it is necessary to connect them. We consider two configurations for coupling the logistic systems. One is the linear diffusive coupling, expressed by

$$x_{n+1}^i = (1 - \sigma)f(x_n^i) + \frac{\sigma}{k_i} \sum_j A_{ji}(x_n^j - x_n^i), \quad (3.6)$$

where $f(x)$ is the logistic function, *i.e.* $f(x) = rx(1 - x)$.

The other is using a coupling term following the Kaneko proposal for coupled map lattices (CML) [112]:

$$x_{n+1}^i = (1 - \sigma)f(x_n^i) + \frac{\sigma}{k_i} \sum_j A_{ji}f(x_n^j). \quad (3.7)$$

The superscripts i, j inform which variable is being considered, and the matrix A is the adjacency⁸ matrix, with $A_{ji} = 1$ for a causal link $x_j \rightarrow x_i$ and zero otherwise. We will refer the system as a coupled logistic map if there are only two elements, meaning A is a 2x2 matrix, and a logistic network if the number of elements is greater. k_i is the incoming degree of the i -th node, *i.e.* 0 if there is no link pointing to i , 1 if there is one link, and so

⁸We adopt here the convention that the columns indicate the receiving ends of the edges. Although this convention might not be the tradition in dynamical systems, it is the standard of Matlab/Octave, which is why we adopt it.

on. By convention, if $k_i = 0$ the whole coupling term is assumed zero.

In Fig. 3.13 it is displayed the phase-space of the coupled logistic map systems of Eqs. 3.6 and 3.7 for $x^1 \rightarrow x^2$, $r = 4$ and increasing σ . Both coupling types show similarities, promoting synchronization as the coupling strength is increased. At $\sigma = 0.5$, the CML system exhibits synchronization of the type $x^1 = x^2$. The linear diffusive system progressively collapse its bands until forming a parabola shape of x^2 with respect to x^1 by $\sigma \approx 0.45$. By $\sigma = 0.5$ the solutions already collapsed, with x^2 diverging to $+\infty$.

We may recap the procedure of symbolic encoding and construction of the partitioned space (Sec. 3.2.2) and observe the location of the partition boxes on the coupled logistic map system. This is displayed on Fig. 3.14, for increasing symbolic length L . In this case we plot the coupled logistic map system for $x^1 \rightarrow x^2$, with both linear diffusive and CML coupling. It is observed that the partition division lines of the *cause* x^1 remain straight, while the lines of the *effects* x^2 start to bend for increasing L forming bubble-shaped features discussed in [113, 114]. By analysing the backward iteration of the division lines (at 0.5), the authors conclude that the formation of bubbles is intrinsically connected to the flow of information between the two variables. On Fig. 3.15, we plot the cases without a physical connection and with a physical connection in both directions ($x^1 \leftrightarrow x^2$). Again, it is observed bubble-shaped features for increasing L in the connected system, this time on both axis, whereas such features are absent when there is no link.

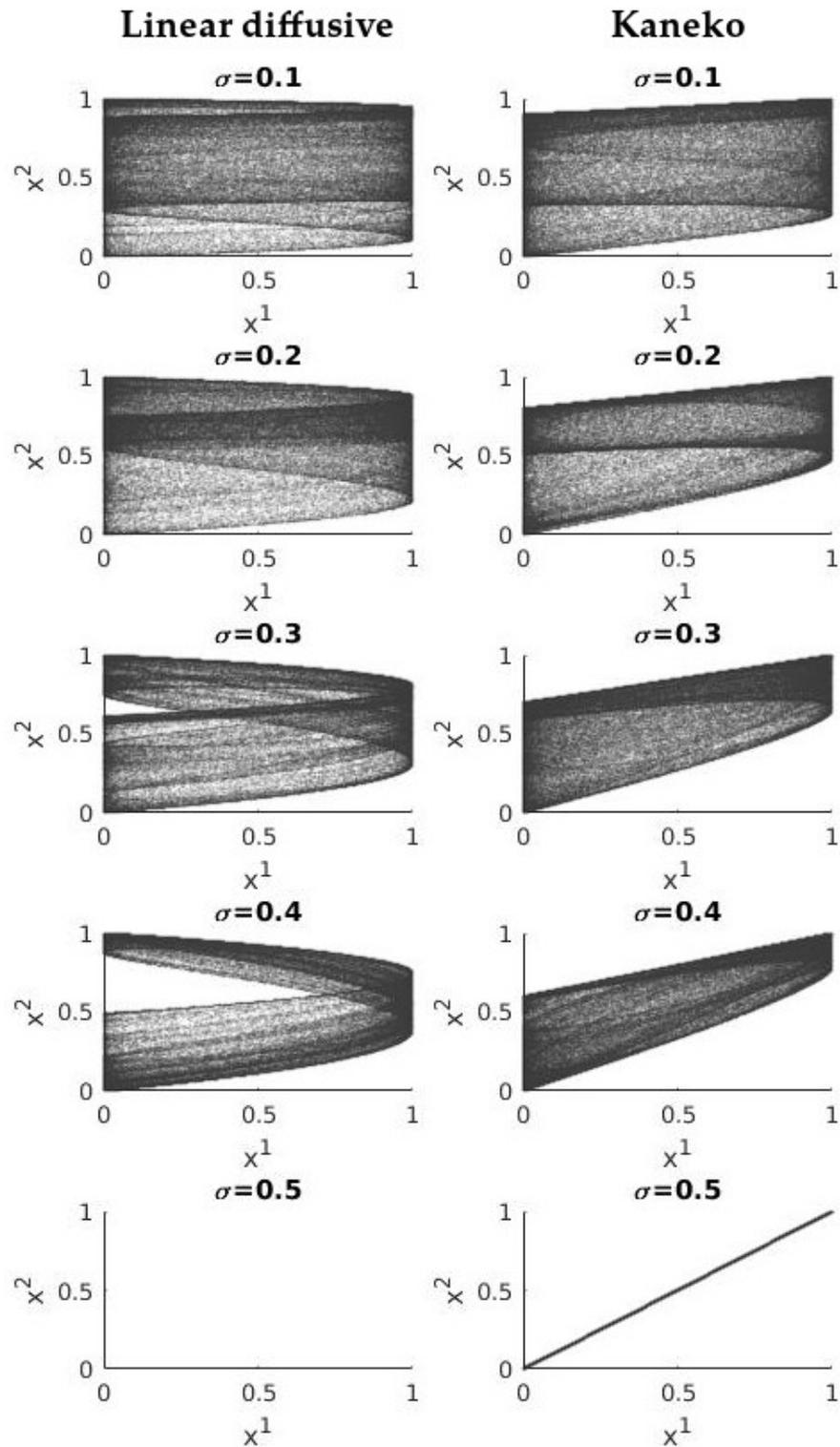


Figure 3.13: Phase-space plots with 100000 non-transient points of the coupled logistic maps $x^1 \rightarrow x^2$, for $r = 4$ and varying coupling strength σ . The first 5000 points of the time-series are discarded as possible transient. A transparency ($\alpha_{ch} = 0.05$) is added.

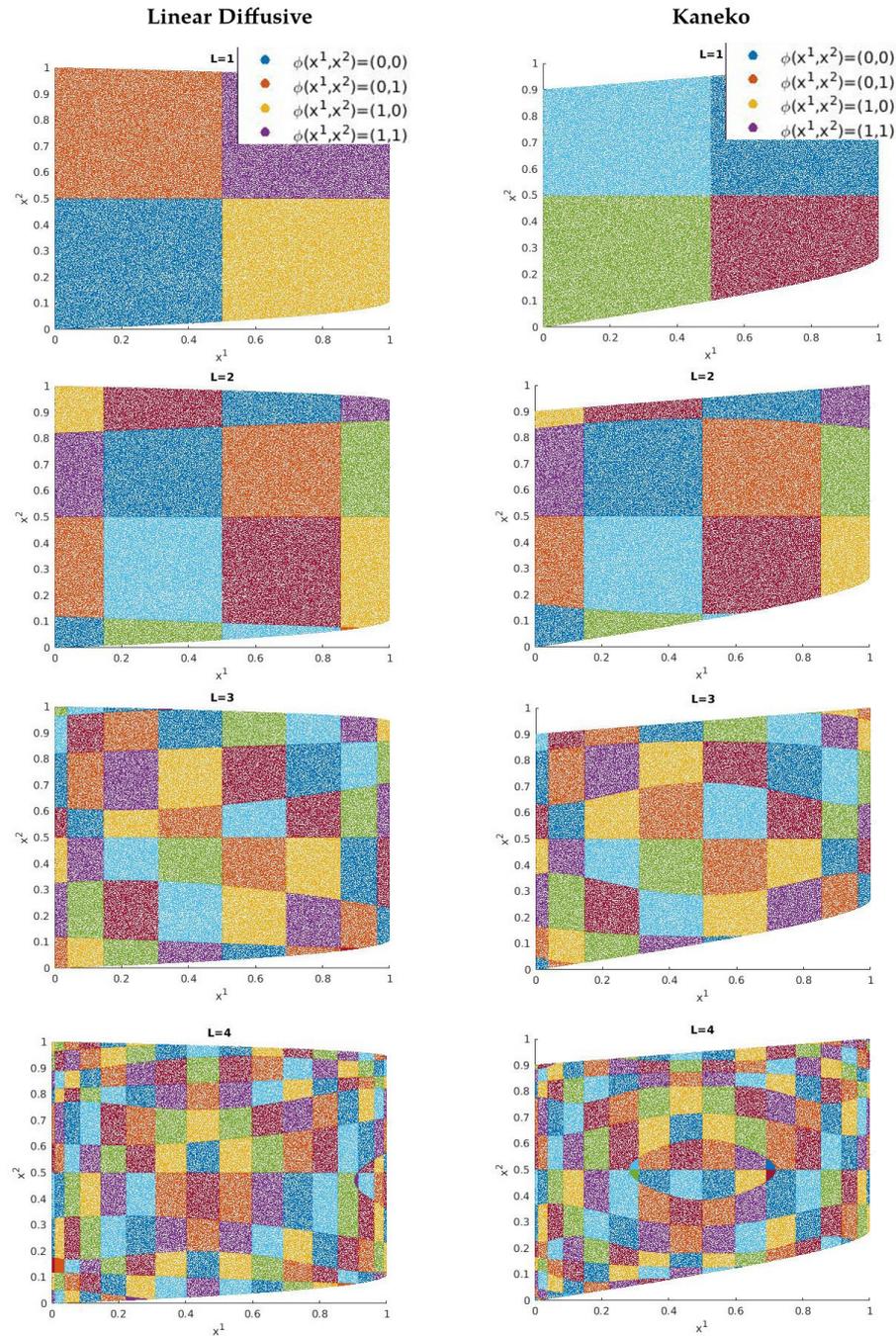


Figure 3.14: Phase-space plots of the coupled logistic map $x^1 \rightarrow x^2$ showing the location of the partitioned space for binary partition. The symbolic sequence splitting line is 0.5 on both axis. Each colour represents a value of the symbolic encoding $\phi(x, y)$. The system parameters are $r = 4$ and coupling strength $\sigma = 0.1$ and a total of 5000000 non-transient points are displayed. The first 5000 points of the time-series are discarded as possible transient.

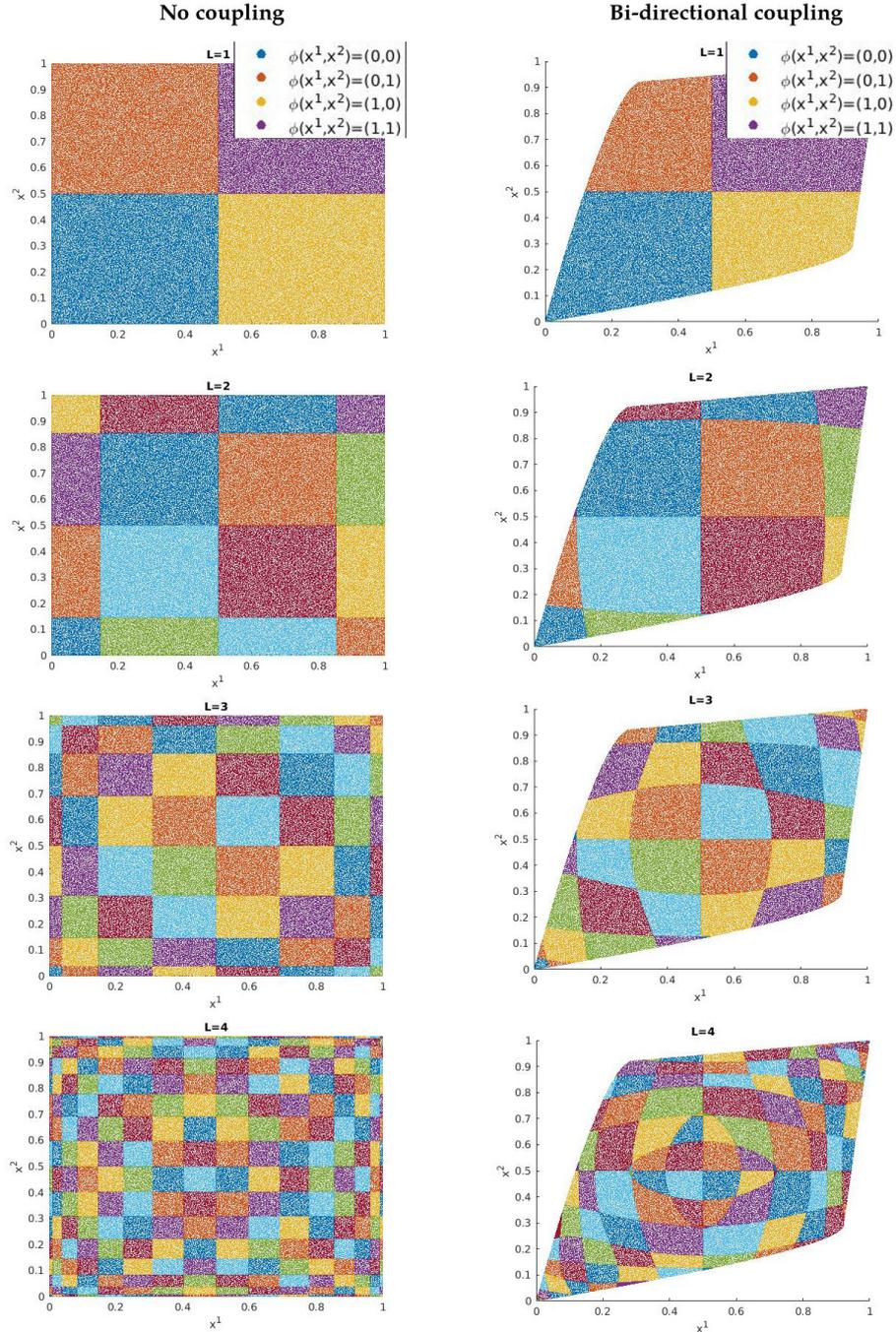


Figure 3.15: Phase-space plots for the system of Eq. 3.7 (logistic with CML coupling), for no coupling between x^1 and x^2 or for bi-directional coupling (*i.e.* $A_{ij} = 1$ for $i \neq j$ and 0 otherwise) with strength 0.1. Again, we show the partitioned space for binary partition, with symbolic sequence splitting line at 0.5 on both axis. Each colour represents a value of the symbolic encoding $\phi(x, y)$. The logistic free parameter is $r = 4$ and a total of 5000000 non-transient points are displayed. The first 5000 points of the time-series are discarded as possible transient.

3.3.2 Logistic networks

For more than two nodes, we refer to the system as a logistic network. This also follows the expressions of Eqs. 3.6 and 3.7 for the linear diffusive or CML coupled systems, but, by increasing the size of the adjacency matrix, more combinations of links can be explored. In particular we investigate the following options: a serial network (Fig. 3.16), a parallel network (Fig. 3.17) and a directed tree network (Fig. 3.18).



Figure 3.16: Example of serial network with 10 nodes

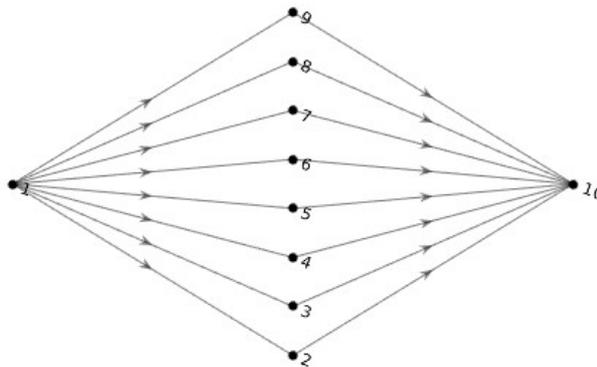


Figure 3.17: Example of parallel network with 10 nodes

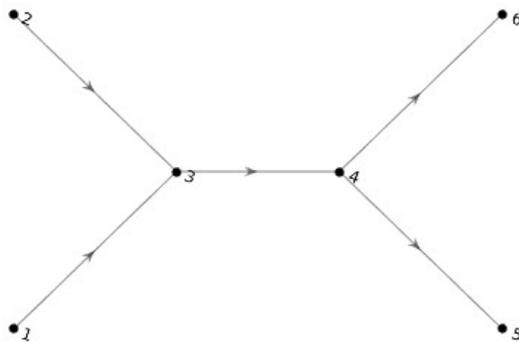


Figure 3.18: Example of a directed tree network with 6 nodes. In this example the extremities are at distance 3.

These three cases attempt to capture the local topologies found in real-world networks and that would constitute fundamental topologies to study how information decays in more larger networks in terms of topological invariants. This analysis is inspired in the

same strategy used by Kirchhoff to calculate equivalent resistances or equivalent capacitances in terms of simple topological rules (resistors in series or in parallel). We aim at understanding what would be these simple rules in regards to the decay of information. The serial case (Fig. 3.16) enables the understanding of how information decays as the topological distance between nodes increases. The parallel case (Fig. 3.17), permits to understand how the information decays or is intensified as information flows simultaneously by parallel paths. Topology in Fig. 3.18 is inspired in the way electricity is traced in power systems [115]. This topology is the simplest case from which one can trace the power flow departing generators and arriving at another nodes, representing, for example, electricity consumers.

Chapter 4

Information theoretical quantities and causality

In this chapter we introduce the theoretical basis for our method of calculating the flow of causal information from time-series, finishing with the description of the developed algorithm in Matlab. With this method we intend to define the arrow of causality between two physical variables, in particular of seismic occurrence with precursor candidates, with results presented in the next chapters.

4.1 Entropy and Mutual Information

Entropy in thermodynamic systems refers to a measure of the number of different microscopy configurations enabling the same macroscopic state. Thus, it can also be interpreted as a measure of how disordered a thermodynamical system is, *i.e.* how much knowledge do we have of the thermodynamic process on the microscopical level. Shannon [116, 117] extended the concept to communication systems, which later become a theory of information, by considering entropy as a measure of the uncertainty or unpredictability of a state. For a random variable X , the entropy is defined by

$$H(X) = - \sum_{i=1}^{n_b} p_x(i) \log p_x(i), \quad (4.1)$$

where $p_x(i) \in [0, 1]$ represents the probability of finding X in one of the possible n_b discrete states.

The logarithm can be taken with base 2 or e , providing a result in *bits* or *nats*, respectively. In this thesis we will adopt bits, but the developed codes enable both cases. The entropy of a fully predictable system (equivalent to a system presenting one state with probability 1 and the remaining with zero probability) is zero, whereas the entropy of a fair coin toss (two equiprobable states) is 1 bit. Shannon entropy is upper bound by $\log(n_b)$, where this bound is reached when the system has equal probabilities and equal to $p(i) = \frac{1}{n_b}$. Furthermore, entropy is always positive, as $p(i) \in [0, 1]$. The concept of entropy can be also applied to a partitioned system, to analyse the uncertainty over symbolic sequences with length L built from measurements of $x(t)$. In this case, the entropy is calculated over the L -order partitioned space, by using the boxes $\phi_x(i)$ representing the trajectory of the points. This way, we only rewrite Eq. 4.1 as

$$H(X_L) = - \sum_{\phi_x} p_x(\phi_x) \log p_x(\phi_x). \quad (4.2)$$

Note that $L = 1$ retrieves the previous case.

So far we have only defined entropy for a univariate system. For the investigation of bivariate systems, the concepts of joint and conditional probabilities can also be applied to define joint and conditional entropies, $H(X, Y)$ and $H(X|Y)$,

$$H(X_L, Y_L) = - \sum_{\phi_x, \phi_y} p(\phi_x, \phi_y) \log p(\phi_x, \phi_y), \quad (4.3)$$

$$\begin{aligned} H(Y_L|X_L) &= - \sum_{\phi_x} p(\phi_x) \sum_{\phi_y} p(\phi_y|\phi_x) \log p(\phi_y|\phi_x) \\ &= - \sum_{\phi_x, \phi_y} p(\phi_x, \phi_y) \log p(\phi_y|\phi_x), \end{aligned} \quad (4.4)$$

given for length- L symbolic sequences based on the variables X and Y . The joint entropy $H(X_L, Y_L)$ informs the overall uncertainty over the complete partitioned system with the two variables, while the conditional entropy $H(Y_L|X_L)$ reflects the uncertainty over the length- L sequences in Y when X is known.

An additional quantity is the Mutual Information shared by the two variables, defined by

$$\begin{aligned} I(X_L; Y_L) &= H(X_L) + H(Y_L) - H(X_L, Y_L), \\ &= \sum_{\phi_x, \phi_y} p(\phi_x, \phi_y) \log \frac{p(\phi_x, \phi_y)}{p(\phi_x)p(\phi_y)}. \end{aligned} \quad (4.5)$$

This is a measure of the amount that the uncertainty of a variable is reduced by knowing the other. Notice, though, that this is equivalent to the intuitive principle of the amount of information that a variable has about the another.

The mutual information is also a positive number, with a maximum value given for the identity case ($X_L = Y_L$), where it is equal to the Shannon entropy. When the knowledge of a variable cannot provide information about the current state of the other variable, the mutual information is zero. Furthermore, mutual information is symmetric, *i.e.*, $I(X;Y)=I(Y;X)$. The consequence is that, although it is a quantity identifying an information link between the two variable, it cannot be used to identify the direction of information flow. The relation between entropy, conditional and joint entropy, and mutual information is given by the Venn diagram of Fig. 4.1.

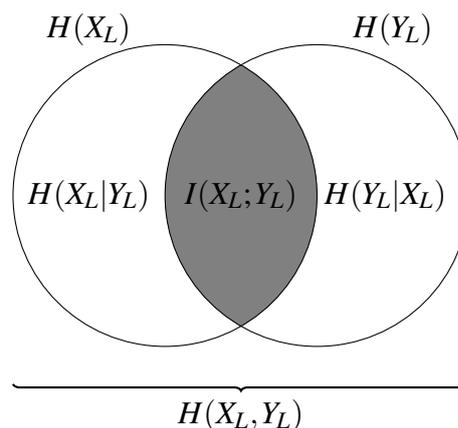


Figure 4.1: Venn diagram showing the entropies, conditional entropies, joint entropy and mutual information for the symbolic sequences X_L and Y_L .

Each quantity is represented by the area comprised by the circles or their intersections. The left circle expresses $H(X_L)$ and the right circle $H(Y_L)$, with their overlapping

grey area being the mutual information between both variables, and the remaining areas the conditional entropies.

4.2 Transfer Entropy

Transfer Entropy was defined by Schreiber [118] to address the question of the amount of information that is not simply mutual, correlated, but being actively exchanged, aiming to indicate the existence and direction of a causal link. For a suspected causation $X \rightarrow Y$, the principle of transfer entropy is to inform how the uncertainty on the future of a variable Y is reduced by knowing the past of X also given the past values of Y . This can be alternatively interpreted as the amount of information contained in the future of Y in relation to the past of X given a known past of Y . As a short-hand notation, we write the length- L sequence of the past of X as ϕ_x^{-t} , the length- L of the past of Y as ϕ_y^{-t} and the length- L of the future of Y as ϕ_y^{+t} . The selection of a sequence representing the past or the future of a time-series is simply done by properly defining the starting time t_0 in the expression $\phi_x(t = t') = \sum_{i=t_0}^{t_0+L} \mathcal{S}_x(t = i) n_s^{t'-(i+1)}$ that builds the length- L symbolic sequence. For example, the past sequence could be given by setting $t_0 = t - L$ and the future by selecting $t_0 = t + 1$. With this notation, the Transfer Entropy from variable X to variable Y can be written as

$$\begin{aligned}
 TE_{X_L \rightarrow Y_L} &= \sum_{\phi_x^{-t}, \phi_y^{-t}, \phi_y^{+t}} p(\phi_x^{-t}, \phi_y^{-t}, \phi_y^{+t}) \log \frac{p(\phi_y^{+t} | (\phi_x^{-t}, \phi_y^{-t}))}{p(\phi_y^{+t} | \phi_y^{-t})}, \\
 &= I(Y_L^{+t}; (X_L^{-t} | Y_L^{-t})), \\
 &= H(Y_L^{+t} | Y_L^{-t}) - H(Y_L^{+t} | (X_L^{-t}, Y_L^{-t})).
 \end{aligned} \tag{4.6}$$

This method was successfully applied in recent years to the investigation of the arrow of causation from time-series in many fields, from Neurology [119] to Econometrics [120] and to Social Systems [121]. In particular, transfer entropy generalises the Granger causality problem [11], of defining the arrow of causation as satisfying the properties of:

cause must happen before the effect, and cause has a unique information about the effect. In particular, it has been shown even to reduce to the Granger method, based on autoregressive modelling¹, for special cases [122].

The conditional probabilities inside the logarithm are responsible for breaking the symmetry between X and Y, hence enabling the determination of the direction of information. However, this comes with a cost: the calculation of conditional probabilities are typically computationally intensive.

4.3 Causal Mutual Information (CaMI)

The Causal Mutual Information (CaMI) is a quantity defined in terms of the physical notion that if X send causal information to Y, then, longer time observations in the variable Y should allow the reduction of the uncertainty about the past states of the variable X [113, 114]. It means, longer observations in Y should allow prediction of the state of the variable X. This measure is defined by

$$\begin{aligned} CaMI_{X_L \rightarrow Y_L} &= \sum_{\phi_x^{-t}, \phi_y^{-t}, \phi_y^{+t}} p(\phi_x^{-t}, \phi_y^{-t}, \phi_y^{+t}) \log \frac{p(\phi_x^{-t}, \phi_y^{-t}, \phi_y^{+t})}{p(\phi_x^{-t})p(\phi_y^{-t}, \phi_y^{+t})}, \\ &= I(X_L^{-t}; (Y_L^{+t}, Y_L^{-t})). \end{aligned} \quad (4.7)$$

The expression evaluates the ability of (the joint) future and past points of Y predicting the past of X. If such ability exists, it is because X causes Y or because they share common (non-causal) information, or both.

The Causal Mutual Information is connected with the Transfer Entropy, as demonstrated:

¹Granger method is based on affirming that X *Granger-causes* Y if an autoregressive model of Y including data from the past of X (i.e., $y_{n+1} = \sum_{i=n-\tau^*}^n a(i)y_i + b(i)x_i$, τ^* the past time considered) fits better to the future of Y than an autoregressive model with data from the past of Y alone (null hypothesis, $y_{n+1} = \sum_{i=n-\tau^*}^n a(i)y_i$).

$$\begin{aligned}
CaMI_{X_L \rightarrow Y_L} &= I(X_L^{-t}; (Y_L^{+t}, Y_L^{-t})), \\
&= H(X_L^{-t}) + H(Y_L^{+t}, Y_L^{-t}) - H(X_L^{-t}, Y_L^{+t}, Y_L^{-t}), \\
&= H(X_L^{-t}) + [H(Y_L^{-t}) + H(Y_L^{+t} | Y_L^{-t})] - [H(X_L^{-t}, Y_L^{-t}) + H(Y_L^{+t} | (X_L^{-t}, Y_L^{-t}))], \\
&= I(X_L^{-t}; Y_L^{-t}) + TE_{X_L \rightarrow Y_L}, \\
&= I(X_L; Y_L) + TE_{X_L \rightarrow Y_L}.
\end{aligned} \tag{4.8}$$

This supports the initial concept that the ability of predicting the past of X from Y is associated with common information and causation from X to Y. The property $H(X, Y) = H(Y) + H(X|Y) = H(X) + H(Y|X)$ is used between the second and third step to rewrite joint into conditional entropies. In a Venn diagram, the relation between CaMI and Transfer Entropy can be expressed as Fig. 4.2.

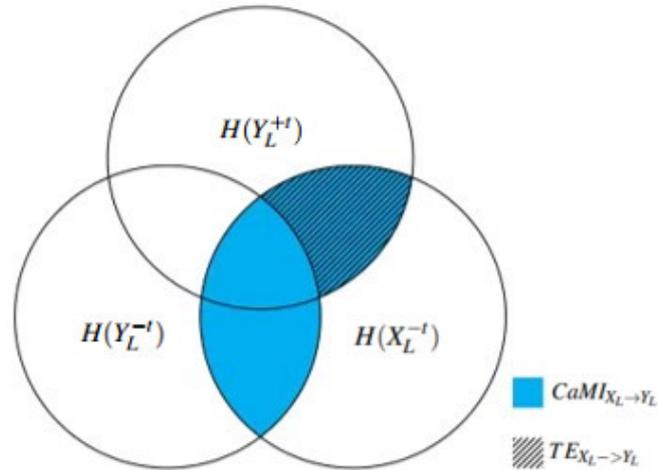


Figure 4.2: Venn diagram of the entropies of the symbolic sequences of past of X and Y and the future of Y, showing the Causal Mutual Information and the Transfer Entropy. The fraction of CaMI which is disjoint of the Transfer Entropy is the Mutual Information $I(X_L; Y_L)$.

Both the mutual information and the causal mutual information involve only joint probabilities, which is easily calculated by the box-binning procedure. Hence it becomes practical to calculate the transfer entropy from these two quantities instead of the original definition. Furthermore, the difference of transfer entropy from X to Y can be solely

calculated from CaMI, as Eq. 4.9. We refer to this net flow of causal information as a Directionality Index (*DirIdx*), once its value indicates the direction of the causality arrow.

$$\begin{aligned}
 DirIdx_{(L)} &= TE_{X_L \rightarrow Y_L} - TE_{Y_L \rightarrow X_L}, \\
 &= CaMI_{X_L \rightarrow Y_L} - I(X_L; Y_L) - CaMI_{Y_L \rightarrow X_L} + I(Y_L; X_L), \\
 &= CaMI_{X_L \rightarrow Y_L} - CaMI_{Y_L \rightarrow X_L}.
 \end{aligned} \tag{4.9}$$

Consider now the case of a small network, where the nodes are dynamical systems producing a measurable time-series and the links are the couplings between each system, as arrows of causation. In this network, the overall structure can in principle be inferred from the time-series of the nodes by evaluating the directionality index in time-series produced by each pair. For that, we use an intuitive notion that if two nodes are directed connected, the direction of their underlying physical connection is responsible to orient the flow of information along the same direction. If the first node events causes events in the second node, $DirIdx > 0$, whereas, in the opposite case, $DirIdx < 0$. However, if $DirIdx \approx 0$ there are two possibilities: or there is no causal connection, and the transfer entropy in both directions is close to zero (as well as the mutual information), or both systems are completely synchronous; or the Y causes X with about the strength that X causes Y, which case transfer entropy will be significant. The latter is the case of identities or bidirectional connection with identical/similar strengths (and similar node dynamics).

Note the dependency of the measures with the symbolic sequence length L . For $L = 1$ we are only evaluating the effect of a single point of the past of X and Y in influencing one point of the future of Y. Highly nonlinear systems often require high order partitions to correctly estimate the quantities and not introducing an artificial memory from the history of the points. Increasing L for an adequate range allows us to approach the results from a Generating Markov Partition (the ‘correct’ partition of the system, which exhibits memoryless property, allowing for the dynamical properties of the system before partitioning being preserved) [123]. This is of particular interest for Seismology. As seen in Chapter 2, the relation between the occurrence of a seismic event and a precursor may

not be a simple connection of short-term events, but an accumulation of past X events, or the lack of them, or even more complex associations impossible to distinguish by simple direct observation of the time-series.

4.4 Local and delayed Causal Mutual Information

In the work that defined CaMI [113] quantities were estimated using a binary encoding ($n_b = 2$), with division line at the centre of the domain, and assessing the whole time-series of a coupled logistic map. The timing of past of X (X_L^{-t}) was chosen to match with the past of Y (Y_L^{+t}), *i.e.* being $t \in [t_0, t_0 + L]$, without any delay between the two variables.

This thesis considers several other configurations for the calculation of CaMI that might be more appropriate to treat complex systems:

1. Allowed for initial partitions with an arbitrary number of bins, thus an encoding with an arbitrary number of symbols;
2. Allowed for the bin division lines to be placed at any location within the domain, so that maximization of the relevant quantities could be seek;
3. Included the possibility of a time-delay between cause and effect;
4. Added the possibility of the analysis being performed in a moving window, so a Local Causal Mutual Information can be evaluated, possibly for real-time measurements.

Standard choices for the location of the initial partition division lines are locations which split the domain in equal sized regions (*e.g.* in the middle of the domain for $n_b = 2$ or every 1/10 of the domain for $n_b = 10$) or positions where the number of points in each bin is equal (so initial symbolic time-series can be approximated by an uniform distribution). However, leaving it flexible for more user options is convenient for seismology. If one of the variables refers to the magnitude of earthquakes in a region along time, it might be of interest to have the flexibility to classify the threshold for which a causal relation holds, especially because most precursor candidates are linked only to the larger seismic events.

The introduction of a time delay d , so that X_L^{-t} ends at $t = t_0 + L$ but Y_L^{+t} only starts at $t = t_0 + L + d$, enables the investigation of the delayed response of the physical system. By varying the value of d , it is expected that the causal measurements peak at the delays corresponding to the natural response time between an event in X and a consequence in Y . In principle, a peak could also be observed for negative delay ($d < 0$) if the coupled system observes anticipated synchronization, a counter-intuitive phenomena in some systems, such as neuron arrays and master-slave lasers, where Y can synchronize with the future of X [124, 125]. By monitoring the causality measures with shifting d , it is possible to question if this is the case for an experimental time-series. The introduction of a moving window enables the understanding of changes in the causal links along time as long as there is sufficient number of points within each window. If the structure of a network is allowed to change through time, then this method could detect these changes.

4.5 Rate of information measures

The Mutual Information Rate (MIR) was firstly mentioned by Shannon as the information exchange between the two variables per unit time, which was later written [126, 127] as

$$MIR = \lim_{L \rightarrow \infty} \frac{I(X_L; Y_L)}{L}. \quad (4.10)$$

It is a convenient measure because the choice of a symbolic length, which is somewhat arbitrary, can be discarded.

In practice, the convergence is fast, with linear relation of $I(X_L; Y_L)$ as a function of L starting as soon as $L = 1$ or $L = 2$. For the largest value of L it is possible that numerical errors occur, due to the partition space having a large number of cells and some being underpopulated. Therefore, MIR is calculated as the slope of the linear interval of the plot of $I(X_L; Y_L)$ in function of L .

We extend the concept of Mutual Information Rate to define a Causal Mutual Information Rate (CaMIR) and a Transfer Entropy Rate (TER), as

$$CaMIR_{X \rightarrow Y} = \lim_{L \rightarrow \infty} \frac{CaMI_{X_L \rightarrow Y_L}}{L}, \quad (4.11)$$

$$TER_{X \rightarrow Y} = \lim_{L \rightarrow \infty} \frac{TE_{X_L \rightarrow Y_L}}{L}. \quad (4.12)$$

The principle for calculating these quantities is to have a measure of rate of the causal information exchange between the variables, and that does not depend on arbitrary parameters such as the length L of the symbolic sequences.

4.6 Pointwise information measures

The pointwise mutual information (PMI) is a measure defined in Computational Linguistics for analysing the co-occurrence of words in a sentence [128]. However, this can be extended to the analysis of co-occurrence of two points in any two time-series. The definition of the pointwise mutual information is given by

$$PMI(\phi_x, \phi_y) = \log \frac{p(\phi_x, \phi_y)}{p(\phi_x)p(\phi_y)}. \quad (4.13)$$

Note that this is the value inside the logarithm term in the original definition of the Mutual Information (Eq. 4.5), and the summation over the partitions is not employed. It means that, for each partition, a PMI value is attributed. Consequently, it is possible to associate for any combination of points x and y from two time-series a value of PMI, representing the information shared by this specific combination. This quantity can be equivalently interpreted as a non-linear ‘correlation’ of events for practical applications. Note that, as the mutual information, the PMI is symmetric. Unlike mutual information, the PMI can be negative. It yields zero for independent variables, negative for rarely occurring (x,y) combinations (with probability below that of independent variables), and increasingly positive as the probability of the co-occurrence increases. A classic example of negative PMI in Computational Linguistics is the case of a preposition being followed by another preposition, which is an event less probable in structured English text than in a randomly generated word sequences.

In analogy to the pointwise mutual information, we define the pointwise transfer entropy (PTE) as

$$\begin{aligned}
PTE_{X_L \rightarrow Y_L}(\phi_x^{-t}, \phi_y^{-t}, \phi_y^{+t}) &= \log \frac{p(\phi_y^{+t} | (\phi_x^{-t}, \phi_y^{-t}))}{p(\phi_y^{+t} | \phi_y^{-t})}, \\
&= \log \frac{p(\phi_x^{-t}, \phi_y^{-t}, \phi_y^{+t}) p(\phi_y^{-t})}{p(\phi_x^{-t}, \phi_y^{-t}) p(\phi_y^{-t}, \phi_y^{+t})}.
\end{aligned} \tag{4.14}$$

This quantity is able to associate, to each set of points of the time-series belonging to a partition cell, the contribution of that cell to the transmission of information from one variable to another. A positive value of $PTE_{X \rightarrow Y}$ means that this point contributes to reduce the uncertainty in the future of Y (thus an optimal location to observe the system), whereas a negative value indicates that such combination augments the uncertainty in the future Y (thus a sub-optimal place to observe the system). By colouring the cells with the value of the PTE, we can see several interesting dynamical properties of the system. Same colour cells would indicate regions that are intrinsically dynamically connected, for example by having orbits that are recurrent to the cell after L iterations. These connected cells would present, as we will show later, a positive PTE. Cells that present a negative PTE correspond to cells whose trajectories do not recur after L iterations, thus knowing that an initial point belong to a cell cannot be used to predict with large accuracy where the point will be after L iterations. For the practical case of earthquake prediction, it means that, for each combination of precursor candidate and seismicity occurrence, it can be associated a pointwise transfer entropy value informing if the precursor is indeed reducing the uncertainty (*i.e.* increasing our knowledge) over a future earthquake event, or if such knowledge cannot be attested from the correlation (*i.e.*, the correlation can be spurious).

For completeness, we also define a pointwise causal mutual information (PCaMI) and a pointwise directionality index (PDirIdx):

$$\begin{aligned}
PCaMI_{X_L \rightarrow Y_L}(\phi_x^{-t}, \phi_y^{-t}, \phi_y^{+t}) &= \log \frac{p(\phi_x^{-t}, \phi_y^{-t}, \phi_y^{+t})}{p(\phi_x^{-t}) p(\phi_y^{-t}, \phi_y^{+t})}, \\
&= PTE_{X_L \rightarrow Y_L}(\phi_x^{-t}, \phi_y^{-t}, \phi_y^{+t}) + PMI(\phi_x, \phi_y),
\end{aligned} \tag{4.15}$$

$$\begin{aligned}
PDirIdx_{(L)}(\phi_x^{-t}, \phi_y^{-t}, \phi_y^{+t}) &= PTE_{X_L \rightarrow Y_L}(\phi_x^{-t}, \phi_y^{-t}, \phi_y^{+t}) - PTE_{Y_L \rightarrow X_L}(\phi_x^{-t}, \phi_y^{-t}, \phi_y^{+t}), \\
&= PCaMI_{X_L \rightarrow Y_L}(\phi_x^{-t}, \phi_y^{-t}, \phi_y^{+t}) - PCaMI_{Y_L \rightarrow X_L}(\phi_x^{-t}, \phi_y^{-t}, \phi_y^{+t}).
\end{aligned} \tag{4.16}$$

Observe that they follow the same relations with the PTE and PMI that occurs regarding CaMI and DirIdx with transfer entropy and mutual information.

4.7 Algorithm of the method and usage

We have developed the *Causality Toolbox* for Matlab, containing functions to calculate the Causal Mutual Information, Mutual Information, Transfer Entropy, Directionality Index, Mutual Information Rate, Causal Mutual Information Rate, Transfer Entropy Rate, Pointwise Causal Mutual Information, Pointwise Transfer Entropy and Pointwise Mutual Information, and, additionally, Multivariate Total Correlation and Multivariate Joint Entropy.

The latter two are global measures (instead of only pairwise) giving a simplified measure for the state of the whole system. Total correlation and the multivariate joint entropy are respectively defined by

$$C(X_L^{(1)}, \dots, X_L^{(n)}) = \sum_{\phi_1, \dots, \phi_n} p(\phi_1, \dots, \phi_n) \log \frac{p(\phi_1, \dots, \phi_n)}{p(\phi_1) \dots p(\phi_n)}, \tag{4.17}$$

$$H(X_L^{(1)}, \dots, X_L^{(n)}) = - \sum_{\phi_1, \dots, \phi_n} p(\phi_1, \dots, \phi_n) \log p(\phi_1, \dots, \phi_n). \tag{4.18}$$

We are not going to apply the total correlation and the multivariate joint entropy in this thesis, as these quantities have been included in the toolbox as the first steps towards a future definition and calculation of multivariate options for all the measures presented so far.

The toolbox is available for download on a GitHub² public repository. It consists

²<https://github.com/artvalencio/causality-toolbox>

of three Matlab functions, *cami.m*, *totals.m* and *camir.m* with procedures given by Figs. 4.3, 4.4 and 4.5. The function *cami.m* calculates the Causal Mutual Information, Transfer Entropy, Mutual Information and Directionality Index by simply providing the two time-series, the location of the initial partitions, the length of the symbolic sequence, embedding delay τ if applicable, and the units (*nats* or *bits*). The function *totals.m* calculates the total correlation and multivariate joint entropy by the definition when the time-series of the nodes, partition locations and the symbolic length are provided, together with the units. Finally, the *camir.m* function obtains the Causal Mutual Information Rate, Transfer Entropy Rate and Mutual Information Rate for a range of partition possibilities in X and Y. The optimal partition is considered to be the one which yields the highest exchange rate of causal information, so the maximum of *CaMIR*. For this, results are provided in separate, for the remaining cases, it is provided as a struct. The inputs are only the time-series and the units. The function *camir.m* depends on *cami.m*.

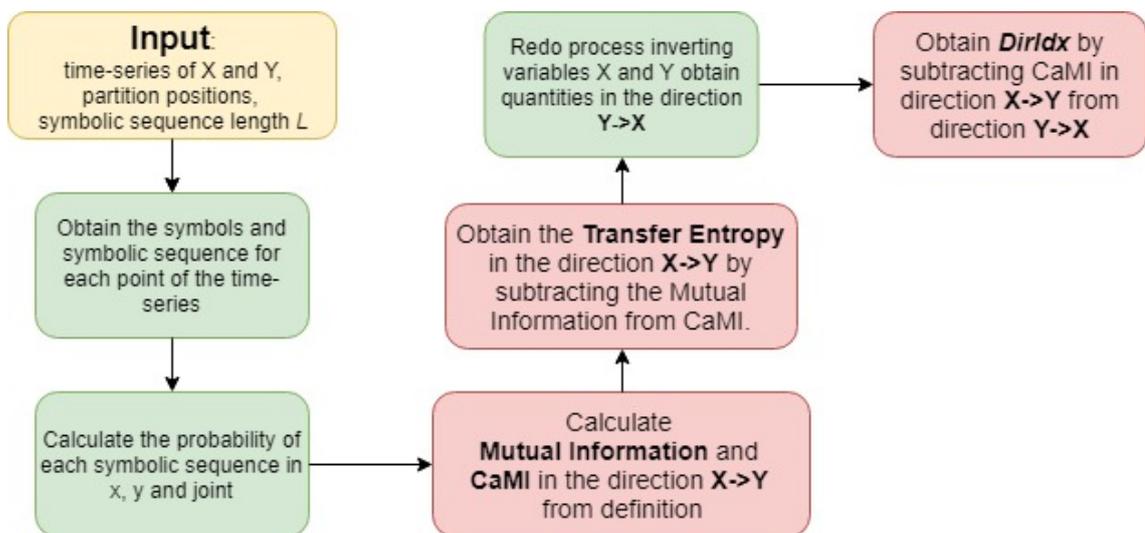


Figure 4.3: Flowchart of the procedure for calculating the informational measures of interest: Mutual Information, Transfer Entropy, Causal Mutual Information and Directionality Index. This is adopted in the Matlab function *cami.m* from the *Causality Toolbox*.

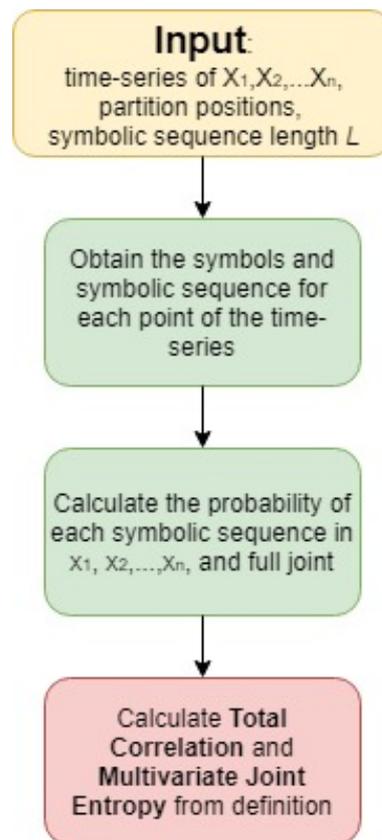


Figure 4.4: Flowchart of the procedure for calculating the Total Correlation and the Multivariate Joint Entropy. This is adopted in the Matlab function *totals.m* from the *Causality Toolbox*.

The Matlab call for the function *cami.m* is:

```
[ cami_xy , cami_yx , mutual_info , diridx , te_xy , te_yx , pointwise ] =
cami ( x , y , L_past , L_fut , xpart , ypart , tau , units , opts );
```

Where x and y are the time-series provided as vector columns. Our implementation allows to have a different symbolic sequence length L for the past of a time-series and for the future of the effect (L_{past} and L_{fut}). In particular, the case analysed by Bianco-Martinez [113, 114], of total symbolic length 2 in X and total symbolic length of 5 in Y is equivalent to $L_{past} = 2$ and $L_{fut} = 3$, *i.e.*, the hypothesis that 2 points in the past of X and Y determine a 3-point trend in the future of Y . The inputs $xpart$ and $ypart$ are vector columns of the initial partition locations in X and Y , tau is the embedding dimension τ if applicable (1 in case of maps) and $units$ is a string ‘bits’ or ‘nats’ as appropriate. Valid options (*opts*) include:

- ‘save’: for saving output to file
- ‘delay’, value (number of points to delay): to consider a delay between cause and effect (delayed CaMI)
- ‘local’, value (number of points in the time-window): to calculate the measures over a moving time-window

The Matlab call for *totals.m* is:

```
[ totalcorrel , totalentropy ] = totals ( x , L , linepos , tau );
```

Where x is a matrix where each column is the time-series of one node. L is the length of the symbolic sequences and $linepos$ is a matrix informing the position of the initial partition divisions in each time-series, with the columns representing the same respective nodes indicated by the columns of x . The input tau , is again the embedding delay if appropriate, or 1 if the time-series comes from a map.

Finally, the Matlab call for *camir.m* is:

```
[ rates , fullresults ] = camir ( x , y , tau );
```

Where x and y are the time-series and τ the embedding delay if appropriate. The outputs $rates$ and $fullresults$ are structs. The first informs the values of main interest: the optimal partition and the $CaMIR$, TER and MIR for this partition. The latter informs all the calculated informational measures and rates for all partitions considered.

4.8 Computational demands

The implementation of the method for calculating CaMI on *camim* has an arithmetic complexity of $O\left(n_b^{(2L_{past}+L_{fut})}\right)$ in the main computational part for the flow of information, and $O(N_T(L_{fut} + L_{past}))$ for the construction of the partitioned space. In comparison, a classical implementation for Transfer Entropy from its original definition would have the same computational complexities in the main part and partitioned space, but with an additional component for determining the matrix of conditional probabilities, with complexity of at least $O(N_T(L_{fut} + L_{past}))$. If converting the conditional probabilities to joint probabilities the extra arithmetic complexity would be reduced, but there still would be a larger number of operations to be performed for obtaining the transfer entropy and the mutual information in comparison to obtaining the transfer entropy from the CaMI and mutual information. The big-O complexity notation informs not the actual performance of the code, but a worst-case scenario of memory/time dependence to the main variables. The actual performance is highly dependable on the machine used.

Timing and memory demand tests were performed in a HP Z220 workstation running 64-bit Ubuntu 14.04LTS, and having 8Mb cache, 4-core 3.4GHz Intel i7-3770 processor, 16Gb 1600MT/s DDR3 RAM memory, and with Swap memory intentionally deactivated, so that the program automatically stops if the memory requirements surpass the RAM availability. The dependence of execution time with the total number of points N_T of the time-series is given in Fig. 4.6, where $n_b = 2$ and $L_{past} = L_{fut} = 3$. Observe that the relation is linear, as expected. Fixing $N_T = 10^6$ and $n_b = 2$, the execution time varies with the symbolic length L ($=L_{past} = L_{fut}$) according to Fig. 4.7. The relation is no longer linear, assuming a more complicated form, as expected. For $N_T = 10^6$ and $L_{past} = L_{fut} = 3$, the relation of execution time with the number of initial bins n_b is described by Fig.

4.8. Note that the execution time grows considerably fast the number of bins, as expected.

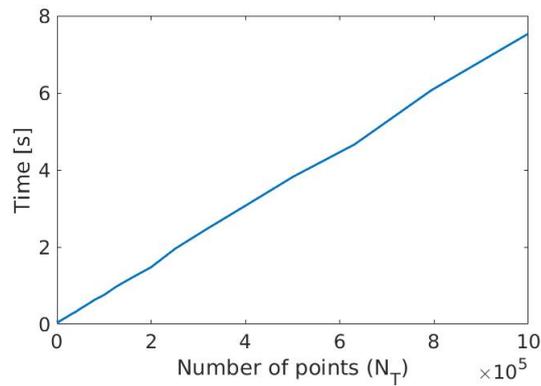


Figure 4.6: Time performance of *camim* in function of the number of points of the time-series.

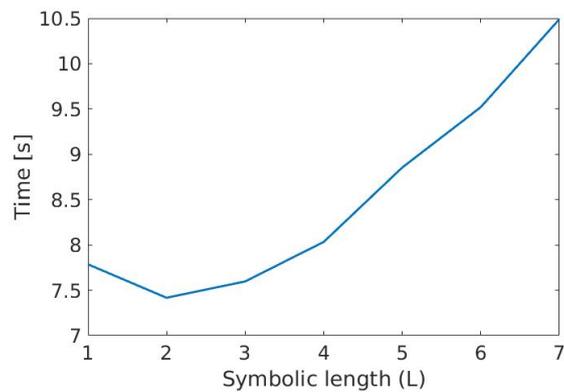


Figure 4.7: Time performance of *camim* in function of the symbolic length $L = L_{past} = L_{fut}$.

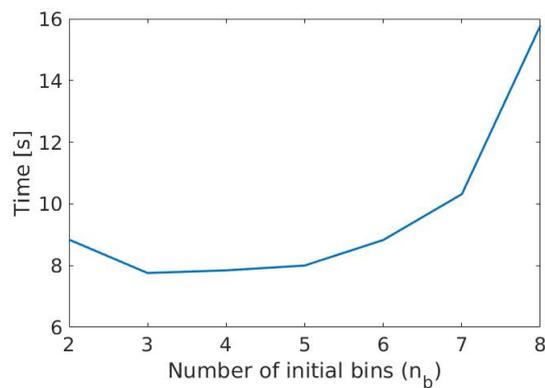


Figure 4.8: Time performance of *camim* in function of the number of initial partition bins n_b .

The same procedure was performed for memory usage. It must be stressed that we are limiting our calculations to cases where there is an average of 30 points or more per

3-dimensional $(\phi_X^{-t}, \phi_Y^{-t}, \phi_Y^{+t})$ box. This is an arbitrary and conservative choice to ensure the computation of the probabilities of each cell of the partitioned space is statistically significant. It was observed no significant variation of memory usage with the total number of points N of the time-series, when n_b and $L_{past} = L_{fut} = L$ are fixed. However, selecting $N_T = 10^6$ and $n_b = 2$, the maximum memory usage changes with the symbolic length $L (= L_{past} = L_{fut})$ as Fig. 4.9 presents. The memory usage is dominated by the step of defining the partitioned space and obtaining the probabilities, expressing a linear growth, as expected. Finally, setting $N_T = 10^6$ and $L_{past} = L_{fut} = 3$, Fig. 4.10 gives the relation of maximum memory usage with the number of initial bins n_b . Once again, note that the memory usage can increase significantly with the number of initial partition bins.

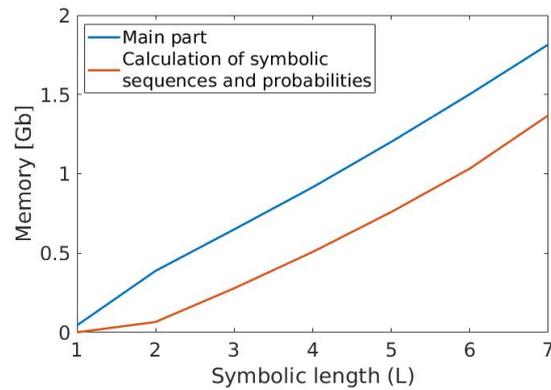


Figure 4.9: Memory usage of *camim* in function of the symbolic length $L = L_{past} = L_{fut}$.

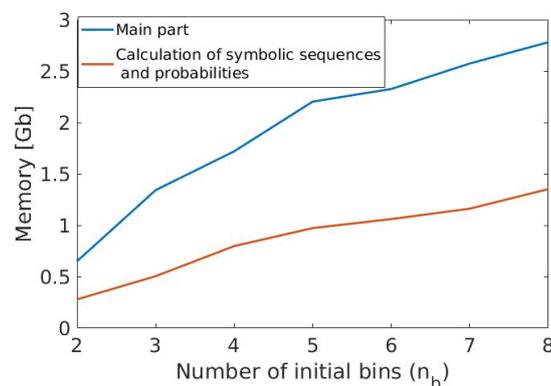


Figure 4.10: Memory usage of *camim* in function of the number of initial partition bins n_b .

The memory values displayed here refer to the observed difference between the total

RAM used and the cached RAM memory of the workstation when running the function. It was also subtracted the memory usage of the operational system, background programs and the Matlab work environment itself, so that only the memory consumption by the *camim* function is presented in the plots. The memory profiling procedure was performed separately from the time profiling, so that this memory usage computation did not affect the results of execution time. Windows users can use the Matlab in-built function *memory* for this procedure, but this is not available for linux users. We have deposited on Github the solution *memmonitor.m* adopted for Ubuntu (<https://github.com/artvalencio/tools>).

Chapter 5

Assessing causality to test-bench systems

This chapter presents the results of a comprehensive causality analysis to characterise test-bench systems. The primary goal is to observe if proposed methodology with the quantities introduced in the previous chapter is capable of correctly identify the arrow of influence in numerically generated time-series. By construction, we know what is the direction of causality, but this is not assumed to be unknown to whom has access to the time-series (emulating a situation where someone is collecting the data from measurements), and the direction of the link is not clear from direct observation. We observe that the directionality index calculated from CaMI not only correctly identify the direction of causality, but also the mutual information provides a reliable parameter for the identification of the strength in coupled logistic maps. It is also shown that the calculation from the information rates (CaMIR, transfer entropy rate and mutual information rate) lead to similar results, with one less arbitrary parameter. The applications are extended to networks, where the directionality index from the rates correctly informs the direction of a link, but the mutual information rate better indicates the strength of a link. In conjunction with transfer entropy rate, the mutual information rate may support the identification of direct or indirect links in a causal network. We consider the information flow when the connections are of CML type or linear diffusive, and the effects when the information from a chaotic node is transmitted in a path of periodic nodes or chaotic nodes (chaotic or a

periodic channel [129]) until reaching final nodes which are also chaotic. Consequences of the introduction of dynamical noise are also explored. In this chapter we will adopt $L_{past} = L_{fut} = L$ for simplification. Numerical values of CaMI, transfer entropy, mutual information and directionality index are in bits, and of CaMIR, transfer entropy rate and mutual information rate in bits/iteration.

5.1 Pseudo-random noise

The first question is how this method to assess causality responds to a fully unconnected and uncorrelated system. A classical example with any method based on calculation of probabilities is the fair-coin toss problem. For two independent fair coin-tosses, we can record the sequential outcomes as the time-series X and Y. The events being completely independent implies that both the mutual information and the transfer entropy are zero by definition, and consequently are the CaMI and the directionality index. Therefore, testing our methodological approach against the fair-coin toss case reveals to which conditions our method to detect causality provides outputs that cannot be discernible from those coming from random uncorrelated systems. In this case, the fair-coin is generated by the Matlab in-built pseudo-random routine (*rand*) producing a [0,1] uniform distribution, which is then encoded as *heads* ('0') or *tails* ('1') by dividing at the 0.5 line in each time-series.

The outcome from the function varies in symbolic length L as in Fig. 5.1, for the number of points N_T fixed in 10^6 . The errors in the mutual information are of about $6 \cdot 10^{-4}$, whereas in the transfer entropy (and consequently in CaMI) are smaller than $5 \cdot 10^{-3}$ until $L = 4$ and about $2.3 \cdot 10^{-2}$ for $L = 5$. If we fix the symbolic length to $L = 3$, the error of informational measures decrease with the number of points N_T as Fig. 5.2. Observe that as the number of points available in the time-series decreases by one order of magnitude, so the accuracy reduces by one decimal place. For example, it means that while for $N_T > 10^6$ errors in CaMI and CaMIR are in the fourth decimal place, if only $N_T = 2 \cdot 10^5$ points are available, these errors increase to the third decimal place, whereas if $N_T = 2 \cdot 10^4$ they go to the second decimal place and so on.

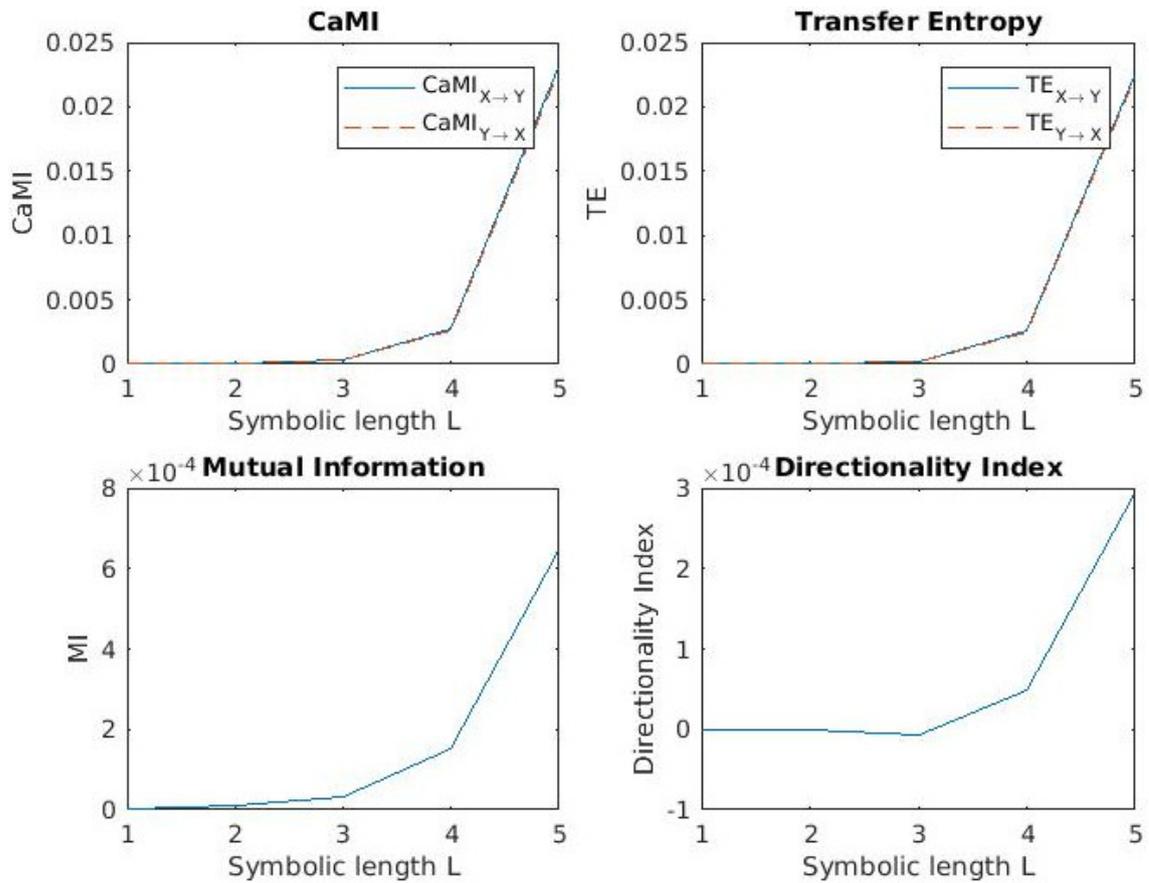


Figure 5.1: Informational measures for a two uncoupled pseudo-random time-series (uniform $[0,1]$ distribution), for varying symbolic length L , $n_b = 2$. A total of 10^6 time-series points are used in the analysis.

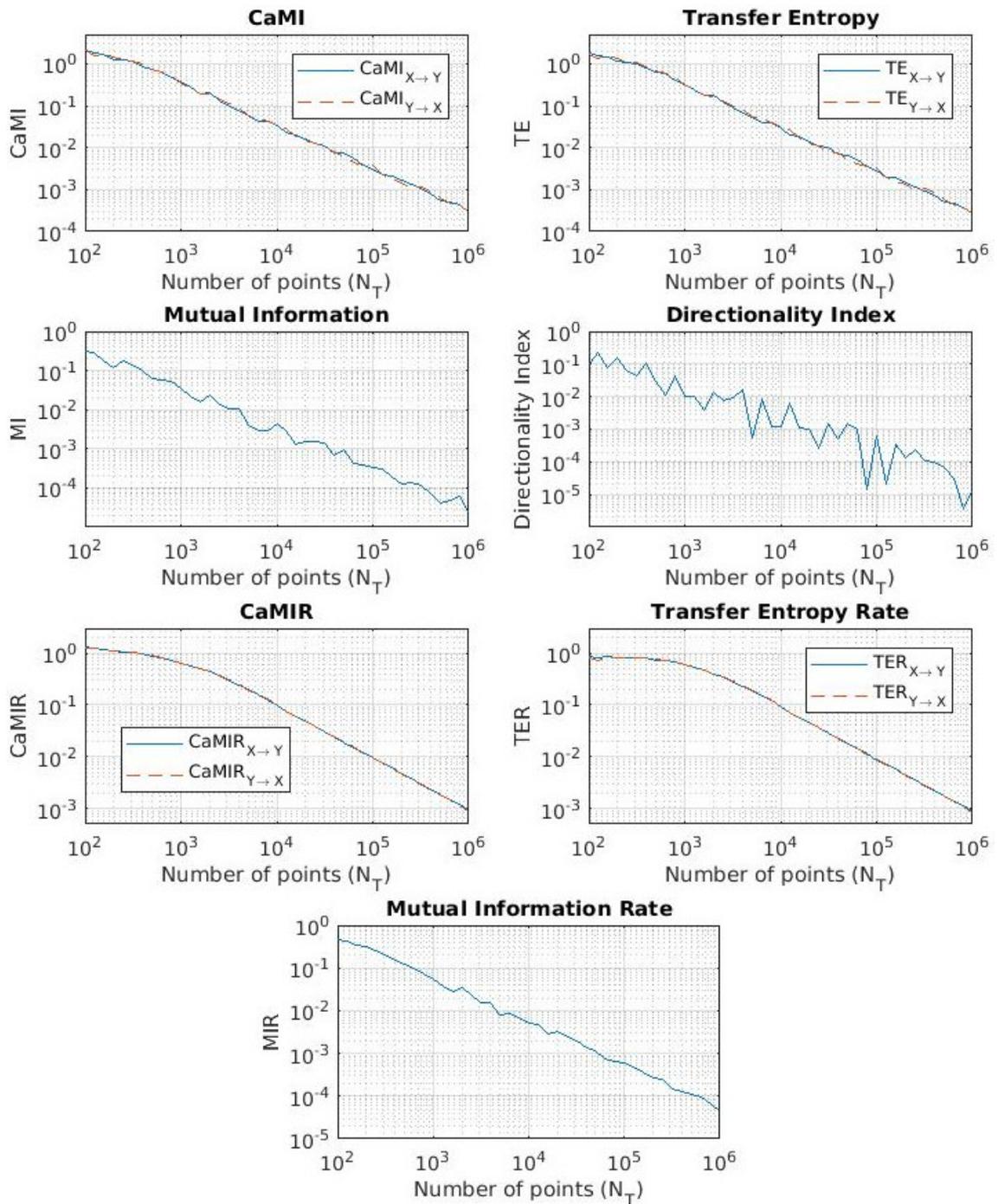


Figure 5.2: Informational measures for a two uncoupled pseudo-random time-series (uniform $[0,1]$ distribution), for varying number of points N_T , with number of initial partitions $n_b = 2$ and symbolic length $L = 3$

The reason that these quantities provides incorrect information for larger L is due to the time-series being finite. Assuming an initial binary partition (with 2^2 cells), when treating symbolic sequences of length L will produce 2^{2L} cells for the calculation of mutual information and 2^{3L} cells for the calculation of CaMI and transfer entropy (due to considering an axis of the past of Y and another for the future of Y). Probabilities in the cells will be evenly distributed, roughly uniform, since the data is random and decorrelated. Because the number of points is finite, the less equal the proportion of number of points in each box, the larger the L . Since cells will not all be visited with exactly the same frequency (except in the limit $N_T \rightarrow \infty$), it creates an artificial illusion for information.

5.2 Application to coupled logistic maps

5.2.1 Error levels: uncoupled system

Moving from a stochastic to a deterministic dynamical system, our starting question is if we observe the same behaviour for a case where it is known the informational measures should yield zero. Hence, we create the time-series from two uncoupled ($A_{i,j} = 0$) logistic maps with free parameter $r = 4$. Indeed, the method produced the same output for the uncoupled logistic maps, as expected, once there is no exchange of information between the two systems. Furthermore, the logistic map for $r = 4$ and a partition line at 0.5 (which is a Generating Markov Partition) leads to a fully random uncorrelated symbolic sequence. This is presented on Fig. 5.3 for dependency on the symbolic length and Fig. 5.4 for dependency on the number of time-series points, where precisely the same error margins for the same informational quantities are observed. In particular, we highlight that for $N_T = 10^6$, the errors in CaMI, CaMIR, transfer entropy and transfer entropy rate are in the third decimal place, while in the mutual information and mutual information rate are in the fourth decimal place. This method is intended to be applicable to time-series in excess of 10^3 points. For the minimal case ($N_T = 10^3$) the errors in CaMI, CaMIR, transfer entropy and transfer entropy rate are in the first decimal place, whereas in the mutual information and mutual information rate are in the second decimal place.

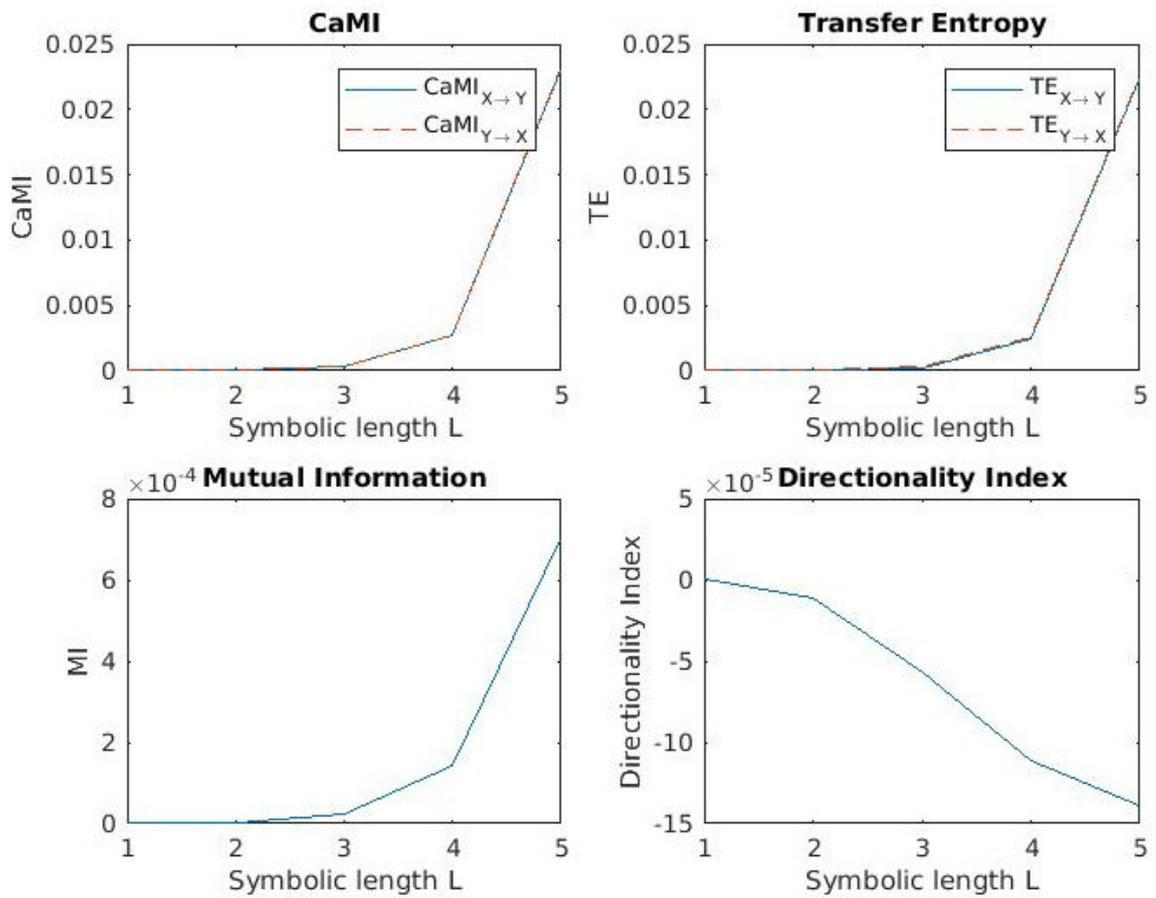


Figure 5.3: Informational measures for a two uncoupled logistic time-series ($r = 4$, $A_{i,j} = 0$), for varying symbolic length L , $n_b = 2$. A total of 10^6 non-transient time-series points are used in the analysis.

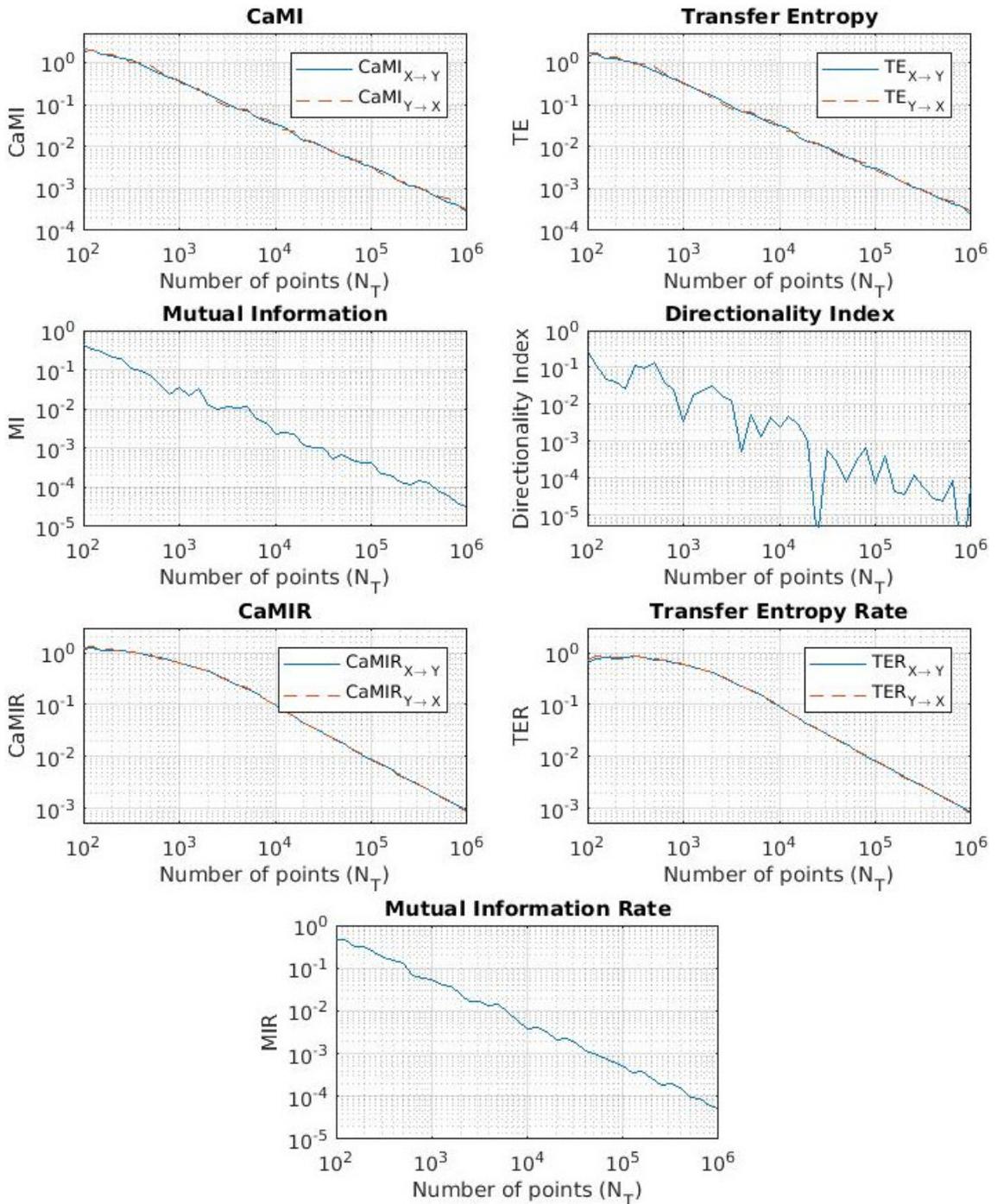


Figure 5.4: Informational measures for a two uncoupled logistic time-series ($r = 4$, $A_{i,j} = 0$), for varying number of points N_T , with number of initial partitions $n_b = 2$ and symbolic length $L = 3$.

5.2.2 Diffusive coupling

The first coupled system we study is the coupled logistic map with linear diffusive (Eq. 3.6) in the configuration $X \rightarrow Y$. The phase-space of this system changes with different coupling strengths such as seen in Chapter 3 (Fig. 3.13 (left)). For a total of 10^6 points, binary partitioning ($n_b = 2$) and a sequence length of $L = 3$, the informational measures vary with the coupling strength σ as Fig. 5.5. In this case, we have fixed the symbolic sequence length to $L = 3$ to reveal the behaviour of the CaMI, transfer entropy, mutual information and directionality index. The directionality index is positive, correctly indicating that the information flows from X to Y.

The mutual information dominates the CaMI contribution except for the very low coupling strength. Moreover, the relation of the mutual information with the coupling strength is almost linear for this type of coupling. For the diffusive coupling it is also observed that the transfer entropy in the direction $X \rightarrow Y$ is at an almost constant level of 0.8 bits for a large interval for the chosen parameters. A small valley is observed near the coupling $\sigma = 0.2$. The transfer entropy and transfer entropy rate in the opposite direction is 1-2 orders of magnitude smaller, increasing linearly for $\sigma > 0.25$. This is an indicator that from this point we are entering the high intensity coupling regime. The CaMIR is dominated by the mutual information rate, which increases almost linearly with the coupling strength. Notice that it is also possible to correctly identify the direction of causality from the difference of CaMIR in each direction. It is preferred to analyse based on the CaMIR, once the numerical values are independent of an user selection of the partition resolution.

The numerical values of these informational quantities depend on the user choice of the length of the symbolic sequence or the number of initial partitions, which, ultimately, are measures of the resolution of the analysis. Figs. 5.6 and 5.7 reveal how these numbers change as the user selects different L and n_b parameters. Although there is an overall trend of increase of the numerical values of the informational quantities for increasing resolutions, the relation is not necessarily linear, so the resolution must be informed. In particular, the rates have to be calculated over a linear portion of the dependence with

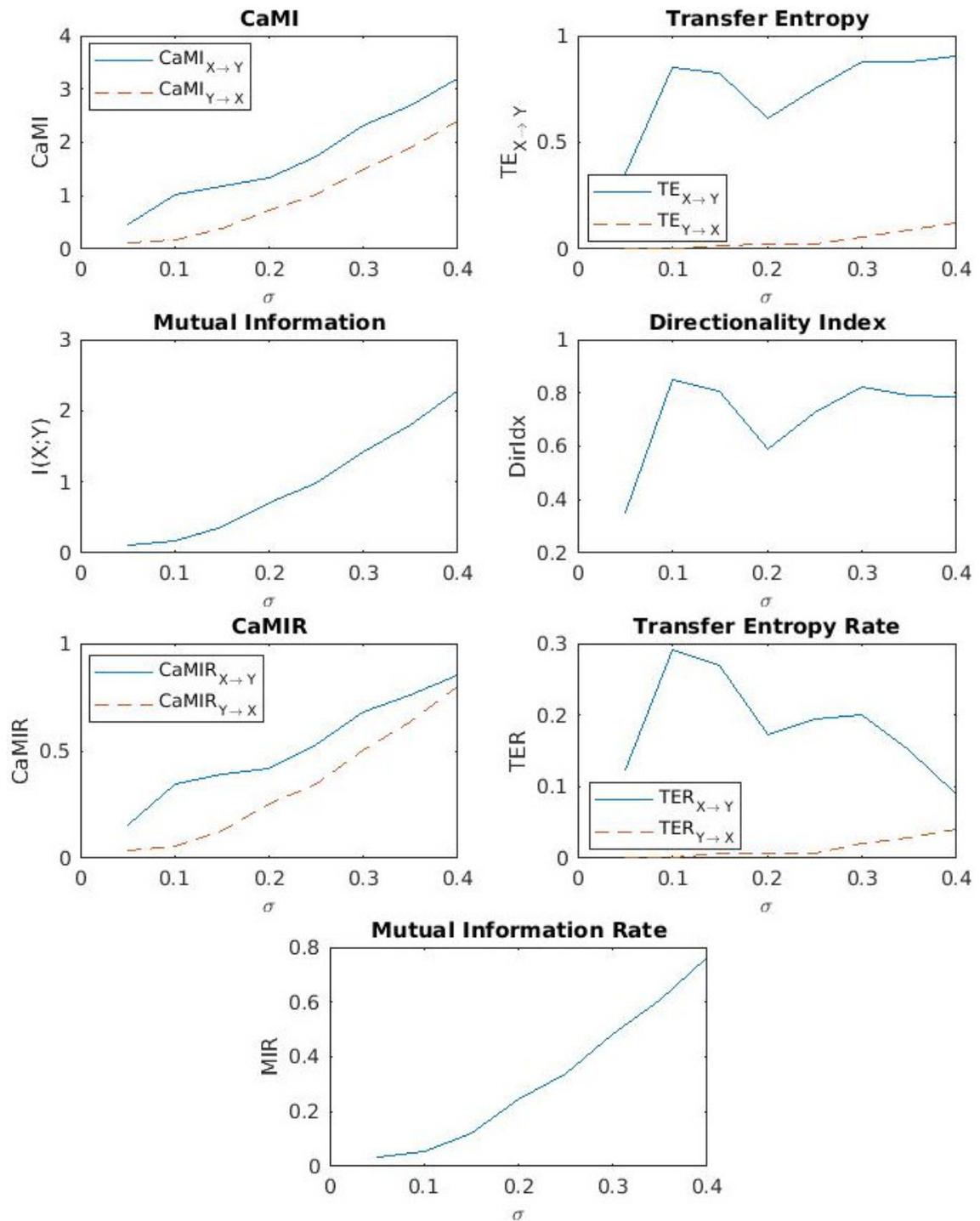


Figure 5.5: Informational measures for a time-series from coupled logistic $X \rightarrow Y$ ($r = 4$) map with diffusive coupling. Values for $L = 3$, $n_b = 2$ and varying coupling strength σ . A total of 10^6 non-transient time-series points are used in the analysis.

L (Fig. 5.6), which in this case is attainable in the interval $L \in [2,4]$. The reason of why $L = 5$ produced a discrepancy to the trend is because at this resolution there will be cells with insufficient points, not enabling the determination of the true distribution of probabilities for a proper statistical count.

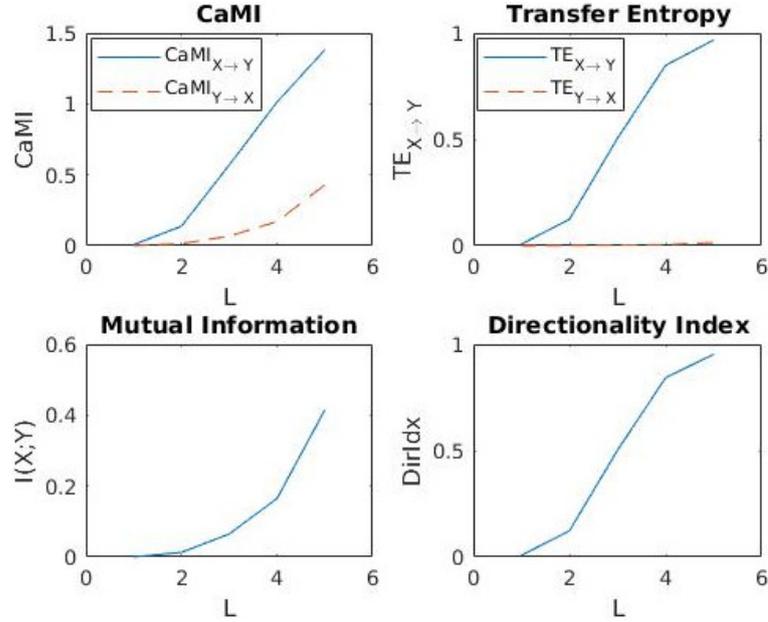


Figure 5.6: Informational measures for a time-series from coupled logistic $X \rightarrow Y$ ($r = 4$) map with diffusive coupling. Values for $n_b = 2$, $\sigma = 0.1$ and varying symbolic length L . A total of 10^6 non-transient time-series points are used in the analysis.

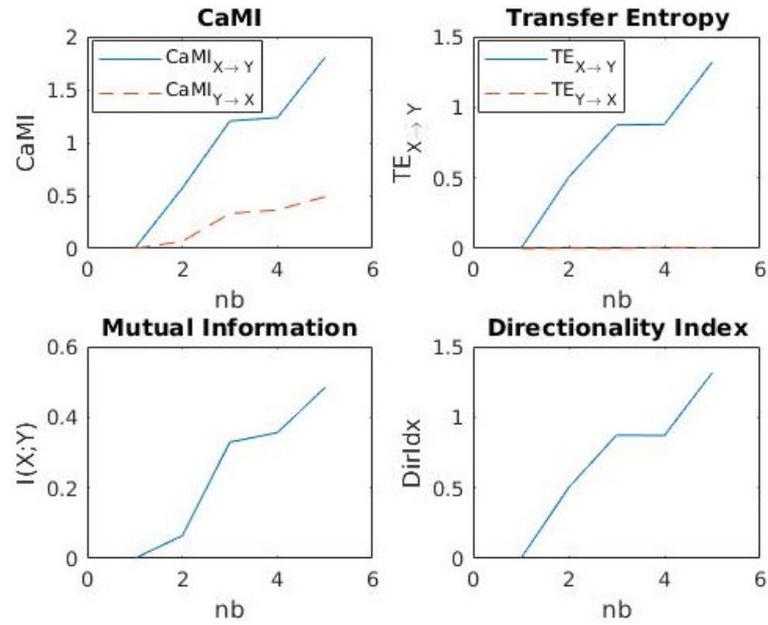


Figure 5.7: Informational measures for a time-series from coupled logistic $X \rightarrow Y$ ($r = 4$) map with diffusive coupling. Values for $L = 3$, $\sigma = 0.1$ and varying number of initial bins n_b . A total of 10^6 non-transient time-series points are used in the analysis.

If a small coupling is included in the opposite direction (i.e. a $X \leftrightarrow Y$ system, with $\sigma_{x \rightarrow y} \gg \sigma_{y \rightarrow x}$) no noticeable effects are observed. This occurs because the partitioned space does not change significantly (causal bubbles only in the main direction of influence). As the coupling strength increases to a similar value in both directions ($\sigma_{x \rightarrow y} \approx \sigma_{y \rightarrow x} = \sigma$), the system changes considerably. Non-synchronous solutions exist only for $\sigma < 0.1$, and the partition space presents a symmetric causal bubble, not allowing for the identification of a predominant net flow of information, as observed in Chapter 3 (Fig. 3.15 (right)). The informational quantities for this system varies with the coupling strength as in Fig. 5.8.

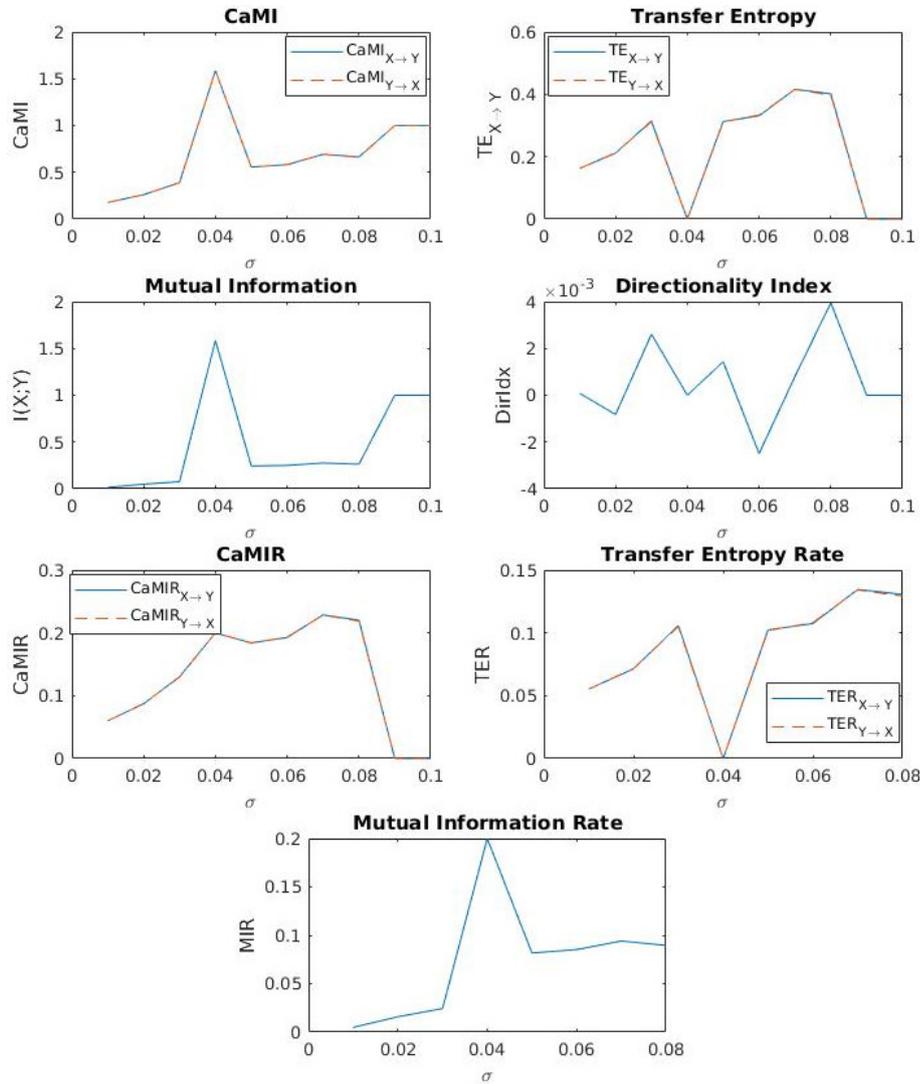


Figure 5.8: Informational measures for a time-series from coupled logistic $X \leftrightarrow Y$ ($r = 4$) map with diffusive coupling. Values for $L = 3$, $n_b = 2$ and varying coupling strength σ . A total of 10^6 non-transient time-series points are used in the analysis.

Notice that the quantities in the direction $X \rightarrow Y$ and $Y \rightarrow X$ overlap, as expected.

The directionality index is within the expected error level in the third decimal place, hence showing no preferential direction, as expected. However, all other informational quantities hold significant values, indicating that there is information exchange between the systems, allowing the conclusion that both systems influence each other equally. This time the mutual information and mutual information rate do not grow linearly. However, it must be noted that from $\sigma = 0.8$ the system begins to synchronize. The system also synchronises at $\sigma = 0.4$, converging to two points in the phase-space, which explains the peak in mutual information. For the intervals far from synchronisation, mutual information and mutual information rate grow linearly with coupling strength.

For a weakly coupled ($\sigma = 0.1$) $X \rightarrow Y$ system we also investigate the distribution of the pointwise information measures in the phase-space. The results for increasing symbolic length L are shown in Figs. 5.9 and 5.10. Although the PMI starts almost homogeneous with high pointwise mutual information (only one outlier partition, equivalent to an independent measure), as the symbolic length increases different centres of high pointwise mutual information start to form, with a comparatively lower PMI in between. These centres are the areas inside the causal bubbles presented in [113, 114]: enclosed areas formed by the backward iterations of the partition lines. The formation of the bubbles is a feature of the flow of information between the two variables.

In the pointwise transfer entropy, not only the bubbles in the same regions are visually evident with lower L , but also it can be observed a ring-formation pattern, of alternating high and low transfer entropy. In these areas, mostly corresponding to the edge of the bubbles, there is significant knowledge on the future of the effect. On the outside of the rings, and sometimes even between the rings, the pointwise transfer entropy can be negative. In these areas the uncertainty of the future effect increases. The pointwise causal mutual information and pointwise directionality index largely follow the results from the pointwise transfer entropy. The pointwise directionality index, in particular, suggests that these causal bubbles are the main contributors for the information flow properly indicating the underlying physical causation. Few areas between these centres indicate the opposite direction, which is an open question.

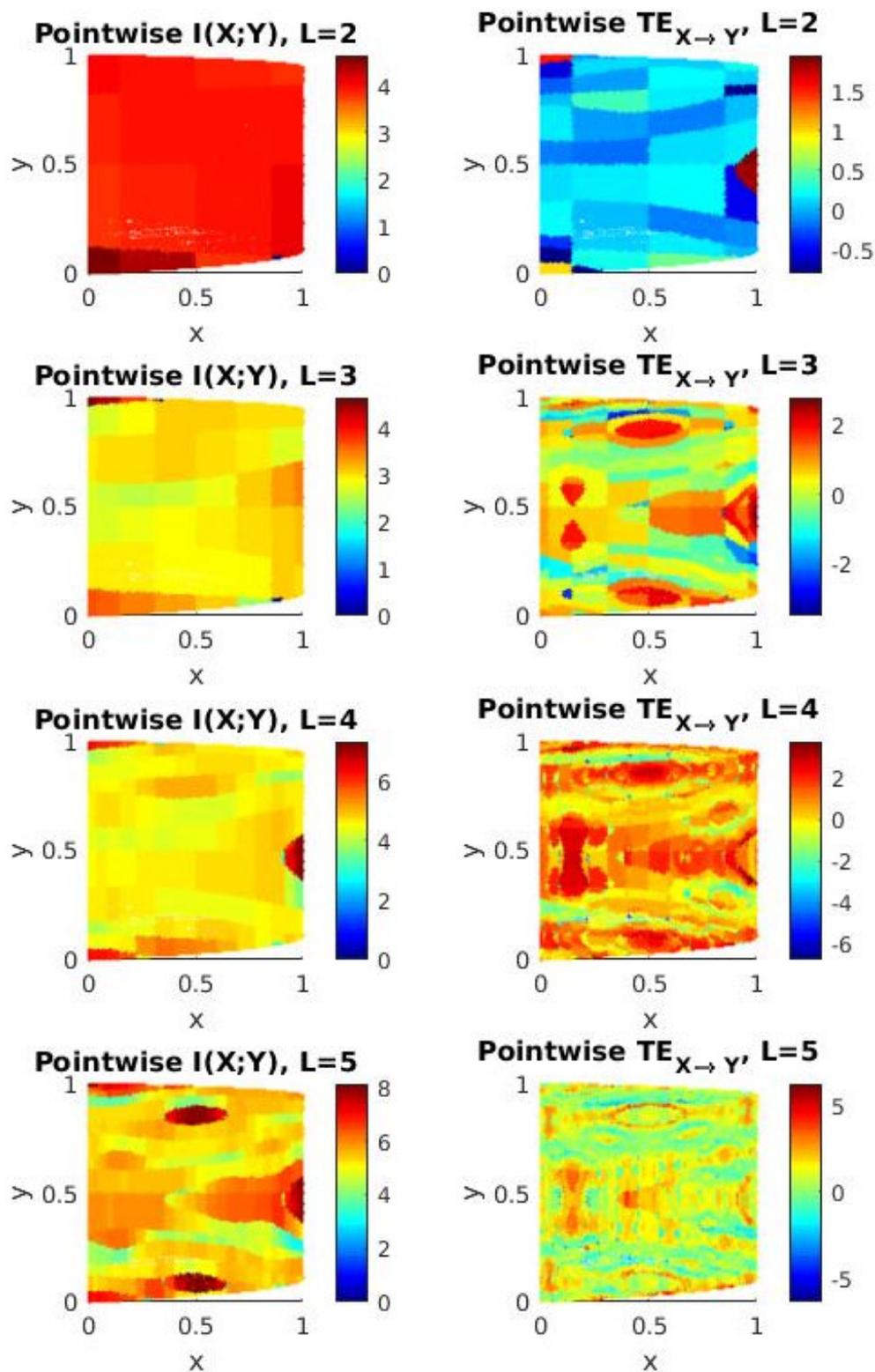


Figure 5.9: Pointwise mutual information (left column) and pointwise transfer entropy (right column) for a time-series from coupled logistic $X \rightarrow Y$ ($r = 4$) map with linear diffusive coupling. Values for $n_b = 2$ and $\sigma = 0.1$. A total of $2 \cdot 10^5$ non-transient time-series points are used in the analysis.

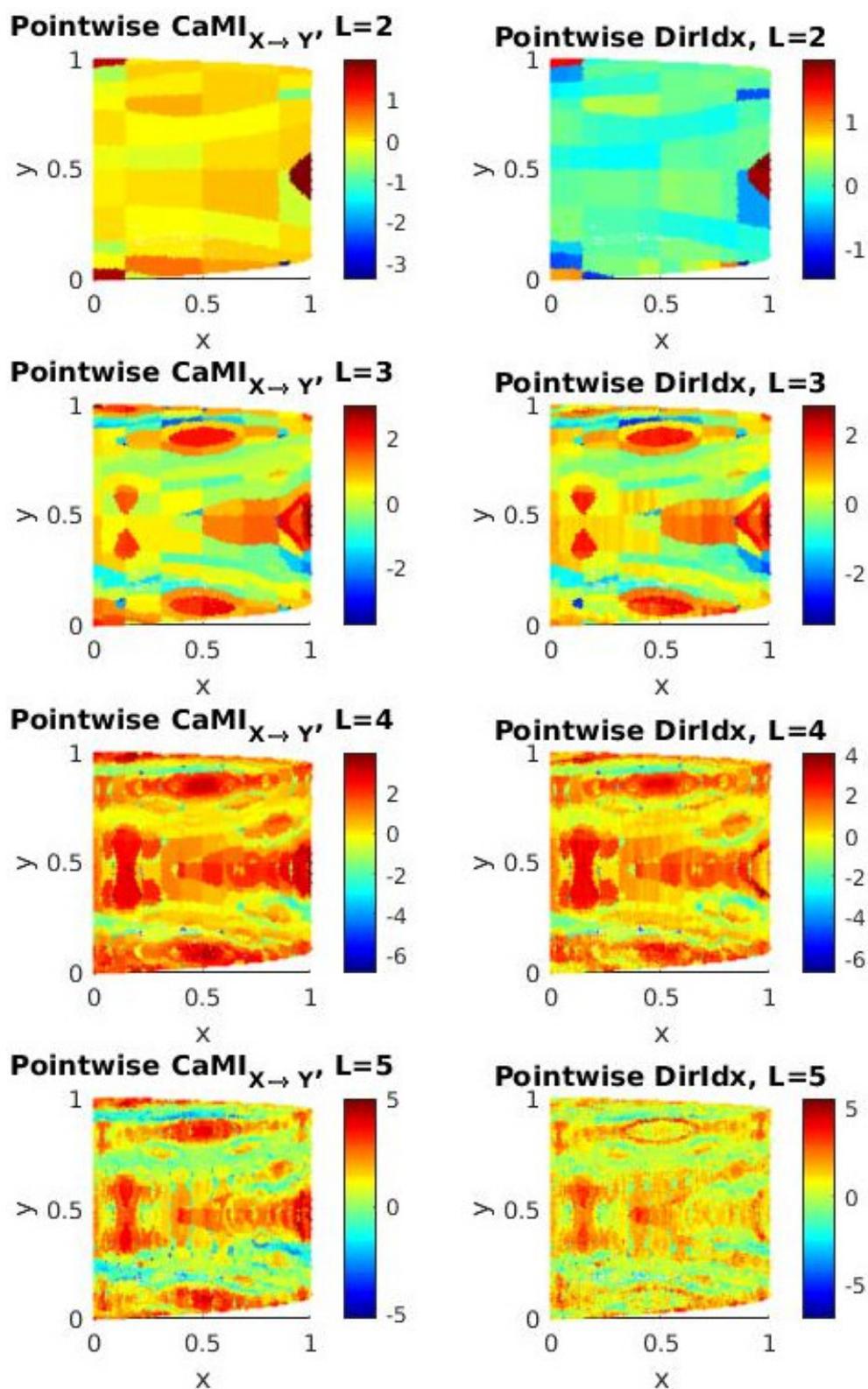


Figure 5.10: Pointwise causal mutual information (left column) and pointwise directionality index (right column) for a time-series from coupled logistic $X \rightarrow Y$ ($r = 4$) map with linear diffusive coupling. Values for $n_b = 2$ and $\sigma = 0.1$. A total of $2 \cdot 10^5$ non-transient time-series points are used in the analysis.

5.2.3 CML coupling

A new question is how the results differ when a different kind of physical coupling mechanism is in place, such as the CML coupling scheme (Eq. 3.7). We first build such a system with $X \rightarrow Y$ causation, for varying coupling strength σ in the $(0,0.4]$ range. The changes in the phase-space with σ was also seen in Chapter 3 (Fig. 3.13 (right)).

Figure 5.11 shows the results for $L = 3$ and $n_b = 2$ using $N_T = 10^6$ points. A few differences with respect to the linear diffusive case are evident. The first is that the transfer entropy in the direction of the causal flow is not almost constant this time. It presents a sharp peak in $\sigma = 0.2$, followed by a valley at $\sigma = 0.25$. Also, the transfer entropy in the opposite direction grows linearly with the coupling strength even for small σ , despite only reaching a significant value for $\sigma > 0.25$. Another aspect is that this time the transfer entropy in the opposite direction can yield values of about half the transfer entropy in the causal direction. The directionality index indicates the correct direction for all σ and the mutual information increases about linearly with the coupling strength. The mutual information rate increases almost linearly, except for the peak on $\sigma = 0.2$. From the difference in CaMIR the correct direction of causality can also be identified. The transfer entropy exhibits a sharp peak at $\sigma = 0.15$.

Once again we question what is the variation of the numerical values obtained with the parameters of choice of the user. For the variation with the symbolic sequence length L (Fig. 5.12) this time we observe an almost linear increase in all informational quantities in the range $L \in [1, 5]$. It means that the whole interval can be considered for the analysis of the informational rates. With respect to the variation with the number of initial partitions n_b (Fig. 5.13), the same behaviour of a general growth (but in most cases not linear) is observed, like the case with diffusive coupling. Hence the analysis is effective for a significant range of parameter L . It is preferred to select the smaller symbolic lengths to which the method is applicable once it is less computationally expensive and carries less errors from underpopulation of the partition cells.

The pointwise informational measures were also obtained for the CML coupling, in the low coupling $\sigma = 0.1$ regime for different L , as shown in Figs. 5.14 and 5.15. This

time the pointwise mutual information shows the causal bubble structure with one central bubble and three other in homoclinic points of the system. These points are dense, so the mutual information increases. The pointwise transfer entropy also exhibits patterns corresponding to these causal bubble areas.

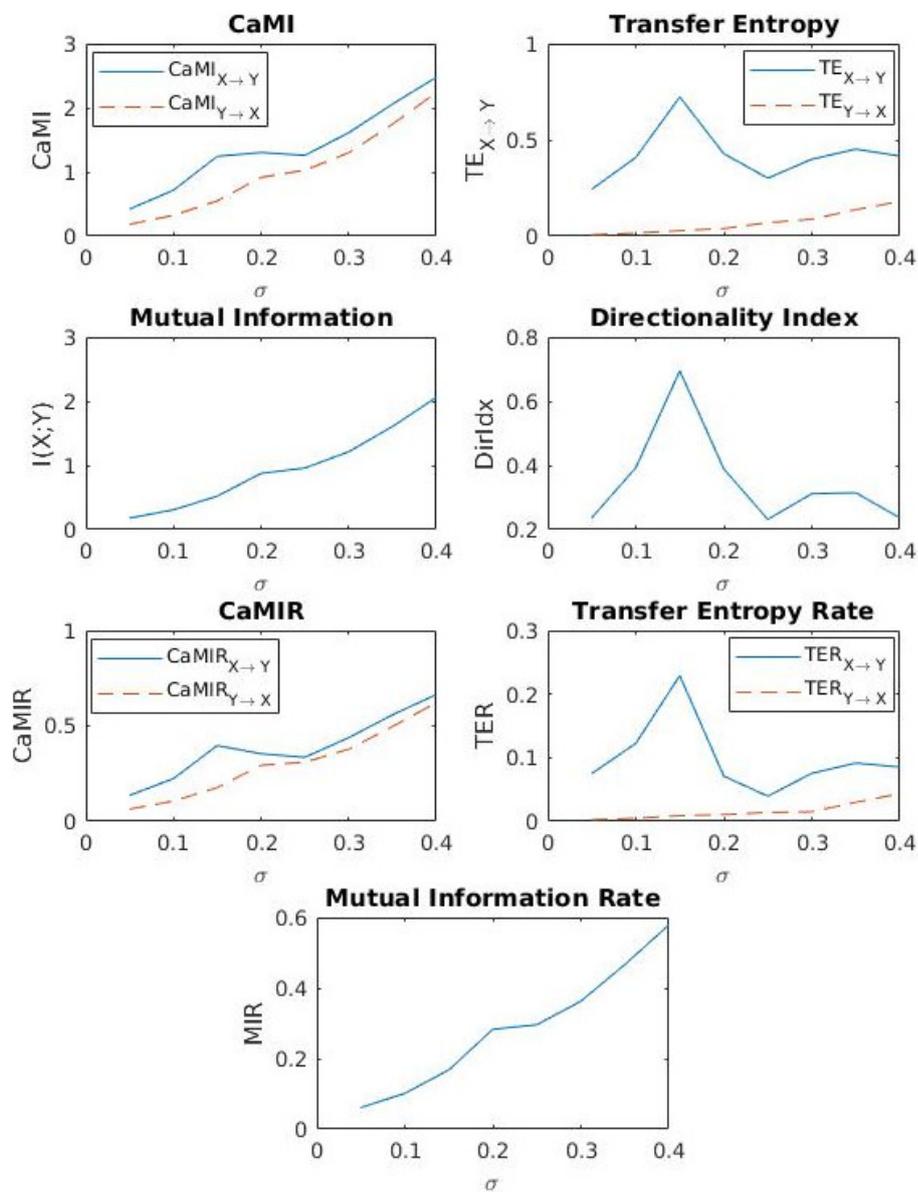


Figure 5.11: Informational measures for a time-series from coupled logistic $X \rightarrow Y$ ($r = 4$) map with CML coupling. Values for $L = 3$, $n_b = 2$ and varying coupling strength σ . A total of 10^6 non-transient time-series points are used in the analysis.

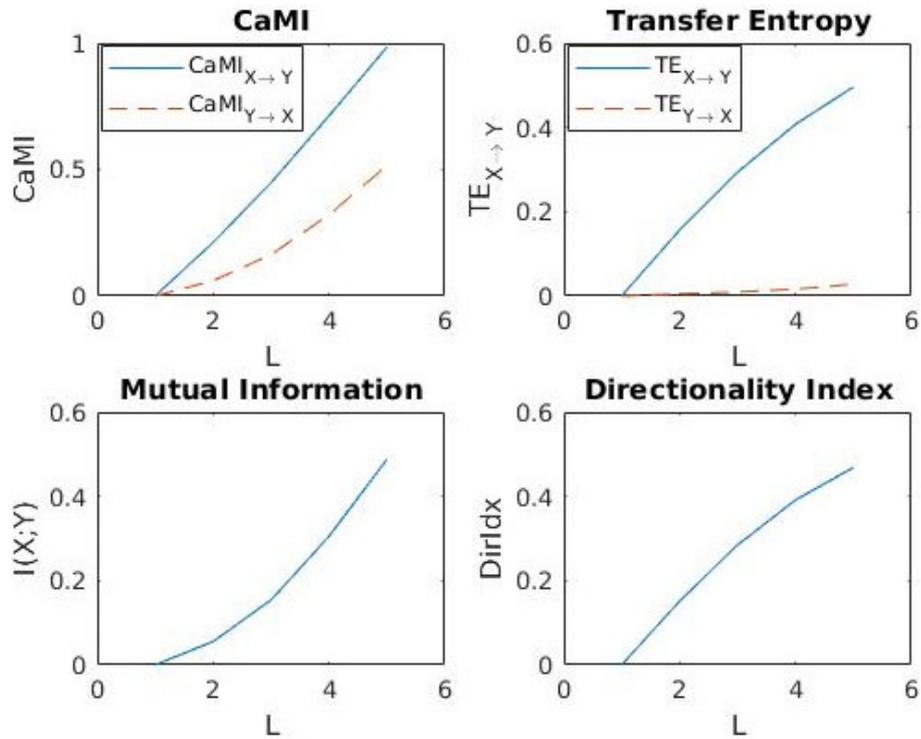


Figure 5.12: Informational measures for a time-series from coupled logistic $X \rightarrow Y$ ($r = 4$) map with CML coupling. Values for $n_b = 2$, $\sigma = 0.1$ and varying symbolic length L . A total of 10^6 non-transient time-series points are used in the analysis.

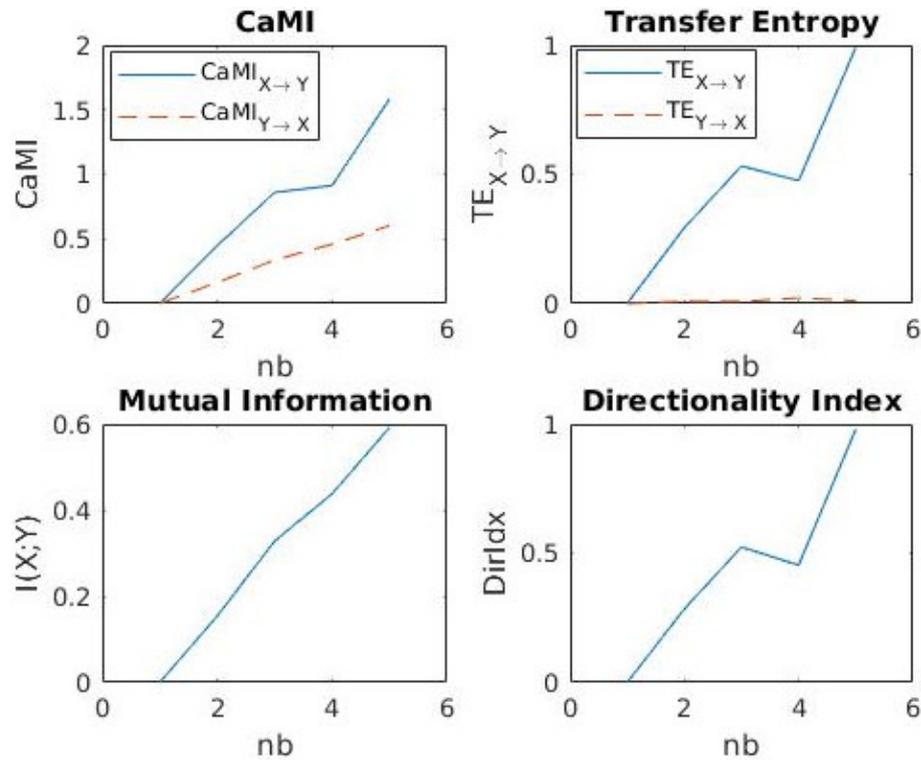


Figure 5.13: Informational measures for a time-series from coupled logistic $X \rightarrow Y$ ($r = 4$) map with CML coupling. Values for $L = 3$, $\sigma = 0.1$ and varying number of initial bins n_b . A total of 10^6 non-transient time-series points are used in the analysis.

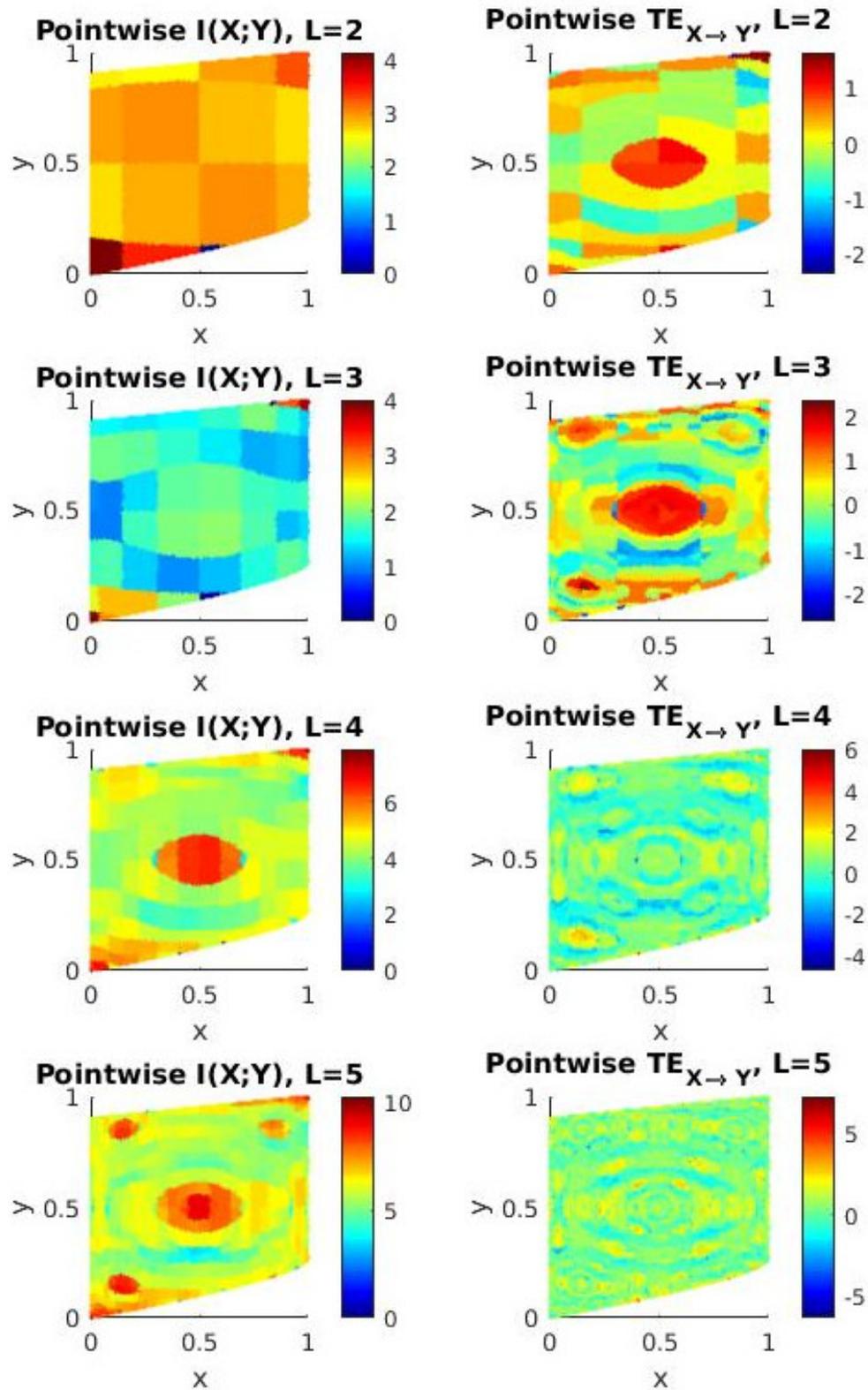


Figure 5.14: Pointwise mutual information (left column) and pointwise transfer entropy (right column) for a time-series from coupled logistic $X \rightarrow Y$ ($r = 4$) map with CML coupling. Values for $n_b = 2$ and $\sigma = 0.1$. A total of $2 \cdot 10^5$ non-transient time-series points are used in the analysis.

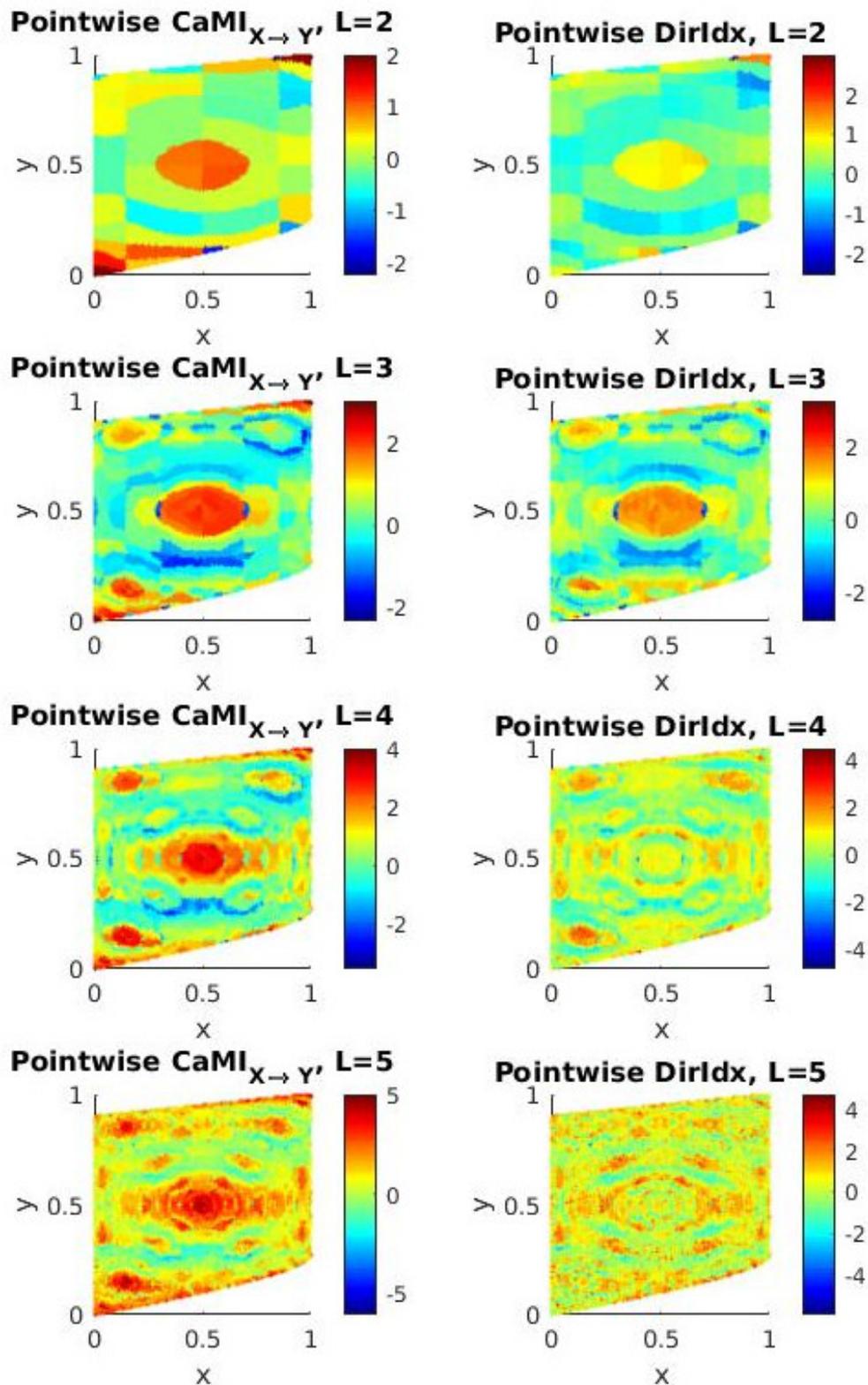


Figure 5.15: Pointwise causal mutual information (left column) and pointwise directional index (right column) for a time-series from coupled logistic $X \rightarrow Y$ ($r=4$) map with CML coupling. Values for $n_b = 2$ and $\sigma = 0.1$. A total of $2 \cdot 10^5$ non-transient time-series points are used in the analysis.

The causal bubble area has high pointwise transfer entropy in this system only up to

$L = 3$. As L increase to [4,5], the pointwise transfer entropy becomes very high on singular points, almost shadowing the formation of ring patterns of alternating positive (yellow) – zero/negative (blue) transfer entropy. On this particular case, the ring patterns is visually more evident on the pointwise causal mutual information and pointwise directionality index. This case also shows that the directionality index is not homogeneously in the causal direction, the causes for which remains to be investigated.

5.2.4 Including dynamical noise

We consider the linear diffusive coupled logistic map system $X \rightarrow Y$ from Eq. 3.6, but modify the logistic function $f(x) = rx(1 - x)$ so to include an additive noise term. This way, the new function is given by $f'(x, n) = (1 - \lambda)rx(1 - x) + \lambda\eta_n$, with λ the level of noise and η_n the specific value that a noise term from an uniform distribution [0,1] holds at iteration n . As the noise term is included in the logistic function, each node yields a different noise value at iteration n . On the low-coupling regime (Fig. 5.16), these quantities decrease very fast towards zero, already halving at $\lambda = 0.05$ noise level for the case $\sigma = 0.1$. The directionality index drops to about 10% of the original value (without noise) when $\lambda = 0.1$, making the determination of the causal direction more difficult for higher noise amplitudes.

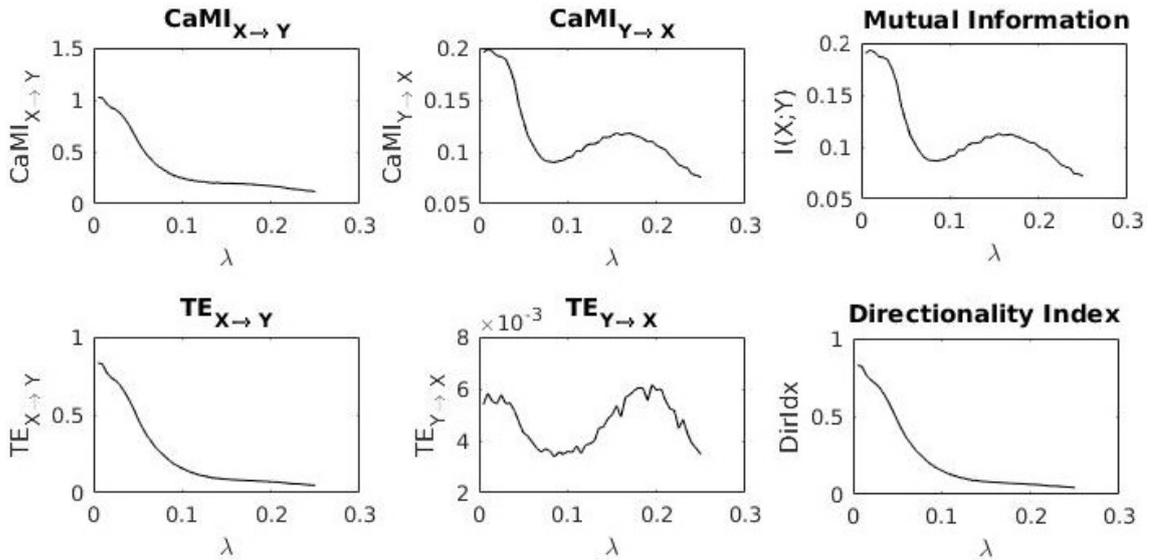


Figure 5.16: Informational measures for a time-series from coupled logistic $X \rightarrow Y$ ($r = 4$) map with diffuse coupling and additive dynamical noise. Values for $L = 3$, $n_b = 2$ and $\sigma = 0.1$ (low coupling strength) and varying noise level λ . A total of 10^6 non-transient time-series points are used in the analysis.

For the high-coupling regime (Fig. 5.17), the system is more resilient. All quantities (except transfer entropy in the opposite direction, which is naturally small) decay linearly with increasing noise λ . The quantities remain sufficiently high so that the direction of causation can be reliably identified even for $\lambda \gg 0.2$. The decay of mutual information with noise, however, proves an intrinsic problem in applications to real-world systems. There can always be an ambiguity of whether the obtained value reflects a true relatively lower coupling nature of the physical system, or if it is so because of addition of noise (in the physical process or in its measurement).

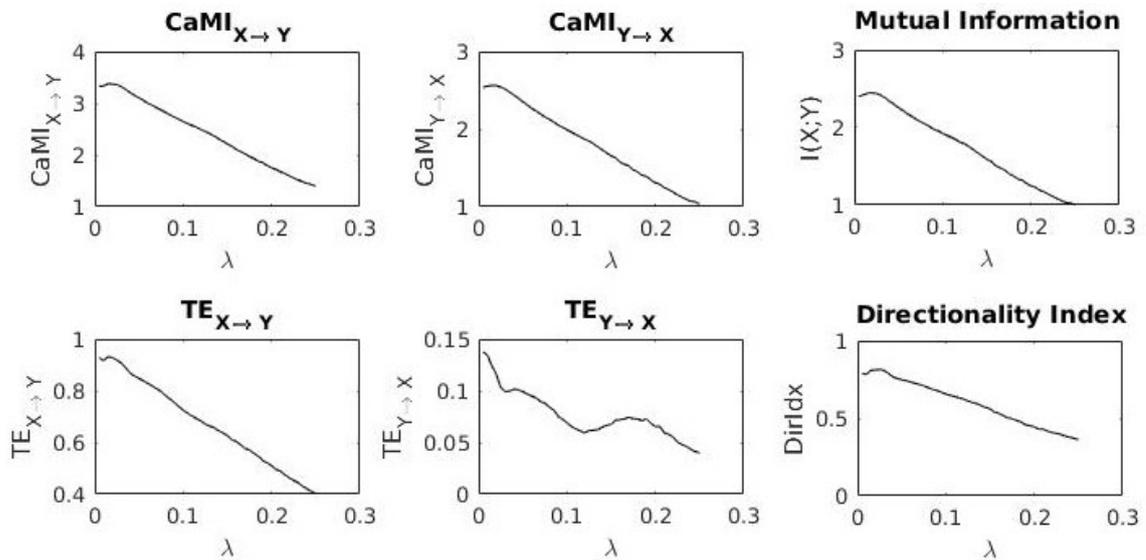


Figure 5.17: Informational measures for a time-series from coupled logistic $X \rightarrow Y$ ($r = 4$) map with diffuse coupling and additive dynamical noise. Values for $L = 3$, $n_b = 2$ and $\sigma = 0.4$ (high coupling strength) and varying noise level λ . A total of 10^6 non-transient time-series points are used in the analysis.

5.2.5 Connecting a chaotic and an intermittent system

Finally, we consider two systems: a chaotic node attempting to drive an intermittent one, and an intermittent driving a chaotic. The latter case might simulate, for a short interval of time during the laminar phases of the intermittent behaviour when the trajectory behaves “almost periodically”, the case of tides influencing a non-periodic system, such as seismic occurrence. The first case might be of interest to potential precursor candidates with the link to the seismic occurrence mediated by a system that is almost periodic at times and chaotic at other times (*e.g.* possibly strain changes). This analogy, however, is only to

illustrate a possible link to the current methodological analysis, since our interest here is only to understand how is given the flow of information when the active channel is not only given by chaotic dynamics. The case of networks with chaotic nodes at the extremes and a periodic channel uniting them will be explored in the next section. Outside seismology, this encounter applications in areas related to communication.

First, we build a linear diffusive $X \rightarrow Y$ coupled logistic map with $r_x = 4$ and $r_y = 3.6871960$ and varying coupling strength σ . This reflects a chaotic system attempting to drive an intermittent system. Fig. 5.18 shows the results. The directionality index indicates the correct causal direction. The mutual information and mutual information rate grow monotonically with the coupling strength. The transfer entropy in the causal direction is around 0.6–0.9 bits for $L = 3$ and $n_b = 2$. In the opposite direction the transfer entropy is much smaller, of 0.05 – 0.2 bits. The value of CaMI is dominated by the mutual information. The CaMIR and MIR follow a similar pattern of CaMI and mutual information. For very high coupling ($\sigma > 0.35$) the transfer entropy rate, as well as CaMIR, becomes very close in both directions, consequence of the increase of the synchronization level. Although transfer entropy rate in the causal direction becomes very close to the value in the opposite direction, the arrow of causality could still be defined in the correct direction with the absolute quantities, *i.e.* CaMI and transfer entropy.

Then, we build a linear diffusive $X \rightarrow Y$ coupled logistic map with $r_x = 3.6871960$ and $r_y = 4$ with varying σ , representing an intermittent system attempting to drive a chaotic. Fig. 5.19 shows the results. The directionality index once again indicates the correct arrow of causality. The transfer entropy in the causal direction is about 5 times larger than in the opposite direction. The relation of mutual information with the coupling strength is not trivial this time, but a growing trend still can be observed of $I(X_L; Y_L)$ with σ . In general terms, the same applies for the the mutual information rate apart from the higher coupling strength. The causal direction could also be inferred by the rates (CaMIR or transfer entropy rate).

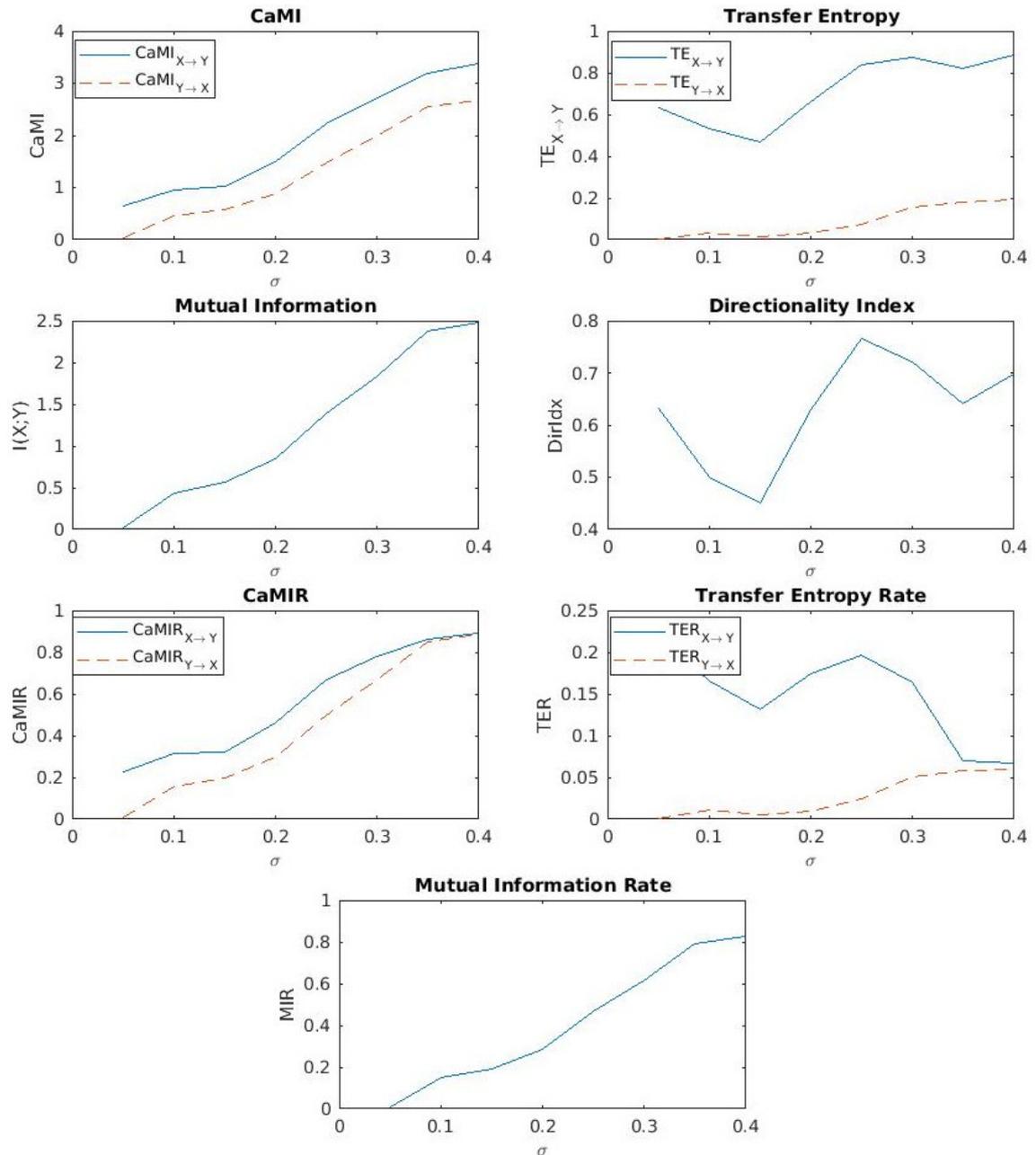


Figure 5.18: Informational measures for a time-series from coupled logistic map $X \rightarrow Y$ from chaotic to almost periodic ($r_x = 4$, $r_y = 3.6871960$), with diffuse coupling. Values for $L = 3$, $n_b = 2$ and varying coupling strength σ . A total of 10^6 non-transient time-series points are used in the analysis.

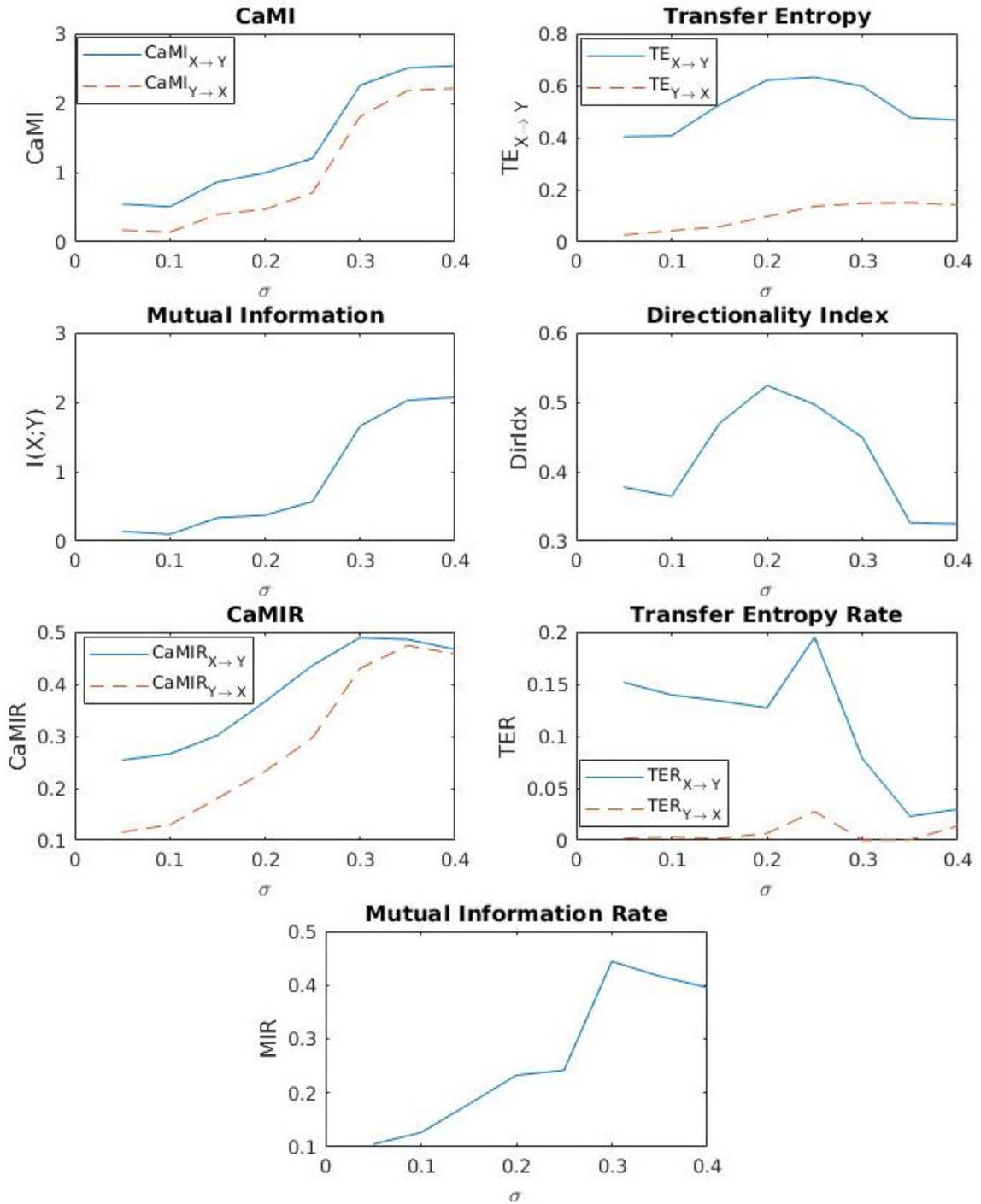


Figure 5.19: Informational measures for a time-series from coupled logistic map $X \rightarrow Y$ from almost periodic to chaotic ($r_x = 3.6871960$, $r_y = 4$), with diffuse coupling. Values for $L = 3$, $n_b = 2$ and varying coupling strength σ . A total of 10^6 non-transient time-series points are used in the analysis.

5.3 Application to logistic networks

We consider three small networks built from coupled logistic maps to investigate the case of mediated causation. This could help explain the cases of precursors which are not direct causes for earthquake occurrence, but mediated by another physical process, for example strain changes. Three topologies are considered: serial (all nodes in sequence), parallel (one ‘cause’ node and one ‘effect’ node, with several mediators connecting them), and a tree network (multiple ‘cause’ nodes connect to a common mediating channel, which then splits to multiple ‘effect’ nodes), which, for simplicity, we keep at maximum distance of 3 nodes from causes to effects. We will conduct the analysis using the information rates (CaMIR, transfer entropy rate and mutual information rate), as we have observed to provide similar results, with one less arbitrary parameter choice (L). A new directionality index was reformulated as the net transfer entropy rate instead. The networks considered were shown in Chapter 3, Figs. 3.16, 3.17, 3.18.

5.3.1 Serial network

The first network to be considered is a serial connection of ten nodes (as shown at the end of Chapter 3, Fig. 3.16) producing logistic maps with $r = 4$. Our intention is to observe how the information from the first node decays as it is transmitted along the network and how this is observed with the information measures considered.

Figure 5.20 reveals the case for when the coupling between the nodes is linear diffusive. Here we evaluate the transmission of information from node 1 to the other nodes of the network. The horizontal axis refers to the coupling strength and the vertical axis the i -th node to which the information quantity is being evaluated (*e.g.*, for $i=4$, we are evaluating the information quantities from node 1 to node 4). The colour indicates the value of the informational measure. In this case, the mutual information rate grows with the coupling strength (as investigated previously), and it is highest at the first connected node, decaying fast for the second connected and being effectively zero on the third connected. The transfer entropy rate, on the other hand, can still hold significant values up to the fifth connected node, as long as the coupling strength is sufficiently high. However,

curiously, for high coupling ($\sigma > 0.2$) the transfer entropy rate in the causal direction is highest at the indirectly connected nodes instead of the directly connected¹. Reminding that, for a $X \rightarrow Y$ system, the phase-space distribution of points starts to squeeze, eventually reaching synchronization, as the coupling strength σ approaches 0.4 (Chapter 3, Fig. 3.13), it is then expected that neighbour nodes would have transfer entropy suddenly dropping to zero due to this phenomena. More distant nodes would have the points in the state space more widespread, enabling non-zero solutions for transfer entropy. Once the transfer entropy rate in the opposite direction is almost zero for all cases, the directionality index follows the transfer entropy rate in the causal direction. The value of CaMIR is dominated by the mutual information rate.

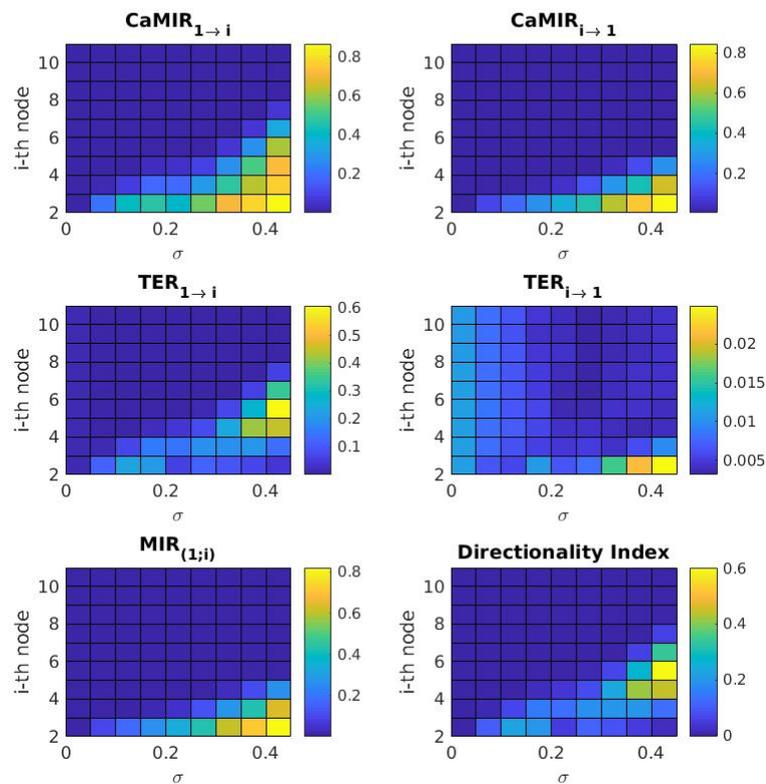


Figure 5.20: Rates of informational measures from node 1 to nodes i for a 10 node serial logistic network of diffusive coupling. The directionality index is based on net transfer entropy rate. All nodes have $r = 4$ and the time-series consists of 10^6 non-transient points. Horizontal axis is the coupling strength, vertical axis refers to the node being analysed and the colours are the values of each informational measure.

¹This phenomenon observed for this simple network is of vital importance, in particular to Neuroscience, since currently it is not fully understood how the information from a signal entering a neural network can excite specific regions of the network that are known to be not connected.

Considering a CML coupling instead, Fig. 5.21 shows that the mutual information rate grows with the coupling strength in the same way as with the linear diffusive coupling, but the transfer entropy rate in the causal direction has a sharp peak in $\sigma = 0.2$, as discussed previously. Also, the values of transfer entropy rate in the causal direction are lower and in the opposite direction are higher than the same measures in the system with linear diffusive coupling. Apart from this, it is still observed the features of the transfer entropy rate having significant values up to the fifth connected node on higher coupling strengths, and of its value not necessarily being highest at the second connected node (a relevant feature also observed in the other coupling configuration). Alas, the directionality index correctly indicates the arrow of causation, but is not as efficient as a parameter for understanding the coupling strength and the directly connected node as the mutual information. The CaMIR once again is dominated by the mutual information rate. For both coupling types the information is lost from the sixth connection onwards for the coupled logistic maps investigated. This represents a practical upper limit for indirect influence. The network with the CML coupling configuration presents an exclusive feature, that TER maximizes its value for two distinct coupling strengths σ . Additionally, mutual information (rate) decays as a functional of a topological distance, in this case, the shortest path length.

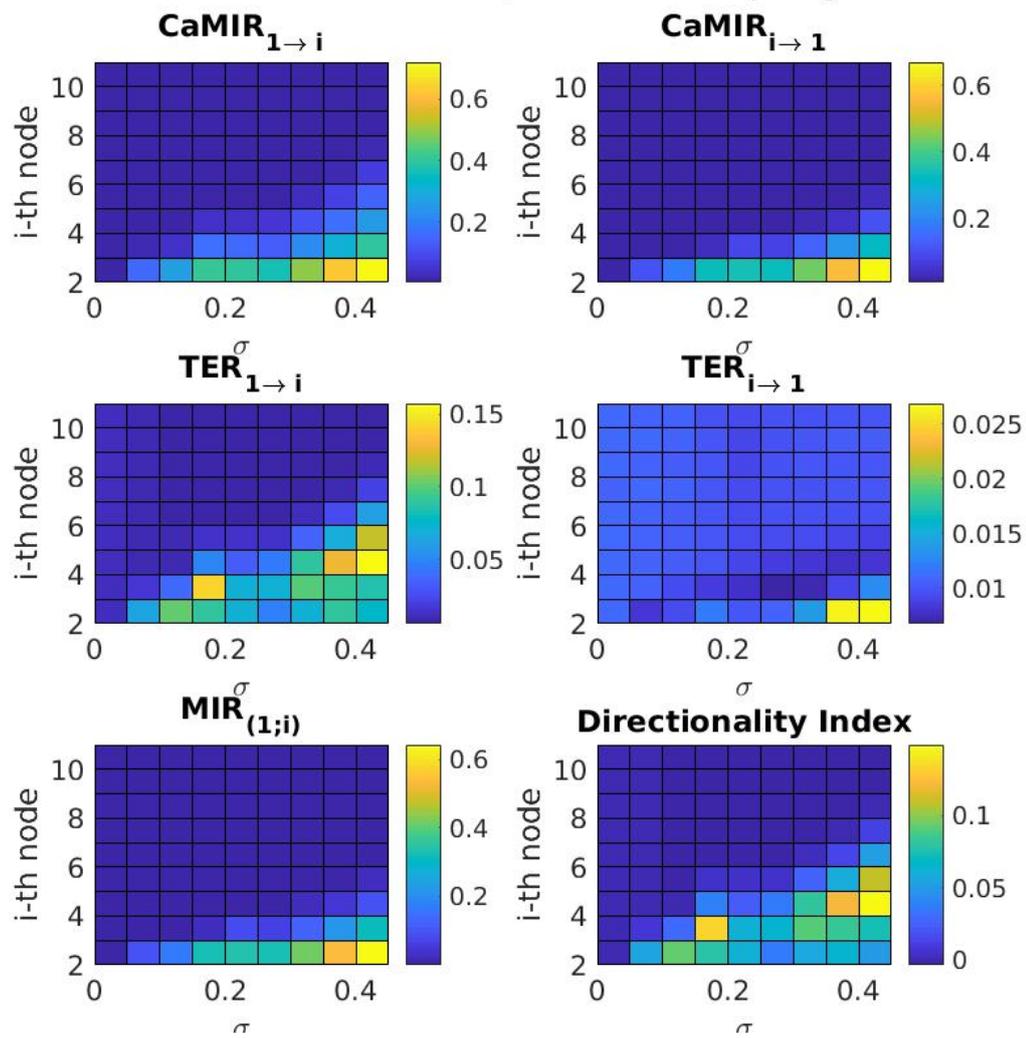


Figure 5.21: Rates of informational measures from node 1 to nodes i for a 10 node serial logistic network with CML coupling. The directionality index is based on net transfer entropy rate. All nodes have $r = 4$ and the time-series consists of 10^6 non-transient points. Horizontal axis is the coupling strength, vertical axis refers to the node being analysed and the colours are the values of each informational measure.

5.3.2 Parallel network

The second network consists of ten nodes made of logistic maps with $r = 4$: one ‘cause’ (node 1), eight ‘mediators’ in parallel (nodes 2 – 9) and one ‘effect’ (node 10). The diagram for this network was presented at the end of Chapter 3, Fig. 3.17. For a diffusive coupling, Fig. 5.22 shows how the information quantities vary with the coupling strength (horizontal axis) and from the along the network (vertical axis, measured as from the node 1 to the i -th node as in the previous case). Once again, the mutual information rate is highest for the direct coupling and increases with coupling strength. The transfer entropy rate in the causal direction presents a clear threshold: for $\sigma \leq 0.2$ the transfer entropy rate to the directly connected nodes is higher, but above that the situation inverts, so the ‘effect’ would hold higher transfer entropy with respect to the ‘cause’ than the mediators. Again, this would be influence of the squeezing of the phase-space of immediate neighbours when the coupling increases (chapter 3, Fig. 3.13).

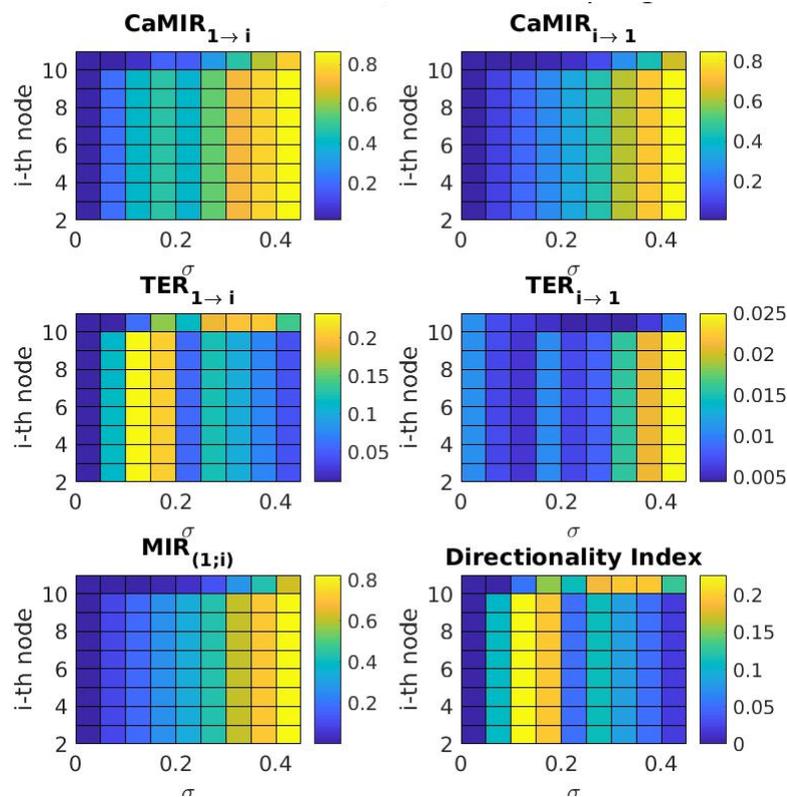


Figure 5.22: Informational measures from node 1 to nodes i for a 10 node parallel logistic network of diffusive coupling. All nodes have $r = 4$ and the time-series consists of 10^6 non-transient points. Horizontal axis is the coupling strength, vertical axis refers to the node being analysed and the colours are the values of each informational measure.

Given the symmetry of the topology chosen, the informational quantities are the same between node 1 and any mediator nodes (2-9) as expected. However, when we measure the flow of information from node 1 to node 10, it is surprising to observe that the mutual information rate from node 1 to nodes 2 to 9 are higher than the one arriving in node 10. This shows that the decay of information due to the path distance (shown in the previous section) is also acting on this case. The amount or rate of information arriving in the last node 10 is increased if the coupling is increased. However, an increase in the coupling above $\sigma = 0.2$ causes the amount of transfer entropy rate from 1 to any mediator to be smaller than the amount or rate arriving at the last node. This is again evidence of the counter-intuitive phenomenon seen in the previous section that nodes not directly connected can exchange larger rates of information, even larger than the amount exchange between nodes that are mediating the information.

Instead, if we initially select the time-series of a ‘mediator’ (*e.g.* node 4) and attempt to observe the relation of this node with the ‘cause’, the ‘effect’ and the other mediating nodes, we obtain the results on Fig. 5.23. The result highlights how information between nodes that are not connected can still be exchanged by a mediation process, which might prevent one from doing topological inference based on informational quantities. It is interesting to note that in this case the mutual information rate is highest with the other mediators rather than with the ‘cause’ or ‘effect’ (where the direct link is). Also, the transfer entropy rate in each direction correctly discerns the causing and the effect nodes, and holds zero for the other mediators. Consequently, the directionality index correctly indicates the causal flow. In general, both the transfer entropy rate and the mutual information rate have increasing values with the coupling strength. However, a peak is observed precisely at $\sigma = 0.2$ followed by a valley at $\sigma = 0.25$, which we can recall to be the threshold between the low and high coupling modes on this system.

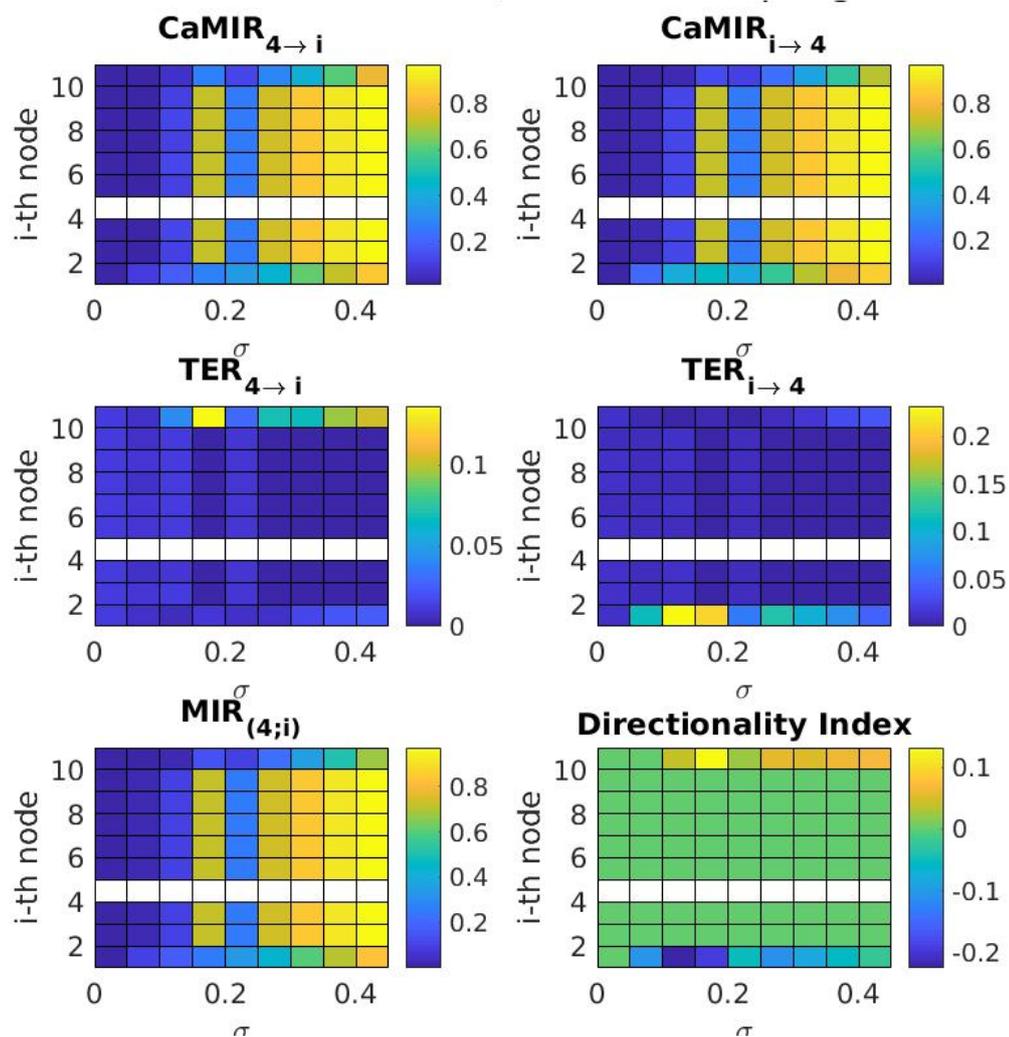


Figure 5.23: Informational measures from node 4 to nodes i for a 10 node parallel logistic network of diffusive coupling. All nodes have $r = 4$ and the time-series consists of 10^6 non-transient points. Horizontal axis is the coupling strength, vertical axis refers to the node being analysed and the colours are the values of each informational measure.

For the CML coupling, the transfer entropy between the first node and the mediators directly connected to it is greater or equal than with the last node, connected indirectly, as Fig. 5.24 reveals, and as was observed in the diffusive coupling configuration. The results on the figure also show that the mutual information is higher between the first node and the directly connected node, increasing proportionally with the coupling strength. The transfer entropy rate, however, does not necessarily increase with the coupling strength. This time the transition between the low and high coupling behaviour occurs more smoothly at $\sigma \in [0.25, 0.3]$. Below this interval the transfer entropy rate is higher for the directly connected node, and above this interval it is highest for the indirectly connected node of the ‘effects’. Note that the outcome for the CML coupling is equivalent to the one obtained for the linear diffusive (Fig. 5.22).

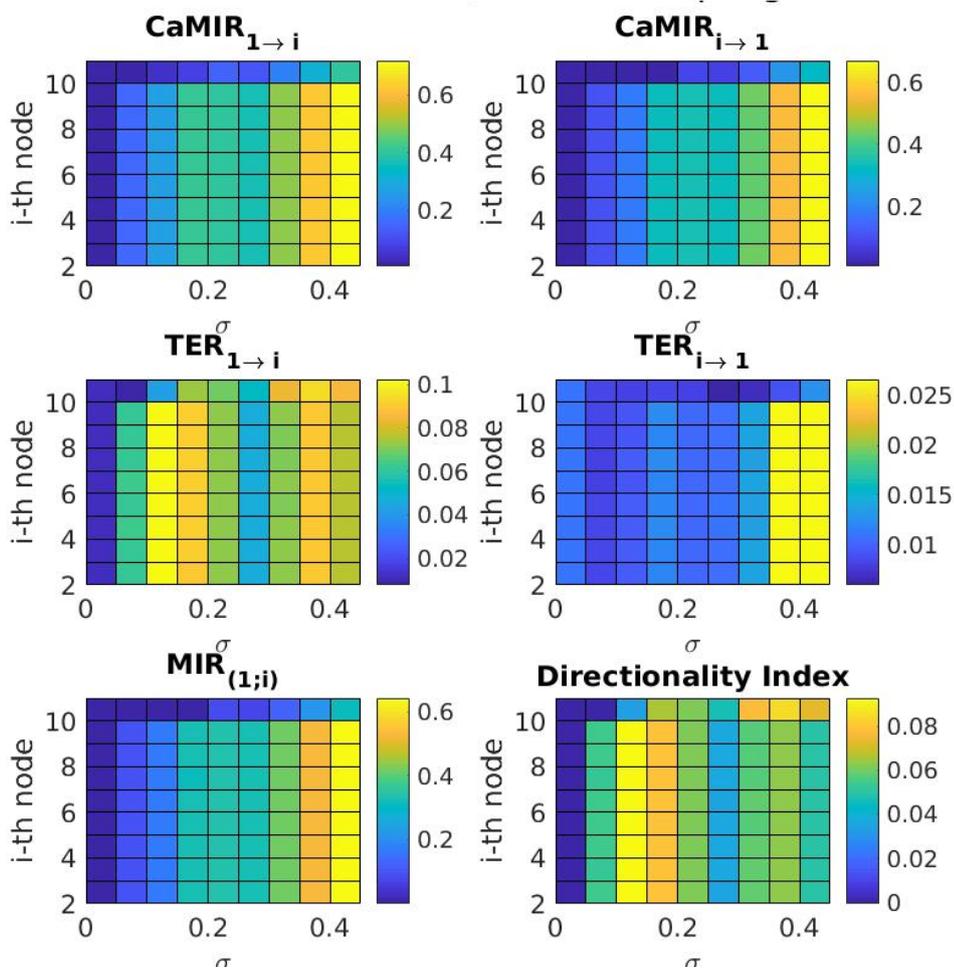


Figure 5.24: Informational measures from node 1 to nodes i for a 10 node parallel logistic network with CML coupling. All nodes have $r = 4$ and the time-series consists of 10^6 non-transient points. Horizontal axis is the coupling strength, vertical axis refers to the node being analysed and the colours are the values of each informational measure.

The information measures between a mediator node (node 4) and the other nodes, for a CML-coupled parallel system, is given on Fig. 5.25. Just as the case with the linear diffusive coupling, the mutual information is higher with the other mediators, the transfer entropy is greater with the directly connected nodes, and the direction of the causal flow is correctly identified. The difference between the CML and the linear diffusive coupling is that the peak in the mutual information rate and valley in transfer entropy rate at $\sigma = 0.2$ is slightly shifted to $\sigma = 0.25$. Also, the transfer entropy rate in the with the ‘effect’ node (node 10) is smaller than in the linear diffusive case.

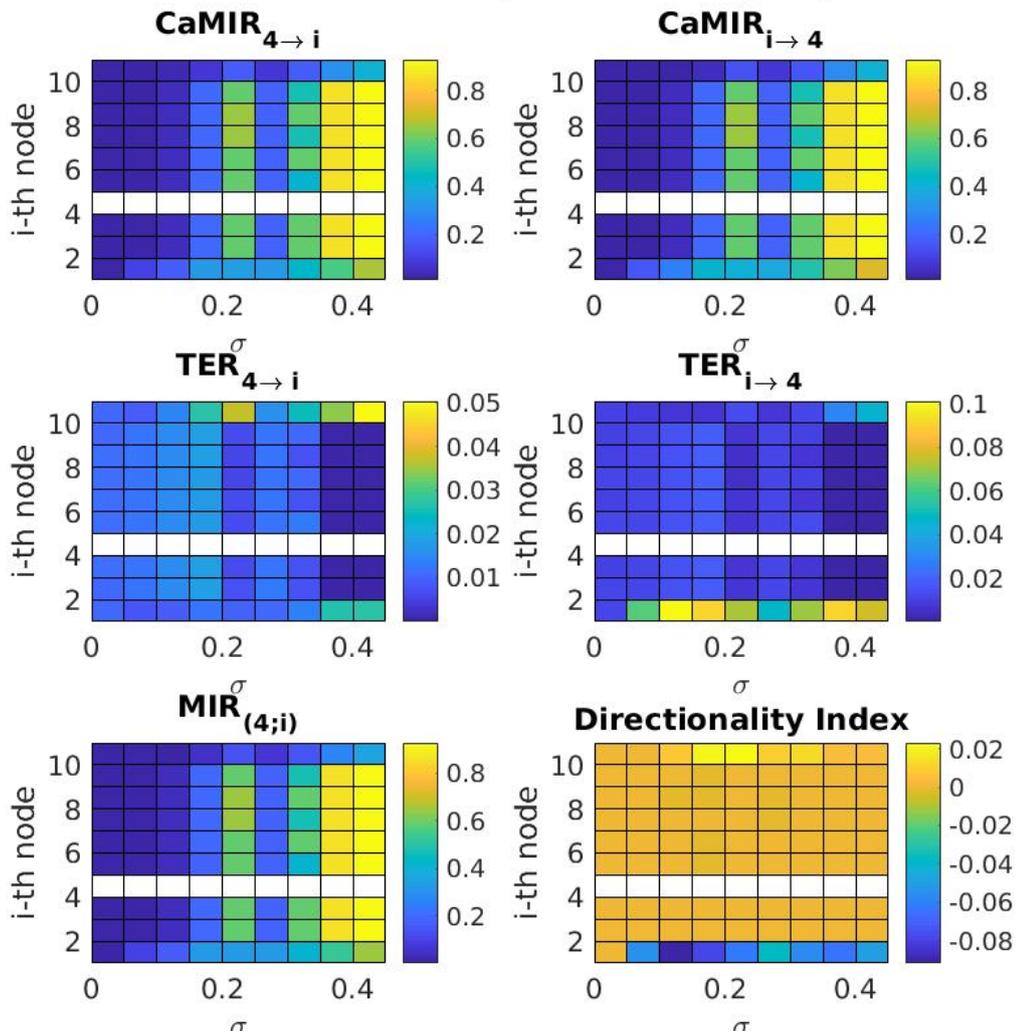


Figure 5.25: Informational measures from node 4 to nodes i for a 10 node parallel logistic network with CML coupling. All nodes have $r = 4$ and the time-series consists of 10^6 non-transient points. Horizontal axis is the coupling strength, vertical axis refers to the node being analysed and the colours are the values of each informational measure.

5.3.3 Directed tree network with maximum distance 3

The last topology to be considered is a tree network where: node 1 and node 2 are two different ‘causes’, node 3 is a common mediator, node 4 is connected in serial to node 3 (hence nodes 3–4 constitute a channel like a chaotic communication link), node 5 and node 6 are ‘effects’ connected to node 4. A diagram was presented at the end of Chapter 3, Fig. 3.18. All nodes are logistic $r = 4$ maps. With this topology we investigate what are the consequences of, in one side, combining two information sources, and on the other, splitting the information transmission into two ‘effect’ components.

For a diffusive coupling, we analyse the case of transmission of information from ‘cause’ node 1 to the other nodes, obtaining results on Fig. 5.26. The results correctly show no information link with node 2, which is an independent ‘cause’. For the remaining nodes the interpretation is similar to the serial network: the mutual information rate is highest in the directly connected node (node 3) quickly dropping to zero for the others, and the transfer entropy rate in the causal direction is highest for the directly connected node on low coupling ($\sigma \leq 0.2$) and for the indirectly connected nodes on higher couplings ($\sigma > 0.2$).

Care must be taken when observing the $TER_{i \rightarrow 1}$ (opposite to the flow): the highest value, which appear to occur for the two disconnected nodes (1 and 2) or for uncoupled system ($\sigma = 0$) is actually the error arising from working with finite time-series, as described in the beginning of the chapter. The fact the other values are lower only confirms that $TER_{i \rightarrow 1} = 0$, as expected. As in the serial case, for the highest coupling strengths the transfer entropy rate in the causal direction ($TER_{1 \rightarrow i} = 0, i > 2$) remain significant until the nodes of the effects. The results of this network being equivalent to the serial case indicates that the presence of additional (possibly unknown) causes or effects does not interfere with the measurement of causality between the (known) variables under investigation.

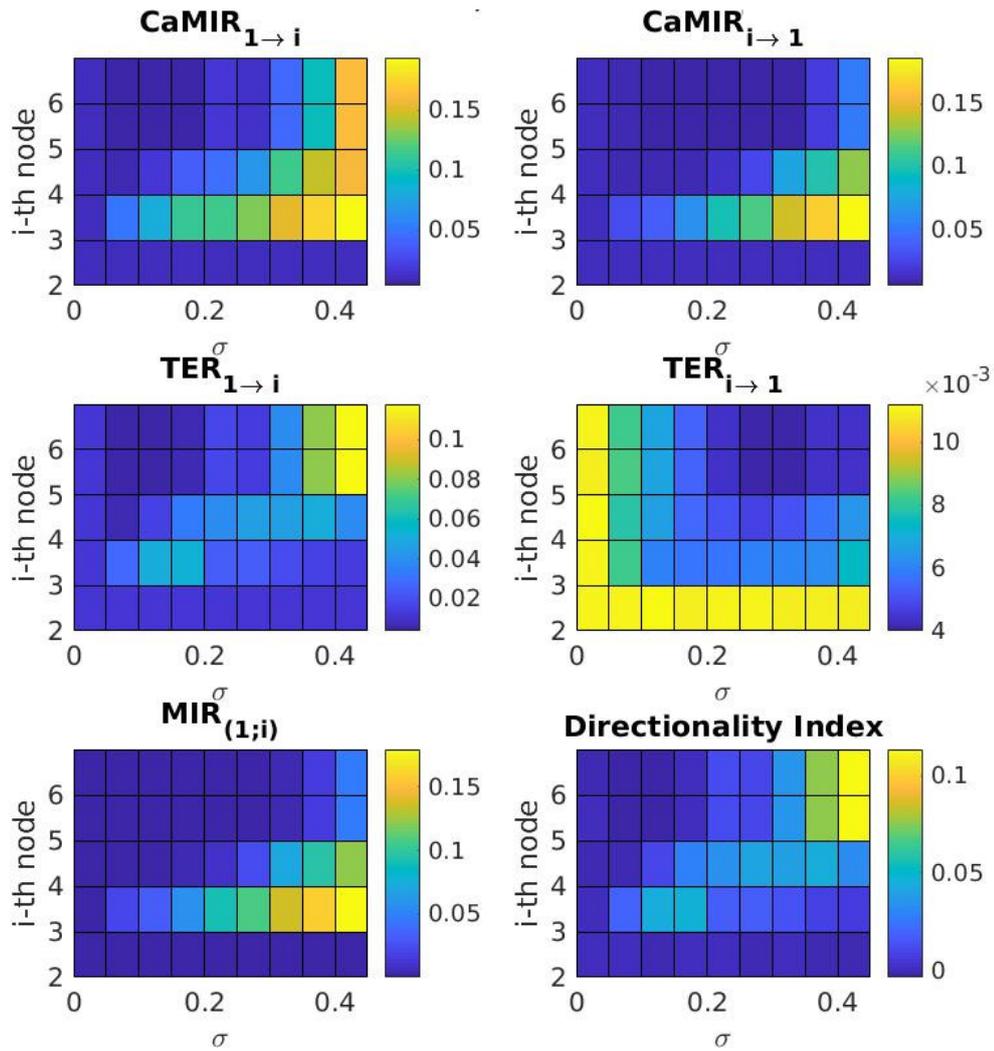


Figure 5.26: Informational measures from node 1 to nodes i for a 6 node distance-3 tree logistic network of diffusive coupling. All nodes have $r = 4$ and the time-series consists of 10^6 non-transient points. Horizontal axis is the coupling strength, vertical axis refers to the node being analysed and the colours are the values of each informational measure.

If the coupling is of CML type, the results (Fig. 5.27) are similar to the linear diffusive case. Node 2 is correctly indicated as unconnected to node 1, the mutual information between node 1 and node 3 is high and decreases significantly for the other nodes. The mutual information rate grows monotonically with the coupling strength, but the transfer entropy rate does not, presenting a peak in $\sigma = 0.2$. A threshold is observed on $\sigma = 0.3$ where the transfer entropy rate is systematically higher for the indirectly connected nodes. For the CML type the results are also similar to the serial connection, meaning that the study of causality will not be impacted by lack of knowledge of all possible ‘cause’ or ‘effect’ physical events involved in a process. In special, it implies that we do not have to

know all other seismic precursors to analyse if a candidate can be a precursor.

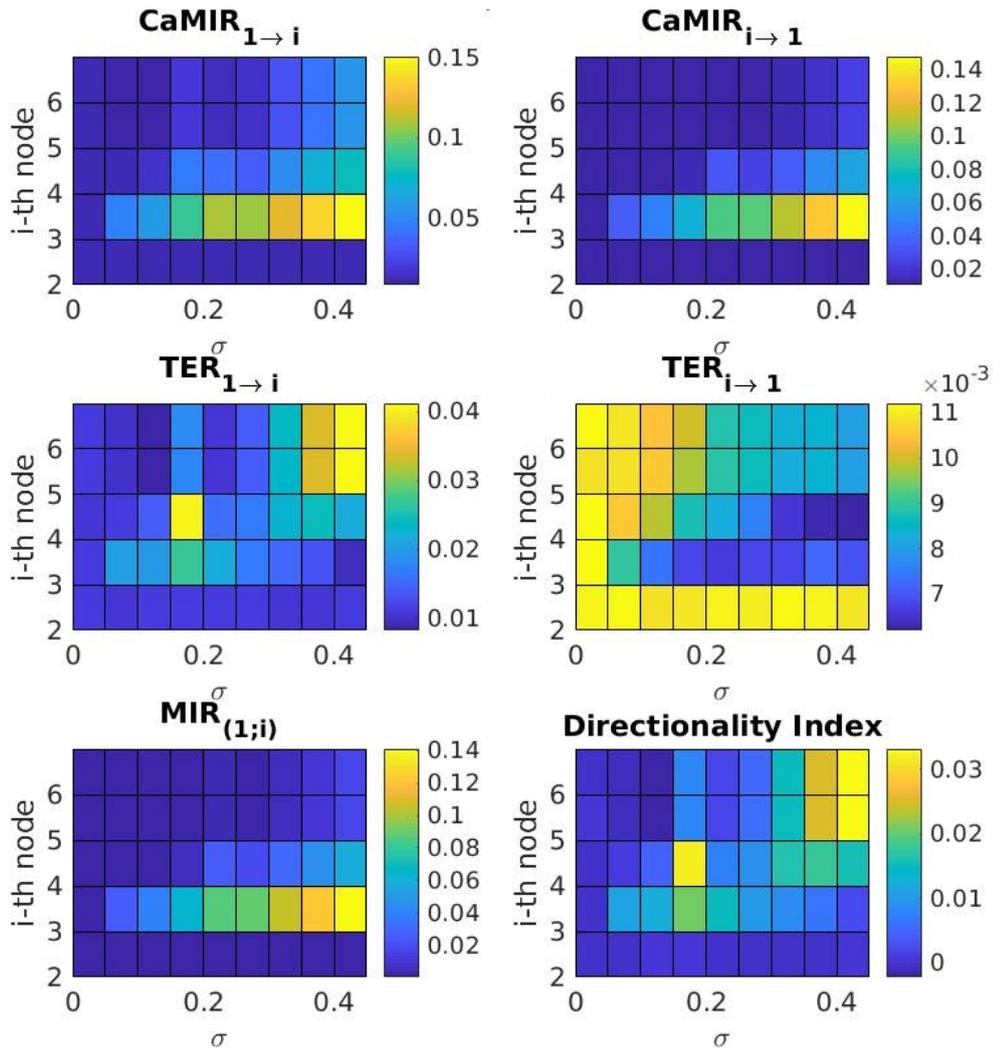


Figure 5.27: Informational measures from node 1 to nodes i for a 6 node distance-3 tree logistic network with CML coupling. All nodes have $r = 4$ and the time-series consists of 10^6 non-transient points. Horizontal axis is the coupling strength, vertical axis refers to the node being analysed and the colours are the values of each informational measure.

There is one point of exclusivity about the CML type, which is that it enhances the “multichannel” phenomenon: large levels of information can be transmitted for different values of the coupling and for different pair of nodes. If the coupling has a particular value, information is transmitted highly from node 1 to 4, for $\sigma = 0.2$, but if σ is 0.4, TER is higher from node 1 to nodes 5 and 6.

5.3.3.1 A variation: periodic channel

A variation is proposed on the tree network with distance 3, consisting of turning the chaotic channel formed by nodes 3–4 into a periodic channel. The main application is

for communication systems with strongly amplitude damping factor², where this channel could be a carrier for the information from the chaotic nodes 1 and 2. However, it is also relevant for our study on seismic precursors if it is considered that the strain/stress changes on the Earth crust is an almost periodic process (mainly dictated by tides, which is almost periodic). Therefore, all seismic precursor candidates considered to be mediated by this link work as a small network with periodic channel.

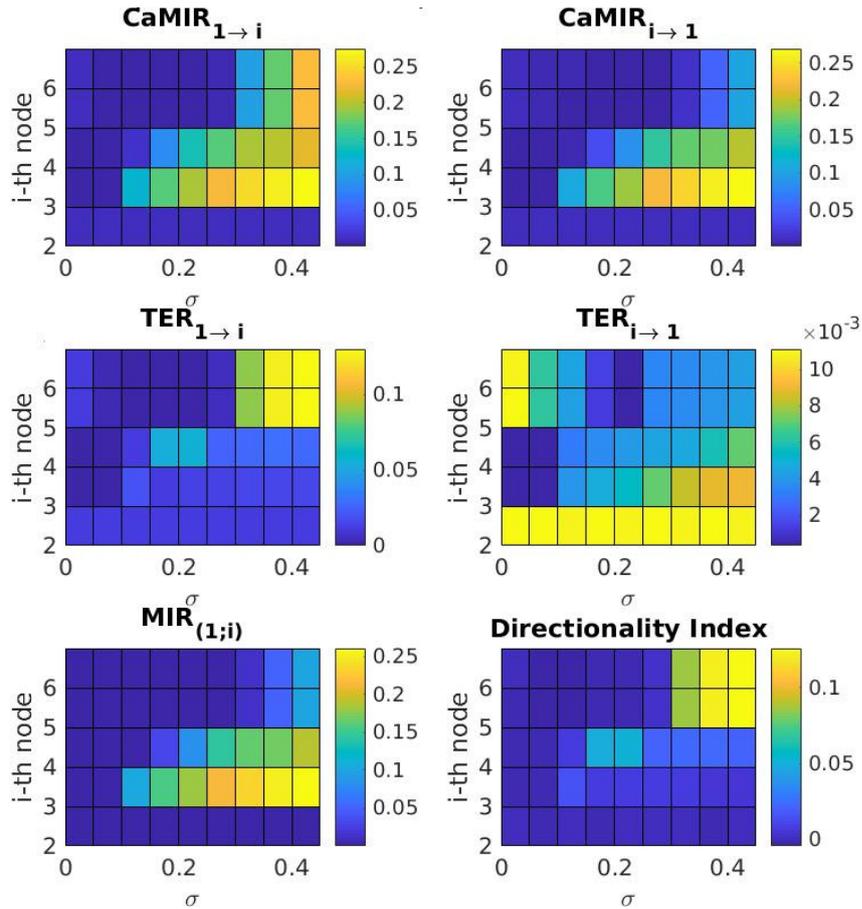


Figure 5.28: Informational measures from node 1 to nodes i for a 6 node distance-3 tree logistic network of linear diffusive coupling. ‘Cause’ nodes (1–2) and ‘effect’ nodes (5–6) have $r = 4$ and the mediating channel nodes (4–5) have $r = 3.3$ (period-2 solution). The time-series consists of 10^6 non-transient points. Horizontal axis is the coupling strength, vertical axis refers to the node being analysed and the colours are the values of each informational measure.

For simplicity, we have selected $r = 3.3$ for the logistic maps of the periodic channel (nodes 3 and 4), which is a period-2 solution, and the chaotic $r = 4$ solution for the ‘cause’ (1 and 2) and ‘effect’ (5 and 6) nodes. The coupling type is linear diffusive. The results

²or to communication in neurosystems, when neurons behave periodically

(Fig. 5.28) are similar from those obtained for the chaotic channel, with the only observed difference of the values of mutual information rate being higher with periodic channel than with chaotic. This is an initial indicator that the type of the mediating channel does not influence the results significantly. The multichannel communication phenomenon observed previously is maintained in a periodic channel. These results are expected, since the periodic channel does not destroy the information, but only transforms it into a fractal set where information is more complicated to be realized [129].

5.4 Discussion of results

In this chapter our theoretical approaches to study causality have been tested against a number of numerically generated systems.

As expected, the numerical error increases with the length of the symbolic sequence L and decreases with the number of time-series points available. We do not advise using this method with fewer than 1000 points, once the errors for CaMI, transfer entropy and their rates would be in the first decimal place, comparable to the expected values.

We notice in coupled logistic network systems that the mutual information increases monotonically with the coupling strength, but the same is not true for the transfer entropy, which is highly sensitive to the coupling type. The dependence of CaMI, transfer entropy and mutual information with L typically present a linear portion for which their rates can be calculated. The variation with the number of initial bins n_b , however, is typically not linear, although generally increasing. The CaMIR and the transfer entropy rate revealed as effective in presenting the correct arrow of causation as the CaMI and transfer entropy, with the advantage that it removes the arbitrary user selection of a symbolic length. The mutual information rate in general follows the same characteristics as the mutual information, hence also enabling substitution.

In all studied cases this directionality index indicated the correct direction of causation, but not necessarily is highest at the directly connected node.

On logistic networks, the mutual information rate is typically highest at the directly connected node and increases with the coupling strength. The transfer entropy rate, can

be highest in distant nodes, especially, but not exclusively, for higher coupling strengths. An exception is when observing a mediator node in a parallel network. It usually has a large amount of shared information with other mediators, despite they not being directly connected.

An adequate way of distinguishing the parallel from the serial case is to observe the mutual information rate in conjunction with the transfer entropy rate. If the transfer entropy rate is zero the nodes are not connected. If the transfer entropy is significant, then the mutual information rate should be observed, and typically it holds significant values only if the link is direct. We expected network inference to be possible by this procedure.

The distance-3 tree networks have shown results similar to the serial network regarding the flow from node 1 to nodes 3–6. This means that there is no significant implications in the analysis of a causing system or precursor if there is the influence of another ‘cause’ or precursor which is unknown or which the time-series could not be acquired. No significant changes are observed if the channel is periodic instead of chaotic, except for larger mutual information rates between the nodes.

When the coupling strength is too high ($\sigma > 0.3 - 0.4$) a coupled logistic map may start approaching synchronisation, as observed by the compression in the phase space plots on Chapter 3 (Fig. 3.13). In this process, the time-series of directly connected nodes become more similar, so the transfer entropy rate between them drops and the mutual information rate increases. However, a node at a distance-3 is also under the effects of the dynamics of the mediator, so the impact of the beginning of the synchronisation is reduced, and the transfer entropy may remain high (or even increasing with σ). Only by observing the time-series without physical indication of the coupling strength of the system at hand, this special case might lead to incorrect interpretations of two nodes being in parallel when they can be two directly connected nodes with high σ .

Chapter 6

Causal analysis of earthquake precursor candidates

In this chapter we apply the information theoretical methods defined on Chapter 4 to the real-world seismic data. Four possibilities are considered:

- Tidal variations triggering large earthquakes,
- Variations of Gutenberg-Richter's b-value preceding the occurrence of large earthquakes,
- Pre-seismic gravity changes,
- Preceding anomalous seismicity leading to new events (*e.g.* foreshocks, quiescence).

For the analysis of the precursor candidates we select data from the area around the Japan trench. This fault system results from the Pacific tectonic plate moving under the Okhotsk plate, and is selected for having one of the highest rates of large earthquakes observed on the planet. More importantly, there are also IGETS gravity stations nearby. Of particular interest is the Matsushiro station (MA), located about 60km from the Japanese east coast, or 200km from the trench. The time-series from this station provides information of locally observed astronomic tides, as the raw gravity signal is mainly composed by them. Also, as discussed in Chapter 3, the gravity residuals can be obtained by removing

the tides from the original signal. Hence a link between these residuals and the occurrence of seismic phenomena can be also investigated (pre-seismic gravity changes).

The data about seismic occurrences is obtained from IRIS catalogue, specifying the area of interest as the box of latitude range [34.98, 42.48] and longitude range [138.75, 147.02] (in decimal degrees). The time-series of the b-value is calculated from the seismic occurrences data as previously described on Sec. 3.1.1. The gravity data is obtained directly from the IGETS database¹, where the CORMIN (minute-sampling, corrected for local errors) data of MA station is selected. The period of analysis is 01 July 1997 to 30 June 2008, which corresponds to the interval of the available gravity data. A total of 321925 seismic events are registered in the period.

A typical qualitative definition to classify the events by magnitude considering large earthquakes as those with $M_w > 5.5$ (122 cases in the period). However, we consider such definition arbitrary, once it is possible that a precursor is associated with events with magnitudes different from this threshold. The data itself should be able to indicate what is the optimal threshold. We will find this on a case-by-case basis by means of performing the calculation for different thresholds of precursor candidate anomaly and earthquake level (selection different locations of the initial partition divisions), and choosing the combination which maximizes CaMI. However, we can only consider an event to be an anomaly if it is at least two standard deviation distant from the average (probability of occurrence smaller than 5% in a normal distribution). So, in the range of thresholds tested we start with this physical constraint, and increase progressively the thresholds both in the precursory signal and in the seismicity indicator.

An outcome in favour or against the existence of the causal flow of information between a precursor candidate and a seismicity behaviour does not unequivocally confirm or reject such precursor candidate. Instead, it only provides an additional layer of analysis, which could assist the arguments in favour or against the hypothesis in discussion. It should be considered that it only refers to properties observed in the time-series, but the (geo)physical theory for the precursor and laboratory testing are still essential. Also,

¹<http://isdc.gfz-potsdam.de/igets-data-base/>

although here we test for only four candidates, there are other precursor candidates that can be investigated with this method, but we do not explore mainly due to lack of data access or insufficient data. Once long time-series of other precursor candidates become easily available, these candidates can be equally tested.

To enable comparison, our sampling time is fixed at 1 day for all time-series, which leads to 4018 data points over the whole period. Two uncorrelated uniformly distributed random datasets with 4018 data points lead to a CaMI and transfer entropy for $L = 2$ in the order of 0.008 bits and a mutual information of the order of 0.002 bits with our method. For $L = 3$, the CaMI and transfer entropy becomes of the order of 0.1 bit, and of mutual information 0.01 bit for this uncorrelated uniform random system. For $L = 4$, the values of CaMI and transfer entropy are of the order of 1 bit and of mutual information of the order of 0.6 bits for such uncorrelated uniform random dataset. These values can be interpreted as conservative error margins, or, most appropriately, the margins of confidence to which we can distinguish a system from a uncorrelated random uniform distributions. Other distributions are expected to lead to values smaller than this, hence the interpretation as a conservative error margins. Further increasing L leads to values (of these error margins) too large, so we will not consider here. For CaMIR and transfer entropy rate calculated in the interval $L \in [2, 4]$, the numerical error for 4018 data points is of the order of 0.5 bits/iteration and of mutual information rate of 0.03 bits/iteration. As the error levels in CaMIR and transfer entropy rate are considered too large, given the number of points available, we adopt CaMI and transfer entropy for this chapter. It would be convenient to perform in future the analysis using the rates, once a time-series of 30 years of precise gravity data (or another measure of interest) becomes available, enabling over 10000 points for the causal analysis.

The precursor candidates we are going to test in the following sections are summarised by the diagram of Fig. 6.1. It will be evaluated the response delay to which observed peaks in the informational measures occur, in an attempt to clarify the time-range to which a precursor candidate may signal a seismic occurrence.

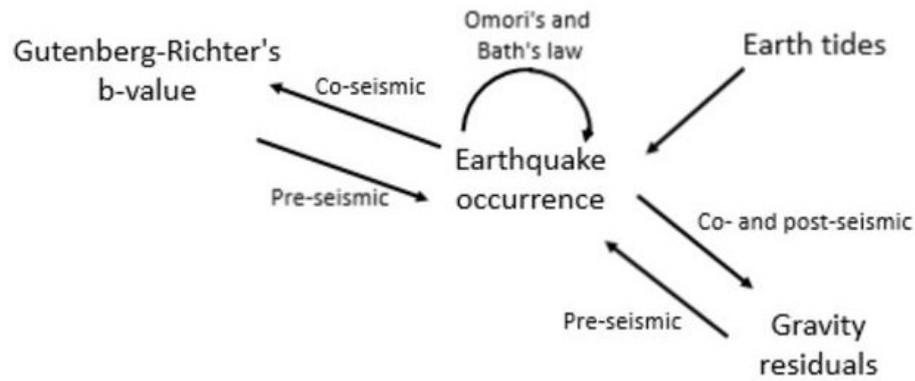


Figure 6.1: Expected links between the earthquake precursor candidates investigated and the earthquake occurrence.

6.1 Tidal triggering of earthquakes

The first hypothesis consists on the maximum push and recoil amplitudes of tides on Earth acting as earthquake triggers. According to the view, the gravitational pull from the Moon and Sun deforms the crust (directly or indirectly, through the weight of the changing ocean levels), leading to strain and stress variations at geological faults that potentially surpass rupture point.

We obtained the locally observed tides in the Japan area from the gravity time-series of Matsushiro MA station. Tidal variations occur on a daily/twice-a-day basis, but the amplitude can vary significantly, for example, at intervals of about 14-days (fortnightly tides), corresponding to half the period of the orbit of the Moon, or with the arrival of a storm surge changing the ocean levels (and influencing gravity). Higher and lower envelopes were designed to fit the daily peak and valleys from the raw gravity data, and the daily tidal amplitude was defined as the maximal difference of these envelope functions in a day (Fig. 6.2). This observed tidal amplitude is considered the precursor candidate signal.

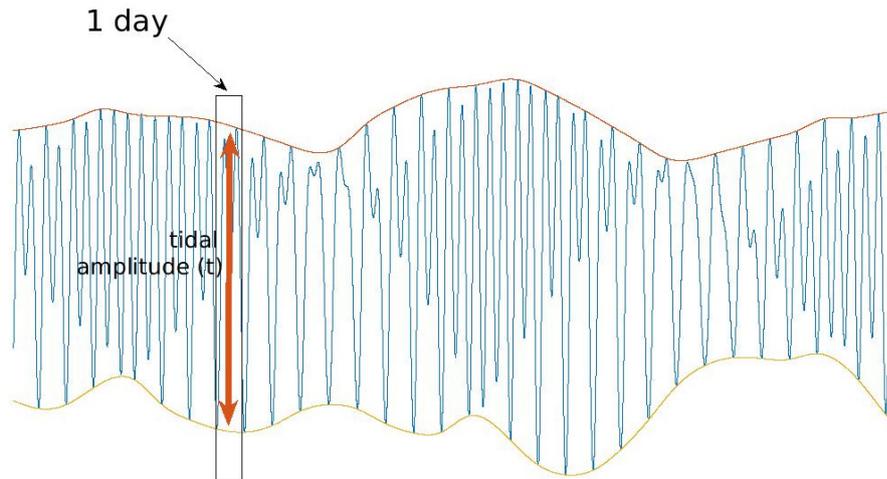


Figure 6.2: Illustration of method for obtaining the daily tidal amplitude for the gravity signal.

The first question is whether high tidal amplitudes lead to high magnitude earthquakes, and, if so, what is the time-delay between a high tide and an earthquake. Our method for determination of information quantities is applied to the observed tidal amplitudes and the time-series of the maximum earthquake magnitude observed in a day. A delay between both time-series is considered, up to 30 days. Also, the magnitude defining a small or a large event and the tidal amplitude defining the threshold of a small and a high tide are varied until finding the combination maximising CaMI. This combination of thresholds is tidal amplitudes of $2.17\mu\text{m/s}^2$ and magnitude $M_w 5.6$, with values of information measures varying with respect to delay as given in Fig. 6.3. Maximal values are observed for delays 5 and 25 days.

Reminding that the CaMI and transfer entropy for uncorrelated uniform random distributions are of 0.008 bits for $L = 2$, 0.1 bit for $L = 3$ and 1 bit for $L = 4$. So all causal information quantities shown on Fig. 6.3 are below this margin of confidence, *i.e.*, in principle consistent with zero. For mutual information, the margin is of the order of 0.002 bits for $L = 2$, 0.01 bit for $L = 3$ and 0.6 bit for $L = 4$. As the results for $L = 2$ are below 0.02 and for the other delays are below 0.01 bit, the mutual information is also consistent with zero. The optimal thresholds leading to maximum CaMI correspond to the minimum allowed threshold possibility (*i.e.*, equal to the average plus 2 standard deviation).

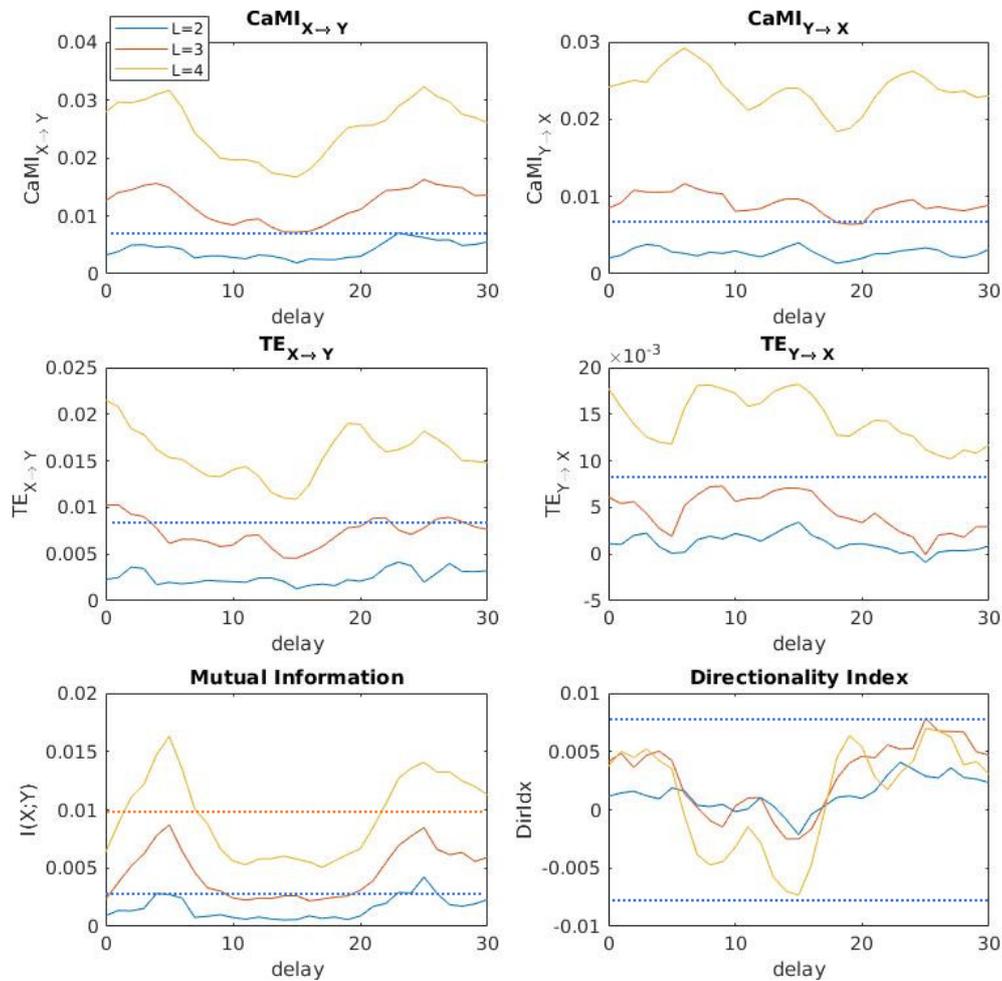


Figure 6.3: Information-theoretical values for the hypothesis that highest tidal amplitudes (X) induce earthquake events (Y). The optimal partition line of gravity tidal amplitude is $2.17 \mu\text{m/s}^2$ and maximum daily magnitude $M_w = 5.6$. The dashed lines represent the confidence margins. If a confidence margin is not shown, its value is beyond the scale of the plot.

The pointwise informational quantities are given in Appendix C. In general the transfer entropy and directionality index between the tidal amplitudes and the maximum daily magnitude is zero. Few individual data points, though, may present higher pointwise mutual information and pointwise transfer entropy, and interestingly they tend to fall in the high daily magnitude range, which is desired. But from the available data is not possible to identify these points as forming a subregion of the precursor domain. Without such localisation, a precursor candidate cannot be considered for prediction.

Another possibility is that, instead of the large tides leading to singular large events, the causality could be over a cumulative energy release (either one large event or many smaller that in the sum represent a large release). Therefore, we consider the same analysis of the tidal amplitudes but now against the time-series of cumulative daily magnitude. The optimal partition divisions are $2.37\mu\text{m/s}^2$ for the tidal amplitude (which is above the mean plus two standard deviation) and $\sum M_w = 102.2$ for cumulative daily amplitude (which is equal to the mean plus two standard deviation). Only magnitudes $M_w \geq 3$ are considered in the cumulative sum, once the catalogue might be incomplete below this level (local events, beyond reach or sensitivity of instruments). For the optimal thresholds, the results of the information quantities in function of the time-delay between the time-series are shown on Fig. 6.4.

The maximum of the flow of causal information is observed in the delay of 10 days. The results of the causal information quantities (CaMI and transfer entropy) are smaller than the comparison with the uncorrelated uniformly distributed random system (0.008 bits for $L = 2$, 0.1 bit for $L = 3$, 1 bit for $L = 4$), hence consistent, in principle, with zero. However, the results for the mutual information are of the same level as the error (0.002 bits for $L = 2$, 0.01 bits for $L = 3$), apart for $L = 4$ due to insufficient data (error 0.6 bits). The possibility of flow of information between these two variables is not discarded, but it is not possible to confirm with the amount of available data. The pointwise information measures for this system, displayed in Appendix C, clarifies that the high mutual information contributors are associated with the high tide (mostly above $2.4\mu\text{m/s}^2$) and high cumulative magnitude (above $\sum M_w = 100$), which is a desired behaviour for a precursor

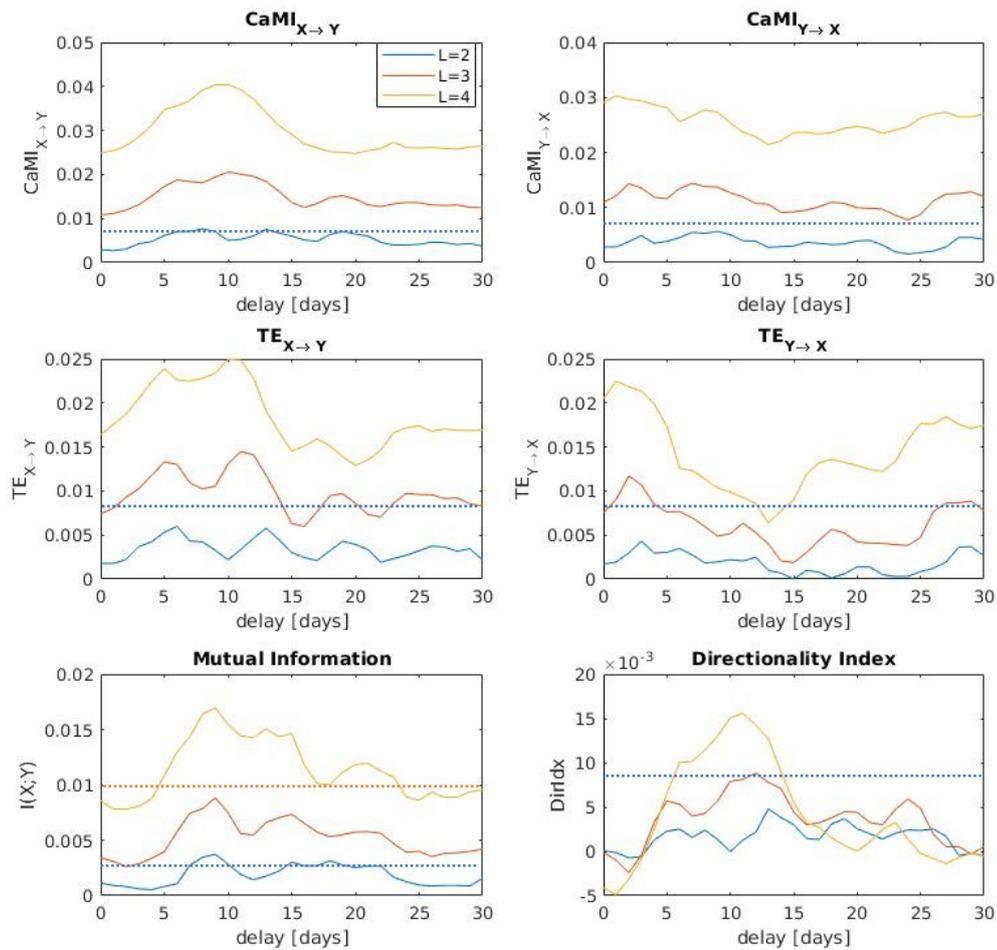


Figure 6.4: Information-theoretical values for the hypothesis that highest tidal amplitudes (X) induce high cumulative magnitudes (Y). The optimal partition line of gravity tidal amplitude is $2.37 \mu\text{m}/\text{s}^2$ and maximum cumulative magnitude $\sum M_w = 102.2$. The dashed lines represent the confidence margins. If a confidence margin is not shown, its value is beyond the scale of the plot.

candidate. However, this behaviour does not appear in the pointwise transfer entropy, so this common information does not appear to have a causal flow origin.

The final possibility considered is whether high tidal amplitudes could trigger a large number of seismic events, as opposed to a specific relation with the magnitude. In this case, it could be that the almost periodic stress changes applied to the fault system do not necessarily lead to the rupture of the larger events, but to a great number of smaller earthquakes, or to a cascade of one large earthquake triggering several smaller events in a short timescale. Hence, we have applied the same procedure for the time-series of the gravity tidal amplitudes against the time-series of the daily seismicity rate. The daily seismicity rate is here defined as the number of earthquakes catalogued in the area around the Japan trench during a 24h interval, considering only events with magnitude $M_w \geq 3$. The optimal partition is found at tidal amplitude of $2.37\mu\text{m/s}^2$ (above the average plus two standard deviation) and seismicity rate of 27 events/day (equal to the average plus two standard deviation). The results are displayed on Fig. 6.5.

The delay leading to maximal causal mutual information is 9 days. The results are very similar to the case of cumulative amplitudes. Once again, the causal informational quantities are below the margins from uncorrelated uniformly distributed random systems, so consistent with zero. For $L=2$, CaMI, transfer entropy are below the 0.008 bit line and are in the second decimal place for $L \in [3, 4]$ (hence below 0.1 and 1 bit confidence intervals). At the peak the mutual information, observed at delay 8–9 days, it equals the error level of 0.002 bit for $L=2$ and almost reach 0.01 bit for $L=3$. It remains significantly below the confidence interval of 0.6 bit for $L=4$. With more data points the confidence interval which enables discerning a system from a uncorrelated random setup could be reduced and these quantities determined more precisely. The pointwise information measures reveal, just as in the case of cumulative magnitude, a localisation in the high pointwise mutual information points in the region of high tidal amplitudes and high seismicity rate. The case for pointwise transfer entropy (PTE) is not so clear, though, apparently forming a region of high PTE in the mean tidal amplitudes and high seismicity rate for $L=2$, but spreading for larger L . With more data points, these structures would be

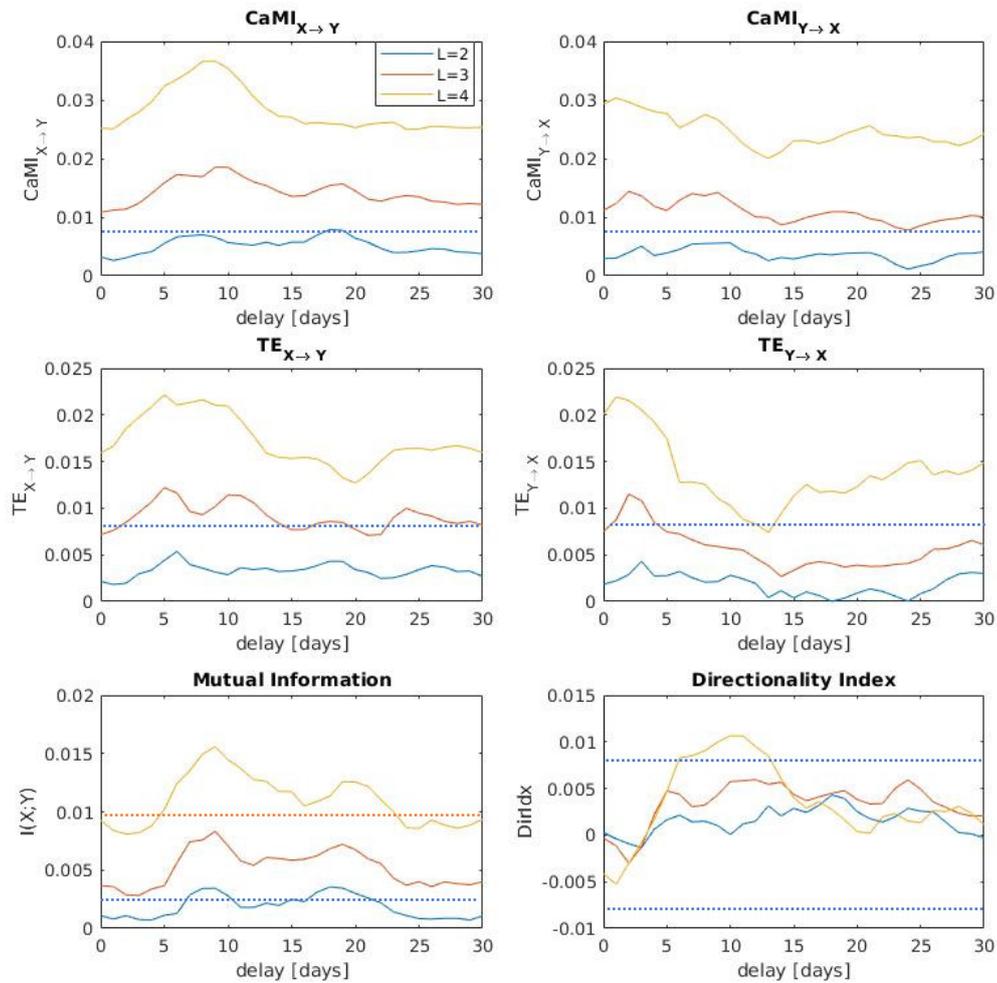


Figure 6.5: Information-theoretical values for the hypothesis that highest tidal amplitudes (X) induce high seismicity rate (Y). The optimal partition line of gravity tidal amplitude is $2.37\mu\text{m/s}^2$ and seismicity rate 27 events/day. The dashed lines represent the confidence margins. If a confidence margin is not shown, its value is beyond the scale of the plot.

more clear.

In summary, the hypothesis of tidal triggering of earthquakes is not verified above the confidence levels from uncorrelated uniformly distributed random systems using our method, for the Japanese trench area in the period from 01 July 1997 to 30 June 2008. However, there are suggestions of a peak of mutual information for the link of observed high gravity tidal amplitudes with high cumulative daily magnitude and high seismicity rate, particularly with a delay of 8–10 days between the signals. This deserves to be further explored once more data is available.

6.2 Gutenberg-Richter's b-value temporal variations

Laboratory experiments on rock mechanics suggest that the Gutenberg-Richter b-value fluctuates before a rupture: for an elastic rupture a sharp decrease occurs before, while for an inelastic rupture a brief peak occurs prior to the decrease [80]. Similar behaviour is expected to occur with real-world data preceding large earthquakes, as these rupture models were designed to approach seismic behaviour. A time-series of the b-value is computed from the IRIS seismic catalogue following procedure on Chapter 3 (Sec. 3.1.1). The selected time-window for the calculation of each data point of b-value is 2 weeks, which is a trade-off enabling the smallest possible window but still allowing a number of earthquake events above 50 at each time, so the error in the computed b-values is 0.1. As in the previous case, the sampling rate in all time-series is 1 day. The time-series of b-values is translated into absolute deviations from the average, so both the possibilities of sharp increase and dip can be considered.

We begin by investigating the causal link between b-value anomalies and maximum daily magnitude (single large events). The optimal threshold found is b-value of 0.32 and maximum daily magnitude of 5.6 (both of which equal to the mean plus two standard deviation). The results of the information measures in function of the time-delay between the time-series are given in Fig. 6.6.

The maximum CaMI is observed at a delay of 17 days. Reminding that for an uncorrelated uniformly distributed system the values of CaMI and transfer entropy are of

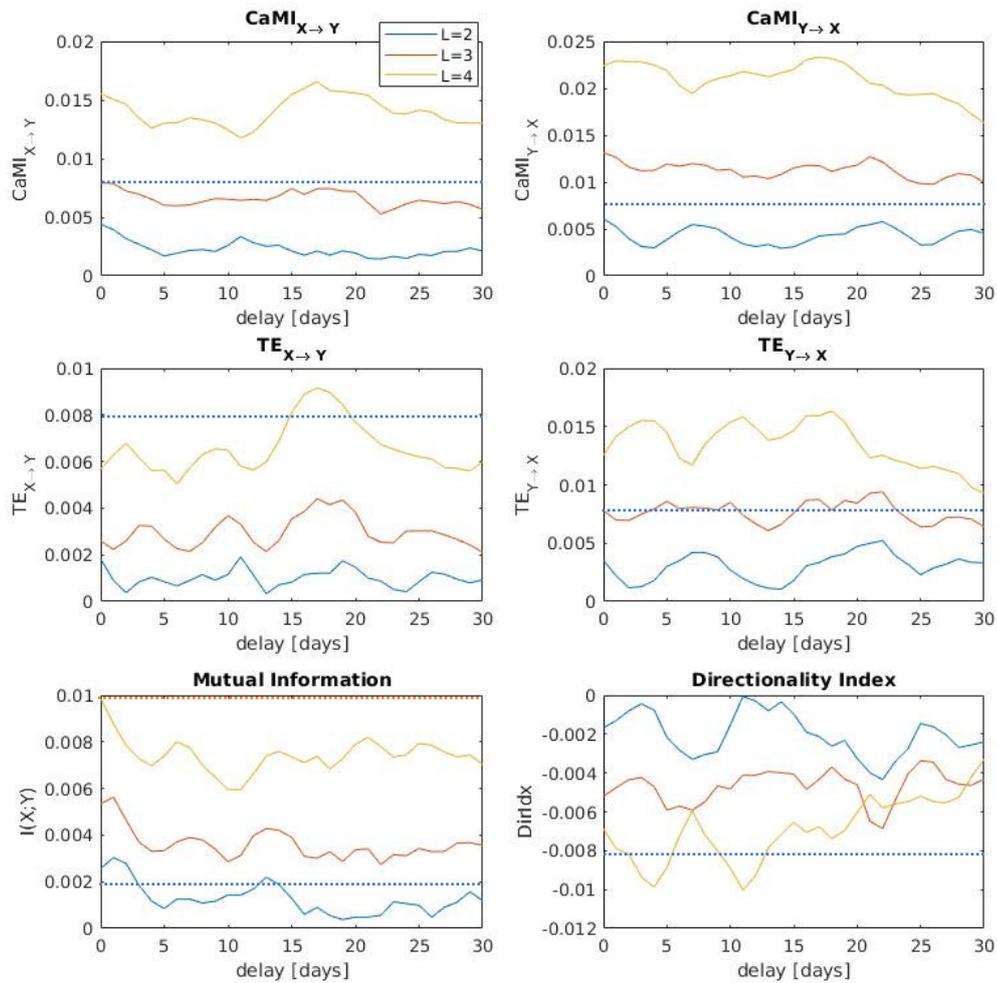


Figure 6.6: Information-theoretical values for the hypothesis that b-value anomalies (X) are precursors of earthquake events of high magnitude (Y). The optimal partition line of b-value is 0.32 and of maximum daily magnitude is $M_w = 5.6$. The dashed lines represent the confidence margins. If a confidence margin is not shown, its value is beyond the scale of the plot.

0.008 bits for $L=2$, 0.1 bit for $L=3$ and 1 bit for $L=4$, all the values shown of Fig. 6.6 for causal transfer of information are below this margin, hence effectively a null result as it is not possible to discern it from an uncorrelated random case. For mutual information, the confidence interval is 0.002 bits for $L=2$, 0.01 bit for $L=3$ and 0.6 bit for $L=4$. The mutual information for $L=2$ is above this confidence interval for delays of 0–3 days and 13 days. For longer symbolic sequences the values are always below these margins. The observed negative directionality index indicates that the flow of information is more significant in the direction of the maximum earthquake magnitude inducing b-value changes than the other way. This is not unexpected, as the b-value is derived from seismicity patterns. In Appendix C we have included the pointwise information measures for the optimal partition and delay, but the values are consistently zero for all points except few randomly scattered. We observe no indication of a possible precursory behaviour of b-value temporal anomalies with the maximal daily magnitude in this dataset.

The second hypothesis consists of assessing the b-value anomalies as a possible precursor of daily cumulative seismic magnitude. As previously mentioned, only magnitudes $M_w \geq 3$ are considered in the cumulative sum, once the catalogue might be incomplete below this level. The optimal initial partition leading to largest CaMI is found at b-value anomaly of 0.35 and cumulative magnitude of $\sum_i M(i) = 102.2$. Results of the information measures for varying time-delay are shown in Fig. 6.7.

The maximum of CaMI is observed for delay of 10 days. The values of CaMI and transfer entropy remain significantly lower than the error bar for all symbolic sequence lengths L (0.008 bits for $L=2$, 0.1 bit for $L=3$, 1 bit for $L=4$). The same occurs for mutual information (errors of 0.002 bits for $L=2$, 0.01 bit for $L=3$, 0.6 bit for $L=4$). The directionality index this time is mostly positive (indicating information flow from the b-value to the cumulative daily magnitude), however its value is negligible when accounting for error. In Appendix C the pointwise information measures are presented for the optimal partition, showing that the peak of mutual information is given by a group of only 5 data points and of transfer entropy for another 3 points. In the case of mutual information there is a suggestion of a higher pointwise mutual information region in the b-value high

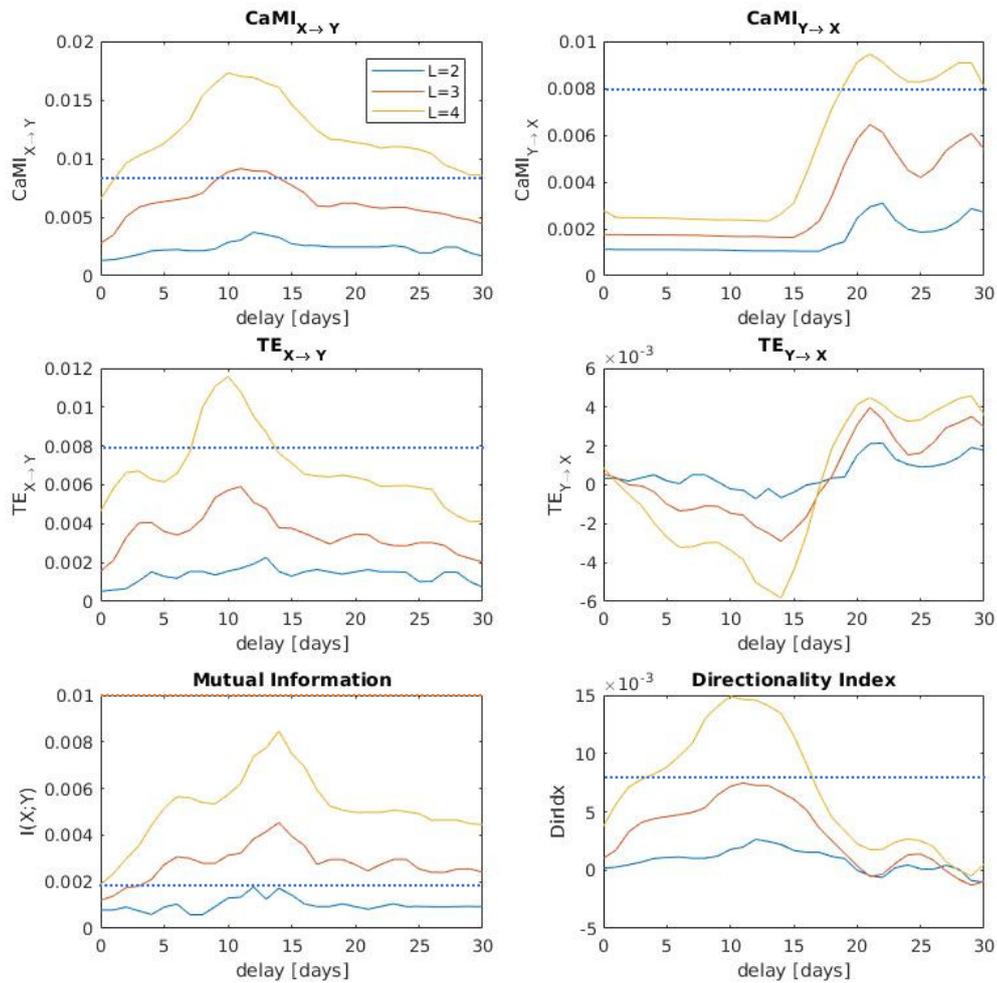


Figure 6.7: Information-theoretical values for the hypothesis that b-value anomalies (X) are precursors of high cumulative magnitudes (Y). The optimal partition line of b-value anomalies is 0.35 and maximum cumulative magnitude $\sum M_w = 102.2$. The dashed lines represent the confidence margins. If a confidence margin is not shown, its value is beyond the scale of the plot.

anomaly and high cumulative daily magnitude quadrant, but 5 points are insufficient to reach a conclusion. For pointwise transfer entropy (PTE), though, the location of the high PTE points in the state-space appear to be arbitrary. There is still insufficient evidence to support or discard this precursory behaviour, mainly due to the lack of data points (higher points imply smaller errors).

The last hypothesis consists of the b-values anomalies being related to occurrence of higher daily seismicity rate, *i.e.* an elevated number of earthquake events as opposed to a specific relation with their magnitude. The optimal partition is b-value anomaly of 0.32 and seismicity rate of 27 events/day. Results of the information measures for varying time-delay between the time-series are shown in Fig. 6.8.

The peak of causal mutual information is observed on 23 days of delay. The causal mutual information and transfer entropy are below error levels, as they are below 0.008 bits for $L=2$ and 0.1 for $L=3-4$. The mutual information almost reach the error level of 0.002 bits for $L=2$ in the delay interval of 23–27 days. However, it is significantly lower than the error levels of 0.01 bit for larger L . Hence, it could not be observed an association between b-value anomalies and high daily seismicity rate above the error levels with this dataset. A pointwise analysis (with the pointwise mutual information and pointwise transfer entropy) also leads to null results. These can be seen in Appendix C.

In summary, the hypothesis of temporal b-value anomalies as precursor of earthquakes is not yet verified above the error margins (from the comparison with uncorrelated uniformly distributed random system) using our method on the Japanese trench area for the period 01 July 1997 to 30 June 2008. More data ² can support the future reanalysis, by reducing the error levels, particularly with respect to a possible relation of b-value anomalies with cumulative daily magnitudes.

²*E.g.* 30 years, enabling over 10000 time-series points

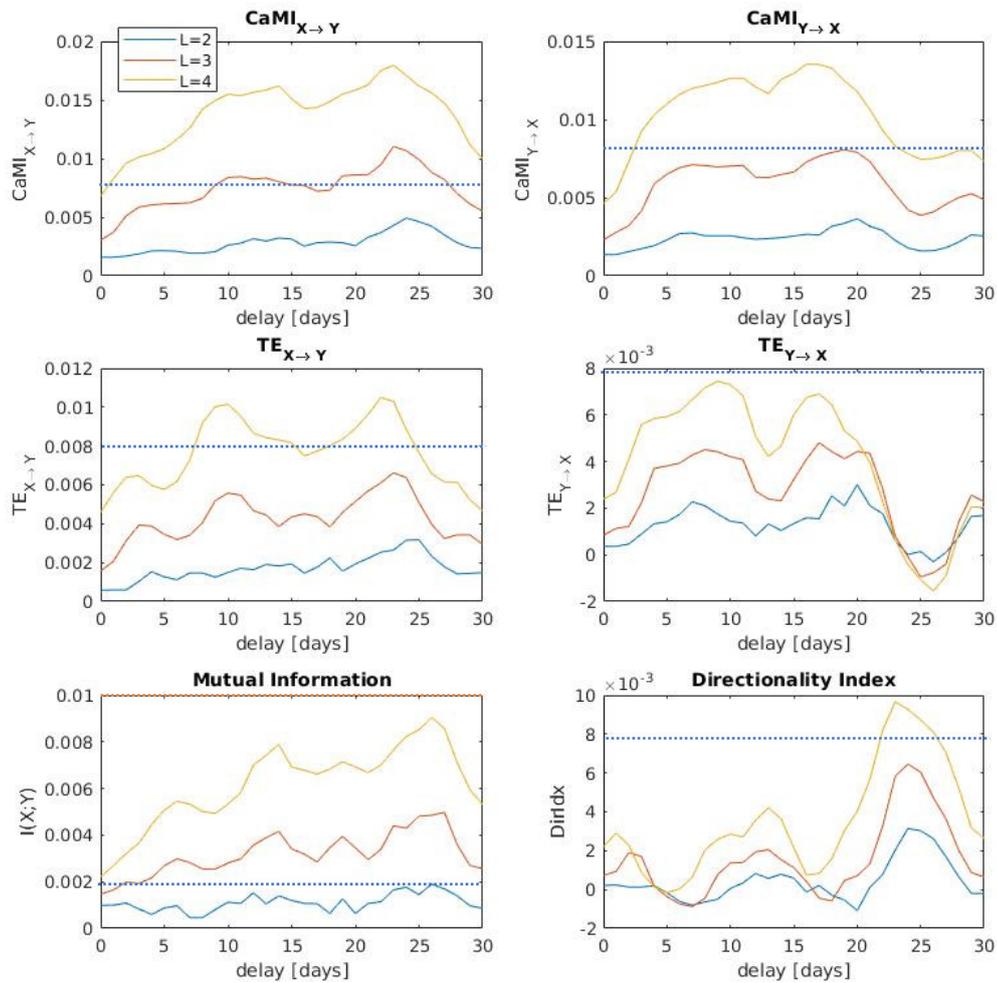


Figure 6.8: Information-theoretical values for the hypothesis that the b-value anomalies (X) are precursors of high seismicity rate (Y). The optimal partition line of b-value anomalies is 0.32 and seismicity rate 27 events/day. The dashed lines represent the confidence margins. If a confidence margin is not shown, its value is beyond the scale of the plot.

6.3 Pre-seismic gravity variations

As the stress accumulates in the fault area, the rock is deformed, increasing density. In principle this lead to small gravity variations detectable at the surface. According to the hypothesis, as the stress builds up, so should the gravity anomaly. However this anomaly is expected to be smaller than the tidal variations, hence this have to be removed from the raw gravity signal, producing a time-series of gravity residuals. Just like in the previous case, anomalies are considered as deviations from the mean with threshold above two standard deviation. The original time-series consisted from the 1 minute-sampled raw gravity data from Matsushiro station, which was treated to remove tides and atmospheric/ocean effects, and re-sampled to 1-day sampling time, where each data point consists of the largest anomaly of the day.

The first hypothesis considered is whether the gravity residuals are associated to the occurrence of future large earthquakes, and, if so, what is the expected time delay. The optimal threshold found for defining a gravity residual anomaly is 23.84nm/s^2 , and for the maximum daily magnitude is $M_w 5.6$ (both of which equal to the average plus two standard deviation). Fig. 6.9 shows the result of the information measures for this combination and varying delay between the time-series.

The maximum of CaMI is found at the delay of 8 days. The CaMI from the gravity residuals to the earthquake occurrence is slightly higher than the error level of 0.008 bits for $L=2$ in the delays 8–11 days. The mutual information is higher than the error level of 0.002 bit for $L=2$ for delay 9–12 and reaches the error level of 0.1 bit for $L=3$. For higher L the values are below the error levels. The values of transfer entropy are below error levels. Increasing the amount of data available would reduce the error levels, which could strongly support this first indications of a possible association with delay of around 10 days. Unfortunately, this result is not further assisted by the pointwise information measures, which exhibit no clear pattern (details on Appendix C)

The second possibility considered is if there are indications of the gravity residuals as precursors of a cumulative daily magnitude, and what is the delay observed between the anomaly in the gravity and occurrence of the peak in seismic energy release. The

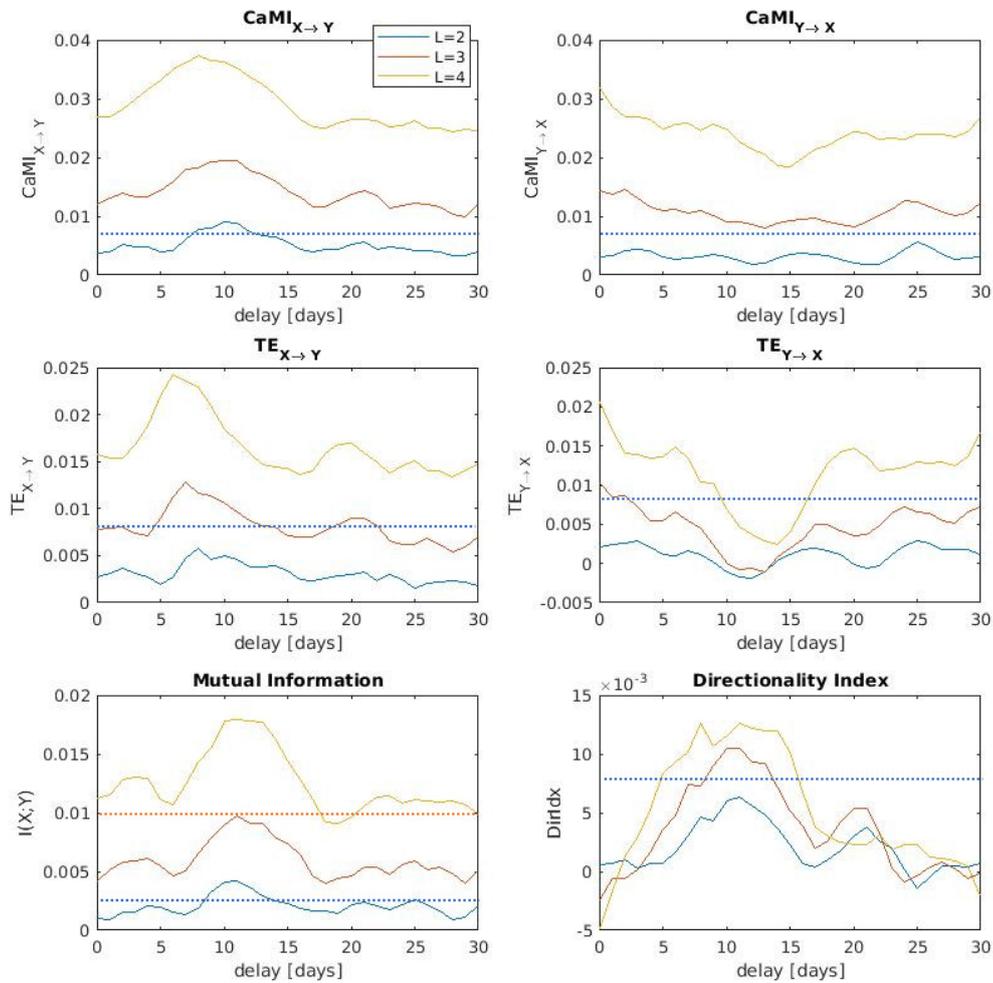


Figure 6.9: Information-theoretical values for the hypothesis that gravity residual anomalies (X) are precursors of earthquake events of high magnitude (Y). The optimal partition line of gravity residuals is 0.32nm/s^2 and of maximum daily magnitude is $M_w = 5.6$. The dashed lines represent the confidence margins. If a confidence margin is not shown, its value is beyond the scale of the plot.

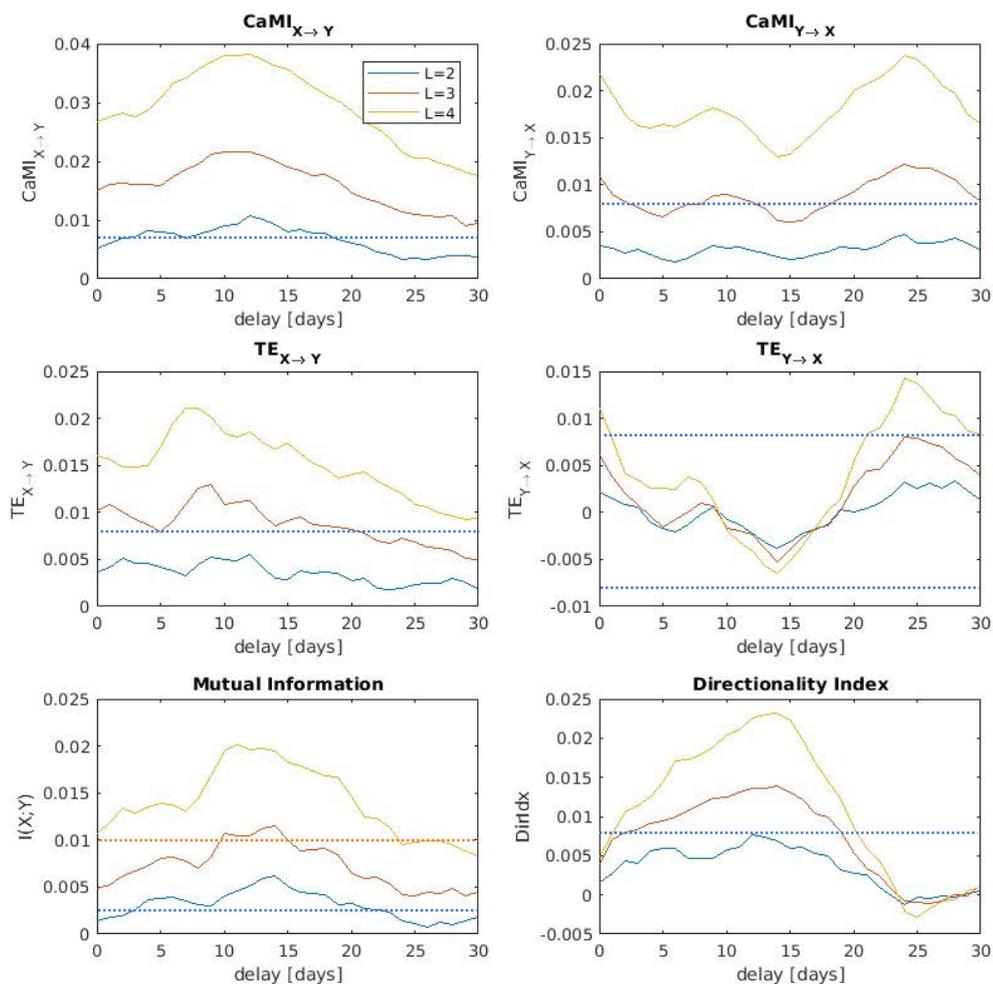


Figure 6.10: Information-theoretical values for the hypothesis that gravity residual anomalies (X) are precursors of high cumulative daily magnitudes (Y). The optimal partition line of gravity residuals is 23.84 nm/s^2 and of cumulative daily magnitude is $\sum M_w = 102.2$. The dashed lines represent the confidence margins. If a confidence margin is not shown, its value is beyond the scale of the plot.

optimal threshold found is gravity residuals of 23.84 nm/s^2 and cumulative magnitude of $\sum M_w = 102.2$. The results of the information measures, in function of the delay between the time-series, are given in Fig. 6.10.

The delay with highest CaMI is 12 days. The value of CaMI in the direction of the precursor candidate to the cumulative magnitude is larger than the error level of 0.008 bits for $L = 2$ and delay of 11–13 days. The mutual information is larger than the error level of 0.002 bit at $L = 2$ and larger than the error level of 0.01 bit for $L = 3$ for delays of 10–15 days. For higher L the error levels are greater than the value obtained. All values of

transfer entropy are below error levels. This result is also a promising first indicator of a possible flow of information from the gravity residuals to the cumulative daily magnitude in the system studied. Furthermore, it is supported by the pointwise information quantities, which apparently exhibit localisation of both high pointwise mutual information and pointwise transfer entropy in the high gravity residual anomaly - high cumulative daily magnitude combination. However, these can only be made clear with further data points, so the preliminary pointwise results are displayed on Appendix C.

The last possibility considered is a link between the gravity residual anomalies and high seismicity rate. The optimal threshold found is gravity residual of 23.84nm/s^2 (which is equal to the average plus two standard deviation) and seismicity rate of 31 events/day (which is above the average plus two standard deviation). Fig. 6.11 shows the results of the information measures for varying delay applied between the time-series.

The delay reaching the maximum CaMI is 11 days. Once more, CaMI is above the error level (from comparison with an uncorrelated uniformly distributed system) for $L=2$ and mutual information is above the error level for $L=2$ and $L=3$ in delay ranges within the interval of 10–15 days. Larger values of L lead to error bars superior to the information values obtained and all the values of transfer entropy are below the current error levels. As in the previous case, these results are promising first indicators that the gravity residual anomalies might be associated to high seismicity rate, with a delay of about 11 days between the gravity anomaly and the occurrence of high seismicity. As in the previous case, this is further supported by the concentration of the points of high pointwise mutual information and high pointwise transfer entropy in the high gravity anomaly - high seismicity area, which can be seen in Appendix C. However, in this the definition of the region is less evident than in the case of cumulative daily magnitude. More data points, when available, would support this analysis and also reduce the error levels.

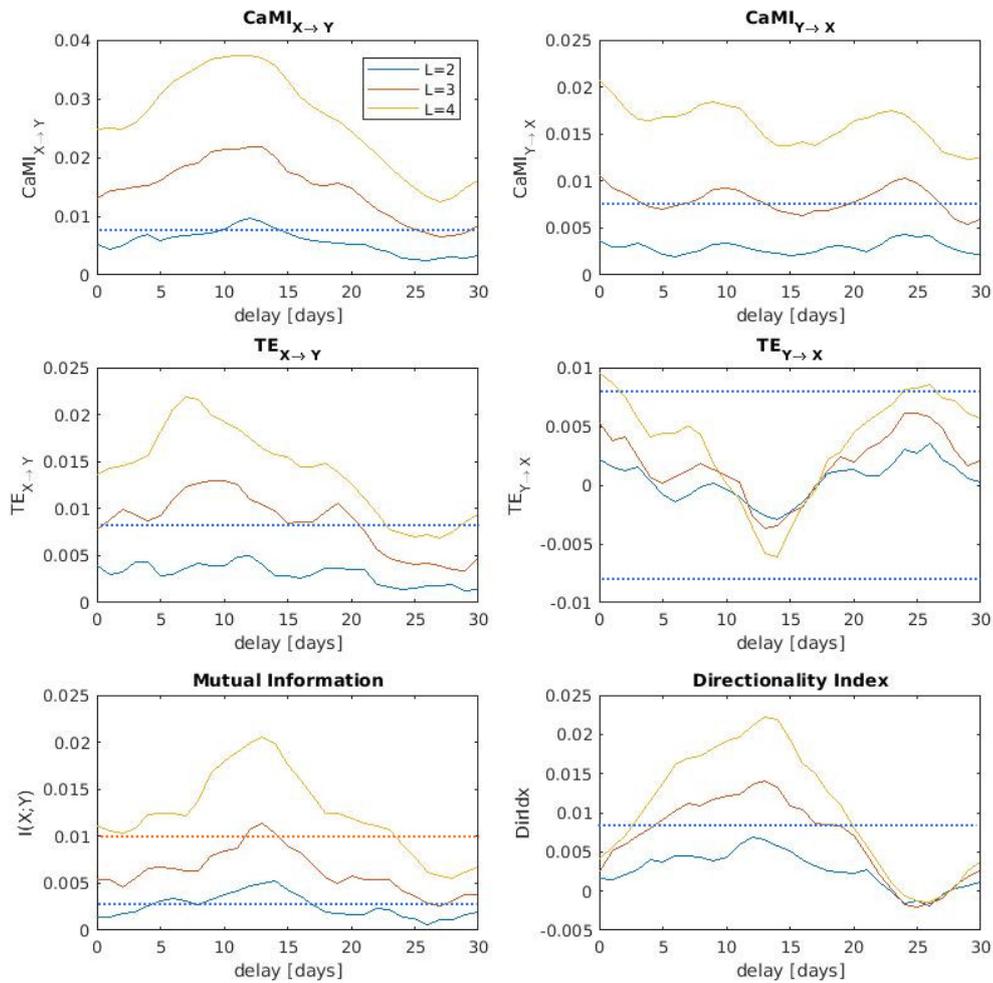


Figure 6.11: Information-theoretical values for the hypothesis that gravity residual anomalies (X) are precursors of high seismicity rate (Y). The optimal partition line of gravity residuals is 23.84nm/s^2 and of seismicity rate is 31 events/day. The dashed lines represent the confidence margins. If a confidence margin is not shown, its value is beyond the scale of the plot.

6.4 Preceding seismicity leading to new events

Finally, we consider other two possible scenarios for preceding seismicity being associated to new events. The first scenario is for a high cumulative magnitude before a large event. This could be the case of a foreshock behaviour – either by a single or a composition of events. The second scenario is for a low seismic rate before a large event. This attempts to model a type of seismic quiescence.

For the first case, we found the optimal thresholds to be $\sum M_w = 140$ for the cumulative daily magnitude (higher than the average plus two standard deviation) and $M_w 5.6$ for the maximal daily magnitude. The values of the information quantities in function of the time-delay applied between the time-series are shown in Fig. 6.12.

In the first 5 days the relation between cumulative and maximal magnitudes is a direct consequence of the definition of these quantities and the aftershock models (*e.g.* ETAS). Our primary interest is for larger delays. The maximum CaMI is then found at the delay of 11 days. For an interval of 11–15 days, CaMI is higher than the error level (from the comparison with uncorrelated uniformly distributed system) of 0.008 bits for $L=2$. Also, for delay of 14 days, the mutual information is above the error levels of 0.002 bits for $L=2$ and 0.01 bits for $L=3$. The transfer entropy is below the error levels in all cases. Pointwise information measures (see Appendix C) reveal that the main contribution to the causal mutual information and the transfer entropy come from the high values of cumulative magnitude. Hence, this is an association worth to be further analysed with more data in future.

The second case, of an unusually low seismicity rate linking to a future occurrence of a large earthquake event, has optimal partition of 2 events/day for the seismicity rate and $M_w 5.6$ for the maximum daily magnitude. The results of the information measures for varying delay between the time-series is shown in Fig. 6.13.

The most relevant maxima of CaMI (shown for all L) is on delay of 23 days. However, all the information quantities are smaller than the error levels, at all delay times. The directionality index, slightly negative, with value very close to zero, is suggestive that this

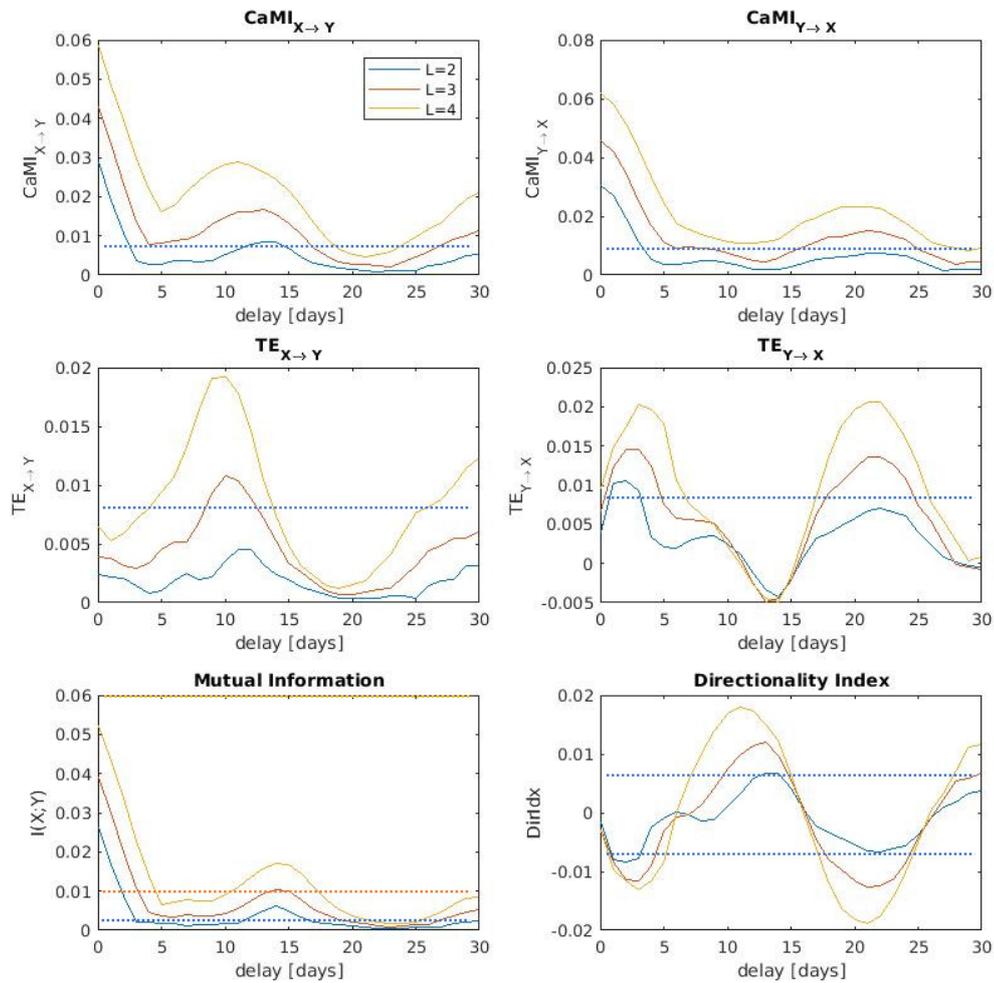


Figure 6.12: Information-theoretical values for the hypothesis that high cumulative daily magnitude (X) are precursors of earthquake events of high magnitude (Y). The optimal partition line of cumulative daily magnitude is $\sum M_w = 140$ and of maximum daily magnitude is $M_w = 5.6$. The dashed lines represent the confidence margins. If a confidence margin is not shown, its value is beyond the scale of the plot.

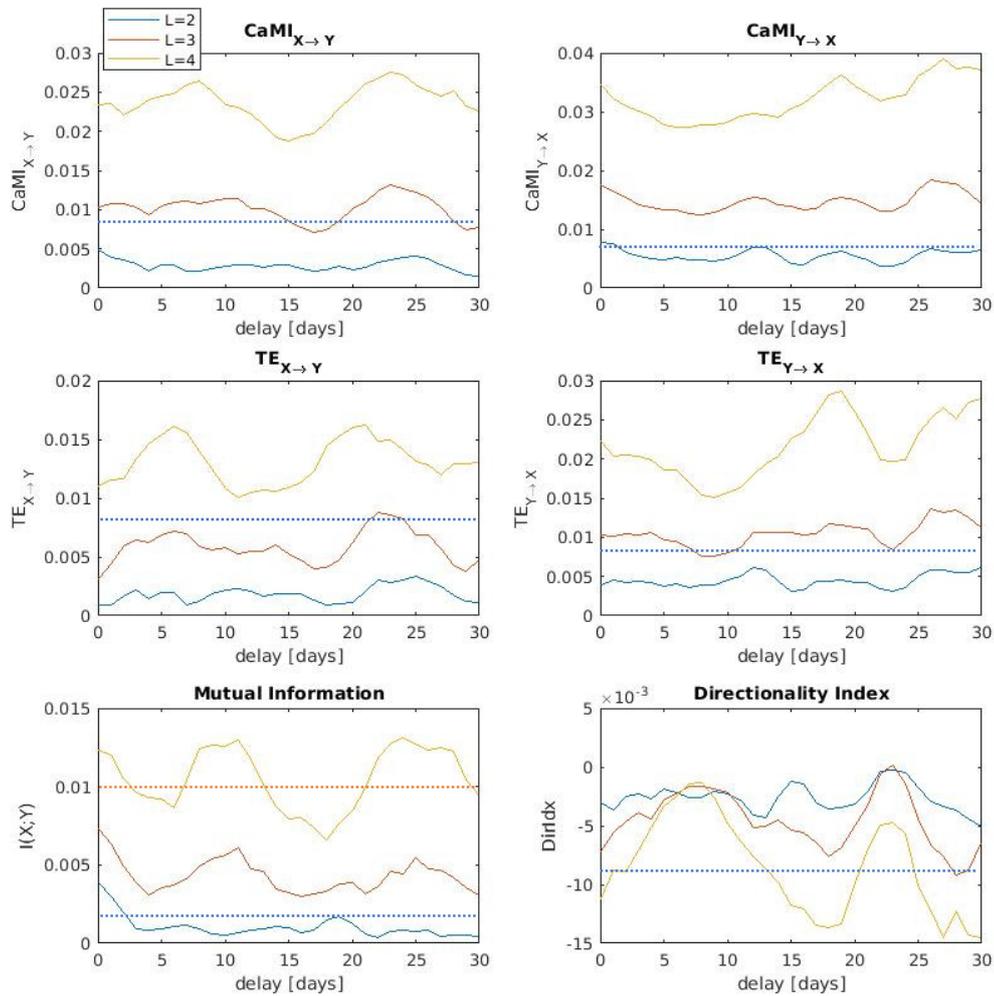


Figure 6.13: Information-theoretical values for the hypothesis that low seismicity rate (X) are precursors of earthquake events of high magnitude (Y). The optimal partition line of the seismicity rate is 2 events/day and of maximum daily magnitude is $M_w = 5.6$. The dashed lines represent the confidence margins. If a confidence margin is not shown, its value is beyond the scale of the plot.

precursor candidate is not verified in the dataset, as there is no significant flow of information from the seismicity rate to the maximum magnitude. This is further supported by the pointwise information having no visually distinguishable structures when plotted in the phase-space (compare Appendix C with the causal bubble features of Figs. 5.9, 5.10, 5.14 and 5.15).

6.5 Summary

Of the possibilities tested, the most promising precursor candidates are summarised in Fig. 6.14. A future analysis with more data points will provide more supporting evidence or discard some of these links.

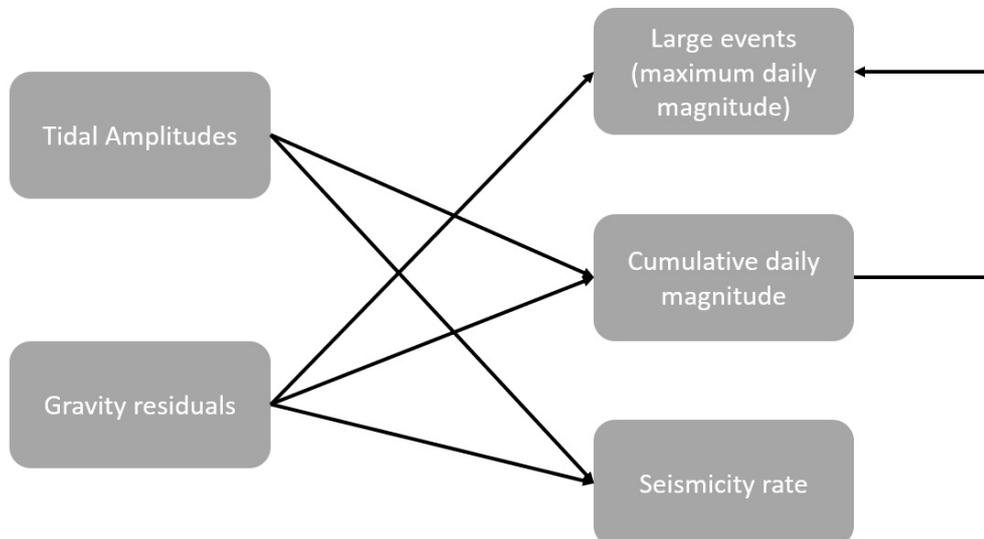


Figure 6.14: Diagram summarising the most promising precursor candidates after information analysis

Chapter 7

Conclusions

7.1 Review of context, objective and findings

The most fundamental question in Seismology is: “Can earthquakes be predicted? If so, how?”. On the other hand, perhaps the most essential epistemic question to Physics is the definition of causation: “What is a cause? What is an effect? How to distinguish them from the observation of a physical system?”. In this work, we have combined the two areas, by developing a method for calculation of the causal flow of information between two systems. This method uses the observed time-series of the physical variables to them distinguish the variable sending information from the variable receiving information.

There are several geophysical variables regarded as candidates of earthquake precursors. This method has the potential of supporting their analysis by indicating the presence of a flow of information from the precursor candidate to the occurrence of seismicity. This is further improved by considering delayed effects, *i.e.* also enabling to determine the amount of time in advance the precursor expressed information about the seismicity occurrence.

We have applied the method first for coupled logistic systems, as a test-bench, retrieving expected results and discovering new ones, such as the effects on logistic networks. We then proceed to apply to 11 years of data of seismicity occurrence and precursor candidates on the Japan trench area. This was not intended to confirm or discard any precursor candidate, but only to present a preliminary view of how the method can be used to earthquake prediction problems.

The method revealed the correct direction of causality for both coupled systems and for networks. In general, it is observed a proportionality of the mutual information or mutual information rate value with the coupling strength, and a decay of the information measures with the amplitude of an applied dynamic noise. The transfer entropy rate and the directionality index might correctly indicate the direction of causation, but, on special circumstances (particularly high coupling), they might not be strongest to the neighbour node, but to a node on distance 2 or distance 3. The mutual information rate, however, is strongest to the neighbour node, usually. The exception is when there are other nodes in parallel, sharing significant mutual information with the analysed node, but not directly connected (both are effects from the same cause). In summary, effective inference of the network structure requires simultaneous analysis of the mutual information or mutual information rate and of the transfer entropy or transfer entropy rate.

At the present moment the information quantities obtained for the geophysical systems considered are of the same levels of our conservative margins of confidence, from the application to random time-series with the same number of points. Hence, more data is still required before reaching a definite conclusion. We roughly estimate that 30 years of continuous data should suffice for clarifying the contributions of some precursor candidates that appear to share or transmit information.

7.2 Open questions

The analysis of the method applied to logistic networks opened up the question of why, under special conditions, the transfer entropy is higher to a node more distant than the immediate neighbour. This has practical consequences, as the transfer entropy is applied for inference of network structures from time-series observations. If the system happens to be in these cases, typically associated to high coupling, then the reconstruction might not reflect the real network topology. For systems like the brain, though, this is a property expected to be found, as regions not directly connected can become strongly functionally connected, forming a channel of communication.

An additional question for future exploration is how the structures of the pointwise

information measures change as the coupling strength is varied. In particular, this could clarify the peak of transfer entropy at coupling strength $\sigma = 0.2$ with the CML coupling.

Regarding the application to seismic data, there are initial indications of earthquakes being causally connected to several precursors. However, an error analysis approach based on uniform random time-series prevents us from rigorously stating that indeed we have found the main precursors. This is not an issue, as the primary goal is not to confirm any hypothesis but to test how the method could be applied to earthquake prediction. The cases where an initial indicator of information sharing and possible causal information transmission between precursor candidates and seismicity occurrence are:

- Tidal amplitudes and cumulative daily magnitude, on a delay of 5–13 days
- Tidal amplitudes and seismicity rate, on a delay of 5–10 days
- Gravity residuals and high magnitude events, on a delay of 8–11 days
- Gravity residuals and cumulative daily magnitude, on a delay of 11–13 days
- Gravity residuals and seismicity rate, on a delay of 10–15 days
- High cumulative daily magnitude and future high magnitude events, on a delay of 11–15 days

It is not yet clear why all the candidates are connected in a similar delay interval of 5–15 days. These candidates deserve to be further explored as soon as more data is available. With more data points, not only the error levels decrease, but also a clearer analysis of pointwise informational measures becomes possible. The number of points available (4018) does not allow for a definition of spatial patterns of these pointwise quantities in the phase space of the data points. Preliminary results were left for the Appendix C.

7.3 Applications to other fields of knowledge

The method presented in this work can be applied to any time-series (as long as the sampling rate is constant and gaps are filled) or sequence of ordered events. It means

that it can also support analysis of:

- Financial systems, by identifying the key variables influencing a market,
- Neuroscience, by identifying the transmission of information in a neuron network or, in the macroscale, the overall flow of information between different areas of the brain,
- Genetics, by identifying how different regions of the DNA/RNA sequencing might be causally related,
- Social sciences, by analysing the flow of information between quantifiable social, demographic and environmental factors, supporting hypothesis of links between different social variables,
- Computational linguistics, by enabling the analysis of how words are causally associated in human language, so to improve natural language computer models,
- Atmospheric sciences and oceanography, by identifying the flow of information between different variables, particularly in the occurrence of extreme events,
- Ecology, by identifying, from collected data such as time-series of fish population, the likely trophic chain and relation with physical environment factors in a ecosystem.

Bibliography

- [1] V.I. Ferronsky and S.V. Ferronsky. *Dynamics of the Earth: Theory of the Planets' Motion Based on Dynamic Equilibrium*. Springer, London, 2010. doi: 10.1007/978-90-481-8723-2.
- [2] F.D. Stacey and P.M. Davis. *Physics of the Earth*. Cambridge University Press, Cambridge, 2008.
- [3] S. Stein and M. Wysession. *An Introduction to seismology, earthquakes and Earth structure*. Blackwell, Oxford, 2003.
- [4] K. Aki and P.G. Richards. *Quantitative Seismology*. University Science Books, South Orange, N.J., 2 edition, 2009.
- [5] K.E. Bullen and B.A. Bolt. *An introduction to the theory of seismology*. Cambridge University Press, Cambridge, 1985.
- [6] P. Shearer. *Introduction to seismology*. Cambridge University Press, Cambridge, 2009.
- [7] F.A. Dahlem and J. Tromp. *Theoretical global seismology*. Princeton University Press, Princeton, N.J., 1998.
- [8] CRED and UNISDR. *Poverty & Death: Disaster mortality 1996-2015*. CRED – Université Catholique de Louvain and UNISDR, Brussels and Geneva, 2016.
- [9] R.J. Geller, D.D. Jackson, Y.Y. Kagan, and F. Mulargia. Earthquakes cannot be predicted. *Science*, 275(5306):1616, 1997. doi: 10.1126/science.275.5306.1616.
- [10] A. Valencio, N. Valencio, and F.Y. Hanai. Keeping the feet on the ground: reflecting, monitoring and taking action in front of seismic events (*Mantendo os pés no chão: refletir, monitorar e agir diante eventos sísmicos*). *Revista Eletrônica de Tecnologia e Cultura*, 20:21–40, 2017. Original document in Portuguese.
- [11] C.W.J. Granger. Investigating causal relations by econometric models and cross-spectral methods. *Econometrica*, 37(3):424–438, 1969. doi: 10.2307/1912791.
- [12] C.W.J. Granger. Testing for causality: a personal viewpoint. *Journal of Economic Dynamics and Control*, 2:329–352, 1980. doi: 10.1016/0165-1889(80)90069-X.
- [13] A. Valencio, C. Grebogi, and M.S. Baptista. Methods for removal of unwanted signals from gravity time-series: comparison using linear techniques complemented with analysis of system dynamics. *Chaos*, 27(10):103126, 2017. doi: 10.1063/1.4996452.

- [14] N. Valencio and A. Valencio. News coverage of disasters in Brazil: sociopolitical dimensions put at the margins of public debate (*Cobertura jornalística sobre desastres no Brasil: dimensões sociopolíticas marginalizadas no debate público*). *Disertaciones*, 12(2):165–185, 2017. doi: 110.12804/revistas.urosario.edu.co/disertaciones/a.4791. Original document in Portuguese.
- [15] N. Valencio and A. Valencio. Vulnerability as social oppression: the traps of risk-prevention actions. In V. Marchezini, B. Wisner, S. Saito, and L. Londe, editors, *Reduction of Vulnerability to Disasters: from Knowledge to Action*, pages 111–137. RiMa, São Carlos, 2017.
- [16] N. Valencio and A. Valencio. Media coverage of the UK flooding crisis: a social panorama. *Disasters*, (early view), in press. doi: 10.1111/disa.12255.
- [17] N. Valencio and A. Valencio. Cities plagued by dangerous waters (*Cidades atormentadas pelas águas perigosas*), In preparation. Original document in Portuguese.
- [18] S. Missiakoulis. Aristotle and earthquake data: A historical note. *International Statistical Review*, 76:130–133, 2008. doi: 10.1111/j.1751-5823.2007.00040.x.
- [19] R. Feng and Y. Wu. Research on history of chinese seismology. *Earthquake Science*, 23(3):243–257, 2010. doi: 10.1007/s11589-010-0720-z.
- [20] J. Dewey and P. Byerly. The early history of seismometry (to 1900). *Bulletin of the Seismological Society of America*, 59(1):183–227, 1969.
- [21] T. Rikitake. *Earthquake Prediction*, volume 9 of *Developments in Solid Earth Geophysics*. Elsevier, Amsterdam, 1976.
- [22] B.L.N. Kennett. *The seismic wavefield*, volume 1: introduction and theoretical development. Cambridge University Press, Cambridge, 2001.
- [23] B.L.N. Kennett. *The seismic wavefield*, volume 2: interpretation of seismograms on regional and global scales. Cambridge University Press, Cambridge, 2002.
- [24] M. Wyss. *Evaluation of Proposed Earthquake Precursors*. American Geophysical Union, Washington, D.C., 2013.
- [25] B.P. Allmann and P. Shearer. Global variations of stress drop for moderate to large earthquakes. *Journal of Geophysical Research*, 114:B01310, 2009. doi: 10.1029/2008JB005821.
- [26] D. Schorlemmer, S. Wiemer, and M. Wyss. Variations in earthquake-size distribution across different stress regimes. *Nature*, 437:539–542, 2005. doi: 10.1038/nature04094.
- [27] S. Ji, S. Sun, Q. Wang, and D. Marcotte. Lamé parameters of common rocks in the earth’s crust and upper mantle. *Journal of Geophysical Research*, 115:B06314, 2010. doi: 10.1029/2009JB007134.

- [28] B.L.N. Kennett and E.R. Engdahl. Travel times for global earthquake location and phase association. *Geophysical Journal International*, 105:429–465, 1991. doi: 10.17611/DP/9991809.
- [29] A.M. Dziewonski and D.L. Anderson. Preliminary reference Earth model. *Physics of the Earth and Planetary Interiors*, 25:297–356, 1981. doi: 10.1016/0031-9201(81)90046-7.
- [30] B.L.N. Kennett, E.R. Engdahl, and R. Buland. Constraints on seismic velocities in the earth from travel times. *Geophysical Journal International*, 122:108–124, 1995. doi: 10.1111/j.1365-246X.1995.tb03540.x.
- [31] C.F. Richter. An instrumental magnitude scale. *Bulletin of the Seismological Society of America*, 25:1–32, 1935.
- [32] H.B. Seed, M.P. Romo, J. Sun, A. Jaime, and J. Lysmer. Relationships between soil conditions and earthquake ground motions in Mexico City in the earthquake of Sept. 19, 1985. Technical report, Earthquake Engineering Research Center, University of California, Berkeley, 1985. Available at <https://nehrpsearch.nist.gov/static/files/NSF/PB88178991.pdf> [Accessed 16 Feb 2018].
- [33] B. Gutenberg and C.F. Richter. Earthquake magnitude, intensity, energy, and acceleration. *Bulletin of the Seismological Society of America*, 32(3):163–191, 1942.
- [34] B. Gutenberg and C.F. Richter. Magnitude and energy of earthquakes. *Nature*, 176:795, 1955. doi: 10.1038/176795a0.
- [35] B. Gutenberg and C.F. Richter. Magnitude and energy of earthquakes. *Annali di Geofisica*, 9:1–15, 1956. doi: 10.4401/ag-5590.
- [36] H. Kanamori. The energy release in great earthquakes. *Journal of Geophysical Research*, 82(20):2981–2987, 1987. doi: 10.1029/JB082i020p02981.
- [37] G. Thompson and C. Reyes. GISMO - a seismic data analysis toolbox for MATLAB [open source software package], 2017. Available at <http://geoscience-community-codes.github.io/GISMO/> [Accessed 6 Feb 2018].
- [38] F. Omori. On the aftershocks of earthquakes. *Journal of the College of Science, Imperial University of Tokyo*, 7:111–200, 1894.
- [39] T. Utsu, Y. Ogata, and R.S. Matsu'ura. The centenary of the Omori formula for a decay law of aftershock activity. *Journal of the Physics of the Earth*, 43:1–33, 1995. doi: 10.4294/jpe1952.43.1.
- [40] C.F. Richter. *Elementary Seismology*. Freeman, San Francisco, CA, 1958.
- [41] M. Bath. Lateral inhomogeneities of the upper mantle. *Tectonophysics*, 2(6):483–514, 1965. doi: 10.1016/0040-1951(65)90003-X.
- [42] Y. Ogata. Detection of precursory relative quiescence before great earthquakes through a statistical model. *Journal of Geophysical Research*, 97(B13):19845–19871, 1992. doi: 10.1029/92JB00708.

- [43] Y. Okada. Surface deformation due to shear and tensile faults in a half-space. *Bulletin of the Seismological Society of America*, 75(4):1135–1154, 1985.
- [44] S. Okubo. Gravity and potential changes due to shear and tensile faults in a half-space. *Journal of Geophysical Research*, 97(B5):7137–7144, 1992. doi: 10.1029/92JB00178.
- [45] E. Serpelloni, L. Anderlini, A. Avallone, V. Cannelli, A. Cavaliere, D. Cheloni, C. D’Ambrosio, E. D’Anastasio, A. Esposito, G. Pietrantonio, A.R. Pisani, M. Anzidei, G. Cecere, N. D’Agostino, S. Del Mese, R. Devoti, A. Galvani, A. Massucci, D. Melini, F. Riguzzi, G. Selvaggi, and V. Sepe. GPS observations of coseismic deformation following the May 20 and 29, 2012, Emilia seismic events (Northern Italy): data, analysis and preliminary models. *Annals of Geophysics*, 55(4):759–766, 2012. doi: 10.4401/ag-6168.
- [46] Y. Imanishi. A network of superconducting gravimeters detects submicrogal coseismic gravity changes. *Science*, 306(5695):476–478, 2004. doi: 10.1126/science.1101875.
- [47] T. Ito, E. Gunawan, F. Kimata, T. Tabei, I. Meilano, Y. Ohta, N. Ismail, I. Nurdin, and D. Sugiyanto. Co-seismic offsets due to two earthquakes (m_w 6.1) along the Sumatran fault system derived from GNSS measurements. *Earth, Planets and Space*, 68(1):57, 2016. doi: 10.1186/s40623-016-0427-z.
- [48] K. Chen, M. Ge, A. Babeyko, X. Li, F. Diao, and R. Tu. Retrieving real-time co-seismic displacements using GPS/GLONASS: a preliminary report from the September 2015 M_w 8.3 Illapel earthquake in Chile. *Geophysical Journal International*, 206(2):941–953, 2016. doi: 10.1093/gji/ggw190.
- [49] F. Beauducel. Okubo92 Gravity change due to shear and tensile faults - Matlab/Octave scripts for geophysical studies and others [open source software package], 2012. Available at Matlab Exchange [Accessed 20 Feb 2018].
- [50] F. Beauducel. Okada85 Surface deformation due to a finite rectangular source - Matlab/Octave scripts for geophysical studies and others [open source software package], 2014. Available at Matlab Exchange [Accessed 20 Feb 2018].
- [51] C. Tolomei, G. Pezzo, C. Bignami, S. Atzori, E. Trasatti, A. Antonioli, S. Stramondo, and S. Salvi. Coseismic deformation field and source modelling for the Amatrice event of August 24, 2016 by means of InSAR and CGPS data. In *Proceedings of the 35th GNGTS, Lecce, 2016*, pages 84–87. OGS – Istituto Nazionale di Oceanografia e di Geofisica Sperimentale, Lecce, Italy, 2016.
- [52] J. Harms, J.-P. Ampuero, M. Barsuglia, E. Chassande-Mottin, J.-P. Montagner, S. N. Somala, and B. F. Whiting. Transient gravity perturbations induced by earthquake rupture. *Geophysical Journal International*, 201(3):1416–1425, 2015. doi: 10.1093/gji/ggv090.
- [53] J. Harms. Terrestrial gravity fluctuations. *Living Reviews in Relativity*, 18(3):1–150, 2015. doi: 10.1007/lrr-2015-3.

- [54] J.P. Montagner, K. Juhel, M. Barsuglia, J.P. Ampuero, E. Chassande-Mottin, J. Harms, B. Whiting, P. Bernard, E. Clévéde, and P. Lognonné. Prompt gravity signal induced by the 2011 Tohoku-Oki earthquake. *Nature Communications*, 7:13349, 2016. doi: 10.1038/ncomms13349.
- [55] M. Vallée, J.P. Ampuero, K. Juhel, P. Bernard, J.P. Montagner, and M. Barsuglia. Observations and modeling of the elastogravity signals preceding direct seismic waves. *Science*, 358(6367):1164–1168, 2017. doi: 10.1126/science.aao0746.
- [56] A.E.V. Lopes, M. Assumpção, A.F. Nascimento, J.M. Ferreira, E.A.S. Menezes, and J.R. Barbosa. Intraplate earthquake swarm in Belo Jardim, NE Brazil: reactivation of a major Neoproterozoic shear zone (Pernambuco Lineament). *Geophysical Journal International*, 180(3):1303–1312, 2010. doi: 10.1111/j.1365-246X.2009.04485.x.
- [57] M. Assumpção, G. Suarez, and J.A. Veloso. Fault plane solutions of intraplate earthquakes in Brazil: some constraints on the regional stress field. *Tectonophysics*, 113(3-4):283–293, 1985. doi: 10.1016/0040-1951(85)90201-X.
- [58] J.M. Ferreira, M. Takeya, J.M. Costa, J.A. Moreira, M. Assumpção, J.A. Veloso, and R.G. Pearce. A continuing intraplate earthquake sequence near João Câmara, Northeastern Brazil: preliminary results. *Geophysical Research Letters*, 14(10):1042–1045, 1987. doi: 10.1029/GL014i010p01042.
- [59] N. Kato. Simulation of seismic cycles of buried intersecting reverse faults. *Journal of Geophysical Research*, 106(B3):4221–4232, 2001. doi: 10.1029/2000JB900417.
- [60] W. Levandowski, M. Zellman, and R. Briggs. Gravitational body forces focus North American intraplate earthquakes. *Nature Communications*, 8:14314, 2017. doi: 10.1038/ncomms14314.
- [61] W.A. Thomas and C.A. Powell. Necessary conditions for intraplate seismic zones in north america. *Tectonics*, 36(12):2903–2917, 2017. doi: 10.1002/2017TC004502.
- [62] G.C. Beroza and S. Ide. Necessary conditions for intraplate seismic zones in North America. *Annual Review of Earth and Planetary Sciences*, 39:271–296, 2011. doi: 10.1146/annurev-earth-040809-152531.
- [63] J.R. Leeman, D.M. Saffer, M.M. Scuderi, and C. Marone. Laboratory observations of slow earthquakes and the spectrum of tectonic fault slip modes. *Nature Communications*, 7:11104, 2016. doi: 10.1038/ncomms11104.
- [64] K. Obara. Nonvolcanic deep tremor associated with subduction in Southwest Japan. *Science*, 296(5573):1679–1681, 2002. doi: 10.1126/science.1070378.
- [65] G. Rogers and H. Dragert. Nonvolcanic deep tremor associated with subduction in Southwest Japan. *Science*, 300(5627):1942–1943, 2003. doi: 10.1126/science.1084783.

- [66] M. Wyss. Evaluation of proposed earthquake precursors. *EOS: Transactions of the American Geophysical Union*, 72(38):411, 1991. doi: 10.1029/90EO10300.
- [67] M. Wyss. Second round of evaluations of proposed earthquake precursors. *Pure and Applied Geophysics*, 149:3–16, 1997. doi: 10.1007/BF00945158.
- [68] K.T. Wu, M.S. Yue, H.Y. Wu, S.L. Chao, H.T. Chen, W.Q. Huang, K.Y. Tien, and S.D. Lu. Case 2: Foreshocks to the Haicheng earthquake of 1975, certain characteristics of the Haicheng earthquake (M=7.3) sequence. In M. Wyss, editor, *Evaluation of Proposed Earthquake Precursors*. American Geophysical Union, Washington, D.C., 2013.
- [69] J.R. Bowman and J.W. Dewey. Relocation of teleseismically recorded earthquakes near Tennant Creek, Australia: implications for midplate seismogenesis. *Journal of Geophysical Research*, 96(B7):11973–11979, 1991. doi: 10.1029/91JB00923.
- [70] R.S. Matsu'ura. Precursory quiescence and recovery of aftershock activity before some large aftershocks. In M. Wyss, editor, *Evaluation of Proposed Earthquake Precursors*. American Geophysical Union, Washington, D.C., 2013.
- [71] H. Wakita, Y. Nakamura, K. Notsu, M. Noguchi, and T. Asada. Radon anomaly: A possible precursor of the 1978 Izu-Oshima-kinkai Earthquake. *Science*, 207(4433):882–883, 1980. doi: 10.1126/science.207.4433.882.
- [72] E. Roeloffs, E. Quilty, and C.H. Scholtz. Case 21 water level and strain changes preceding and following the August 4, 1985 Kettleman Hills, California, earthquake. *Pure and Applied Geophysics*, 149:3–16, 1997. doi: 10.1007/BF00945158.
- [73] A.T. Linde, K. Suyehiro, S. Miura, I.S. Sacks, and A. Takagi. Episodic aseismic strain signal precursors to the Japan Sea earthquake of 1983. In M. Wyss, editor, *Evaluation of Proposed Earthquake Precursors*. American Geophysical Union, Washington, D.C., 2013.
- [74] H. Sato. Precursory land tilt prior to the Tonankai earthquake of 1944, in some precursors prior to recent great earthquakes along the Nankai Trough. In M. Wyss, editor, *Evaluation of Proposed Earthquake Precursors*. American Geophysical Union, Washington, D.C., 2013.
- [75] H. Ishii, S. Miura, and A. Takagi. Large-scale crustal movements before and after the 1983 Japan Sea earthquake. *Journal of the Physics of the Earth*, 34(Suppl.):159–174, 1986. doi: 10.1029/GL014i010p01042.
- [76] Council of European Union. Council Directive 2013/51/Euratom of 22 October 2013 laying down requirements for the protection of the health of the general public with regard to radioactive substances in water intended for human consumption, 2013. Available at <http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32013L0051> [Accessed 21 Feb 2018].
- [77] R.A. Baragiola, C.A. Dukes, and D. Hedges. Ozone generation by rock fracture: earthquake early warning? *Applied Physics Letters*, 99:204101, 2011. doi: 10.1063/1.3660763.

- [78] International Commission on Earthquake Forecasting for Civil Protection (ICEF). Operational earthquake forecasting: state of knowledge and guidelines for utilization. *Annals of Geophysics*, 54(4):315–391, 2011. doi: 10.4401/ag-5350.
- [79] C.H. Scholz. The frequency-magnitude relation of microfracturing in rock and its relation to earthquakes. *Bulletin of the Seismological Society of America*, 58(1): 399–415, 1968. doi: 10.1029/JB082i020p02981.
- [80] I.G. Main, P.G. Meredith, and C. Jones. A reinterpretation of the precursory seismic b -value anomaly from fracture mechanics. *Geophysical Journal International*, 96 (1):131–138, 1989. doi: 10.1111/j.1365-246X.1989.tb05255.x.
- [81] M. Imoto. Changes in the magnitude-frequency b -value prior to large ($M \geq 6.0$) earthquakes in Japan. *Tectonophysics*, 193(4):311–325, 1991. doi: 10.1016/0040-1951(91)90340-X.
- [82] E.S. Cochran, J.E. Vidale, and S. Tanaka. Earth tides can trigger shallow thrust fault earthquakes. *Science*, 306(5699):1164–1166, 2004. doi: 10.1126/science.1103961.
- [83] S. Tanaka. Tidal triggering of earthquakes precursory to the recent Sumatra megathrust earthquakes of 26 December 2004 (M_w 9.0), 28 March 2005 (M_w 8.6), and 12 September 2007 (M_w 8.5). *Geophysical Research Letters*, 37(2):L02301, 2010. doi: 10.1029/2009GL041581.
- [84] S. Ide, S. Yabe, and Y. Tanaka. Earthquake potential revealed by tidal influence on earthquake size-frequency statistics. *Nature Geoscience*, 9:834–837, 2016. doi: 10.1038/ngeo2796.
- [85] S. Tanaka. Tidal triggering of earthquakes prior to the 2011 Tohoku-Oki earthquake (M_w 9.1). *Geophysical Research Letters*, 39:L00G26, 2012. doi: 10.1029/2012GL051179.
- [86] M. Kennedy, J.E. Vidale, and M.G. Parker. Earthquakes and the moon: Syzygy predictions fail the test. *Seismological Research Letters*, 75(5):607–612, 2004. doi: 10.1785/gssrl.75.5.607.
- [87] F. Zhan, Y. Zhu, J. Ning, J. Zhou, W. Liang, and Y. Xu. Gravity changes before large earthquakes in China: 1998-2005. *Geo-spatial Information Science*, 14(1): 1–9, 2011. doi: 10.1007/s11806-011-0440-0.
- [88] M. Shahrisvand, M. Akhoondzadeh, and M. A. Sharifi. Detection of gravity changes before powerful earthquakes in GRACE satellite observations. *Annals of Geophysics*, 57(5):A0543, 2014. doi: 10.4401/ag-6612.
- [89] J. Havkov and G. Aguacil. *Instrumentation in Earthquake Seismology*. Springer, Cham, Switzerland, 2016. doi: 10.1007/978-3-319-21314-9.
- [90] A. Valencio, C. Grebogi, and M.S. Baptista. A fast-track glimpse to the life-cycle of modelling complex systems: hands-on, 2017. Presentation at the ICTP-SAI FR School on Nonlinear Time Series Analysis and Complex Networks in the Big Data Era, available at: <http://www.ictp-saifr.org/school-on-nonlinear-time-series-analysis-and-complex-networks-in-the-big-data-era/>.

- [91] T.M. Niebaudel, G.S. Sasagawa, J.E. Fallel, R. Hilt, and E. Klopping. A new generation of absolute gravimeters. *Metrologia*, 32(3):159–180, 1995. doi: 10.1088/0026-1394/32/3/004.
- [92] J.M. Goodkind. The superconducting gravimeter. *Review of Scientific Instruments*, 70(11):4131–4152, 1999. doi: 10.1063/1.1150092.
- [93] D. Cheloni, R. Giuliani, E. D’Anastasio, S. Atzori, R.J. Walters, L. Bonci, N. D’Agostino, M. Mattone, S. Calcaterra, P. Gambino, F. Deninno, R. Maseroli, and G. Stefanelli. Coseismic and post-seismic slip of the 2009 L’Aquila (Central Italy) M_w 6.3 earthquake and implications for seismic potential along the Campotosto fault from joint inversion of high-precision levelling, InSAR and GPS data. *Tectonophysics*, 622:168–185, 2014. doi: 10.1016/j.tecto.2014.03.009.
- [94] D.C. Agnew. Strainmeters and tiltmeters. *Reviews of Geophysics*, 24(3):579–624, 1986. doi: 10.1029/RG024i003p00579.
- [95] H.J. Kämpel. About the potential of wells to reflect stress variations within inhomogeneous crust. *Tectonophysics*, 211:317–336, 1992. doi: 10.1016/0040-1951(92)90068-H.
- [96] D.E. Cartwright and A.C. Eden. Corrected tables of tidal harmonics. *Geophysical Journal International*, 33(3):253–264, 1973. doi: 10.1111/j.1365-246X.1973.tb03420.x.
- [97] Y. Tamura. A Harmonic Development of the Tide-Generating Potential. *Bulletin d’Informations Mareés Terrestres*, 99:6813–6855, 1987.
- [98] T. Hartmann and H.G. Wenzel. The hw95 tidal potential catalogue. *Geophysical Research Letters*, 22(24):3553–3556, 1995. doi: 10.1029/95GL03324.
- [99] E. Spiridonov, O. Vinogradova, E. Boyarskiy, and L. Afanasyeva. Atlantida3.1_2014 for Windows: a software for tidal prediction. *Bulletin d’Informations Mareés Terrestres*, 149:12062–12081, 2015.
- [100] L. Carrere, M. F. Lyard, L. A. Guillot, and F. L. Roulou. FES2012: A new global tidal model taking advantage of nearly 20 years of altimetry. In *Proceedings of meeting "20 Years of Altimetry"*, page 5, Venice, 2012.
- [101] H.G. Wenzel. The nanogal software: Earth tide data processing package ETERNA 3.30. *Bulletin d’Informations Mareés Terrestres*, 124:9425–9439, 1996.
- [102] A. Venedikov, J. Arnosó, and R. Vieira. Program VAV-2000 for Tidal Analysis of Unevenly Spaced Data with Irregular Drift and Colored Noise. *Journal of the Geodetic Society of Japan*, 47(1):281–286, 2001.
- [103] Y. Tamura and D. C. Agnew. *Baytap08 User’s Manual*. Scripps Institution of Oceanography, UC San Diego, La Jolla (CA), 2008. Available at <http://escholarship.org/uc/item/4c27740c>. [Accessed 03 Mar 2018].
- [104] D.L. Codiga. Unified tidal analysis and prediction using the UTide Matlab functions. Technical report, Graduate School of Oceanography, University of Rhode Island, Narragansett, RI, 2011.

- [105] F. Takens. Detecting strange attractors in turbulence. In F. Takens, editor, *Dynamical Systems and Turbulence – Warwick 1980*, volume 898 of *Lecture Notes in Mathematics*, pages 366–381. Springer, Berlin, 1981.
- [106] M.B Kennel, R. Brown, and H.D.I. Abarbanel. Determining embedding dimension for phase-space reconstruction using a geometrical construction. *Physical Review A*, 45:3403, 1992. doi: 10.1103/PhysRevA.45.3403.
- [107] H.D.I. Abarbanel. *Analysis of observed chaotic data*. Springer, New York, 1996.
- [108] M.T. Rosenstein, J.J. Collins, and C.J. Luca. A practical method for calculating largest lyapunov exponents from small data sets. *Physica D*, 65(1-2):117–134, 1993. doi: 10.1016/0167-2789(93)90009-P.
- [109] R.M. May. Simple mathematical models with very complicated dynamics. *Nature*, 261:459–467, 1976. doi: 10.1038/261459a0.
- [110] C. Grebogi, E. Ott, and J.A. Yorke. Crises, sudden changes in chaotic attractors, and transient chaos. *Physica D*, 7(1-3):181–200, 1983. doi: 10.1016/0167-2789(83)90126-4.
- [111] Y. Pomeau and P. Manneville. Intermittent transition to turbulence in dissipative dynamical systems. *Communications in Mathematical Physics*, 74(2):189–197, 1980. doi: 10.1007/BF01197757.
- [112] K. Kaneko. Simulating physics with coupled map lattices. In A. Onuki, K. Kyozi, and S. Masuo, editors, *Formation, Dynamics and Statistics of Patterns*, volume 1, page 1. World Scientific, Singapore, 1990. doi: 10.1142/9789814368223_0001.
- [113] E. Bianco-Martinez. *Information, causality, and observability approaches to understand complex systems*. PhD thesis, University of Aberdeen, Aberdeen, 2015.
- [114] E. Bianco-Martinez and M.S. Baptista. Space-time nature of causality. arXiv:1612.05023. (To appear in Chaos).
- [115] C. Wang, C. Grebogi, and M.S. Baptista. Uncovering hidden flows in physical networks. *EPL (Europhysics Letters)*, 118:58001, 2017. doi: 10.1209/0295-5075/118/58001.
- [116] C. Shannon. A mathematical theory of communication. *Bell System Technical Journal*, 27(3):379–423, 1948. doi: 10.1002/j.1538-7305.1948.tb01338.x.
- [117] C. Shannon. A mathematical theory of communication. *Bell System Technical Journal*, 27(4):623–666, 1948. doi: 10.1002/j.1538-7305.1948.tb00917.x.
- [118] T. Schreiber. Measuring information transfer. *Physical Review Letters*, 85(2):461–464.
- [119] J.G. Orlandi, O. Stetter, J. Soriano, T. Geisel, and D. Battaglia. Transfer entropy reconstruction and labeling of neuronal connections from simulated calcium imaging. *PLoS One*, 9(6):e98842, 2014. doi: 10.1371/journal.pone.0098842.

- [120] R. Marschinski and H. Kantz. Analysing the information flow between financial time series. *European Physical Journal B*, 30(2):275–281, 2002. doi: 10.1140/epjb/e2002-00379-2.
- [121] J. Borge-Holthoefer, N. Perra, B. Gonçalves, S. González-Bailón, A. Arenas, Y. Moreno, and A. Vespignani. The dynamics of information-driven coordination phenomena: a transfer entropy analysis. *Science Advances*, 2(4):e1501158, 2016. doi: 10.1126/sciadv.1501158.
- [122] L. Barnett, A.B. Barrett, and A.K. Seth. Granger causality and transfer entropy are equivalent for gaussian variables. *Physical Review Letters*, 203:238701, 2009. doi: 10.1103/PhysRevLett.103.238701.
- [123] N. Rubido, C. Grebogi, and M.S. Baptista. Entropy-based generating markov partitions for complex systems. *Chaos*, 28:033611, 2018. doi: 10.1063/1.5002097.
- [124] F.S. Matias, P.V. Carelli, C.R. Mirasso, and M. Copelli. Anticipated synchronization in neuronal circuits unveiled by a phase-response-curve analysis. *Physical Review E*, 95:052410, 2017. doi: 10.1103/PhysRevE.95.052410.
- [125] C. Masoller and D.H. Zanette. Anticipated synchronization in coupled chaotic maps with delays. *Physica A*, 300(3-4):359–366, 2001. doi: 10.1016/S0378-4371(01)00362-4.
- [126] R.L. Dobrushin. General formulation of Shannon’s main theorem in information theory. *Uspekhi Matematicheskikh Nauk*, 14(6(90)):3–104, 1959.
- [127] R.M. Gray and J.C. Kieffer. Asymptotically mean stationary measures. *The Annals of Probability*, 8(5):962–973, 1980.
- [128] K.W. Church and P. Hanks. Word association norms, mutual information and lexicography. *Computational Linguistics*, 16(1):22–29, 1990.
- [129] M.S. Baptista and J. Kurths. Transmission of information in active networks. *Physical Review E*, 77:026205, 2008. doi: 10.1103/PhysRevE.77.026205.
- [130] E.L. Quarantelli. *What is a disaster?: perspectives on the question*. Routledge, London, 1998.
- [131] C. Gilbert. Studying disaster: changes in the main conceptual tools. In E.L. Quarantelli, editor, *What is a disaster?: perspectives on the question*, pages 11–18. Routledge, London, 1998.
- [132] N. Valencio. Disasters, social order and planning in civil defence: the Brazilian context (Desastres, ordem social e planejamento em defesa civil: o contexto brasileiro). *Saúde e Sociedade*, 19(4):748–762, 2010. doi: 10.1590/S0104-12902010000400003. Original text in Portuguese.
- [133] N. Valencio. From the risk area to the temporary shelter: analysis of the conflicts due to a precarious territoriality (*Da ‘área de risco’ ao abrigo temporário: uma análise dos conflitos subjacentes a uma territorialidade precária*). In N. Valencio,

- M. Siena, and V. Marchezini, editors, *Sociology of Disasters: construction, interfaces and perspectives in Brazil (Sociologia dos Desastres: construção, interfaces e perspectivas no Brasil)*, volume 1. RiMa, São Carlos, Brazil, 2009. Original text in Portuguese.
- [134] P. Bourdieu. *The symbolic power (O poder simbólico)*. Bertrand Brasil, Rio de Janeiro, 2004. Original text in Portuguese.
- [135] A. Peters, K.Y. Chung, and S. Chu. Measurement of gravitational acceleration by dropping atoms. *Nature*, 400(6747):849–852, 1999. doi: 10.1038/23655.
- [136] B. Canuel, F. Leduc, D. Holleville, A. Gauguet, J. Fils, A. Viridis, A. Clairon, N. Dimarcq, C.J. Bordé, A. Landragin, and P. Bouyer. Six-axis inertial sensor using cold-atom interferometry. *Physical Review Letters*, 97(1):010402, 2006. doi: 10.1103/PhysRevLett.97.010402.
- [137] M. de Angelis, A. Bertoldi, L. Cacciapuoti, A. Giorgini, G. Lamporesi, M. Prevedelli, G. Saccorotti, F. Sorrentino, and G.M. Tino. Precision gravimetry with atomic sensors. *Measurement Science and Technology*, 20(2):022001, 2009. doi: 10.1088/0957-0233/20/2/022001.
- [138] H.J. Metcalf and P. van der Straten. *Laser Cooling and Trapping*. Springer, London, 2001.
- [139] C. Wieman, G. Flowers, and S. Gilbert. Inexpensive laser cooling and trapping experiment for undergraduate laboratories. *American Journal of Physics*, 63(4): 317–330, 1995. doi: 10.1119/1.18072.
- [140] M. Kasevich and S. Chu. Atomic interferometry using stimulated raman transitions. *Physical Review Letters*, 67(2):181–184, 1991. doi: 10.1103/PhysRevLett.67.181.
- [141] T. Farah, C. Guerlin, A. Landragin, P. Bouyer, S. Gaffet, F. Pereira Dos Santos, and S. Merlet. Underground operation at best sensitivity of the mobile Ine-syrte cold atom gravimeter. *Gyroscopy and Navigation*, 5(4):266–274, 2014. doi: 10.1134/S2075108714040051.
- [142] S.B. Cahn, A. Kumarakrishnan, U. Shim, T. Sleator, P.R. Berman, and B. Dubetsky. Time-domain de broglie wave interferometry. *Physical Review Letters*, 79(5):784–787, 1997. doi: 10.1103/PhysRevLett.79.784.
- [143] B. Barrett, I. Chan, and A. Kumarakrishnan. Atom-interferometric techniques for measuring uniform magnetic field gradients and gravitational acceleration. *Physical Review A*, 84:063623, 2011. doi: 10.1103/PhysRevA.84.063623.

Appendix A

Comment on definitions of disasters and media representation

Disasters, hazards, risks, vulnerability and susceptibility are terms used interchangeably in common language, but which hold different meanings in the field of Disaster Studies. Being it an interdisciplinary area, to appropriately make this distinction is of vital importance, so to appropriately identify the contribution from each actor given its competence. Misidentification of the concepts and roles lead to an understanding of the field limited to the common sense and to inadequate strategies. A typical example is of technical fields imposing solutions for the “social part” without regard for the social composition, social norms, interpretations of a message, relationship between different social groups and state actors, and different levels of damaging depending on the societal structure.

The coordination of the disaster operational aspects is typically done by the Civil Defence or Protection, which is the institution or set of institutions responsible for crisis management, emergency planning and safeguarding of the civil society. It can be a permanent institution, such as the Brazilian “Proteção e Defesa Civil”, the German “Technisches Hilfswerk” (THW) and the US Federal Emergency Management Agency (FEMA), or called on time, such as in the UK, where the Civil Protection is a shared responsibility of many actors (Civil Contingencies Secretariat, NHS, Police, Firefighters, Environment Agency) but a central crisis cabinet can be defined in midst of an event.

Although there is no consensus on the specific definition of a *disaster*, it is commonly

accepted to be consisted of a tragic, disruptive, social event [130]. Furthermore, a disaster can only be referred as such after its occurrence. A hazard, on the other hand, refers to a physical event with potential to cause damage. Note that a hazard can have no social consequences, therefore, it is possible to have a hazard without a disaster. It is the case, for example, of the occurrence of large earthquakes in low-populated areas. The hazard level may be considerable, as measured by parameters such as the earthquake magnitude or else, but likely to have no disaster consequences. Additionally, a hazard is also only defined when the physical event has already started to occur. On contrary, when referring to the probability of future occurrence of a physical event, the quantity being analysed is the *risk*. Depending on the case, a different set of tools might be employed for risk analysis.

The distinction of hazard, which is in essence a physical phenomena, and disaster, which is a social crisis, is essential to avoid conflicts between different technical operators and scientific areas. The body of this thesis concerns with the study of hazards, more specifically the seismic hazard, using tools and definitions from Physics and Geophysics. It does not analyse disaster cases (involving in-depth social analysis) neither does it evaluate risk scenarios (which involves forecasting models). The focus, instead, is in improving the state-of-art knowledge on the seismic hazard and its relations with other elements (precursory signal candidates). This may support both the technical actors concerned with the disasters to make informed choices by associating this new knowledge with their own expertise of the societal structure, and the technical actors engaged in risk analysis to redesign their forecasting analysis tools and improve potential differential damaging scenarios.

The last distinction is between vulnerability and susceptibility. Susceptibility is the exposition of the subject to a threat. For example, an individual is susceptible to the earthquake risk if he is in an earthquake-prone area. This condition alone is insufficient to render this individual vulnerable, as he can be, for example, in a well designed building, able to resist to the strongest ground motions. Hence, *vulnerability*, although it is term (like disaster) of difficult consensual definition [131], conveys this idea of the state of

an individual in face of a threat (either a risk or an incoming hazard). When comparing vulnerability and susceptibility, also the general principle of *resilience* arises, as it relates to this ability of the individuals to resist to a hazard or cope with its adverse effects.

Three paradigms have historically developed to interpret disasters [131]: disaster as a war, disaster as social vulnerability and disaster as uncertainty. The first consists in comparing the hazard as an enemy, threatening the citizens. This is typically the first approach implemented by Civil Defence systems, supported by media language (*e.g.* “earthquake kills”) and imagery. The consequence is militarised and/or technocrat civil defence institutions, which might not always understand the intricacies of the structure of civil society, especially in cases requiring long-term solution [132]. An example of such failure in understanding the civil society is in the management of public sheltering for the affected, not always that temporary, where the logistics of barracks is strictly applied, not rarely splitting families and controlling (or not allowing) private spaces [133].

The second paradigm, disaster as social vulnerability, spread from the late 1970’s as a criticism of the first approach. In this view, disaster is not simply an event, with defined start and end chronological times, but a social process that began with the social mechanisms producing the differential vulnerability in a societal structure. This enables the construction of a Civil Defence system operated by civilians, and more engaged with sectors such as Civil Service and Health institutions. However, it is not perfect: first it is difficult, especially for newcomers in the field, to abandon the notion of a causing ‘agent’ (hazard), and second, it is not simple to measure what is vulnerability, as it can be different for different societal background. Moreover, if it is interpreted only half-way (disaster=vulnerability without accounting for the historical construction of vulnerability), it is simple to fall in the traps of victim-blaming. In this case, affected groups are blamed for not being cautious enough, when the reasons that lead to the vulnerability in first place (social inequality etc) remain unquestioned. This mistake occurs when the vulnerability is seen as a state (fixed in time) instead of a process (built over history) [15].

The third paradigm, developed from the late 1980’s, of disaster as uncertainty, focuses in the problems of evaluation, communication and institutional action in the modern

world. This approach does not constitute an opposition to the other two, like the previous case. In this view, the disaster is the rupture of the common social norm: suddenly there is the presence of a danger threatening the built environment and the citizens on it, the very complexity of society sends no or several conflicting messages of actions to be taken or not, and the individuals lose their common parameters to understand and interact with the environment. This approach brings the Civil Defence planning closer to the sciences of complexity, particularly, Architecture, Communications and Information Technology, Sociology, and Physics of Complex Systems.

It is a common misconception that the disaster scenario consists of affected individuals and emergency services working in cooperation to save them. This conception is not rarely amplified by the discourse of emergency personnel as “heroes” by government sectors and media outlets. The real scenario is more complex, defined by a field of action [134] where numerous actors not only fulfil their role but dispute for position, visibility, recognition, status, and, in ultimate instance, power. Not only collaboration occurs at the scene, but also conflicts, particularly attribution of responsibilities of damages and other negative effects between the actors. The media, expressed by the press vehicles, is in charge of observing the scenario and the disputes in the field and providing an informed panorama for the public opinion. However, the media itself is an element in the field, and must be assessed if it is, intentionally or unintentionally, privileging the voice of certain actors over others, or hiding the full picture.

As an illustration, we have analysed the portrait of the 2015-2016 UK flooding crisis case by the BBC News online reports [16], revealing at least 67 types of social actors involved in the disaster (Fig. A.1). These form complex relations of cohesion but also of conflict or concern (neutral), such as shown on Fig. A.2. A complete perspective of all the represented social relations is given in Fig. A.3. These reveal both the complexity of the field of action of a disaster and the ability of this media outlet in presenting the panorama. However, regarding the balance of voice given to individual actors, expressed by the direct/reported speech ratio, affected group members had a given voice ratio of 30%, whereas the government and emergency services had a voice ratio of around 50%.

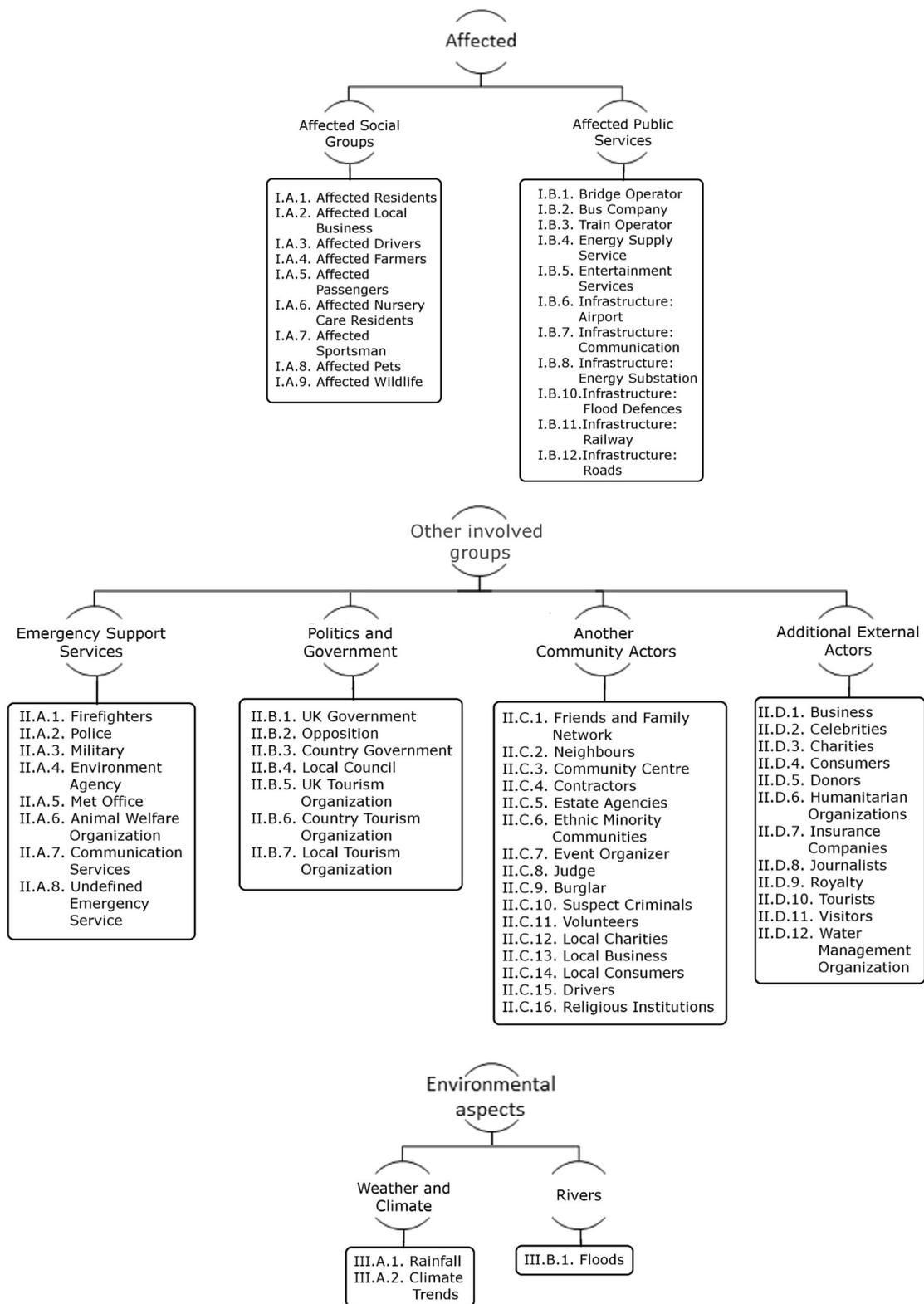
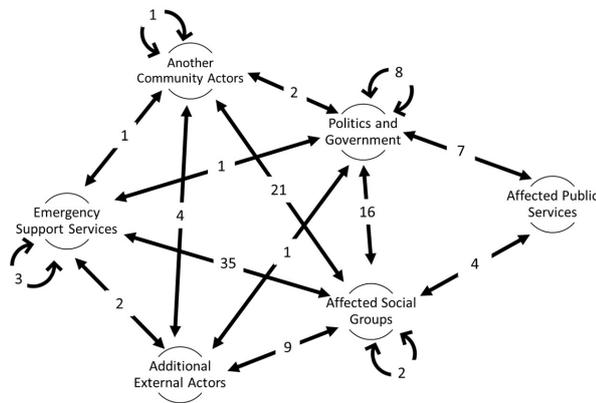
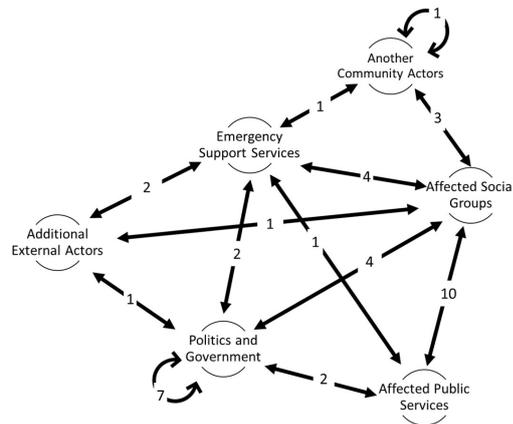


Figure A.1: Classification of the identified social actors in a sample of 51 news reports of the 2015-2016 UK flooding crisis. “Country” is a shorthand descriptor for the devolved administrations: Scotland, Wales and Northern Ireland. Figure published by the author in [16]

Interactions of Cohesion



Interactions of Conflict/Obstruction



Interactions of Concern

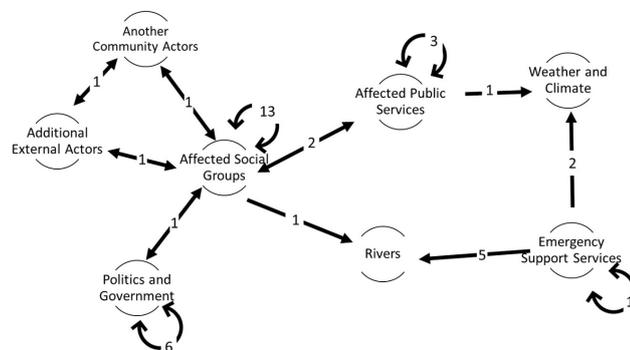


Figure A.2: Networks of social interactions for the actors involved in the 2015-2016 UK flooding crisis, according to sample of 51 news reports from BBC News. Figure published by the author in [16].

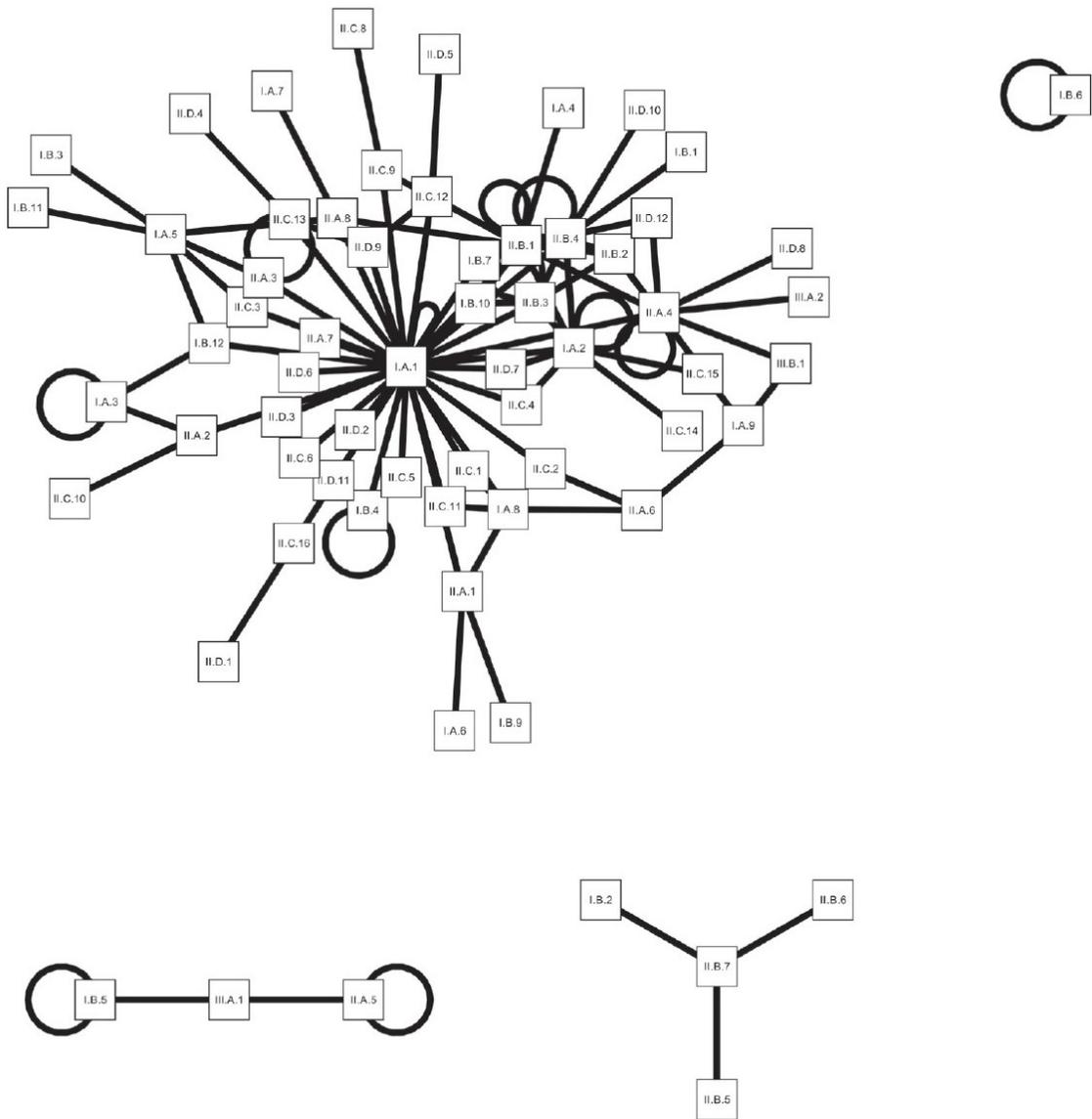


Figure A.3: Detailed network of social interactions containing the 67 social actors identified in the 2015-2016 UK flooding crisis, according to sample of 51 news reports from BBC News. Figure published by the author in [16].

Appendix B

Steps in the development of a Cold Atom Gravimeter

Gravimeters are precise accelerometers, tailored to measure the absolute value or the variations in the value of Earth's gravity acceleration. On Chapter 2 (Sec. 2.3.2) we have introduced the optical and the superconductor gravimeters. With the advent of techniques in laser cooling and guidance of atoms, there have been proposals for designs of gravimeters based on matter-wave interferometry [135–137]. Here we outline the main steps in the development of an atom-interferometry gravimeter, which was part of the research activities in the first year of the research programme. Currently, the main virtue of these proposals is the potential of such devices having high accuracy whilst remaining relatively compact, practical for field operations and possible satellite missions.

Before an atom-interferometry experiment takes place it is necessary to prepare and trap a cold atom sample. In this case, there is a high-vacuum chamber with an entry for the source of the atoms being used and windows or fibre-optical entries for the application of laser beams. Typical choices of atoms are neutral Rubidium and Strontium, as commercial glass cells with Sr/Rb vapour cells are available as atomic sources for operation at environment temperature (no need for significant heating to create a gas), these elements exhibit hyperfine structure (key for trapping), and they have strong spectral lines in the visible to near-IR bands (where it is easier to find/build precise lasers).

Once an atomic sample has been released into the chamber, six low-intensity laser

beams frequency-locked just below the peak of a strong and thin spectral line of the element are applied in the area of the sample, forming standing waves in each of the three Cartesian directions. This allows for a cooling mechanism referred as 3D optical molasses. When an atom thermally moves in any direction, the Doppler-shifted frequency of the counter-propagating laser beam matches with the atom spectral line, which undergoes a cycle of photon absorption and re-emission in a random direction (spontaneous emission), leading to a net damping force (light pressure force).

If the beams form standing waves and the beams have opposing circularly polarisation, a space-dependant potential pattern is created, making the atoms suddenly shift from “favourable” to “unfavourable” energy state as move from one area to another, release energy in the process. This additional process is the Sisyphus cooling. Further applying an asymmetric magnetic field, such as a quadrupole configuration (easily created by two circular coils with current in opposite directions), the atom spectral lines also shift slightly depending on the spatial location, and again a light pressure force makes them converge to the central point (atom trapping). A detailed theoretical account for these cooling and trapping processes can found in [138] and a simple description of how to build a device with low cost materials is presented in [139]. The result is a trapped vapour of Rubidium or Calcium atoms cooled down to μK level. Mechanisms of selection of the coldest atoms (evaporative cooling) can also be applied for further cooling to nK temperatures, reaching phase-transition to Bose-Einstein condensation, but this is not strictly needed.

After obtaining the cold atomic sample and switching off the trap, the application of additional laser pulses can be used to transition a fraction of the atomic sample into a excited state and vice-versa. This can effectively act as atom-optical components (beam-splitter, mirrors). If the process of choice is a stimulated Raman transition (two-photon absorption), a pulse of time $T = \pi/2\Omega$ (called a $\pi/2$ -pulse, Ω is the Rabi frequency of the atom-light coupling) creates a superposition of states $|energy, momentum\rangle$, $|1, p\rangle$ (original) and $|2, p + 2\hbar k\rangle$ (excited), so the atomic populations will also become spatially separated with time due to their different momentum. It means that a $\pi/2$ -pulse effectively operates as an atom beam-splitter. If, however, the pulse has twice that time,

i.e. it is a π -pulse, then populations will make the transitions: $|1, p\rangle \rightarrow |2, p + 2\hbar k\rangle$ and $|2, p + 2\hbar k\rangle \rightarrow |1, p\rangle$. This population inversion is effectively an atomic mirror, and, if the populations were previously set spatially apart by a beam-splitting process, now they will merge. This is detailed in [140] in the construction of the one of the first Mach-Zehnder atom-interferometer designs.

In the simplest atom-gravimeter design (release type), there is a vertical high-vacuum chamber and the atomic sample is initially cooled and trapped on the top. Immediately after release the Raman $\pi/2$ -pulses are applied, splitting the cloud into two populations which fall under gravity, and mid-way on the trajectory the Raman π -pulses are applied, acting as a mirror. When the two clouds combine, another Raman $\pi/2$ -pulses are applied and the total interferometer phase difference observed in the clouds is proportional to gT_{fall}^2 [141]. This process is shown in Fig. B.1. Other designs are possible, such as driving the atoms in upwards motion before the fall under gravity (fountain type) or using time-domain interferometry [142, 143], which have the advantage of enabling devices even more compact for satellite missions. However, the mechanisms for building such device and observing the interferometry can be less intuitive.

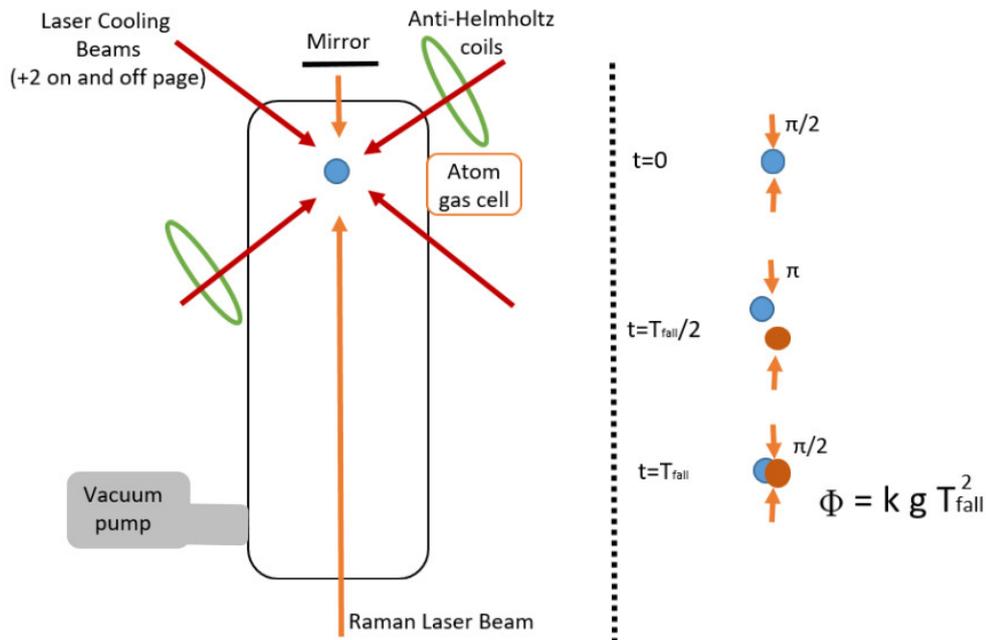


Figure B.1: General scheme of a Mach-Zehnder atom gravimeter with the atom-release configuration. Note that it retrieves the absolute gravity value. The current accuracy of a device from the Observatory of Paris is on the order of 40 nm/s^2 [141].

Appendix C

Pointwise information measures applied to precursor candidates

In this appendix, we present the pointwise information quantities from the earthquake precursor candidates: tidal variations, b-value anomalies, pre-seismic gravity changes and preceding seismicity. The definition of the pointwise information measures is presented in Chapter 4 (Sec. 4.6).

This appendix complements the findings of Chapter 6, where the causal mutual information, mutual information, transfer entropy and directionality index were applied to identify the overall information sharing and possible flow of information between seismic precursor candidates and earthquake occurrence. With the pointwise quantities shown here, it is possible to distinguish the individual data points contributing most to the overall information measures. The pointwise values are displayed next as a colour-code attributed to each data point in the phase-space plot of the precursor with the seismicity occurrence. Very complex structures can potentially be formed, such as seen for coupled logistic systems in Chapter 5, referred as causal bubbles in [113, 114] (for the case of pointwise mutual information). Such structures should not be formed in systems not causally connected.

For a small amount of geophysical data (4018 points), we are most likely unable to delineate the boundaries of such causal bubbles, but we may be able to observe if a concentration of points (in the phase-space) with high pointwise information measures

is suggested. In particular, if such concentration is also in a subregion of the precursor domain, it may potentially support earthquake prediction, by defining a range of values in the precursory activity highly linked to the seismic occurrence.

Tidal triggering of earthquakes

High tidal amplitudes leading to high magnitude events

Two peaks of CaMI were observed, on a delay of 5 days and 25 days.

Delay: 5 days

Results in Figs. C.1 and C.2 show that, although there is a concentration of higher pointwise mutual information points in the area around $2.1\mu\text{m/s}^2$ tidal amplitude and maximum magnitude above 5.6 for $L=2$, this disappears for higher symbolic lengths L . No pattern is formed for the other pointwise information quantities.

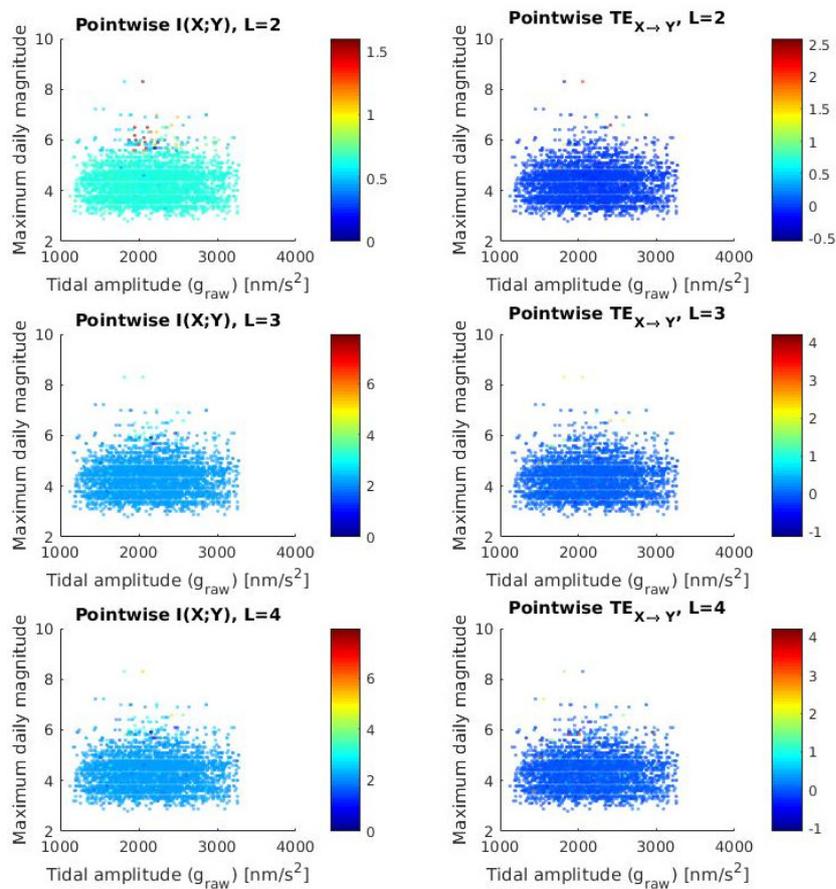


Figure C.1: Pointwise mutual information and pointwise transfer entropy for tidal amplitude (X) leading to high magnitude events (Y). Partition of tidal amplitude is $2.17\mu\text{m/s}^2$ and of magnitude is 5.6. Time-delay between the time-series is 5 days.

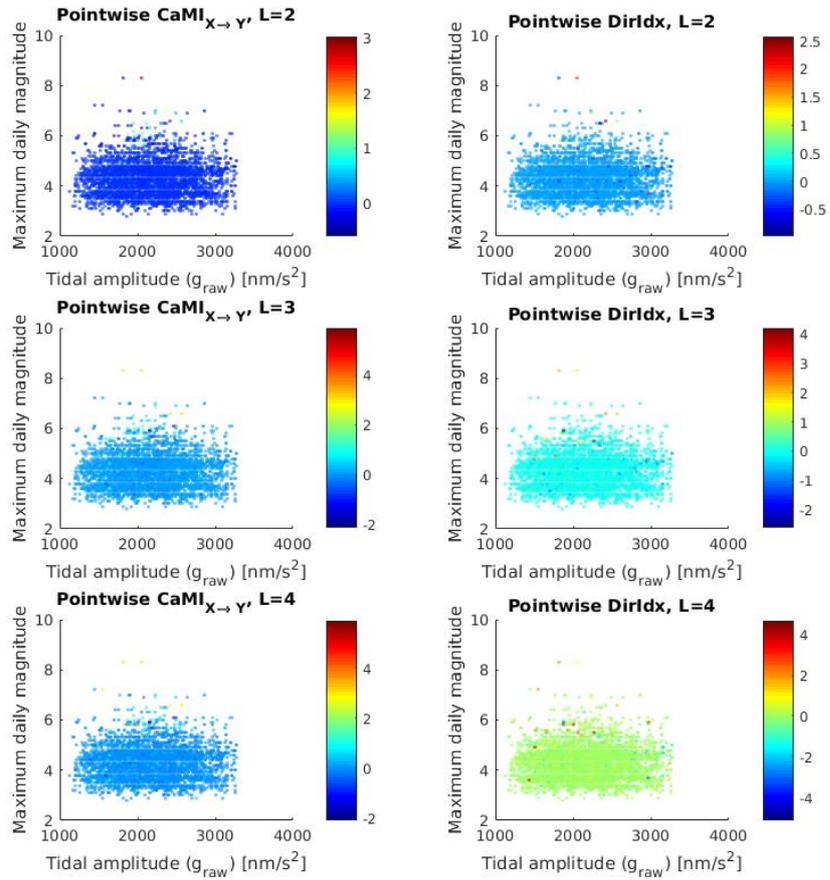


Figure C.2: Pointwise causal mutual information and pointwise directionality index for tidal amplitude (X) leading to high magnitude events (Y). Partition of tidal amplitude is $2.17 \mu\text{m/s}^2$ and of magnitude is 5.6. Time-delay between the time-series is 5 days.

Delay: 25 days

Results in Figs. C.3 and C.4 show that higher pointwise mutual information occurs for tidal amplitudes below $2.2\mu\text{m/s}^2$ associated to high earthquake magnitude ($M_w > 5.6$), whereas it is smaller than the average for tidal amplitudes above $2.2\mu\text{m/s}^2$ associated to high earthquake magnitude. This pattern occurs for all L. There were only few points with high pointwise transfer entropy, but they were located at high earthquake magnitudes. Pointwise CaMI followed the pointwise mutual information for $L=2$, but could not reveal a clear pattern for higher L. Pointwise directionality index shows no clear pattern.

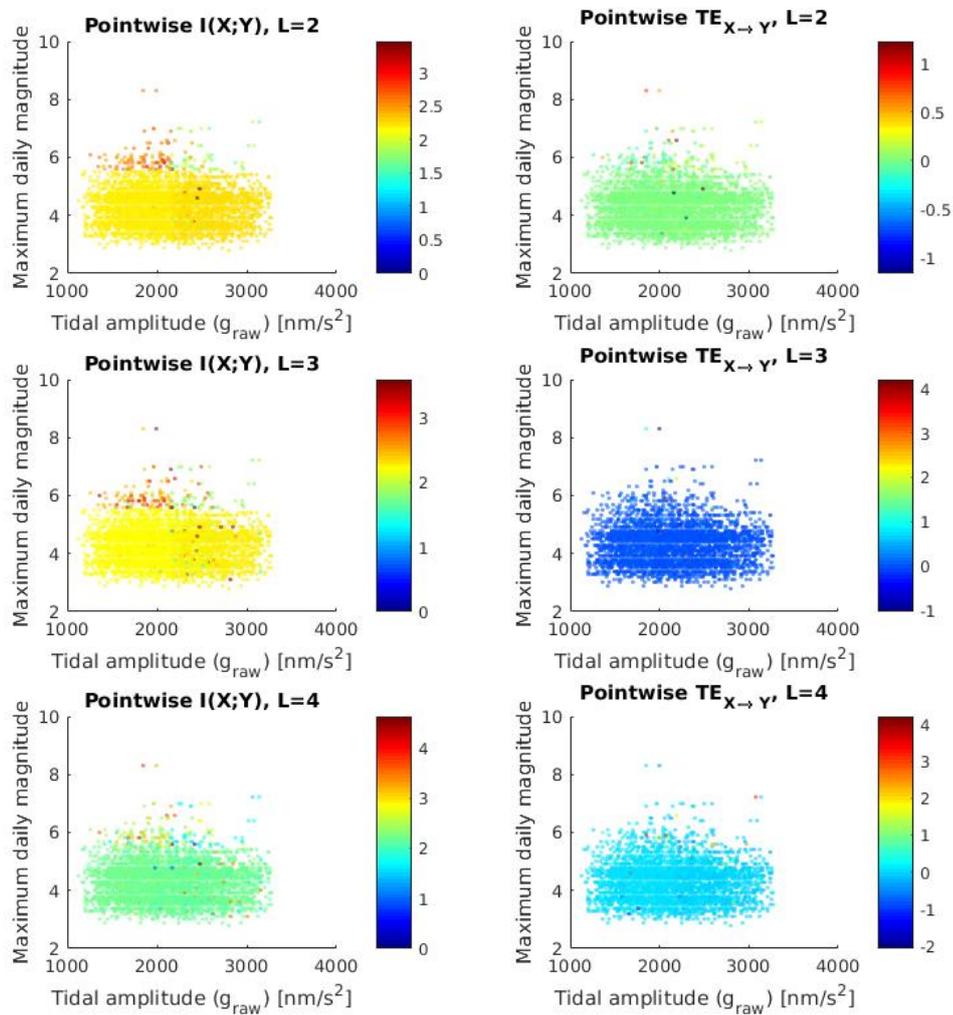


Figure C.3: Pointwise mutual information and pointwise transfer entropy for tidal amplitude (X) leading to high magnitude events (Y). Partition of tidal amplitude is $2.17\mu\text{m/s}^2$ and of magnitude is 5.6. Time-delay between the time-series is 25 days.

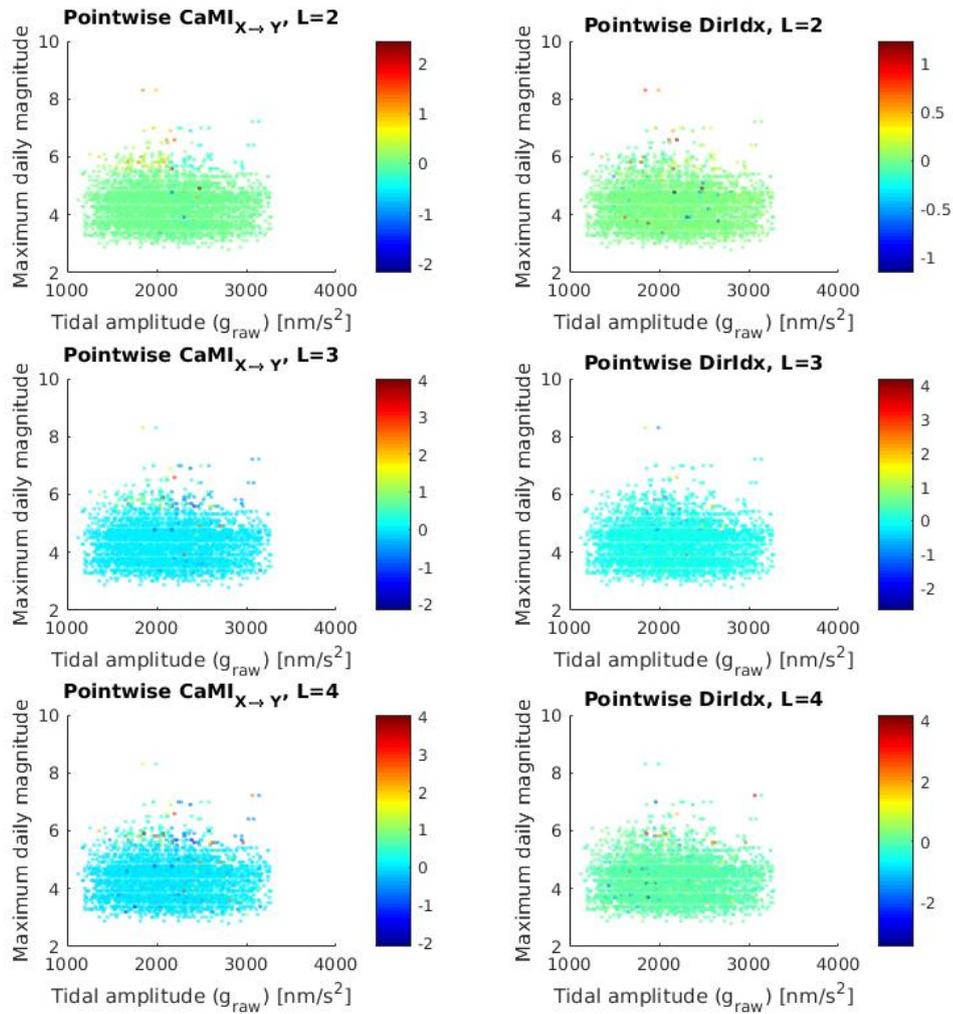


Figure C.4: Pointwise causal mutual information and pointwise directionality index for tidal amplitude (X) leading to high magnitude events (Y). Partition of tidal amplitude is $2.17 \mu\text{m/s}^2$ and of magnitude is 5.6. Time-delay between the time-series is 25 days.

High tidal amplitudes leading to high cumulative daily magnitudes

In this case the results (Figs. C.5 and C.6) show a higher pointwise mutual information for the high tidal amplitudes and high cumulative daily magnitude and lower pointwise mutual information for tidal amplitudes and high cumulative daily magnitude, for $L=2$. The pattern is progressively lost as L increases. No pattern is observed for the other information quantities, except the pointwise CaMI for $L=2$ which is dominated by the pointwise mutual information.

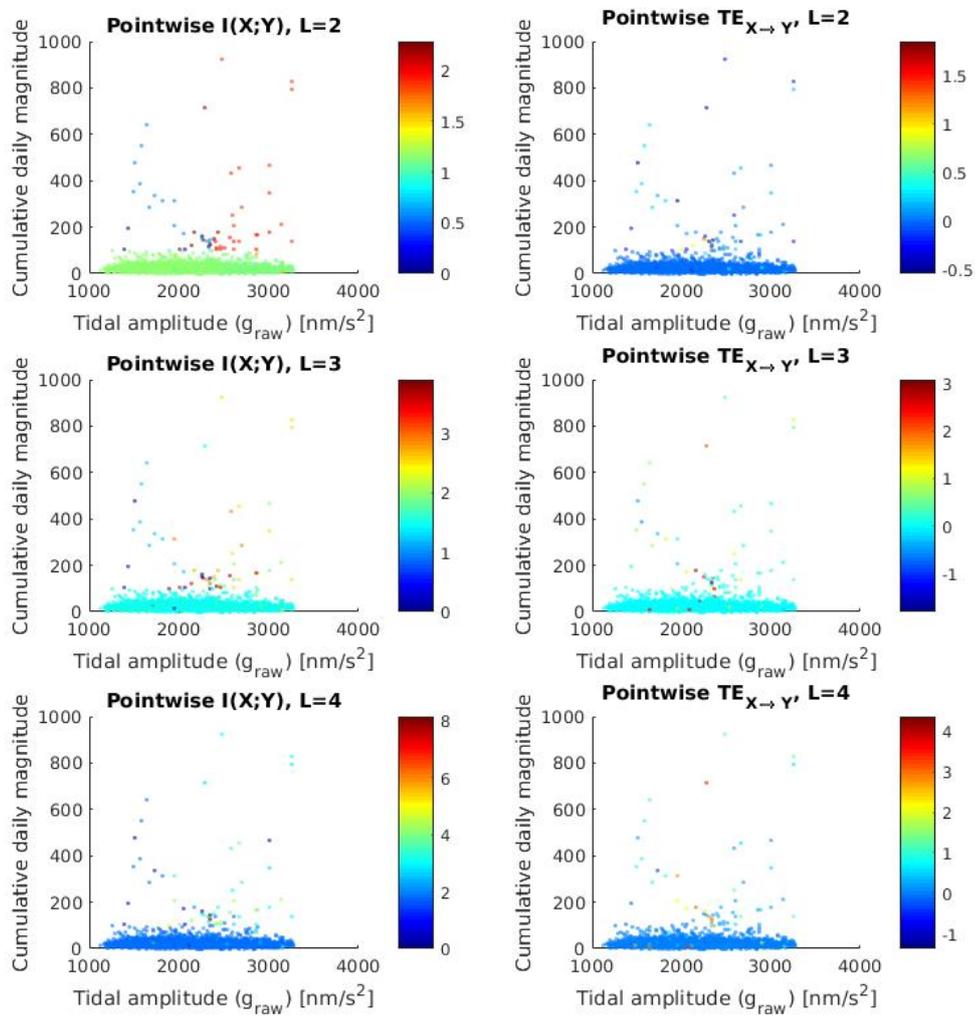


Figure C.5: Pointwise mutual information and pointwise transfer entropy for tidal amplitude (X) leading to high cumulative daily magnitude (Y). Partition of tidal amplitude is $2.37\mu m/s^2$ and of cumulative daily magnitude is 102.2. Time-delay between the time-series is 10 days.

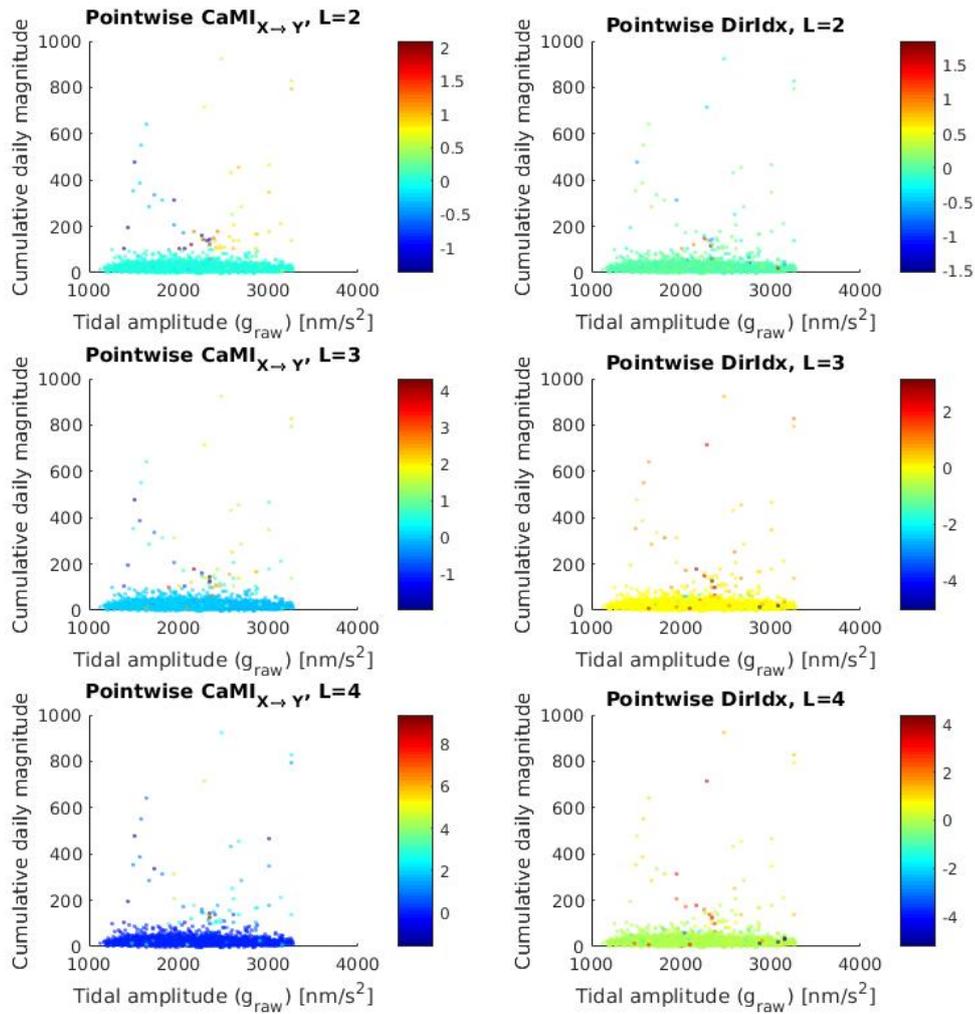


Figure C.6: Pointwise causal mutual information and pointwise directionality index for tidal amplitude (X) leading to high cumulative daily magnitude (Y). Partition of tidal amplitude is $2.37 \mu\text{m/s}^2$ and of cumulative daily magnitude is 102.2. Time-delay between the time-series is 10 days.

High tidal amplitudes leading to high seismicity rate

The results, on Figs. C.7 and C.8, show only few points with high pointwise mutual information, which tend to locate at the region of middle to high tidal amplitude associated with high seismicity rate. The combination of low tidal amplitude and low seismicity rate usually has pointwise mutual information lower than the average. Some of the points of high pointwise mutual information also have high pointwise transfer entropy. Both in pointwise CaMI and in the pointwise directionality index the combination of high tidal amplitudes and high seismicity rates yields high values. The effect is more pronounced for smaller L .

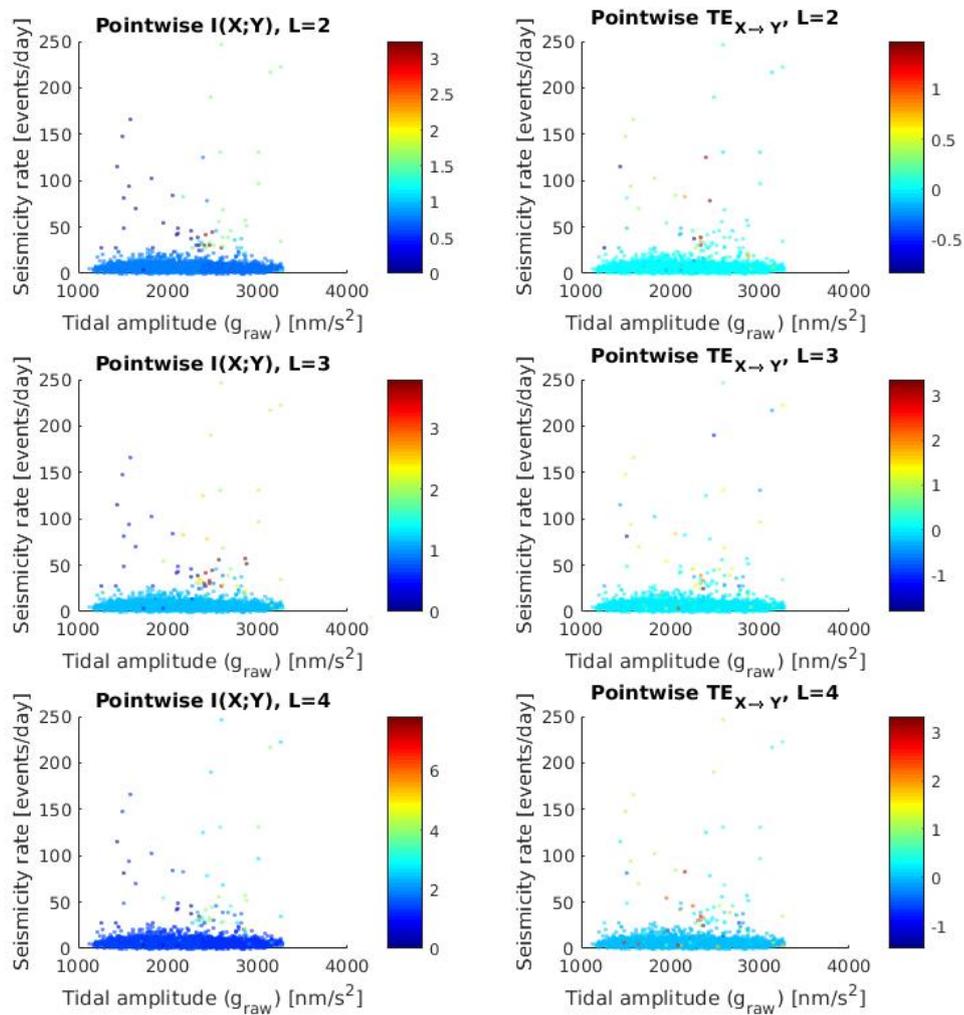


Figure C.7: Pointwise mutual information and pointwise transfer entropy for tidal amplitude (X) leading to high seismicity rate (Y). Partition of tidal amplitude is $2.37\mu\text{m/s}^2$ and of seismicity rate is 27 events/day. Time-delay between the time-series is 9 days.

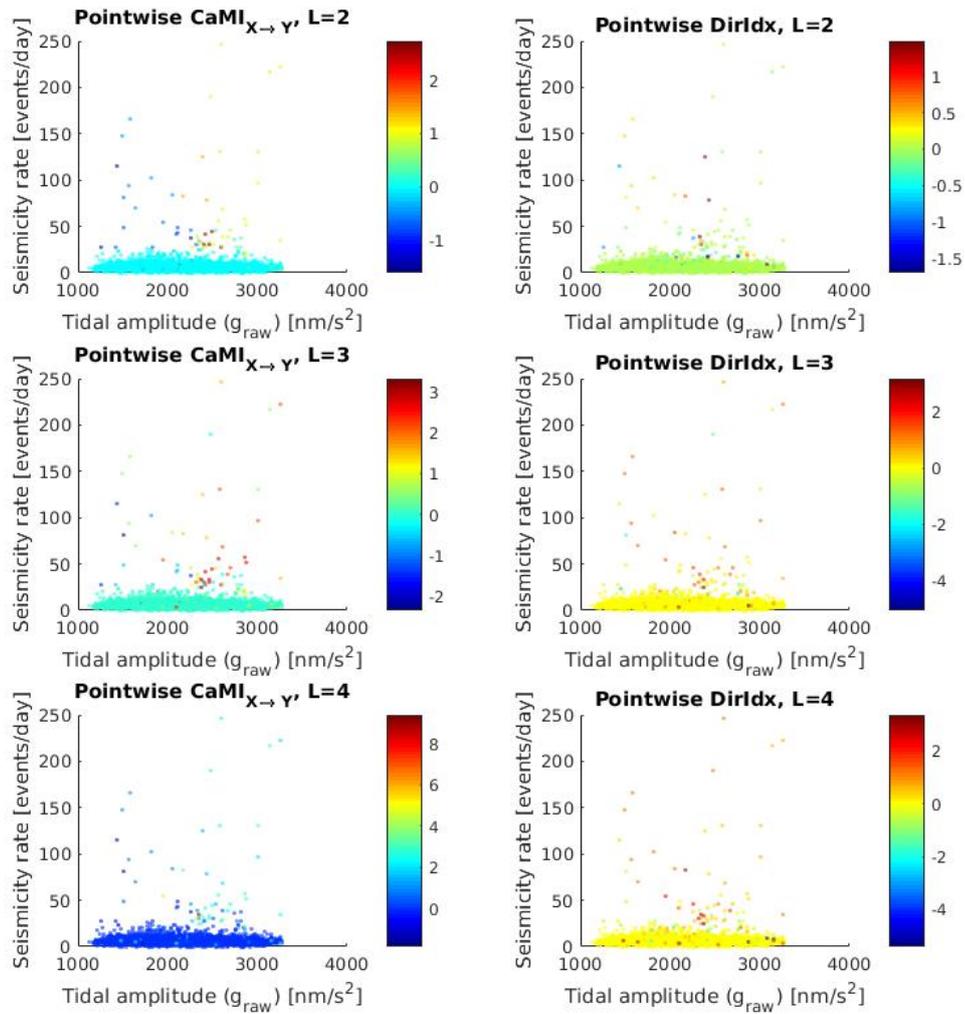


Figure C.8: Pointwise causal mutual information and pointwise directionality index for tidal amplitude (X) leading to high seismicity rate (Y). Partition of tidal amplitude is $2.37\mu m/s^2$ and of seismicity rate is 27 events/day. Time-delay between the time-series is 9 days.

Anomalies in b-value

Anomaly in b-value as precursor of high magnitude events

The results of the pointwise information measures for this hypothesis are displayed on Figs. C.9 and C.10. The low values of pointwise mutual information and transfer entropy are almost uniformly distributed. The few outliers do not exhibit any noticeable pattern. The same follows for pointwise CaMI and pointwise directionality index.

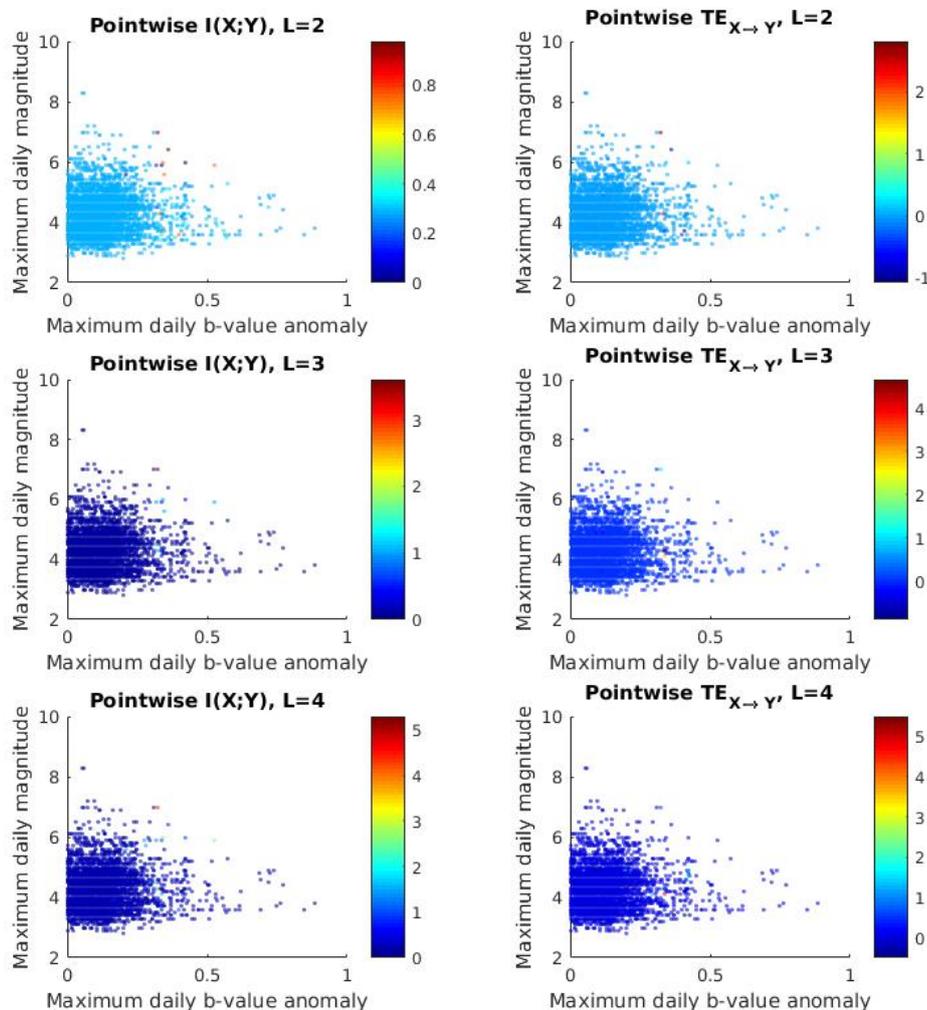


Figure C.9: Pointwise mutual information and pointwise transfer entropy for b-value anomaly (X) as precursor of high magnitude events (Y). Partition of b-value anomaly is 0.32 and of magnitude is 5.6. Time-delay between the time-series is 17 days.

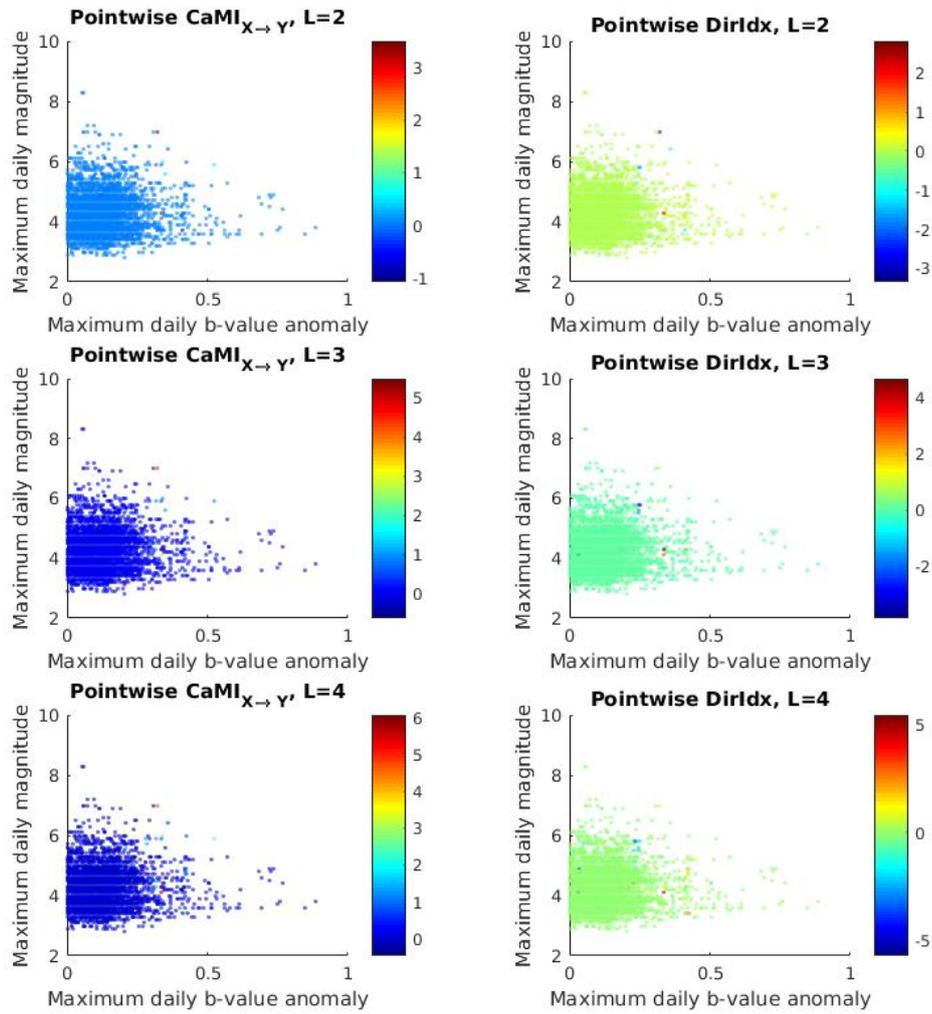


Figure C.10: Pointwise causal mutual information and pointwise directionality index for b-value anomaly (X) as precursor of high magnitude events (Y). Partition of b-value anomaly is 0.32 and of magnitude is 5.6. Time-delay between the time-series is 17 days.

Anomaly of b-value as precursor of high cumulative daily magnitudes

Figs. C.11 and C.12 present the results of pointwise information measures for this case. Only about three points of high pointwise mutual information are identified. Although they reside in the high b-value anomaly – high cumulative daily amplitude quadrant, it is insufficient to define any pattern. The pointwise transfer entropy remain low (close to zero) for these points. Pointwise CaMI follows the pointwise mutual information, but the pointwise directionality index is almost homogeneously low for all points in the dataset.

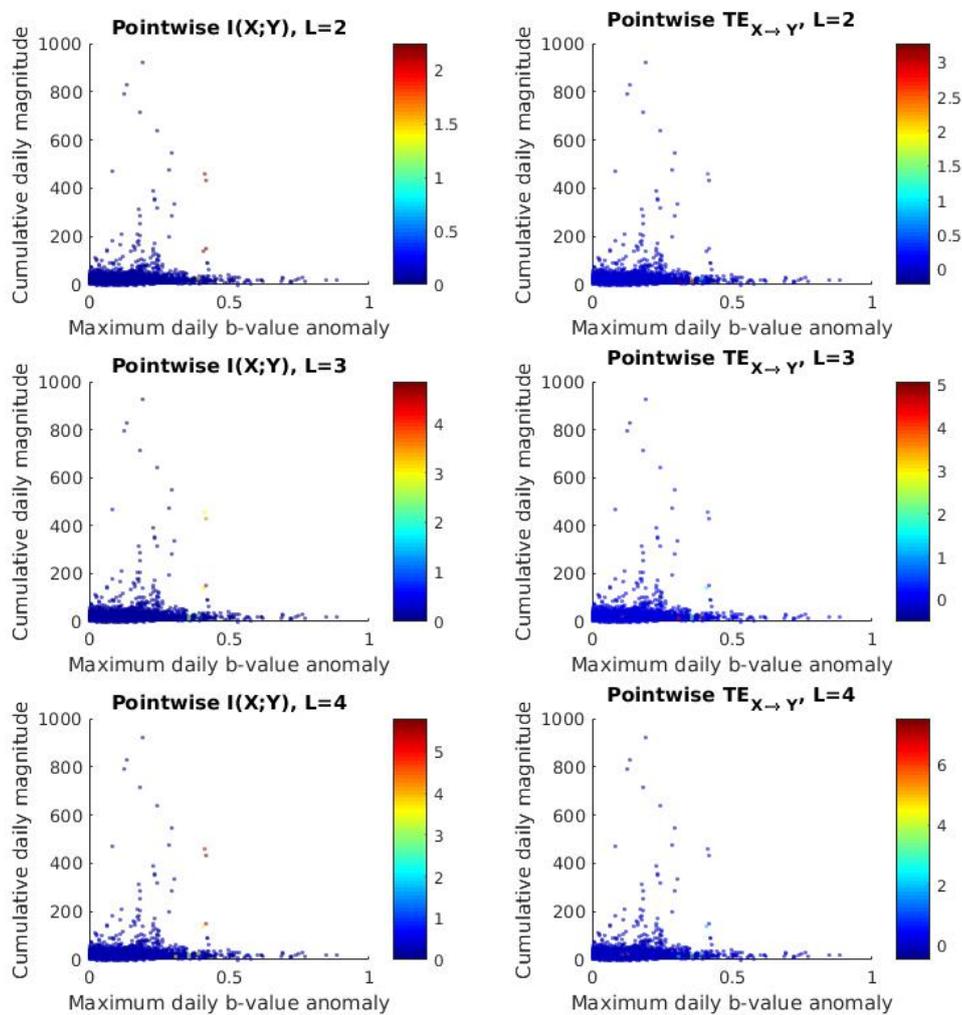


Figure C.11: Pointwise mutual information and pointwise transfer entropy for b-value anomaly (X) as precursor of high cumulative daily magnitude (Y). Partition of b-value anomaly is 0.35 and of cumulative daily magnitude is 102.2. Time-delay between the time-series is 10 days.

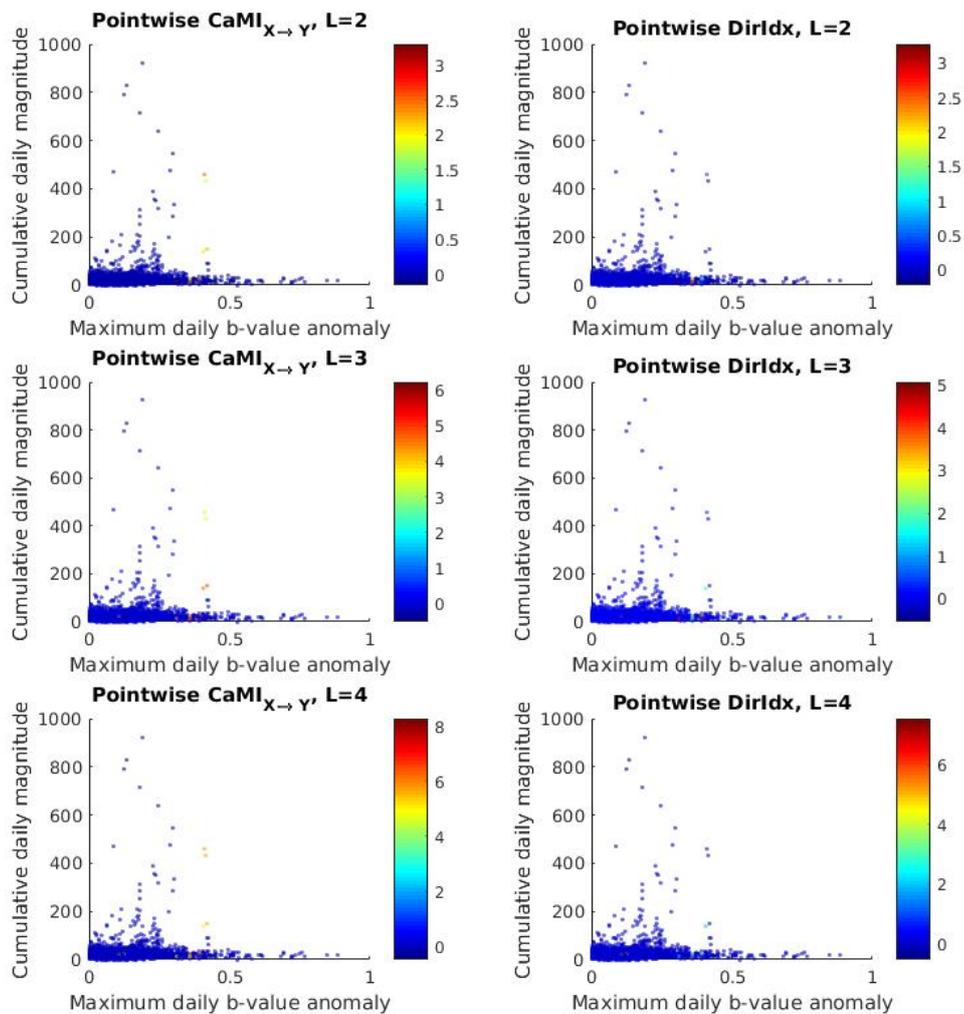


Figure C.12: Pointwise causal mutual information and pointwise directionality index for b-value anomaly (X) as precursor of high cumulative daily magnitude (Y). Partition of b-value anomaly is 0.35 and of cumulative magnitude is 102.2. Time-delay between the time-series is 10 days.

Anomaly of b-value as precursor of high seismicity rate

The results of this case are shown in Figs. C.13 and C.14. No pattern is observed for pointwise mutual information and pointwise transfer entropy. The few points diverging

from the zero mutual information and transfer entropy seem to be located at random in in the phase space. The same happens for pointwise CaMI and directionality index.

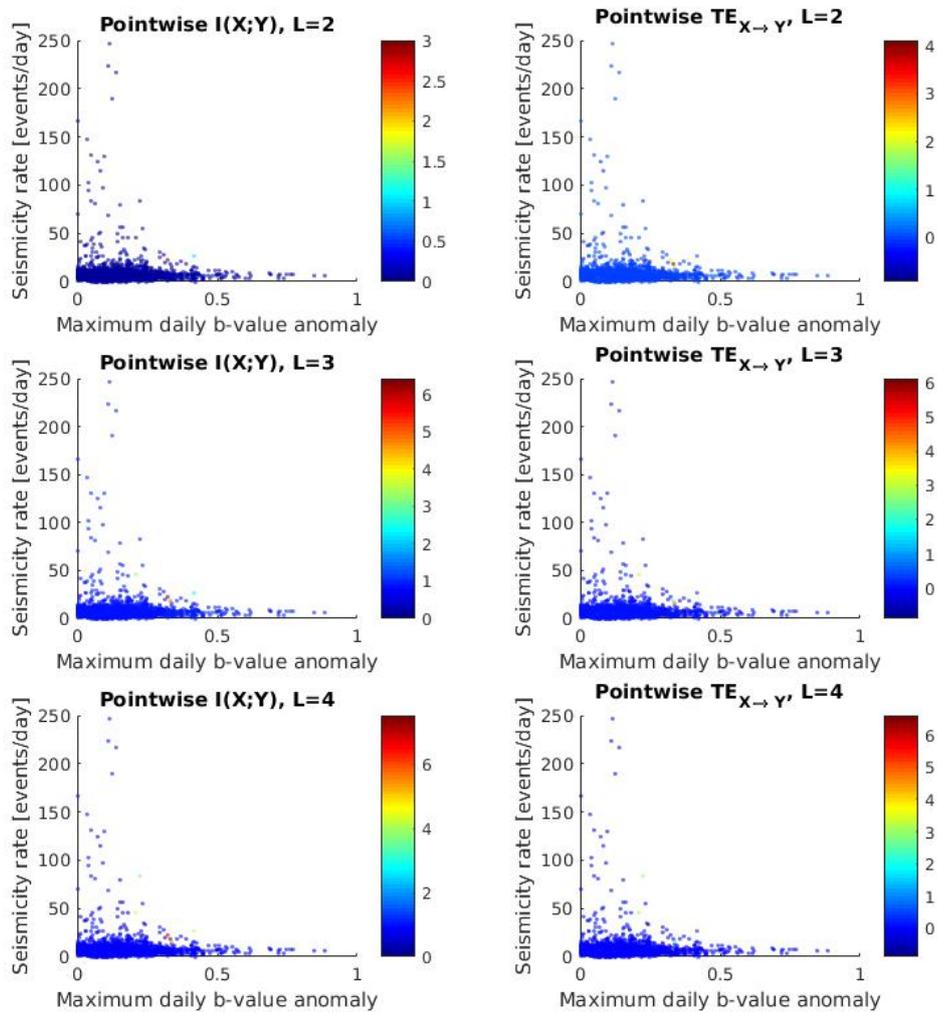


Figure C.13: Pointwise mutual information and pointwise transfer entropy for b-value anomaly (X) as precursor of high seismicity rate (Y). Partition of b-value anomaly is 0.32 and of seismicity rate is 27 events/day. Time-delay between the time-series is 23 days.

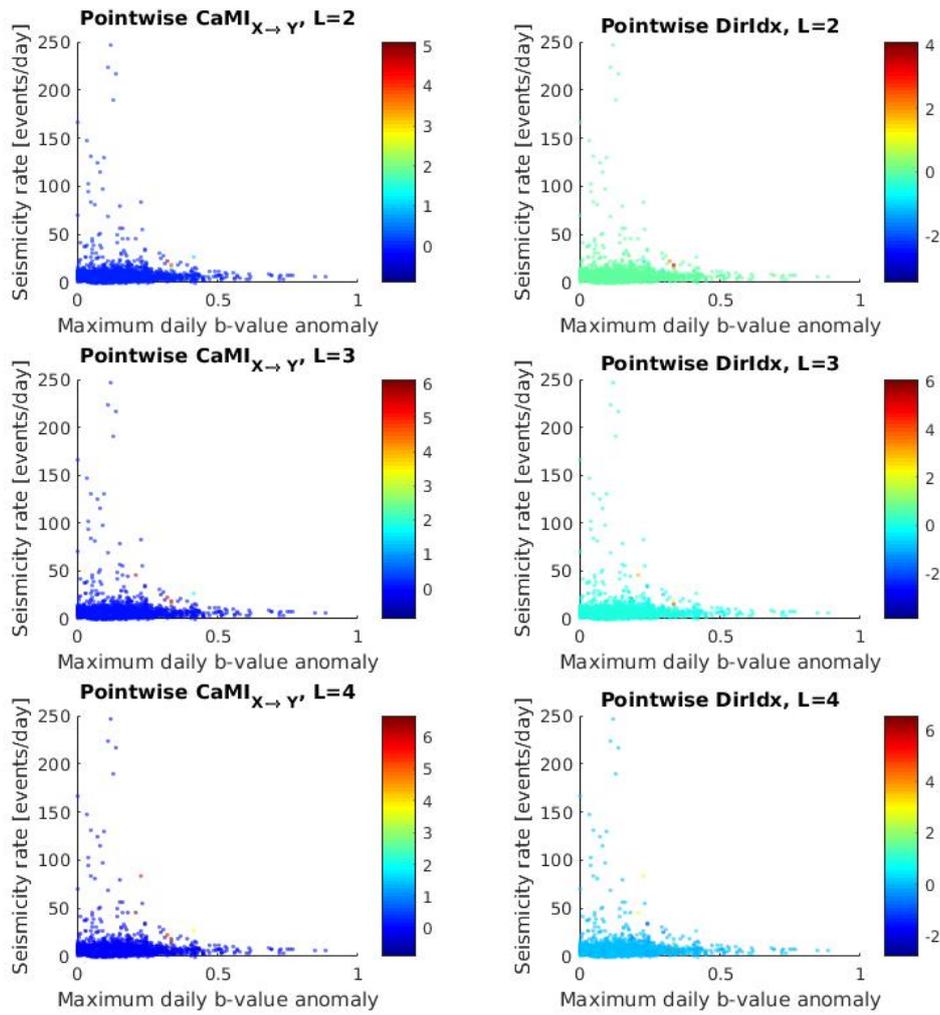


Figure C.14: Pointwise causal mutual information and pointwise directionality index for b-value anomaly (X) as precursor of high seismicity rate (Y). Partition of b-value anomaly is 0.32 and of seismicity rate is 27 events/day. Time-delay between the time-series is 23 days.

Pre-seismic gravity changes

Anomaly in gravity residuals as precursor of high magnitude events

The results of pointwise information measures for this case are shown in Figs. C.15 and C.16. The points with higher pointwise mutual information are found in the region of high gravity residuals and high maximum daily magnitude. Interestingly, the region becomes more evident as L increases, with these points concentrating in a region with gravity residuals of 30–40 nm/s^2 . Some of these points also show high pointwise transfer entropy for $L=2$. As L increases the pointwise transfer entropy pattern becomes less evident. The pointwise CaMI and directionality index follows a similar pattern of the transfer entropy.

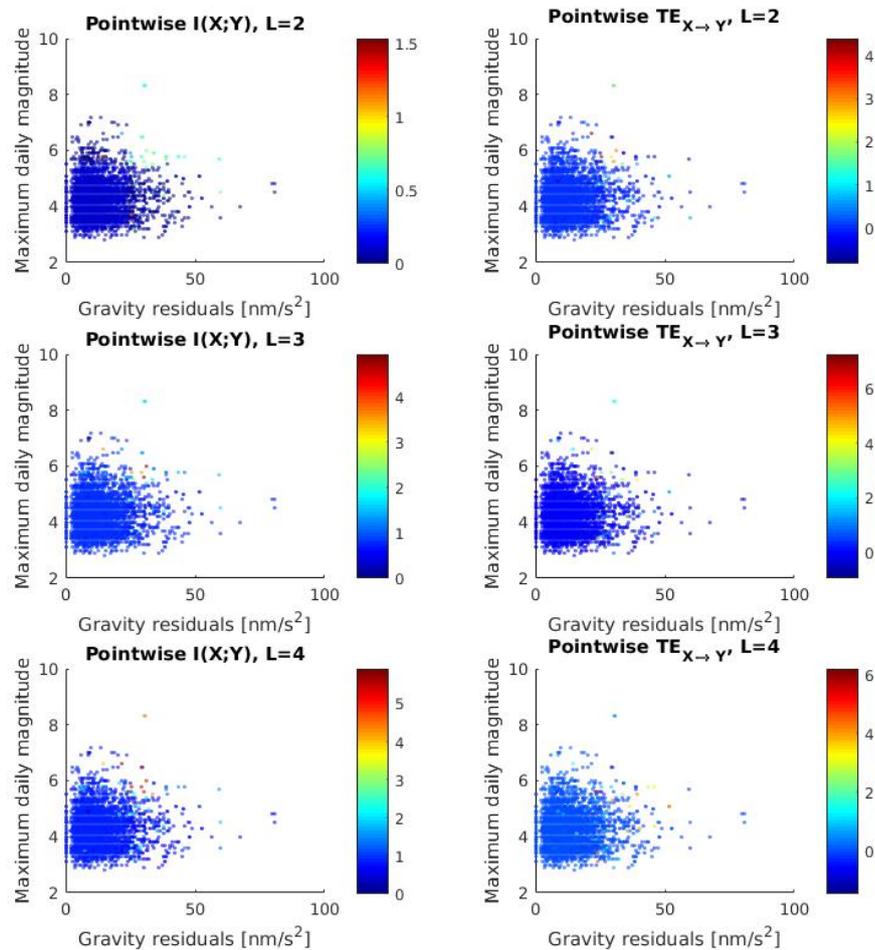


Figure C.15: Pointwise mutual information and pointwise transfer entropy for gravity residuals anomaly (X) as precursor of high magnitude events (Y). Partition of gravity residuals anomaly is 23.84nm/s^2 and of magnitude is 5.6. Time-delay between the time-series is 8 days.

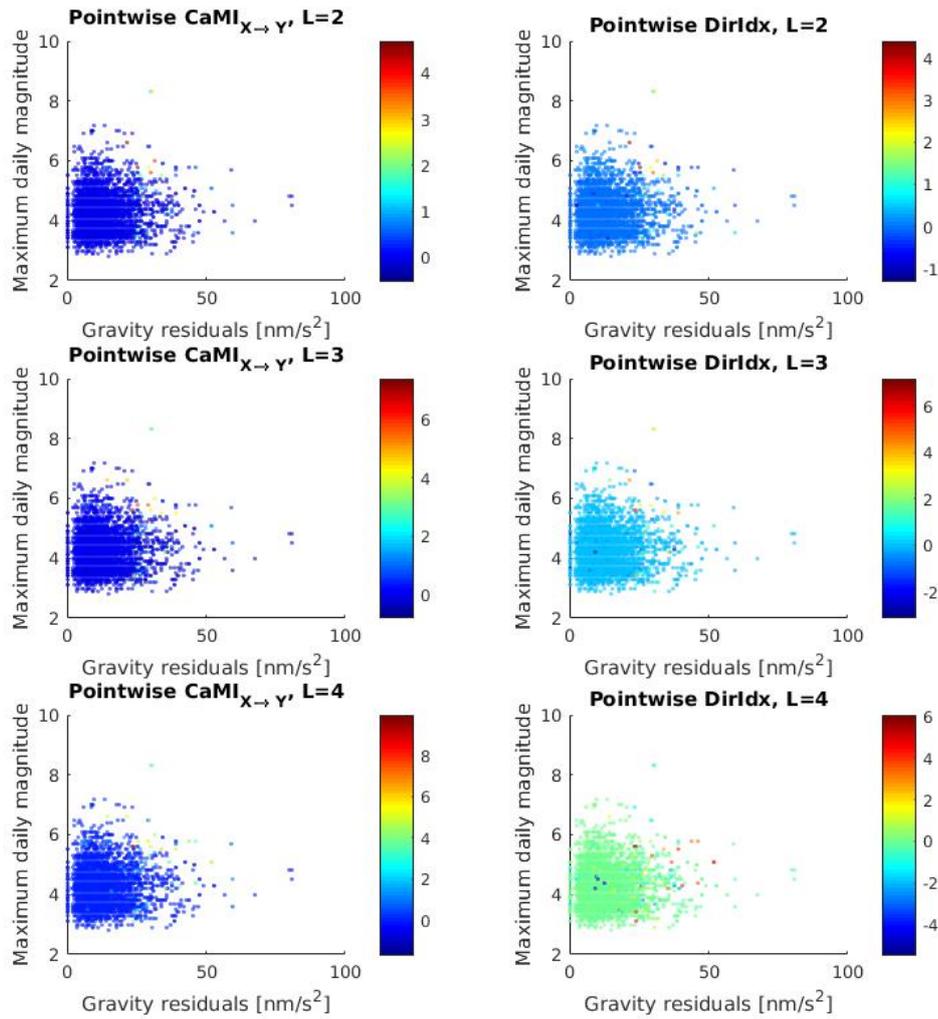


Figure C.16: Pointwise causal mutual information and pointwise directionality index for b-value anomaly (X) as precursor of high magnitude events (Y). Partition of gravity residuals anomaly is 23.84nm/s² and of magnitude is 5.6. Time-delay between the time-series is 8 days.

Anomaly of gravity residuals as precursor of high cumulative daily magnitudes

Figs. C.17 and C.18 reveal the results of the pointwise information measures for this hypothesis. The pointwise mutual information is higher in a region between 20–40nm/s² for the gravity residuals associated to high cumulative daily magnitude. Also, points of gravity residuals above 20nm/s² and high cumulative daily magnitude may have high pointwise transfer entropy. The pointwise CaMI and directionality index follows similar pattern.

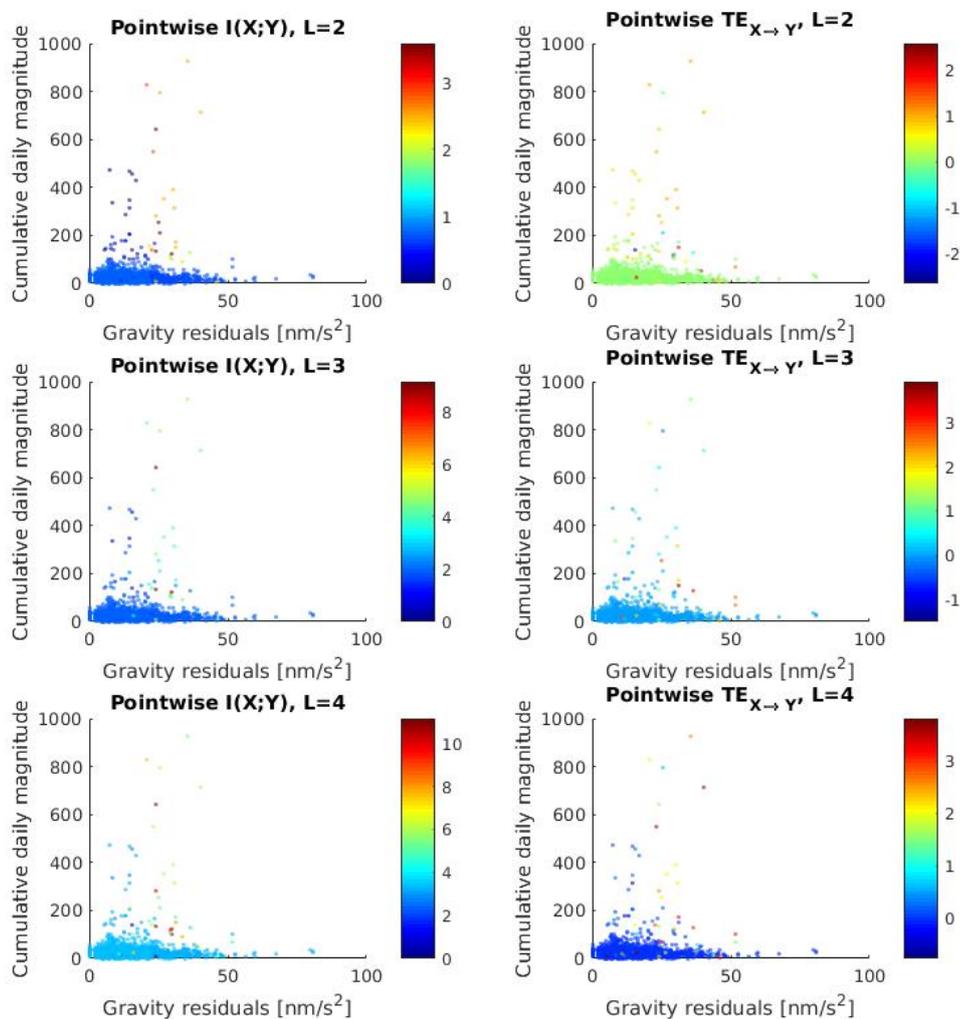


Figure C.17: Pointwise mutual information and pointwise transfer entropy for gravity residuals anomaly (X) as precursor of high cumulative daily magnitude (Y). Partition of gravity residuals anomaly is 23.84nm/s² and of cumulative daily magnitude is 102.2. Time-delay between the time-series is 12 days.

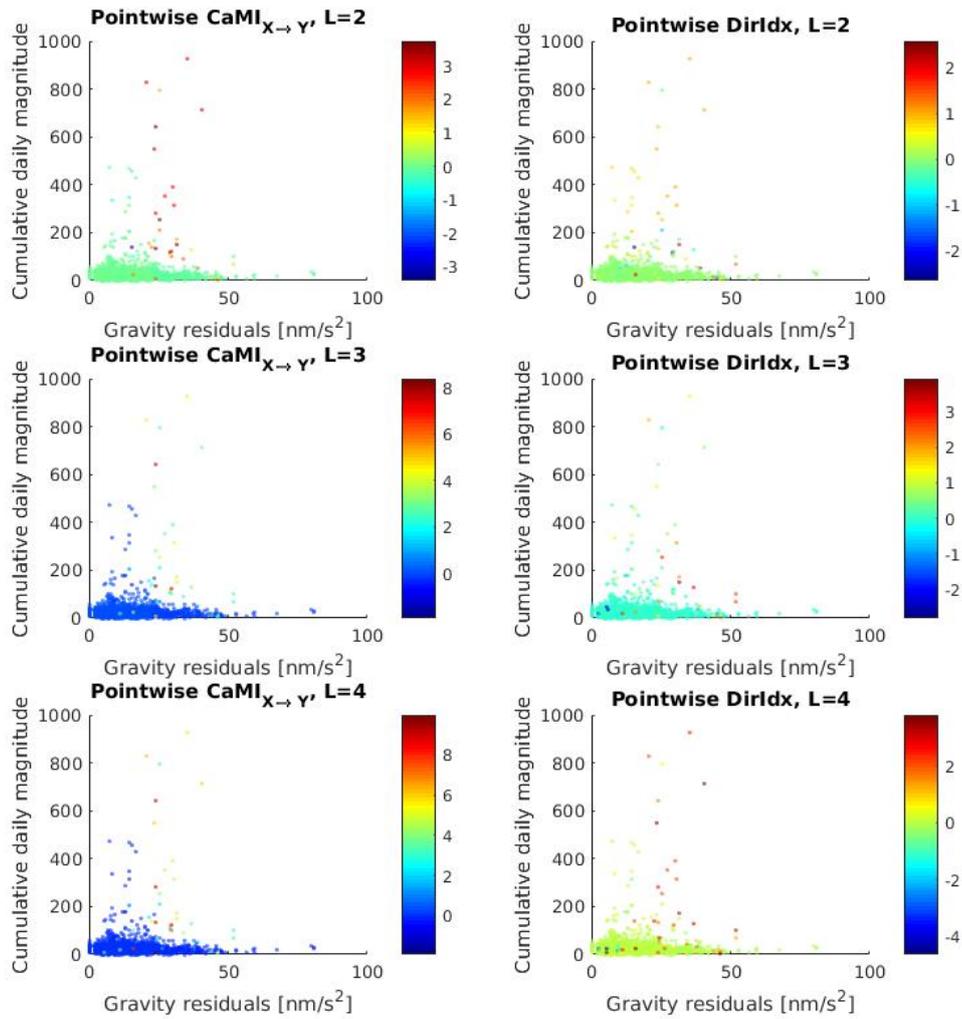


Figure C.18: Pointwise causal mutual information and pointwise directionality index for gravity residuals anomaly (X) as precursor of high cumulative daily magnitude (Y). Partition of gravity residuals anomaly is 23.84nm/s² and of cumulative magnitude is 102.2. Time-delay between the time-series is 12 days.

Anomaly of gravity residuals as precursor of high seismicity rate

The results of the pointwise measures for this case are shown in Figs. C.19 and C.20. Once again, high pointwise mutual information is observed for some points above 20nm/s^2 associated with high seismicity rate. This time this is not clearly associated with high pointwise transfer entropy. The pointwise CaMI largely follows the pointwise mutual information. The directionality index reveals no clear pattern.

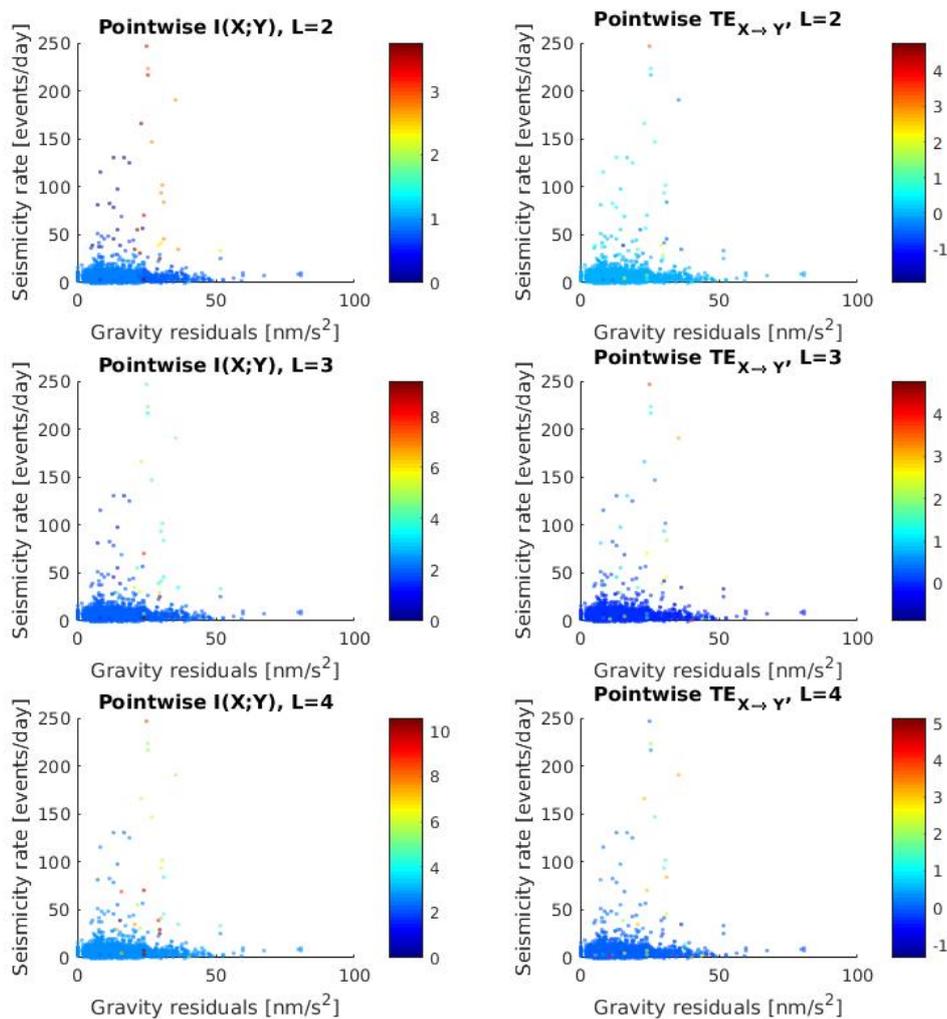


Figure C.19: Pointwise mutual information and pointwise transfer entropy for gravity residuals anomaly (X) as precursor of high seismicity rate (Y). Partition of gravity residuals anomaly is 23.84nm/s^2 and of seismicity rate is 31 events/day. Time-delay between the time-series is 11 days.

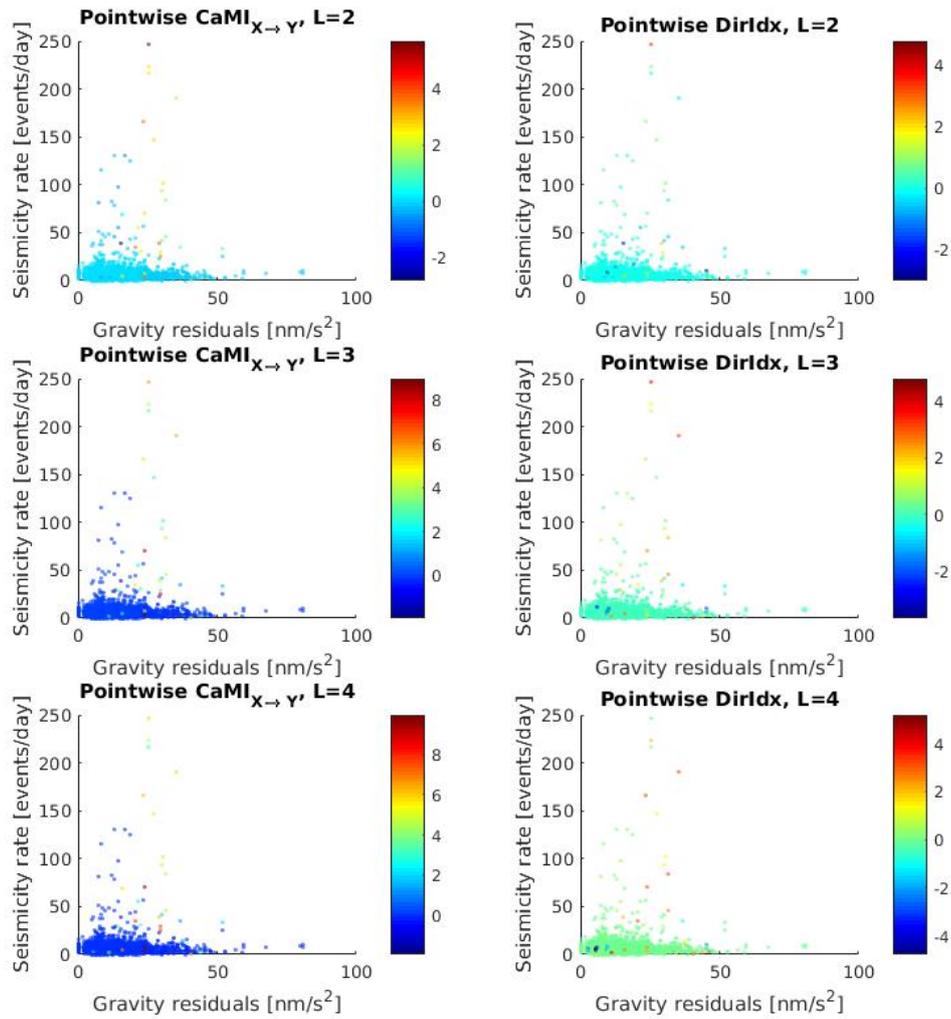


Figure C.20: Pointwise causal mutual information and pointwise directionality index for gravity residuals anomaly (X) as precursor of high seismicity rate (Y). Partition of gravity residuals anomaly is 23.84nm/s² and of seismicity rate is 31 events/day. Time-delay between the time-series is 11 days.

Preceding seismicity

High cumulative daily magnitude as precursor of high magnitude events in future

Figs. C.21 and C.22 are the results of the pointwise information measures for this hypothesis. There are only few points of high pointwise mutual information and pointwise transfer entropy, so no clear pattern can be discerned. However, in both cases, these points have shown for the anomalous high cumulative daily magnitude. The same applies for the pointwise CaMI and directionality index.

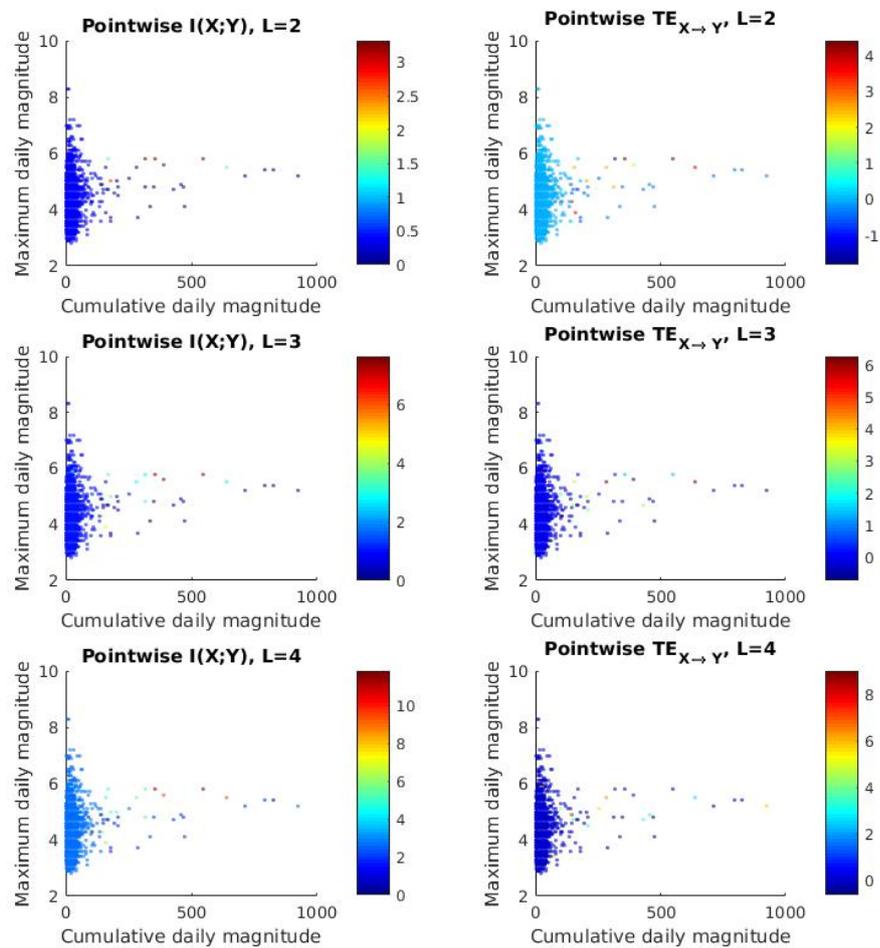


Figure C.21: Pointwise mutual information and pointwise transfer entropy for cumulative daily magnitude (X) as precursor of high magnitude events (Y). Partition of cumulative daily magnitude is $\sum M_w = 140$ and of magnitude is $M_w 5.6$. Time-delay between the time-series is 11 days.

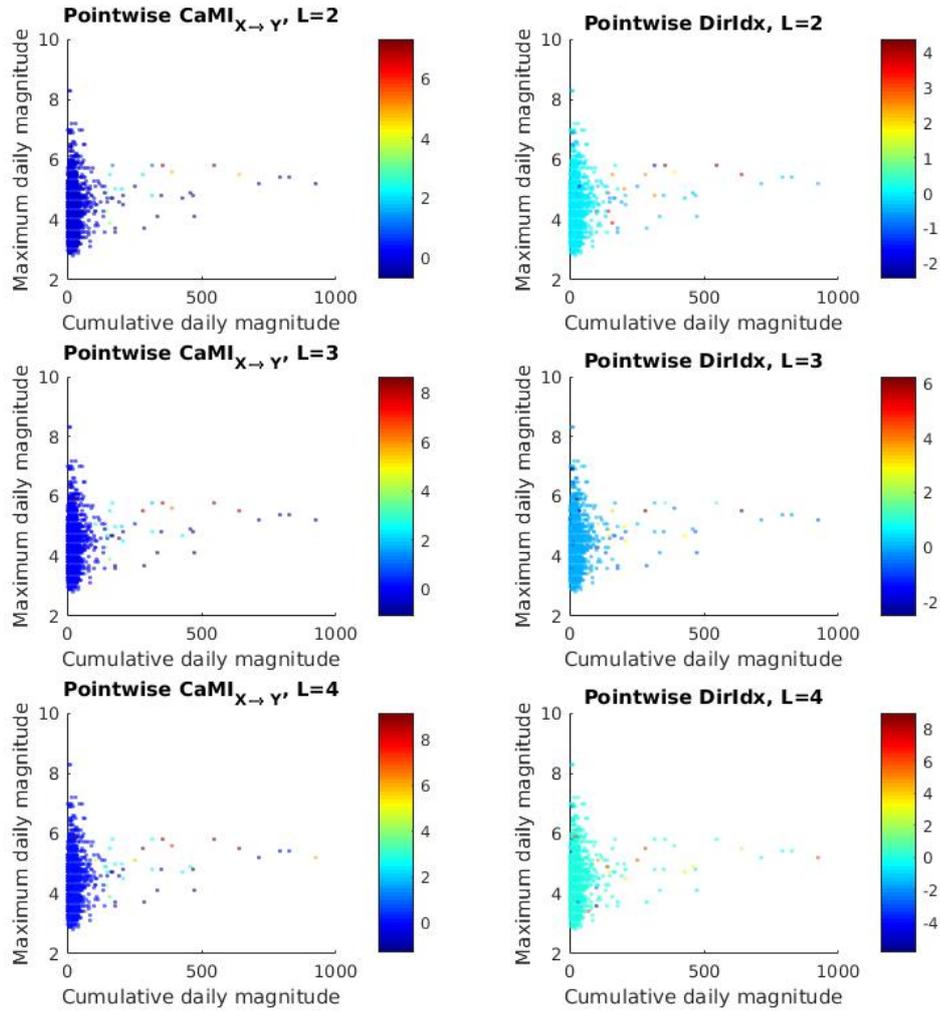


Figure C.22: Pointwise causal mutual information and pointwise directionality index for cumulative daily magnitude (X) as precursor of high magnitude events (Y). Partition of cumulative daily magnitude is $\sum M_w = 140$ and of magnitude is $M_w 5.6$. Time-delay between the time-series is 11 days.

Low seismicity rate as precursor of high magnitude events in future

The results of this hypothesis are shown in Figs. C.23 and C.24. For all pointwise informational quantities the points receive all about the same value (around 1 for pointwise mutual information and CaMI and 0 for pointwise transfer entropy and directionality index). The few outliers do not exhibit any recognisable pattern.

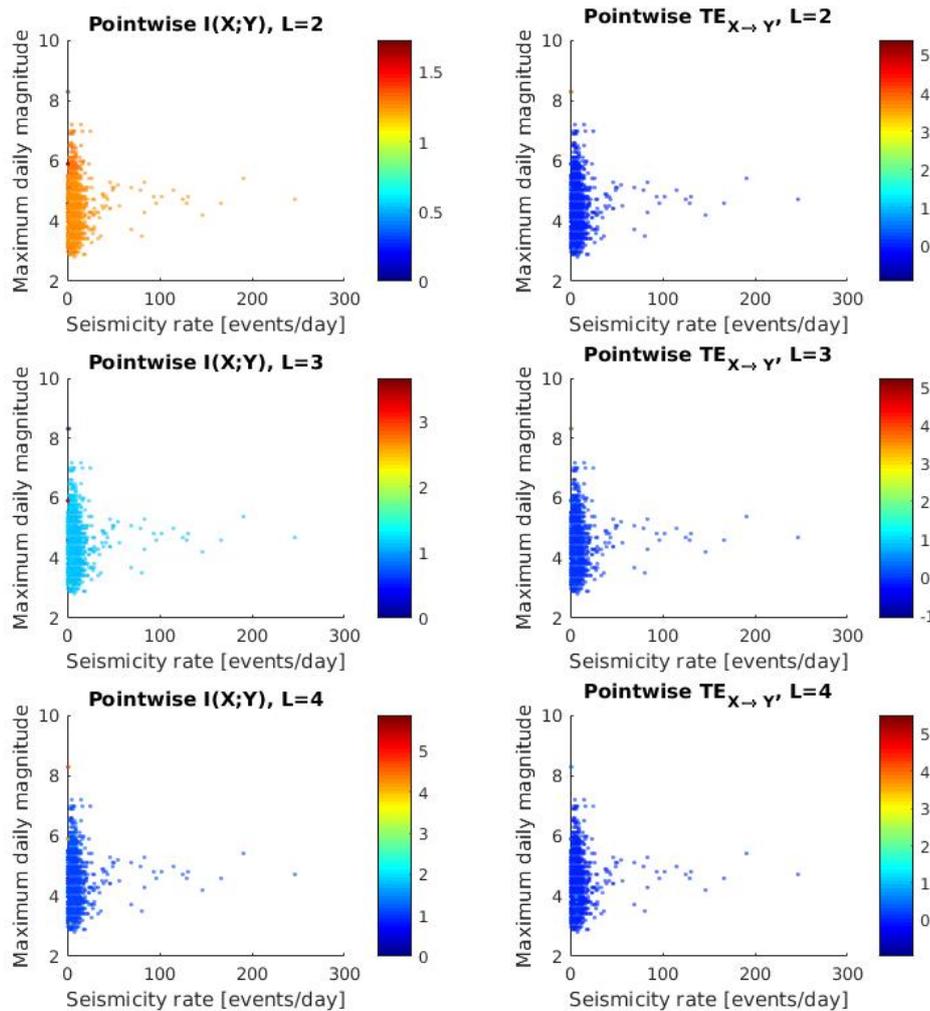


Figure C.23: Pointwise mutual information and pointwise transfer entropy for seismicity rate (X) as precursor of high magnitude events (Y). Partition of seismicity rate is 2 events/day and of magnitude is $M_w 5.6$. Time-delay between the time-series is 23 days.

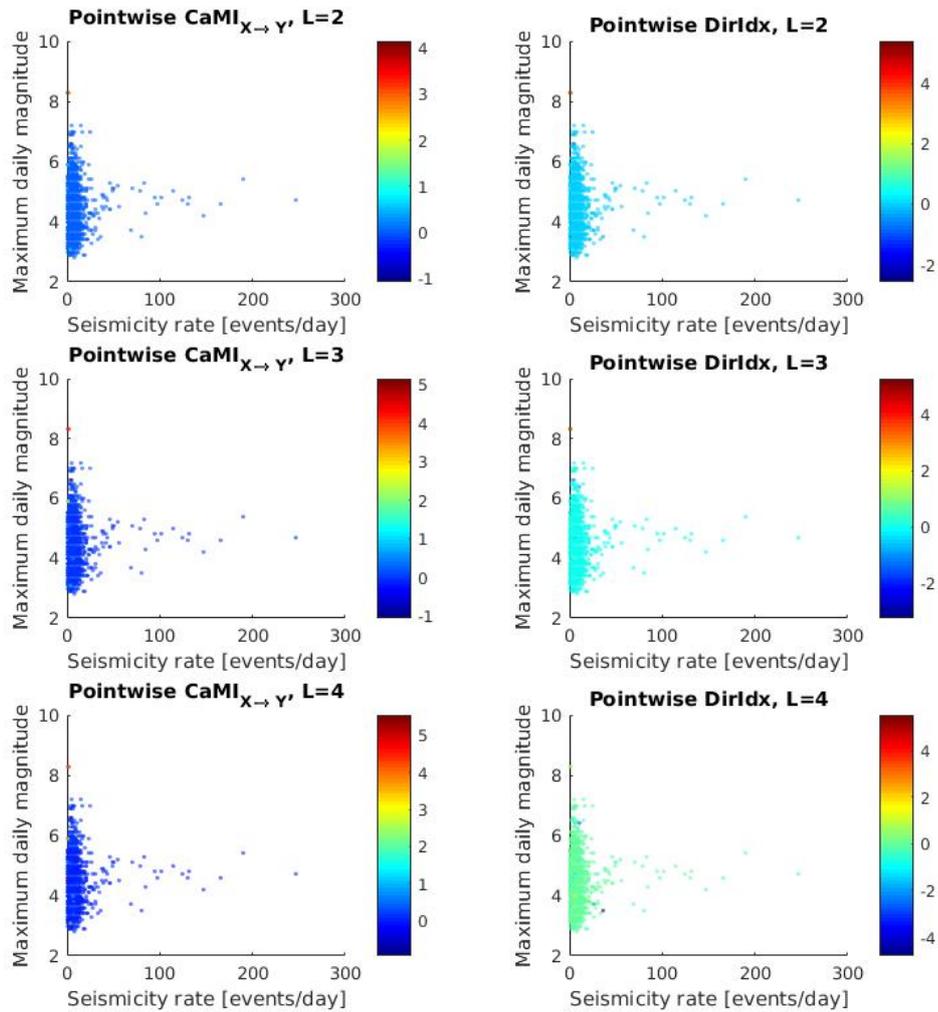


Figure C.24: Pointwise causal mutual information and pointwise directionality index for seismicity rate (X) as precursor of high magnitude events (Y). Partition of seismicity rate is 2 events/day and of magnitude is $M_w 5.6$. Time-delay between the time-series is 23 days.

Conclusions

The results presented in this appendix refer to the pointwise information measures of the earthquake precursor candidates investigated (tidal variations, b-value anomalies, pre-seismic gravity changes and anomalous preceding seismicity) with respect to the earthquake occurrences (large event, large cumulative magnitude, high seismicity rate). The delays and partitions shown here are those of the peaks of CaMI, as shown in Chapter 6. In principle, our goal was to observe the formation of the causal bubbles, *i.e.* well defined regions of the phase space with high pointwise information quantities, similarly to shown in Chapter 5, Figs. 5.9, 5.10, 5.14 and 5.15, when the measures were applied to coupled logistic maps. Some factors were already expected to make the causal bubble structures in the plots in this Appendix less evident than those in Chapter 5, such as the presence of a natural background or measurement noise factor and, especially, the limitation in the number of points (4018 in this case, against 200000 for our simulations with coupled logistic maps). Indeed, any initial indications of patterns in this appendix still require further analysis, whenever further data becomes available. Of particular interest is the case when a specific range of values of the precursor has higher pointwise information quantities to the portion of the earthquake occurrences associated to high magnitude or seismicity. It means that this range of values of the precursor has more associated information or exchange more information with the potentially destructive seismic events, therefore we should monitor them more closely. Interestingly there is an initial indication of this phenomena in our dataset for pre-seismic gravity changes on the order of $20 - 40\text{nm/s}^2$. This has not been observed for the other precursor candidates.