Alma Mater Studiorum · University of Bologna

School of Science
Department of Physics and Astronomy
Master Degree in Physics

Unraveling ocean-atmosphere coupled variability with causal methods: Transfer Entropy and Information Flow

Supervisor:
Prof. Alberto Carrassi

Submitted by:
Chiara Zelco

Co-supervisors:
Prof. Michael Ghil
Prof. Daniele Marinazzo
Prof. Stéphane Vannitsem

Academic Year 2023/2024

Alla mia famiglia e alla mia città,
alle persone che ho incontrato qui.

Abstract

Causal questions are fundamental to scientific pursuit. The study of causality and its applications has followed a nonlinear trajectory, shaped by diverse methodological developments and debates about their interpretations. Here, we unravel the evolution of these approaches, from Judea Pearl's formal framework of causal inference to methods based on the reduction of informational surprise, multivariate probability, and dynamical systems. Although principled causal inference ideally relies on Pearl's framework, its application is often unfeasible. Instead, methods grounded in information theory, combined with prior knowledge of the system, are widely used to assist in the causal inference process. Recent advances include nonlinear, higher-order information-theoretic approaches. These methods are increasingly applied in Earth and climate sciences to address questions such as the causes of extreme events and global warming, or to explore the mutual influences between the ocean and atmosphere in driving the climate system. A key unresolved question concerns the nature of this interaction. Does atmospheric weather drive the ocean, does the ocean steer the atmosphere, or does a coupled mode of variability govern the system?

In this context, we investigate the reciprocal influences of ocean and atmosphere using a low-order coupled ocean-atmosphere model that includes realistic thermal and mechanical coupling. By applying Transfer Entropy and the Liang-Kleeman Information Flow, we analyze the dynamical directions within the coupled system. We uncover the directed dynamics of information exchange, adding insight on the emergence of low-frequency variability in the atmosphere. These results offer a new perspective on interannual and decadal-scale climate prediction.

Contents

Acronyms

AMO Atlantic Multidecadal Oscillation.

AO Atlantic Oscillation.

CMI Conditional Mutual Information.

ENSO El Niño-Southern Oscillation.

FCI Fast Causal Inference algorithm.

KSG technique Kraskov-Stögbauer-Grassberger technique for TE estimation.

LKIF Liang-Kleeman Information Flow.

MI Mutual Information.

NAO North Atlantic Oscillation.

ODE Ordinary differential equation.

PC algorithm Peter and Clark algorithm.

PDE Partial differential equation.

PDO Pacific Decadal Oscillation.

PNA Pacific North American Index.

QG Quasi geostrophic.

SOI Southern Oscillation Index.

TE Transfer Entropy.

TNA Tropical Northern Atlantic Index.

Introduction

Causality offers a compelling way to unravel interdependent relationships in complex systems, especially when it is important to understand the direction of influence among components or to attribute causes to observed events. This approach is particularly valuable in climate science, where the inherent complexity and multiscale nature of the climate system pose significant challenges.

The concept of causality has deep philosophical roots and evolved over time amid debates about its meaning and practical application. Early attempts to use causal methods often met with limited success. In the 20th century, however, two main approaches emerged. Clive Granger introduced a data-driven method based on the idea that a variable X causes Y if incorporating past values of X improves the prediction of Y. In parallel, Judea Pearl's influential work in artificial intelligence and probabilistic reasoning established a formal framework for causality. Pearl's approach employs interventions and causal graphs to resolve many of the ambiguities that had long surrounded the concept, providing a common mathematical foundation across disciplines. Although Pearl's interventional framework is not directly applicable to every inference problem — such as the causal analysis of non-reproducible time series — it has paved the way for further developments.

The gap between theory and application spurred the evolution of Granger's ideas into multivariate and nonlinear methods and inspired other data-driven approaches based on different principles. Notably, information-theoretic methods have achieved considerable success. Transfer Entropy (TE) identifies a causal link as the reduction in uncertainty about one variable's future provided by another, while the Liang-Kleeman Information Flow (LKIF) quantifies causality in terms of the increase in a target variable's entropy induced by another. Other approaches, such as Convergent Cross Mapping (CCM) and Empirical Model Reduction (EMR), also show promise by drawing on dynamical systems theory to capture the physical nature of the problems more closely.

Although these methods often explain causality in different ways, as they stem from different theoretical principles, some common ground has been found—for instance, Granger causality and Transfer Entropy are equivalent for Gaussian variables, and similarities have been noted between approaches aligned with Pearl's framework and those based on information theory. Still, in practical applications such as reconstructing causal

graphs or quantifying the strength of causal links, different methods frequently yield varying results.

Causal methods are increasingly applied to climate studies, a field defined by nonlinear, complex, and multiscale interactions. Climate research encompasses a wide range of approaches—from simple models that capture essential dynamics, as pioneered by Edward Lorenz, to high-resolution Earth System Models (ESMs), at the top of the *hierarchy of models* spectrum. In this work, we add a new perspective by applying data-driven causal methods to explore the nonlinear interactions within a prototypical climate model.

Causal inference is used in climate science for tasks ranging from attributing climate change and extreme events to disentangling complex interactions among climate components. One enduring challenge is to determine the dominant direction of influence between the ocean and the atmosphere. This issue is often described as "*the dog and the tail*" problem, where the scenarios include the ocean driving the atmosphere - the dog wagging its tail, the atmosphere steering the ocean - the fast movement of the tail driving the dog, or a coupled mode in which both mutually interact, the dog biting its own tail.

Motivated by this challenge, we investigate ocean–atmosphere interactions using MAOOAM, a low-order coupled spectral model. We analyze causal relationships via multivariate and interregional approaches based on Transfer Entropy and the Liang-Kleeman Information Flow. We will see that the multivariate TE and LKIF approaches prove effective at revealing the transition of the system from a strong to a weak coupling configuration. However, the high dimensionality of MAOOAM and the challenges inherent in interpreting causal links within its spectral formulation limit this approach.

The interregional causal analysis will highlight a strong ocean-to-atmosphere dependency in the strong coupling scenario, and a predominant atmosphere-to-ocean direction in the weak coupling situation, qualitatively aligning with the physical expectations. TE and LKIF produce noticeably distinct results, likely reflecting the different principles underlying each method.

The structure of this thesis is as follows. In Chapter 1, we review the historical development of causality and the formal framework established by Judea Pearl. Chapter 2 details methods for causal inference from time series and discusses the current state of research on their causal foundations. In Chapter 3, we present the climate system as a complex, multiscale phenomenon, with a focus on ocean–atmosphere coupled variability, and describe the MAOOAM spectral model. Chapter 4 presents our application of Transfer Entropy and Liang-Kleeman Information Flow to analyze causal relations between oceanic and atmospheric Fourier modes in MAOOAM. Finally, Chapter 4.4 summarizes our findings and outlines directions for future research.

Chapter 1

An overview on causality

Though it is basic to human thought, causality is a puzzling concept, shrouded in controversy and caution, because scientists and philosophers have had difficulties defining when one event *truly causes* another. It is well understood that "the rooster's crow does not cause the sun to rise" (Judea Pearl, 2009), but even this obvious statement has long been difficult to express into a mathematical equation. In this chapter we examine the reasons behind this historical limitation and we review the historical development of the discipline of causality - from the early debates on the predominance of statistics to the rise of different causal approaches in the 20th century. We discuss the rise of methods rooted in time series observation. In parallel, we present the evolution and the fundamental ideas of the formal framework for causal inference, which has established a hierarchical structure for framing causal questions, either providing answers when feasible or delineating the mathematical limitations to doing so. We discuss its merits in addressing the fundamental research question of performing robust causal inference from time series. We examine the required assumptions, the context-specific limitations, and the methods developed in this domain.

1.1 Causality: historical setting of a long-lasting problem

The need to ask *why* questions and the capacity to find causal explanations came very early in human development. In the beginning, causality was not problematic. Causes were generally attributed to angry deities, a naive but clear and unproblematic notion of causality. Something changed with the rise of engineering: it introduced the possibility to intervene on a broken piece to make it work again, instead of blaming the problem on the divine. At that point in history, deities and humans ceased to be the sole agents of causal forces - lifeless objects and processes began acquiring causal character. Causality held a dual role: causes were the targets of credit and blame on one side, and the carriers

of physical flow of control on the other.

An important rupture was brought about during the Renaissance, with Galileo's revolutionary thought. It consisted of two maxims, expounded in his 1638 book *Discorsi* (Galilei, 1638): (i) *Description first, explanation second, "how" precedes "why", and* (ii) *Description is carried out in the language of mathematics: equations.* Instead of asking if an object falls because it is pulled from below or pushed from above, Galileo suggested asking *how well you can predict* the time it takes for the object to travel a certain distance. Following this hook, physics saw the proliferation of empirical laws, extremely useful, but mainly ignoring the more fundamental principles that explained why they worked so well, e.g., Hooke's law, Ohm's law and Joule's law.

While physics was momentarily satisfied with practical success, philosophy kept searching for the origin and justification of causality. Hume, about one hundred years after Galileo's contributions, took his ideas to an extreme. He argued convincingly that the *why* is not merely second to the *how*, but that the why is totally superfluous as it is subsumed by the how. Causal connections according to Hume are the product of observations: causation is a learnable habit of the mind. However, this argument could trick us into thinking that the Sun comes up because of the rooster's crow. To account for such an obviously untrue explanation, a new maxima was developed and is still very popular nowadays: "*correlation does not imply causation*". Not every time we observe correlation between two phenomena those phenomena are found to be causally linked, or not in the direction that we had guessed based on observation.

So how do we acquire knowledge of causation, without stumbling on spurious correlations and inferring the wrong direction of causal links? It is clear that the regularity of succession is not sufficient to identify correctly a causal relationship. What would be sufficient instead? What patterns of experience convince people that a connection is *causal*? One line of thought is that if causal information has an empirical meaning beyond temporal succession, then we would expect it to show up in the laws of physics.

1.1.1 The mechanistic critique

It was first noted by Ernst Mach, and then strongly argued by Bertrand Russell, that, in fact, this is not true. In 1913, Bertrand Russell wrote "*All philosophers imagine that causation is one of the fundamental axioms of science, yet oddly enough, in advanced sciences, the word "cause" never occurs... The law of causality, I believe, is a relic of bygone age, surviving, like the monarchy, only because it is erroneously supposed to do no harm.*" The main argument to support this is that the laws of physics are invariant for time reversal. If a process can happen in one direction, it can happen in the opposite direction too - at least in classical mechanics. These laws are not formulated in terms of cause and effect, they describe the regularity that we observe among phenomena. This argument makes sense in Russell's classical mechanistic view, assuming that determinism holds.

However, Russell's argument does not hold when we face other disciplines of physics, for instance thermodynamics and statistical mechanics, where determinism does not hold anymore as it did in classical mechanics. The breakthrough was ignited by the introduction of the concept of *arrow of time*: the fact that even if the microscopical laws that govern the dynamics of single infinitesimal parts of a system are time reversible, the average behavior of the system inevitably shows a temporal asymmetry, a precise direction of time (Prigogine and Rysselberghe, 1963). This is true every time we adopt a macroscopic view of the system at study. Thus, we find out that Russell's argument is sound as long as we restrict our study to classical mechanics, but it is incomplete when we study macroscopic systems: from brain connections to teleconnection phenomena in the climate system (Contu, 2024).

This argument establishes that causality exists and that it plays an identifiable role in macroscopic systems, but it does not solve all the problems that concern causality and the applied search of causal connections. Another important debate took place in another field of science: statistics.

1.1.2 The debate in Statistics: correlation versus causation

Statistics introduced the concept of correlation largely thanks to the pioneering work of Francis Galton, who, around the end of the 19th century, investigated the heritability of physical traits in families. Galton observed that when a father was significantly taller than average, his son would also tend to be taller than average—but not as extremely so as the father. This tendency, which applies to any anthropometric characteristic, became known as *regression toward the mean*. Today we understand that his is not a natural law but rather a statistical phenomenon observed in correlated datasets. We explain why following Contu, 2024.

We know from statistics that to test for a linear relationship, $y = Ax + B$, between two random variables x and y we have to perform a regression test on sample data. This leads us to the best estimates for the coefficients A and B:

$$A = \frac{\sum x^2 \sum y \; - \; \sum x \sum xy}{\Delta} \qquad B = \frac{N \sum xy \; - \; \sum x \sum y}{\Delta}, \tag{1.1}$$

where $\Delta = N \sum x^2 - (\sum x)^2$. At the same time we can compute the linear correlation coefficient to test the goodness of the linear hypothesis we made: $r_{xy} = \frac{cov(x,y)}{\sigma_x \sigma_y} = r_{yx}$. The correlation coefficient can be used to express the estimate of A:

$$A = \frac{\mathrm{cov}(x,y)}{\sigma_x^2} = r_{xy} \frac{\sigma_x}{\sigma_y}. \tag{1.2}$$

If we standardize the random variables, that is, transform them to have unit standard deviation and zero mean, the correlation coefficient remains unvaried: $r_{XY} = r_{xy}$. This

makes sense, as the standardization of random variables does not change the physical meaning of their correlation or independence. Assuming the two standardized variables are related through a linear model, $Y = \beta X$, then we also have $\beta = r_{XY} = r_{xy}$. Therefore it's immediate to see that the regression of X on Y is the same as the regression of Y on X

$$\mathbb{E}(Y|X) = \mathbb{E}(X|Y). \tag{1.3}$$

This means that if $Y = \beta X$, it is also true that $X = \beta Y$. Since the correlation coefficient lives in the $[-1, 1]$ interval, this means that $\mathbb{E}(Y|X) = r_{xy}X \leq X$.

Now let's go back to Galton's investigations. With these tools, he would correctly note that the height of the sons regresses towards the mean height of the fathers, but he would find the same thing in the opposite direction, too: $\mathbb{E}[X|Y] = r_{xy}Y \leq Y$, the height of the fathers regressing to the mean height of the sons. This confirms the purely noncausal nature of this phenomenon, as a son's height cannot causally determine his father's. An example is reported in Figure 1.1

Realizing the essentially statistical nature of the phenomenon, Galton abandoned the search for a causal explanation and instead emphasized that variables are simply "correlated." Galton concluded "*It is easy to see that co-relation must be the consequence of the variations of the two organs being partly due to common causes.*" The first objective measure of how two variables are *related* to each other, based only on data, clear of human judgment. This research inspired one of Galton's student, Karl Pearson, who would later become one of the central figures in statistics. He interpreted his Professor's discoveries to mean that "*there was a category broader than causation, namely correlation, of which causation was only the limit*"[1] (Pearson, 1892) and that would allow to introduce mathematical approaches to soft disciplines. Pearson firmly believed causes and effects to be archaic and smoky concepts, to be substituted by the rigorous concept of correlation. He denied the need for an independent concept of causal relation beyond correlation. His tall stand in the field consented him to substantially forbid causation from statistics before it had a chance to take root[2].

Statistics followed along the path carved by Pearson and became a powerful and popular discipline. However, not everything found an easy answer with statistical explanation. A classical example: very high correlation is observed between data of ice-cream consumption and shark attacks. An easy explanation to this strange correlation might be that they share a common cause, in this case the month of the year. During warm summer months ice-cream consumption is higher than during winter, and since more people go to the beach, it is also more likely that there will be shark attacks. In this

[1] In 1911 Pearson published the third edition of his book "*The Grammar of Science*", containing a new chapter titled *Contingency and Correlation - The Insufficiency of Causation*".

[2] Only one causal concept would be allowed in the field of statistics, developed years later by Sir Ronald Fisher: *randomized experiments*. This technique still plays a fundamental role in testing causal relationships (Fisher, 1992).

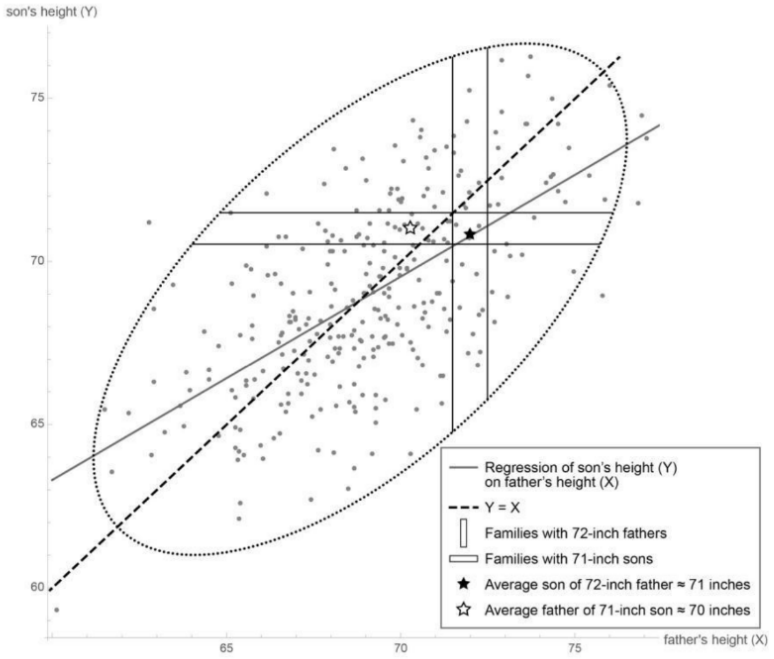


Figure 1.1: The scatter plot shows a data set of heights, with each dot representing the height of a father (on the x-axis) and his son (on the y-axis). The dashed line coincides with the major axis of the ellipse, while the solid line (called the regression line) connects the rightmost and leftmost points on the ellipse. The difference between them accounts for regression to the mean. For example, the black star shows that 72" fathers have, on the average, 71" sons. (That is, the average height of all the data points in the vertical strip is 71".) The horizontal strip and white star show that the same loss of height occurs in the noncausal direction (backward in time). Source: J. Pearl and Mackenzie, 2018.

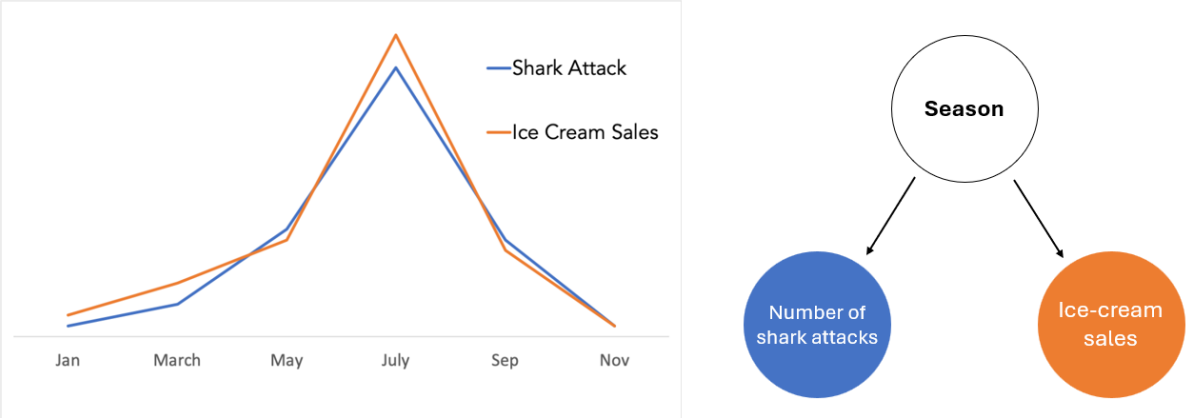


Figure 1.2: The correlation between the number of shark attacks and ice-cream sales per month of the year (left). A causal explanation of the strong correlation between the two variables considering the variable of the season, which acts here as a confounding variable (right). Left image credits: Candela Iglesias

example the season is a *confounding variable*: it causes both ice-cream sale and shark attack numbers, and if not considered it leads to a wrong guess of the causal relationship between the two. Statistics had a hard time defining what a confounding variable is, and could not come up with a satisfactory statistics-based definition (J. Pearl and Mackenzie, 2018). To correctly define what a confounding is, one needs to have a causal scheme in mind, correlation only is not sufficient. The scientific community at the beginning of the '900 did not have an appropriate language to express this, and would not have it for a long time, with an exception.

1.1.3 Wright's coefficients method

In the 1920s Sewall Wright, an American geneticist, introduced the method of *path analysis* to study the transmission of traits in populations of guinea pigs. The starting point of this method is to put on mathematical grounds the relationship between the variables at study, relating them through an equation. However, his revolutionary idea was to *associate to the mathematical equations a graphical description*, able to express a directionality between the variables, a feature that mathematical equations alone do not have. An example is found in Figure 1.3. The graphical representation later became the central point of Pearl's research on causality, but it would remained neglected for many years before coming back in vogue. Wright also made an important point at which we will often hint at in the following chapters: *causal conclusions cannot be drawn without some causal hypotheses on the system.* Having some prior hypothesis about the causal structure of the system makes the problem of inferring the single causal relationships among variables easier and more sound.

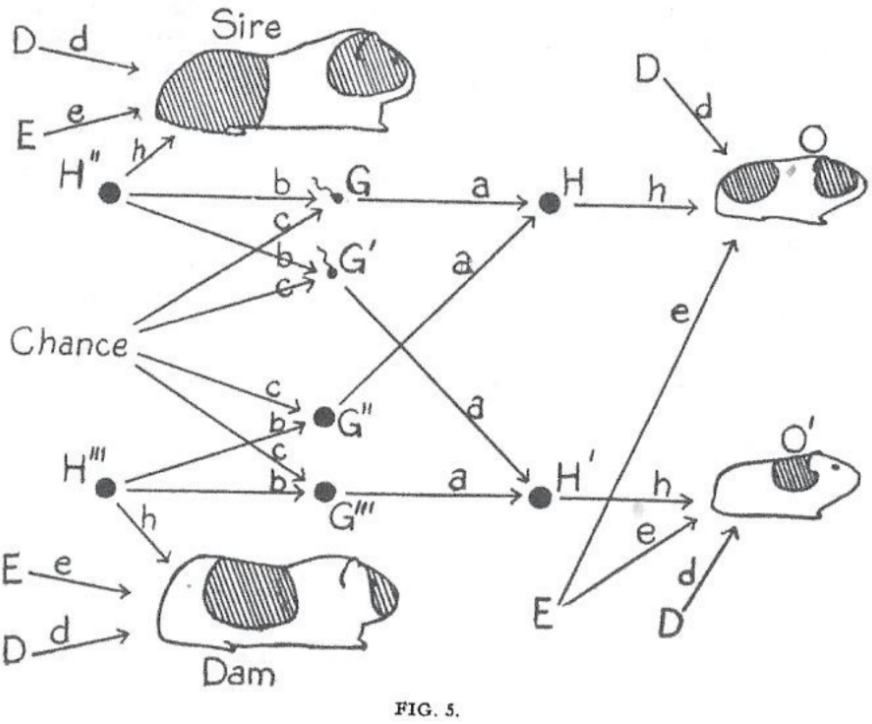


FIG. 5.

Figure 1.3: Sewall Wright's first path diagram, illustrating the factors leading to coat color in guinea pigs. D = developmental factors (after conception, before birth), E = environmental factors (after birth), G = genetic factors from each individual parent, H = combined hereditary factors from both parents, O, O' = offspring. The objective of analysis was to estimate the strength of the effects of D, E, H (written as d, e, h in the diagram). Source: Wright, 1920

Wright's experiments with causal graphs went mostly unnoticed when not explicitly opposed by Pearson and other statisticians (J. Pearl and Mackenzie, 2018).

Statisticians would remain - and still are - very cautious regarding causality. The reason for this can be partly attributed to the fact that causation is much harder to measure than correlation. Correlations can be estimated from observational studies, while the simple way to estimate causality requires controlled experiments, which in many situations are not viable. Another important reason for the under-development of causal inference has to be attributed to the substantially separate proliferation of approaches in different fields. The lack of a shared, complete language to express certain concepts makes the task of developing scientific activity exceptionally difficult.

Despite the strong views of the Statistics world on causal inference and the other difficulties, research on the topic spurred in different domains (J. Pearl and Mackenzie, 2018). However, it developed with little organization and it did not build a corpus of shared definitions to guide researchers from different fields. With these premises, the

discipline of causality was not able to progress much.

The situation changed in the second half of the 20^{th} century, on one side with the growing interest in artificial intelligence and Bayesian networks, and on the other side with Clive Granger's efforts in the development of a practical method to estimate causal relationships from observational data. The latter line of research would first become extremely popular in econometrics, and later on in many other disciplines, from neuroscience to climate sciences. Granger's approach to causal inference - *Granger causality* - states that a time series X is said to Granger-cause Y if it can be shown that X provides statistically significant information about future values of Y. It has long been argued that this predictive notion does not correspond to *real causality*, and we will discuss in Chapter 2 the possible merits of this critique. The first approach, stemming from interest in artificial intelligence, took a completely different path. It set itself to rethink from scratch the categories of statistics and of causality, and propose a formalization of causality that would be accessible to researchers notwithstanding their specific background. This framework, initially developed by computer scientist Judea Pearl, led to a revolution in causal inference and in science in general; it would later extend its appeal to a broad variety of scientific fields.

1.1.4 Defining causality through interventions on the system

The rising interest in artificial intelligence that was happening in the last decades of the 20^{th} century made it possible to lay the philosophical groundwork that would allow the causal breakthrough to happen. The effort to build a new type of "intelligent" machine from scratch, programming every aspect of it, forces us to wonder about the basic blocks of intelligence: what makes us humans intelligent, in the first place, and how can we transfer this kind of ability onto a machine? Causal thinking plays an important role in human intelligence, it was argued by Judea Pearl that the ability to pose causal questions - and possibly to answer them - is what makes human thought so special and powerful.

The overriding ideas of the new framework proposed by Judea Pearl were

(i) to treat causation as a summary of behavior under interventions, and

(ii) to use equations *and* graphs as mathematical language within which causal thoughts can be represented and manipulated.

These two concepts are put together by a third one: interventions act by cutting specific parts of the equations, as a "surgery".

An example

We can see the framework's prescriptions in practice looking at the simple circuit in Figure 1.4, featuring two independent gates: a multiplier and an adder.

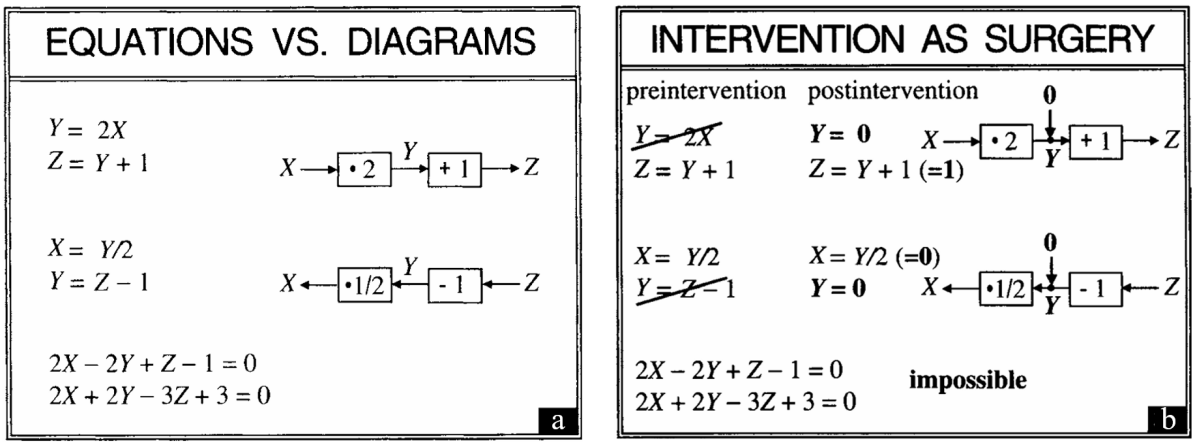


Figure 1.4: A simple visualization of the necessity of a graphical representation to encode causal information. Source: Pearl, 1996 Judea Pearl, 1996.

We could try to represent what the circuit encodes using the equations on the left, in the picture. Are the two representations equivalent? They are not, otherwise the two circuits would be exactly the same, but they clearly do not obtain the same results. The top circuit tells us that if we manipulate Y, Z will be affected, and X will not. The bottom one encodes that manipulating Y will affect X, but not Z. The intervention process is depicted in Figure 1.4, on the right. This directional sense is not present in the equations on the left, which can be rearranged to implement also actions that are forbidden on the physical circuits. Moreover, if we performed some additional algebraic operations on our equations, we could obtain two new equations, shown at the bottom, which point to no structure at all, and don't tell us how the variables affect each other.

The interventional definition of causality

The explained model of intervention leads us to a formal definition of causation: "*Y is a cause of Z if we can change Z by manipulating Y*". Having a clear idea of the causal diagram is vital in this process: it tells us which equation is to be deleted when Y is manipulated. This information is lost when the equations are transformed into algebraically equivalent form, at the bottom of Figure 1.4.

To sum up, the causal model prescribed by the line of research in computer science, at the end of the past century, contains symmetric equations, a distinction between *in* and *out*, an assumption that each equation corresponds to an independent mechanism and thus has to be preserved as a separate mathematical sentence, and that interventions are interpreted as surgeries over those mechanisms. This structure calls for a specific language, able to implement in mathematical language the act of performing an intervention - a surgery - and to keep together all the relevant information about the system. An essential element of the language that would have to be developed is the *do-calculus* tool,

which allows to describe mathematically interventions on the system. An example of the action of this tool in climate modeling is the prescription of sea-surface temperatures to precise values in climate models to study their effect on atmospheric dynamics.

The language of causality has been developed and formalized in the concept of Structural Causal Models (SCMs). Hence, SCMs are tools designed by this line of research as necessary element for a rigorous study of causality. An overview of SCMs has been given in Judea Pearl (2009), and other references are summarized in Peters, Janzing, and Schlkopf (2017). In the following section we present the ground theory of Pearl's causality framework and some fundamental notions about SCMs, to understand the wide applicability of these tools and the power of the causal inference discipline this way formalized.

We note that even when in possession of Pearl's framework, testing for cause and effect remains a difficult task, and discovering causes of effects is even more daunting. The aim of the formalization of causal inference performed in the last decades is to prove that causality is not a metaphysical concept, instead, it can be understood in terms of simple processes, and expressed in mathematical language.

1.2 Causal Modeling and Learning

The contemporary view on causal modeling is that a causal structure entails a probability model, but it contains additional information not contained in the latter, as reported in the scheme in Figure 1.5. There are fundamentally two types of research in causality: *causal reasoning*, the process of drawing conclusions from a causal model, and *causal learning*, also referred to as structure learning or causal discovery: the problem of inferring causal structures from their empirical implications (observational data and/or interventions) Peters, Janzing, and Schlkopf, 2017. Together, these two categories form the discipline of *causal inference*, a larger term that indicates the general effort of describing causal relationships between variables.

Large interest is dedicated to understanding which parts of the causal structure can be in principle inferred from the joint distribution, a problem called *structure identifiability*. The solution of this problem is in no way trivial, and it requires the dataset to satisfy strong assumptions. In the following we expose the basics of SCMs and introduce the assumptions that datasets need to respect so that we can infer causal structures from data. We note that other popular approaches in causal inference, e.g. Granger causality, are not tightly bound to the formal structure of SCMs. Nevertheless, they have evolved to integrate important concepts rooted in such domain. We will further discuss it at the end of Chapter 2.

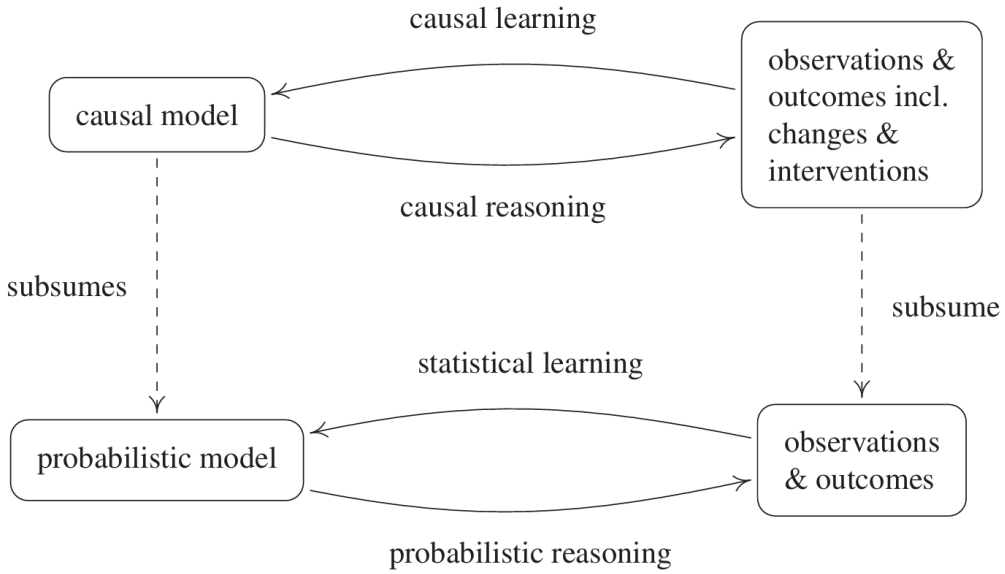


Figure 1.5: A scheme of probabilistic and causal inference. Courtesy of Peters, Janzing, and Schlkopf, 2017.

1.2.1 Causal Graphs

The language of causality relies on the representation of causal knowledge through graphs. A graph is a collection of vertices and edges, where edges can be directed or undirected. In the latter case they convey no designated direction of causal information.

Definition 1.2.1 (**Causal Graph**). *Consider finitely many random variables* $\mathbf{X} = (X_1, \ldots, X_d)$ *with index set* $\mathbf{V} := \{1, \ldots, d\}$*, joint distribution* $P_{\mathbf{X}}$*, and density* $p(\mathbf{x})$*. A graph* $\mathcal{G} = (\mathbf{V}, \mathcal{E})$ *consists of (finitely many) nodes or vertices* $\mathbf{V}$ *and edges* $\mathcal{E} \subseteq \mathbf{V}^2$ *with* $(v, v) \notin \mathcal{E}$ *for any* $v \in \mathbf{V}$*. We further have the following definitions:*

- *A node* i *is called a parent of* j *if* $(i, j) \in \mathcal{E}$ *and* $(j, i) \notin \mathcal{E}$ *and a child if* $(j, i) \in \mathcal{E}$ *and* $(i, j) \notin \mathcal{E}$*. The set of parents of* j *is denoted by* $\mathbf{PA}_j^{\mathcal{G}}$*, and the set of its children by* $\mathbf{CH}_j^{\mathcal{G}}$*. A node without parents is called a source node, a node without children a sink node. Two nodes* i *and* j *are adjacent if either* $(i, j) \in \mathcal{E}$ *or* $(j, i) \in \mathcal{E}$*. We call* $\mathcal{G}$ *fully connected if all pairs of nodes are adjacent. We say that there is an undirected edge between two adjacent nodes* i *and* j *if* $(i, j) \in \mathcal{E}$ *and* $(j, i) \in \mathcal{E}$*. An edge between two adjacent nodes is directed if it is not undirected. We then write* $i \to j$ *for* $(i, j) \in \mathcal{E}$*. We call* $\mathcal{G}$ *directed if all its edges are directed.*

- *A path in* $\mathcal{G}$ *is a sequence of (at least two) distinct vertices* $i_1, \ldots, i_m$*, such that there is an edge between* i_k *and* i_{k+1} *for all* $k = 1, \ldots, m-1$*. If* $i_{k-1} \to i_k$ *and* $i_{k+1} \to i_k$*,* i_k *is called a collider relative to this path. If* $i_k \to i_{k+1}$ *for all* k*, we speak*

of a directed path from i_1 to i_m and call i_1 an ancestor of i_m and i_m a descendant of i_1.

- *A graph $\mathcal{G}$ is called a partially directed acyclic graph (PDAG) if there is no directed cycle, that is, if there is no pair (j, k) with directed paths from j to k and from k to j. $\mathcal{G}$ is called a directed acyclic graph (DAG) if it is a PDAG and all edges are directed.*

Any graph structure can be broken down to three essential three-node structures, shown below: *chains* (a), *forks* (b) and *colliders* (c).

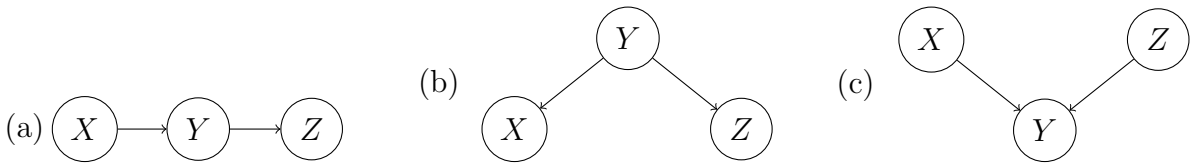


Figure 1.6: The essential three-node graph structures: *chains* (a), *forks* (b) and *colliders* (c).

1.2.2 Structural causal models

SCMs have been used for a long time in social sciences, econometrics and agriculture, before being formalized by the definition we hereby present, following the formulation of Peters, Janzing, and Schlkopf, 2017. SCMs are the key tools used to formalize causal reasoning and causal learning. They describe the relevant features of the system and their reciprocal interactions.

Definition 1.2.2 (**Structural Causal Models**). *A structural causal model (SCM) $\mathcal{C} := (\mathbf{S}, P_{\mathbf{N}})$ consists of a collection $\mathbf{S}$ of d (structural) assignments*

$$X_j := f_j(\mathbf{PA_j}, N_j) \quad j = 1, ..., d \tag{1.4}$$

where $\mathbf{PA}_j \subseteq \{X_1, \ldots, X_d\} \setminus \{X_j\}$ are called parents of X_j; and a joint distribution $P_{\mathbf{N}} = P_{N_1,...,N_d}$ over the noise variables, which we require to be jointly independent; that is, $P_{\mathbf{N}}$ is a product distribution.
The graph $\mathcal{G}$ of an SCM is obtained by creating one vertex for each X_j and drawing directed edges from each parent in $\mathbf{PA}_j$ to X_j, that is, from each variable X_k occurring on the right-hand side of equation 1.4 to X_j (see Figure 1.7). We henceforth assume this graph to be acyclic.
We sometimes call the elements of $\mathbf{PA}_j$ not only parents but also direct causes of X_j, and we call X_j a direct effect of each of its direct causes.

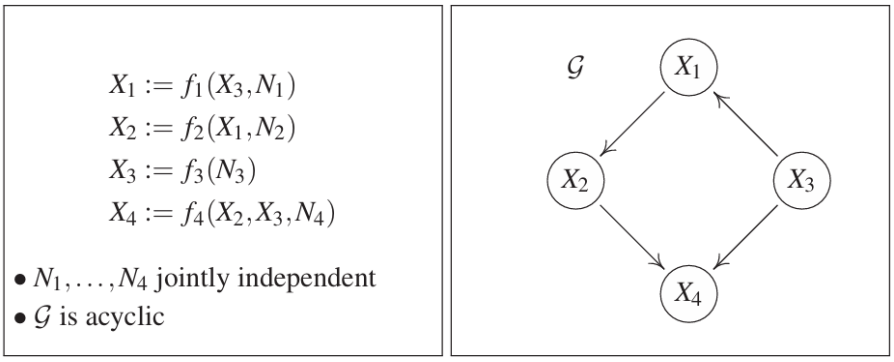


Figure 1.7: Example of an SCM (left) with the corresponding graph (right). Courtesy of Peters, Janzing, and Schlkopf, 2017

With the definition of SCM, we are able to describe the system's behavior under intervention, and how the system changes when we pose counterfactual questions. In mathematical terms, we say that the SCM entails an observational distribution, but also interventional and counterfactual distributions, which are obtained by making modifications to the SCM $\mathcal{C}$ (Peters, Janzing, and Schlkopf, 2017). In the case of interventions, the SCM is modified through the action of the *do*-operator.

The *do*-operator

Intervening on the system amounts to fixing the value of one variable, e.g. imposing $X = x$. This operation is denoted by the expression $do(X = x)$, which prescribes the deactivation of the original causal mechanism of variable X and the imposition of $X = x$. On the level of causal graphs, an intervention $do(X = x)$ removes all edges that point into X, and the intervention's effect propagates along proper causal paths starting at X. Essentially, the *do*-operator is the tool we use to perform interventions "as a surgery", as described in Paragraph 1.1.4.

As doing is fundamentally distinct from observing, it is understood that $do(X = x)$ is distinct from $X = x$. In terms of probabilities, $P(Y = y|X = x)$ is the probability that $Y = y$ conditional on finding $X = x$, while $P(Y = y|do(X = x))$ is the probability that $Y = y$ when we intervene to make $X = x$. In the distributional terminology, $P(Y = y|do(X = x))$ represents the population distribution of Y if everyone in the population had their X value fixed at x (J. Pearl, M. Glymour, and Jewell, 2016).

1.2.3 Graphical causal models

Every SCM is associated with a *graphical causal model*, informally, a graph (Lauritzen, 1996). Graphical models consist of a set of nodes representing the cause and effect variables, and a set of edges between the nodes, representing the functions in f_j (see

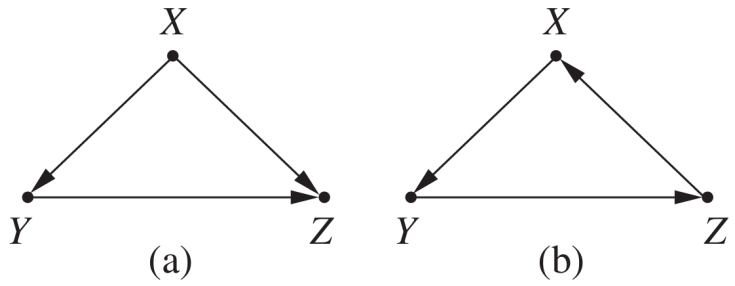


Figure 1.8: An acyclic graph (a) and a cyclic graph (b).

Eq. 1.4). The most common type of graphical causal model is the *directed acyclic graph* (DAG). The association between SCMs and graphical causal models allows to give a graphical definition of causation: if, in a graphical model, a variable X descends directly from another variable Y, then Y is a direct cause of X. If X is a descendant of Y, then Y is a potential cause of X, as it is higher in the graph hierarchy but not necessarily it will have causal effect on X (J. Pearl, M. Glymour, and Jewell, 2016). Note that in applications *cyclic* graphs where arrows exit and enter the same node have also become popular, even if they cannot directly be handled in this framework, see Figure 1.8.

Graphical models contain less information than SCMs, but they work well in situations when we only possess information about the presence or absence of a causal link, rather than the strength of the connection. For instance, the concept of independence can be expressed visually using DAGs, and the graphical representation allows us to capture the probabilistic information that is embedded in a structural equation model in an easy way. As a result, if we have the information about the structure of an SCM, we can predict patterns of independencies in the data, without relying on any quantitative information carried by the equations or by the distributions of the errors (J. Pearl, M. Glymour, and Jewell, 2016). An alternative to graphical causal models are *time series graphs*, introduced more recently in Eichler, 2011.

One important tool to study graphical causal models is d-separation, which allows to predict which dependencies will be shared by all possible data-sets generated by the graph at study.

d-separation

d-separation, where the d stands for "directional", allows us to determine, for any pair of nodes, whether the nodes are d-connected, meaning there exists a connecting path between them, or d-separated, no such path exists. If a pair of nodes are d-separated the variables they represent are independent. A pair of nodes are d-connected if they are possibly, or most likely, dependent. Two nodes X and Y are d-separated if every path between them is blocked. If even one path between X and Y is unblocked, X and Y are

Level (Symbol)	Typical Activity	Typical Questions	Examples
1. Association $P(y\|x)$	Seeing	What is? How would seeing X change my belief in$Y$?	What does a symptom tell me about a disease? What does a survey tell us about the election results?
2. Intervention $P(y\|do(x), z)$	Doing	What if? What if I do X?	What if I take aspirin, will my headache be cured? What if we ban cigarettes?
3. Counterfactuals $P(y_x\|x', y')$	Imagining, Retrospection	Why? Was it X that caused Y? What if I had acted differently?	Was it the aspirin that stopped my headache? Would Kennedy be alive had Oswald not shot him? What if I had not been smoking the past 2 years?

Figure 1.9: The ladder of causation. A classification of causal information in terms of the questions each class is capable of answering.

d-connected (J. Pearl, M. Glymour, and Jewell, 2016).

1.2.4 Framing causal questions: the Ladder of Causation

Causal information can be organized in a hierarchy, in terms of what type of questions we can give an answer to using a certain type of causal information. In this section, we study the different levels of this hierarchy and the types of research questions associated to them, following the schematization elaborated in Jakob Runge, Gerhardus, et al., 2023.

Association, Intervention and Counterfactuals

Causal information can be sharply classified in terms of classes, and in terms of what type of questions each class is capable of giving an answer to. The approach was originally developed by Judea Pearl, 2009. The classification forms a three-level hierarchy, shown in Figure 1.9, in the sense that questions at level i ($i = 1, 2, 3$) can only be answered if information from level j ($j \geq i$) is available. We call the first level *Association*, because it invokes purely statistical relationships, defined by the raw data. For instance, observing a customer who buys toothpaste makes it more likely that they buy floss; such association can be inferred directly from the observed data using conditional expectation. Questions at this layer are placed at the bottom level on the hierarchy, because they require no causal information. The second level, *Intervention*, ranks higher than Association: it involves not just seeing what is, but changing what we see. A typical question at this level would be: What happens if we double the price? Such questions cannot be answered from sales data alone, because they involve a change in customers behavior, in reaction to the new pricing. Customer choices under the new price structure

may differ substantially from that prevailing in the past. Finally, the top level contains the *Counterfactuals*. The term has long philosophical tradition[3], and it has been given structural semantics in the SCM framework, as we have already mentioned. A typical question in the counterfactual category is "*What if I were to act differently?*", a question that calls for retrospective reasoning. Counterfactuals are placed at the top of the hierarchy because they subsume interventional and associational questions. If we have a model that can answer counterfactual queries, we can also answer questions about interventions and observations.

The translation does not work in the opposite direction. The ladder of causation states that interventional questions cannot be answered from purely observational information (i.e. from statistical data alone). No counterfactual question involving retrospection can be answered from purely interventional information, such as that acquired from controlled experiments. The hierarchy is therefore directional, with the top level being the most powerful one (Judea Pearl, 2009).

Syntactic signatures of levels

Each layer in the hierarchy has a syntactic signature that characterizes the sentences admitted into that layer, and synthetically reported in the first column of Figure 1.9. The associational layer is characterized by conditional probability sentences. At the interventional layer we find sentences of the type $P(y|do(x), z)$, which denotes "The probability of event $Y = y$ given that we intervene and set the value of X to x and subsequently observe event $Z = z$". Such expressions can be estimated experimentally from randomized trials or analytically using Causal Bayesian Networks Judea Pearl, 2009. The last level features expressions of the type $P(y_x|x', y')$ which stand for "The probability of event $Y = y$ had X been x, given that we actually observed X to be x' and Y to be y'".

1.2.5 Applications of the Ladder of Causation

The possibility to separate causal questions in this sharp hierarchy is helpful both in theoretical and practical terms. In the latter case, it allows to know exactly what kind of tools we need to answer a certain type of question: is observational data enough, or do we need to intervene on the system, or will we have to resort to counterfactual methods? In Jakob Runge, Gerhardus, et al., 2023 the ladder of causation is used to build a practical template, the Questions-Assumptions-Data template, which guides researchers to phrase and solve their problems in the framework of causal inference. In the following we see the main categories of causal problems.

[3]One historical counterfactual definition of causality was given by philosopher David Hume: "*We may define a cause to be an object followed by another, where, if the first object had not been, the second never had existed*" Hume, 2004.

Causal learning / Causal discovery

Causal discovery, causal structure learning or causal network reconstruction (Spirtes, C. Glymour, and Scheines, 2001; Peters, Janzing, and Schlkopf, 2017; J. Runge, 2018), aims to reconstruct qualitatively the causal graph, or at least partial information about it, between a set of variables from samples of the observational distribution $p(\cdot)$ and/or from interventional data. In section 1.3.3 we will see some approaches at causal discovery.

Causal effects

The goal of causal effect estimation is to quantify the total causal effect of one variable or set of variables on another. These questions require that we already have knowledge about the causal graph determining the presence or absence of causal links between variables. Causal effects can be defined as the quantity

$$\Delta_{\mathbf{X}\to Y}(\mathbf{X}', \mathbf{X}) = \mathbb{E}[Y|do(\mathbf{X} = \mathbf{x}')] - \mathbb{E}[Y|do(\mathbf{X} = \mathbf{x})], \tag{1.5}$$

that is, as the difference in the expected value of Y when setting $\mathbf{X}$ by intervention to $\mathbf{x}'$ as opposed to $\mathbf{x}$. Other definitions, based on post-interventional distributions, are possible. In addition, also Granger causality-like methods are used to identify the magnitude of causal effects.

There is no general agreement on which method gives the most accurate answer. For instance, it has been discussed in Smirnov, 2022 how different methods evaluate different magnitudes of effects, but similar conclusions are drawn in every study that compares different methods applying them to the same system (Krakovská et al., 2018). We expand this point at the end of Chapter 2.

Counterfactual questions

We have seen that counterfactual questions ask what would have happened to Y if X had been set to a different value, and correspond to the higher step of Pearl's ladder of causation. An example of counterfactual questions in climate science is the causal attribution of weather and climate-related events, which has been explored in A. Hannart et al., 2016; Alexis Hannart, Carrassi, et al., 2016 Carrassi et al., 2017; Alexis Hannart and Naveau, 2018. The approach that is there followed is to evaluates the extent to which a given external climate forcing, for instance solar irradiation, greenhouse gas emissions, or ozone concentrations, has changed the probability of occurrence of event Y. The probabilities of the factual world, where the forcing is present ($X = 1$), and of the counterfactual world, where the forcing has been removed ($X = 0$), are then compared and stored in the fraction of attributable risk indicator, $FAR = 1 - \frac{p_0}{p_1}$. The FAR is interpreted as the fraction of the likelihood of an event that is attributable to

the external forcing. The conventional framework of counterfactuals and the FAR were initially adapted from best practices in epidemiology (Rothman and Greenland, 1998).

Causal claims follow from the FAR and its uncertainty, associated with model and sampling errors. They result in statements such as "*It is very likely that over half the risk of European summer temperature anomalies exceeding a threshold of 1.6°C is attributable to human influence*" (Stott, Stone, and Allen, 2005). These works apply Pearl's theory in its interventional sense, as simulating the counterfactual system amounts to a (simulated) intervention on the system.

In many applications, including in climate sciences, we find ourselves in this second category. In the next section we understand the context and the tools of causal inference from time-series.

1.3 Learning causal models from time series

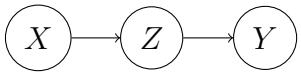
To learn a causal model it means to estimate the causal model underlying some observational and/or interventional data. In this work we dedicate particular attention to the first, and harder, task: can we infer the causal structure of the model by using *purely observational information*? We said that this problem is faced when intervening on the system is not possible, and simulations are too expensive or not accurate enough, which is often the case in Earth system and climate sciences. Moreover, the ever growing diffusion of data-driven methods increases the appeal of developing safe data-based causal inference. In some cases, additional physical information about the relationships between variables is also available, for example the data being time-indexed, and it should be used to aid the process of causal learning.

Given the compelling nature of the previous question, many different methods, grounded in different philosophical approaches, have been developed to answer the question of causal inference from observational data. We have seen that Clive Granger had started wondering about a solution to this problem already in the 60s. Ever since then, his method has been improved and extended, and the interest in data-driven causal inference has become of extremely wide interest. Pearl's school of thought, instead, strongly links causality to intervention on the system. Therefore in this framework, strictly speaking and with the ladder of causation in mind, all we can measure from observational data are the statistical dependencies among the variables and their time lags, more precisely called *conditional independence relations*.

Definition 1.3.1 (**Conditional independence**). *Consider three events X, Y and Z. X and Y are said to be conditionally independent given Z if and only if $P(Z) > 0$ and*

$$P(X|Y,Z) = P(X|Z) \tag{1.6}$$

We denote the property as $X \perp\!\!\!\perp Y|Z$ and graphically as



The knowledge about conditional (in)dependencies gives us precious information about the system at study, and sometimes causal information, too. This possibility has been practically implemented with the development of conditional independencies-based causal algorithms, and the literature has grown to be quite vast. It has been proven that also Granger-like approaches can be re-conducted to this framework (Jakob Runge, Gerhardus, et al., 2023). The correct application of methods that exploit conditional independencies in the data requires that the system satisfies a set of assumptions, so that (part of) the causal structure can be recovered from the data. These assumptions have a very general meaning, that goes beyond the algorithms we have cited in this paragraph. They are also valid - even if less faithfully applied - for Granger-derived methods of causality.

In the next paragraph we see how the inference of causal relationships from conditional dependencies can be formalized.

1.3.1 From correlation to causation: Reichenbach's principle

Questions that belong to the first level of the ladder of causation in Figure 1.9, are questions of correlation, and the famous topos recites that *correlation is not causation*: statistical properties alone do not determine causal structures. However, it can be argued that while "*one may postulate that while we cannot infer a concrete causal structure, we may at least infer the existence of causal links from statistical dependencies.*" (Peters, Janzing, and Schlkopf, 2017). This concept was first formulated by the science philosopher Hans Reichenbach.

Principle 1.3.1 (**Reichenbach's common cause principle**)**.** *If two random variables X and Y are statistically dependent, then there exists a third variable Z that causally influences both. (As a special case, Z may coincide with either X or Y.) Furthermore, this variable Z screens X and Y from each other in the sense that given Z, they become independent, $X \perp\!\!\!\perp Y|Z$.*

In practice, the principles hints that if we find a correlation between two variables, they will also be causally related, in one of the ways shown in Figure 1.10.

This principle does not end the debate about correlation not implying causation: the random variables might be conditioned on others, or they might only *appear* to be dependent, for example if they both follow an exponential law in time, but do not have shared physical meaning. This case goes under that of spurious correlations. Nonetheless, the principle gives us a kind of operative definition, while warning us that we will need more assumptions to be able to state something about the causal structure between the variables.

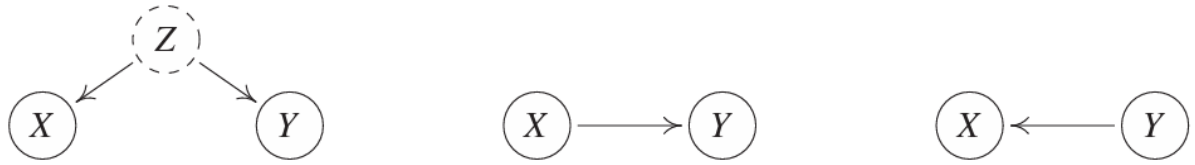


Figure 1.10: Reichenbach's common cause principle establishes a link between statistical properties and causal structures. A statistical dependence between two observables X and Y indicates that they are caused by a variable Z, a confounder. The figure simplifies if Z coincides either with X or with Y Peters, Janzing, and Schlkopf, 2017.

A variety of different assumptions have been shown to be sufficient to estimate the true causal graph from time series. Such assumptions are not necessarily easily respected by datasets in real situations, which further motivates research into whether alternative assumptions, more naturally encountered in practice, could yield the same results (Spirtes, C. Glymour, and Scheines, 2001; Peters, Janzing, and Schlkopf, 2017). We will trace the assumptions for causal discovery from observational time series in the next section.

1.3.2 Required assumptions

Causal inference from data is an emerging topic in many fields of science. In the case in which the dataset is represented by time series, we can rely on the fact that the causes precede the effects, but we have to deal with some difficulties linked to the nature of the acquisition of the data: for instance the coarse graining of the time series and thus the possible presence of instantaneous effects or of missing information. In the following, we present the main assumptions under which the time series graph represents causal relations, as traced in J. Runge, 2018.

Causal Sufficiency

When trying to infer causal relationships from data, we have to make sure that all the relevant causal variables have been taken into consideration. It was first suggested by Clive Granger that methods should consider "the time series $\mathbf{U}_t^-$ representing all available information [in the universe]" (Granger, 1969), however, we often need much less than the whole universe

Definition 1.3.2 (**Causal sufficiency**). *A set $W \in V \times \mathbb{Z}$ of variables is causally sufficient for a process $\mathbf{X}$ if and only if in the process every common cause of any two or more variables in W is in W or has the same value for all units of the population.*

In this way no potentially relevant causes for the process X are left out. Indeed, if the set of causal variables satisfies causal sufficiency, it means there are no unobserved

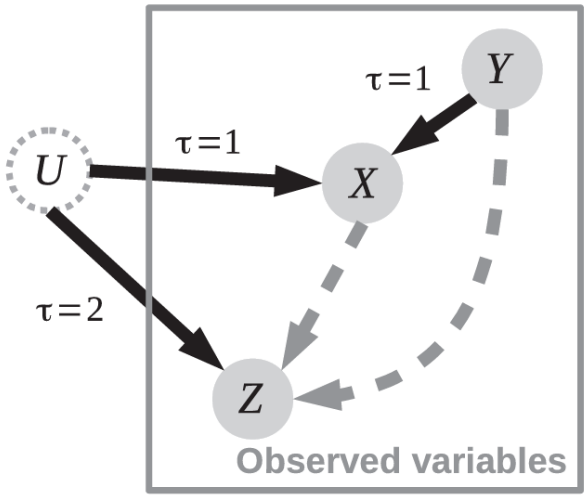


Figure 1.11: The problem of unobserved variables. Here, U is a latent confounder and affects the correct recovery of causal links between (X, Y, Z), leading to spurious links. All causal links are determined at time lag τ. Courtesy of J. Runge (2018).

variables, which could substantially change the inferred causal graph if not considered inside the set of causally relevant variables.

In observational time series analysis it is not easy to satisfy causal sufficiency: the assumption could be violated if one or more variables are forgotten, but also if all variables are observed but are sampled at too coarse time intervals relative to the causal links. Research on algorithms that account for sub-sampled is a currently active field (J. Runge, 2018).

Causal Markov Condition

The second property relates statements about graph *separation* to conditional independencies.

Definition 1.3.3 (**Markov property**). *Given a DAG $\mathcal{G}$ and a joint distribution $P_{\mathbf{X}}$, this distribution is said to satisfy the **global Markov property** with respect to the DAG $\mathcal{G}$ if*

$$\mathbf{A} \perp\!\!\!\perp_{\mathcal{G}} \mathbf{B}|\mathbf{C} \implies \mathbf{A} \perp\!\!\!\perp \mathbf{B}|\mathbf{C} \tag{1.7}$$

for all disjoint vertex sets $\mathbf{A}, \mathbf{B}, \mathbf{C}$ (the symbol $\perp\!\!\!\perp_{\mathcal{G}}$ denotes d-separation).

Intuitively, the Causal Markov Condition implies that once we know the values of Y_t's parents, all other variables in the past ($t - \tau$ for $\tau > 0$) become irrelevant for predicting Y_t. If Causal Sufficiency is not fulfilled, also the Markov condition will generally not hold. Examples of graphs that violate the Markov condition are given in J. Runge (2018).

A distribution P_{X_1,X_2,X_3,X_4} is Markovian with respect to the graph $\mathcal{G}$ shown in Figure 1.7 if according to Definition 1.3.3,

$$X_2 \perp\!\!\!\perp X_3|X_1 \quad \text{and} \quad X_1 \perp\!\!\!\perp X_4|X_2, X_3 \tag{1.8}$$

This comes for free in the SCM structure: the distributions entailed from an SCM is Markovian with respect to the graph of the SCM. But what if we want to solve the inverse problem of inferring the causal graph from the entailed conditional independencies?

Faithfulness

The Causal Markov Condition guarantees that separation in the graph implies independence in the process. But something can be concluded from the reverse direction too, from an estimated conditional independence relation. *Faithfulness* guarantees that the graph entails all conditional independence relations that are implied by the Markov condition.

Definition 1.3.4 (**Faithfulness**). *Consider a distribution $P_{\mathbf{X}}$ and a DAG $\mathcal{G}$. $P_{\mathbf{X}}$ if faithful to the DAG $\mathcal{G}$ if*

$$\mathbf{A} \perp\!\!\!\perp \mathbf{B}|\mathbf{C} \implies \mathbf{A} \perp\!\!\!\perp_{\mathcal{G}} \mathbf{B}|\mathbf{C} \tag{1.9}$$

for all disjoint vertex sets $\mathbf{A}, \mathbf{B}, \mathbf{C}$.

The Markov condition and Faithfulness can be used to justify Reichenbach's common cause principle. Assuming there exists an SCM over the collection $\mathbf{X}$ of random variables that contains X and Y with graph $\mathcal{G}$, then Reichenbach's common cause principle follows from the Markov property. A proof can be found in Peters, Janzing, and Schlkopf (2017). It has also been shown in Peters, Janzing, and Schlkopf, 2017 that some causal conclusions can be made from conditional dependencies without using Faithfulness, in the case of time series data.

An example of a distribution that is Markovian but violates Faithfulness is obtained by making two paths cancel each other "by accident", when the computation of a conditional independence vanishes for a numerical coincidence. This creates an independence that is not implied by the graph structure, and it means that the distribution is not Faithful.

Intuitively, Faithfulness together with the Causal Markov Condition allow us to conclude that, in the limit of infinite sample size, a measured independence, given any set of conditions, implies that no *direct* causal mechanisms exists. On the contrary, a measured statistical dependency is actually due to some causal mechanism, although *not necessarily direct*.

This is an important point and a problem in causal network reconstruction. The Markov condition relates statements about graph separation to conditional independencies. This means that *different graphs might encode the exact same set of conditional*

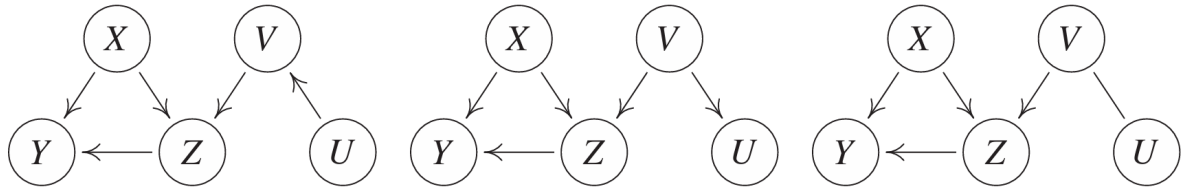


Figure 1.12: The left and center DAGs are Markov equivalent: they entail the same set of conditional independencies. The graph on the right represents their *Markov equivalence class*: all the DAGs sharing the set of conditional independencies. It corresponds to the structure we would infer from the conditional independencies in a dataset. Courtesy of Peters, Janzing, and Schlkopf (2017).

independencies (i.e. different graphs can be *Markov equivalent*). On the other hand, Faithfulness tells us we can retrieve *undirected* graphical connections from conditional independencies. It means that starting from observed conditional independencies, we rarely end up inferring only one possible causal graph. This might pose a problem if the direction of the causal links is relevant in the problem at study, which is often the case. We show this case in Figure 1.12.

Other assumptions

Causal inference from time series requires some additional assumptions. Consider Granger causality or the definitions of time series graphs: they rely on lagged definitions of causality.

To make the definitions work, we need to assume that there are **no instantaneous causal effects**, i.e. that we can observe a time lag between the cause and the effect. If we allow instantaneous effects in the system we cannot leverage the advantage of the direction of time to understand which variable is a cause and which an effect. This problem becomes particularly relevant in the case of sub-sampling of the time series.

The propriety of **stationarity** is also often required. A common example of non-stationarity is found in climate time series, which are subject to many non-stationary forcings. A classical example is the increasing trend of greenhouse gas emissions in the atmosphere in the last two centuries. If we don't consider the prior knowledge that different variables might be driven by such common signal, we might end up stating that those variables are causally related, when in reality they just share a common pattern due to an external forcing J. Runge, 2018. It could be argued that non-stationarity is a problem of violation of causal sufficiency: we violate causal sufficiency in the greenhouse gas examples because we do not possess point-wise information on the time-changing forcings acting on a system. In principle we would solve the problem if we could include that information in the causal inference process.

Assumptions have to be made also regarding the **type of dependency structure**

underlying the system. Knowing if the relationships between variables in a system are linear or nonlinear, bivariate or multivariate, permits to choose methods that are more accurate to analyze them. While classical statistical methods are often based on the assumptions of linearity, modern statistics, the physics community and machine learning have developed non-parametric or model-free methods that allow to better capture the nonlinearity of the systems at study, at the cost of weaker theoretical results.

One last but relevant issue is that of **measurement error**, which contaminates the variables between which we seek to reconstruct dependencies and makes the causal graph reconstruction difficult. More details and practical examples about these assumptions can be found in J. Runge, 2018.

Assumptions on the absence of a causal link

It is very easy to violate causal sufficiency by not measuring or considering a causally relevant variable; indeed, real situations rarely meet all the requirements above. A conclusion on the existence of a causal link, thus, rests on a number of strong assumptions. Something more straightforward can be said about the *absence of a causal link*, instead.

Definition 1.3.5. *Let $\tilde{\mathbf{X}}$ be measurements of a stochastic process $\mathbf{X}$. Assuming Faithfulness (Definition 1.3.4) and that all variables in $\tilde{\mathbf{X}}$ are measured without error we have that for $\tilde{X}, \tilde{Y}, \tilde{\mathbf{Z}} \in \tilde{\mathbf{X}}$ with $\tilde{X}, \tilde{Y} \notin \tilde{\mathbf{Z}}$ if*

$$\tilde{X}_{t-\tau} \perp \tilde{Y}_t \mid \tilde{\mathbf{Z}} \quad \textit{for any subset } \tilde{\mathbf{Z}} \in \tilde{\mathbf{X}}_t^- \Rightarrow X_{t-\tau} \not\to Y_t, \tag{1.10}$$

that is, if independence is measured given any subset of conditions, then there is no direct causal link between $X_{t-\tau}$ and Y_t in $\mathcal{G}$.

This assert rests on much weaker assumptions than the existence of a causal link, it does not require Causal Sufficiency or the Causal Markov Condition J. Runge, 2018.

1.3.3 Learning causal models with SCMs: algorithms

Currently, time-series causal inference counts a large variety of different approaches, which have been reviewed in Jakob Runge, Gerhardus, et al., 2023. In the next chapter we will deepen Granger causality, information-theoretic and other data-driven approaches. However, some methods have been developed based on the framework we have so far described. These methods have strong theoretical grounds, but they are not free of limitations. Above all, the assumptions we have reviewed above are not easily met in datasets.

One category is that of *constraint-based approaches*, which assume that all considered data is generated by an SCM. With this hypothesis, the Markov property holds and the connectivity pattern of the causal graph imprints corresponding marginal and conditional

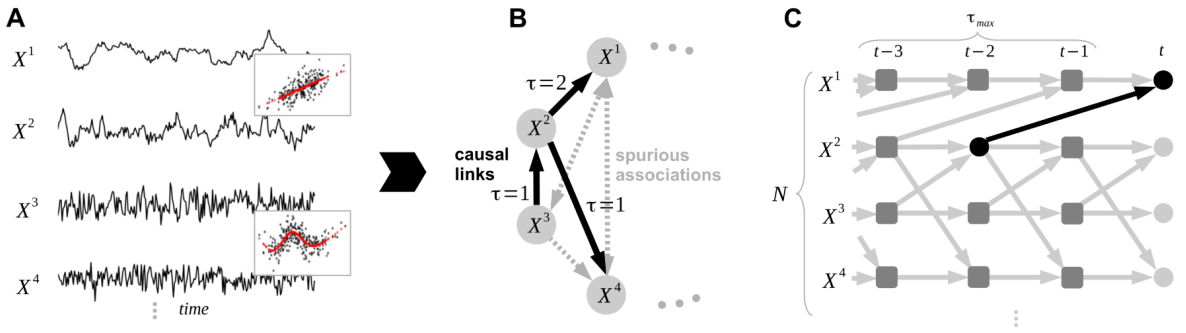


Figure 1.13: Causal model reconstruction. Consider a time series dataset (panel **A**) from a complex system of which we try to reconstruct the underlying causal dependencies (panel **B**), accounting for linear and nonlinear dependencies and including their time lags (link labels). Causal discovery aims to unveil spurious associations (gray arrows) which necessarily emerge due to common drivers (e.g., $X_1 \leftarrow X_2 \rightarrow X_4$) or transitive indirect paths (e.g., $X_3 \rightarrow X_2 \rightarrow X_1$). (C) The time series graph defined in Definition 1 resolves also the time-dependence structure up to some maximum time lag τ_{max}. A link $X^i_{t-\tau} \rightarrow X^j_t$ (black edge) exists if $X^i_{t-\tau}$ and Y^j_t are not conditionally independent on the past of the whole process (gray boxes). Courtesy of J. Runge (2018).

dependencies and independencies into the observational distribution. The basic idea of these approaches is to perform a sequence of statistical tests of independence for all pairs of variables (X, Y) in the system and appropriately chosen sets of conditioning variables $\mathbf{Z}$. Based on the results of these tests, the structure of the causal graph is constrained, as shown in Figure 1.13 where the causal graph is inferred from time series.

Notable examples are the Peter-Clark (PC) algorithm (Spirtes, C. Glymour, and Scheines, 2001), the Fast Causal Inference (FCI) algorithm, which drops the assumption of causal sufficiency (Spirtes, C. Glymour, and Scheines, 2001), and extensions of these two techniques specifically designed for time series (Jakob Runge, Gerhardus, et al., 2023). Constraint-based methods have found large application in numerous Earth sciences problems, for example in Chu, Danks, and C. Glymour, 2005, which also connects the PC algorithm to Granger causality. Importantly, it can be argued that Granger causality and information-theoretic extensions also essentially utilize independence relations (Peters, Janzing, and Schlkopf, 2017; J. Runge, 2018), we will deepen this point in the next chapter.

Key concepts

- The concept of causality has been widely discussed and has evolved from debates in philosophy and statistics to the modern approaches used today. One early contribution was the development of the first *graphical method* in the early 1900s, by Sewall Wright.

- The field gained significant attention in the mid-20th century with the emergence of the *interventional definition of causality*, alongside the development of the first data-driven method, known as Granger causality.
- The formal framework for causality, based on the interventional definition, was established toward the end of the 20th century. It uses causal graphs as the foundation of Structural Causal Models (SCMs), which describe the causally relevant features of a system and their interactions.
- Causal questions can be organized into a hierarchy based on the type of causal information available. The main categories include *learning causal models*, *estimating causal effects*, and addressing *counterfactual questions*.
- A major challenge in causal inference is *reconstructing a causal graph from time series* or observational data, a problem frequently encountered in Earth system sciences.
- Reichenbach's principle allows for the inference of partial causal information from *conditional independencies* among variables, provided that the key assumptions of *causal sufficiency*, the *causal Markov condition*, and *faithfulness* are met.

Chapter 2

Causal inference from time series: methods in a nutshell

2.1 Granger Causality

The idea of Granger causality was introduced by the econometrist Clive Granger in the 1960s. Granger's initial interest in causality was ignited by a question regarding a pair of stochastic processes, which seemed to be inter-related. Was it possible to extract one-way dependence from the output time series? Granger gave a practical answer, based also on Norbert Wiener's work Wiener, 1956, which became extreme successful, first in econometrics, and later in many other disciplines, including Earth system studies. The underlying idea of measuring whether X *Granger causes* Y is that there is some unique information in X relevant for Y that is not contained in Y's past, as well as the whole system's past Granger, 1969.

Conditioning on the whole system, if we do not know any better, coincides with conditioning on the whole universe. This assumption can be substituted by the more reasonable request of accounting for the *minimal sufficient* set of variables, recalling the definition of Causal Sufficiency given in the previous Chapter. However, even this weakened assumption can be very hard to satisfy in practice. In fact, Granger causality and all the other methods that are based on observational data strongly suffer from the violation of the Causal Sufficiency assumption.

The most common framework to operationally compute Granger causality are vector auto-regressive models (VAR),

$$\mathbf{X}_t = \sum_{\tau=1}^{\tau_{max}} A(\tau)\mathbf{X}_{t-\tau} + \eta_t, \tag{2.1}$$

where $\mathbf{X}_t = (X_t^1, ..., X_t^N)$ is an N-dimensional multivariate time series, $A(\tau)$ is the $N \times N$ coefficient matrix at lag τ, τ_{max} some maximum time lag, and η denotes an

independent noise term.

For instance, considering a 2-dimensional system and discrete-time lags, we have

$$\begin{pmatrix} X_t^1 \\ X_t^2 \end{pmatrix} = \sum_{\tau=1}^{\tau_{max}} \begin{pmatrix} A_{11}(\tau) & A_{12}(\tau) \\ A_{21}(\tau) & A_{22}(\tau) \end{pmatrix} \begin{pmatrix} X_{t-\tau}^1 \\ X_{t-\tau}^2 \end{pmatrix} + \eta_t \tag{2.2}$$

If we find $A_{21} \neq 0$ for any time lag τ, A_{21} is a causal link from X^1 to X^2.

In general, a variable X_t^i in the set *Granger causes* another variable X_t^j if any of the coefficients $A_{ji}(\tau)$ at lag τ is non-zero. A non-zero component $A_{ji}(\tau)$ can then be denoted as a causal link $X_{t-\tau}^i \to X_t^j$ at lag τ.

Granger's method views causality as a form of *increased predictability*. The definition relies on two concepts: (i) *the cause should occur before the effect*, and (ii) *the cause should contain information about the caused that is not available otherwise*. A relationship that fulfills both requirements is thus defined as causal Sun, Taylor, and Bollt, 2015. This view of causality has seen strong oppositions. Judea Pearl argued that the approach should be classified as *statistical* rather than causal, i.e. it belongs on the first level of the ladder of causation (Judea Pearl, 2009). On more pragmatic grounds, Granger argued "*Applied economists found the definition [of Granger causality] understandable and useable and applications of it started to appear*" (A. Seth, 2007). Indeed, Granger causality has been applied to a variety of different contexts, and it often gives better insights with respect to correlation. This highlights an important point: Granger causality may not adhere to the interventional definition of causality, and it requires certain assumptions about the set of variables to function properly. However, when these assumptions are met, it provides information that simple correlation cannot provide, as it yields a directed measure of influence rather than the symmetric association captured by a correlation coefficient.

The original mathematical formulation of Granger causality was based on linear regression modeling, involving only two variables. Many extensions of Granger-causality have been later developed, extending the applicability to more than two variables, to applications with partial or conditional characteristics (Smirnov, 2022), and making it possible to study non-linear relationships (Marinazzo, Pellicoro, and Stramaglia, 2008). In Bueso, Piles, and Camps-Valls, 2020 the kernel nonlinear generalization of Granger causality is applied to the prototypical climate phenomenon of El Niño Southern Oscillation, disclosing its impact on soil moisture globally.

Nonlinear dependencies have also been modeled using information-theoretic analogs of Granger causality, in particular *transfer entropy*. We dedicate the next sections to information-theoretic based causal inference.

2.2 (Generalities on) Information theoretic approaches to causal inference

Information theory is largely used to study complex systems, due also to the increasingly-popular idea that commonalities between complex systems may be found in the way they handle information (J. Lizier et al., 2011). The large interest in traditional information-theoretic measures, such as entropy and mutual information, and in complex systems has led to the development of more advanced measures suitable to study such complexity, the characterization of order-chaos phase transitions and measures of network structure.

These measures are used to understand how the global behavior of complex systems emerges from the collective processing of information between the individual elements of a system. Therefore, one has to look at mainly three things when studying complex systems: how information is *transferred* in the interaction between elements, how it is *stored* by elements, and how the information-sources are non-trivially *combined* (Joseph T. Lizier, 2014).

A number of measures have been introduced to quantify these aspects. Information-based causal inference measures represents a systematic way of overcoming the model-dependent limitation of the linear Granger causality test (Sun, Taylor, and Bollt, 2015). An advantage of information theoretic measures, as compared to standard Granger causality, is that they are intrinsically sensitive to *nonlinear* signal properties and they do not require a model to be specified for the data. Therefore, they are a natural domain to measure information transfer. The most diffused information-theoretic measures is *Transfer Entropy*, developed by Schreiber (2000), a finite version of the *directed information* derived by James Massey a few years before (Massey, 1990).

2.2.1 Transfer Entropy

Schreiber opened this new line of research in causal inference by introducing an information theoretic measure that quantifies the *statistical coherence* between systems evolving in time, introducing the possibility to retrieve causal information in a system using *information theory* tools. This method has been widely applied in several contexts, e.g. in computational neuroscience, bio-informatics (Sun, Taylor, and Bollt, 2015), and social sciences (Steeg and Galstyan, 2011). Recently it has seen a considerable diffusion in the study of climate (Silini et al., 2022; Stramaglia et al., 2024; Delgado-Bonal et al., 2020; Tongal and Sivakumar, 2021; Dhifaoui et al., 2023).

In stochastic or deterministically chaotic systems one can use entropy to quantify the information produced in the time evolution of a system. When the system is made of more than one component, information is diversified. On one side, the components individually produce or lose information, and they exchange information among themselves on the other. Measuring the relative rates of these two information types gives us knowledge

on the structure of the system. From a causal perspective, we are interested in knowing the *rate of information exchange* between the components, which is usually identified as the causal influence of a component on another component, even if this point is object of controversy as we will discuss in Section 2.5. To address the search of causal links from the transfer of information, we introduce some information-theoretic quantities and consider how their characteristics might make them suitable to estimate the rate of information exchange between variables.

2.2.2 Elements of Information theory

The fundamental quantity in information theory is *Shannon entropy*, which measures the average number of bits needed to optimally encode outcomes of a discrete variable X following a probability distribution $p(x)$:

$$H(X) = \sum_x p(x) \log_a p(x), \tag{2.3}$$

where the sum extends over all possible states x of the process. The base of the logarithm determines the units in which information is measured. For example, base $a = 2$ leads to information measured in bits. The formulation we use in this section is valid for discrete variables and for coarse grained states of continuous systems. The limit for the resolution r going to zero is unproblematic (finite and independent of the partition) for Transfer Entropy, but has to be handled carefully in general.

In order to construct an optimal encoding that uses just as many bits as given by the entropy, it is necessary to know the probability distribution $p(x)$. What if we pick another probability distribution $q(x)$ instead? The excess number of bits that will be encoded if a different distribution $q(x)$ is used is given by the *Kullback entropy* or Kullback-Leibler divergence, which measures how *distant* the guessed distribution $q(x)$ is different from from the true one $p(x)$:

$$K(p(x) \mid\mid q(x)) = \sum_x p(x) \log \frac{p(x)}{q(x)}. \tag{2.4}$$

The Kullback entropy can be computed also for conditional probabilities $p(x|y)$. In this case we can compute the Kullback entropy for a single state y with a marginalization:

$$K_y(p(x|y) \mid\mid q(x|y)) = \sum_x p(x|y) \log \frac{p(x|y)}{q(x|y)}, \tag{2.5}$$

We can also compute the complete Kullback entropy, by also summing over y with respect to $p(y)$.

Mutual information as a symmetric measure of common information

The mutual information of two processes X and Y with joint probability $p_{XY}(x, y)$ can be seen as the excess amount of code produced by making the wrong assumption that the two processes are independent, i.e. using $q_{XY}(x, y) = p_X(x)p_Y(y)$ instead of the generic $p_{XY}(x, y)$.

The mutual information is then identified by the corresponding Kullback entropy

$$K(p(x, y)|p(x)p(y)) = I(X; Y) = \sum p(x, y) \log \frac{p(x, y)}{p(x)p(y)}. \tag{2.6}$$

Here and in the following we omit the subscripts of the probabilities and the summation on states x and y. We infer from Eq. 2.6 that Mutual Information is a natural way to quantify the *deviation from independence* of two processes. It measures the average reduction of uncertainty about x that results from knowing the value of y, or vice versa. $I(X; Y)$ is symmetric under exchange of the two indices X and Y; it does not contain any directional sense. Though it has been used to measure directed information transfer from one variable to another, this is not to be valid, as MI is a static and symmetric measure of shared information (J. Lizier et al., 2011).

Mutual information can be given a directional sense in a somewhat *ad hoc* way by introducing a time lag in either one of the variables. However, it is more natural to introduce *transition probabilities*, that naturally introduce time lags in the system (Schreiber, 2000), something which is instead naturally implemented in transfer entropy.

Another useful quantity is the *conditional mutual information* (CMI), $I(X; Y|Z)$, the MI between X and Y when Z is known:

$$I(X; Y|Z) = \sum p(x, y, z) \log \frac{p(x, y|z)}{p(x|z)p(y|z)}. \tag{2.7}$$

In terms of Shannon entropies:

$$I(X; Y|Z) = H(X|Z) - H(X|Y, Z). \tag{2.8}$$

From another perspective, using the chain rule for the MI from two variables Y_1, Y_2 jointly to another variable X:

$$I(X; Y_1, Y_2) = I(X; Y_1) + I(X; Y_2|Y_1), \tag{2.9}$$

which can be generalized to multivariate $\mathbf{Y}$ if system dimension is greater than 2. The CMI is always positive for discrete, jointly distributed random variables, and is also non-negative for continuous random variables under certain regularity conditions. Conditioning on a third random variable may *either increase* or *decrease* the mutual information: that is, the difference $I(X; Y) - I(X; Y|Z)$, may be positive, negative, or zero.

Transfer entropy as a directed measure of information transfer

To incorporate a dynamical structure one needs to study *transition probabilities* instead of static probabilities. For example we might want to know what happens if we add an additional state to the system, once we know all the previous states. To understand this scenario and compute relevant quantities, let us consider a system that may be approximated by a stationary Markov process of order k. The conditional probability to find X in state x_{n+1} at time $n+1$ is independent of the state i_{n-k}:

$$p(x_{n+1} \mid x_n, \ldots, x_{n-k+1}) = p(x_{n+1} \mid x_n, \ldots, x_{n-k+1}, x_{n-k}). \tag{2.10}$$

The shorthand notation $x_n^{(k)} = (x_n, \ldots, x_{n-k+1})$ indicates words of length k. The average number of bits needed to encode one additional state of the system if all previous states are known is given by the *entropy rate*

$$h(X) = -\sum p(x_{n+1}, x_n^{(k)}) \log p(x_{n+1} \mid x_n^{(k)}). \tag{2.11}$$

As $p(x_{n+1} \mid x_n^{(k)}) = p(x_{n+1}^{(k+1)})/p(x_n^{(k)})$, this can be shown to corresponds to the difference between Shannon entropies of the processes given respectively by $k+1$ and k dimensional delay vectors constructed from the process X:

$$h(X) = H_{X^{(k+1)}} - H_{X^{(k)}}. \tag{2.12}$$

If X is obtained by coarse graining a continuous system $\mathcal{X}$ at resolution r, the entropy $H_{\mathcal{X}}(r)$ and entropy rate $h_{\mathcal{X}}(r)$ depend on the partitioning and in general behave like $-\log r$ when $r \to 0$. For the special case of deterministic systems, this limit might exist and takes the name of Kolmogorov-Sinai entropy. The situation if opposite for the MI: for generic noisy interdependence, $\lim_{r\to 0} MI_{\mathcal{XY}}(r)$ is finite and independent of the partition, but for deterministically coupled processes, $M_{\mathcal{XY}}(r)$ grows without bound as $r \to 0$ (Schreiber, 2000).

To study the dynamics of shared information between processes one needs to generalize the entropy rate to more than one system. Therefore, we build a mutual information rate by generalizing $h(X)$ to the case of two processes (X, Y). This can be achieved with the MI, as seen before; however, the resulting Kullback entropy is still symmetric under the exchange of X and Y. Therefore, to obtain a *directed* quantity, we measure the deviation from the generalized Markov property:

$$p(x_{n+1} \mid x_n^{(k)}) = p(x_{n+1} \mid x_n^{(k)}, y_n^{(l)}). \tag{2.13}$$

where k and l indicate two possible lengths for y_n memory.

If there is no information flowing from Y to X, the state of Y has no influence on the transition probabilities of process X. We can quantify by how much this assumption is

violated using a Kullback entropy with the generalized Markov properties, by which we define the *transfer entropy*:

$$T_{k,l}(Y \to X) = \sum p(x_{n+1}, x_n^{(k)}, y_n^{(l)}) \log \frac{p(x_{n+1} \mid x_n^{(k)}, y_n^{(l)})}{p(x_{n+1} \mid x_n^{(k)})}. \tag{2.14}$$

This formula computes the distance between the assumption that state x_{n+1} is determined both by its past and by process Y, with respect to assumption that the state n_{n+1} is only "caused" by its past states. Natural choices for l are $l = k$ or $l = 1$. Transfer entropy quantifies the information provided by a source node about a destination's next state that was not contained in the past of the destination. It can be equivalently expressed as conditional mutual information, or as the sum of two mutual information quantities:

$$\begin{aligned} T_k(Y \to X) &= I\big(Y; X_{n+1} \mid X_n^{(k)}\big), \\ beginequation1mm]T_k(Y \to X) &= I\big(Y; (X_{n+1}, X_n^{(k)})\big) - I\big(Y; X_n^{(k)}\big), \end{aligned} \tag{2.15}$$

where X_{n+1} refers to the next state of the destination, and $X_n^{(k)}$ refers to its past k states. TE is computed in the last equation using two mutual information estimators, as developed in Kraskov, Stögbauer, and Grassberger (2004) and Kraskov, 2004. The resulting quantity $T_k(Y \to X)$ is now non-symmetric. It measures the degree of dependence of X on Y and not vice versa.

> Transfer Entropy is usually understood in terms of *resolution of uncertainty*. The TE from Y to X is the degree to which Y disambiguates the future of X beyond the degree to which X already informs its future.

Unlike mutual information, transfer entropy is designed to ignore static correlations due to the common history or common input signals Schreiber, 2000, which means that it is studied to marginalize the impact of confounding variables.

2.2.3 Granger Causality and Transfer Entropy

Granger causality is framed in terms of *prediction*. On the other hand, we said that Transfer Entropy is understood in terms of resolution of uncertainty. This indicates an attractive symmetry between the notions of *prediction* and that of *disambiguation*. In fact, it has been proven that Granger causality and Transfer Entropy are equivalent for Gaussian variables (Barnett, Barrett, and A. K. Seth, 2009). In practice, numerical equivalence between the results computed for the two methods will depend on the method

used to estimate transfer entropy in sample. Even if the equivalence is restricted to Gaussian variables, we note that the Gaussian assumption is valid in many cases, and it is often used as an analytical benchmark even for systems that are not Gaussian.

Both Granger causality and transfer entropy, in turn, are essentially tests for *conditional independence* (J. Runge, 2018; Peters, Janzing, and Schlkopf, 2017). Tests for causality are based on checking whether a particular conditional mutual information is greater than zero. Let us consider the conditional mutual independence of X and Y given Z (Eq. 2.7). The conditional independent test

$$X \perp\!\!\!\perp Y \mid Z \iff p(x, y \mid z) = p(x \mid z)p(y \mid z) \quad \forall\, x, y, z \tag{2.16}$$

substituted in 2.7 then gives

$$X \perp\!\!\!\perp Y \mid Z \iff I(X; Y \mid Z) = 0, \tag{2.17}$$

which implies that the TE is zero, based on its definition via CMI (2.15). This formalization allows to link the three frameworks, which are often considered to be separate and counter-posed.

2.2.4 Multivariate Transfer Entropy

Transfer entropy was originally developed by Schreiber to be applied only to bivariate scenarios. This lack of generality entails many limitations for the method. Without proper conditioning, inference based on bivariate transfer entropy tends to produce systematic errors due to, for example, the effects of indirect influences and dominance of neighbors.

Conveniently, the theory permits the extension of information transfer measures to be achieved trivially, by generalizing the scalar random variables to variables in $\mathbb{R}^N$. Therefore, for the mutual information we have $I(\mathbf{X}; \mathbf{Y})$ and the transfer entropy becomes

$$T_k(\mathbf{Y} \to \mathbf{X}) = I(\mathbf{Y}; \mathbf{X_{n+1}} | \mathbf{X}_n^{(k)}), \tag{2.18}$$

where $\mathbf{X}$ and $\mathbf{Y}$ are variables in $\mathbb{R}^N$, $\mathbf{X}_{n+1}$ refers to the next state of the destination and $\mathbf{X}_n^{(k)}$ refers to its past k states. The *multivariate mutual information* $I(\mathbf{X}; \mathbf{Y})$ measures the amount of information shared between a set of source variables $\mathbf{Y}$ and a set of destination variables $\mathbf{X}$. The *multivariate transfer entropy* $T(\mathbf{Y} \to \mathbf{X})$ measures the amount of information that a set of source variables $\mathbf{Y}$ provides about a set of destination variables $\mathbf{X}$, that was not contained in the past of the destination set.

J. Lizier et al. (2011) describe a method to estimate numerically the TE between subsystems of a multivariate system, called *interregional* Transfer Entropy. This implementation accounts for *collective interactions* resulting from the combined activity of multiple driving elements, other than directional and non-linear relationships. This

method is particularly suitable to study the interactions between subsystems of variables, and we will employ it in Chapter 4 to study the causal relations between ocean and atmosphere in the MAOOAM model.

In practice, several different approaches have been developed to extend the Transfer Entropy to multivariate systems. Sun, Taylor, and Bollt (2015) have generalized $T(Y \to X)$ to *causation entropy* taking into account the state of Z as $T(Y \to X|Z)$. A similar approach is used in Ay and Polani, 2008. Smirnov (2022) extends the dynamical causal effect formalism to the case of more than two subsystems, also proving how the causation entropy is naturally produced as a generalization of the TE within this formalism. Other examples of multivariate extensions are found in and in Jakob Runge, Heitzig, Marwan, et al. (2012), Jakob Runge, Heitzig, Petoukhov, et al. (2012), and Jakob Runge, Donner, and Kurths (2015). Joseph T. Lizier, Mikhail Prokopenko, and Zomaya (2008) and J. T. Lizier and M. Prokopenko (2010) introduce the concept of *complete* TE to take into account the state of a third or multiple subsystems, as opposed to *pairwise* or*apparent* TE, which corresponds to bivariate TE. In numerical experiments, we rename the complete TE as *multivariate TE*, for better coherence with other methods.

Another line of research investigates the use of Transfer Entropy to study interactions in complex systems, where a bivariate approach is not sufficient to capture the overall behavior of the system. The focus is thus placed on *higher-order effects*, compared to two body effects. Transfer Entropy is decomposed in unique, redundant and synergistic atoms (Rosas et al., 2022; Stramaglia et al., 2024; Martínez-Sánchez, Arranz, and Lozano-Durán, 2024).

2.3 Liang-Kleeman Information Flow

The Liang-Kleeman information flow (LKIF) was developed in a series of papers, starting in 2005 with the 2D formulation for deterministic systems (X. San Liang and Kleeman, 2005). Later generalizations formalized and extended the method to multidimensional stochastic systems (X. San Liang, 2016; X. San Liang, 2021). Recent advancements developed the tools to compute the information flow between subsystems (X. San Liang, 2022). This line of work has been applied to real world problems in different disciplines, including applications to the Earth system, such as identifying dynamical dependencies in the climate system (Stéphane Vannitsem, Carlos A. Pires, and David Docquier, 2024; S. Vannitsem, X. S. Liang, and C. A. Pires, 2024), the interaction between ocean and atmosphere (D. Docquier, S. Vannitsem, and Bellucci, 2023; Yang et al., 2024), soil moisture-precipitation interaction (Hagan et al., 2019), the interaction between the land surface and the atmosphere (Shao et al., 2024).

The Liang-Kleeman information flow (LKIF) is reputed to be a theoretically motivated, formalism-driven notion, well grounded on the mathematical basis provided by

the Liouville equation and a hydrodynamics basis (Smirnov, 2022).

> The rate of information flow, or information flow, from a component X_j to another component X_i, is defined as the contribution of entropy from X_j per unit time in increasing the marginal entropy of X_i.

Carlos A. Pires, David Docquier, and Stéphane Vannitsem (2024) developed an extension of the formalism to non-linear systems. In the following we will present the method in the general formulation for d-dimensional systems, which we will refer to as *multivariate* LKIF. This term is used to indicate the study of the information flow between *two variables*, X_i and X_j, in a *multi-dimensional* system. This is different from the *bivariate* approach, as in the multivariate context, the other $d-2$ variables also influence X_i and X_j, thus the information flow between them is different from what we would compute in the bi-dimensional system, where only these two variables are present.

We also present the extension of the LKIF to the study of complex subsystems, and discuss the causal interpretation of the general approach.

2.3.1 Information Flow-based causal analysis

The starting point of the information flow-based causal inference framework is a causal graph, analogue to those we have seen in Chapter 1. The underlying idea of this method is that the problem of causal inference in a system of variables organized in a causal graph can be recast within the framework of dynamical systems, and that the causal inference can be carried forth to that between the coordinated in a dynamical systems. Causality is then identified by studying a physical quantity, the *information flow*, which can easily be cast within the dynamical systems framework. The central physical quantity is Shannon entropy, which is interpreted as self-information. One can then try to find a related quantity that measures the propagation of entropy or uncertainty from one component to another, and interpret is as closely related to causality. We can formalize this in the following definition.

Definition 2.3.1. *In a dynamical system (Ω, Φ_t) on the d-dimensional phase space Ω, where Φ_t may be a continuous-time flow ($t \in \mathbb{R}^+$) or discrete-time mapping ($t \in \mathbb{Z}^+$), the information flow from a component X_j to another component X_i of the variables system $\mathbf{X} = (X_1, ..., X_d)$, written $T_{j\to i}$, is defined as the contribution of entropy from X_j per unit time (or per step) in increasing the marginal entropy of X_i.*

Causality can be defined, at this point, exploiting information flow. The quantitative definition, given in X. San Liang, 2021 follows.

Definition 2.3.2. *X_j is causal to X_i if and only if $T_{j\to i} \neq 0$. The magnitude of the causality from X_j to X_i is measured by $|T_{j\to i}|$.*

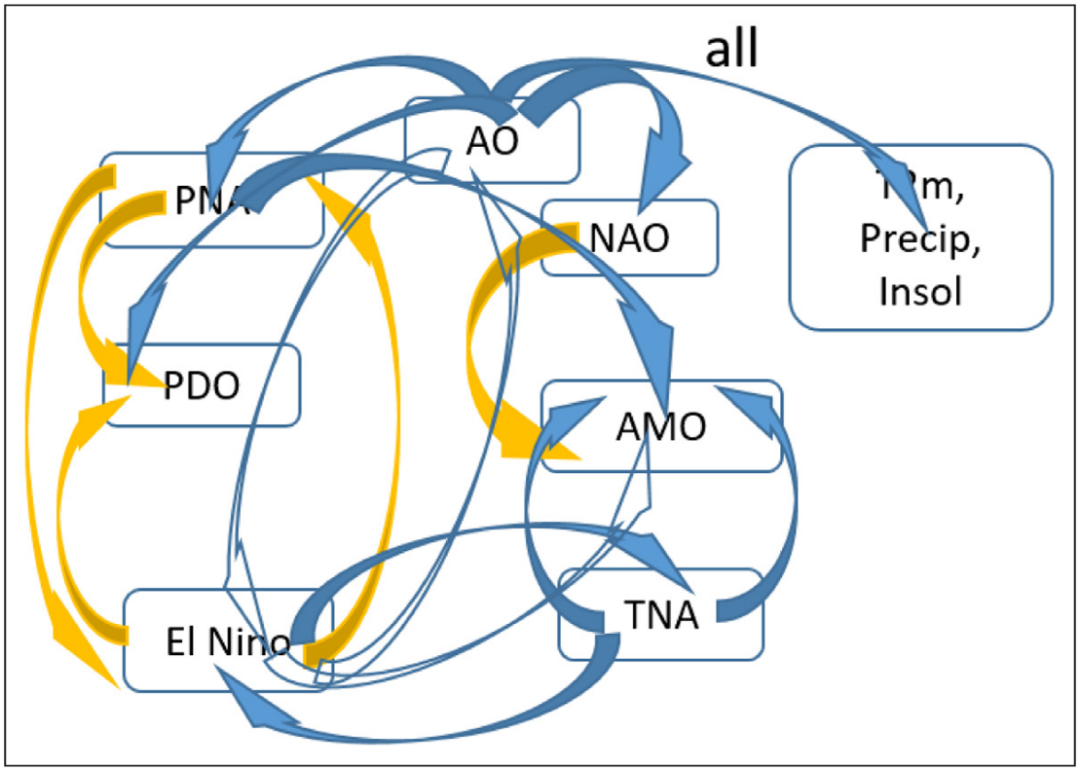


Figure 2.1: Graph reconstruction of the dynamical dependencies among climate indices. Dependencies between different climatic observables at short (blue arrows) and all time scales (yellow arrows). The open blue arrows indicate weak dependencies as compared to the other ones. The label "all" indicates that all local variables are influenced by AO. The observables in the right box are temperature, precipitation and insolation. The meaning of the other observables is reported in the Acronyms page. Courtesy of Stéphane Vannitsem and X. San Liang (2022).

By evaluating the information flow within all the components of a dynamical system, we are thus able to reconstruct the underlying causal graph.

2.3.2 An overview of the theory

In the following we give an overview of the theory of the Liang-Kleeman Information Flow (LKIF) as a causal notion. This method considers the continuous flow case, which implies the differentiability of the vector field that is the object of the analysis that follows. For discrete-time mappings, the information flow is in a more complicated form. Furthermore, it is assumed that the noise is well described by a Wiener process (white noise).

LKIF assumes that the system at study can be approximated by a d-dimensional continuous-time stochastic system. For $\mathbf{X} = (X_1, \ldots, X_d)$ we have

$$\mathrm{d}\mathbf{X} = \mathbf{F}(\mathbf{X}, t)\mathrm{d}t + \mathbf{B}(\mathbf{X}, t)\mathrm{d}\mathbf{W}, \tag{2.19}$$

Where $\mathbf{F} = (F_1, ..., F_d)$ may be arbitrary nonlinear functions of $\mathbf{X}$ and t, $\mathbf{W}$ is a vector of standard Wiener processes, and $\mathbf{B} = (b_{ij})$ is the matrix of perturbation amplitudes, which may also be any functions of $\mathbf{X}$ and t. Assume that $\mathbf{F}$ and $\mathbf{B}$ are both differentiable with respect to $\mathbf{X}$ and t. The following theorem then holds.

Theorem 2.3.1. *For the system 2.19, the rate of information flowing from (X_j) to X_i is*

$$\begin{aligned} T_{j\to i} &= -\mathbb{E}\left[\frac{1}{\rho_i}\int_{\mathbb{R}^{d-2}} \frac{\partial(F_i\rho_{\setminus j})}{\partial x_i} d\mathbf{x}_{\setminus i\setminus j}\right] + \frac{1}{2}\mathbb{E}\left[\frac{1}{\rho_i}\int_{\mathbb{R}^{d-2}} \frac{\partial^2(g_{ii}\rho_{\setminus j})}{\partial x_i^2}\, d\mathbf{x}_{\setminus i\setminus j}\right], \\ &= -\int_{\mathbb{R}^d} \rho_{j|i}(x_j|x_i)\frac{\partial(F_i\rho_{\setminus j})}{\partial x_i}\, d\mathbf{x} + \frac{1}{2}\int_{\mathbb{R}^d} \rho_{j|i}(x_j|x_i)\frac{\partial^2(g_{ii}\rho_{\setminus j})}{\partial x_i^2}\, d\mathbf{x}, \end{aligned} \tag{2.20}$$

where $d\mathbf{x}_{\setminus i}$ means $dx_1 \dots dx_{i-1}dx_{i+1} \dots dx_n$, $\mathbb{E}$ stands for mathematical expectation, $g_{ii} = \sum_{k=1}^{n} b_{ik}b_{ik}$, $\rho_i = \rho_i(x_i)$ is the marginal probability density function (pdf) of X_i, $\rho_{j|i}$ is the pdf of X_j conditioned on X_i, and $\rho_{\setminus j} = \int_{\mathbb{R}} \rho(\mathbf{x})dx_j$. The information flow is measured in nats per unit time.

The bivariate analysis is based on the following corollary.

Corollary 2.3.1. *When $d = 2$,*

$$T_{2\to 1} = -\mathbb{E}\left[\frac{1}{\rho_1}\frac{\partial(F_1\rho_1)}{\partial x_1}\right] + \frac{1}{2}\mathbb{E}\left[\frac{1}{\rho_1}\frac{\partial^2(g_{11}\rho_1)}{\partial x_1^2}\right]. \tag{2.21}$$

From Theorem 2.3.1 one can derive the following property.

Theorem 2.3.2. *If in Definition. 2.3.1 neither F_1 nor g_{11} depend on X_2, then $T_{2\to 1} = 0$.*

Thus, we see that this formulation of information flow puts on rigorous footing an important aspect of causality. If the evolution of an event, say X_1, is independent of another one, X_2, then the causal information from X_2 to X_1 is zero. This is the only quantitatively stated fact about causality, it is also called the *principle of nil causality.*

Further, it has been established that

Theorem 2.3.3. *$T_{2\to 1}$ is invariant under arbitrary nonlinear transformation of $(X_3, X_4, ..., X_d)$.*

Invariance under a change of variables is a highly desired property in physics. It indicates that information flow, being invariant under the representation we choose for the system, has to be an intrinsic property of it. This result will come in handy when we will treat causal graph reconstruction using the information flow.

For linear systems, the information can be greatly simplified,

Theorem 2.3.4. *In Eq. 2.19, if* $\boldsymbol{F}(\boldsymbol{X}) = \boldsymbol{f} + \boldsymbol{A}\boldsymbol{X}$ *and* $\boldsymbol{B}$ *is a constant matrix, then*

$$T_{j\to i} = a_{ij}\frac{\sigma_{ij}}{\sigma_{ii}} \tag{2.22}$$

Where a_{ij} *is the* $(i,j)^{th}$ *entry of* $\mathbf{A}$*, and* σ_{ij} *the population covariance between* X_i *and* X_j*.*

From this expression we can see that the vanishing of σ_{ij} implies $T_{j\to i} = 0$. But we can still have $T_{j\to i} = 0$ even if $\sigma_{ij} \neq 0$. We can thus say that, in the linear sense, causation implies correlation, but correlation does not imply causation. This does not generalize to nonlinear systems.

Exploiting the previous formalism, one can compute the maximum likelihood estimators of the information flow, in the bi-dimensional and in the multi-dimensional cases. We see the relevant calculations in Appendix C, which reveal how the estimates of the LKIF can be quite easily computed from data using sample covariances. The reduction of the problem to the computation of simple statistical objects makes the causal inference process with this method particularly convenient.

Normalization of the information flow

In many problems it is important to be able to compare the magnitudes of the information flow between variables, other than verifying if the causal link exists or not. This calls for the implementation of the normalized information flow, addressed in X. San Liang, 2021. The normalized version of the information flow is

$$\tau_{2\to 1} = \frac{T_{2\to 1}}{Z}, \tag{2.23}$$

where Z is the normalizer. This quantity is defined as

$$Z \equiv \left|\frac{dH_1^*}{dt}\right| + \sum_{j=2}^{d} |T_{j\to 1}| + \left|\frac{dH_1^{\text{noise}}}{dt}\right|. \tag{2.24}$$

The first term, dH_1^*/dt is the information from variable X_1 to itself, which corresponds to a self-loop in the graph. The second element is the contribution from all other variables to X_1, and the remaining one is the effect of noise. The normalized information flow from X_2 to X_1 is:

$$\tau_{12} = \frac{T_{2\to 1}}{Z}. \tag{2.25}$$

$\tau_{2\to 1}$ lies on $[-1, 1]$ and is often expressed in percentages. So, when $|\tau_{2\to 1}|$ is 100%, X_2 has the maximal impact on X_1.

2.3.3 Information flow between complex subsystems

This section addresses the Information Flow formalism applied to the study of causal relations between sub-sections of a system, a problem that has been addressed in X. San Liang (2022). The idea is to consider a system divided in multivariate sub-components, and measure the information flow between the sub-components. Systems of this type often emerge in electrical engineering or in climate, where different components, for example the ocean and the atmosphere, evolve on different time scales. Indeed, we will use this approach to study the ocea-atmosphere coupling in Chapter 4.

With respect to the multivariate formulation, where we compute the information flow between the single variables of the multivariate system, this *interregional* approach allows to give a "bulk" description of the interaction between subsystems, to capture collective dynamics and keep the method computationally feasible.

Consider an n-dimensional dynamical system

$$A: \quad \begin{cases} \dfrac{dx_1}{dt} = F_1(x_1, x_2, \ldots, x_n; t) + \displaystyle\sum_{k=1}^{m} b_{1,k}(x_1, x_2, \ldots, x_n; t)\, \dot{\omega}_k, \\ \quad \vdots \\ \dfrac{dx_r}{dt} = F_r(x_1, x_2, \ldots, x_n; t) + \displaystyle\sum_{k=1}^{m} b_{r,k}(x_1, x_2, \ldots, x_n; t)\, \dot{\omega}_k, \end{cases} \tag{2.26}$$

$$B: \quad \begin{cases} \dfrac{dx_{r+1}}{dt} = F_{r+1}(x_1, x_2, \ldots, x_n; t) + \displaystyle\sum_{k=1}^{m} b_{r+1,k}(x_1, x_2, \ldots, x_n; t)\, \dot{\omega}_k, \\ \quad \vdots \\ \dfrac{dx_s}{dt} = F_s(x_1, x_2, \ldots, x_n; t) + \displaystyle\sum_{k=1}^{m} b_{s,k}(x_1, x_2, \ldots, x_n; t)\, \dot{\omega}_k, \end{cases} \tag{2.27}$$

$$\begin{cases} \dfrac{dx_{s+1}}{dt} = F_{s+1}(x_1, x_2, \ldots, x_n; t) + \displaystyle\sum_{k=1}^{m} b_{s+1,k}(x_1, x_2, \ldots, x_n; t)\, \dot{\omega}_k, \\ \quad \vdots \\ \dfrac{dx_n}{dt} = F_n(x_1, x_2, \ldots, x_n; t) + \displaystyle\sum_{k=1}^{m} b_{n,k}(x_1, x_2, \ldots, x_n; t)\, \dot{\omega}_k, \end{cases} \tag{2.28}$$

where $\mathbf{x} \in \mathbb{R}^n$ denotes the vector of state variables $(x_1, x_2, \ldots, x_n)$, $\mathbf{F} = (F_1, \ldots, F_n)$ are differentiable functions of $\mathbf{x}$ and time t, and $\mathbf{w}$ is a vector of m independent standard Wiener processes, and $\mathbf{B} = (b_{ij})$ is an $n \times m$ matrix of stochastic perturbation amplitudes. From the components $(x_1, \ldots, x_n)$, we separate out two sets, $(x_1, \ldots, x_r)$ and $(x_{r+1}, \ldots, x_s)$, and denote them as $\mathbf{x}_1, ..., r$ and $\mathbf{x}_{r+1}, ..., s$, respectively. The remaining

components $(x_{s+1}, \ldots, x_n)$ are denoted as $\mathbf{x}_{s+1,\ldots,n}$. The subsystems so distinguished are referred to as A and B.

The results obtained by extending the information flow formalism to subsystems A and B are summarized by the following theorem.

Theorem 2.3.5 (Information Flow for subsystems)**.** *For the dynamical system Equations 2.26 - 2.28, if the probability density function (pdf) of* $\mathbf{x}$ *is compactly supported, then the information flow from* $\mathbf{x}_{1,\ldots,r}$ *to* $\mathbf{x}_{r+1,\ldots,s}$ *and that from* $\mathbf{x}_{r+1,\ldots,s}$ *to* $\mathbf{x}_{r+1,\ldots,s}$ *are, in nats per unit time, respectively,*

$$\begin{aligned} T_{A\to B} =& - E\left[\sum_{i=r+1}^{s} \frac{1}{\rho_{r+1,\ldots,n}} \int_{\mathbb{R}^{n-s}} \frac{\partial F_i \rho_{r+1,\ldots,n}}{\partial x_i}\, d\mathbf{x}_{s+1,\ldots,n}\right] \\ &+ \frac{1}{2} E\left[\sum_{i=r+1}^{s}\sum_{j=r+1}^{s} \frac{1}{\rho_{r+1,\ldots,s}} \int_{\mathbb{R}^{n-s}} \frac{\partial^2 g_{ij}\rho_{r+1,\ldots,n}}{\partial x_i \partial x_j}\, d\mathbf{x}_{s+1,\ldots,n}\right], \end{aligned} \tag{2.29}$$

$$\begin{aligned} T_{B\to A} =& - E\left[\sum_{i=1}^{r} \frac{1}{\rho_{1,\ldots,r}} \int_{\mathbb{R}^{n-s}} \frac{\partial F_i \rho_{1,\ldots,r,s+1,\ldots,n}}{\partial x_i}\, d\mathbf{x}_{s+1,\ldots,n}\right] \\ &+ \frac{1}{2} E\left[\sum_{i=1}^{r}\sum_{j=1}^{r} \frac{1}{\rho_{1,\ldots,r}} \int_{\mathbb{R}^{n-s}} \frac{\partial^2 g_{ij}\rho_{1,\ldots,r,s+1,\ldots n}}{\partial x_i \partial x_j}\, d\mathbf{x}_{s+1,\ldots,n}\right], \end{aligned} \tag{2.30}$$

where $g_{ij} = \sum_{k=1}^{m} b_{ik} b_{jk}$*, and* E *signifies mathematical expectation.*

When $r = 1$, $s = n = 2$, Equation 2.30 reduces to

$$T_{B\to A} = -E\left[\frac{1}{\rho_1}\frac{\partial F_1 \rho_1}{\partial x_1}\right] + \frac{1}{2} E\left[\frac{1}{\rho_1}\frac{\partial^2 g_{11}\rho_1}{\partial x_1^2}\right], \tag{2.31}$$

which is precisely the expression for information transfer between two variables previously obtained in X. San Liang, 2008 in the context of two-dimensional stochastic dynamical systems. The same holds for Equation 2.29. The maximum likelihood estimators have also been obtained under a Gaussian assumption in X. San Liang (2022). We will see an application of this method to the ocean-atmosphere interaction in the MAOOAM model in Chapter 4.

2.4 Other methods

The application of causal methods to disentangle directed dynamical relations is a growing research topic. We have studied methods that assume that the interactions in the system arise from an underlying stochastic or Markov process. Nonlinear state-space

methods assume that interactions occur in an underlying dynamical system, and attempt to uncover causal relationships using tools rooted in *dynamical systems theory*. These methods have better results when it can be shown that a *deterministic nonlinear attractor* can be recovered from the data (Jakob Runge, Bathiany, et al., 2019). An example of state-space method is the *Convergent Cross Mapping* (CCM). A data-driven method with strong dynamical foundations is the *Empirical Model Reduction*. Other related methods are collected in Jakob Runge, Bathiany, et al. (2019).

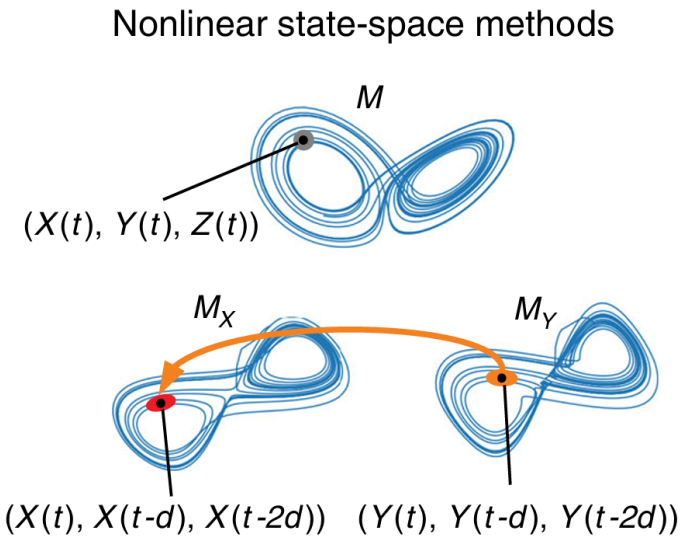


Figure 2.2: The convergent cross-mapping (CCM), illustrated for the chaotic Lorenz system. CCM reconstructs the variables' state spaces (M_X, M_Y) using *time-lagged coordinate embedding* which lays out the conditions under which a chaotic dynamical system can be reconstructed from a sequence of observations of the system's state, creating a "shadow" of its true dynamics in the form of a topological diffeomorphism. The method concludes that causality exists in the $X \to Y$ direction if points on M_X can be predicted using nearest neighbors in M_Y (orange ellipse) and the prediction improves the more points on the attractor are sampled. Courtesy of Jakob Runge, Bathiany, et al. (2019).

Convergent Cross Mapping

The Convergent Cross Mapping (CCM) method, developed in Sugihara et al. (2012), is a nonlinear state-space method that makes the assumption of an underlying (reconstructable) deterministic nonlinear dynamical system to the dynamics. CCM looks for the signature of a time-series variable X in Y's time series by seeing whether there is a correspondence between the points in the attractor manifold built from Y, M_Y, and points in the X manifold, M_X, where these two manifolds are constructed from lagged coordinates of the time-series variables Y and X. Essentially, the idea is to see whether

the time indices of nearby points on the Y manifold can be used to identify nearby points on M_X. If so, then one can use Y to estimate X and vice versa.

> CCM measures if points on X's state space, M_X, can be predicted using nearest neighbors in Y's state space, M_Y, and the prediction improves with more sampled points. If this happens, we can say that X causes Y.

when one variable X is a stochastic driver of a population variable Y, information about the states of X can be recovered from Y, but not vice versa. The CCM method is depicted in Figure 2.2, as a nonlinear state space method. The *delay coordinate embedding* in the caption is a form of state space reconstruction, where one dimensional data, an observation of the system state, is reconstructed in a higher dimensional space using lagged observations. This means that each value of the time series, $X(t)$, is plotted against its value at $X(t-\tau), X(t-2\tau), ...$ with τ the lag, in a higher dimensional embedding space. This idea relies on Taken's theorem (Takens, 1981)

This method has been applied in the context of climate science, to analyze the same model that is the object of study in this work, in S. Vannitsem and Ekelmans (2018).

Empirical Model Reduction

The Empirical Model Reduction (EMR) methodology was introduced in S. Kravtsov, D. Kondrashov, and M. Ghil (2005) to reconstruct nonlinear models from a set of time-series data, either observational or obtained from large-model simulations. The underlying idea is to retrieve the simplest possible dynamical model from data, without losing the complexity of the original dynamical system.

EMR is an implementation of the Mori-Zwanzig formalism of statistical physics for effectively reducing a complex system to its dominant slow variables (Dmitri Kondrashov, Chekroun, and Michael Ghil, 2015). It has been applied with success to modeling and predicting nonlinear climate processes. For instance, it has been used to reconstruct and forecast atmospheric flow regimes and to develop reduced models for phenomena like ENSO and air-sea interaction (Sergey Kravtsov, D. Kondrashov, and Michael Ghil, 2010).

The EMR technique iteratively models residual errors through additional regression levels until the remaining noise is statistically white, accounting also for serial correlations. It uses principal component regression and partial least squares to reduce the dimensionality of the model underlying the original data, yielding a robust, smaller-dimensional model.

In the EMR, the evolution of the time series X_i, represented by $\mathrm{d}X_i$, is modeled by linear and quadratic contributions of the variables provided in the time series, including

a *noise residue*. The EMR method can be formulated theoretically, in continuous time, as a stochastic differential equation (SDE):

$$dZ = \left(\sum_k \alpha_k Z_k + \sum_k \sum_j \beta_{jk} Z_j Z_k \right) dt + \sigma \, dW, \tag{2.32}$$

here $Z = Z(t)$ is the continuous-time stochastic process that gives rise to a particular time serie X_j and $W = W(t)$ is a Wiener process. The EMR method in discrete time can be written as:

$$\Delta x_i^{(n)} = \left(N_{ijk} x_j^{(n)} x_k^{(n)} + L_{ij} x_j^{(n)} + F_i \right) \Delta t + r_{0,i}^{(n)} \Delta t, \tag{2.33}$$

$$\Delta r_i^{(n)} = L_{ij}^{(1)} [\mathbf{X}; \mathbf{r}]_j \Delta t + r_{1,i}^{(n)} \Delta t, \tag{2.34}$$

$$\vdots \tag{2.35}$$

$$\Delta r_{L-1,i}^{(n)} = L_{ij}^{(L)} [\mathbf{X}; \mathbf{r}; \mathbf{r}_1; \cdots; \mathbf{r}_{L-1}]_j \Delta t + \Delta \xi_i(t). \tag{2.36}$$

Here $\mathbf{X}$ is an M-vector with components $\{x_i : i = 1, 2, \ldots, M\}$, with $\mathbf{X}$ observed at $n = 1, 2, \ldots, N$ equidistant points in time, and

$$\Delta x_i^{(n)} = x_i^{(n+1)} - x_i^{(n)}. \tag{2.37}$$

Equation 2.33 is the main level of this recursive scheme and the only one that includes explicitly a nonlinearity. The levels $\ell = 1, 2, \ldots, L$ of Eqs. 2.34-2.36 indicate the iterative procedure that removes the non-white part of the noise from the residues $r_{0,i}^{(n)}$. This recursion links the EMR to the Mori–Zwanzig principle for decoupling leading, slow variables from the noisier, fast ones (Dmitri Kondrashov, Chekroun, and Michael Ghil, 2015).

The EMR has been proposed as a causal method (O. de Viron et al., personal communication, July 2024), where its potential stems from the underlying dynamic foundation of the model. By analyzing the structure of the regression coefficients - including feedback terms - one can infer both direct and indirect causal relationships, accounting for linear and nonlinear effects, and assess how changes in one part of the system propagate through its dynamics.

EMR fits a nonlinear model predicting the rate of change of X ($\mathrm{d}X/\mathrm{d}t$) as a function of Y. The method is successful if the coefficient for Y in X's rate of change is significantly nonzero.

Method	Core idea	Structure	Key Ref.
Granger Causality	X Granger-causes Y if X contains some information relevant for the prediction of Y that is not contained in Y's past.	Linear, with extensions	Granger, 1969
Transfer Entropy	$TE_{X\to Y}$ is the amount of uncertainty reduced in future values of Y by knowing the past values of X, given past values of Y.	Model free	Schreiber, 2000; J. Lizier et al., 2011
Liang-Kleeman Information Flow	$LKIF_{X\to Y}$ is the contribution of entropy from X per unit time in increasing the marginal entropy of Y.	Linear	X. San Liang, 2021
Convergent Cross Mapping	Measures if points on X's state space, M_X, can be predicted using nearest neighbors in Y's state space, M_Y, and the prediction improves with more sampled points.	Nonlinear	Sugihara et al., 2012
Empirical Model Reduction	Fits a nonlinear model predicting the rate of change of X ($\mathrm{d}X/\mathrm{d}t$) as a function of Y. The method is successful if the coefficient for Y is significantly nonzero.	Nonlinear	Dmitri Kondrashov, Chekroun, and Michael Ghil, 2015

Table 2.1: A summary of the causal methods described in this chapter that take time series data as input.

2.5 Reconciling different causal methods: which perspectives?

In the previous sections we have seen methods used in causal inference, rooted in different theoretical grounds. The fact that these methods entail the same notion of causality, or that they have any causal meaning at all, is not universally agreed upon.

Transfer entropy, for instance, is a measure of information transfer. *Is information transfer akin to causal effect? If not, what is the distinction between them?* J. T. Lizier and M. Prokopenko, 2010. We have understood that causality is a slippery concept, even more so when the aim is to infer it from observations. The general knowledge seems to agree that the only operational definition that is able to express the true idea of causality is given by Pearl's interventional approach.

In information-theoretic approaches no intervention is performed on the system, instead, they try to address the problem leveraging the fact that if there is a causal link

between two variables, information is exchanged between the two. Thus, we could hope to approximate causal relationships by carefully disentangling the transfer of information between sources and sinks of a system. Several authors have reflected on this matter and on what kind of causal content data-driven methods are able to retrieve, if any. Having already hinted at some of these reflections, we expand on the topic in the coming paragraphs.

2.5.1 Transfer Entropy and Causality

The definition of causal relationship and of information transfer, and even more so the difference between the two, is subtle. This can be inferred from the confusion that spread about the interpretation of the Transfer Entropy. Exemplarily, the same quantity has been interpreted as "information flow" (Bossomaier et al., 2016), "information transfer" (Schreiber, 2000) (J. T. Lizier and M. Prokopenko, 2010), "information transport" (Schreiber, 2000), or again, "directed statistical coherence" (Mikhail Prokopenko and Joseph T. Lizier, 2014). This short but eloquent list highlights a broader problem in the causal community. The proliferation of different concepts, newly introduced often without proper linkage to previous work, and the interpretation difficulties that hereby derive.

We follow the clarification effort carried forth in J. T. Lizier and M. Prokopenko (2010), which explicitly researches the connection between the concepts of causality and of transfer entropy. The main point made is that causal effect is an intrinsic property of a system. Since often we do not have previous knowledge about this, we have to infer it through suitable quantities.

Causal effects can be estimated through the *information flow* (Ay and Polani, 2008), which should be used as a primary tool (where possible) to establish the presence of - and quantify - causal relationships. This quantity (which is different from the LKIF) essentially derives from Pearl's intervention-based definition of causality. It results from considerations on how to measure causal information flow via interventional conditional probabilities. For instance, an interventional conditional probability $p(a|\hat{s})$ considers the distribution of a resulting from *imposing* the value of $\hat{s}$. *Imposing* means intervening intervening in the system to set the value of the imposed variable, and it is very similar to what the *do*-operator does in Pearl's framework.

On the other hand, transfer entropy represents the quantity that one *measures* when looking for *information transfer*, that is, the amount of information that a source variable adds to the next state of a destination variable[1]. In this sense, TE remains a measure of observed (conditional) correlation rather than direct causal effect. In a sense, the transfer entropy is focused on *computation*, it does not have direct access to the physically

[1]A thermodynamic interpretation of transfer entropy, as being proportional to external entropy production and connected to the concept of irreversibility is given in Faes, Nollo, and Porta (2012) and Mikhail Prokopenko, Joseph T. Lizier, and Price (2013).

encoded causal dependencies in a system. Therefore, if we want to measure causal effect, we should prefer to use information flow.

There are situations where measuring the intervention-derived information flow is not possible, e.g. where one has no ability to intervene in the system or no knowledge of the underlying dynamics. Then, under certain parameter settings, i.e. setting the history length of the process in consideration to include only the causal contributors from the destination's past, the complete transfer entropy *converges* with the information flow, and may still provide a reasonable inference where these conditions are approximated.

Thus, the quantity measured with transfer entropy might or might not be causal depending on how we set the computation. J. T. Lizier and M. Prokopenko, 2010 state that the distinction between this quantity and causal effect has often been neglected, with the two sometimes problematically directly equated, as in the case of the Liang-Kleeman Information Flow.

2.5.2 Towards causal inference

We have seen that Transfer Entropy, under certain conditions, can give causal information about a system at study. The method has been popularly used in applications even when those conditions are not fully respected, as sometimes the focus of research is not to fulfill a definition, but to *say something more* or *better* than other methods - for example to go beyond the information retrieved from correlation studies. As we pose the matter in these terms, the road is open to improvement, in multiple ways. On one side there is space for **technical improvement**, as a part of the limitations are on the implementation side. For instance, efforts in this sense have brought to the generalization of early methods to *nonlinear* or *model-free* versions, and to *multivariate* extensions, which are better suited for nonlinear and complex systems.

In another sense, considerable scope remains for investigating **whether different methods share common roots, how they differ from one another, and whether these differences can be leveraged to examine distinct aspects or types of systems**. The panorama of fields in which causal inference is pursued is very diverse in terms of disciplines and applications. Through the years, this has opened the way to a large proliferation of methods, and consequently to the development of specific causal jargons. This results in a restricted access to - and understanding of - the results produced in the various communities. It is likely that it might be considered easier to implement a new method from scratch rather than going through the painful process of translating a specific method conceived for applications in a different context. These problems of language and of communication between fields appear quite clearly in the context of information theory, as we have argued in this work, and as clearly stated in J. T. Lizier and M. Prokopenko, 2010.

An important unification attempt has been pursued by the mathematised causality framework of Judea Pearl. However, despite the development of practical algorithms for

causal inference from time series in the context of the formal framework, the gauntlet of having to respect the many assumptions required to correctly apply these methods still makes it a daunting task. Indeed, the concrete applications of SCMs to physical sciences are still quite rare. Nevertheless, we argue that even when not directly applicable, Pearl's framework holds an important position in the domain of causal inference.

2.5.3 The legacy of Pearl's framework

Pearl's framework of causality has largely influenced the whole panorama of causal inference, including those methods that are do not directly refer to the intervention-based definition. One example of this is the spread of the use of *graphical representation* of causal problems. Causal diagrams are now unconsciously employed by researchers belonging to any causality school of thought.

Additionally, this is (partly) true for the awareness on the assumptions for causal inference too, which we have outlined in Chapter 1. Any paper in the field of causal inference, no matter the methods it uses, will warn its readers about the possible presence of *confounders*, which were correctly defined only in Pearl's framework, or about the dangers of violating the *causal sufficiency* requirement. One interesting example of this natural integration is given by the evolution of Clive Granger's work after the first implementation in Granger (1969). For instance, some works dedicate to equipping the original idea with graphical models (graphical Granger models) Arnold, Liu, and Abe, 2007, or to the more arguable task of reconciling Granger's approach with Pearl's framework White, Chalak, and Lu, 2011.

Another noteworthy development is the pursuit of novel methods to analyze processes and time series, while remaining consistent with Pearl's framework, widely regarded as the most convincing definition of causality. In this respect, we mention the line of research on causal effect (Ay and Polani, 2008). In Baldovin, Cecconi, and Vulpiani (2020) it is argued that a intervention-like quantity can be derived through the fluctuation-response formalism, with the physical response function quantifying the interventional causal link. Tian and Judea Pearl (2001) proposes a new method of discovering causal relations in data, based on the detection and interpretation of local spontaneous changes in the environment, exploiting dynamic changes in the statistical data distribution.

We recall the line of research on causal inference algorithms developed as direct and applied consequence of Pearl's research (Spirtes, C. Glymour, and Scheines, 2001; Peters, Janzing, and Schlkopf, 2017). These algorithms open the possibility of inferring causal information from time series, provided the assumptions seen in Section 1.3.2 hold. In Dechter, Geffner, and Halpern (2010) Spirtes, Glymour, Scheines and Tillman, who have been at the forefront of the causal framework formalization, summarize the content of this paragraph:

> *The discovery of d-separation, and the development of several related no-*

> *tions, has made possible principled search for causal relations from observational and quasi-experimental data in a host of disciplines. New insights, algorithms and applications have appeared almost every year since 1990, and they continue. We are seeing a revolution in understanding of what is and is not possible to learn from data, but the insights and methods have seeped into statistics and applied science only slowly. We hope that pace will quicken.*

2.5.4 Comparing causal methods

So far we have largely discussed different methods and the merit of their causal interpretation. Assuming these methods to correctly estimate some aspect of causality, we would expect them to output equal or similar results when applied to the same datasets. This is often not the case. Despite the research in this sense being still quite limited, the works that compare more than one method have to conclude that different methods give different results, both in terms of which causal links exist or do not exist and in terms of the magnitude of the retrieved links. In part, this can be attributed to the fact that methods are not universal, e.g. they might be linear or nonlinear, bivariate or multivariate. Therefore, they might perform better on certain types of data: "*It is important to choose the right method for a particular type of data. [...] blind application of any causality test easily leads to incorrect conclusions.*" (Krakovská et al., 2018). Some works that provide comparison of methods in applied cases are Krakovská et al. (2018), D. Docquier, S. Vannitsem, Ragone, et al. (2022), Jakob Runge, Bathiany, et al. (2019), Paluš et al. (2018), and Shao et al. (2024) and (O. de Viron et al., pers. comm., July 2024). The contrasting results obtained in this last work are shown in Figure 2.3.

Smirnov (2022) gives a theoretical perspective on the comparison of causal methods, designing a generative formalism that allows to formulate in a unified manner and interrelate a variety of causal inference methods used in time series analysis. A general notion of dynamical causal effect is implemented with this formalism, with causality intended in the interventional way. Specific *causal quantifiers*, which quantitatively estimate the causal relation between two variables in a system can thus be derived as its realizations. In this contex, the author provides an example by studying the generalization of Transfer Entropy and the Liang-Kleeman information flow, which proves necessary in view of the unclear interpretation of many popular causal methods. It has been noted that different approaches are prone to estimating different magnitudes for the causal couplings in the same system, to the point that "*any causal coupling in a complex system may be stronger than couplings in other directions according to one quantifier, and weaker according to another quantifier. Which of the two information flows is more appropriate and where?*" (Smirnov, 2022). The Transfer Entropy and the Liang Kleeman Information Flow are shown to be related as opposite limit cases in a specific dynamical causal effect family. However, their evaluation of causal links magnitudes are shown to differ from each other essentially, and to describe different manifestations of a directional coupling in dynamics.

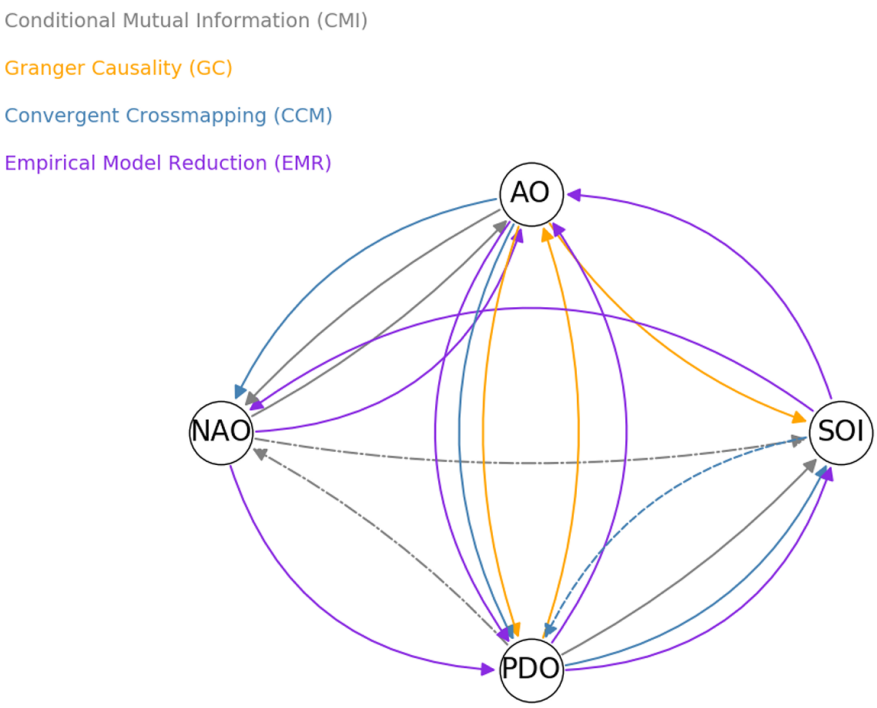


Figure 2.3: Comparison of the causal links obtained with four distinct methods: Conditional Mutual Information (CMI), Granger Causality (GC), Convergent Cross Mapping (CCM) and Empirical Model Reduction (EMR), in their linear versions. The meaning of the acronyms is reported in the Acronyms page. Courtesy of O. de Viron, D. Delforge and M. Ghil, pers. comm., July 2024.

For instance, the general framework is used to analyze the evolution of two overdamped oscillators X and Y, a simple model for irregular processes. The evaluated coupling $Y \to X$ is arbitrarily strong when evaluated with TE rate and zero when quantified with the LKIF. On the other hand, in a wide range of situations the two methods result in similar values, proportional to each other with a moderate factor. We will see the comparison between these two methods when applied to an ocean-atmosphere coupled model in Chapter 4.

A note of warning on the application of data-driven causal methods

The uncertainties in the causal interpretation across different methodologies, among other issues, imply that the application of information-theoretic methods for causal inference remains an evolving field. While interesting results have been achieved, it is debatable whether the quantitative outputs can be fully trusted without a previous intuition of the underlying physical mechanisms, a well-founded hypothesis regarding the causally relevant variables, and a critical perspective on the magnitude of the inferred

causal effects. We will discuss this in practice when studying the causal coupling in MAOOAM, in Chapter 4.

Key concepts

- Popular methods for causal inference from time series are *Granger causality* (GC), *Transfer Entropy* (TE), the *Liang Kleeman Information Flow* (LKIF). The *Convergent Cross mapping* (CCM) and *Empirical Model Reduction* (EMR) are also promising data-driven methods rooted in dynamical systems theory.

- Granger causality relies on the idea that X causes Y if X can be shown to increase the predictability of Y. Initially developed as a linear method, nonlinear extensions have been developed in the last years.

- Transfer Entropy is a model-free method developed in the context of information theory. One can say that the TE from Y to X is the degree to which Y disambiguates the future of X beyond the degree to which X already informs its future. Transfer Entropy is equivalent to Granger causality for Gaussian variables, and is usually understood as a closely related evolution of it. The method is multivariate and features an interregional implementation.

- The Liang-Kleeman Information Flow is a linear method rooted in stochastic dynamical systems. It retrieves the asymmetric information flow from a variable X to another variable Y as the contribution of entropy from X per unit time in increasing the marginal entropy of Y. The method is multivariate and features an interregional implementation.

- Other methods for causal inference from time series are the Convergent Cross Mapping (CCM), which assumes that a *deterministic nonlinear attractor* underlies the time-series data and searches for causality in the increment of predictability that knowing points on the X manifold makes us gain on the points on the manifold of Y. Another method is the EMR, a data-driven model reduction technique, which can be used to infer nonlinear causal relationships between model variables.

- Different methods do not always agree among each other in the results they obtain, particularly in terms of estimated causal effects and reconstructed causal graph.

- There have been several attempts to reconcile the causal meaning of different approaches. Granger causality has been proven to be equivalent to Transfer Entropy up to a multiplicative factor 2 for Gaussian variables.

- Methods in this chapter do not leverage the interventional definition of causality, but they do significantly better than correlation methods in retrieving directed

relations dependencies in data. Moreover, they incorporate important aspects of the formal causality framework, despite being rooted in different philosophies.

Chapter 3

Disentangling the direction of the Ocean-Atmosphere coupling

In the previous Chapter we described the general framework and the methods of causality, and in particular the problem of causal inference from time series. We have seen that this effort has been applied in Earth system and climate sciences. We now expand on this topic and deepen specific features of climate-related causal research context.

3.1 Outlook on causality in the climate sciences

The climate system is a forced, dissipative, nonlinear, complex and heterogeneous system that is out of thermodynamic equilibrium (Michael Ghil and Lucarini, 2020). Natural variability emerges in the system on many scales of motion, in time as well as space, which also evolves under the action of external forcings, natural as well as anthropogenic. These features make the study of the climate system an extraordinarily complex and fascinating task.

Many questions about the complex interactions in the climate system, its evolution over time, or the attribution of events to climate change are inherently causal questions, for instance

- What drives the observed variability in the climate system, and how do large-scale circulations affect local weather patterns?
- Can the global warming trend be attributed to anthropogenic forcing?
- Can specific extreme weather events be attributed to climate change?
- How do key climate components interact to shape regional or systemic behavior?

The traditional approaches used to answer these questions are often rooted in regression, also called *optimal fingerprinting* (Alexis Hannart, Ribes, and Naveau, 2014), and, more recently in machine learning. The use of causal methods to tackle these questions has diffused only recently, but it has become quite popular in the last years. The idea is that better results could be achieved once the listed questions about the functioning of the Earth system, being causal, are tackled inside the framework of causality.

In this domain, causality has been employed to better understand and identify physical processes at different scales, large scale atmospheric teleconnections, to evaluate and improve the development of climate models and for causally robust forecasting. The social aspect of climate studies also calls for a causal understanding, for example regarding the attribution of climate change and extreme events through counterfactuals, or in the context of policy making, to determine effective land-use strategies Jakob Runge, Gerhardus, et al., 2023. Algorithms derived from the formal framework of causality have recently been applied to Earth system sciences, mainly focusing on the study of the climate system (Kretschmer et al., 2016; Jakob Runge, Petoukhov, and Kurths, 2014; Jakob Runge, Nowack, et al., 2019; Ebert-Uphoff and Deng, 2012; Melkas et al., 2021). Methods rooted in Granger's idea of predictive causality or derived from information theory and dynamical systems theory have also found large application in this field (McGraw and Barnes, 2018; Papagiannopoulou et al., 2017; Silva et al., 2021), and in the context of climate change attribution (Risser, Ombadi, and Wehner, 2024). The Convergent Cross Mapping method has been applied to climate science in Nes et al. (2015) and in S. Vannitsem and Ekelmans (2018), for instance.

There are still many open methodological problems in the development and application of causal methods for the Earth system. One goal is to improve the handling of uncertain expert knowledge and the spatio-temporal complexity of the underlying dynamical phenomena, and to develop more robust and statistically efficient algorithms (Jakob Runge, Gerhardus, et al., 2023).

A lively line of research currently aims at the integration of causality in machine learning. In this context, this cross-fertilization could lead to incorporating large language models into the framework of causal hypothesis testing - combined with expert knowledge and data - could yield deeper insights into the causal structure of the climate system, improve data-driven forecasting, development of more powerful methods for process understanding, causal attribution of extreme events, and improvement of the physics embedded in climate models (Jakob Runge, Gerhardus, et al., 2023; Cohrs et al., 2025).

An important challenge for the causal endeavor in Earth system science, also seen in a more general scope in the previous chapter, is the broad language gap between the methodological and domain science communities with their individual scientific language. A complete overview of the open problems in causal inference for Earth system science is given in Jakob Runge, Bathiany, et al. (2019).

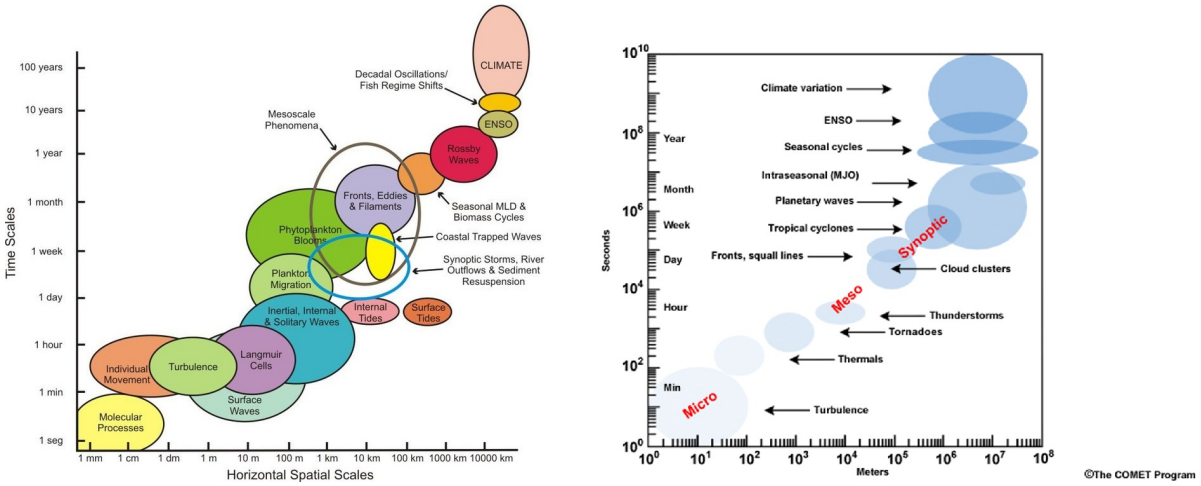


Figure 3.1: Chelton diagram of ocean variability and COMET diagram of atmospheric variability. Image on the right: courtesy of Michael Ghil and Lucarini (2020)

3.2 Interactions between ocean and atmosphere: the direction of the coupling

Although significant progress has been achieved (Michael Ghil, 2019), understanding and predicting the dynamics of the climate system remains challenging. In addition to the difficulties inherent in any nonlinear, complex system out of equilibrium, several specific obstacles persist. These include the presence of well-defined subsystems - the atmosphere, the oceans, and the cryosphere - each characterized by distinct physical and chemical properties and operating on widely differing spatial and temporal scales; the complex processes that couple these subsystems; continuously varying natural and anthropogenic forcings; the absence of scale separation among processes; the lack of detailed, homogeneous, high-resolution, long-term observations of climatic fields; and the fact that we have only one realization of the processes driving climate evolution (Michael Ghil and Lucarini, 2020).

The multiple spatial and temporal scales in the climate system necessitate both model reduction methods, which retain only the essential physical features, and complex parameterizations of subgrid-scale processes in numerical models (Michael Ghil and Lucarini, 2020). The multiscale nature of climate is often summarized with *Stommel diagrams*, which capture the spatial-temporal variability of a climate subsystem by associating it with phenomenologically well-defined dynamical features. Figure 3.1 presents two qualitative Stommel diagrams for the ocean and the atmosphere.

Different classes of numerical models - each based on distinct dynamical balances - can explicitly simulate only one or a few dynamical ranges, which together span approximately 15 orders of magnitude in space and time. This limitation led to the development

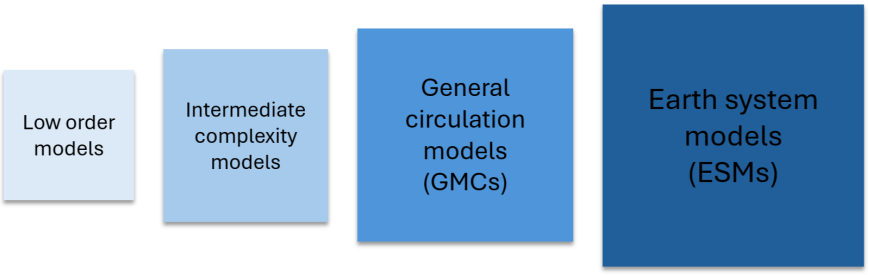


Figure 3.2: *Hierarchy of climate models*, ranging from low-order to comprehensive Earth System Models (ESMs). Lower-order and intermediate-complexity models isolate key dynamical processes at minimal computational cost, while global climate models (GCMs) simulate a broad array of physical processes at high spatial and temporal resolutions. Modern ESMs, building on advances in fundamental scientific knowledge as well as in computational and data storage capabilities, incorporate an increasing number of physical, chemical, and biological processes (Michael Ghil and Lucarini, 2020).

of the so-called *hierarchy of models* (Dijkstra and Michael Ghil, 2005; Michael Ghil and Lucarini, 2020). The hierarchy begins with the simplest models, such as low-order models that capture essential dynamics through differential equations. This approach, pioneered by Barry Saltzman (Saltzman, 1962) and Edward Lorenz (Lorenz, 1963), is reviewed in Stéphane Vannitsem, Jonathan Demaeyer, et al., 2015. Global Climate Models (GCMs) and Earth System Models (ESMs), illustrated in Figure 3.2, aim to incorporate as many physical processes as possible, with ESMs continually extending the spatial and temporal scales. There is also an increasing effort to achieve "seamless prediction" across scales (Meehl et al., 2021).

The oceans and the atmosphere dominate the Earth's hydroclimate system and interact across different scales through various physical processes. The atmosphere influences ocean dynamics by providing momentum, heat, and moisture fluxes, while the ocean regulates weather and climate by supplying moisture and heat to the atmosphere. Understanding these ocean–atmosphere modes is crucial for advancing our comprehension of the climate system. The atmosphere, ocean, and their coupled system exhibit many modes of variability, the most well-known being the El Niño Southern Oscillation (ENSO).

Despite growing knowledge of these variability modes, a comprehensive understanding remains elusive. In particular, the nature of the coupling between the ocean and the atmosphere at different latitudes is still unresolved. It remains unclear whether one medium predominantly drives the other or if both act as a coupled mode. In other words: *does atmospheric weather drive the ocean, does the ocean steer the atmosphere, or does a coupled mode of variability govern the system?* This issue can be illustrated with a metaphor: imagine the slower, heavier ocean as a dog and the faster, lighter atmosphere as its tail. The question then becomes whether the dog wags its tail, the tail

drives the dog, or if the system is better represented by the dog biting its tail.

Aim of this causal research

Is it weather noise driving a passive ocean, the slower ocean steering the atmosphere, or does a mode of coupled variability exist? We investigate the nature of the ocean-atmosphere variability through the use of causal methods.

If atmospheric forcing is dominant, wind stress over the ocean should be the primary driver. Conversely, if the ocean exerts a stronger influence, its thermal fluxes - through the release of heat and water vapor - would significantly affect atmospheric temperature and composition. Alternatively, both media may influence each other through these and other mechanisms arising from their coupled interaction.

This topic has been extensively studied through observations and modeling. In this work, we contribute to the understanding of ocean-atmosphere coupling by introducing a causal inference perspective. Specifically, we apply the Liang-Kleeman and Transfer Entropy methods, as reviewed in the previous chapter, to investigate the directed dependencies within a low-order coupled ocean–atmosphere model.

3.3 The Modular Arbitrary-Order Ocean-Atmosphere Model

The Modular Arbitrary-Order Ocean-Atmosphere Model (MAOOAM), developed by Vannitsem *et al* (De Cruz, J. Demaeyer, and S. Vannitsem, 2016), is based on a *quasi-geostrophic* atmospheric framework proposed by Jules Charney and David Straus, further elaborated in Reinhold and Pierrehumbert, 1982, and on the ocean model of Stefano Pierini Pierini, 2011. The model solves for the vorticity and the temperature in both media, and is written in spectral Fourier modes, whose full number can be adjusted to the desired resolution.

MAOOAM is part of a series of studies in which incremental improvements progressively integrated relevant physical mechanisms, with the last addition being the introduction of an energy balance between the ocean and the atmosphere. This model is part of a larger python implementation, `qgs`, developed by Lesley De Cruz, Jonathan Demaeyer and Stéphane Vannitsem Jonathan Demaeyer, Cruz, and Stéphane Vannitsem, 2020.

MAOOAM successfully identifies the low frequency variability (LFV) associated with the coupling between the ocean and the atmosphere. The model allows for an arbitrary

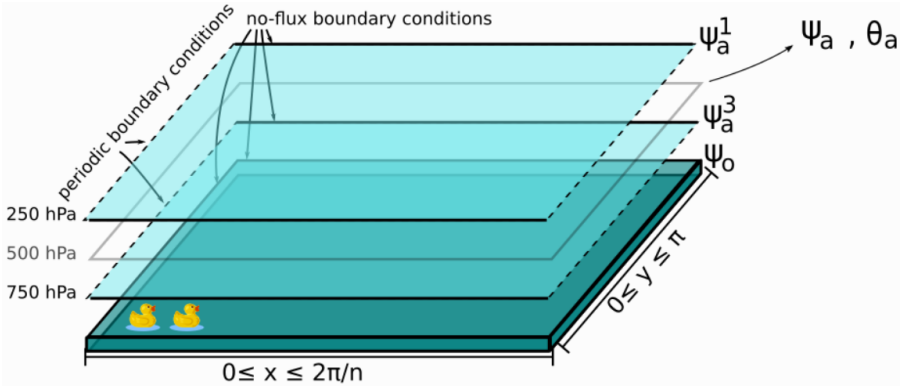


Figure 3.3: Sketch of the ocean-atmosphere coupled model layers. The ocean is a closed basin. The atmosphere is a periodic channel. The domain (β-plane) zonal and meridional coordinates are labeled as the x and y variables. Source.

level of resolution. The LFV structure develops both in the simplest version of the model, and in the versions with higher resolution. Higher resolution versions are good candidates when investigating more realistic dynamics. In this work we will refer to the simplest model formulation, which can be integrated on local machines while, still interesting physical features (Stéphane Vannitsem, Jonathan Demaeyer, et al., 2015; De Cruz, J. Demaeyer, and S. Vannitsem, 2016).

3.3.1 Features of MAOOAM

MAOOAM features the dynamics of a two-layer channel atmosphere and a *shallow-water* active closed-basin ocean layer, meaning the horizontal length scale is much greater than the vertical length scale, and the vertical velocity scale is small compared to the horizontal velocity scale. The dynamics is approximated to take place on a tangent plane to the Earth at a given latitude ϕ: the Coriolis parameter f is Taylor expanded about a given latitude ϕ, retaining only the linear terms: $f = f_0 + \beta y$, with $\beta = (\mathrm{d}f/\mathrm{d}y)|_{\phi_0}$. This approximation is called β-plane approximation. The atmosphere is described by quasi-geostrophic equations, meaning that the Coriolis force and pressure gradient forces are almost in balance, in the limit of small Rossby number, the ratio of the inertial force to the Coriolis force. A simple sketch of MAOOAM is reported in Figure 3.3.

The atmospheric and ocean components are coupled both *thermally* and *mechanically*. The coupling between the two components includes wind forcings, radiative and heat exchanges. The model equations for the dynamics, described in Stéphane Vannitsem, Jonathan Demaeyer, et al. (2015), are reported in the next paragraphs.

Atmospheric dynamics equations

The atmospheric model is based on the vorticity equations of a two-layer, quasi-geostrophic flow defined on a β-plane. The dynamical evolution is thus described in pressure coordinates by the equations of motion of the atmospheric streamfunction fields[1] ψ_a^1 at geopotential height of 250hPa and ψ_a^3 at 750hPa, and the vertical velocity $\omega = \mathrm{d}p/\mathrm{d}t$. The equations read

$$\begin{aligned}
\frac{\partial}{\partial t}\left(\nabla^2\psi_a^1\right) + J(\psi_a^1, \nabla^2\psi_a^1) + \beta\frac{\partial\psi_a^1}{\partial x} &= -k_d'\nabla^2(\psi_a^1 - \psi_a^3) + \frac{f_0}{\Delta p}\omega, \\
\frac{\partial}{\partial t}\left(\nabla^2\psi_a^3\right) + J(\psi_a^3, \nabla^2\psi_a^3) + \beta\frac{\partial\psi_a^3}{\partial x} &= +k_d'\nabla^2(\psi_a^1 - \psi_a^3) - \frac{f_0}{\Delta p}\omega - k_d\nabla^2(\psi_a^3 - \psi_o)
\end{aligned} \tag{3.3}$$

where ∇^2 is the horizontal Laplacian, f_0 is the Coriolis parameter at latitude ϕ_0, with $\beta = \frac{\mathrm{d}f}{\mathrm{d}y}$ its meridional gradient here[2]. The parameters k_d and k_d' multiply the surface friction term and the internal friction between the layers, respectively, and $\Delta p = 500\,\mathrm{hPa}$ is the pressure difference between the two atmospheric layers. Finally, J is the Jacobian $J(S, G) = \partial_x S\,\partial_y G - \partial_y S\,\partial_x G$. The last term of the second equation accounts for the presence of a surface boundary velocity of the oceanic flow (denoted by ψ_o), which represents the mechanical contribution of the interaction between the ocean and the atmosphere. The equations have been nondimensionalized as explained in appendix A of Ref. Stéphane Vannitsem, Jonathan Demaeyer, et al., 2015.

These equations can also be rewritten in terms of the barotropic and baroclinic streamfunctions $\psi_a = (\psi_a^1 + \psi_a^3)/2$ and $\theta_a = (\psi_a^1 - \psi_a^3)/2$. The barotropic streamfunction is linked to the average component of the flux, while the baroclinic streamfunction represents the shear component of the flux.

[1] The stream function ψ in a 3-dimensional space is defined as:

$$\psi(x, y, z) = \int_A^P (u\mathrm{d}y - v\mathrm{d}x), \tag{3.1}$$

where u and v are the x and y components of the flow velocity $\mathbf{u}$, satisfying the continuity equation $\nabla\cdot\mathbf{u}$ and A and P are two points in the xy plane (Lamb, 1945; Batchelor, 1967). The streamfunction ψ is the volumetric flux $Q(x, y, t)$ through the test surface per unit thickness b, where thickness is measured perpendicular to the plane of flow. The flow velocity components are derived as

$$u = \frac{\partial\psi}{\partial y}, \quad v = \frac{\partial\psi}{\partial x}. \tag{3.2}$$

This representation is valid in the case in atmospheric and shallow water oceanic flows, where the thickness of the atmosphere and of the ocean are very small compared to the zonal and meridional extensions, in the x and y directions, respectively (Smyth, 2020).

[2] More precisely, the Coriolis parameter f is linearized around a value f_0 evaluated at latitude ϕ_0, so that $f_0 = f_{\phi_0}$, $f = f_0 + \beta y$, with β the meridional gradient at latitude ϕ_0.

The time evolution of the atmospheric temperature T_a obeys a radiative and heat flux scheme reflecting the exchanges of energy between ocean and atmosphere:

$$\gamma_a \left(\frac{\partial T_a}{\partial t} + J(\psi_a, T_a) - \sigma w \frac{p}{R} \right) = -\lambda (T_a - T_o) + E_{a,R} \tag{3.4}$$

with

$$E_{a,R} = \epsilon_a \sigma_B T_o^4 - 2\epsilon_a \sigma_B T_a^4 + R_a. \tag{3.5}$$

In Eq.3.4, $E_{a,R}$ is the net radiative flux in the atmosphere, R the gas constant, ω the vertical velocity in pressure coordinates, and

$$\sigma = -\frac{R}{p} \left(\frac{\partial T_a}{\partial p} - \frac{1}{\rho_a c_p} \right) \tag{3.6}$$

is the static stability, here taken to be constant, with p the pressure, ρ_a the air density, and c_p the specific heat at constant pressure.

As for the radiative flux in the ocean, in Eqs. 3.5, the net radiative flux $E_{a,R}$ within the atmosphere is composed of three terms: the ingoing flux $\epsilon_a \sigma B T_o^4$ of radiative energy effectively absorbed; the outgoing flux $-2\epsilon_a \sigma B T_a^4$ re-emitted to the ocean and to space; and the shortwave radiative flux R_a absorbed directly by the atmosphere. The rates of radiative emission and absorption are assumed to be equal. This assumption is strictly valid only at thermodynamic equilibrium, but it can be safely applied to systems in local thermodynamic equilibrium, like the lower atmosphere, in which molecular collisional processes dominate the radiative processes Stéphane Vannitsem, Jonathan Demaeyer, et al., 2015.

Oceanic dynamics equations

The oceanic component of the MAOOAM model is a *shallow-water* active ocean layer superimposed on a deep ocean layer at rest. The dynamics is given by the reduced-gravity quasi-geostrophic vorticity equation on a β-plane. The equation of motion for the streamfunction ψ_o of the active ocean layer reads

$$\frac{\partial}{\partial t} \left(\nabla^2 \psi_o - \frac{\psi_o}{L_R^2} \right) + J(\psi_o, \nabla^2 \psi_o) + \beta \frac{\partial \psi_o}{\partial x} = -r \nabla^2 \psi_o + \frac{\mathrm{curl}_z \tau}{\rho h}. \tag{3.7}$$

The active oceanic layer has density ρ, depth h, the quiescent deep layer has density ρ_∞. $g' = g(\rho_\infty - \rho)/\rho$ is referred to as the reduced gravity felt by the fluid in the active layer, and $L_R = \sqrt{\frac{g'h}{f_0}}$ is the reduced Rossby deformation radius. The friction coefficient at the bottom of the active layer is r, and $\mathrm{curl}_z \tau$ is the vertical component of the curl of the wind stress. This forcing is provided by the wind stress generated by the atmospheric component of the coupled system. Assuming, as done in Stéphane Vannitsem, Jonathan Demaeyer, et al., 2015, that the wind stress is given by $(\tau_x, \tau_y) = C(u - U, v - V)$, where

$(u = -\partial\psi_a^3/\partial y, v = \partial\psi_a^3/\partial x)$ are the horizontal components of the geostrophic wind, and (U, V) the corresponding components of the geostrophic currents in the ocean, one gets

$$\frac{\mathrm{curl}_z \tau}{\rho h} = \frac{C}{\rho h}\nabla^2(\psi_a^3 - \psi_o). \tag{3.8}$$

The *drag coefficient* $d := C/(\rho h)$ characterizes the strength of the mechanical coupling between the ocean and the atmosphere. In the next Chapter we will choose two values of d to study the causal relationship between the two fluids in strong and weak mechanical coupling conditions. Equation 3.8 is also nondimensionalized as done for the atmospheric equations.

The temperature T_o is assumed to be a passive scalar transported by the ocean currents, interacting with the atmospheric temperature through radiative and heat exchange, according to a scheme detailed in Barsugli and Battisti, 1998:

$$\gamma_o \left(\frac{\partial T_o}{\partial t} + J(\psi_o, T_o) \right) = -\lambda(T_o - T_a) + E_{o,R} \tag{3.9}$$

with

$$E_{o,R} = -\sigma_B T_o^4 + \epsilon_a \sigma_B T_a^4 + R_o \tag{3.10}$$

In Eqs. 3.9 and 3.10, $E_{o,R}$ is the net radiative flux at the ocean surface. Eq. 3.10 contains the outgoing longwave radiation flux, the longwave radiation flux re-emitted to the ocean and the shortwave radiation entering the ocean, respectively. ϵ_a is the emissivity of the atmosphere and σ_B is the Boltzmann constant. The parameter γ_o is the heat capacity of the ocean, and λ is the heat transfer coefficient between the ocean and the atmosphere that combines both the latent and sensible heat fluxes. It is assumed that the combined heat transfer is proportional to the temperature difference between the atmosphere and the ocean. The radiation scheme of the model is graphically depicted in **this** documentation.

By decomposing the temperatures as $T_a = T_{a,0} + \delta T_a$ and $T_o = T_{o,0} + \delta T_o$, where $T_{a,0}$ and $T_{o,0}$ are the climatological reference temperatures, and doing the same for the short-wave radiation fluxes entering the atmosphere and the ocean, R_a and R_o, one can also obtain the evolution equations for the temperature anomalies δT_a and δT_o. The climatological reference temperatures are set to $T_{a,0} = 270$ K and $T_{o,0} = 285$ K so to have the model operating in realistic conditions.

3.3.2 Low-order spectral formulation of the model

The model we described in the previous section is expressed in terms of partial differential equations (PDEs). The solution of these equations is obtained by projecting them onto a finite orthogonal basis of functions. From this operation we obtain the spectral expansion

of the model. The physical fields are expanded in terms of the basis functions and of Fourier components, e.g. for the atmospheric streamfunction

$$\psi_a(x,y) = \sum_i^{n_a} \psi_{a,i} F_i(x,y), \tag{3.11}$$

where $\psi_{a,i}$ is the i^{th} Fourier component and $\{F_i(x,y)\}_i$ is the set of basis functions used to expand the physical fields. In this way, one obtains a system of ordinary differential equations (ODEs) for the spectral coefficients of the fields, in place of the PDEs. This procedure is sometimes referred to as a *Galerkin expansion*, and it is explained in detail in Appendix A. Eventually, the integrated approximated physical fields of the model can also be retrieved by computing, for instance, Eq. 3.11, and in similar ways for the other physical fields.

An arbitrary level of resolution can be chosen for the model by setting the number of Fourier modes in the spectral decomposition. In this work we set $n_a = 10$ and $n_o = 8$. The model thus consists of n_a Fourier components for the baroclinic streamfunction ψ_a, namely $\{\psi_{a,i}\}_{i=1}^{n_a}$, n_a Fourier components for the atmospheric temperature anomaly δT_a, n_o Fourier components for the ocean streamfunction ψ_o, and n_o Fourier components for the oceanic temperature anomaly δT_o. The full coupled model is thus based on a set of 36 ODEs. The time-dependent solutions of this system are obtained by numerical integration, using a Runge-Kutta method with a fixed time step of around 11 days, and an overall integration time of 10^6 days, approximately.

3.3.3 Dynamical behavior of the system

The dynamical behavior of the system changes depending on the values assigned to the three main parameters of the model, C_o [$\mathrm{Wm^{-2}}$], λ [$\mathrm{Wm^{-2}K^{-1}}$] and d [$\mathrm{s^{-1}}$], we will focus specifically on the last one. These parameters correspond to the meridional variation in the radiative input from the Sun, the strength of the ocean-atmosphere coupling, and the intensity of the heat fluxes, respectively. We focus on two values for the atmosphere-ocean mechanical coupling parameter d, which is set to $d = 1.0e^{-8}$ in the *weak* configuration, and $d = 1.1e^{-7}$ in the *strong* configuration. An illustration of the long term dynamical behavior of the weak and strong configurations, is given in Figure 3.4. The panels show the trajectory solution of the model for around 10^6 days, projected onto the 3-dimensional portion of the phase space spanned by three key modes $(\psi_o^2, \theta_o^2, \psi_a^1)$, i.e. the second Fourier modes of the ocean streamfunction and temperature, and the first one of the atmospheric streamfunction.

The marked difference between shapes of the attractors in Figure 3.4 manifests the two mechanical coupling strength scenarios. In the weakly coupled configuration, the attractor has an approximately regular large scale shape, densely but discontinuously filled by the trajectory, due to chaos. The solutions of the coupled model are well

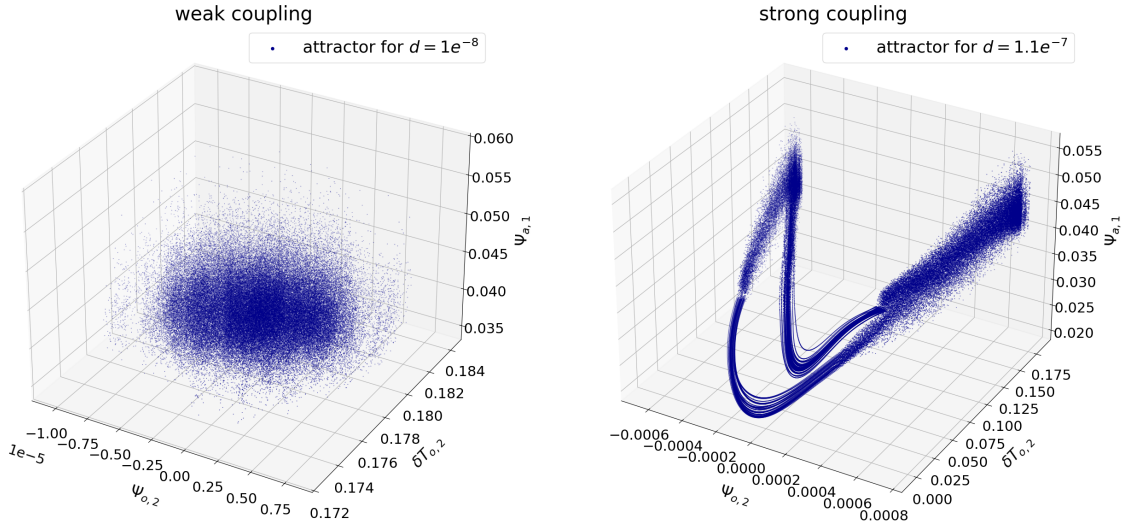


Figure 3.4: Illustration of the MAOOAM attractor in the weak (left) and strong (right) configurations for the three variables $(\psi_o^2, \theta_o^2, \psi_a^1)$. The model is integrated forward for 10^6 days.

localized around small values of $\psi_{o,2}$. The attractor for the strong configuration is still the one of a chaotic dynamics, however now it features organized trajectories around an unstable periodic orbit, around which the solution is wandering. The atmospheric streamfunction in this configuration displays low frequency dynamics, shown in Figure 3.5. This LFV is accompanied by a succession of recurrences in regions of lower or higher values of the atmospheric streamfunction, ψ_a^1, with low and high variability respectively, the "passive" and "active" regimes, respectively (Tondeur et al., 2020).

The chaotic behavior is still dominated by the LFV on decadal and multi-decadal time scales, that are typical of oceanic processes. Once the decadal-scale periodic orbits develop, the coupled system's short-term instabilities are drastically reduced, indicating the stabilizing role of the ocean on the atmospheric dynamics. The LFV in the system is no longer present when d is set to zero, suggesting how the mechanical coupling is at the origin of the development of the long-periodic oscillation. The C_o radiative-flux parameter plays a partial role too, as the coupled ocean-atmosphere mode is found in correspondence of high values of both d and C_o.

In Chapter 4 we set the heat flux parameter λ to a fixed value of $15.06 Wm^{-2}K^{-1}$, an intermediate value with respect to the extremes $\lambda = 0\ Wm^{-2}K^{-1}$ that produces very chaotic trajectories, and $\lambda = 100\ \mathrm{Wm}^{-2}K^{-1}$, which corresponds to purely periodic trajectories, suggesting that the larger is the heat flux, the stabler is the dynamics Stéphane Vannitsem, Jonathan Demaeyer, et al., 2015. We will also fix the radiative input meridional variation C_o to 310 Wm^{-2}, a high enough value so that the LFV develops in the $d = 1.1e^{-7}$ case and does not in the $d = 1.0e^{-8}$ case.

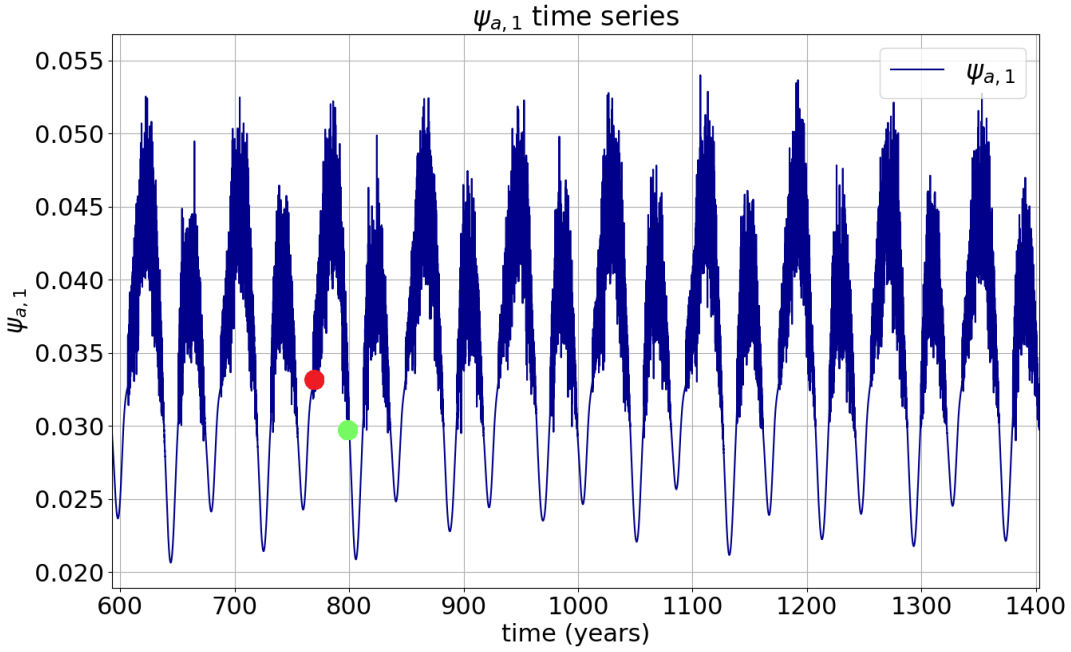


Figure 3.5: Time series of the first component of the atmospheric streamfunction, ψ_a^1, for the strong coupling model configuration. The red and green points indicate the start of active and passive regimes respectively.

Physical features of the model

The model incorporates several key physical ingredients of the coupled ocean-atmosphere system at mid-latitudes. It shows the presence of dry atmosphere baroclinic instability, as the solar forcing induces a horizontal North-South temperature gradient in the atmosphere, in turn keeping the vertical gradient of the wind consistent. The wind gradient then generates a shear force which is responsible for the development of eddies at the interface of the two layers; causing the instability in the ocean. Concurrently, the ocean transports the heat to counteract the initial gradient of temperature Tondeur et al., 2020.

The model also features upper-ocean mass and heat transport in a shallow layer, an energy balance scheme incorporating radiative and heat fluxes, and a mechanical coupling mechanism between the atmosphere and the ocean. It displays zones of predominant low and high pressure and a subtropical jet, having some similarity with the Azores High and the Icelandic Low. These elements recall realistic climatological properties of the atmosphere over the ocean. The mechanical coupling parameter also allows for the development of vigorous, albeit smooth oceanic gyres Stéphane Vannitsem, Jonathan Demaeyer, et al., 2015.

Key points

- Causality has been increasingly applied to study the climate system, an extraordinarily complex, multiscale system. In this context it is used for causal attribution of climate change and extreme events, and to disentangle complex nonlinear interactions between components of the system.
- Open methodological problems in investigating the climate system with causal methods include better handling incomplete domain knowledge, developing methods that can handle high degrees of complexity and data sparsness, and the integration of causality in machine learning for more powerful forecasting and events attribution.
- One open problem in the understanding of the climate system is the interaction between ocean and atmosphere: is it the dog (ocean) that wags its tail (atmosphere), the tail driving the dog, or does the dog bite its tail (coupled ocean-atmosphere mode)?
- We study this problem within a low-order ocean-atmosphere coupled spectral model: MAOOAM. MAOOAM includes realistic thermal and mechanical ocean-atmosphere coupling.
- Two mechanical coupling configurations are selected: strong and weak coupling. In the strong coupling configurations MAOOAM develops low frequency variability (LFV) in the atmosphere.

Chapter 4

Numerical Results

4.1 Causal analysis of the MAOOAM model

In this chapter, we address the problem of the direction of the atmosphere-ocean coupling from a causal perspective, using two of the causal methods introduced in the previous chapters: Transfer Entropy (TE) and Liang–Kleeman Information Flow (LKIF). We apply these methods to the MAOOAM spectral model to investigate the causal relationships among its Fourier modes. We conduct two types of analysis: the first focuses on the causal links in *pairs* of spectral modes in the full multivariate system; the second examines the *interregional* causal links between the ocean and the atmosphere, treating them as distinct subsystems.

In Chapter 2 we have studied the theoretical descriptions of TE and LKIF, whose mathematical formulations are relatively simple. However, the empirical estimation of these quantities from N samples of time-series can be challenging, as it depends on the type of data, it is susceptible to the presence of bias and variance due to finite sample size, and it suffers from the curse of dimensionality. In the coming paragraphs we give details of the numerical implementation of the Transfer Entropy and of the Liang-Kleeman Information Flow.

4.1.1 Numerical estimation of Transfer Entropy

The numerical implementation of the TE relies on the Kraskov-Stögbauer-Grassberger (KSG) technique (Kraskov, 2004), making it suitable to handle multivariate nonlinear systems and long time series. In the experiments we adopt the multivariate and the interregional implementations of the method. The first analyzes pairs of variables in multivariate contexts. We use it to compute the directional link from mode X_i to mode X_j, indicating two of the 36 modes of MAOOAM, applying the method to the time-series that result from the model integration. The interregional approach aims at evaluating

the TE between two multivariate subsystems within a complex multivariate system, in our case: the ocean and the atmosphere.

The method for TE estimation is part of the library *Java Information Dynamics Toolkit* (JIDT) (Joseph T. Lizier, 2014). JIDT provides an open-source code Java implementation of information-theoretic measures in complex systems, focusing on higher level measures for information dynamics, such as MI and TE.

KSG estimation builds on the non-linear and model-free capabilities of *kernel estimation*[1]. More specifically, the method uses a dynamically altered (box-)kernel of width r to adjust to the density of samples in the vicinity of any given observation. For each sample of random variables (x, y), one locates the k-th nearest neighbor in the full (x, y) space, using the max norm to compare distances in x and y. This yields distances r_x and r_y. Two different algorithms are proposed for determining r_x and r_y from the k-th nearest neighbor, we report them in Appendix B. In this work we use the second algorithm, less biased and more accurate for large dimensions of the system.

Using kernel estimation allows to do bias correction, which grants better data efficiency and accuracy, and makes the method effectively parameter-free. As such, it is widely-used to compute MI, conditional MI and TE for continuous data. The method is rather computationally expensive, but it can be made faster through fast nearest neighbor techniques Joseph T. Lizier, 2014.

Interregional Transfer Entropy in JIDT

In many applications we are interested in understanding the TE flow between two or more regions within the variable space. It is generally impractical to compute this *interregional information transfer* as the multivariate TE (see subsection 2.2.4) between the complete set of variables in the two regions, say $\mathbf{R}_a$ and $\mathbf{R}_b$, as these regions might contain a high number of variables. A practical way to retain the benefits of the TE measure in this scenario is to compute the interregional information transfer from region to region as

[1]When we use kernel-estimators, the joint pdfs that we need to compute information-theoretic quantities are estimated with *kernel functions*, which measure "similarity" between pairs of samples $\{x_n, y_n\}$ and $\{x'_n, y'_n\}$ using a resolution r, also referred to as *kernel width*. For example, suppose we need to estimate $\hat{p}(x, y)$, this can be computed as

$$\hat{p}_r(x_n, y_n) = \frac{1}{N} \sum_{n'=1}^{N} \Theta(|(\frac{x_n - x'_n}{y_n + y'_n})| - r). \tag{4.1}$$

There are different possibilities as per the shape of the function Θ and for the norm $|\cdot|$. JIDT implements the "box kernel" combination, where Θ is the step kernel: $\Theta(x > 0) = 0$, $\Theta(x \leq 0) = 1$, and the norm is the maximum distance. These estimates for the pdfs are then used directly in evaluating a local measure for each sample $n \in [1, N]$ and averaging over all samples. Kernel estimation can measure non-linear relationships and is model-free, but it is sensitive to the choice of parameter r, and it is biased and less time-efficient than other methods like Gaussian estimators.

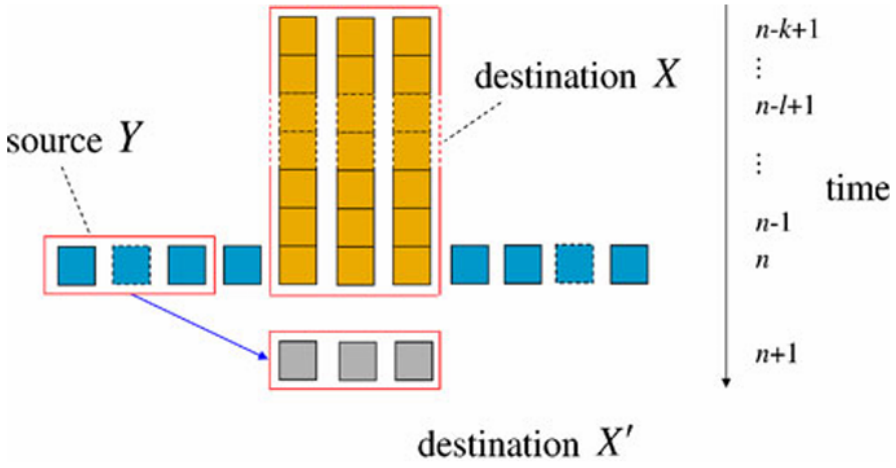


Figure 4.1: Multivariate transfer entropy $T_k(\mathbf{Y} \to \mathbf{X})$ from the set of source variables $\mathbf{Y}$ to the set of destination variables $\mathbf{X}$. Courertsy of J. Lizier et al. (2011).

the multivariate TE *averaged* over a large number S of *samples* of pairs of subsets of v variables in each region:

$$T_{k,v}\left(\mathbf{R}_a \to \mathbf{R}_b\right) = \left\langle T_k\left(\mathbf{R}_{a,i} \to \mathbf{R}_{b,i}\right)\right\rangle_{i,j}. \tag{4.2}$$

Here, i and j label the sample subsets $\mathbf{R}_{a,i}$ and $\mathbf{R}_{b,j}$ of size v in each region. Further technical details can be found in J. Lizier et al., 2011. The method allows to identify the underlying directed information structure between multivariate regions of the system in consideration, capturing collective interactions resulting from the combined (nonlinear and directional) activity of multiple driving elements. The practical implementation of the method features Kraskov's method (Kraskov, Stögbauer, and Grassberger, 2004) and is further discussed in J. Lizier et al., 2011.

Significance testing

In theory, the TE between two unrelated variables are equal to 0. In practice, since the TE is empirically measured from finite data, there will be some error to take into account. In JIDT this issue is handled through hypothesis testing. A null hypothesis H_0 is formulated: the state changes $x_n^{(k)} \to x_{n+1}$ of the destination X have no temporal dependence on the source Y. A test of statistical significance is made in support of H_0 by generating a number of surrogate time series Y^P, permuting the elements of the source time series y_n assuming $H_0$to hold, and then performing a one-sided p-value test.

Significance testing for the multivariate measure $T_k(\mathbf{Y} \to \mathbf{X})$ is a straightforward extension. Importantly, in generating the surrogates $\mathbf{Y}^P$ the component time series $Y_1, Y_2, \ldots$ of $\mathbf{Y}$ are not permuted *individually*, but the vectors $\mathbf{y}_n$ are permuted at each time point n in $\mathbf{Y}$ *as a whole*. This ensures that the only difference in making the

surrogate measurements $T_k(\mathbf{Y}^P \to \mathbf{X})$ is the temporal relationship $p(x_{n+1}|x_n^{(k)}, \mathbf{y}_n)$ (J. Lizier et al., 2011). Significance testing can similarly be extended to the *interregional* measure $T_{k,v}(\mathbf{R}_a \to \mathbf{R}_b)$, though there are some differences that are taken into account in the code and explained in J. Lizier et al., 2011.

4.1.2 Numerical implementation of the Liang-Kleeman Information Flow

The LKIF theoretical approach has been implemented in two versions to study both bivariate and multivariate system of time series, as seen in Section 2.3. The multivariate approach is used to compute the information flow between the elements of certain pair of variables of the multidimensional system, where variables are time series X. San Liang, 2021. The implementation of the multivariate method uses the estimator formula derived in Appendix C. We use this method to compute the directional link from mode X_i to mode X_j of the MAOOAM spectral model, analogously to the use of the multivariate TE. Importantly, we note that unlike the TE, which is model-free, the LKIF is a *linear* method.

The LKIF has been further extended to compute the information flow between two multivariate subsystems within a larger complex system (X. San Liang, 2022). The goal of the method is analogous to that of the interregional TE, despite the underlying approach being different. We will refer to this method as *interregional LKIF*.

In the numerical experiments, we use the normalized LKIF introduced in Paragraph 2.3.2, denoted by $\tau_{i\to j}$ and expressed in *nats*. When $\tau_{j\to i}$ is statistically different from 0 (either positive or negative), X_j has an influence on X_i, while if $\tau_{j\to i} = 0$, there is no influence. A value of $|\tau| = 100\%$ indicates that X_j has the maximum influence on X_i. A positive (negative) value $\tau_{j\to i}$ means that the variability in X_j makes the variability in X_i more uncertain (certain), i.e., it increases (decreases) the variability in X_i (X. San Liang, 2014). The numerical implementation of this method has been developed by D. Docquier, S. Vannitsem, Ragone, et al. (2022).

The *interregional* approach analyzes the interactions between all 36 MAOOAM time series at the same time. We differentiate the first 20 atmospheric variables from the other 16 oceanic variables, thereby considering the atmosphere and ocean as distinct subsystems. For this approach, we use a numerical Python implementation developed by Daniel Hagan, previously applied in Shao et al. (2024).

The statistical significance of $\tau_{j\to i}$ and for the interregional LKIF is computed via bootstrap resampling with the replacement of all the terms included in the LKIF max likelihood estimators.

4.2 Experimental setup

In Table 4.1 we report the specific settings we used for the multivariate TE, and in Tab. 4.2 those for the interregional TE. The meaning of the parameters is fully explained in subsection 4.1.1. We consider a significance value of $\alpha_{TE} = 0.1$.

Properties	REORDERINGS	CALCULATOR	HIST	NORM	k	ALGO
Multivariate	100	Kraskov	1	max norm	2	2

Table 4.1: Set up parameters for the *multivariate TE*. REORDERINGS: number of surrogate pairs for statistical significance computation. CALCULATOR: which underlying transfer entropy or mutual information implementation to use for the calculation. HIST: considered history length. NORM: type of norm used. k: number of nearest neighbors K to use in KSG estimation. ALGO: which algorithm to select for the computation.

Properties	v	S	REORDERINGS	CALCULATOR	HIST	NORM	ALGO
Interregional	3	100	50	Kraskov	1	max norm	2

Table 4.2: Set up parameters for the *interregional TE*. **v**: number of joint values to consider in each region. **S**: number of subsets of v variables to select from each region. REORDERINGS: number of surrogate pairs for statistical significance computation. CALCULATOR: which underlying transfer entropy or mutual information implementation to use for the calculation. HIST: considered history length. NORM: type of norm used. ALGO: which algorithm to select for the computation.

For the multivariate LKIF statistical significance computation we used 1000 bootstrap iterations and a significance level $\alpha_{LKIF,mv} = 0.01$. For the interregional LKIF we lowered the reorderings number to 100 and set the significance level to $\alpha_{LKIF,int} = 0.05$, due to computational power limitations. We chose less surrogates for the TE iterations, as the algorithm is slower. We will take this into account when analyzing the results.

4.3 Experiment results: multivariate analysis

In this section we perform numerical experiments using the multivariate TE and the LKIF, applying them with the parameters specified above. The spectral nature of MAOOAM makes it appealing to use multivariate methods that output variable-to-variable results. This type of analysis presents some limitations, as it is not immediate to retrieve physical meaning from the single Fourier modes. However, we will see that the multivariate analysis is nonetheless informative about relevant aspects of the problem at study.

Both methods take as time-series input the 36 modes of MAOOAM, where they are used to identify the causal links between each pair of variables (X_i, X_j) with $i, j =$

$1, ..., 36$. While the LKIF is able to account for self-interactions (X_i, X_i), the TE is only implemented to study the information transfer between different variables. The "self-loop" for information-theoretic quantities is usually referred to as *information storage*, and not treated with the TE.

We also compare the results obtained with the TE and the LKIF with the *instantaneous correlation*. The basic difference between the first two methods and correlation is that the latter is a linear symmetric measure, as can be seen in Figure 4.5, while the other are directed measures. In all the numerical experiments of this section we take the absolute value of the measured quantities and we normalize them to the unity, as we wish to compare only the relative magnitudes of the causal couplings retrieved with the different methods. We also implement a masking to set non-significant values to zero based on the respective significance tests.

In the figures, the first 20 variables are atmospheric variables, and the other 16 are oceanic variables. The causation patterns are meant to be read in the following way, also shown in Figure 4.2: the *top-left* panel (20×20 square) indicates the causal relations among atmospheric variables $a_j \to a_i$ in the following), the *top right* panel (16×20 rectangle) shows the causality *from* the atmospheric variables *to* the oceanic ones ($a_j \to o_i$), the *bottom-left* (20×16 rectangle) shows the causal links from oceanic modes to atmospheric modes ($o_j \to a_i$), and the *bottom-right* (16×16 square) gathers the causal links among oceanic variables ($o_j \to o_i$).

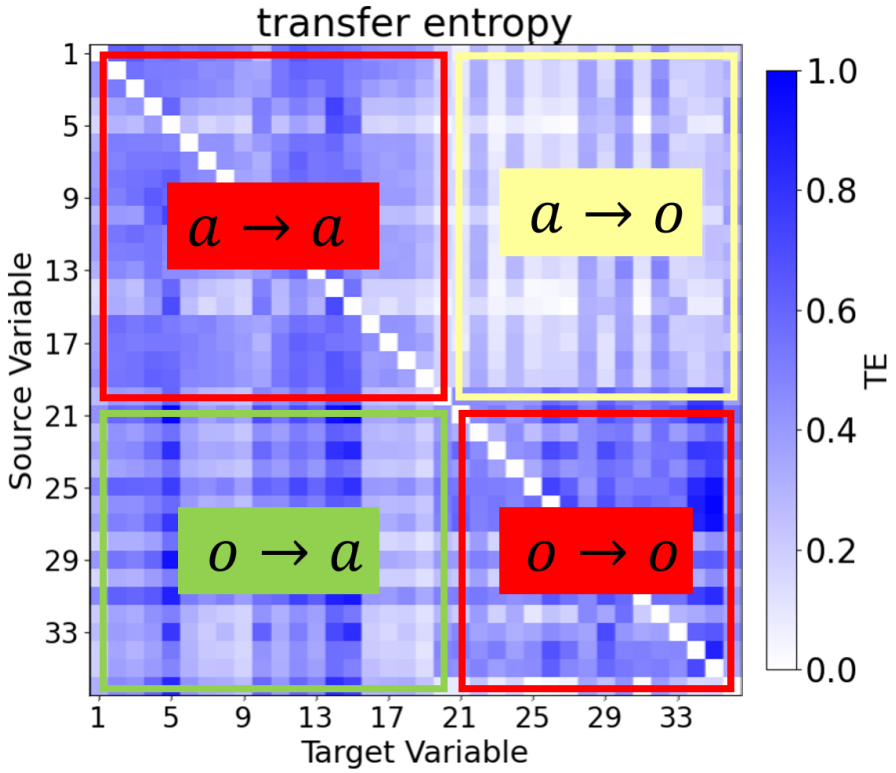


Figure 4.2: The structure used to display causal links in the multivariate analysis experiments. a indicates the atmospheric variables and o the oceanic variables. Causality source (target) variables are represented on the y-axis (x-axis). The same scheme is applied to all experiments in this section.

Strong and weak coupling comparison: TE, LKIF, correlation

We analyze the output of TE, LKIF and correlation when applied to the strong coupling and weak coupling configurations of MAOOAM. The two causal methods and the correlation all identify that the system undergoes a marked change when the friction coefficient changes from strong to weak. TE exhibits the greatest responsiveness to this change, likely because its nonlinear behavior captures dynamics that linear measures such as LKIF and correlation cannot.

The **TE** particularly highlights the difference between the two scenarios, as shown in Figure 4.3. As said, the self-interactions, corresponding to the diagonal in the plot, are not computed. The strong coupling analysis indicates strong $o_j \to o_i$ causal links, and slightly less strong $a_j \to a_i$ interactions. More remarkably, it retrieves strong $o_j \to a_i$ interactions, which instead are less marked for the causal link from the atmosphere to the ocean. The weak coupling output shows considerable $o_j \to o_i$ interactions, while almost zero $a_j \to a_i$ causality. A weak causal interaction is identified in the $a_j \to o_i$

direction, almost entirely absent in the opposite direction.

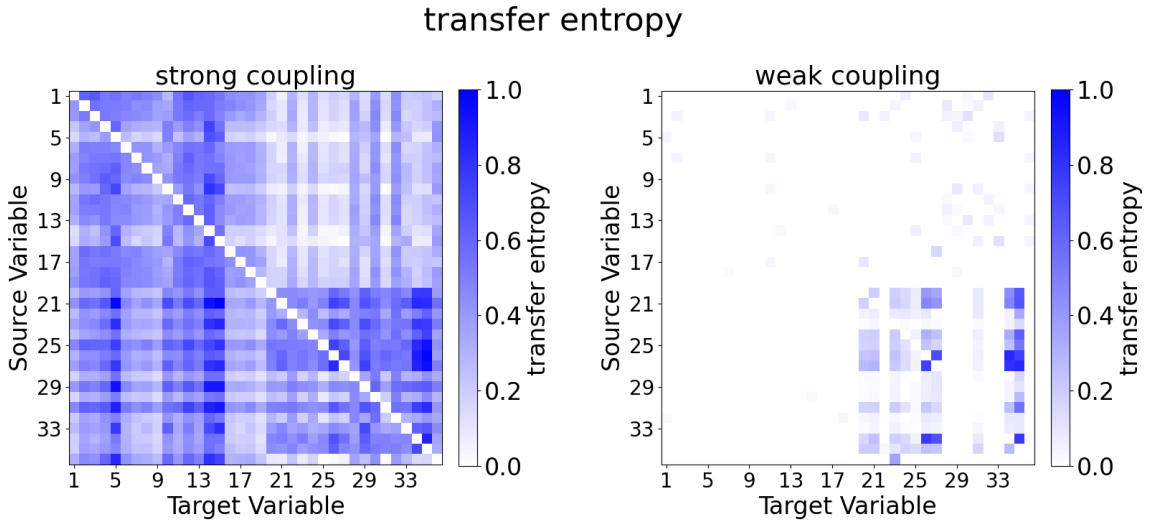


Figure 4.3: Comparison of the bivariate Transfer Entropy in absolute value for MAOOAM in the strong (left) and weak (right) coupling scenarios; the axes display the system's state index. To facilitate direct comparison, the values in both scenarios are globally normalized to unity, with $|TE_{j\to i}| = 1$ in the picture indicating the global maximum value obtained with the two computations.

The **LKIF** method, similarly to the correlation (Figure 4.5) retrieves strong self-interactions, displayed in the full-colored diagonal in Figure 4.4. While this is an expected behavior for the correlation, there is no such prescription for the LKIF, although it is reasonable that variables retain strong information flow from their own past. The causal links appear to be generally less intense with respect to the TE results, but we can still identify a clear difference between the strong and weak coupling cases. For strong coupling, the causal structure is similar to the one retrieved with the TE, despite displaying a less marked difference in the $a_j \to o_i$ direction with respect to the $o_j \to a_i$ direction. Similar considerations apply in the weak coupling case. The $o_j \to o_i$ interactions are marked. Some marked structures are identified in the $a_j \to a_i$ panel, which were not visible for the TE. Also in this case the $a_j \to o_i$ causality is greater than in the opposite direction, while still feeble.

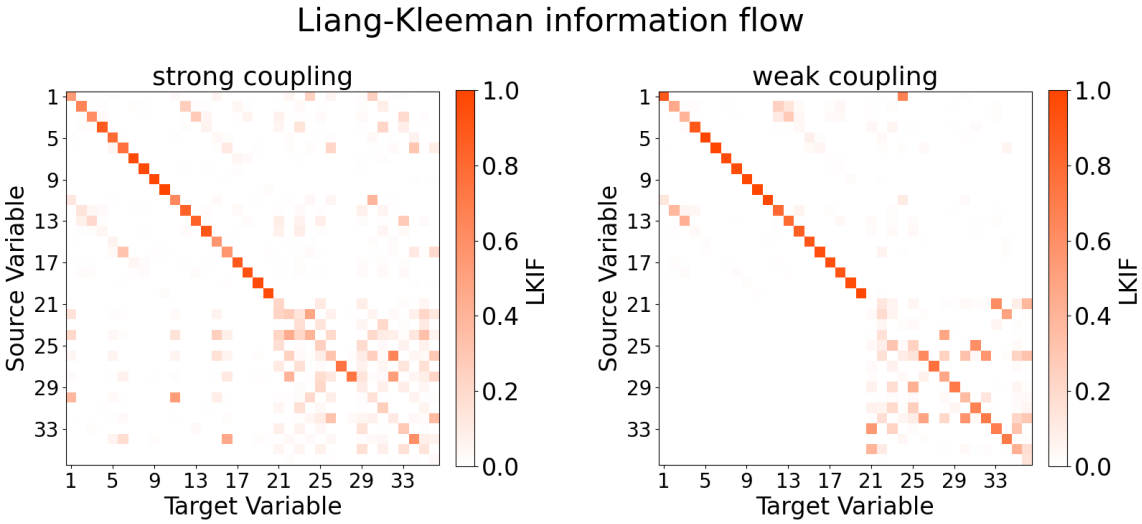


Figure 4.4: Comparison of the normalized Liang–Kleeman information flow τ (in absolute value) for MAOOAM in the strong (left) and weak (right) coupling scenarios; the axes display the system's state index. To facilitate direct comparison, the values in both scenarios are globally normalized to unity, with $|\tau_{j\to i}| = 1$ in the picture indicating the global maximum value obtained with the two computations.

We compare the outputs of the two causal methods with the **correlation**, shown in Figure 4.5. Similarly to the LKIF, the diagonal is dominant with respect to the other correlations, as expected. In this case the picture is symmetric, since the correlation coefficient is a symmetric scalar. We observe much stronger correlations between the atmospheric and the oceanic modes in the strong coupling condition with respect to the weak coupling one. Some interesting hard-colored diagonal structures are observed in the atmosphere-atmosphere panel. The correlation correctly captures the influence of the mechanical coupling strength, but does not differentiate the directed $o_j \to a_i$ and $a_j \to o_i$ interactions.

The causal methods are able to differentiate the $o_j \to a_i$ interactions from the $a_j \to o_i$ ones, retrieving information the correlation does not have, and they correctly identify the physical transition from strong to weak coupling in MAOOAM. However, interpreting causal links between the individual Fourier modes presents significant limitations, for two main reasons. The first is that the system comprises a very large number of variables, which might lead to errors in the conditioning mechanisms inside the methods. The other is that the physical meaning of a causal relationship between Fourier modes is not straightforward.

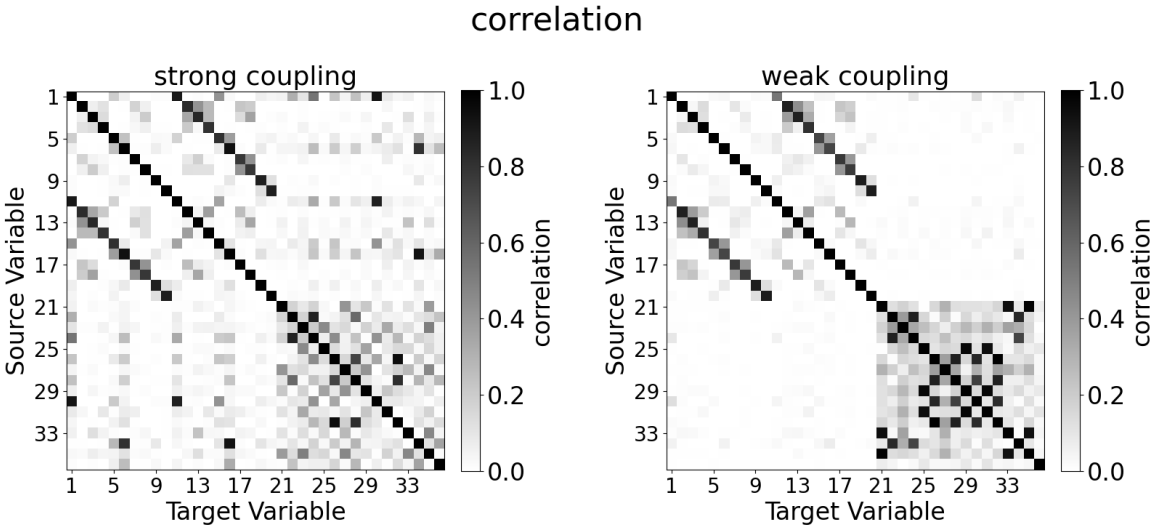


Figure 4.5: Comparison of the instantaneous correlation for the spectral model under the strong (left) and weak (right) coupling scenarios; the axes display the system's state index. To facilitate the comparison with the other methods, the correlation is normalized to the unity and the absolute value is considered.

Smoothing the atmospheric signal

In Section 3.3 we have seen that the atmospheric streamfunctions display regions of fast variability on top of the LFV (the *active* regions in Figure 3.5), in the strong coupling MAOOAM configuration, and fast chaotic dynamics in the slow coupling configuration. We are interested in knowing if the ocean-atmosphere causal relations change once we smooth out the fast variability of the atmosphere. This operation amounts reducing the evolution time-scales gap between the two media. Therefore, we perform a running mean with increasing values of the averaging window over atmospheric data (instantaneous values, 100 days, 1000 days) similarly to what done in Tondeur et al. (2020). We then apply the TE and the LKIF to the resulting time series. Figure 4.6 displays results when the coupling is strong, and Figure 4.7 when the coupling is weak.

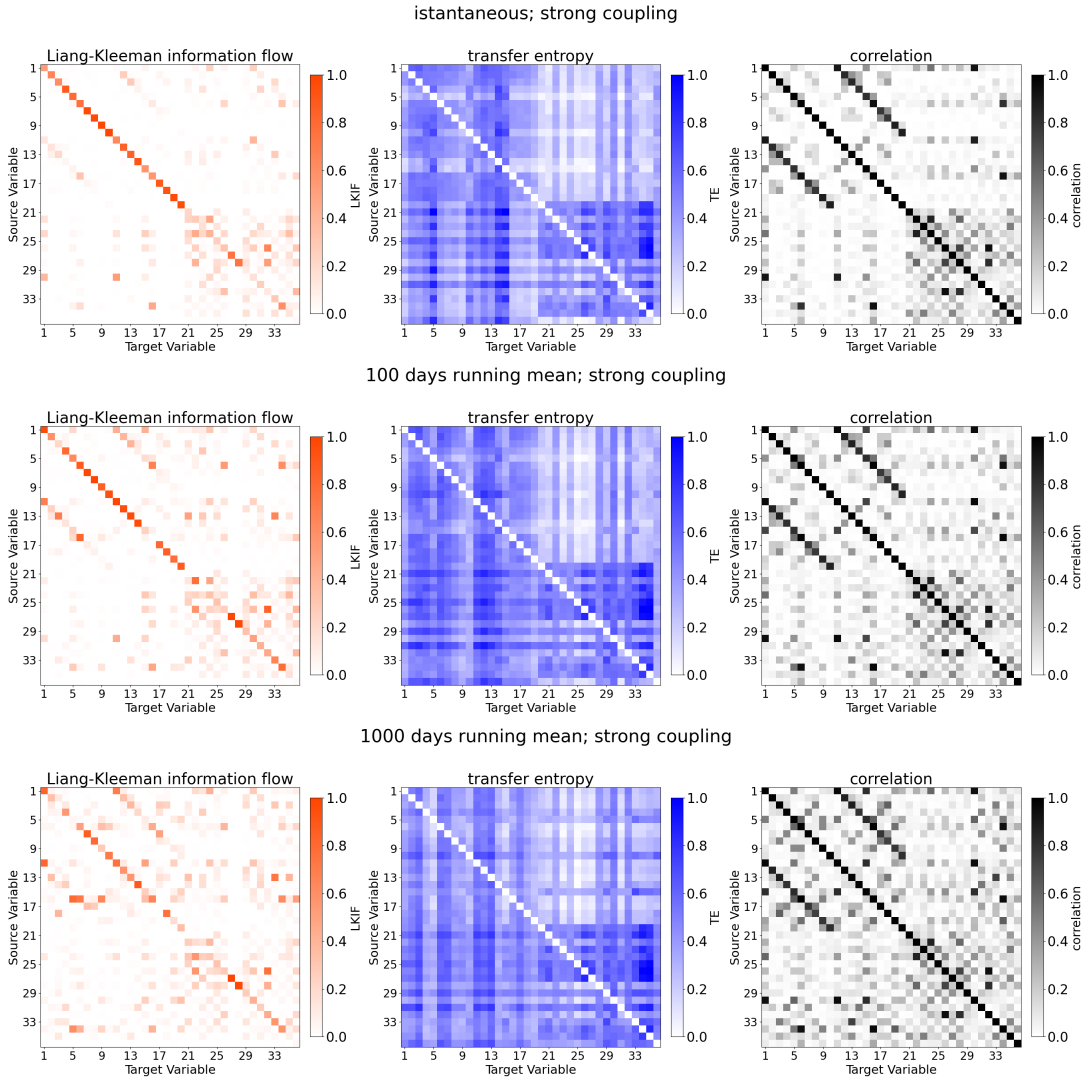


Figure 4.6: Comparison of the Liang–Kleeman information flow τ (left), Transfer Entropy (center) and correlation (right) for the spectral model in the strong coupling scenario. The top figure shows the instantaneous analysis results; the other figures show the results for the atmospheric time-averaged values over the averaging period of 100 days (center) and 1000 days (bottom).

We first look at the correlation in the **strong coupling** scenario. We clearly see the intensity of the correlation increasing as the size of the running mean window increases. This increase is particularly marked in the $o_j - a_i$ areas, showing how the smoothing operation effectively makes the ocean-atmosphere coupling emerge more clearly. We observe similar changes for the LKIF, with an increase of the self-causation areas $o_j \to o_i$ and $a_j \to a_i$, and an increase in the $o_j \to a_i$ and $a_j \to o_i$ areas too. The difference is particularly marked in the latter direction, indicating that the smoothing of the atmo-

spheric fast variability makes the driving of the atmosphere on the ocean emerge more neatly. The TE appears to increment with the size of the averaging window, too, but the change from the case with 100 days-window size to the case with window size equal to 1000 days is not so straightforward. We observe a change in the structure of the causation in the whole system, which however does not affect the understanding that the $o_j \to a_i$ causal link is more marked than the opposite direction.

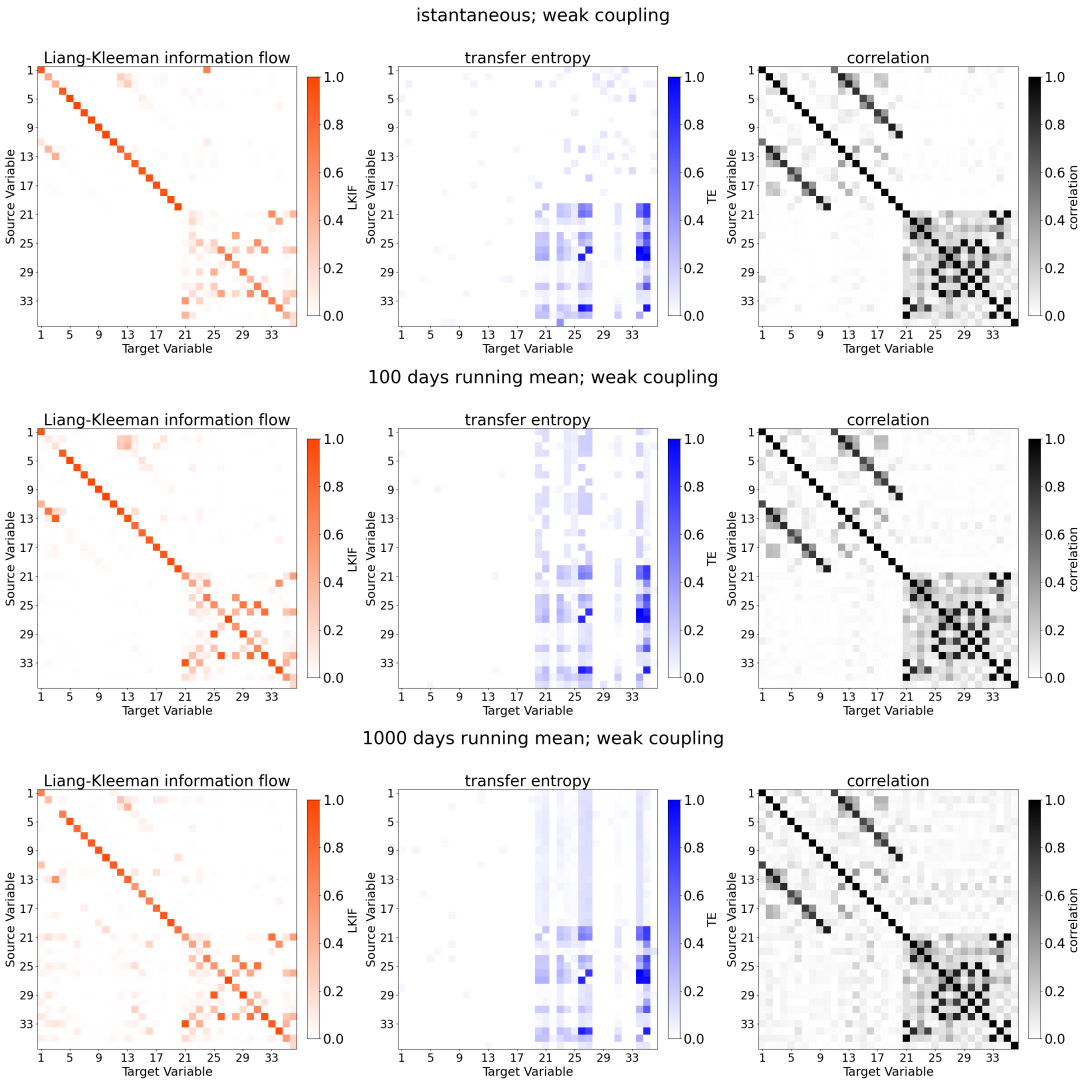


Figure 4.7: Comparison of the Liang–Kleeman information flow τ (left), Transfer Entropy (center) and correlation (right) for the spectral model in the weak coupling scenario. The top figure shows the instantaneous analysis results; the other figures show the results for the atmospheric time-averaged values over the averaging period of 100 days (center) and 1000 days (bottom).

In the **weak coupling** scenario, it is interesting to note that in this case the instantaneous values of correlation are close to zero. The progressive smoothing of the atmospheric variability thus makes the correlation in the $o_j - a_i$ areas emerge, rather than increase, as the window size grows. We observe something similar for the causal methods too, with important differences.

The progressive smoothing makes a considerable $o_j \to a_i$ causal link emerge for the LKIF, particularly in the 1000 days case, while the opposite-direction link appears to remain about zero. For the TE, we observe a neat increase in the $a_j \to o_i$ area, limited to a fraction of the oceanic modes. This increase involves only a part of the atmospheric modes in the 100 days case, and it extends to all 20 of them in the 1000 days case. The $o_j \to a_i$ displays zero causation in both scenarios.

We stress the same considerations on the limitations of performing multivariate causal analysis on MAOOAM. Despite these difficulties, we can say that the smoothing operation appears to make the ocean-atmosphere coupling emerge more clearly, although the LKIF and the TE appear to make the causal links emerge in opposite directions, $o_j \to a_i$ and $a_j \to o_i$ respectively.

4.4 Experiment results: interregional analysis

In this section we report the results of the analysis of the MAOOAM model applying the interregional TE and LKIF methods. Both methods consider two or more subsystems of a certain given system, and compute the causal link between them. The natural choice in the case of MAOOAM is to divide the first 20 atmospheric modes of the model, subsequently called subsystem A, from the other 16 spectral modes, subsystem O. Therefore, the interregional methods allow to compute the complex interactions from $A \to O$ and from $O \to A$, in the different configurations of the model.

Strong and weak coupling comparison

In the first analysis we compare the output of the two methods applied to the strong coupling scenario and the weak coupling scenario, shown in Figure 4.8. In the strong coupling scenario (left panel) both the TE and the LKIF identify a strong causal link $O \to A$. In the opposite direction the TE retrieves a slightly weaker causal link, while the LKIF finds a non-significant value, corresponding to the absence of the causal link. This numerical result is compatible with what we expect from the model, where the ocean reduces the fast variability of the atmosphere in the strong coupling scenario, as outlined in Chapter 3. In the weak coupling case the situation is reversed. The link is stronger in the $A \to O$ direction for both methods. The TE retrieves a weaker but significant link in the $O \to A$ direction, while the LKIF reports a non-significant link.

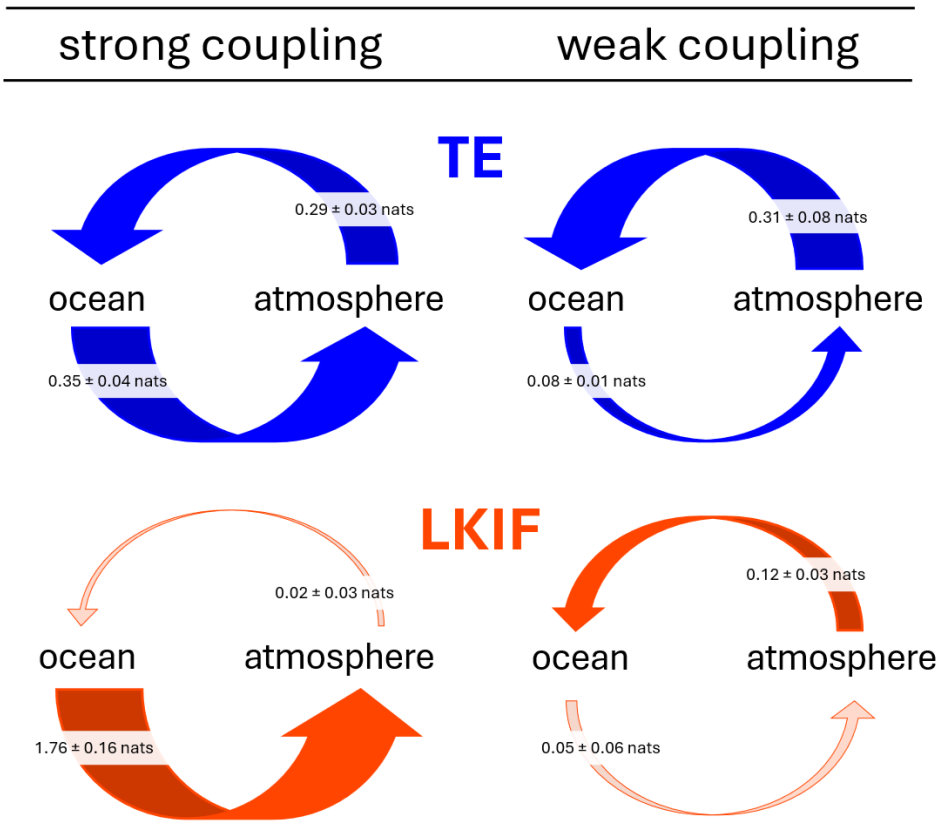


Figure 4.8: Interregional analysis of the ocean-atmosphere coupled system using the TE output (top) and LKIF output (bottom), in the strong coupling scenario (left), and the weak coupling scenario (right). The fully opaque arrows indicate significant causal links; the transparent arrows indicate there is no causation in that direction (non-significant causal link). The estimated strength of the causal link is reported on each arrow, with the associated error. The arrow dimensions are normalized with respect to the method used: dimensions can be compared between the strong-weak coupling scenarios of the same method, but cannot be compared between LKIF and TE.

From these results we observe how the interregional causal analysis is more informative with respect to correlation methods, as it allows to separate the causal influence from the ocean to the atmosphere and vice versa. The TE reports smaller values with respect to the significant links obtained with the LKIF, which makes nonviable a direct comparison. The two methods agree in the relative magnitudes of the inferred causal links: both methods identify a large causal influence from $O \to A$ in the strong coupling scenario, and detect significant $A \to O$ causation in the weak coupling scenario. However, we note that the LKIF is more prone to labeling causal links as non-significant, while the TE has higher resolution. This could be due to the linear nature of the LKIF, compared to the model-free approach of the TE. The overall result is the detection of non-symmetric causation between the O and A systems, with qualitative compatibility

of the $O \rightarrow A$ and the $A \rightarrow O$ causal links in the strong and weak coupling scenarios, respectively. This indicates that both methods, albeit carrying significant limitations, agree on the physical content of the model.

We compare the interregional results with the multivariate results reported in the previous section. In the strong coupling scenario, TE results obtained with the interregional approach are compatible with multivariate TE results, as an approximate estimate of the causal strength between $o_j \rightarrow a_i$ values, compared to the causal strength between $a_j \rightarrow o_i$ values agrees with the bigger dimension of the $O \rightarrow A$ arrow with respect to the thinner $A \rightarrow O$ arrow. The comparison is less clear for the LKIF, as we do not observe a marked difference between the two areas in the multivariate LKIF, whereas the interregional approach highlights a strong difference.

In the weak coupling scenario we observe the multivariate and interregional TE results are again compatible, as the $a_j \rightarrow o_i$ causal direction of the multivariate approach is predominant over the opposite direction, as also reflected by the arrow dimensions in the interregional case. A similar consideration can be done for the LKIF, since the only causal links found by the multivariate method appear to be in the $a_j \rightarrow o_i$ direction, and the same result is found with the interregional LKIF.

The interpretation of the TE color intensity in the multivariate case is not straightforward, as the sharp reduction of the blue intensity does not correspond to a marked change in the dimension of the arrows in the interregional approach.

Smoothing the atmospheric signal

We here analyze the TE and LKIF results obtained when a running mean operation is applied to atmospheric modes, exponentially increasing the values of the averaging window from instantaneous values to approximately 40 years. The oceanic variables are not averaged. The rationale behind this operation is to observe how the causal relations change when the fast variability of the ocean is smoothed out. The results for the TE are shown in Figure 4.9, LKIF results are shown in Figure 4.10 . The top panels correspond to the strong coupling case and the bottom panels to the weak coupling case, in the $A \rightarrow O$ and $O \rightarrow A$ directions, respectively.

In general, we observe that the smoothing of the atmospheric variability affects the obtained results, and some interesting behaviors are observed both for the TE and the LKIF. In the strong coupling scenario, the $A \rightarrow O$ TE causal link smoothly diminishes as the atmospheric variability is increasingly smoothed out. The values of the causal link become non-significant, i.e. the error bars of the data points intersect the axis $y = 0$, around a widow size of 1 year. This indicates that the influence of the atmosphere on the ocean goes to zero when the averaging window is larger than 1 year, approximately. Instead, the LKIF causal link, while initially not significant, progressively stabilizes to a small but significant strength. In the $O \rightarrow A$ direction, the TE causal link changes magnitude abruptly a few times before stabilizing around a value of 0.2 $nats$. A similar

behavior is observed for the LKIF, but the causation values for very high averaging periods stabilize around almost-zero values.

The weak coupling scenario shows an interesting behavior for the $A \to O$ link with both methods. For TE, the connection becomes non-significant following a cusp-like behavior for a few months of running mean window size, and then becoming significant with approximately constant values for higher average values. In the opposite direction $O \to A$, the causal link is initially small but significant, then increases up to a peak, which is meat around a 1 year running mean over the atmosphere, then slowly decreasing again while remaining significant. For the LKIF, the entity of the causal link is initially stable, then enters a region of minima and starts increasing for averaging values larger than 1 year. It peaks around a running mean window size of 10 years approximately and starts dropping again. In the opposite direction, the $O \to A$ connection is initially not significant, with large error bars for the first values of running mean window. For higher values of the averaging window the causal connection becomes small but significant.

The comparison of the two methods as the atmospheric dynamics is progressively smoothed returns that TE and LKIF is not immediate, as the TE systematically outputs smaller values with respect to the TE, and the two methods do not always agree on wether a causal link is significant or not. Also in the case of the multivariate analysis, we stress that an immediate physical interpretation of the output is hardly achievable, as the analysis is performed on a simplified spectral model.

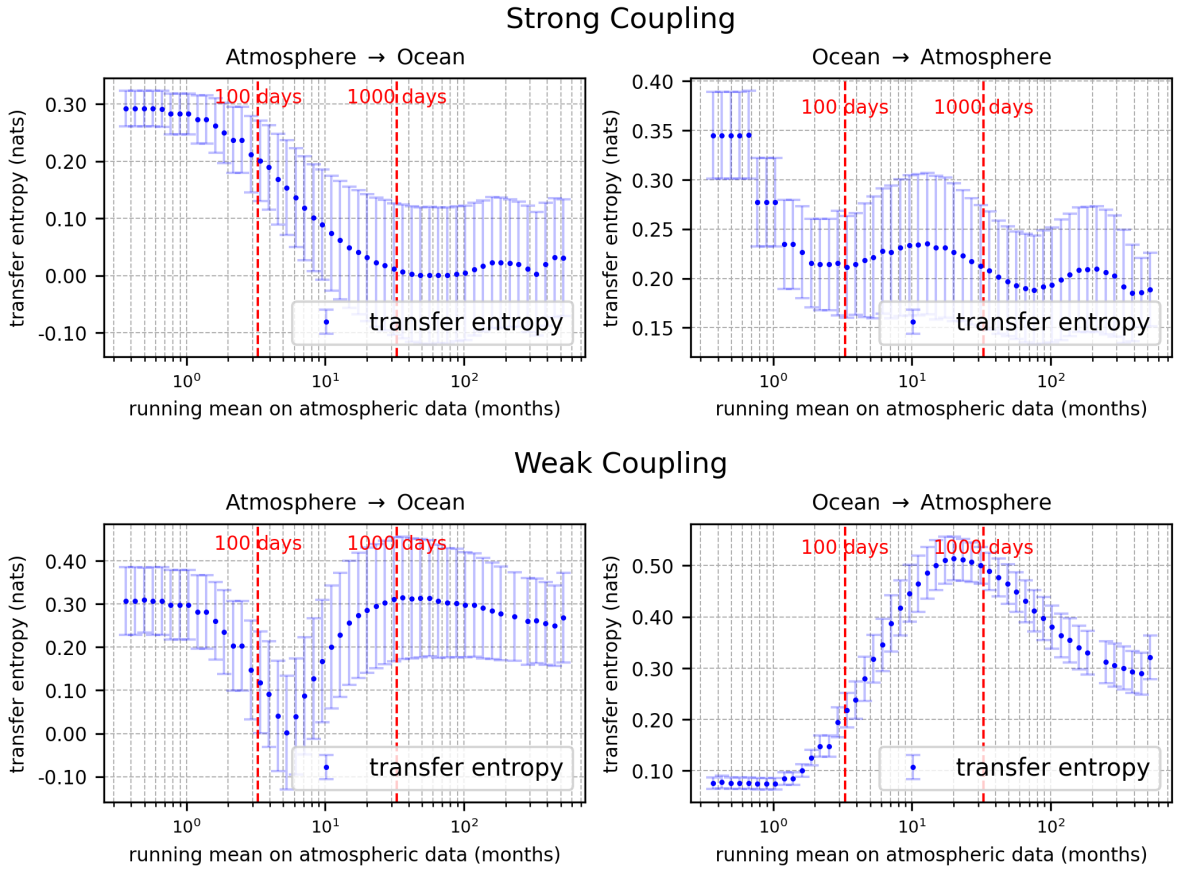


Figure 4.9: Analysis of the ocean-atmosphere coupling using the interregional TE in function of a running mean over atmospheric modes. The averaging period (in months) increases exponentially on the x axis, shown in log scale. Results for the strong coupling scenario are shown in the top panel: $A \to O$ (left) and $O \to A$ (right). Results for the weak scenario are shown in the bottom panel: $A \to O$ (left) and $O \to A$ (right). Error bars are reported on each data point. Red dashed lines indicate running mean window sizes of 100 and 1000 days.

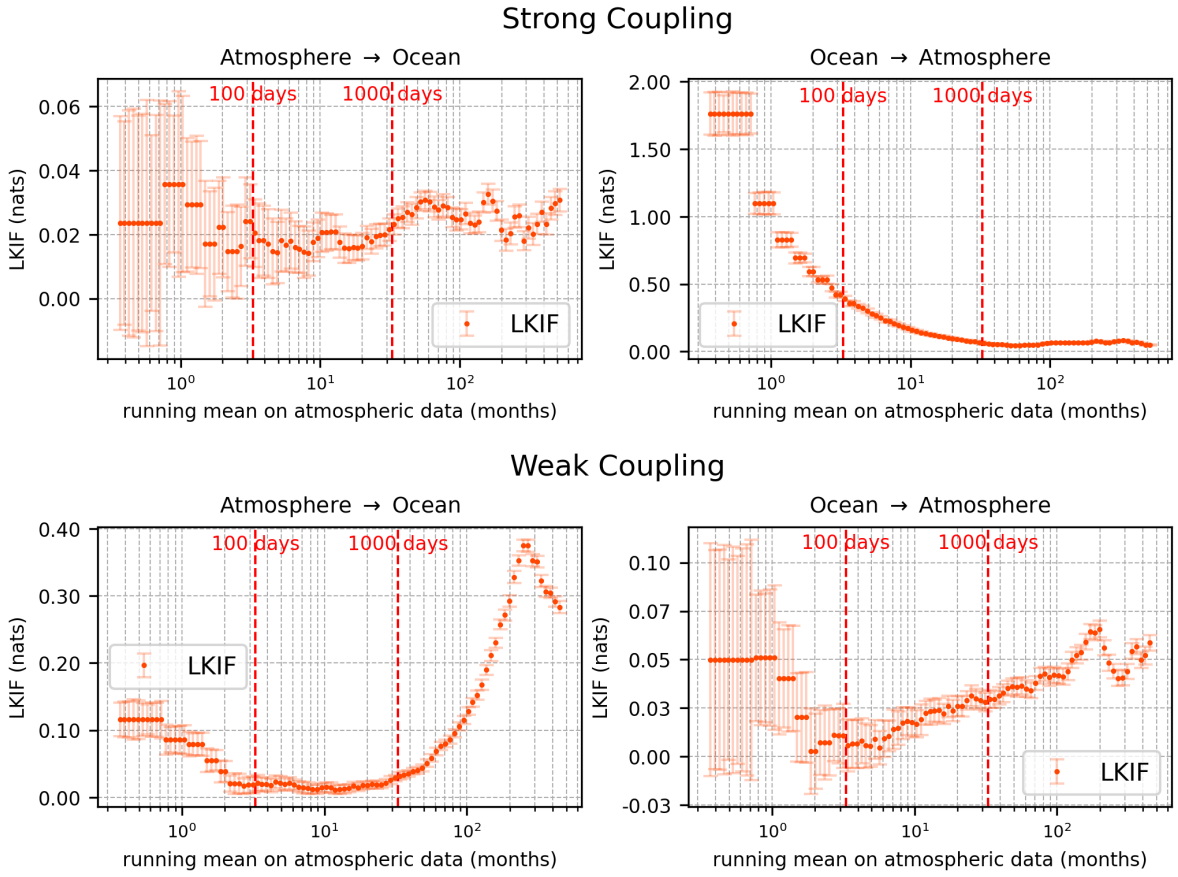


Figure 4.10: Analysis of the ocean-atmosphere coupling using the interregional LKIF in function of a running mean over atmospheric modes. The averaging period (in months) increases exponentially on the x axis, shown in log scale. Results for the strong coupling scenario are shown in the top panel: $A \to O$ (left) and $O \to A$ (right). Results for the weak scenario are shown in the bottom panel: $A \to O$ (left) and $O \to A$ (right). Error bars are reported on each data point. Red dashed lines indicate running mean window sizes of 100 and 1000 days.

Similarly to what we did at the end of the previous chapter, we now compare the results obtained with the interregional and the multivariate approaches. In the strong coupling case, while the interregional TE approach shows a decreasing trend for the $A \to O$ causal link, we had observed an apparent intensification of the causal link in the multivariate approach. Instead, we confirm the stabilization of the pattern in the $o_j \to a_i$ area with increasing smoothing window value, as we see that the two values belong in a region of minima in the interregional approach. For the LKIF, the two approaches seem to find similar results, although in the $o_j \to a_i$ region we don't clearly observe the drop shown by the interregional approach.

Moving to the weak coupling scenario, we confirm the increase of TE values in the

$a_j \to o_i$ region, also observed with the interregional approach. The high values reached with the interregional approach in the $O \to A$ panel, instead, hardly appear with the multivariate approach. The two LKIF approaches are more in line, as we observe stable low LKIF values in the $a_j \to o_i$ region as well as in the $A \to O$ panel, and we also observe a slight increase in the opposite direction, with both methods.

Introducing a time-lag between ocean and atmosphere

We analyze the output of the interregional TE and LKIF with the application of increasing time-lag between the atmosphere and the ocean. The LKIF has not been designed to work with time lags by default, although they can be used in principle (X. San Liang, 2021). To implement the time-lagged analysis, we can shift in time the time series of the leading variable and recompute LKIF between oceanic and atmospheric variables based on the lagged time series. We follow the same approach for the TE. For positive values of the x-axis the atmosphere is pushed forward with respect to the ocean, while it lags behind it for negative x values. The goal of this analysis is to study how the causality in the $A \to O$ and $O \to A$ directions changes if we increase the temporal separation between the oceanic and atmospheric dynamics. We reported the results of this analysis in Figure 4.11 for the TE and in Figure 4.12 for the LKIF.

Notably, the behavior for positive and negative values of the atmospheric lag with respect to the ocean appears to be approximately symmetric. In general the presence of a lag does not change the strength of the retrieved causal link up until high values of the time lag (both positive and negative). An exception to this is represented by the weak coupling LKIF scenario, where the evaluated links are not symmetric with respect to the y-axis for small to medium values of the time lag, and stabilize only after a lag of several months. In the strong coupling case we observe a drop followed by an increase around 100 months of lag, in the results of both methods.

An overall look at the results of both methods says something similar to what we observed in the multivariate analysis in Figure 4.8 : the two methods agree in finding strong $O \to A$ causal link in the strong coupling case and both retrieve very small to non-significant link in the weak $O \to A$ scenario. However, the TE has generally higher resolution and finds significant links also in the strong and the weak coupling $A \to O$ case, which are both non-significant for the LKIF.

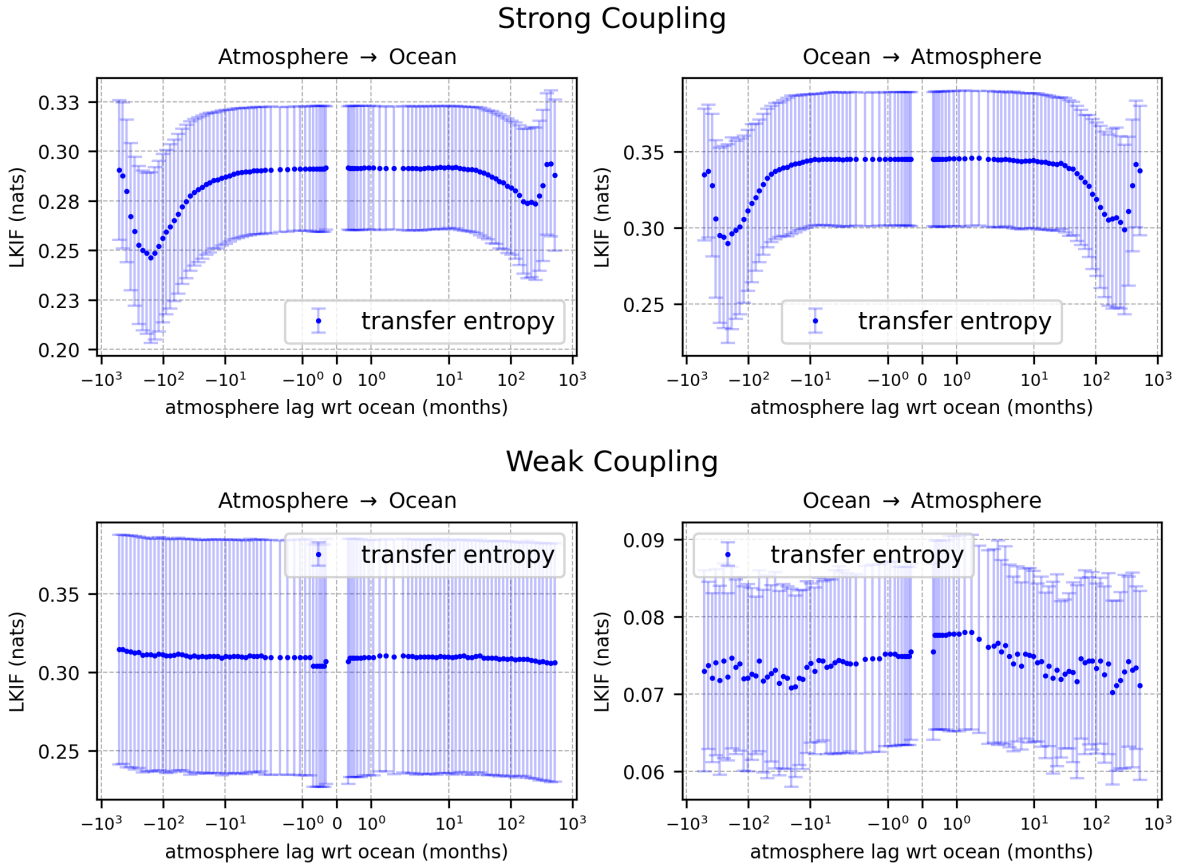


Figure 4.11: Analysis of the ocean-atmosphere coupling with the interregional TE. The lag between the atmosphere and the ocean, in months, increases exponentially on the x axis, shown in log scale. For positive values of the x axis the atmosphere lags forward with the respect to the ocean, and it lags behind the ocean for negative x values. Results for the strong coupling scenario are shown in the top panel: $A \to O$ (left) and $O \to A$ (right). Results for the weak scenario are shown in the bottom panel: $A \to O$ (left) and $O \to A$ (right). Error bars are reported on each data point. Red dashed lines indicate time lags of 100 and 1000 days.

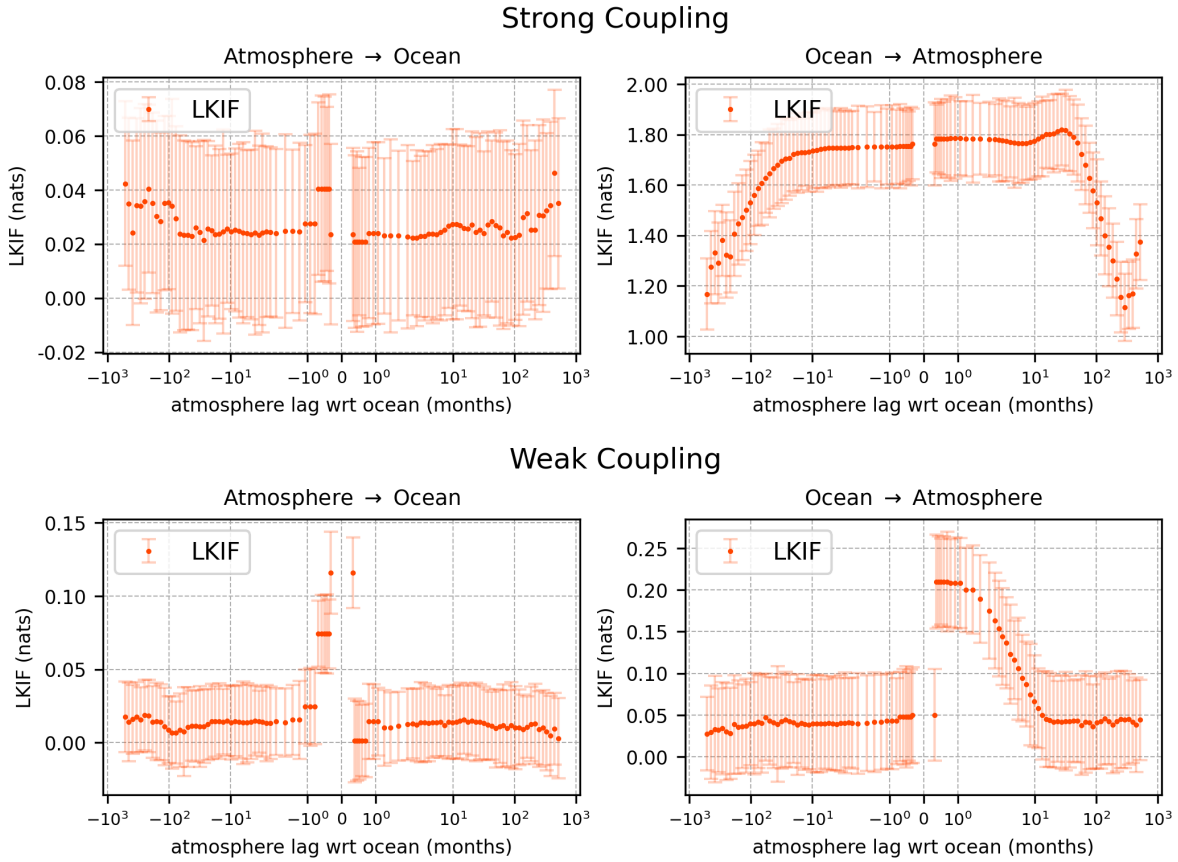


Figure 4.12: Analysis of the ocean-atmosphere coupling with the interregional LKIF. The lag between the atmosphere and the ocean, in months, increases exponentially on the x axis, shown in log scale. For positive values of the x axis the atmosphere lags forward with the respect to the ocean, and it lags behind the ocean for negative x values. Results for the strong coupling scenario are shown in the top panel: $A \to O$ (left) and $O \to A$ (right). Results for the weak scenario are shown in the bottom panel: $A \to O$ (left) and $O \to A$ (right). Error bars are reported on each data point. Red dashed lines indicate time lags of 100 and 1000 days.

Key points

- A causal analysis is performed on the 36 spectral modes of MAOOAM through the Liang-Kleeman Information Flow (LKIF) and the Transfer Entropy (TE), using *multivariate* and *interregional* approaches.
- Multivariate analysis is particularly effective for examining mode-to-mode causal links within a multivariate system. However, its application is limited by the system's high dimensionality and the inherent challenges in interpreting causal links in the spectral formulation of MAOOAM.

- *Multivariate* TE and LKIF clearly underline the change from strong to weak coupling in the system and successfully retrieve different causal information in the *atmosphere* $\rightarrow$ *ocean* versus *ocean* $\rightarrow$ *atmosphere* directions. Significant local differences are retrieved in the output of the two methods, and more substantial ones with respect to the correlation.

- *Interregional causal analysis* highlights the advantage of causal methods over a symmetric correlation measure, as it marks a clear difference in the causal information between the *ocean* $\rightarrow$ *atmosphere* and the *atmosphere* $\rightarrow$ *ocean* directions. The TE is more sensitive, as it retrieves significant causal links also in the directions which the LKIF marks as non-significant.

- *Smoothing the Atmospheric Signal.* Reducing fast fluctuations in the atmospheric data makes the ocean–atmosphere connection emerge more clearly.

- *Time-lag analysis.* Shifting the timing between ocean and atmosphere signals shows that the basic directional influences stay consistent unless the lag is very large (about 10 years), except for the weak coupling case of the atmosphere that shows higher instability.

- *Connections to the Physics of MAOOAM.* The results match physical expectations: when the ocean-atmosphere coupling is strong, the ocean damps the faster atmospheric variability; in the weak coupling case, the atmosphere increasingly drives the ocean. Significant foundational differences between TE and LKIF limit further interpretations.

Conclusions

In this work, we have explored the foundations of causality with a focus on the challenges of performing causal inference from time series. We examined both its philosophical and scientific grounds and described the two main approaches that have emerged over the last century. One approach is based on a mathematically rigorous framework largely introduced by Judea Pearl (Spirtes, C. Glymour, and Scheines, 2001; Judea Pearl, 2009; Peters, Janzing, and Schlkopf, 2017), which employs causal graphs and the notion of interventions. The other approach is rooted in data-driven methods aimed at inferring causal relations from observations and time series, with Granger causality (Granger, 1969) being its best-known representative.

These two lines of work have evolved considerably in recent decades. On one side, algorithms based on conditional independencies have advanced the interventional framework, while on the other, several data-driven methods rooted in information theory and dynamical systems have been developed. In our work, we paid special attention to the information-theoretic method Transfer Entropy (Schreiber, 2000; J. Lizier et al., 2011) and to the Liang–Kleeman Information Flow (X. San Liang, 2021). We also described dynamical systems-based methods such as Convergent Cross Mapping (CCM) (Sugihara et al., 2012) and Empirical Model Reduction (EMR) (Dmitri Kondrashov, Chekroun, and Michael Ghil, 2015). Current research aims to generalize linear and/or bivariate methods to higher-order nonlinear versions, to capture complex interactions more faithfully (Marinazzo, Pellicoro, and Stramaglia, 2008; Stramaglia et al., 2024). In addition, recent research has explored data-driven methods that attempt to reconnect with the interventional definition of causality, for example through the fluctuation-response formalism (Baldovin, Cecconi, and Vulpiani, 2020). Theoretical studies have examined possible common ground between these methods. For instance, it has been argued that Granger causality and Transfer Entropy are equivalent up to a factor of 2 for Gaussian variables (Barnett, Barrett, and A. K. Seth, 2009), and that these methods can be recast in terms of testing for conditional independencies, much like algorithms based on the formal interventional framework (Peters, Janzing, and Schlkopf, 2017; J. Runge, 2018). However, a broader generalization has not yet been achieved.

Different methods do not always yield convergent results, particularly with respect to estimated causal effects and reconstructed causal graphs. This may be due to the

different theoretical foundations of the methods and to their specific implementation characteristics - such as linear versus nonlinear, or bivariate versus multivariate - and it underscores the need to choose the appropriate method for a particular type of data (Krakovská et al., 2018).

Causal methods have seen increasing application in climate studies, driven by the belief that a causal perspective offers valuable insights for disentangling complex dynamical relationships and attributing causes to events (A. Hannart et al., 2016; Alexis Hannart, Carrassi, et al., 2016), particularly compelling tasks in this field. This is particularly true in climate science, where we deal with one of the most fascinating, complex, and multiscale nonlinear systems (Michael Ghil and Lucarini, 2020).

Open methodological challenges in applying causality to the climate system include handling incomplete domain knowledge and developing methods that can cope with high degrees of complexity and missing data (Jakob Runge, Gerhardus, et al., 2023). A growing line of research aims to integrate machine learning with causality. For example, incorporating large language models into the framework of causal hypothesis testing - combined with expert knowledge and data - could yield deeper insights into the causal structure of the climate system and improve data-driven forecasting (Cohrs et al., 2025).

We have positioned our work within the field of causal studies in climate science by investigating ocean–atmosphere coupled variability. Although this problem has been extensively studied through observations and modeling, the direction of the coupling between the ocean and the atmosphere remains not fully understood. In this work, we have investigated the causal interactions of the ocean–atmosphere system using the low-order coupled spectral model MAOOAM (De Cruz, J. Demaeyer, and S. Vannitsem, 2016) and two methods: Transfer Entropy (TE) and Liang–Kleeman Information Flow (LKIF).

First, we examined mode-to-mode causal relationships via multivariate approaches, which proved effective at revealing how the system transitions from strong to weak coupling and at distinguishing the directional influences between the atmosphere and the ocean. The two methods qualitatively agree but present significant differences. The high dimensionality of MAOOAM and the challenges inherent in interpreting causal links within its spectral formulation limit this approach.

To complement our findings, we adopted an interregional perspective by analyzing the ocean and atmosphere as distinct multivariate subsystems. This interregional causal analysis highlights the advantage of causal methods over symmetric correlation measures, as it clearly differentiates the two causal directions. Both methods retrieve a strong $ocean \rightarrow atmosphere$ causal link in the strong coupling scenario, while the $atmosphere \rightarrow ocean$ becomes predominant in the weak coupling scenario. Notably, TE demonstrated higher sensitivity by retrieving significant causal links in both directions, whereas LKIF sometimes marked weaker links as non-significant. Furthermore, reducing fast fluctuations in the atmospheric data (smoothing) made the ocean–atmosphere connection emerge more clearly, and a time-lag analysis showed that the basic directional

influences remain consistent unless the lag becomes very large (around 10 years), with the weak coupling case exhibiting higher instability.

Overall, the results obtained with TE and LKIF show some relevant differences, likely due to their distinct theoretical foundations. Nevertheless, they qualitatively align with physical expectations: under strong coupling, the ocean dampens rapid atmospheric variability, while under weak coupling, the atmosphere increasingly drives the ocean. These findings warrant further investigation.

The analyses presented in this work are limited to the spectral modes of MAOOAM. A natural extension of this research would be to investigate the causal dependencies in real space by integrating the model to obtain atmospheric and oceanic physical fields, thereby providing more straightforward physical intuition.

Further insight into the ability of TE and LKIF to capture physical information could be obtained by studying higher-dimensional models. Moreover, applying these methods to observational or reanalysis time series would allow us to test them under conditions of maximum complexity and directly compare their outputs with real-world data.

The analysis could also be extended to include other methods. An evolution of this work—not implemented due to time constraints—would involve testing the EMR and CCM methods. This would help assess whether dynamical systems-based approaches offer greater insight into the causal dependencies of the system, as they operate in a framework that closely aligns with the system's own dynamics.

Acknowledgements

I would like to deeply thank my supervisor at the University of Bologna, Alberto Carrassi, for his constant presence throughout the work, for introducing me to the other supervisors - Michael Ghil, Stéphane Vannitsem, and Daniele Marinazzo - and for his great advice and genuine interest in the topics we studied, which I find inspiring.

I am grateful to Sabrina Speich, who made the five-month bilateral exchange between the Collegio Superiore of the University of Bologna and the École Normale Supérieure in Paris possible, and to Fabio di Andrea, who welcomed me at the Department of Geosciences and ensured that I was doing well at the university during my stay.

I am exceptionally thankful to Michael Ghil, who introduced me to a beautiful way of making science and to a universe of stories and anecdotes that I think should someday end up in a book. I thank him for sharing his knowledge with me and for coming up with the idea of studying the dog-and-tail problem in the experiments section of this work.

I sincerely thank Stéphane Vannitsem for introducing me to MAOOAM, for the insightful exchanges on causality, and for his hospitality at the Institut Royal Météorologique in Bruxelles. I also thank David Docquier for helping me set up the Liang-Kleeman Information Flow.

Very big thanks to Daniele Marinazzo, who shared many ideas on causality, devoted plenty of time to thorough discussions, suggested the use of Transfer Entropy in the experiments, and cleared my mind on the state of the art of this field.

I also thank the University of Bologna, the Collegio Superiore, and the École Normale Supérieure, as well as the many great people with whom I had the chance to exchange ideas about this work.

Appendix A

Derivation of the MAOOAM spectral model

Set of Basis Functions

The atmospheric field is defined in a zonally periodic channel with no-flux boundary conditions in the meridional direction: $\partial \cdot / \partial x \equiv 0$ at the meridional boundaries. The fields are projected on Fourier modes that take into account these boundary conditions, with a *Galerekin expansion*.

Galerkin expansion

The Galerkin expansion is spectral method belonging to the general class of Weighted Residual Methods (Oliveira and Rodrigues, 2011), developed to solve numerically differential equations, for which the solution is approximated by a finite series expansion in some basis of functions. The series expansion is such that some quantity - a measure of error or residual - which should be exactly zero is forced to be zero in an approximate sense. Let us give an example. We consider a function $f(x)$ that is approximated by the series expansion $f_N(x) = \sum_{i=0}^{N} a_i F_i(x)$, where a_i are the unknown coefficients and $F_i(x)$ are the basis functions. The coefficients are determined by requiring that the residual $R_N \equiv f(x) - f_N(x)$ is set to zero in some approximate sense, that is

$$\langle R_N, \phi_j \rangle = \int_a^b R_N \phi_j w \mathrm{d}x = 0, \tag{A.1}$$

where $j = 0, ..., N$, $\phi_j(x)$ are the test functions and w is the weight. The choice of the spectral function defines the spectral method. The Galerkin method is retrieved when $\phi_i(x) = F_i(x)$, with the basis functions satisfying the boundary conditions. Spectral methods transform a partial differential equation in a finite set of ordinary differential equations, with considerable economy of the computational resources Oliveira and Rodrigues, 2011.

Spectral expansion of MAOOAM

The basis functions for the atmosphere chosen in Stéphane Vannitsem, Jonathan Demaeyer, et al. (2015) are defined as follows

$$\begin{aligned} F^A_P(x,y) &= \sqrt{2}\cos(Py), \\ F^K_{M,P}(x,y) &= 2\cos(Mnx)\sin(Py), \\ F^L_{H,P}(x,y) &= 2\sin(Hnx)\sin(Py), \end{aligned} \tag{A.2}$$

where M, H, and P are integers. The apices A, K, L come from the original naming of the functions assigned in Lorenz (1963). The functions with the apex A represent the zonal component of the flow. Functions with apices K and L represent superposed waves, and are called waved of the first mode. The coordinates x and y are the horizontal adimensionalised coordinates, rescaled by dividing the dimensional coordinates by the characteristic length L (L_x and L_y for zonal and meridional extents, respectively). The model's domain is then defined as:

$$0 \leq x \leq \frac{2\pi}{n}, \quad 0 \leq y \leq \pi, \tag{A.3}$$

where n is the aspect ratio between its meridional and zonal extents L_y (L_y) and L_x (L_x).

In the application we will refer to the functions as $F_1(x,y), ..., F_{n_a}(x,y)$, numbering them along increasing values f $M = H$ and then P, where n_a is the number of modes in the spectral expansion. In the version of the model we used in this work, we have $n_a = 10$.

For the basis functions we have

$$\nabla^2 F_i(x,y) = -a_i^2 F_i(x,y) \tag{A.4}$$

with eigenvalues $a_i^2 = P_i^2 + n^2 M_i^2$ or $a_i^2 = P_i^2 + n^2 H_i^2$. The atmospheric Fourier modes are orthonormal with respect to the inner product on the domain.

We now turn to the oceanic basis functions. Both oceanic fields ψ_0 and δT_0 are defined in a closed basin with no-flux boundary conditions at both meridional and zonal boundaries. This condition implies that the modes will only involve sine functions. The set of basis functions counts n_o elements, corresponding to the number of oceanic modes one wants to implement. In this work we used $n_o = 8$. The basis functions read

$$\psi_{H_o,P_o}(x,y) = 2sin(\frac{H_o n}{2}x)sin(P_o y) \tag{A.5}$$

with integer values of H_o, P_o. The basis functions are numbered along increasing values of H_o and then P_o. For these functions we have

$$\nabla^2 \phi_i(x,y) = -m_i^2 \phi_i(x,y) \tag{A.6}$$

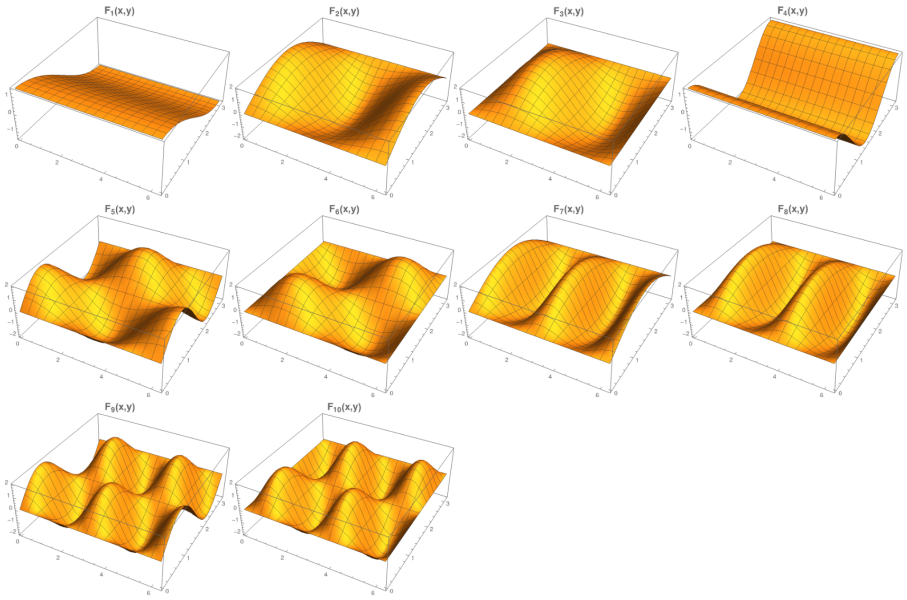


Figure A.1: The first 10 basis functions for the atmosphere, F_i, evaluated on the nondimensional domain of the model. Courtesy of Jonathan Demaeyer, Cruz, and Stéphane Vannitsem (2020)

with eigenvalues $m_i^2 = P_{o,i}^2 + n^2 H_{o,i}^2/4$. These Fourier modes are orthonormal with respect to the inner product over the domain. Atmospheric and oceanic basis functions, instead, are not orthonormal to each other.

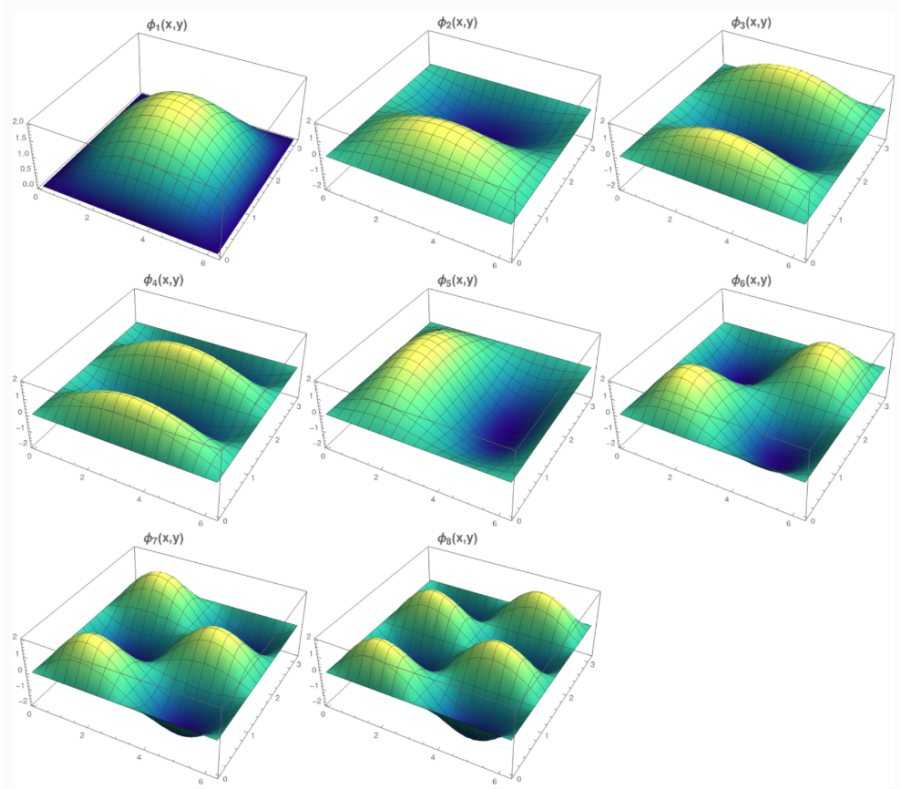


Figure A.2: The first 8 basis functions for the ocean, ϕ_i, evaluated on the nondimensional domain of the model. Courtesy of Jonathan Demaeyer, Cruz, and Stéphane Vannitsem (2020).

Fields expansions

$$
\begin{aligned}
\psi_a(x,y) &= \sum_{i=1}^{n_a} \psi_{a,i} F_i(x,y), \\
\theta_a(x,y) &= \sum_{i=1}^{n_a} \theta_{a,i} F_i(x,y), \\
\delta T_a(x,y) &= \sum_{i=1}^{n_a} \delta T_{a,i} F_i(x,y), \\
\delta T_a(x,y) &= \frac{2f_0}{R} \sum_{i=1}^{n_a} \theta_{a,i} F_i(x,y), \\
\psi_o(x,y) &= \sum_{j=1}^{n_o} \psi_{o,j} \left(\phi_j(x,y) - \overline{\phi_j}\right), \\
\delta T_o(x,y) &= \sum_{j=1}^{n_o} \delta T_{o,j} \phi_j(x,y).
\end{aligned}
\tag{A.7}
$$

In the expansion for ψ_o, a term $\bar{\phi_j}$ is added to the oceanic basis function ϕ_j in order to get a zero spatial average: this ensures mass conservation in the ocean, and does not affect the dynamics. The mass conservation is automatically satisfied for ψ_a, instead, as the spatial average of the atmospheric basis functions is zero.

The short-radiation for atmosphere and ocean, and the vertical velocity in the atmosphere are decomposed as well.

Ordinary differential equations

The fields, parameters and variables are non-dimensionalized by divinding time by f_0^{-1}, distance by the characteristic length scale L, pressure by the pressure difference Δp, temperature by $f_0^2 L^2/R$, and streamfunction by $L^2 f_0$. The result of the adimensionalization is the identification of the field θ_a with δT_a: $2\theta_a \equiv \delta T_a$.

Having adimensionalized, one can write the ordinary differential equations of the truncated model. These equations are at most quadratic in the fields $\psi_{a,i}, \theta_{a,i}, \psi_{o,i}, \omega_i, \delta T_{o,i}$. The vertical velocity field ω_i can be eliminated. These equations include coefficients that result from the Galerkin expansion. The coefficients are correspond to the inner products of the Fourier modes F_i with F_i, ϕ_i with ϕ_i and F_i with ϕ_i. We will not write explicitly the resulting ODEs of the model, but they can be explicitly found in the model's documentation page, here.

The obtained ordinary differential equations can then be integrated numerically, so that the solution for the Fourier components of the fields are obtained as time series. The numerical integration is performed through a Runge-Kutta method; the scheme is detailed here.

The `qgs` implementation also gives the possibility to obtain several physical fields of the model, by using the formulae A.7. One can obtain the atmospheric and oceanic streamfunctions and temperature fields, the wind velocities, and several other physical components of the model. The spatial resolution of these fields can be chosen by the user.

Appendix B

Kraskov-Stögbauer-Grassberger technique for TE estimation

The Kraskov-Stögbauer-Grassberger technique computes Transfer Entropy via Conditional Mutual Information. In the following we explain the rationale of this technique.

Estimation of MI

Consider a set of N bivariate measurements, $z_i = (x_i, y_i),\ i = 1, ..., N$, which are assumed to be independent identically distributed (i.i.d.) realizations of a random variable $Z = (X, Y)$ with density $\mu(x, y)$. Here, x and y can be scalar or vector quantities. The density is assumed to be a proper smooth function, so that the integrals written below exist in some sense. The Mutual Information is defined as

$$MI(X, Y) = \int\int \mathrm{d}x\mathrm{d}y\ \mu(x, y) \log \frac{\mu(x, y)}{\mu_x(x)\mu_y(y)}, \tag{B.1}$$

where the terms at the denominator are the marginal densities of X and Y. The MI can be decomposed iteratively in the MI for subsystems, leveraging its definition. So one can decompose $MI(X_1...X_n)$ for any $n > 2$ and for any partitioning of the set $(X_1...X_n)$ into the MI between elements within one cluster and MI between clusters. This property is numerically useful and it derives from the information-theoretic ground of MI, representing an advantage of this method over others that do not have these theoretical grounds. in case of changes of high-dimensional redundancies under reparametrization of some subspace, since it is sufficient to calculate the MIs in that subspace only.

There are different approaches to practically compute MI. The KSG technique estimates MI from k-nearest neighbor statistics. This approach is frequently used to compute estimators for Shannon entropy, and it has been extended to the estimation of MI in Kraskov, Stögbauer, and Grassberger, 2004.

Assume some metrics to be given on the spaces spanned by X, Y and $Z = (X, Y)$. We can then rank, for each point $z = (x, y)$, its neighbors by the maximum norm distance

$\|z - z'\| = \max\{\|x - x'\|, \|y - y'\|\}$, while any norm can be used for $\|x - x'\|$ and $\|y - y'\|$. Let us denote by $\epsilon(i)/2$ the distance from z_i to its k^{th} neighbor, and by $\epsilon_x(i)/2$ and $\epsilon_y(i)/2$ the distance between the same points projected into the X and Y subspaces. We have $\epsilon(i) = \max\{\epsilon_x(i), \epsilon_y(i)\}$. One can then compute the MI following two different algorithms, detailed in Kraskov, Stögbauer, and Grassberger, 2004.

Estimation algorithms

The first algorithm to compute MI counts the number $n_x(i)$ of points x_j whose distance from x_i is strictly less than $\epsilon(i)/2$, and similarly for y instead of x. This is shown in Figure B.3(a). Notice that $\epsilon(i)$ is a random fluctuating variable, and therefore also $n_x(i)$ and $n_y(i)$ fluctuate. We denote by $\langle ... \rangle$ averaged both over all $i \in [1, ..., N]$ and over all realizations of the random samples,

$$\langle ... \rangle = N^{-1} \sum_{i=1}^{N} \mathbb{E}[...(i)]. \tag{B.2}$$

The estimate for MI is then

$$MI^{(1)}(X, Y) = \psi(k) - \langle \psi(n_x + 1) + \psi(n_y + 1) \rangle + \psi(N). \tag{B.3}$$

In the second algorithms $n_x(i)$ and $n_y(i)$ by the number of points are replaced by $\|x_i - x_j\| \leq \varepsilon_x(i)/2$ and $\|y_i - y_j\| \leq \varepsilon_y(i)/2$. The estimate for MI is then

$$MI^{(2)}(X, Y) = \psi(k) - 1/k - \langle \psi(n_x) + \psi(n_y) \rangle + \psi(N). \tag{B.4}$$

The algorithms estimate the mutual information in *nats*. In general, both formulas give very similar results. For the same k, Eq. B.3 gives slightly smaller statistical errors but have larger systematic errors. The latter algorithm should be applied if interested in very high dimensional systems.

The adaptation of this method to compute the Transfer Entropy has been derived in Wibral, Vicente, and Lindner, 2014.

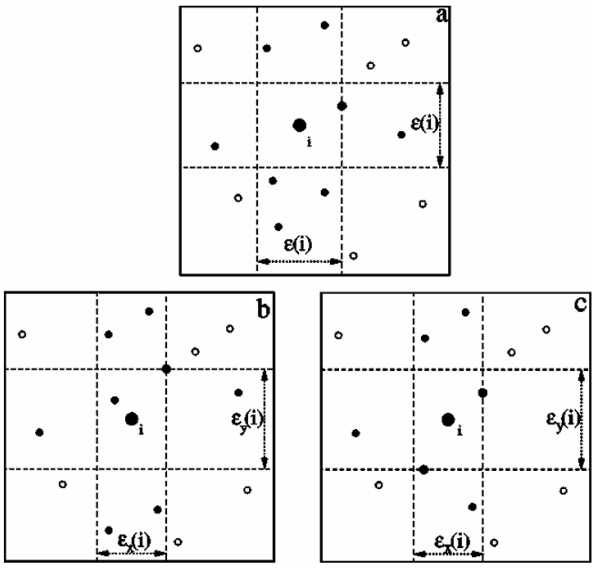


Figure B.1: Panel (a): Determination of $\epsilon(i)$, $n_X(i)$, and $n_Y(i)$ in the first algorithm, for $k = 1$ and some fixed i. In this example, $n_X(i) = 5$ and $n_Y(i) = 3$. Panels (b),(c): Determination of $\epsilon_X(i)$, $\epsilon_Y(i)$, $n_X(i)$, and $n_Y(i)$ in the second algorithm for $k = 2$. Panel (b) shows a case in which $\epsilon_X(i)$ and $\epsilon_Y(i)$ are determined by the same point, while panel (c) shows a case in which they are determined by different points. Courtesy of Kraskov, Stögbauer, and Grassberger, 2004.

Appendix C

Estimation of the Liang-Kleeman Information Flow

C.1 Estimation of the bivariate information flow

The following theorem, proven in X. San Liang, 2014, gives the maximum likelihood estimator for the information flow in a two-dimensional system ($d = 2$).

Theorem C.1.1. *Given two time series X_1 and X_2, under the assumption of a linear model with additive noise, the maximum-likelihood estimator of 2.21 is*

$$\hat{T}_{2\to 1} = \frac{C_{11}C_{12}C_{2,d1} - C_{12}^2 C_{1,d1}}{C_{11}^2 C_{22} - C_{11}C_{12}^2} \tag{C.1}$$

where C_{ij} is the sample covariance between X_i and X_j, and $C_{i,dj}$ is the sample covariance between X_i and a series derived from X_j using the Euler forward differencing scheme: $\dot{X}_{j,n} = (X_{j,n+k} - X_{j,n})/(d\Delta t)$, with $k \geq 1$ some integer.

C.2 Estimation of the multivariate information flow

The possibility to compute the information flow for multivariate time series has been implemented in X. San Liang, 2021, which develops a formalism for time series. We give here the details for the estimation of 2.20, given observations of the d components. Without loss of generality, it suffices to examine $T_{2\to 1}$. We hence examine a linear stochastic differential equation:

$$\mathrm{d}\mathbf{X} = \mathbf{f} + \mathbf{A}\mathbf{X}\mathrm{d}t + \mathbf{B}\mathrm{d}\mathbf{W}, \tag{C.2}$$

where $\mathbf{f}$ is a constant vector, and $\mathbf{A} = (a_{ij})$ and $\mathbf{B} = (b_{ij})$ are constant matrices. $\mathbf{B}$ is assumed to be diagonal to simplify the calculations and as it is a quite reasonable assumption X. San Liang, 2021.

The estimation of the information flow under the observation of d time series is again obtained with maximum-likelihood estimation. Suppose that the series are equal-distanced with a time stepsize Δt, and let N be the sample size. Consider an interval $[n\Delta t, (n+1)\Delta t]$, and let the transition pdf be $\rho(X_{n+1}|X_n;\theta)$, where θ stands for the vector of parameters to be estimated. So, the log likelihood is:

$$\ell_N(\theta) = \sum_{n=1}^{N} \log \rho(\mathbf{X}_{n+1}|X_n;\theta) + \log \rho(\mathbf{X}_1). \tag{C.3}$$

Following some passages reported in X. San Liang, 2014, one obtains the maximum likelihood estimators:

$$\begin{aligned} \hat{a}_{1i} &= \frac{1}{\det \mathbf{C}} \sum_{j=1}^{d} \Delta_{ij} C_{jd,1}, \\ \hat{g}_{11} &= \frac{Q_{N,1}\Delta t}{N}, \\ \hat{f}_1 &= \overline{\dot{X}_1} - \sum_{i=1}^{d} \hat{a}_{1i} \overline{X}_i, \end{aligned} \tag{C.4}$$

where

$$\begin{aligned} C_{ij} &= \overline{(X_i - \overline{X}_i)(X_j - \overline{X}_j)}, \\ C_{i,dj} &= \overline{(X_i - \overline{X}_i)(\dot{X}_j - \overline{\dot{X}_j})}, \end{aligned} \tag{C.5}$$

are the sample covariances, Δ_{ij} are the cofactors of the matrix $\mathbf{C} = (C_{ij})$, and

$$\begin{aligned} Q_{N,1} &= \sum_{n=1}^{N} \left[\dot{X}_{1,n} - \left(\hat{f}_1 + \sum_{j=1}^{d} \hat{a}_{1j} X_{j,n} \right) \right]^2 \\ &= \sum_{n=1}^{N} \left[(\dot{X}_{1,n} - \overline{\dot{X}_1}) - \sum_{i=1}^{d} \hat{a}_{1i} (X_{i,n} - \overline{X}_i) \right]^2 \\ &= N(C_{d1,d1} - 2\sum_{i=1}^{d} \hat{a}_{1i} C_{d,1i} + \sum_{i=1}^{d}\sum_{j=1}^{d} \hat{a}_{1i}\hat{a}_{1j} C_{ij}). \end{aligned} \tag{C.6}$$

By 2.22, this yields an estimator of the information flow from X_2 to X_1, which concludes our calculations:

$$\hat{T}_{2\to 1} = \frac{1}{\det \mathbf{C}} \cdot \sum_{j=1}^{d} \Delta_{2j} C_{j,d1} \cdot \frac{C_{12}}{C_{11}}, \tag{C.7}$$

where $C_{j,d1}$ is the sample covariance between X_j and the derived series $\dot{X}_1$. When $d = 2$ the formula coherently reduces to 2.21, the 2D estimator.

Bibliography

Arnold, Andrew, Yan Liu, and Naoki Abe (2007). "Temporal causal modeling with graphical granger methods". In: *Proceedings of the 13th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. KDD '07. San Jose, California, USA: Association for Computing Machinery, pp. 66–75. ISBN: 9781595936097. DOI: 10.1145/1281192.1281203.

Ay, Nihat and Daniel Polani (2008). "Information Flows in Causal Networks". In: *Advances in Complex Systems* 11.01, pp. 17–41. DOI: 10.1142/S0219525908001465.

Baldovin, Marco, Fabio Cecconi, and Angelo Vulpiani (Dec. 2020). "Understanding causation via correlations and linear response theory". In: *Phys. Rev. Res.* 2 (4), p. 043436. DOI: 10.1103/PhysRevResearch.2.043436.

Barnett, Lionel, Adam B. Barrett, and Anil K. Seth (Dec. 2009). "Granger Causality and Transfer Entropy Are Equivalent for Gaussian Variables". In: *Physical Review Letters* 103.23. ISSN: 1079-7114. DOI: 10.1103/physrevlett.103.238701.

Barsugli, Joseph J. and David S. Battisti (1998). "The Basic Effects of Atmosphere–Ocean Thermal Coupling on Midlatitude Variability". In: *Journal of the Atmospheric Sciences* 55.4, pp. 477–493. DOI: 10.1175/1520-0469(1998)055<0477:TBEOAO>2.0.CO;2.

Batchelor, G.K. (1967). *An Introduction to Fluid Dynamics*. Cambridge Mathematical Library. Cambridge University Press. ISBN: 9780521098175.

Bossomaier, Terry et al. (Jan. 2016). *An introduction to transfer entropy: Information flow in complex systems*. English. Includes bibliographical references and index. Springer International Publishing. ISBN: 9783319432212. DOI: 10.1007/978-3-319-43222-9.

Bueso, Diego, Maria Piles, and Gustau Camps-Valls (Dec. 2020). "Explicit Granger causality in kernel Hilbert spaces". In: *Physical Review E* 102.6. ISSN: 2470-0053. DOI: 10.1103/physreve.102.062201.

Carrassi, A. et al. (2017). "Estimating model evidence using data assimilation". In: *Quarterly Journal of the Royal Meteorological Society* 143.703, pp. 866–880. DOI: https://doi.org/10.1002/qj.2972.

Chu, Tianjiao, David Danks, and Clark Glymour (2005). "Data Driven Methods for Nonlinear Granger Causality: Climate Teleconnection Mechanisms". In.

Cohrs, Kai-Hendrik et al. (Jan. 2025). "Large language models for causal hypothesis generation in science". In: *Machine Learning: Science and Technology* 6.1, p. 013001. DOI: 10.1088/2632-2153/ada47f.

Contu, Francesco (2024). "Correlazione e causalità: le radici fisiche di una distinzione cruciale". PhD thesis. University of Bologna.

De Cruz, L., J. Demaeyer, and S. Vannitsem (2016). "The Modular Arbitrary-Order Ocean-Atmosphere Model: MAOOAM v1.0". In: *Geoscientific Model Development* 9.8, pp. 2793–2808. DOI: 10.5194/gmd-9-2793-2016.

Dechter, Rina, Hector Geffner, and Joseph Y. Halpern (2010). *Heuristics, Probability and Causality. A Tribute to Judea Pearl*. College Publications. ISBN: 1904987656.

Delgado-Bonal, Alfonso et al. (2020). "Analyzing changes in the complexity of climate in the last four decades using MERRA-2 radiation data". In: *Scientific Reports* 10.1, p. 922. ISSN: 2045-2322. DOI: 10.1038/s41598-020-57917-8.

Demaeyer, Jonathan, Lesley De Cruz, and Stéphane Vannitsem (2020). "qgs: A flexible Python framework of reduced-order multiscale climate models". In: *Journal of Open Source Software* 5.56, p. 2597. DOI: 10.21105/joss.02597.

Dhifaoui, Zouhaier et al. (2023). "Exploring the effect of climate risk on agricultural and food stock prices: Fresh evidence from EMD-Based variable-lag transfer entropy analysis". In: *Journal of Environmental Management* 326, p. 116789. ISSN: 0301-4797. DOI: https://doi.org/10.1016/j.jenvman.2022.116789.

Dijkstra, Henk A. and Michael Ghil (2005). "Low-frequency variability of the large-scale ocean circulation: A dynamical systems approach". In: *Reviews of Geophysics* 43.3. DOI: https://doi.org/10.1029/2002RG000122.

Docquier, D., S. Vannitsem, and A. Bellucci (2023). "The rate of information transfer as a measure of ocean–atmosphere interactions". In: *Earth System Dynamics* 14.3, pp. 577–591. DOI: 10.5194/esd-14-577-2023.

Docquier, D., S. Vannitsem, F. Ragone, et al. (2022). "Causal Links Between Arctic Sea Ice and Its Potential Drivers Based on the Rate of Information Transfer". In: *Geophysical Research Letters* 49.9. e2021GL095892 2021GL095892, e2021GL095892. DOI: https://doi.org/10.1029/2021GL095892.

Ebert-Uphoff, Imme and Yi Deng (2012). "Causal Discovery for Climate Research Using Graphical Models". In: *Journal of Climate* 25.17, pp. 5648–5665. DOI: 10.1175/JCLI-D-11-00387.1.

Eichler, Michael (Feb. 2011). "Graphical modelling of multivariate time series". In: *Probability Theory and Related Fields* 153.1–2, pp. 233–268. ISSN: 1432-2064. DOI: 10.1007/s00440-011-0345-8.

Faes, Luca, Giandomenico Nollo, and Alberto Porta (2012). "Non-uniform multivariate embedding to assess the information transfer in cardiovascular and cardiorespiratory variability series". In: *Computers in Biology and Medicine* 42.3. Computing complexity in cardiovascular oscillations, pp. 290–297. ISSN: 0010-4825. DOI: https://doi.org/10.1016/j.compbiomed.2011.02.007.

Fisher, R. A. (1992). "Statistical Methods for Research Workers". In: *Breakthroughs in Statistics: Methodology and Distribution*. Ed. by Samuel Kotz and Norman L. Johnson. New York, NY: Springer New York, pp. 66–70. ISBN: 978-1-4612-4380-9. DOI: 10.1007/978-1-4612-4380-9_6.

Galilei, G. (1638). *Discorsi e dimostrazioni matematiche intorno a due nuove scienze attenenti alla mecanica ed i movimenti locali*. Gli Elsevirii.

Ghil, Michael (July 2019). "A Century of Nonlinearity in the Geosciences". en. In: *Earth and Space Science* 6.7, pp. 1007–1042. ISSN: 2333-5084, 2333-5084. DOI: 10.1029/2019EA000599.

Ghil, Michael and Valerio Lucarini (July 2020). "The physics of climate variability and climate change". In: *Reviews of Modern Physics* 92.3. ISSN: 1539-0756. DOI: 10.1103/revmodphys.92.035002.

Granger, Clive William John (1969). "Investigating causal relations by econometric models and cross-spectral methods". In: *Essays in econometrics: collected papers of Clive W. J. Granger*.

Hagan, Daniel Fiifi Tawia et al. (2019). "A Time-Varying Causality Formalism Based on the Liang–Kleeman Information Flow for Analyzing Directed Interactions in Nonstationary Climate Systems". In: *Journal of Climate* 32.21, pp. 7521–7537. DOI: 10.1175/JCLI-D-18-0881.1.

Hannart, A. et al. (2016). "Causal Counterfactual Theory for the Attribution of Weather and Climate-Related Events". In: *Bulletin of the American Meteorological Society* 97.1, pp. 99–110. DOI: 10.1175/BAMS-D-14-00034.1.

Hannart, Alexis, A. Carrassi, et al. (May 2016). "DADA: Data Assimilation for the Detection and Attribution of Weather- and Climate-related Events". In: *Climatic Change* 136.

Hannart, Alexis and Philippe Naveau (2018). "Probabilities of Causation of Climate Changes". In: *Journal of Climate* 31.14, pp. 5507–5524. DOI: 10.1175/JCLI-D-17-0304.1.

Hannart, Alexis, Aurélien Ribes, and Philippe Naveau (2014). "Optimal fingerprinting under multiple sources of uncertainty". In: *Geophysical Research Letters* 41.4, pp. 1261–1268. DOI: https://doi.org/10.1002/2013GL058653.

Hume, David (2004). *An Enquiry Concerning Human Understanding*. Dover, p. 226.

Kondrashov, Dmitri, Mickaël D. Chekroun, and Michael Ghil (Mar. 2015). "Data-driven non-Markovian closure models". In: *Physica D: Nonlinear Phenomena* 297, pp. 33–55. ISSN: 0167-2789. DOI: 10.1016/j.physd.2014.12.005.

Krakovská, Anna et al. (Apr. 2018). "Comparison of six methods for the detection of causality in a bivariate time series". In: *Phys. Rev. E* 97 (4), p. 042207. DOI: 10.1103/PhysRevE.97.042207.

Kraskov, Alexander (2004). "Synchronization and Interdependence Measures and their Applications to the Electroencephalogram of Epilepsy Patients and Clustering of Data (PhD Thesis)". PhD thesis. John von Neumann Institute for Computing (NIC).

Kraskov, Alexander, Harald Stögbauer, and Peter Grassberger (June 2004). "Estimating mutual information". In: *Phys. Rev. E* 69 (6), p. 066138. DOI: 10.1103/PhysRevE.69.066138.

Kravtsov, S., D. Kondrashov, and M. Ghil (2005). "Multilevel Regression Modeling of Nonlinear Processes: Derivation and Applications to Climatic Variability". In: *Journal of Climate* 18.21, pp. 4404–4424. DOI: 10.1175/JCLI3544.1.

Kravtsov, Sergey, D. Kondrashov, and Michael Ghil (Dec. 2010). "Empirical model reduction and the modelling hierarchy in climate dynamics and the geosciences". In: *Stochastic Physics and Climate Modelling*.

Kretschmer, Marlene et al. (2016). "Using Causal Effect Networks to Analyze Different Arctic Drivers of Midlatitude Winter Circulation". In: *Journal of Climate* 29.11, pp. 4069–4081. DOI: 10.1175/JCLI-D-15-0654.1.

Lamb, H. (1945). *Hydrodynamics*. Dover Books on Physics. Dover publications. ISBN: 9780486602561.

Lauritzen, Steffen L (May 1996). *Graphical Models*. Oxford University Press. ISBN: 9780198522195. DOI: 10.1093/oso/9780198522195.001.0001.

Liang, X. San (Sept. 2008). "Information flow within stochastic dynamical systems". In: *Phys. Rev. E* 78 (3), p. 031113. DOI: 10.1103/PhysRevE.78.031113.

— (Nov. 2014). "Unraveling the cause-effect relation between time series". In: *Phys. Rev. E* 90 (5), p. 052150. DOI: 10.1103/PhysRevE.90.052150.

— (Nov. 2016). "Information flow and causality as rigorous notions ab initio". In: *Phys. Rev. E* 94 (5), p. 052201. DOI: 10.1103/PhysRevE.94.052201.

— (2021). "Normalized Multivariate Time Series Causality Analysis and Causal Graph Reconstruction". In: *Entropy* 23.6. ISSN: 1099-4300.

— (2022). "The Causal Interaction between Complex Subsystems". In: *Entropy* 24.1. ISSN: 1099-4300. DOI: 10.3390/e24010003.

Liang, X. San and Richard Kleeman (Dec. 2005). "Information Transfer between Dynamical System Components". In: *Phys. Rev. Lett.* 95 (24), p. 244101. DOI: 10.1103/PhysRevLett.95.244101.

Lizier, J. T. and M. Prokopenko (Jan. 2010). "Differentiating information transfer and causal effect". In: *The European Physical Journal B* 73.4, pp. 605–615. ISSN: 1434-6036. DOI: 10.1140/epjb/e2010-00034-5.

Lizier, Joseph T et al. (Aug. 2011). "Multivariate information-theoretic measures reveal directed information structure and task relevant changes in fMRI connectivity." odefinierat/okänt. In: *Journal of Computational Neuroscience* 30, pp. 85–107. ISSN: 1573-6873. DOI: 10.1007/s10827-010-0271-2.

Lizier, Joseph T. (2014). "JIDT: An Information-Theoretic Toolkit for Studying the Dynamics of Complex Systems". In: *Frontiers in Robotics and AI* 1. ISSN: 2296-9144. DOI: 10.3389/frobt.2014.00011.

Lizier, Joseph T., Mikhail Prokopenko, and Albert Y. Zomaya (Feb. 2008). "Local information transfer as a spatiotemporal filter for complex systems". In: *Phys. Rev. E* 77 (2), p. 026110. DOI: 10.1103/PhysRevE.77.026110.

Lorenz, Edward Norton (1963). "Deterministic nonperiodic flow". In: *Journal of the Atmospheric Sciences* 20, pp. 130–141.

Marinazzo, Daniele, Mario Pellicoro, and Sebastiano Stramaglia (Apr. 2008). "Kernel Method for Nonlinear Granger Causality". In: *Phys. Rev. Lett.* 100 (14), p. 144103. DOI: 10.1103/PhysRevLett.100.144103.

Martínez-Sánchez, Á., G. Arranz, and A. Lozano-Durán (2024). "Decomposing causality into its synergistic, unique, and redundant components". In: *Nat Commun.* DOI: doi:10.1038/s41467-024-53373-4.

Massey, James L. (1990). "Causality, Feedback and Directed Information". In: *Proceedings of the 1990 International Symposium on Information Theory and Its Applictations*.

McGraw, Marie C. and Elizabeth A. Barnes (2018). "Memory Matters: A Case for Granger Causality in Climate Variability Studies". In: *Journal of Climate* 31.8, pp. 3289–3300. DOI: 10.1175/JCLI-D-17-0334.1.

Meehl, Gerald A. et al. (May 2021). "Initialized Earth System prediction from subseasonal to decadal timescales". In: *Nature Reviews Earth & Environment* 2.5, pp. 340–357. ISSN: 2662-138X. DOI: 10.1038/s43017-021-00155-x.

Melkas, Laila et al. (2021). *Interactive Causal Structure Discovery in Earth System Sciences*. arXiv: 2107.01126 [physics.data-an].

Nes, Egbert H. van et al. (2015). "Causal feedbacks in climate change". In: *Nature Climate Change* 5, pp. 445–448.

Oliveira, H P de and E L Rodrigues (Nov. 2011). "Numerical evolution of axisymmetric vacuum spacetimes: a code based on the Galerkin method". In: *Classical and Quantum Gravity* 28.23, p. 235011. DOI: 10.1088/0264-9381/28/23/235011.

Paluš, Milan et al. (July 2018). "Causality, dynamical systems and the arrow of time". In: *Chaos: An Interdisciplinary Journal of Nonlinear Science* 28.7, p. 075307. ISSN: 1054-1500. DOI: 10.1063/1.5019944.

Papagiannopoulou, C. et al. (2017). "A non-linear Granger-causality framework to investigate climate–vegetation dynamics". In: *Geoscientific Model Development* 10.5, pp. 1945–1960. DOI: 10.5194/gmd-10-1945-2017.

Pearl, J., M. Glymour, and N.P. Jewell (2016). *Causal Inference in Statistics: A Primer*. Wiley. ISBN: 9781119186847.

Pearl, J. and D. Mackenzie (2018). *The Book of Why: The New Science of Cause and Effect*. Penguin Books Limited. ISBN: 9780241242643.

Pearl, Judea (1996). *The Art and Science of Cause and Effect*. Public lecture delivered as part of the UCLA Faculty Research Lectureship Program.

— (2009). *Causality: Models, Reasoning and Inference*. 2nd. USA: Cambridge University Press. ISBN: 052189560X.

Pearson, Karl (1892). *The Grammar of Science*. Cambridge Library Collection - Physical Sciences. Cambridge University Press.

Peters, Jonas, Dominik Janzing, and Bernhard Schlkopf (2017). *Elements of Causal Inference: Foundations and Learning Algorithms*. The MIT Press. ISBN: 0262037319.

Pierini, Stefano (2011). "Low-Frequency Variability, Coherence Resonance, and Phase Selection in a Low-Order Model of the Wind-Driven Ocean Circulation". In: *Journal of Physical Oceanography* 41, pp. 1585–1604.

Pires, Carlos A., David Docquier, and Stéphane Vannitsem (2024). "A general theory to estimate Information transfer in nonlinear systems". In: *Physica D: Nonlinear Phenomena* 458, p. 133988. ISSN: 0167-2789. DOI: https://doi.org/10.1016/j.physd.2023.133988.

Prigogine, I. and Pierre Van Rysselberghe (Apr. 1963). "Introduction to Thermodynamics of Irreversible Processes". In: *Journal of The Electrochemical Society* 110.4, p. 97C. DOI: 10.1149/1.2425756.

Prokopenko, Mikhail and Joseph T. Lizier (2014). "Transfer Entropy and Transient Limits of Computation". In: *Scientific Reports* 4.

Prokopenko, Mikhail, Joseph T. Lizier, and Don C. Price (2013). "On Thermodynamic Interpretation of Transfer Entropy". In: *Entropy* 15.2, pp. 524–543. ISSN: 1099-4300. DOI: 10.3390/e15020524.

Reinhold, Brian B. and Raymond T. Pierrehumbert (1982). "Dynamics of Weather Regimes: Quasi-Stationary Waves and Blocking". In: *Monthly Weather Review* 110.9, pp. 1105–1145. DOI: 10.1175/1520-0493(1982)110<1105:DOWRQS>2.0.CO;2.

Risser, Mark D., Mohammed Ombadi, and Michael F. Wehner (2024). *Granger causal inference for climate change attribution*. arXiv: 2408.16004 [stat.AP].

Rosas, Fernando E. et al. (Mar. 2022). "Disentangling high-order mechanisms and high-order behaviours in complex systems". In: *Nature Physics* 18.5, pp. 476–477. ISSN: 1745-2481. DOI: 10.1038/s41567-022-01548-5.

Rothman, K.J. and S. Greenland (1998). *Modern Epidemiology*. Lippincott-Raven. ISBN: 9780316757805.

Runge, J. (July 2018). "Causal network reconstruction from time series: From theoretical assumptions to practical estimation". In: *Chaos: An Interdisciplinary Journal of Nonlinear Science* 28.7, p. 075310. ISSN: 1054-1500. DOI: 10.1063/1.5025050.

Runge, Jakob, Sebastian Bathiany, et al. (2019). "Inferring causation from time series in Earth system sciences". In: *Nature Communications* 10.

Runge, Jakob, Reik V. Donner, and Jürgen Kurths (May 2015). "Optimal model-free prediction from multivariate time series". In: *Physical Review E* 91.5. ISSN: 1550-2376. DOI: 10.1103/physreve.91.052909.

Runge, Jakob, Andreas Gerhardus, et al. (2023). "Causal inference for time series". In: *Nature Reviews Earth & Environment* 4, pp. 487–505.

Runge, Jakob, Jobst Heitzig, Norbert Marwan, et al. (Dec. 2012). "Quantifying causal coupling strength: A lag-specific measure for multivariate time series related to trans-

fer entropy". In: *Physical Review E* 86.6. ISSN: 1550-2376. DOI: 10.1103/physreve.86.061121.

Runge, Jakob, Jobst Heitzig, Vladimir Petoukhov, et al. (June 2012). "Escaping the Curse of Dimensionality in Estimating Multivariate Transfer Entropy". In: *Phys. Rev. Lett.* 108 (25), p. 258701. DOI: 10.1103/PhysRevLett.108.258701.

Runge, Jakob, Peer Nowack, et al. (Nov. 2019). "Detecting and quantifying causal associations in large nonlinear time series datasets". In: *Science Advances* 5.11. ISSN: 2375-2548. DOI: 10.1126/sciadv.aau4996.

Runge, Jakob, Vladimir Petoukhov, and Jürgen Kurths (2014). "Quantifying the Strength and Delay of Climatic Interactions: The Ambiguities of Cross Correlation and a Novel Measure Based on Graphical Models". In: *Journal of Climate* 27.2, pp. 720–739. DOI: 10.1175/JCLI-D-13-00159.1.

Saltzman, Barry (1962). "Finite Amplitude Free Convection as an Initial Value Problem—I". In: *Journal of Atmospheric Sciences* 19.4, pp. 329–341. DOI: 10.1175/1520-0469(1962)019<0329:FAFCAA>2.0.CO;2.

Schreiber, Thomas (July 2000). "Measuring Information Transfer". In: *Phys. Rev. Lett.* 85 (2), pp. 461–464. DOI: 10.1103/PhysRevLett.85.461.

Seth, A. (2007). "Granger causality". In: *Scholarpedia* 2.7. revision #127333, p. 1667. DOI: 10.4249/scholarpedia.1667.

Shao, Yuhao et al. (2024). "The Many Shades of the Vegetation–Climate Causality: A Multimodel Causal Appreciation". In: *Forests* 15.8. ISSN: 1999-4907. DOI: 10.3390/f15081430.

Silini, Riccardo et al. (2022). "Assessing causal dependencies in climatic indices". In: *Research Square*. DOI: 10.21203/rs.3.rs-1691544/v1.

Silva, Filipi N. et al. (2021). "Detecting Climate Teleconnections With Granger Causality". In: *Geophysical Research Letters* 48.18. e2021GL094707 2021GL094707, e2021GL094707. DOI: https://doi.org/10.1029/2021GL094707.

Smirnov, Dmitry (Mar. 2022). "Generative formalism of causality quantifiers for processes". In: *Physical Review E* 105. DOI: 10.1103/PhysRevE.105.034209.

Smyth, Bill (Feb. 2020). *5.2: The Streamfunction*. en. URL: https://eng.libretexts.org/Bookshelves/Civil_Engineering/Book%3A_All_Things_Flow_-_Fluid_Mechanics_for_the_Natural_Sciences_(Smyth)/05%3A_Fluid_Kinematics/5.02%3A_The_Streamfunction (visited on 03/02/2025).

Spirtes, Peter, Clark Glymour, and Richard Scheines (Jan. 2001). *Causation, Prediction, and Search*. The MIT Press. ISBN: 9780262284158. DOI: 10.7551/mitpress/1754.001.0001.

Steeg, Greg Ver and Aram Galstyan (2011). *Information Transfer in Social Media*. arXiv: 1110.2724 [cs.SI].

Stott, P., DA Stone, and M Allen (Jan. 2005). "Human Contribution to the European Heatwave of 2003". In: *Nature* 432, pp. 610–4. DOI: 10.1038/nature03089.

Stramaglia, Sebastiano et al. (July 2024). "Disentangling high-order effects in the transfer entropy". In: *Phys. Rev. Res.* 6 (3), p. L032007. DOI: 10.1103/PhysRevResearch.6.L032007.

Sugihara, George et al. (2012). "Detecting Causality in Complex Ecosystems". In: *Science* 338.6106, pp. 496–500. DOI: 10.1126/science.1227079.

Sun, Taylor, and Bollt (2015). "Causal Network Inference by Optimal Causation Entropy". In: *SIAM Journal on Applied Dynamical Systems* 14. DOI: 10.1137/140956166.

Takens, Floris (1981). "Detecting strange attractors in turbulence". In: *Dynamical Systems and Turbulence, Warwick 1980*. Ed. by David Rand and Lai-Sang Young. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 366–381. ISBN: 978-3-540-38945-3.

Tian, Jin and Judea Pearl (2001). "Causal Discovery from Changes". In: *Conference on Uncertainty in Artificial Intelligence*.

Tondeur, Maxime et al. (2020). "On Temporal Scale Separation in Coupled Data Assimilation with the Ensemble Kalman Filter". In: *Journal of Statistical Physics* 179, pp. 1161–1185.

Tongal, Hakan and Bellie Sivakumar (2021). "Forecasting rainfall using transfer entropy coupled directed–weighted complex networks". In: *Atmospheric Research* 255, p. 105531. ISSN: 0169-8095. DOI: https://doi.org/10.1016/j.atmosres.2021.105531.

Vannitsem, S. and P. Ekelmans (2018). "Causal dependences between the coupled ocean–atmosphere dynamics over the tropical Pacific, the North Pacific and the North Atlantic". In: *Earth System Dynamics* 9.3, pp. 1063–1083. DOI: 10.5194/esd-9-1063-2018.

Vannitsem, S., X. S. Liang, and C. A. Pires (2024). "Nonlinear causal dependencies as a signature of the complexity of the climate dynamics". In: *EGUsphere* 2024, pp. 1–22. DOI: 10.5194/egusphere-2024-3308.

Vannitsem, Stéphane, Jonathan Demaeyer, et al. (2015). "Low-frequency variability and heat transport in a low-order nonlinear coupled ocean–atmosphere model". In: *Physica D: Nonlinear Phenomena* 309, pp. 71–85. ISSN: 0167-2789. DOI: https://doi.org/10.1016/j.physd.2015.07.006.

Vannitsem, Stéphane and X. San Liang (Mar. 2022). "Dynamical Dependencies at Monthly and Interannual Time Scales in the Climate System: Study of the North Pacific and Atlantic Regions". en. In: *Tellus A: Dynamic Meteorology and Oceanography* 74.2022, pp. 141–158. ISSN: 1600-0870. DOI: 10.16993/tellusa.44.

Vannitsem, Stéphane, Carlos A. Pires, and David Docquier (2024). "Causal dependencies and Shannon entropy budget: Analysis of a reduced-order atmospheric model". In: *Quarterly Journal of the Royal Meteorological Society* 150.764, pp. 4066–4085. DOI: https://doi.org/10.1002/qj.4805.

White, Halbert, Karim Chalak, and Xun Lu (Dec. 2011). "Linking Granger Causality and the Pearl Causal Model with Settable Systems". In: *Proceedings of the Neural Information Processing Systems Mini-Symposium on Causality in Time Series*. Ed.

by Florin Popescu and Isabelle Guyon. Vol. 12. Proceedings of Machine Learning Research. Vancouver, Canada: PMLR, pp. 1–29.

Wibral, Michael, Raul Vicente, and Michael Lindner (2014). "Transfer Entropy in Neuroscience". In: *Directed Information Measures in Neuroscience*. Ed. by Michael Wibral, Raul Vicente, and Joseph T. Lizier. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 3–36. DOI: 10.1007/978-3-642-54474-3_1.

Wiener, N. (1956). *Modern Mathematics for Engineers*. New York: McGraw-Hill.

Wright, Sewall (1920). "The Relative Importance of Heredity and Environment in Determining the Piebald Pattern of Guinea-Pigs". In: *Proceedings of the National Academy of Sciences of the United States of America* 6.6, pp. 320–332. ISSN: 00278424, 10916490.

Yang, Yang et al. (Nov. 2024). "The Causal Relation within Air–Sea Interaction as Inferred from Observations". In: *Journal of Climate* 37. DOI: 10.1175/JCLI-D-23-0742.1.