VNIVERSITAT DE VALÈNCIA Escola Tècnica Superior d'Enginyeria

Escola Tècnica Superior d'Enginyeria ETSE-UV

**MÁSTER UNIVERSITARIO EN CIENCIA DE DATOS**


**TRABAJO DE FIN DE MÁSTER**

# EXPLORING CAUSAL INFERENCE METHODS TO DISCOVER CAUSAL RELATIONS AND ASSESS THE IMPACT OF HUMAN ACTIVITIES FROM CLIMATIC DATA

**AUTHOR:**
**ORIOL DE LA PAU CHACÓN ALBERO**

**ADVISOR:**
**JORDI MUÑOZ MARI**

**SEPTEMBER 2024**

**Presidente del Tribunal Evaluador** / *President del tribunal avaluador* via **CCA del Máster** / *CCA del Màster*

**Abstract**

This master thesis investigates causal discovery and inference in temporal climate series, bridging foundational data science concepts with advanced causal modeling techniques.

First, different theoretical causal concepts and frameworks are introduced, establishing the foundation for the application of different methods to climatic datasets. The research uses the Bayesian Causal Impact method to assess the influence of human activities on climate variables, such as quantifying $CO_2$ emission reductions during COVID-19 lock-downs. Subsequently, the PC-MCI algorithm is utilized to uncover causal relationships among climate variables like temperature, soil moisture, radiation, and precipitation in spatio-temporal datasets.

On the one hand, the results obtained show the rapidly growing relevance and potential of causal methods, specially in the context of complex systems like climate. On the other hand some of the most common limitations on causal reasearch, also encountered during these experiments, are extensively discussed and framed as an opportunity for future research.

# Contents

# 1 Introduction

Since the beginning of the scientific method, the main motivation across scientific disciplines has been to uncover the underlying causes behind all kinds of natural and social phenomena. However, the process of causal attribution has faced historical limitations, both in terms of theoretical representation and experimental implementation. Consequently, researchers from diverse fields have turned to the utilization of assotiational statistical tools to extract models from data [1]. This tools (such as correlation, regression, likelihood...) have proven themselves as very useful to advance scientific knowledge. Nevertheless, they are only able to infer the mathematical relationships between variables and do not provide insight into the causal connections between them. For example, using standard probability calculus, we can figure out from data that diseases and symptoms have a strong relationship, which we can even quantify. However, this formalism does not allow us to say that symptoms do not cause diseases. We have all heard the slogan 'correlation does not imply causation' [2]. Instead, from a causal point of view, the relationship between dependent variables can be understood in one of two ways: either they have a causal influence on each other in a specific direction, or they are both influenced by a common driver (Reichenbach's common cause principle [3]). And we want to unveil these relationships.

Causal inference deals with a wide range of problems where we need to understand the underlying data-generating process. However, the task of establishing causality is often challenging due to various factors, such as confounding variables, selection bias, and the inability to conduct controlled randomized experiments in certain situations[4]. For instance, if we seek to make predictions about how a given population will evolve after an specific intervention, in a context where a randomized experiment is not possible. This problem remains hard to solve as we may not have samples from a manipulated population [5].

In order to unveil causal relationships from data it was necessary to extend the probability formalism and to establish a set of assumptions, these being the foundation of causal inference. These requirements, as well as the inherent difficulty of proving some causal results in absence of a groundtruth, were the reasons why historically some staticians remained skeptical when it came to accept this discipline [2]. However, causal inference is a rapidly evolving discipline, and recent years have witnessed significant progress in different causal formalisms. This fact, combined with enhanced computational power, and the recent exponential growth of available data, has contributed to the maturity of the field. Causal inference now holds the potential to assist us in tackling unresolved scientific questions, offering valuable insights and facilitating our quest for knowledge [3]. Furthermore, it will be interesting to deepen in the relationship between causal inference and Machine Learning. There are already some advancements on how causal approaches can help to improve the performance of Machine Learning models [5]; and further on this project we will also show an example on Machine Learning simplifying a causal problem.

Focusing on the case of dynamical complex systems, such as the Earth climate system or the human brain, where multiple variables interact giving rise to unexpected behaviours, causal approaches remain truly promising in order to uncover the underlying physical relationships between variables. There are two main reasons to support this statement. Firstly, conducting interventional experiments in these domains often raises ethical concerns or practical limitations, as they may jeopardize human lives, for example. Secondly, the quantity and quality of temporal series data has substantially increased coming both from remote sensing observations and simulations, in the case of Earth system [3]. For example, on [6] is shown how causal methods are able to unveil, from temporal data, how tropical Pacific

surface temperatures (el Niño) affect land temperatures in North America, while the reverse relationship does not hold. This phenomenon is physically well established but cannot be inferred from traditional methods such as correlation.

In the context of Climate Sciences there are numerous possible applications of causal methods such as: understanding nonlinear dynamical interactions between variables which are partially not adequately represented by current models; uncovering the causes of climatic extreme events in order to mitigate its consequences; evaluating the impact of human intervention on climatic anomalies; or assessing different physical models and refining model parameters. However, there are a lot of challenges to be solved if we want to harness all the potential of causal methods on Climate Sciences. Some of the most important would be: the existence of significant unobserved variables, the computational cost of these models in a growing set of variables, the relative absence of datasets with known causal groundtruth, or the fact that causal relationships on climate events usually manifest on different spatio-temporal frequencies [3]. Considering all the mentioned factors, the implementation of causal methods on climate science remains a significant challenge with a huge potential to help us to understand the planet where we live and to address the complex climatic disruptions we will face on the coming years, which motivated me to undertake this project.

In this master thesis I will focus on two different models, each of them being an example of one of the main paradigms on Causal Inference from time series: structural equation models, and graphical models based on conditional independence [5]. For each case we will first provide an overview of the theoretical principles lying on its foundation. Subsequently, these models will be applied to various climate datasets and the results obtained will be compared with the existing literature. This comparative analysis aims to shed light on both the potential and limitations associated with these models in the context of climate science, paving the way for future investigations.

To start with, our attention will be on CausalImpact, an R library developed by Google Research [7]. This method integrates a Bayesian approach with a Structural state-space Equation Model in order to estimate the effect of a given intervention on the evolution of a temporal variable. The key idea is to predict how the variable would have evolved if the intervention had not occurred, leveraging its pre-intervention behavior alongside that of comparable, unmanipulated control variables. This conterfactual series can be compared with the actual observed outcome to estimate the causal impact of this event. First we will use this method on a simple dataset ([8]) to illustrate how it works. In this example, we will try to infer if the global temperature increase in the XX century can be explained just by other natural variables such as radiation, or if we can causally attribute it to anthropogenic interventions such as the exponential growth on $CO_2$ emissions. Once the foundation of this method is understood, the impact of Covid-19 lockdowns on the observed $CO_2$ emissions reduction will be assessed for different countries. The results will be compared with those obtained in [9].

On the second part of this project the focus will shift to PC-MCI, an algorithm implemented in the Python library Tigramite by [10]. PC-MCI aims to estimate the temporal causal graph from high-dimensional time series datasets. This approach combines two stages, the first one being based on the standard PC algorithm [11]. It starts with a fully connected graph and iteratively removes irrelevant links, although the resulting set may also include some false positives. The second stage involves the MCI (Momentary Conditional Independence) test, which examines whether two variables are conditionally independent given a subset of other variables, all of them selected during the first stage. In cases where the test

confirms conditional independence, under certain assumptions, we can infer the absence of a causal relationship between the variables. This modification of the standard PC algorithm aims to enhance its computational efficiency and tackles certain challenges posed by the original version. Furthermore, linear and nonlinear conditional independence tests can be implemented, accounting for generally nonlinear causal links observed on complex systems.

In order to analyze the performance of this method, it will be tested on a comprehensive global climatic dataset compiled in [12]. Following an exploratory data analysis and preprocessing, we will have four temporal series (representing temperature, radiation, soil moisture and precipitation) for each one of Earth's 13072 pixels. Our objective will be to estimate the temporal causal graph of these variables within different climate zones across the Earth. To begin, we will delve into the realm of temporal series clustering to better understand the geographical distribution of different climatic patterns. Then we will explore the PCMCI algorithm, both with linear and nonlinear links for each pixel. The resulting outcomes will be used to estimate the strength and sign of causal effects for each pair of causally related variables. We will also use a clustering algorithm in order to explore how climatic causal relationships may change across the planet. This analysis will allow us to infer some interesting conclusions that can be qualitatively compared with existing literature, providing a comprehensive analysis of the performance and capabilities of the employed model. The results obtained highlight the potential of causal inference to deepen our understanding of climate models. However, they also showcase the limitations that can be found when dealing with this class of algorithms. Finally, an extense discussion on these drawbacks lays the groundwork for a future experiment that could significantly improve our comprehension of the patterns presented hereafter.

# 2 Causal Impact

On social sciences, such as econometrics, a common objective is to assess the impact of a given intervention on the evolution of a temporal variable. For example, we may want to quantify the effect of a marketing campaign on the sales figures of a particular product or the number of clicks a website receives. One conventional approach would be conducting an experiment where the marketing campaign is implemented in a designated market A, and then comparing the results with those of a market B with similar characteristics, where the campaign was not launched (for example, two different geographical regions). However, in today's interconnected world, where the influence of a campaign can extend across zones, and considering the limited resources that companies can allocate to marketing initiatives, there is a growing interest in employing cost-efficient causal inference methods to answer these kind of questions.

In a study by Google Research [7], this problem was addressed using a Bayesian state-space equations model. In this section, we will delve into the theoretical principles and assumptions of this model, and evaluate its potential utility in analyzing climatic data, focusing on assessing the impact of human activities on temporal climatic variables.

## 2.1 Theoretical approach: Bayesian structural time-series models

The causal impact of a treatment can be defined as the difference between the observed value and the unobserved value that would have been obtained under the alternative treatment (e.g. absence of intervention). Hence, our objective is to develop a model able to predict the behaviour of our temporal variable in case the intervention would have not taken place, which will be called counterfactual. To construct this counterfactual, we can utilize three sources of information: the pre-intervention dynamics of our target variable; the behaviour of similar unmanipulated variables that were predictive of the response series before the treatment; as well as the existing prior knowledge regarding the model parameters if a Bayesian approach is used. In this case, structural state-space time-series models will be used to infer the counterfactual, which are based on the following two equations:

$$y_t = Z_t^T \alpha_t + \epsilon_t \tag{1a}$$

$$\alpha_{t+1} = T_t \alpha_t + R_t \eta_t \tag{1b}$$

The state equation (1b) represents the underlying dynamics of the system and describes how it evolves over time. It is typically specified as a set of linear equations, where the state at each time step depends on the previous state and some random noise. The state variables ($\alpha_t$) can represent unobservable quantities or latent variables that drive the observed data. The observation equation (1a) relates the observed data to the underlying state of the system. It specifies how the observations are generated from the state variables at each time step.

In this case, $y_t$ will be a scalar observation and $\alpha_t$ is a $d$-dimensional state vector. The random noise in both equations is usually assumed to follow a multivariate normal distribution and to be independent of all other unknowns ($\epsilon_t \sim \mathcal{N}(0, \sigma_t^2)$ and $\eta_t \sim \mathcal{N}(0, Q_t)$). Then we have a $d \times d$ transition matrix $T_t$ which defines the linear relationship between the latent states at consecutive time steps; and the observation matrix $Z_t^T$ that relates the latent states to the observed data (in our case being a $d$-dimensional vector). $\epsilon_t$ is a scalar observation error with noise variance $\sigma_t$, and $\eta_t$ is a $q$-dimensional system error with a $q \times q$ state-diffusion matrix $Q_t$, where $q \leq d$. $Q_t$ captures the uncertainty or variability in the evolution of the

latent states. The structure of the state equation's error ($R_t\eta_t$) allows us to incorporate state components of less than full rank, such as seasonality [7].

Structural state-space models have two key characteristics that make them highly appealing for addressing time-series problems. In the first place, they exhibit flexibility, as they can accommodate a broad range of models, such as ARIMA models, within the state-space framework. Secondly, they offer modularity, allowing different components, such as trend and seasonality, to be modeled independently and then combined to form a comprehensive final model.

This problem will be solved using a Bayesian approach. First we have the target data previous to the intervention $y_{1:n}$, and we define a prior distribution $p(\theta)$ for all the model parameters previously explained. This parameters usually only depend on a small set of variance terms, which prior distribution is often assumed as a Gamma distribution, although it may be modified to include expert knowledge. We also choose a distribution for the initial values of the state vector given the model parameters $p(\alpha_0|\theta)$. Information from covariate series is also included through a regression term. Our goal will be to find the posterior distribution of the model parameters and state vector, given the data $p(\alpha, \theta|y_{1:n})$, which will allow us to predict the behaviour of the target variable through time in case the intervention would have not taken place. A Markov Chain Monte-Carlo (MCMC) simulation will be used for this purpose.

The main idea of Monte-Carlo simulations would be to characterize a distribution without requiring complete knowledge of its mathematical properties by randomly sampling values out of the distribution. The key contribution of Markov chain based Monte-Carlo simulations is that each sample depends on the one just before it, but it does not depend on any other previous sample. It can be shown that this property guarantees that, once the simulation converges after enough iterations, all drawn samples come from a stationary distribution that is a good approximation of the target distribution. Hence, samples obtained prior to convergence must be discarded in a process known as burn-in [13]. In this particular case the MCMC method used is a Gibbs sampler [14]. Once a proper distribution is achieved for the model parameters and the state vector we can predict the counterfactual (including its error interval) and compute the causal impact of the intervention as the difference between the counterfactual and the observed values. We can also calculate the posterior distribution of cumulative impact using the same results.

In order to implement this method, some strong assumptions must be adopted, as is usually the case for causal inference algorithms. First, we assume the existence of a control group of time series that were unaffected by the intervention. Inaccuracies may result in underestimating or overestimating the actual impact, or in erroneously inferring an effect when none truly exists. Secondly, the model assumes that the relationship between covariates and the treated time series, established during the pre-intervention period, remains constant throughout the post-intervention period. Although in some cases it might not be possible to check if these assumptions hold, there are some tests that may help. A simple plot can inform us if covariates actually have been affected by the intervention (first assumption). It will also be useful to confirm that covariates are a good predictor of the target variable prior to the intervention. This model is implemented in R by [15].

## 2.2 Can global warming be explained only by natural variables?

Once the theoretical foundations behind Causal Impact have been reviewed we will show how to work with this method using a simple example. The Coupled Model Intercomparison Project (CMIP) is a collaborative effort within the climate science community to compare and evaluate global climate models. It aims to improve our understanding of climate change and its potential impacts by coordinating model simulations and sharing data among participating research institutions worldwide. Its projections and data have played a crucial role on the Intergovernmental Panel on Climate Change (IPCC) assessments. On this section, we will be working with a simplified version of a dataset simulated using the ModelE2 version of the CMIP (phase 5), developed by the Goddard Institute for Space Studies (GISS) in [8].

We will be working with five different climatic variables representing global average values. Each variable has an annual temporal frequency, ranging from 1850 (when early measurements enable the evaluation of simulated data, while the climate system remained unaffected by substantial anthropogenic forcings) up to 2005. The selected variables include two primary natural drivers of Earth's temperature: solar irradiance, which quantifies the power per unit area received from the sun by our planet; and volcanic emissions, which influence the balance between incoming and outgoing radiation on our planet. Additionally, we consider two key anthropogenic drivers of Earth's temperature: greenhouse gas emissions, particularly $CO_2$; and land-use emissions flux which measures emissions and removals of greenhouse gases due to human-induced land use changes. Finally, we examine the global temperature anomaly relative to a baseline computed for the entire study period. The variables of the dataset can be visualized in Figure 1.

As observed in Figure 1(e), global temperature has undergone a constant and significant increase since the 1920s, commonly referred to as global warming. Extensive scientific research has been dedicated to understanding the contribution of natural climatic dynamics and human-induced modifications to this phenomenon ([16] and [17] for example). A comprehensive analysis of this question is beyond the scope of this master thesis, as some very complex physical models are required. However, we can treat the quasi-exponential growth in anthropogenic emissions as a given intervention on a temporal variable (temperature). This manipulation leaves unaffected other variables that were primary drivers of the target variable before the intervention (solar irradiance and volcanic activity do not depend on emissions). Therefore, we can use Causal Impact to quantify the potential effect of human emissions on the temperature anomaly. This way, we aim to address the following question: Can global warming be solely explained by natural variables, or some other factors must be taken into account?

Thus, data from 1850 to 1925 will be considered the pre-intervention period, the post-intervention period being all subsequent data. Solar irradiance and volcanic activity are considered covariates, predictors of the target variable (temperature) before the treatment. We will try to construct the counterfactual behaviour of global temperature if it was only driven by natural phenomena and compare it with its real behaviour. The Causal Impact implementation in R [15] is quite straightforward, including methods to plot and summarize the results.

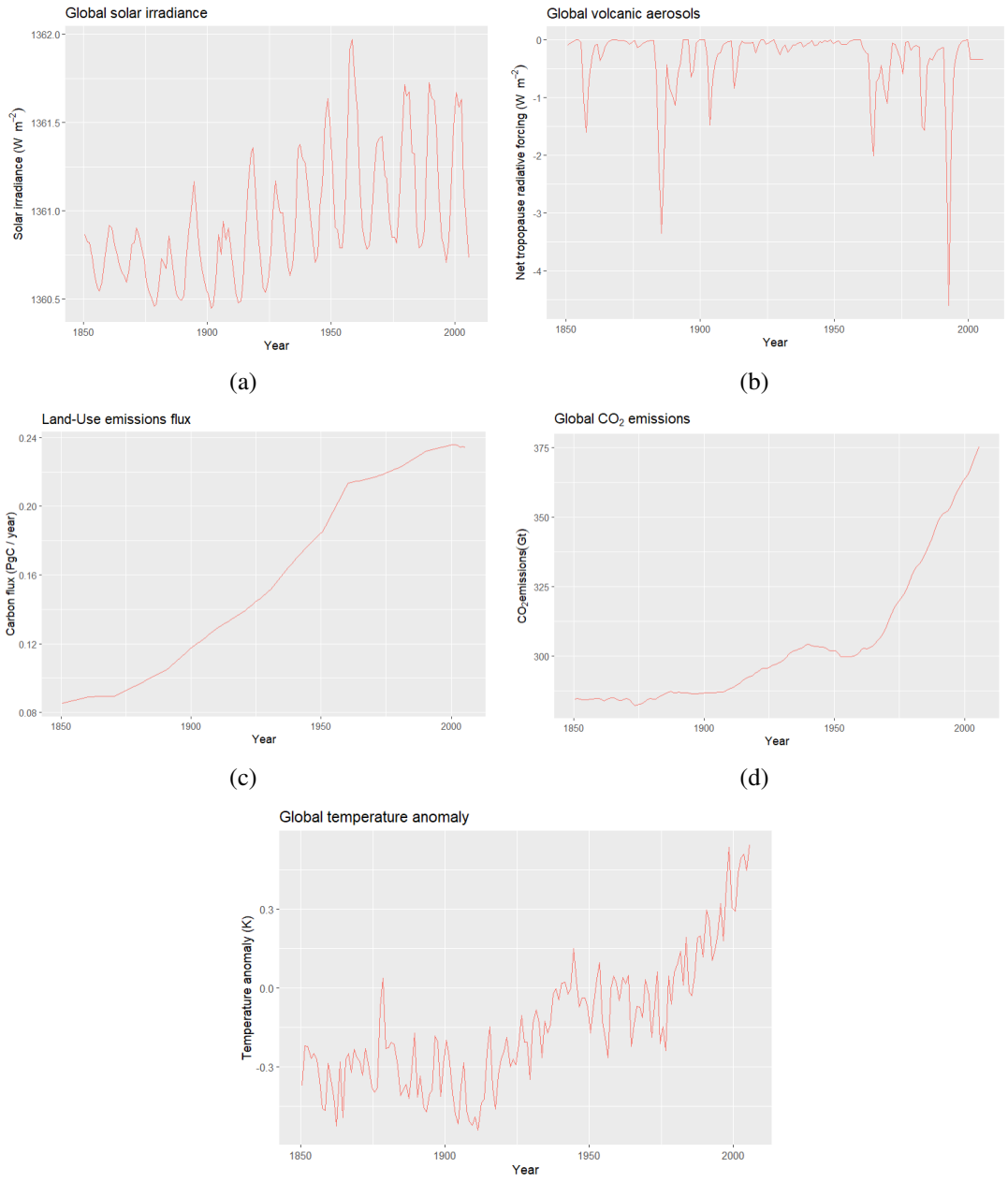
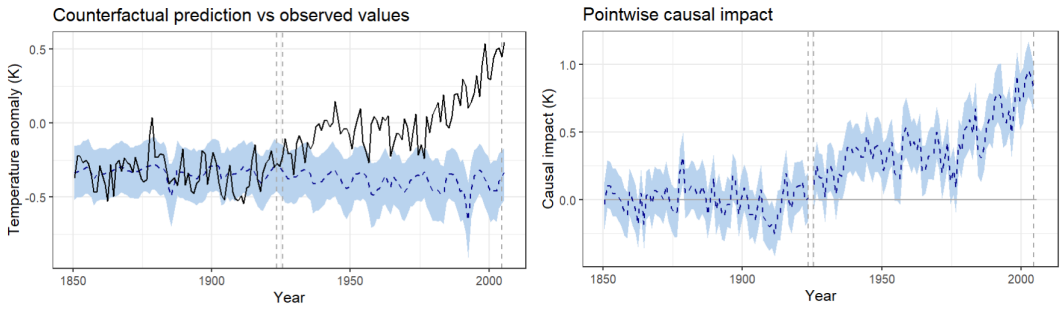



Figure 1: (a) Power per unit area received from the sun. We can see the 11 years cycle on its magnetic field. (b) Net tropopause radiative forcing induced by volcanic activity. It measures the balance between incoming and outgoing radiation between the troposphere and the stratosphere, which can be modified by aerosols, gases and particles emitted by volcanic eruptions. (c) Flux of carbon between natural reservoirs and the atmosphere due to land-use changes. For example, deforestation has a huge negative impact on $CO_2$ natural recapturing. (d) Global $CO_2$ emissions due to industrial production and human mobility. (e) Global temperature variations with respect to the average baseline.




(a) (b)

Figure 2: Visualization of our first experiment results. (a) Observed values of temperature anomaly (black) and our counterfactual prediction with its 95 % confidence interval (blue). (b) Difference between observed data and counterfactual predictions (pointwise causal impact).

The obtained results (see Figure 2) show that in the absence of an intervention, just taking into account the natural drivers of Earth's temperature, it would have stayed in a constant negative range. The observed temperature anomaly increasing evolution stays far beyond the 95% confidence interval for the counterfactual. In fact, the probability of obtaining this effect by chance, in this model, is very small (p = 0.001). This means the causal effect can be considered statistically significant. It is important to note that this method does not allow us to attribute this effect solely to $CO_2$ emissions directly. However, we can clearly assess that natural temperature drivers, such as volcanic emissions and solar irradiance, are not enough to explain the constant increase in global temperature, and some other factors must be considered. Furthermore, the cumulative effect of the intervention on the target variable accounts for a 1.7 K increase in global temperature over the baseline. These results are consistent with the consensus of the scientific community, as reflected in the last IPCC assessment [18], which asseverates that natural climatic dynamics are not enough to explain global warming, and anthropogenic interventions on the climate system account for a 1.1 K increase of temperature since 1850. The difference between the cumulative impact obtained on our experiment and the latter can be explained by the fact that our model only takes into consideration 3 variables whereas the IPCC results come from complex physical simulations. Moreover, after the completion of this experiment, we found a similar study [19] using this algorithm on the same dataset. The conclusions obtained are similar to the ones presented here. A 0.3°increase in global temperature was attributed to $CO_2$ emissions, although a larger list of covariates was used, and the post-treatment period was selected after 1960.

This experiment served as a valuable illustration to understand how to work using Causal Impact, and which kinds of problems, beyond economic sciences, can be addressed with this method. Further work on this model will be developed in the next section.

## 2.3 Assessing the impact of COVID-19 on $CO_2$ emissions reduction

The COVID-19 pandemic lockdowns in 2020 imposed heavy restrictions on all kinds of economic activities and general population mobility. Consequently, a substantial decrease in global $CO_2$ emissions was observed. Considering that annual reductions of greenhouse gas emissions at a similar scale are required to restrict global warming to 1.5 ºC (in the best-case scenario), some questions may arise about the feasibility and pace of the energy transition. In this context, it becomes interesting to assess which precise impact did lockdowns have on the cited $CO_2$ emissions reduction. This problem has been studied on [9]. On the one hand, real emissions were estimated for different countries from daily activity data (such as energy consumption) on different economical sectors. On the other hand, emissions that would have occurred in the absence of the pandemic were simulated using a linear regression model trained on data from the period of 2015-2019. The need to model a counterfactual temporal variable, representing the emissions under the absence of the intervention, using similar covariates, presents a good example where the method described in the previous sections could be an interesting approach. The cited publication also includes a dataset with daily emissions by country and economic sector (power, aviation, total...) produced during 2019 and 2020. Simulated daily emissions in absence of the pandemic are also shared. Therefore, Causal Impact will be applied to the aforementioned dataset in order to compare the achieved results.

First, the methodology used will be exemplified with a very familiar case. We will show how Causal Impact can be implemented to quantify the effect that COVID-19 confinement and mobility restrictions had on Spain's $CO_2$ emissions. Next, the results obtained for all the other studied countries will be displayed and discussed. On Figure 3 the difference between 2019 and 2020 Spanish emissions can be visualized. Through a Bayesian structural state-space equations model, we will be able to predict how much daily $CO_2$ emissions would have been produced by Spain in normal conditions (without pandemic), based on 2019 daily emissions and 2020 pre-intervention data.

In Spain, general population confinement and different mobility and economic restrictions were declared on March 14 [20]. Thus, this date was chosen as the intervention starting point. Although some restrictions were eased on June, in this experiment we considered that the intervention lasted for the hole year, as some new limitations were enforced on the following months, and social and economic normality was not fully recovered that year. Furthermore, this hypothesis simplifies our calculations and allows to leverage the hole dataset. However, in further analysis it could be interesting to assess only the causal impact of the first lockdown.

As we can see on Figure 4, we find a significant effect of the intervention (COVID-19 restrictions) over the daily $CO_2$ emissions data. The probability of obtaining this effect by chance, in this model, is very small ($p = 0.001$). At first glance (see Figure 4(a)) it would seem that our hypothesis (the intervention lasting for the hole year) might be wrong, as the causal impact is concentrated on the first lockdown period. However, if we consider the cumulative causal effect (Figure 4(c)), we clearly see that following months also contribute to the overall decrease in total $CO_2$ emissions. From this cumulative sum we calculate the total amount of $CO_2$ emissions reduction in Spain due to the pandemic in ($28.75 \pm 9.30$) Gt with a 95% confidence interval. This represents a 11.76% reduction relative to 2019 total $CO_2$ emissions.

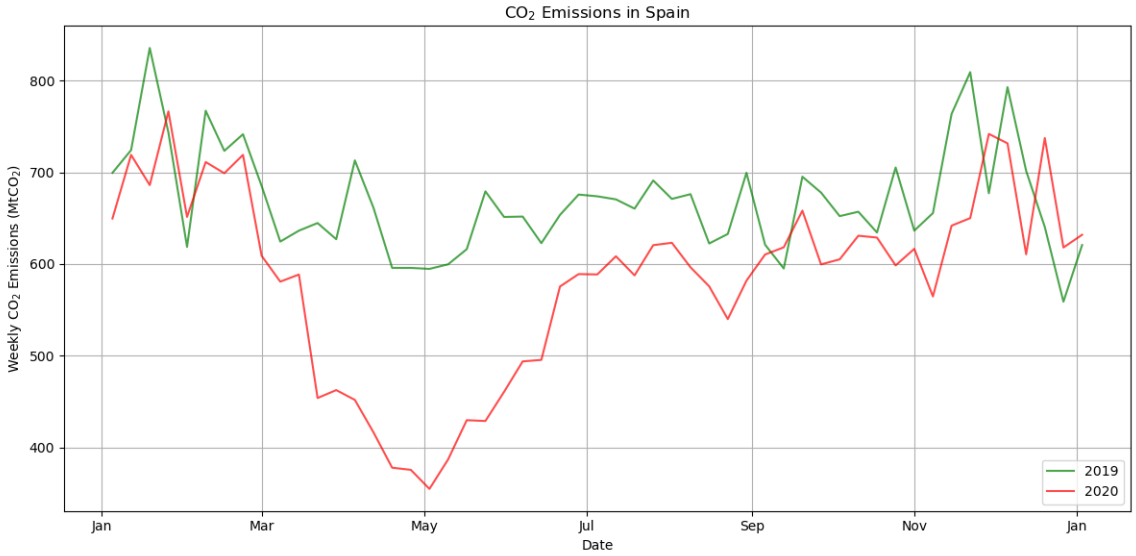



Figure 3: Graphical comparison of $CO_2$ emissions in Spain between 2019 and 2020. We can clearly see an abrupt decrease in emissions between March and June on 2020 data, due to the COVID-19 lockdowns. Emissions remain significantly lower for the following months. Data has been averaged to a weekly frequency to improve the visualization. From this plot we see that both assumptions necessary to implement Causal Impact, described in section 2.1, seem to hold true. First, the behaviour of both variables appears to be very similar besides the March-June confinement, so 2019 emissions should be a good predictor of 2020 counterfactual emissions. This hypothesis is also assumed in [9]. Secondly, the covariate data (2019 emissions) was not affected by the intervention.

We have implemented the same analysis for different countries included in the dataset. Only aggregated emissions accounting for all economic sectors have been studied. The results are shown in Table 1. We have calculated the total amount of emissions reduction per country due to the COVID-19 pandemic as the cumulative sum of daily causal impact, as in the previous example. Then, the emissions reduction relative to 2019 is computed following equation 2. This results can be compared with the ones obtained on [9]. In this case we are not talking about the same exact variable, given the fact that in the original paper the target variable represented the difference between predicted and real emissions for the hole year whereas in our case only data post confinement is taken into account. However, this comparison can show interesting insights. Results in [9] are obtained using equation 3. This equation does not explicitly appear in the paper, but it allowed us to obtain similar results.

$$Em_{rel1}^{CO_2} = \frac{\sum Em_{daily\,counterfactual}^{2020\,postint} - \sum Em_{daily\,real}^{2020\,postint}}{\sum Em_{daily}^{2019}} \tag{2}$$

$$Em_{rel2}^{CO_2} = \frac{\sum Em_{daily\,counterfactual}^{2020} - \sum Em_{daily\,real}^{2020}}{\sum Em_{daily}^{2019}} \tag{3}$$

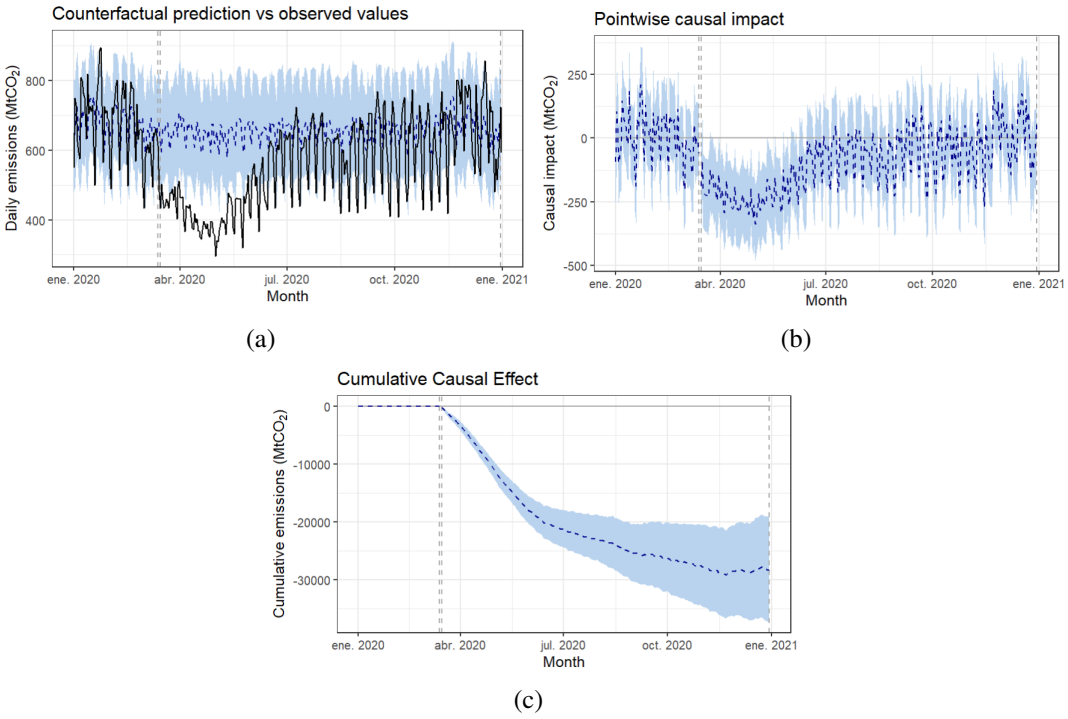
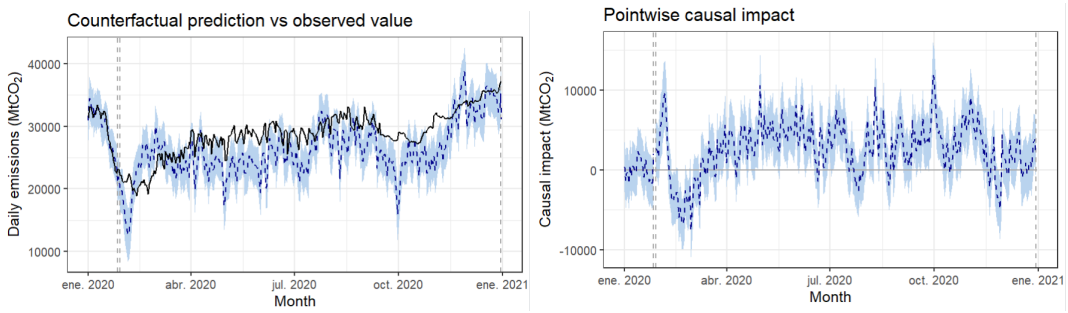



Figure 4: Visualization of our second experiment results with Spanish data. (a) Observed values of daily $CO_2$ emissions (black) and our counterfactual prediction with its 95 % confidence interval (blue). (b) Difference between observed data and counterfactual predictions (pointwise causal impact). (c) Cumulative causal effect as the sum of all previous individual points in (b).

| Country | Emissions red. ($GtCO_2$) | $Em_{rel1}^{CO_2}$ (CI) | $Em_{rel2}^{CO_2}$ (paper) |
|---|---|---|---|
| Brazil | (26.69 ± 10.04) | 6.23% | 9.66% |
| China | (-893.81 ± 411.71) | -8.54% | -0.85% |
| France | (49.42 ± 13.57) | 17.05% | 8.99% |
| Germany | (52.39 ± 40.42) | 7.96% | 7.21% |
| India | (113.50 ± 227.64) | 4.61% | 7.30% |
| Italy | (32.42 ± 15.36) | 10.25% | 7.14% |
| Japan | (32.66 ± 33.78) | 2.99% | 4.69% |
| Russia | (131.31 ± 34.55) | 8.50% | 7.96% |
| Spain | (28.75 ± 9.30) | 11.76% | 12.71% |
| UK | (39.28 ± 15.08) | 11.42% | 8.82% |
| US | (348.94 ± 120.11) | 6.94% | 9.54% |

Table 1: For each studied country we display the increase on $CO_2$ emissions that would have been produced in absence of the pandemic computed using our method (second column); the percentage of reduction of emissions relative to 2019 data obtained with Causal Impact, only after the lockdown starting date (third column); and the relative emissions reduction for the whole year, obtained in [9] (last column).




(a) (b)

Figure 5: Visualization of our second experiment results with Chinese data. (a) Observed values of daily $CO_2$ emissions (black) and our counterfactual prediction with its 95 % confidence interval (blue). (b) Difference between observed data and counterfactual predictions (pointwise causal impact).

## 2.4 Discussion on Causal Impact

On the one hand, we showed how Causal Impact can be helpful in order to assess the impact that some human interventions may have over climatic variables. This versatile tool holds promise beyond its original application, particularly in our current era when human actions can lead to unexpected and dangerous consequences for the climate, as we saw in the previous section. Being able to predict the outcomes of interventions without the need for physical experiments presents significant potential. In fact, our experiment yielded results consistent with existing literature for some cases (Russia, Spain). On the other hand, we also found some significant limitations both in our data and in the model. For other countries results significantly differ from existing literature and confidence intervals are very wide, underscoring our trust on the conclusions obtained.

There are different reasons that I believe could explain these limitations. Firstly, [9] utilized data from the period of 2015-2019 to model the counterfactual. However, only 2019 daily emissions were publicly available. Maybe our experiment would have benefited from more covariates. Furthermore, in some countries like China and the US, the restrictions imposed during the lockdowns varied greatly across regions and over time, from strict confinement measures to mere recommendations. Finally, if we look at China's results (see Figure 5), we see that our model is not capable to account for other unexpected variables or interventions. In this case, although probably $CO_2$ emissions did decrease in some regions where strict lockdowns did apply, other regions did not suffer these consequences and its economy recovered rapidly. Consequently, attributing the observed increase in emissions solely to COVID-19 could be risky. Considering these limitations, we recognize the need for more sophisticated models in some cases, as we will explore in the following section.

# 3 Discovering temporal causal graphs with Tigramite

In this section, our focus shifts to causal discovery. Unlike causal impact estimation, which assumes a known causal structure, our goal is now to uncover the causal relationships between variables directly from data.

## 3.1 Theoretical introduction

In this section, the theoretical framework necessary for discovering the causal structure of a given system from temporal data will be developed. First, the basic ideas and assumptions behind the graphical causal framework will be exposed. Then, the algorithm used PC-MCI will be introduced along an intuitive example. Finally, linear and non-linear conditional independence tests used by the algorithm will be exposed.

### 3.1.1 PC-MCI algorithm

As observed in the previous section, climatic problems usually arise from complex dynamical systems where numerous variables interact across diverse temporal and spatial scales. In scenarios of this nature, where we want to learn the causal structure directly from data, employing the graphical causal model framework becomes interesting. This approach enables the identification of causal connections among variables and the determination of the temporal frequency of these links. It also provides an intuitive manner to visualize these relationships. The framework is primarily based on concepts from graph theory: the causal structure of our system will be represented by a Directed Acyclic Graph (DAG). Nodes represent variables and edges are directed arrows connecting nodes, which designate causal links. Edges must be directed to clearly state which variable is cause and which is consequence. Graphs must be acyclic to preserve consistency and temporal order. In our case, dealing with temporal series, this framework can be modified to better represent the temporal nature of these relationships. Thus, each node will designate a variable for a given time step. This graph should be infinite, but in practice, we define a maximum time lag $\tau_{max}$ [21].

In order to understand causal relations in this framework, we must introduce the concept of conditional independence. Conditional independence is a fundamental concept in probability and statistics. It refers to the idea that the relationship between two variables might be independent of each other when a third variable is taken into consideration. We say that variables A and B are conditionally independent given C if:

$$A \perp\!\!\!\perp B \mid C \iff P(A \mid B, C) = P(A \mid C) \iff P(A, B \mid C) = P(A \mid C) \cdot P(B \mid C) \qquad (4)$$

That is, if we consider variable C, B contributes nothing to the certainty of A. For example, let's only examine variables height and vocabulary complexity. We will find some correlation as children's vocabulary is more rudimentary, and we may even think that there is a causal relationship. However, if we take into consideration the variable age, we will find that height does not contribute any extra information. Thus, vocabulary complexity is conditionally independent from height, given age. There exist different statistical tests that allow to flexibly assess conditional independence between variables from data, including both linear and non-linear conditional independence [3]. Thus, it is possible to construct a DAG depicting the conditional independence relations between the different variables of a dataset. It has been shown that, under some assumptions, if two variables are conditionally independent given a third one, they are also causally independent. Therefore, if these assumptions hold, we can assert that the conditional independence graph represents the causal relations graph [21].

The first assumption is 'Causal Sufficiency'. A set of variables is causally sufficient if all variables that may influence any other variable in the set are also included. Therefore, we must assume that there are no unobserved causally relevant variables. The second assumption is the 'Causal Markov Condition'. It states that any variable $X_t$ in the graph is independent of all other variables conditioned on its parents (nodes that have a directed arrow pointing at $X_t$, its direct causes). This basically means that if we know the values of the parents of $X_t$, then any information about that variable's value that might be contained in other variables is already captured by the parents. From this condition follows the fact that separation in the graph means independence, and connectedness means dependence.

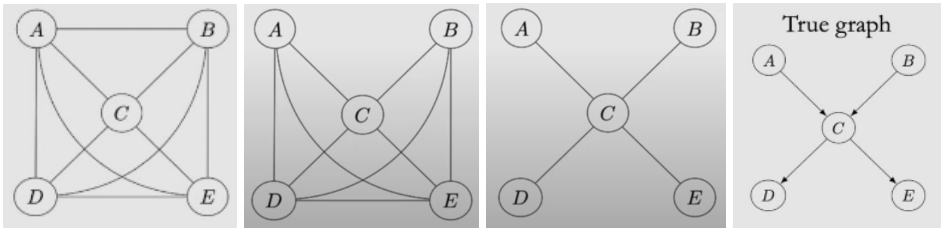
The third one is the 'Faithfulness Assumption'. It states that if a conditional independence relationship is observed in the data, then it is represented in the graph. If we observe a conditional independence relationship in the data that is not represented in the graph, this assumption does not hold, which suggests that there might be missing causal links or unmodeled confounding variables. From the combination of these assumptions, we can derive the equivalence between a measured statistical dependence and the existence of a causal process. These three assumptions are key and common to all models within the graphical causal model framework.

A widespread approach for determining the causal graph of a system is the PC algorithm, named after its developers Peter and Clark [11]. Initially, a fully connected graph is established. In the first stage, conditional independence is tested for all pairs of nodes while iteratively growing the number of conditions, thereby removing spurious links. Connections between nodes at different time steps are directed in accordance with temporal order. During the final stage, contemporaneous links are oriented using a set of logical rules. In Figure 6 an intuitive example of how this algorithm works is displayed.

However, while the PC algorithm is effective in many cases, it has shortcomings, especially when dealing with high-dimensional temporal data or data with complex non-linear causal relationships. In order to address some of these limitations the PC-MCI algorithm has been introduced by [10]. This method comprises two different steps. The first one is based on the PC algorithm. The idea would be to achieve a set of relevant conditions for each variable at each time step where the causal parents are included. Some false positives may also be part of this set, as this step only tests for conditional independence with a given significance level $\alpha_{PC}$. On each iteration, possible parents are ordered taking into account the conditional independence test value. On the next iteration, we will only condition on the strongest parents following the order previously established, with an increasing number of conditions. The algorithm converges once the number of possible parents is equal to or smaller than the number of conditions to be tested. The main difference with the original PC algorithm would be the fact that with this method we do not need to test all possible combinations of conditions, which significantly reduces the execution time. It is important to note that the goal of this step is solely to find the relevant conditions.

The second step is referred to as Momentary Conditional Independence test. In the previous step we achieved a set S of possible parents of the variable X at a given time. This result allows us to identify and remove spurious links. In this second stage, independence will also be tested conditioning on the parents of S, which has been shown to solve the problem of strong auto-correlation (in the standard PC algorithm the presence of strong auto-correlated time series, which is often the case, usually induces a high proportion of false positives). Finally, links are oriented following the same rules as in the standard PC algorithm. The PC-MCI algorithm is implemented in the Python library Tigramite [6].

This method builds upon the assumptions previously explained. We also assume that the




(a) (b) (c) (d)

Figure 6: Illustration of how the PC algorithm is able to recover an atemporal causal graph from data [22]. (a) We start with a fully connected graph. (b) In the first step, we can remove the link between variables A and B as they are independent without conditioning on any variable ($A \perp\!\!\!\perp B \mid \{\}$). (c) In the second step, we test for independence any possible pair of variables, conditioning on a single different variable. Thus, we can remove the links (A,D), (A,E), (B,E), (B,D) and (D,E), as they all are independent conditioned on C (ex. $D \perp\!\!\!\perp A \mid \{C\}$). The next step would be to test independence conditioning on pairs of variables. However, all remaining connected pairs of variables are dependent. This way we achieved the skeleton of the graph. (d) Finally, links must be oriented. First, we can orient A and B pointing towards C. Any different configuration (such as $A \rightarrow C \rightarrow B$ or $A \leftarrow C \rightarrow B$) would have implied the need for conditioning on C in order to find the independence between A and B. We can also orient $D \leftarrow C \rightarrow E$. If this were not the case, we would have found other independent relationships in the first step. For example, $D \rightarrow C \rightarrow E$ implies that A and D are independent. This algorithm has a main limitation: in real-world problems, causal graphs may be way more complex, and this method must test all possible combinations, which is computationally very expensive.

causal structure and the conditional independence relations between variables remain constant through time, representing a weaker definition of stationarity. Finally, some assumptions about the dependencies between variables must be made. If a given causal relation $X \rightarrow Y$ exists on data, it means the existence of an unknown function $Y = f(X)$. The goal of this causal framework is not to find this function, a task that may be pursued in subsequent research, although it can be estimated for some cases. We only want to assess if this relationship truly exists. However, a crucial assumption lies in whether this function is linear or non-linear. The algorithm provides great flexibility, requiring only the selection of a specific conditional independence test to be applied across various steps. Yet, some tests are only able to reveal linear relations, while others perform better when handling non-linear cases. Therefore, incorporating expert knowledge about the nature of dependencies can significantly facilitate our task. In this project, our focus will be on the Partial Correlations test for linear scenarios, and the Conditional Mutual Information test, based on the nearest neighbors estimator, for non-linear cases.

### 3.1.2 Linear dependencies: Partial Correlation

Partial Correlation is a statistical test used to investigate the linear relationship between two variables while controlling for the influence of one or more additional variables. Imagine the following situation. We have a data-generating process with three variables $X \leftarrow Z \rightarrow Y$ where Z is the common cause of both X and Y, which can be written as:

$$X = f_x(Z) + \epsilon_x \tag{5a}$$

$$Y = f_y(Z) + \epsilon_y \tag{5b}$$

with independent normally distributed noise terms $\epsilon_x$ and $\epsilon_y$. In this section, the relation between variables is linear so that both functions take the form $f(z) = Az + B$ with A and B being constant terms. Pearson Correlation is a statistical test usually used to find linear relations between variables. In this case, a significant correlation will be found between variables X and Y, as they are both driven by a common confounder. However, we know that these variables must become independent when conditioning on Z. Partial correlation builds upon these ideas and tests whether X and Y still have a linear relation once the effect of Z is subtracted. Therefore, a linear regression model $\hat{f}(Z)$ is fitted for equations 5a and 5b, and the Pearson Correlation Coefficient is computed between the residuals $r_x = X - \hat{f}_x(Z)$ and $r_y = Y - \hat{f}_y(Z)$ [21]. A numerical example [1] of the previously explained situation can be visualized in Figure 7.

This straightforward conditional independence test can help us to better understand how the PC-MCI algorithm works. We illustrate this in the following example inspired in [23] and examples found in [6]. This example has been programmed in a Jupyter Notebook [2]. With this example, we want to first get acquainted with different Tigramite's classes and functions, used for data generation, visualization, and causal temporal graph inference. And second, to gain insight into the algorithm's inner workings by replicating its results using more basic Python.

We consider a linear data-generating process, described by the following equations:

$$\begin{aligned} A_t &= 0.7A_{t-1} - 0.8B_{t-1} + \eta_t^A \\ B_t &= 0.8B_{t-1} + 0.8D_{t-1} + \eta_t^B \\ C_t &= 0.5C_{t-1} + 0.5B_{t-2} + 0.6D_{t-3} + \eta_t^C \\ D_t &= 0.7D_{t-1} + \eta_t^D \end{aligned} \tag{6}$$

where $\eta^x$ are independent zero-mean unit variance random terms. In a real-world problem, we would have 4 temporal series (see Figure 8) and would try to unveil their drivers.

The algorithm is also able to try different significance level values $\alpha_{PC}$ in order to find a good trade-off between false positives and false negatives, using information theory criteria. However, a significance threshold of $\alpha_{PC} = 0.05$ and a maximum time lag of $\tau_{max} = 4$ have been defined to simplify the example. We also forget about contemporaneous links in this case. In the first step, correlation is tested for all possible combinations between $X_t$ and any other variable at any lag.

Focusing on $D_t$, only two significant correlations have been found. Therefore, the set of possible parents has been quickly reduced to $P(D_t) = \{D_{t-1}, D_{t-2}\}$, with $D_{t-1}$ having the strongest correlation. Consequently, independence between $D_t$ and $D_{t-2}$ conditioning on $D_{t-1}$ is tested using Partial Correlation. The Pearson Correlation Coefficient between residuals obtained from regressing $D_t$ and $D_{t-2}$ as a linear function of $D_{t-1}$ is $\rho(r_{D_t}, r_{D_{t-2}}) = -0.01$ with p-value = 0.7. Thus, we conclude that they are conditionally independent and $D_{t-1}$ is the only possible parent of $D_t$.

[1] Available in `https://github.com/oriolus98/climate_classifier_causal_discovery/blob/main/notebooks/partial_correlations.ipynb`

[2] Available at `https://github.com/oriolus98/climate_classifier_causal_discovery/notebooks/pcmci_intro.ipynb`

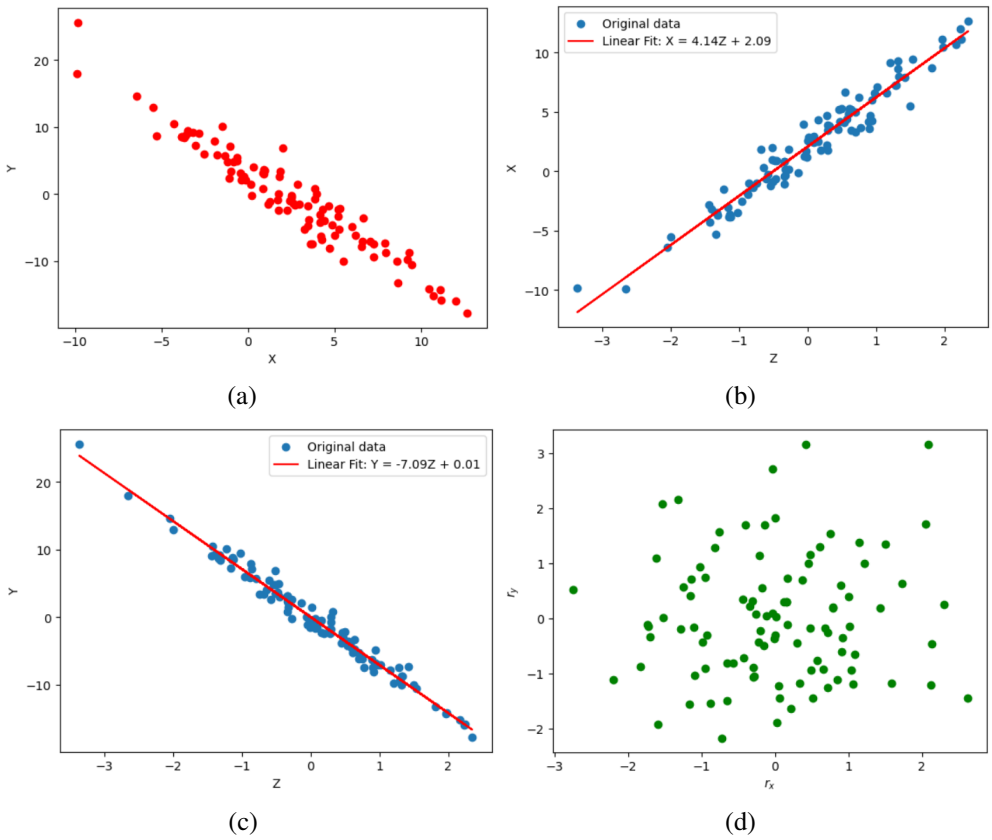



Figure 7: Numerical example of a linear data generating process $X \leftarrow Z \rightarrow Y$. (a) Variables X and Y appear linearly correlated due to a common cause Z, with $\rho(X, Y) = -0.96$ and p-value = $3.2 \cdot 10^{-56}$. We want to test whether this dependency is only due to the presence of a confounder. Therefore, we can fit a linear regression (b) $X = \hat{f}_x(Z)$ and (c) $Y = \hat{f}_y(Z)$. (d) Finally, we can observe the residuals from both regressions are not correlated. This is equivalent to saying there is no linear relation between variables X and Y that cannot be explained by Z. Our conditional independence test yields results $\rho(r_x, r_y) = 0.06$ and p-value = 0.58. Thus, X and Y are independent given Z.

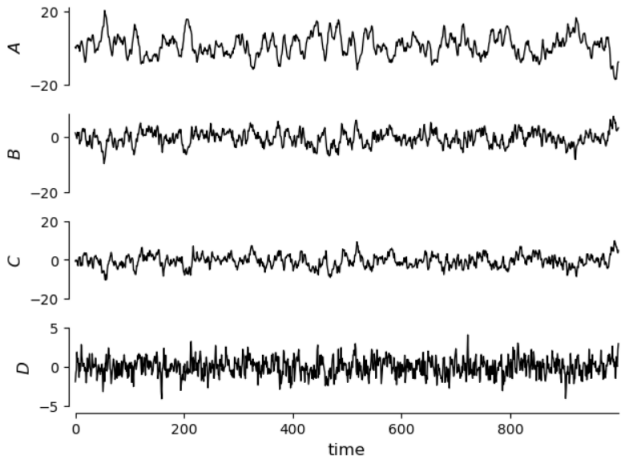



Figure 8: Plot of the variables generated following equation 6 using Tigramite's visualization tools.

Focusing now on $B_t$, all possible parents are significantly correlated on the first stage, with $B_{t-1}$ having the strongest correlation. Next, Partial Correlation is tested for all possible combinations conditioning on $B_{t-1}$. Conditional independence is also tested between $B_t$ and $B_{t-1}$ conditioning on the second strongest correlated possible parent $B_{t-2}$. Results are ordered by the test statistic value as in Table 2. With a significance threshold of $\alpha_{PC} = 0.05$, the set of possible parents is reduced to the seven first rows in Table 2. In the next iteration, Partial Correlation between $B_t$ and any variable in $P(B_t)$ is computed conditioning on the two strongest possible parents $B_{t-1}$ and $D_{t-1}$ (multilinear regression now). Finally, the algorithm would converge, and the only significant positive test results correspond to the set of possible parents $P(B_t) = \{B_{t-1}, D_{t-1}\}$. Following the same methodology, we were able to compute a set of possible parents for each variable:

$$\begin{aligned} P(A_t) &= \{A_{t-1}, B_{t-1}\} \\ P(B_t) &= \{B_{t-1}, D_{t-1}\} \\ P(C_t) &= \{C_{t-1}, B_{t-2}, D_{t-3}\} \\ P(D_t) &= \{D_{t-1}\} \end{aligned} \tag{7}$$

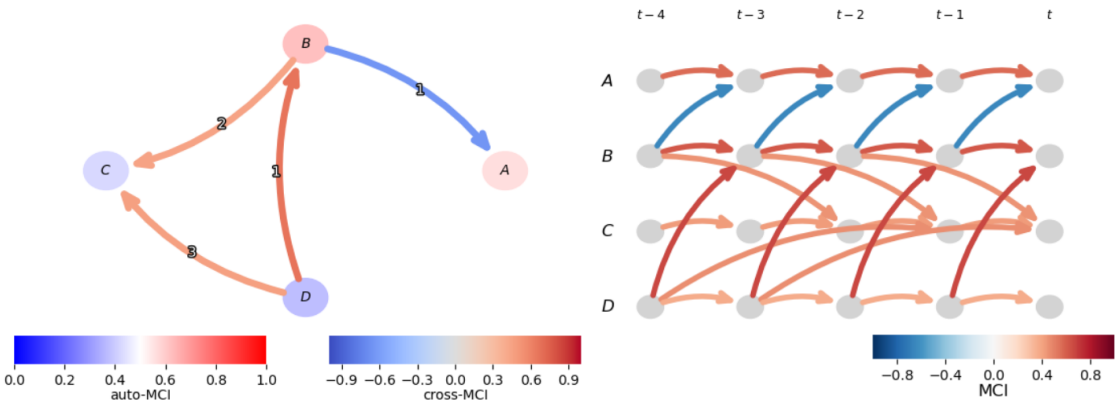
Now, in the second step of the algorithm (MCI), Partial Correlation is calculated between each variable and its parents, conditioning on the set of parents of its parents. For example, we can test if $D_t$ and $D_{t-1}$ remain dependent when conditioned on $D_{t-2}$, which is the case ($\rho = 0.36$, p-value $= 2.7 \cdot 10^{-33}$). The MCI test for $B_t$ would be:

$$\begin{aligned} \rho\left(B_t, B_{t-1} \,|\, B_{t-2}, D_{t-2}\right) &= 0.49, p - value = 9.8 \cdot 10^{-63} \\ \rho\left(B_t, D_{t-1} \,|\, D_{t-2}\right) &= 0.36, p - value = 1.9 \cdot 10^{-31} \end{aligned} \tag{8}$$

In this example, all possible parents pass the MCI test. The algorithm has been able to find the causal drivers of our data, as can be shown by comparing equations 6 and 7. In the case of linear dependencies between variables, the causal effect (expected change in the outcome variable due to a given intervention on its causal parents) can be estimated through a multivariate linear regression of $X_t^j$ on just its parents $P(X_t^j)$. This approach has been shown to outperform other methods [10]. This analysis can be replicated and visualized using Tigramite (see Figure 9).

| Variable | $\rho$ | p-value |
|---|---|---|
| $B_{t-1}$ | 0.69 | $3.9 \cdot 10^{-144}$ |
| $D_{t-1}$ | 0.68 | $9.3 \cdot 10^{-140}$ |
| $D_{t-2}$ | 0.25 | $2.5 \cdot 10^{-15}$ |
| $B_{t-2}$ | -0.13 | $2.9 \cdot 10^{-5}$ |
| $B_{t-3}$ | -0.09 | $2.5 \cdot 10^{-3}$ |
| $A_{t-1}$ | 0.09 | $3.7 \cdot 10^{-3}$ |
| $A_{t-2}$ | 0.08 | 0.013 |
| $A_{t-3}$ | 0.05 | 0.12 |
| $B_{t-4}$ | -0.04 | 0.17 |
| $A_{t-4}$ | 0.04 | 0.17 |
| $C_{t-1}$ | -0.04 | 0.19 |
| $C_{t-2}$ | -0.03 | 0.31 |
| $D_{t-3}$ | 0.03 | 0.36 |
| $C_{t-4}$ | -0.02 | 0.37 |
| $C_{t-3}$ | -0.02 | 0.47 |
| $D_{t-4}$ | -0.01 | 0.68 |

Table 2: Partial Correlation test results between $B_t$ and any other variable conditioning on the strongest possible parent obtained in the previous iteration, ordered by test result.




(a) (b)

Figure 9: Reconstruction of the causal structure of data generated following equation 6, using Tigramite's plotting package. (a) Node color denotes the strength of autocorrelation for each variable. Link color indicate the estimated causal effect. The numerical label designate the strongest link's temporal lag. (b) Although the visual appeal of the previous graph seems preferable, this time series graph more accurately depicts the spatio-temporal dependency structure, enabling the identification of causal pathways.

### 3.1.3 Non-linear dependencies: CMI-knn test

Although the conditional independence test presented in the previous section is quite straightforward, it is not capable of detecting non-linear relationships between variables, usually present in dynamic complex systems. Therefore, various conditional independence tests have been formulated using a non-parametric approach, avoiding assumptions about data distribution or dependencies among variables. Generally, these tests should be able to capture any kind of conditional relations (linear or non-linear). However, these algorithms are very data-intensive, as well as time and computationally expensive. Therefore, if expert knowledge allows us to make assumptions about data dependencies, simpler tests are recommended [21]. In this section our focus will be on a non-parametric Conditional Mutual Information test developed by [24] and included in Tigramite.

Some concepts from Information Theory must be introduced in order to understand this test. Shanon entropy provides a quantitative measure of the amount of uncertainty or surprise associated with a set of possible outcomes from a random variable. It may be understood as the average amount of information needed to specify the value of this variable. If a random variable has low entropy, it means it has low uncertainty, and we can make more accurate predictions about its value. Shanon entropy $H(X)$ of a random variable $X$ with probability distribution $p(x)$ is defined as:

$$H(X) = -\sum_{x \epsilon X} p(x) \log(p(x)) \tag{9}$$

Kullback-Leibler Divergence is a measure of the difference between two different probability distributions $p(x)$ and $q(x)$:

$$D_{KL}(p \,||\, q) = \sum_{x \epsilon X} p(x) \log\left(\frac{p(x)}{q(x)}\right) \tag{10}$$

KL Divergence is non-negative and zero if and only if the two distributions are identical.

From the previous definition, the concept of Mutual Information between random variables $X$ and $Y$ can be introduced as the KL Divergence between the joint distribution of $X$ and $Y$, and the product of their marginal distributions:

$$MI(X, Y) = \sum_{x \epsilon X} \sum_{y \epsilon Y} p(x, y) \log\left(\frac{p(x, y)}{p(x)p(y)}\right) \tag{11}$$

Mutual Information measures the amount of information that knowing the value of one random variable provides about the other one. When the two variables are independent the joint distribution of $X$ and $Y$ equals the product of their marginal distributions and, therefore, $M(X, Y) = 0$. Mutual Information provides an accurate metric for assessing the relationship between variables, surpassing alternative statistical tests such as Pearson Correlation. Its advantage lies in its capability to handle both discrete and continuous variables, while also unveiling non-linear dependencies [25]. It can also be written in terms of entropy:

$$MI(X, Y) = H(X) + H(Y) - H(X, Y) \tag{12}$$

The definition of Mutual Information for discrete variables as in Equation 11 can be extended to the continuous case:

$$MI(X, Y) = \int_X \int_Y p(x, y) \log\left(\frac{p(x, y)}{p(x)p(y)}\right) dx\, dy \tag{13}$$

In the discrete case, joint and marginal probabilities can be estimated from the frequencies at which different possible outcomes appear in a sample. However, in the continuous case, estimating the probability distribution remains challenging due to the infinite nature of possible outcomes. One interesting approach is based on k-nearest neighbours statistics. An article by [26] showed that the entropy of a random variable $X$ can be estimated from the distance between neighbor observations in a sample of size $N$ as:

$$H(X) \approx \frac{1}{N-1} \sum_{i=1}^{N-1} log(x_{i+1} - x_i) + \psi(1) - \psi(N) \tag{14}$$

where $\psi(x) = \Gamma'(x) \cdot \Gamma^{-1}(x)$ is the digamma function. An approximation of the Mutual Information between two random continuous variables can then be computed by combining Equations 12 and 14 [25]. Building upon these ideas, a non-parametric Conditional Independence test has been developed by [24]. Conditional Mutual Information between continuous variables $X$ and $Y$, conditioned on $Z$ can be defined as:

$$CMI(X, Y \,|\, Z) = \int_X \int_Y \int_Z dxdydz\, p(x, y, z) \log\left(\frac{p(x, y \,|\, z)}{p(x \,|\, z)\, p(y \,|\, z)}\right) = H_{XZ} + H_{YZ} - H_Z - H_{XYZ} \tag{15}$$

It can be interpreted as the amount of information shared by $X$ and $Y$ not present in $Z$. From the definition of Conditional Independence (see Equation 4) follows the fact that CMI will be zero if and only if $X$ and $Y$ are independent conditioned on $Z$. Therefore, if we are able to find a good estimate of CMI from our sample data, then it allows us to test if the null hypothesis $H_0 : X \perp\!\!\!\perp Y \,|\, Z$ holds true. This test consists of two different steps. First, the test statistic value $CMI(X, Y \,|\, Z)$ must be computed for our data. From the k-nearest neighbors entropy approximation, CMI can be estimated as:

$$CMI(X, Y \,|\, Z) \approx \psi(k_{CMI}) + \frac{1}{N} \sum_{i=1}^{N} [\psi(k_i^z) - \psi(k_i^{xz}) - \psi(k_i^{yz})] \tag{16}$$

Once fixed the only parameter $k_{CMI}$ (number of nearest neighbors), the distance (see Equation 17) from each point to its k-th nearest neighbor $\epsilon_i$ can be computed. Then, $k_i^{xz}$ is defined as the number of neighbors within this distance $\epsilon_i$ in the subspace $X \otimes Z$; $k_i^{yz}$ will be the number of neighbors in the subspace $Y \otimes Z$; and $k_i^z$ in the subspace $Z$. This estimator is more efficient than previous alternative approaches (for example, discretizing continuous variables with a fixed bandwidth) because it is data-adaptive: hypercubes are smaller within more densely populated areas. The distance between two points $\vec{x} = (x, y, z)$ and $\vec{x'} = (x', y', z')$ is calculated as the maximum norm, as proposed in [25]:

$$\epsilon_i = max\{||x - x'||, ||y - y'||, ||z - z'||\} \tag{17}$$

In the second step, the obtained value must be compared with the expected distribution of the test statistic value under the assumption that the null hypothesis is true. However, to this day, there is no theoretical definition of this distribution. Therefore, a permutation-based approach is employed to simulate it. The goal of this procedure is to generate samples of the null distribution by randomly shuffling our original dataset in a way that the conditions expressed in $H_0$ are preserved, and subsequently computing the CMI value for each instance. Thus, after permutation, any possible dependence between $x$ and $y$ must be destroyed, while the relationship with $z$ must be preserved. The distribution of the test statistic under the null

hypothesis can be empirically estimated by repeating this process multiple times. The p-value of the test can then be derived as the fraction of simulated CMI's larger than the value obtained in the first step.

In the present algorithm [24], permutation is also achieved with a k-nearest neighbors approach. Only $x_i$ values are substituted by other $x_j$ values among their nearest neighbors in the subspace $Z$, in order to preserve $H_0$. Therefore, another parameter $k_{perm}$ indicating the number of neighbors in the second step must be fixed. The mentioned publication [24] also offers insights into how to determine the optimal parameters $k_{CMI}$ and $k_{perm}$. It also presents simulated results suggesting that it is a robust option when dealing with non-linear relations between variables and small samples. Further theoretical and empirical research is however recommended. This Conditional Independence test is also included in the Tigramite package and can be used on all steps of PC-MCI algorithm, although its computational cost remains its main limitation. The causal structure of our toy example (Equation 6) was also unveiled using this test, requiring an execution time of 78 minutes. In contrast, a similar result was achieved in a mere 0.7 seconds interval employing Partial Correlation.

Now that the theoretical foundations of the PC-MCI algorithm have been comprehended, the next section will delve into its application in real climatic data.

## 3.2 Obtaining causal graphs from climatic data

Once the theoretical background has been introduced, the PC-MCI algorithm will be used to try to uncover causal relationships between different climatic variables.

### 3.2.1 Working with climatic data

For our first experiment, a dataset ensembled by [27] has been selected. This dataset contains several different variables, based on *in situ* and satellite measurements, with a monthly temporal resolution and $1^\circ \times 1^\circ$ latitude-longitude spatial resolution, spanning from 1981 to 2010. For the sake of simplicity, only four detrended and deseasonalized variables, without missing values, have been chosen. Each of them represents one of the major climatic variables present in the dataset: `max_5consec_Ix5dayRN_ERA_Residuals` (Radiation), `T_CRU_Residuals` (Temperature), `monthly_Rx1day_P_CPCU` (Precipitation), and `max_5consec_Sx5daySM_GLEAM_Residuals` (Soil Moisture). This results in a very complex dataset with dimensions $13072 \times 384 \times 4$, corresponding to *Earth pixel* $\times$ *time index* $\times$ *climatic variable*. A common approach in this cases consists on summarizing data into non-spatial components. Either incorporating external expert knowledge or through the use of Varimax-rotated Principal Component Analysis [28], some specific regions (known as principal modes of climatic variability) are selected, and causal relations are studied between these regions. A simplified version of this approach will be used on this project. Causal relationships will be evaluated for each separate pixel, and then different causal structures corresponding to different geographical regions will be compared. This method does not account for the probable causal influences between different geographical areas. However, a comprehensive analysis of this complex question lays beyond the scope of this master thesis. In fact, there exists some controversy regarding if current approaches truly take advantage of spatio-temporal relations. For example, [29] suggests that the previously described methods have demonstrated strong performance in well-studied problems (such as El Niño Southern Oscillation), but may not be suitable for new cases.

The Tigramite package includes different visualization tools. However, a simple visual inspection of the temporal series (e.g. Figure 10), cannot provide useful information in this context, besides possible problems within our dataset (such as non-stationarity, missing data, etc.), and is not practical given the high amount of geographical pixels. In this scenario, an interesting approach to better understand and visualize our dataset consists on temporal series clustering.

### 3.2.2 Temporal Series clustering

Clustering is a common technique in data analysis that groups similar data points together, making it easier to identify patterns within a dataset. One common clustering algorithm is K-means, which divides a dataset into a pre-determined number of clusters K by minimizing the distance between data points and their corresponding cluster centroids. The algorithm iteratively assigns data points to the nearest centroid (usually measured using Euclidean distance) and then updates the centroids as the mean of the points in each cluster. While K-means is effective for static data, it can be less suitable for time series data.

Using Euclidean distance to measure similarity between time series can be problematic because it requires the data to be aligned perfectly, which is often not the case. Small misalignments in time can lead to large Euclidean distances even if the overall shapes of the

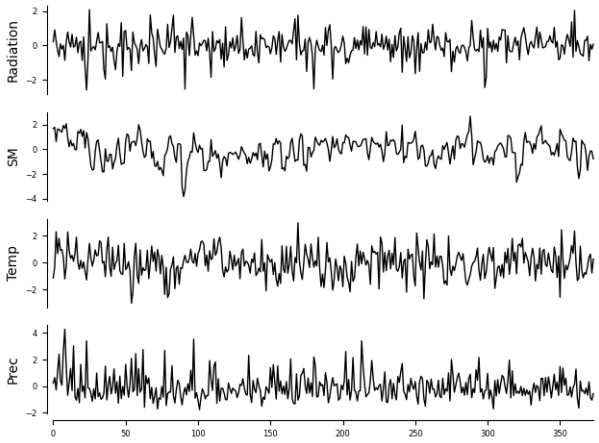



Figure 10: Example climatic temporal series corresponding to coordinates (47.5 °, 33.5 °)

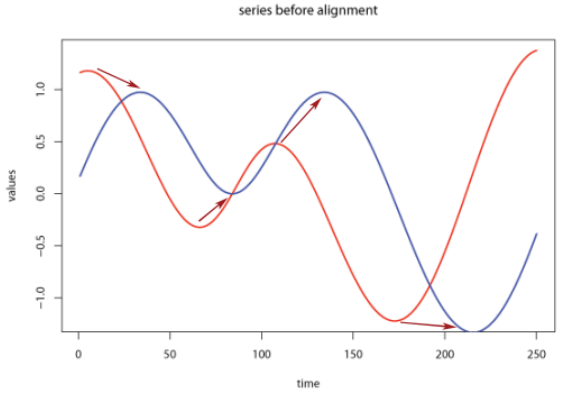
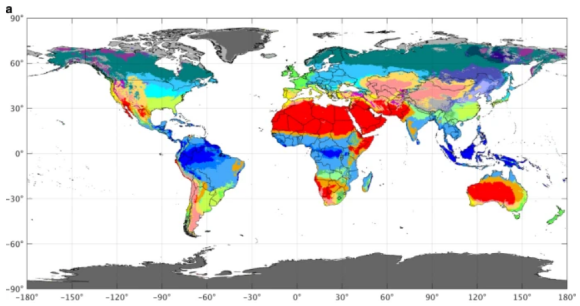



Figure 11: These two temporal series follow a very similar pattern. However, Euclidean distance between data points at the same temporal index cannot find this similarity due to the time lag between both series. Arrows show the desirable points of alignment, which can be found using the DTW algorithm [32].

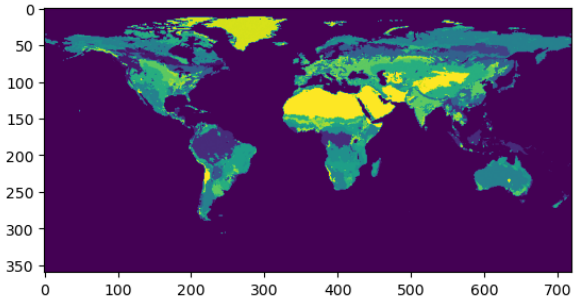
series are similar. This makes Euclidean distance inappropriate for detecting the true similarity between time series (see Figure 11).

Dynamic Time Warping (DTW) is a more suitable distance metric for time series data. It was first designed for speech recognition in the 60s, and then introduced to temporal series analysis in 1994 [30]. DTW measures similarity by allowing for non-linear alignments between time series. Instead of comparing each point directly in time, DTW warps the time axis, aligning points that are similar but may occur at different times. This flexibility makes DTW effective at capturing patterns in time series that Euclidean distance would miss, making it a valuable tool for time series clustering. Both univariate and multivariate time series k-means clustering algorithms based on DTW metric are implemented in the Python package tslearn [31].

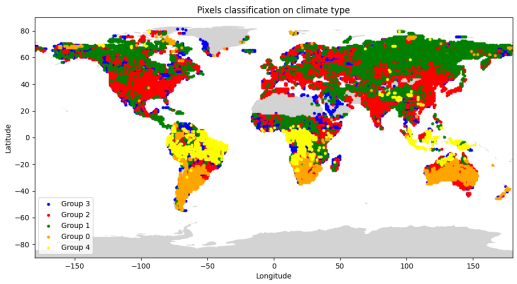
On this project, time series clustering will allow us to identify groups or geographical regions where climate variables behave differently, even in our current unsupervised scenario where we don't have any labeled data. The proposed algorithm requires us to predefine the number of clusters we expect to find. While there are data-driven methods available to help determine this number, such as Silhouette analysis, the guidance of expert knowledge may be very helpful in complex problems. Luckily, climate science research has established




(a)



(b)




(c)

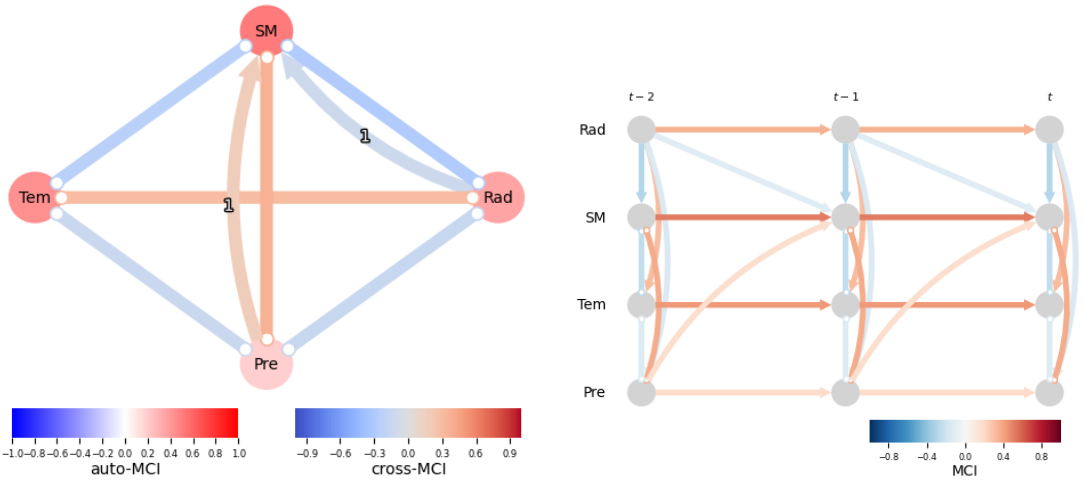
Figure 12: Climate clusters comparison (a) Present Köppen-Geiger climatic classification [34]. The five main categories are divided in subcategories (b) IGBP landcover classification around the world, obtained from Moderate-Resolution Imaging Spectroradiometer data following [35] (c) Climatic groups obtained from clustering our temporal dataset

five main climate categories: tropical, arid, temperate, continental and polar, as presented by Köppen climate classification [33, 34]. Figure 12 compares the obtained clusters with established climatic and vegetation zones. It is interesting to see how the results obtained with our unsupervised approach are similar to accepted climatic and vegetation zones, determined with more complex methods. Especially if we consider that this classifications are usually based just on temperature and precipitation, where seasonality plays an important role, where our time series are stationary.

### 3.2.3 Causal Discovery

Building upon the information developed in the previous sections, we are now able to recover the causal structure of our dataset directly from data.

One interesting feature included in Tigramite allows to introduce link assumptions in




(a) (b)

Figure 13: Example of the obtained causal graph computed for the average series of ten different randomly selected pixels on the tropical cluster, with a 0.01 significance level. In order to obtain these graphs, both linear and non-linear conditional independence tests where compared. However non-linear CMI-knn test was only able to identify the causal link between precipitation and soil moisture, whereas the causal graph obtained using Partial Correlation is more consistent with physical principles. Moreover, the Partial Correlation method yielded results substantially more rapidly. Therefore, the linear approach has been selected for subsequent analysis.

the causal discovery process. Therefore, available expert knowledge can be used by the algorithm to simplify the process, as some links may be directly removed or directed. For example, in this scenario we can clearly assume that radiation comes from the sun, and therefore it cannot be physically caused by temperature or precipitation on Earth. Thus, we introduce in the model an assumptions graph where any link between radiation and any other variable must be directed from the former to the latter, while all other possible links must be checked by the algorithm.

Regarding the fact that we have 13071 different pixels to analyze, the first idea was to average time series within the same climatic cluster obtained in the previous section. We assume that the causal structure may be different for each climatic zone. However, averaging over thousands of temporal series resulted in smoothed series with a significant loss of information. Therefore, Figure 13 shows an example of the resulting causal graph computed for the average series of ten different randomly selected pixels on the tropical cluster (group 3 in Figure 12(c)).

At first glance, we can appreciate that the algorithm correctly identifies the positive causal relationship between radiation and temperature, or precipitation and soil moisture, as well as the negative relationship between temperature and soil moisture. However, it stands unclear how can this graph provide more useful information, as all variables remain coupled, and the majority of the links are undirected. This question will be addressed in the following sections.

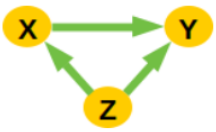



Figure 14: Causal graph of the example system described by Equations 18a, 18b and 18c

### 3.2.4 Causal Effects estimation

Once the causal graph of a given system has been established, a typical problem in Causal Inference consists on estimating the causal effect of a variable on another one.

In the causal graphical framework developed by Pearl [36], the existence of a causal DAG like Figure 14 implies the existence of an underlying Structural Causal Model such as:

$$X := f_x(Z, \epsilon_x) \tag{18a}$$

$$Y := f_y(X, Z, \epsilon_y) \tag{18b}$$

$$Z := f_z(\epsilon_z) \tag{18c}$$

It is important to note that these structural causal equations fundamentally differ from traditional equations in their lack of symmetry. Unlike traditional equations where variables can be freely moved from one side to the other, structural causal equations represent directed relationships. They explicitly model how changing one variable affects another, but not vice versa.

In the defined model, given random variables X and Y, an intervention $do(X = x)$ assigns a fixed value to the variable, modifying its structural equation to $X := x$, which is equivalent to removing the causal link from Z to X in the graph. In this scenario we define the interventional probability distribution $P(Y \mid do(X = x))$, which can only be computed from data in some cases, and is fundamentally different from the conditional probability distribution $P(Y \mid X = x)$.

Thus, the average causal effect comparing interventions $do(X = x)$ and $do(X = x')$ is defined as:

$$\Delta_{yxx'} = \mathbb{E}[Y|do(x)] - \mathbb{E}[Y|do(x')] \tag{19}$$

Generally, it is not possible to compute interventional probability distributions from data (e.g. we cannot apply an intervention over a given population and compare it to the same population without the intervention). However, the graphical framework establishes criteria and assumptions to determine if a given interventional expression can be rewritten in terms of traditional probabilities. Therefore, in these cases the average treatment effect can be estimated directly from observational data.

As commented in Section 3.1.2, for the linear case, assuming a known causal graph, the average treatment effect defined as the expected change in the outcome variable caused by a one unit increment in the cause variable can be estimated through a linear regression on the outcome variable parents, being equivalent to the regression coefficients [10]. These causal effect can also be estimated using Tigramite package's class CausalEffects.

Therefore, for each pixel of the dataset introduced in Section 3.2.1, first the causal graph has been identified, and then, causal effects between all linked variables have been estimated.

For the sake of simplicity, a maximum time lag of $\tau_{max} = 2$ months has been established. Figure 15 shows the geographical distribution of causal effects strength between some relevant causally linked variables.

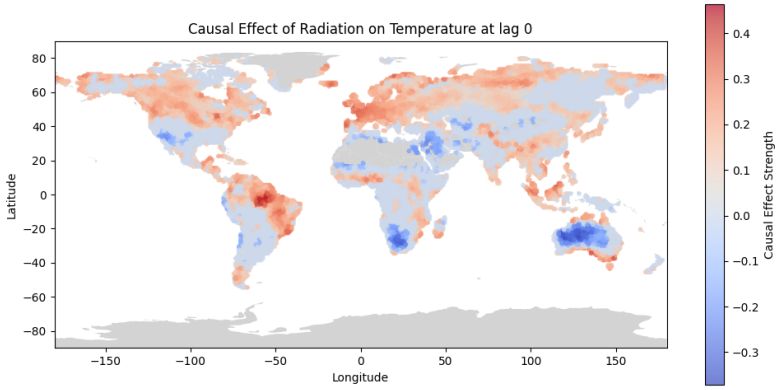
When analyzing these results, we face one of the main limitations when dealing with causal inference: the scarcity of validated benchmarks that allow us to numerically evaluate the performance of our algorithm. In this sense, an study by [37] has been used in order to qualitatively compare the results obtained. This paper presents similar geographical causal forcings using a Robust Convergent Cross-Mapping method. However, their dataset has a significantly higher temporal and spatial resolution, and includes more variables, which can impact the quality of the results (its implications will be described in the next section). Furthermore, results for each causal link are not presented individually but indicating which forcing is predominant for each variable, preventing a direct comparison with our results.

The presented results generally seem reasonably consistent with established physical principles. For example, Figure 15(c) shows a negative impact of Soil Moisture on Temperature, with this relationship being stronger in the geographical areas where a significant relationship was also found in the comparative paper. Effectively, different studies suggest that an increase in Soil Moisture results in cooler temperatures [38] [39].

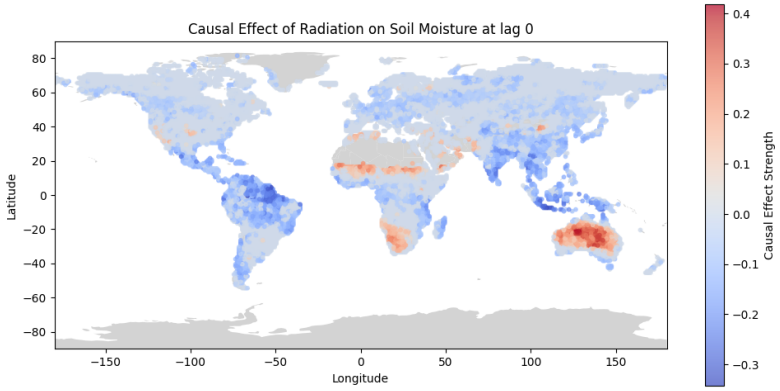
We can also observe a predominantly positive effect of Radiation over Temperature (Figure 15(a)), as well as a predominantly negative effect of Radiation on Soil Moisture (Figure 15(b)). However, as established in literature, the strength of this forcings, and even the sign of the relationship, can vary depending on the geographical region. The reason is the fact that Climate is a highly coupled complex system with numerous interconnected feedback loops that can amplify or diminish the effects of individual factors. For example, radiation is the primary mechanism by which the sun warms the Earth. However, some studies [40] suggest that surface temperatures might diminish over desertified areas, due to reduced absorption of short-wave radiation by the brighter surface. This could explain the negative causal effect on desertic zones observed in Figure 15(a).

Nevertheless, our confidence in the algorithm and the obtained results is undermined by the absence of quantitative benchmarks and metrics, as well as the presence of some unexpected results. As shown in Figure 15(d), the causal effect of Precipitation on Soil Moisture found by the algorithm is generally very weak and negative, which does not sound reasonable. Further discussion on possible reasons for these disagreements, as well as interesting subsequent research directions are presented in the next section.

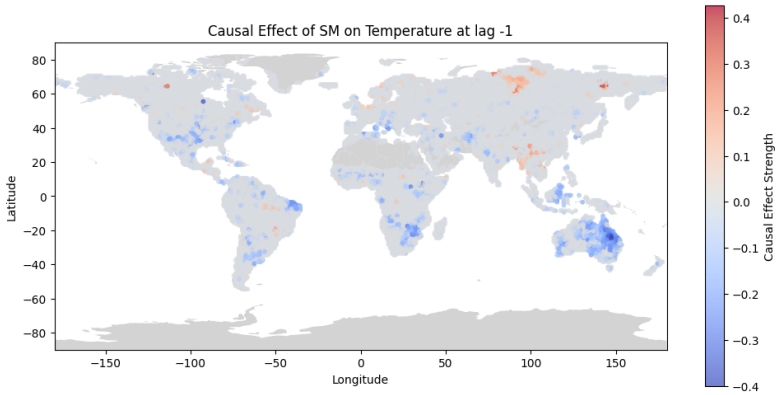
Following the example of [41], where causal effects between variables driving different European rivers discharges have been classified using a clustering algorithm, a similar analysis has been performed in order to conclude our study. The main objective consists on identifying and visualizing differences between the causal relationships among climatic variables depending on geographical zones. Therefore, a k-means analysis using the Euclidean distance has been applied to our dataset containing the causal effects between all existing links for each geographical pixel. Figure 16 shows the results obtained. First, our assumption that causal structure may differ depending on the climatic zone seems to hold, as geographical regions with similar causal effects generally correspond to regions with different climatic patterns delimited in Section 3.2.2. However, we still see some different patterns such as tropical regions in Brazil and desertic regions in Arabian peninsula sharing the same cluster. If we take a look at clusters centroids (Figure 16(b)), we see that the algorithm gave too much importance to autocorrelation links, whereas other relationships causal effects have been averaged to near zero values (which suggests that both positive and negative causal effects in these cases have been included in the same cluster). In order to explore



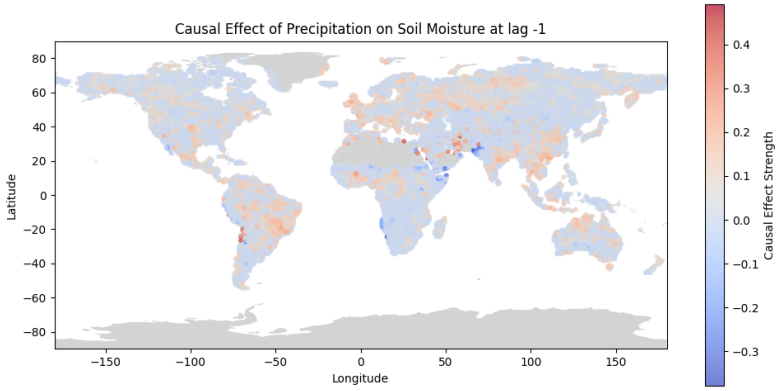

(a)




(b)




(c)




(d)

Figure 15: Geographical distribution of relevant causal relations forcing's strength

these differences, the same analysis has been performed retaining only cross-variable causal effects (see Figure 17). Now we are able to identify some interesting patterns: the negative influence of radiation in temperature over desertic areas, compared to its positive influence over clusters 0 and 2 (including different tropical and continental zones); or the existence of some specific areas where Soil Moisture has a strong positive influence on precipitation. The geographical distribution of these maps might suggest that the distribution of different causal structures around the globe could be driving the existence of separated climatic zones. Still, the difference from traditional climatic zones is explained by the fact that these traditional regions are defined only by temperature, precipitation and vegetation patterns, but do not consider the interaction between these variables. However, this second approach still seems to be losing information, as most possible links causal effects' centroids remain near to zero.

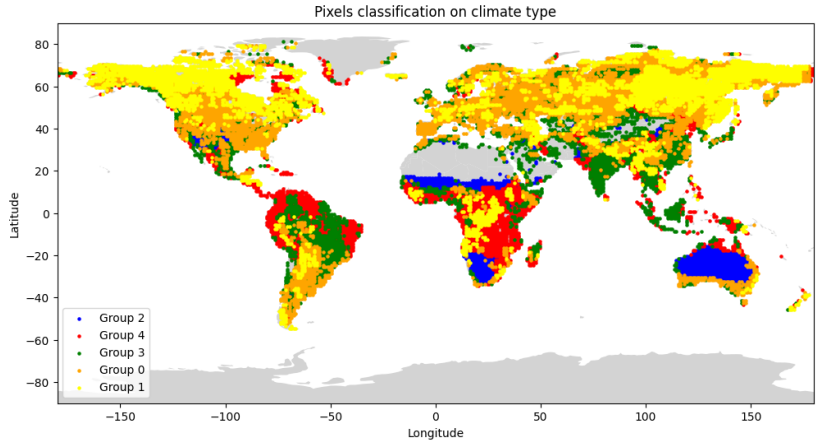
### 3.2.5 Discussion on Tigramite and future research

On the one hand, this extensive analysis of both PC-MCI's theoretical foundations and applied features show how the algorithm is able to identify causally relevant links from raw time-series data, and even quantify the strength of this relationships. The results obtained are generally plausible and consistent with established physical principles, and show how some of the most important climatic variables interact between them at different geographical zones.

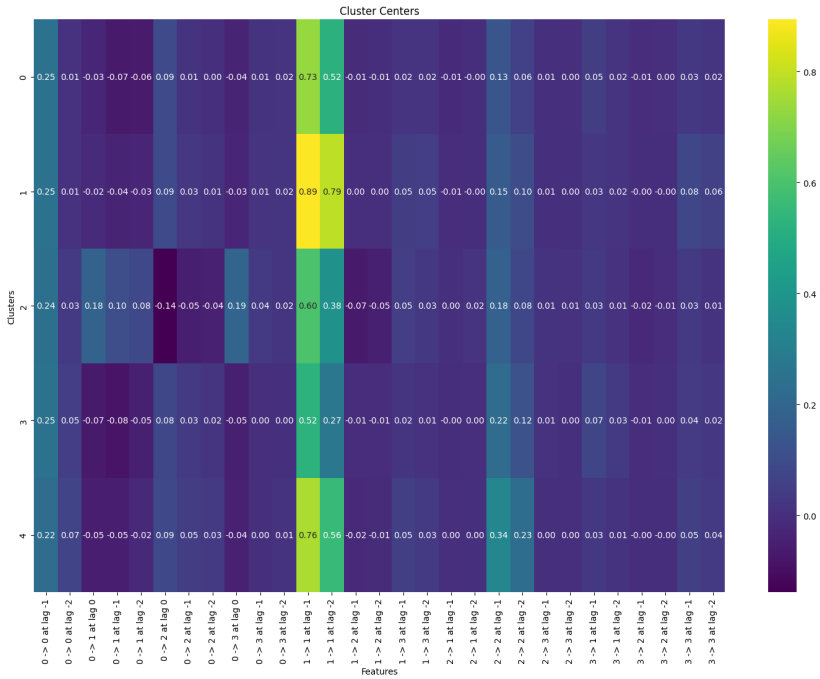
Therefore, this method is a promising advancement in the field of temporal causal discovery. There are multiple studies quantifying causal effects in Climate Sciences using methods like Robust Convergent Cross-Mapping [37], non-linear Granger-causality [27], etc. However, PC-MCI incorporates a powerful previous step consisting on identifying the causal graph, which besides being necessary for a correct causal identification, contributes to the explainability and visualization of the results. Other significant advantages of this algorithm lie in its ability to detect causal relations at different time lags, as well as the possibility of intuitively introducing experts' assumptions.

Several questions remain open in Climate Science, and this algorithm may shed light when identifying the drivers of extreme events or when estimating simulation parameters from data, for example. In fact, different studies are already using it for classifying European rivers characteristics [41], discovering drivers of Indian monsoon rainfall [23], establishing the causal and temporal relationship between whether, vegetation and wildfires [42], etc.

On the other hand, some limitations have been encountered during the process. First, we should acknowledge the fact that some of the fundamental assumptions of the algorithm may not hold in our experiment. Specifically, it does not seem reasonable to assume that all causally relevant variables for this problem are included in our dataset, which violates the assumption of Causal Sufficiency. For example, [37] shows how also vegetation and latent heat (latent energy that is absorbed or released due to water phase transitions or plants transpiration) are also causally connected to Temperature, Soil Moisture and Precipitation. Furthermore, our analysis was conducted using a linear conditional independence test. As a result, we could expect that some patterns may have not been detected by our method, as non-linear interactions between variables are often present in complex systems like Climate. Finally, Tigramite's tutorials [6] also show how time-series sampling frequencies below the underlying time-dependency result in the detection of false and counter-intuitive links. In the context of Climate system, interactions appear at different time scales. Therefore, a monthly sampling frequency should be enough to detect true causal links. However, a dataset with higher frequency probably will reveal new significant causal links, that could even be

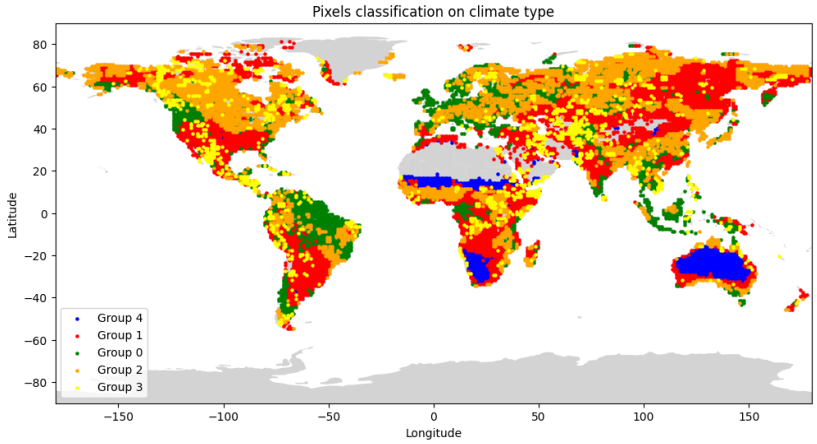



(a)

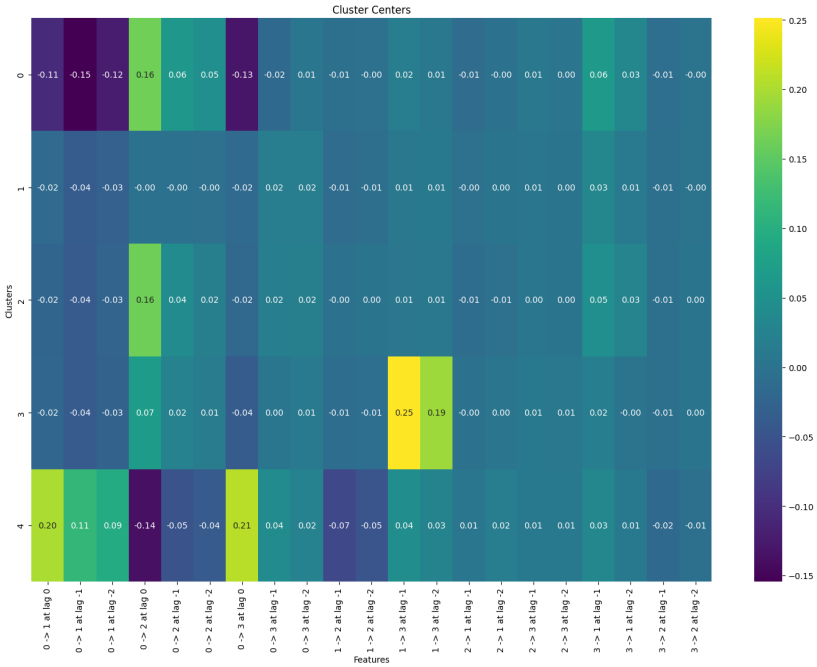



(b)

Figure 16: (a) Geographical classification of causal effects using k-means (b) Average value of each causal effect for each cluster centroid.




(a)




(b)

Figure 17: (a) Geographical classification of causal effects using k-means, for cross-variable links (b) Average value of each causal effect for each cluster centroid, only for cross-variable links.

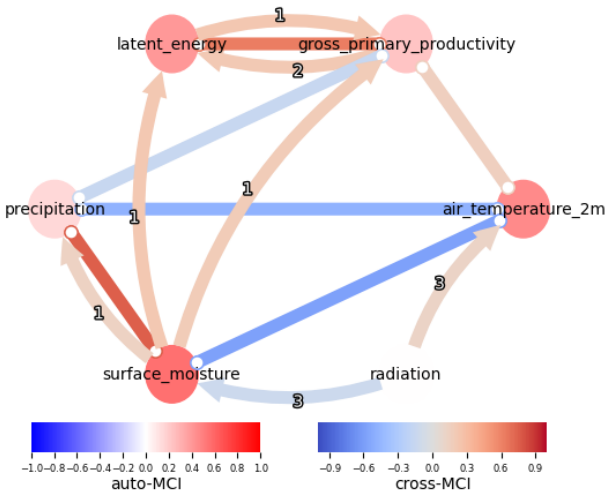



Figure 18: Causal graph of averaged tropical pixels time-series, extracted from the dataset proposed for future research [37]. All identified links seem consistent with physical principles and with results from the paper where the dataset is collected. We also find a strong positive relationship between precipitation and surface moisture, which was not found in our experiment. The algorithm is also able to orient a bigger fraction of the detected links. These exploratory outcomes suggest that this future experiment would identify veryinteresting patterns.

dominant over the ones discovered in our experiment. Moreover, results could be more exact, as contemporaneous unoriented links that appear for example in Figure 13(a), would probably be oriented.

The partially consistent results obtained suggest that the algorithm is relatively robust against these limitations. However, all of this information must be considered in order to conduct an improved version of the experiment. The dataset collected by [37] could be an interesting starting point, as it addresses the violation of Causal Sufficiency (it includes Latent Heat and Gross Primary Productivity,in addition to the four variables used in our study), and it also has a higher sampling frequency (8 days). Non parametric conditional independence tests capable of detecting non-linear relationships can also be analyzed making use of Tigramite's features. Exploratory results (Figure 18) seem very promising. Additionally, the experiments described in the previous sections have been partially automated by developing various Python classes and functions [3], making the process quite straightforward once the new dataset is loaded. Nevertheless, considering the high computational requirements of non-parametric conditional independence tests, the significant increase in dataset size ($1036800 \times 506 \times 6$, compared to $13072 \times 384 \times 4$), and the possibility of exploring larger maximum time lags, substantially more computational resources and efficiency optimization strategies will be required.

In addition to the points discussed, it is relevant to mention that a recent paper by [43] proposed a new algorithm for Causal Discovery in time-series data called Causal Discovery from Nonstationary Time Series (CD-NOTS). The authors consider that the hypothesis of causal stationarity rarely holds in the case of financial systems, motivating the need for methods with weaker assumptions. There are no evidences that this assumption is true neither for climatic systems, and this new algorithm outperformed PC-MCI on almost all simulations.

[3] Available at `https://github.com/oriolus98/climate_classifier_causal_discovery`

Therefore, it could be interesting to compare the results of both methods, although when the authors where contacted recently, they still did not have permission to open source their model.

# 4 Conclusions

This master thesis has explored the complex and evolving field of causal discovery and inference within temporal series in climate systems, building upon foundational concepts from data science such as information theory and statistics, while venturing into new theoretical and practical territories. Through an extensive review and application of causal inference methods, this work has demonstrated both the potential and the challenges of applying causal models to climate datasets.

The first part of this work focused on understanding and applying the theory behind Causal Impact, particularly the Bayesian state-space model developed by Google. This method, which models counterfactuals to quantify the difference between observed and alternatively intervened outcomes, has shown great potential in assessing the impact of human activities on climatic variables. For example, we have applied it to demonstrate that the global increase in temperature cannot be only attributed to natural drivers and to quantify the reduction in $CO_2$ emissions during COVID-19 lock-downs. These experiments showcase the power of causal inference in addressing real-world climate issues, while also highlighting the ability of these models to quantify uncertainty—a critical aspect in climate studies.

Moving forward, we introduced a graphical framework for causal discovery, focusing on the PC-MCI algorithm implemented in the Python Tigramite package. By applying this method to a spatio-temporal climate dataset, we were able to identify causal relationships between important variables: temperature, soil moisture, radiation, and precipitation. The results, visualized through global distributions and clustered for deeper insights, generally aligned with established physical principles and existing literature. However, some anomalies were observed (such as the weak and negative effect of precipitation on soil moisture) underscoring the inherent limitations of causal inference methods. These findings reinforce the need for careful interpretation of results and lay the groundwork for future research.

As this work demonstrates, the field of causal inference is marked by several notable limitations, which often stop researchers from fully engaging with it. The absence of a ground truth in many applications means that results can often only be assessed qualitatively rather than quantitatively. Furthermore, the computational resources required for causal discovery are substantial, and the mathematical and abstract nature of the field can be intimidating. Despite these challenges, causal inference offers powerful tools for understanding complex systems. By explicitly modeling assumptions, causal methods can help identify when causal conclusions can actually be achieved, detect errors on our problem setting, and improve models, as discussed in the Section 3.2.5. It is worth noting that traditional machine learning and statistical models also rely on numerous assumptions, many of which are often overlooked. Causal models, in contrast, might provide more explainability and transparency, making them valuable for both academic research and practical applications.

While large language models (LLMs) have gathered significant attention in recent years, major technology companies such as Google [7], Netflix[44], Microsoft[45] or Booking[46] are also quietly investing in causality research. These examples show that, while this discipline was once largely confined to theoretical research within universities, companies are now actively deploying it and finding valuable applications across various domains. The growing interest in this field stems from the realization that traditional machine learning approaches can sometimes lead to erroneous conclusions when used for decision-making. For instance, in many machine learning applications, practitioners may identify a feature as important based on Shapley values or similar metrics, and then infer that manipulating this feature can change outcomes. However, such inferences are often flawed, as they ignore the

interventional nature of these probabilities [47]—a gap that causal modeling aims to fill.

In conclusion, causal inference remains an open research field with significant challenges, but it is also evolving rapidly and holds tremendous potential. Recent theoretical advances in academia [43, 48], alongside practical applications in various domains are driving this growth. In the actual context of climate change, and considering the inherent complexity of understanding Climate dynamics, this work also highlights the crucial role that causal inference can play in advancing our knowledge of climate systems. While there is still much work to be done, the increasing recognition of causality's importance suggests that this field will continue to expand, offering new insights and tools for understanding and intervening in complex systems. The methods and results presented in this master thesis contribute to this ongoing effort, highlighting both the promise and the challenges of causal discovery in climate science.

# 5 Acknowledgements

The journey to completing this master thesis has been long and challenging, and I am deeply grateful to everyone who supported me along the way. I want to sincerely thank my family and friends for their encouragement, even during times when I struggled to stay motivated. Although it wasn't always easy, their belief in me, even when uncertain, was invaluable. I also wish to express my gratitude to my advisor, Jordi Muñoz Mari, whose guidance and positive feedback reassured me that I was on the right path. This process has not only helped me grow but has also inspired me to consider a possible future in research, a path I had not fully realized before embarking on this work. Thank you all for being part of this journey.

# References

[1] Judea Pearl. An introduction to causal inference, 2010.

[2] Judea Pearl. Causal inference in statistics: An overview. *Statistics Surveys*, 3:96–146, 2009.

[3] Jakob Runge, Sebastian Bathiany, Erik Bollt, Gustau Camps-Valls, Dim Coumou, Ethan Deyle, Clark Glymour, Marlene Kretschmer, Miguel D. Mahecha, Jordi Muñoz-Marí, Egbert H. van Nes, Jonas Peters, Rick Quax, Markus Reichstein, Marten Scheffer, Bernhard Schölkopf, Peter Spirtes, George Sugihara, Jie Sun, Kun Zhang, and Jakob Zscheischler. Inferring causation from time series in earth system sciences. *Nature Communications*, 10, 12 2019.

[4] Hernán MA and Robins JM. Causal inference: What if. *Boca Raton: Chapman Hall/CRC*, 2020.

[5] Peter Spirtes and Kun Zhang. Causal discovery and inference: concepts and recent methodological advances. *Applied Informatics*, 3, 12 2016.

[6] Jakob Runge et al. Tigramite 5.2.1.25. `https://github.com/jakobrunge/tigramite`, 2023.

[7] Kay H. Brodersen, Fabian Gallusser, Jim Koehler, Nicolas Remy, and Steven L. Scott. Inferring causal impact using bayesian structural time-series models. *Annals of Applied Statistics*, 9:247–274, 2015.

[8] Gavin A. Schmidt, Max Kelley, Larissa Nazarenko, Reto Ruedy, Gary L. Russell, Igor Aleinov, Mike Bauer, Susanne E. Bauer, Maharaj K. Bhat, Rainer Bleck, Vittorio Canuto, Yong Hua Chen, Ye Cheng, Thomas L. Clune, Anthony Del Genio, Rosalinda De Fainchtein, Greg Faluvegi, James E. Hansen, Richard J. Healy, Nancy Y. Kiang, Dorothy Koch, Andy A. Lacis, Allegra N. Legrande, Jean Lerner, Ken K. Lo, Elaine E. Matthews, Surabi Menon, Ron L. Miller, Valdar Oinas, Amidu O. Oloso, Jan P. Perlwitz, Michael J. Puma, William M. Putman, David Rind, Anastasia Romanou, Makiko Sato, Drew T. Shindell, Shan Sun, Rahman A. Syed, Nick Tausnev, Kostas Tsigaridis, Nadine Unger, Apostolos Voulgarakis, Mao Sung Yao, and Jinlun Zhang. Configuration and assessment of the giss modele2 contributions to the cmip5 archive. *Journal of Advances in Modeling Earth Systems*, 6:141–184, 3 2014.

[9] Zhu Liu, Zhu Deng, Biqing Zhu, Philippe Ciais, Steven J. Davis, Jianguang Tan, Robbie M. Andrew, Olivier Boucher, Simon Ben Arous, Josep G. Canadell, Xinyu Dou, Pierre Friedlingstein, Pierre Gentine, Rui Guo, Chaopeng Hong, Robert B. Jackson, Daniel M. Kammen, Piyu Ke, Corinne Le Quéré, Crippa Monica, Greet Janssens-Maenhout, Glen P. Peters, Katsumasa Tanaka, Yilong Wang, Bo Zheng, Haiwang Zhong, Taochun Sun, and Hans Joachim Schellnhuber. Global patterns of daily co2 emissions reductions in the first year of covid-19. *Nature Geoscience*, 15:615–620, 8 2022.

[10] Jakob Runge, Peer Nowack, Marlene Kretschmer, Seth Flaxman, and Dino Sejdinovic. Detecting and quantifying causal associations in large nonlinear time series datasets, 2019.

[11] Peter Spirtes and Clark Glymour. An algorithm for fast recovery of sparse causal graphs. *Social Science Computer Review*, 9:62–72, 1991.

[12] Christina Papagiannopoulou, DIego G. Miralles, Stijn Decubber, Matthias Demuzere, N. E.C. Verhoest, Wouter A. Dorigo, and Willem Waegeman. A non-linear granger-causality framework to investigate climate-vegetation dynamics. *Geoscientific Model Development*, 10:1945–1960, 5 2017.

[13] Don van Ravenzwaaij, Pete Cassey, and Scott D. Brown. A simple introduction to markov chain monte–carlo sampling. *Psychonomic Bulletin and Review*, 25:143–154, 2 2018.

[14] George Casella and Edward I. George. Explaining the gibbs sampler. *American Statistician*, 46:167–174, 1992.

[15] Brodersen et al. Causalimpact 1.2.7. `http://google.github.io/CausalImpact/`, 2015.

[16] Jiansong Zhou and Ka-Kit Tung. Deducing multidecadal anthropogenic global warming trends using multiple regression analysis. *Journal of Atmospheric Sciences*, 70:3–8, 01 2013.

[17] Lukas Gudmundsson, Josefine Kirchner, Anne Gädeke, Jeannette Noetzli, and Boris K Biskaborn. Attributing observed permafrost warming in the northern hemisphere to anthropogenic climate change. *Environmental Research Letters*, 17(9):095014, sep 2022.

[18] Ipcc, 2023: Climate change 2023: Synthesis report. a report of the intergovernmental panel on climate change. contribution of working groups i, ii and iii to the sixth assessment report of the intergovernmental panel on climate change. *IPCC, Geneva, Switzerland, (in press)*, 2023.

[19] Chiara Binelli. Estimating causal effects when the treatment affects all subjects simultaneously: An application. *Big Data and Cognitive Computing*, 5(2), 2021.

[20] Real decreto 463/2020, de 14 de marzo, por el que se declara el estado de alarma para la gestión de la situación de crisis sanitaria ocasionada por el covid-19. `https://www.boe.es/eli/es/rd/2020/03/14/463`. Accessed: 2023-07-25.

[21] J. Runge. Causal network reconstruction from time series: From theoretical assumptions to practical estimation. *Chaos*, 28, 7 2018.

[22] Brady Neal. Introduction to causal inference. *Course Lecture Notes (draft)*, 2020.

[23] Giorgia Di Capua, Marlene Kretschmer, Reik V. Donner, Bart Van Den Hurk, Ramesh Vellore, Raghavan Krishnan, and Dim Coumou. Tropical and mid-latitude teleconnections interacting with the indian summer monsoon rainfall: A theory-guided causal effect network approach (supplement 1). *Earth System Dynamics*, 11:17–34, 1 2020.

[24] Jakob Runge. 'conditional independence testing based on a nearest-neighbor estimator of conditional mutual information', in proceedings of the 21st international conference on artificial intelligence and statistics., 2018.

[25] Alexander Kraskov, Harald Stögbauer, and Peter Grassberger. Estimating mutual information. *Phys. Rev. E*, 69:066138, Jun 2004.

[26] L. Kozachenko and N. Leonenko. Sample estimate of the entropy of a random vector. *Problemy Peredachi Informatsii*, 23:9–16, 1987.

[27] Papagiannopoulou, Christina and Miralles, Diego and Decubber, Stijn and Demuzere, Matthias and Verhoest, Niko and Dorigo, Wouter A and Waegeman, Willem. A non-linear Granger-causality framework to investigate climate-vegetation dynamics. *Geoscientific Model Development*, 10(5):1945–1960, 2017.

[28] Jakob Runge, Vladimir Petoukhov, Jonathan F. Donges, Jaroslav Hlinka, Nikola Jajcay, Martin Vejmelka, David Hartman, Norbert Marwan, Milan Paluš, and Jürgen Kurths. Identifying causal gateways and mediators in complex spatio-temporal systems. *Nature Communications*, 6, 2015.

[29] Jeffrey Jake Nichol, Michael Weylandt, Mark Smith, and Laura P. Swiler. Benchmarking the pcmci causal discovery algorithm for spatiotemporal systems. 6 2023.

[30] Donald J. Berndt and James Clifford. Using dynamic time warping to find patterns in time series. In *AAAI - KDD Workshop*, 1994.

[31] Romain Tavenard, Johann Faouzi, Gilles Vandewiele, Felix Divo, Guillaume Androz, Chester Holtz, Marie Payne, Roman Yurchak, Marc Rußwurm, Kushal Kolar, and Eli Woods. Tslearn, a machine learning toolkit for time series data. *Journal of Machine Learning Research*, 21(118):1–6, 2020.

[32] Pavel Senin. Dynamic time warping algorithm review. 2008.

[33] Wladimir K̈oppen, Esther Volken, and Stefan Brönnimann. The thermal zones of the earth according to the duration of hot, moderate and cold periods and to the impact of heat on the organic world. *Meteorologische Zeitschrift*, 20:351–360, 6 2011.

[34] Hylke E. Beck, Niklaus E. Zimmermann, Tim R. McVicar, Noemi Vergopolan, Alexis Berg, and Eric F. Wood. Present and future köppen-geiger climate classification maps at 1-km resolution. *Scientific Data*, 5, 2018.

[35] Gustau Camps-Valls, Manuel Campos-Taberner, Álvaro Moreno-Martínez, Sophia Walther, Grégory Duveiller, Alessandro Cescatti, Miguel D. Mahecha, Jordi Muñoz-Marí, Francisco Javier García-Haro, Luis Guanter, Martin Jung, John A. Gamon, Markus Reichstein, and Steven W. Running. A unified vegetation index for quantifying the terrestrial biosphere. *Science Advances*, 7(9):eabc7447, 2021.

[36] Judea Pearl. The do-calculus revisited, 2012.

[37] Emiliano Díaz, Jose E. Adsuara, Álvaro Moreno Martínez, María Piles, and Gustau Camps-Valls. Inferring causal relations from observational long-term carbon and water fluxes records. *Scientific Reports*, 12, 12 2022.

[38] Junkai Liu and Zhaoxia Pu. Does soil moisture have an influence on near-surface temperature? *Journal of Geophysical Research: Atmospheres*, 124, 06 2019.

[39] Alexis Berg, Ben Lintner, Kirsten Findell, Sergey Malyshev, Paul Loikith, and Pierre Gentine. Impact of soil moisture–atmosphere interactions on surface temperature distribution. *Journal of Climate*, 27:7976–7993, 11 2014.

[40] Paul A Dirmeyer and J Shukla. The effect on regional and global climate of expansion of the world's deserts. *Quarterly Journal of the Royal Meteorological Society*, 122(530):451–482, 1996.

[41] Wiebke Günther, Peter Miersch, Urmi Ninad, and Jakob Runge. Clustering of causal graphs to explore drivers of river discharge. *Environmental Data Science*, 2:e25, 2023.

[42] Yuquan Qu, Carsten Montzka, and Harry Vereecken. Causation discovery of weather and vegetation condition on global wildfire using the pcmci approach. In *2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS*, pages 8644–8647, 2021.

[43] Agathe Sadeghi, Achintya Gopal, and Mohammad Fesanghary. Causal discovery in financial markets: A framework for nonstationary time-series data. *SSRN Electronic Journal*, 2024.

[44] Yixin Wang, Dawen Liang, Laurent Charlin, and David M. Blei. Causal inference for recommender systems. In *Proceedings of the 14th ACM Conference on Recommender Systems*, RecSys '20, page 426–431, New York, NY, USA, 2020. Association for Computing Machinery.

[45] Eleanor et al Dillon. Causal inference and machine learning in practice with econml and causalml: Industrial use cases at microsoft, tripadvisor, uber. `https://causal-machine-learning.github.io/kdd2021-tutorial/`, 2021.

[46] Philipp Bach, Victor Chernozhukov, Carlos Cinelli, Lin Jia, Sven Klaassen, Nils Skotara, and Martin Spindler. Sensitivity analysis for causal ml: A use case at booking.com. In *Proceedings of ACM KDD 2024*, New York, NY, USA, 2024. Association for Computing Machinery.

[47] Eleanor et al Dillon. Be careful when interpreting predictive models in search of causal insights. https://shap.readthedocs.io/en/latest/example_notebooks/overviews/Be%20careful%20when%20interpreting%20predictive%20models%20in%20search%20of%20causal%20insights.html#, 2018.

[48] Siyuan Guo, Chi Zhang, Karthika Mohan, Ferenc Huszár, and Bernhard Schölkopf. Do finetti: On causal effects for exchangeable data. https://arxiv.org/abs/2405.18836, 05 2024.