ENSO dynamics under strong Anthropogenic warming: A causal network reconstruction study using CMIP6 model simulations

Benjamin D. Weggelaar (VU Student Number: 2682475)

supervisor: prof. dr. Dim Coumou
co-supervisor: prof. dr. Michael Massmann

June 30, 2023

Abstract

The El Niño-Southern Oscillation (ENSO) is a prominent climate phenomenon that significantly influences global weather patterns and has wide-ranging impacts on socio-economic systems. In recent decades, the Earth's climate has undergone unprecedented warming due to anthropogenic activities, raising concerns about the potential impact on ENSO dynamics. This study aims to investigate the causal network structure of ENSO under strong anthropogenic warming using state-of-the-art climate model simulations from the Coupled Model Intercomparison Project Phase 6 (CMIP6). This study aims to unravel the complex interactions and causal relationships among various components of the ENSO system, by first estimating a simple three-variable network representing Walker circulation dynamics, and then expanding with more variables. The study shows mixed results, with many components of the ENSO system being well-represented while others are not, and high variability between ensemble runs and CMIP6 models. The network structure between past and future sees limited changes, and there is some evidence for a reduced causal strength in future networks. The outcomes of this study highlights the difficulties of applying network reconstruction methodologies to climate data.

Key words: ENSO, Climate Change, CMIP6 models, PCMCI.

Contents

1 Introduction

Climate change is one of the most pressing challenges facing our planet today, and its impacts are already being felt around the world. The burning of fossil fuels, deforestation, and other human activities have caused global temperatures to rise, resulting in more frequent and severe heat waves, droughts, and extreme weather events (IPCC, 2022). These changes cause shifts in precipitation patterns, leading to more frequent floods and wildfires (Beniston et al., 2007) , and are threatening the existence of countless plant and animal species (Foley et al., 2005). These changes also have significant and far-reaching consequences for natural systems, including ocean currents and atmospheric patterns (Trenberth and Hurrell, 1994).

One of the most significant impacts of climate change on the Earth's climate system is the alteration of the El Niño-Southern Oscillation (ENSO). The ENSO phenomenon is the strongest natural climate mode, and is responsible for a significant inter-annual variability in the climate system (Latif and Keenlyside, 2009). Trade winds across the equator push water towards the west pacific, causing upwelling [1] in the east in order to replace the water moving westwards. This creates a large gradient in sea surface temperature (SST), with warmer temperatures in the west pacific compared to the east. The warmer SST create low pressure zones with high precipitation that can cause extreme climate events such as flooding. The colder SST create high pressure and very low precipitation, causing droughts and wildfires. ENSO not only affects climate around the equatorial pacific, but is known to be a major climate teleconnection, affecting climate in East Asia (Wang et al., 2000), North America (Ropelewski and Halpert, 1986) and Eastern Africa (Marchant et al., 2007), among others.

ENSO alternates between three phases: Neutral, El Niño and La Niña [2]. The neutral phase refers to average conditions in the strength of the trade winds, and therefore average SST gradient. La Niña is a strengthening of the neutral phase, creating colder SST in the east. El Niño is characterized by weaker or even reversing trade winds, causing warm ocean water in the central and east-central equatorial pacific (NOAA, 2023).

The response of ENSO to greenhouse warming is still challenging and not well understood. Firstly, there is an increasing consensus that warming leads to more extreme events of both El Niño and La Niña (Cai et al., 2015; Latif and Keenlyside, 2009). It is also known that, due to the increasing variability, ENSO wil become harder to predict Zheng et al. (2016). Furthermore, Climate simulation models predict an increase in the frequency of El Niño events in response to global warming

---

[1]Upwelling refers to the process in which cold, nutrient-rich water from the deep sea rises to the surface.

[2]La Niña is often referred to as the cooling phase, while El Niño is referred to as the warming phase.

(Timmermann et al., 1999; Cai et al., 2014), even though there has been an increase in La Niña events in recent decades (Clem et al., 2018).

The aim of this report is to study the ENSO dynamics under strong global warming, using a novel causal network approach in climate sciences. This approach was first reviewed by Ebert-Uphoff and Deng (2012), who set the stage for further research, and has been applied to a wide range of topics including the identification of drivers of midlatitude winter circulation (Kretschmer et al., 2016), establishing the causes of surface ozone variability in Antarctica (Kumar et al., 2021), and assessing causal pathways of different climate teleconnections (Samarasinghe et al., 2021). This paper will aim to answer the following research questions:

1. How well do causal discovery algorithms perform in representing ENSO dynamics?

2. How do ENSO dynamics change under strong global warming?

This study will start by providing a literature review on research using causal discovery algorithms in Climate and Earth science. Section 3 provides an in-depth description of the methodology. Section 4 will follow with an explanation of ENSO dynamics, which are necessary to correctly interpret the results. Section 5 discloses the data sources and variables, while Section 6 presents the results. This study finalizes in section 7, by providing a discussion and conclusion.

2 Literature Review

This section provides a literature review on causal network reconstruction studies in the Climate/Earth sciences. Ebert-Uphoff and Deng (2012) set the stage by reviewing causal discovery methods in climate sciences, and from there, multiple methods have been designed to tackle the different challenges in Climate/Earth data, such as the PCMCI methods by Runge (2018); Runge et al. (2019b,a, 2023). Several studies will be presented, notably those using the PCMCI algorithm, in order to place the current study in context.

Kretschmer et al. (2016) was one of the first studies to use causal network algorithms in climate science. The authors use causal effect network (CEN) methodology, which is equivalent to the PC algorithm for time series data, and multiple regression to estimate the causal strength. The aim of the study is to unravel causal relationships of arctic mechanisms on midlatitude winter circulation. Barents and Kara sea ice are found to be important external drivers for winter circulation.

Capua et al. (2020) study interactions between the tropics and mid-latitudes in boreal summer, and the role ENSO plays in these interactions. In general a two-way positive feedback between the tropics and mid-latitudes is found.

Saggioro and Shepherd (2019) quantify the stratosphere-troposphere interaction by evaluating the relation between the polar vortex and the mid-latitude jet stream. The polar vortex is a large region of cold, rotating air in the Earth's polar regions. Its breakdown during spring impacts the hydrological cycle and ecosystem functioning, and is therefore crucial to better understand its relation to the mid-latitude jet stream. Since the polar vortex time series is highly autocorrelated, the PCMCI algorithm is a favorable choice, for its ability to handle this type of confounding. Polar vortex variability is able to explain 40% of jet stream variability.

Lehmann et al. (2020) estimate early forecasts of Moroccan wheat yield. Since most of the wheat grown in Morocco is rain-fed, its variability highly depends on meteorological conditions. The forecast model is is divided into three steps: The first two steps correspond to the RG-CPD method from Kretschmer et al. (2017). This methods consists of identifying potential precursors through spatial clustering of correlation maps, followed by the time series-adapted PC algorithm to identify which of the potential precursors can be interpreted as causal precursors. The last step constructs the forecasting model as a multiple linear regressions, where dependent variables are identified trough the last step 2's PC algorithm. The results show that forecasts can be made as early as December.

Samarasinghe et al. (2021) examine the changes in the relation between the Madden-Jullian osciallation (MJO) and the North Atlantic oscillation (NAO) under global warming. MJO is a traveling intraseasonal pattern over the Indian and Pacific ocean characterised by enhanced and suppressed rainfall, while NAO represents changes in

sea-level pressure (SLP) over the North-Atlantic. The oscillations are linked trough processes in the stratosphere and troposphere, which are likely to change under anthropogenic warming. The teleconnection is found to be strengthened through the stratospheric pathway.

Verma et al. (2022) study climate drivers of the variation of vegetation productivity in India. More specifically, they assess the effect of precipitation and total water storage (TWS) on gross primary production (GPP). The effect was found for rainfall, but not for TWS, the reason being that most of India's vegetation is cropland, with shallow-rooted plants. The analysis is then performed on cropland and forests separately.

Malloy and Kirtman (2023) use the PCMCI algorithm to improve subseasonal rainfall forecast over the US Great Plains. Remote climate teleconnections such as the East Asian monsoon (EAM) can improve rainfall predictability in the region. The authors finds that that EAM above average rainfall, and the boreal summer intraseasonal oscillation (BSISO) are related to Great Plains convection and rainfall, which provide a valuable forecast opportunity for subseasonal prediction.

The eight studies considered in this review were chosen based on their methodology, since all use the PC algorithm, PCMCI algorithm or a modified/extended version of it. These studies serve as examples and guidelines for the application of causal discovery algorithms for Earth sciences. These studies show how to uncover causal relations among meteorological variables and weather patterns, even though they are geographically very apart. This task becomes more challenging taking into account the numerous spurious relations that are inherent in climate data, such as autocorrelation, common drivers and indirect paths.

Table 1: Summary of studies utilizing causal discovery algorithms in Climate/Earth sciences

Reference	Objective	Method(s)	Main results
Kretschmer et al. (2016)	Identify drivers of Midlatitude winter circulation (Arctic oscillation (AO) index and the upper tropospheric jet stream pattern)	CEN	Barents and Kara sea ice concentration are important drivers of midlatitude winter ciruclation, influencing winter AO. Moreover, no causal link has been found between a weak polar vortex and negative AO.
Capua et al. (2020)	Assess how tropical convection interacts with mid-latitude summer circulation and how ENSO affects these interactionsb	MCA & PCMCI	The tropics and mid-latitudes interact causally in both directions, but the strength and occasionally the direction of the causative relationship depends on the timescale. These teleconnection patterns are slighlty affected by ENSO.
Saggioro and Shepherd (2019)	Quantify the effect of the polar breakdown on the jet stream variability during the spring-to-summer transition	PCMCI	Polar vortex variability is able to explain about 40% of monthly mean jet variability, taking into account the confounding caused by the strong autocorrelation of the stratospheric polar vortex time series.
Lehmann et al. (2020)	Estimating early forecasts of Moroccan wheat yield	RG-CPD & Linear regression	SST and geopotential height at 500 hPa (commonly used to low and high pressure systems in the mid-troposphere) are used to forecast wheat yield. They explain 50% of the variation for forecasts made in December, much earlier than the operational forecasts relased in March-April.
Samarasinghe et al. (2021)	Assessing changes in the causal pathways between the Madden-Jullian oscillation (MJO) and North Atlantic oscillation (NAO) under anthropogenic warming	PC-stable	The teleconnection between MJO and NAO are found to be strengthened under SSP585. This strengthening develops through the stratospheric pathways, with a delay of 20 days.
Muñoz et al. (2020)	Determine the combined effect of local climate and ENSO on Dengue cases in Colombia	Wavelet analysis & (nonlinear) PCMCI	El Niño is related to an increase in cases, while la Niña is associated to a decrease. This link is mainly explained through ENSO-driven changes in temperature and rainfall.
Verma et al. (2022)	Establish the effect of precipitation and total water storage (TWS) on atmospheric carbon uptake by vegetation in India	PCMCI	No causal connections were found from TWS to gross primary production (GPP) or net photosynthesis. This relation was found with precipitation instead of TWS. Since shallow-rooted cropland dominate India's green cover, the effect on vegetation is immediate, and thus not via TWS.
Malloy and Kirtman (2023)	The effect of subseasonal monsoon variability on precipitation over the US Great Plains	PCMCI	Boreal summer intraseasonal oscillation (BSISO) is related to rainfall over the US Great Plains with a lag of 14 days. More specifically East Asian Monsoon causally affects Great Plains' low-level jet anomalies, influencing rainfall over the region.

3 Methodology

3.1 Causal inference

3.1.1 Brief history

The concept of causality has had a long and complex history, dating back to the 18th century. John Stuart Mill was one of the first economists to formalize a definition for causality, in terms of a counterfactual (Imbens and Rubin, 2015). In order to know whether a certain treatment caused an specified outcome, we would need to observe an individual receiving and not receiving that treatment at the same time. The reality is that we cannot observe both states at the same time, and thus, this definition of causality was not practically useful (Imbens and Rubin, 2015).

Neyman (1923) rediscovered this notion of causality, and gave it formal notation (Splawa-Neyman et al., 1990) [3]. He also defined the concepts of potential and realized outcomes, akin to the two states of the world defined by John Stuart Mill. During the 1970s, Rubin (1974) formalized the work from Neyman, introduced some more notation, and the first causal inference framework, which today we know as potential outcomes framework or Rubin causal model.

An important development in the area of causal inference, was the work of Pearl (1993, 1994), who introduced the concept of Directed Acyclic Graphs (DAGs) as a tool for representing causal relationships between variables. DAGs can be seen as a graphical representation of a chain of causal effects (Cunningham, 2021).DAGs are often used as an identification strategy, since it gives a clear picture on which variables to include based on previous research and hypotheses. They provide a direct way of thinking about the causal system and strategies to estimate the causal effect (Morgan and Winship, 2014).

3.1.2 Concepts, terminology & notation

The notion of DAGs has its basis in graph theory, which is a sub-field in mathematics. A graph is often defined as an ordered pair $\mathcal{G} = (V, E)$, where the set V is the collection of vertices (also called nodes or points) which represent the variables in the causal system. The set E represents the set of edges (also called links or lines) that represent the association between two edges (Eichler, 2007).

A DAG is essentially a special case of a graph. A directed graph is one in which all arrows point to the same direction (e.g. from left to right). An acyclic graph is one where no cycles are present (Morgan and Winship, 2014). The immediate drawback of this is the absence of contemporaneous feedback loops in the causal graph, which do exist in the climate system. Lagged feedback loops can still be represented in

---

[3]This study was done in 1923, but was later translated and published in 1990.
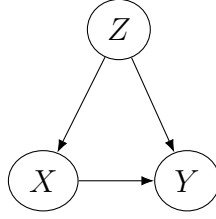
Figure 1: A simple DAG

DAGs for time series. Another drawback is because of the directed nature of the graph, we do not account for simultaneity between variables.

The terminology used in DAGs to describe relationships among variables is equivalent to those used in family trees. For instance, ancestors and descendants refer to the nodes that are connected through directed edges. An ancestor of a given node is any node that has a directed edge pointing to it. Conversely, a descendant of a given node is any node that can be reached by following the directed edges in the graph. If there is a direct arrow from one node to another then this node is a parent of the second node. We can see this in figure 1, where X is a parent of Y. Conversely, we can also say that Y is a child of X since the arrow is pointing towards it. More generally, two nodes are adjacent if an edge connects the two (Wasserman, 2004).

Imagine we are interested in calculating the causal effect of X on Y in the DAG shown in figure 1. There is a total of two paths that goes from X to Y. The first one is the direct path $X \rightarrow Y$. The second is the path $X \leftarrow Z \rightarrow Y$. This path is called a backdoor path, and it introduces spurious correlation between X and Y that are driven solely due to fluctuations in Z. The variable Z would be called a confounder, since it affects the relationship between X and Y (Morgan and Winship, 2014). The idea of the backdoor path is one of the most important things we can learn from the DAG. It is similar to the concept of omitted variable bias (Cunningham, 2021). In order to take the effect of Z into account, we need to condition on this variable when calculating the effect of X on Y. This conditioning is also called "closing the backdoor", and when all backdoors are closed, we have a research design that satisfies the backdoor criterion (Pearl, 1993, 2003).
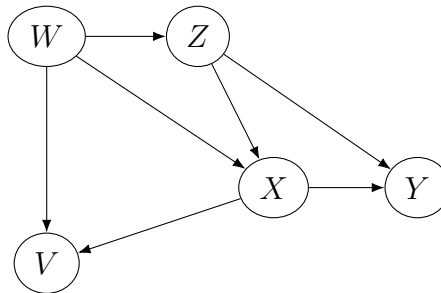

Figure 2: A more complicated DAG

Consider the DAG in figure 2. We want to determine the causal effect of X on Y.

Nevertheless, there are a total of four paths:

- The direct path $X \to Y$

- Backdoor path 1: $X \leftarrow Z \to Y$

- Backdoor path 2: $X \leftarrow W \to Z \to Y$

- Backdoor path 3: $X \to V \leftarrow W \to Z \to Y$

In this case, W and Z would be considered confounders, and we would need to condition on these variables in order fulfill the backdoor criterion, so that we can interpret the effect of X on Y as causal. Nevertheless, notice that the third backdoor path has one variable where all arrows are pointing towards ($X \to V \leftarrow W$). This variable is called a collider, which happens when two variables cause a third variable along a path (Wasserman, 2004). The term "collider" refers to the fact that in graphical models, the arrow heads from variables that lead into the collider appear to "collide" on the node. Such variables are important to consider, because they close backdoor paths automatically. Therefore, they cause the opposite to happen compared to confounders. Conditioning on a collider would introduce collider bias (Cunningham, 2021).

3.2   Causal discovery algorithms

Due to increasing computer power, a range of causal discovery algorithms have been presented in the last decades. Causal discovery aims to infer causal structure from the data. These algorithms typically work by examining correlations between variables and attempting to rule out associations between them.

There two main classes of causal discovery algorithms are constrained-based and score-based approaches (Glymour et al., 2019). To find causal connections, constraint-based approaches rely on the idea of conditional independence. The assumption is that there is no direct causal relationship between two variables if they are conditionally independent given a collection of additional factors. Thus, constraint-based approaches are able to infer causal links by discovering sets of factors that make two variables conditionally independent (Glymour et al., 2019).

Score-based approaches, on the other hand, make use of statistical models to gauge the strength of the connections between variables that cause them to occur. These techniques often include fitting a model to the data before comparing several causal models using criteria like the Bayesian Information Criterion (BIC) or the Akaike Information Criterion (AIC) (Huang et al., 2018).

3.2.1   Causal discovery assumptions

Three assumptions are often brought up in the context of causal discovery (also called axioms in Spirtes et al. (2001). These are necessary to interpret the time

series graph as a true causal graph. The three assumptions are (Runge, 2018):

- **Causal sufficiency**: Any pair of nodes in graph $\mathcal{G}$ has no common external cause. In other words, there is no unobserved variable causing at least two variable in the causal system. If this assumption is not fulfilled, it would result in an incomplete causal graph.

- **Causal Markov condition**: All nodes V, conditioned on their parents, are independent on their non-descendants. Since we have the general assumption that the cause precedes the effect in time (no contemporaneous effects), the intuition of the Markov condition follows that if we know the values of the X's parents, all variables further in the past become irrelevant for predicting X. Note that if the causal sufficiency assumption is not fulfilled, then generally the causal Markov condition will not hold (Spirtes et al., 2001).

- **Faithfulness**: The conditional independence relationships implied by the causal graph are faithfully represented in the data. In other words, if two variables are conditionally independent given a set of other variables in the causal graph, then they will be independent in the observed data, assuming that the data is a representative sample.

3.2.2 Peter-Clark (PC) algorithm

Let $V = \{X^1, X^2, ..., X^N\}$ be the set of vertices (or variables) in graph $\mathcal{G}$. i j and k are indices in $1, 2, ..., N$. Furthermore, we define $Adj(X_i)$ as the set of adjacent nodes of X^i [4].

The PC algorithm has the following steps:

- **Step 1**: Set up the fully connected undirected graph.

- **Step 2**: Delete edges when nodes are unconditionally independent.

- **Step 3**: Set the number of conditions $K = 1$ and select a pair of adjacent vertices X^i and X^j such that $Adj(X^i)\backslash\{X^j\}$ has $\geqslant K$ elements. A conditional independence test is performed between X^i and X^j given the conditional set S. The conditional set S is a subset in $Adj(X^i)\backslash\{X^j\}$, where S has K number of elements. If X^i and X^j are conditionally independent then we record the subset of S in $Sepset(X^i, X^j)$ and $Sepset(X^j, X^i)$ [5]. This is done for all pairs of nodes that have at least K elements in $Adj(X^i)\backslash\{X^j\}$. Next, we increase K by one and we repeat the process. Since K represents the number of elements in S, we first condition on one variable, then two variables, until we reach the

---

[4]Spirtes et al. (2001) uses slightly different notation. They use $Adjacencies(C, A)$ to refer to the set of vertices adjacent to A in the directed acyclic graph C. We avoid this notation by not specifying the DAG, since there is only one in this setting.
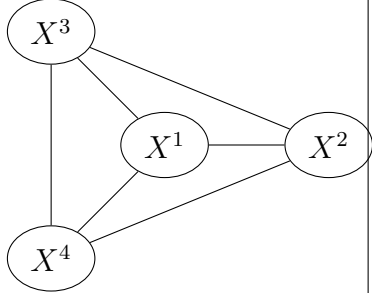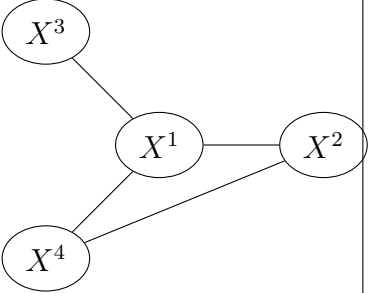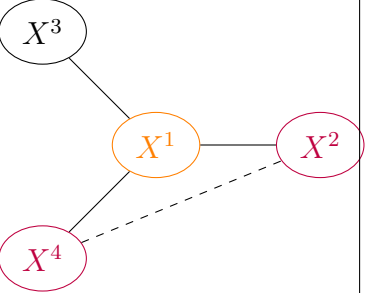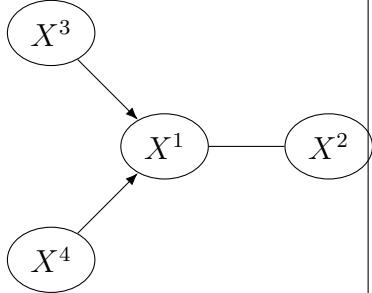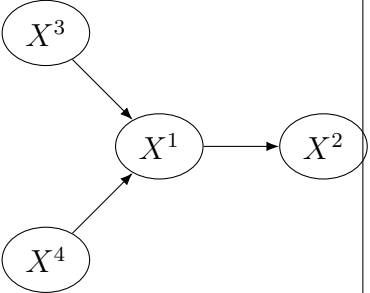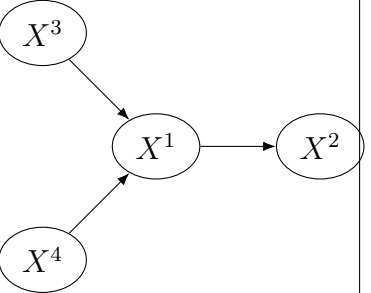
[5]Spirtes et al. (2001) denotes the separation set as $Sepset(\cdot)$, which is used in the next step to determine colliders.

maximum number of elements in $Adj(X^i)\backslash\{X^j\}$ for all adjacent pairs X^i and X^j.

- **Step 4**: For each triple X^i, X^j and X^k, such that the pair X^i and X^j and the pair X^j and X^k are each adjacent, but the pair X^i and X^k are not, orient $X^i - X^j - X^k$ as $X^i \rightarrow X^j \leftarrow X^k$ only if X^j is not in $sepset(X^i, X^k)$. This is because if X^j is not in $sepset(X^i, X^k)$ but is part of the conditioning set, then it must be a collider by definition.

- **Step 5**: Change the undirected edges to directed edges according to the constraints imposed by DAGs: No cycles and all edges must be directed.

- **Step 6**: Final output: estimated causal graph.

Table 2 shows the steps in the PC algorithm in a simplified way. The second diagram shows that X^4 and X^2 are dependent. Nevertheless, in step 3, I illustrate a conditional independence test between X^4 and X^2 (in red) given X^1 (in orange). Since the edge disappeared, X^4 and X^2 are independent given X^1 and thus X^1 becomes part of $Sepset(X^4, X^2) = \{X^1\}$.

Table 2: Steps in PC algorithm. Based on Spirtes et al. (2001) and Glymour et al. (2019)

3.2.3 PC-stable algorithm

Although the PC algorithm is computationally effective and asymptotically reliable, it takes unneeded risks with sample data. This is because If an edge is mistakenly removed from the true graph at an early stage of step 3, then other edges which are not in the true graph may be included in the output. For instance, if the $X^4 - X^1$ edge is mistakenly removed, then the $X^4 - X^2$ edge will not be removed in a future stage, since X^4 and X^1 are not adjacent anymore. The causal relationships between X^i and X^j may be completely unaffected by X^i and X^j's independence or dependence on many of these subsets of variables. The PC-stable algorithm only uses a subset $\mathcal{S}$ as the conditioning set, and is particularly useful for small to medium number of variables (Spirtes et al., 2001).

3.2.4 PCMCI algorithm

The PC algorithm was not specifically designed for time series data (Runge, 2018). Therefore, the Peter-Clark momentary conditional independence (PCMCI) algorithm is an extension of the PC-stable algorithm, which expands it in two ways. Firstly, it is designed to tackle time-series data, and thus, lag effects of different variables can be included in the algorithm. Secondly, the PCMCI algortihm consists of two phases: The first phase is the original PC-stable algorithm from section 3.2.3. The second stage consists of conditional independence tests, that evaluate the strength of the edges between variables. The tests are performed on all variables conditioning on their parents [6], which aims to account for autocorrelation.

Furthermore, the PCMCI algorithm needs two additional assumptions, on top of those explained in section 3.2.1. These assumptions are:

- **Causal stationarity**: The time series $V = (X^1, X^2, ..., X^N)$ must be all stationary series.

- **Temporal precedence**: We assume that the cause has occurred before the effect. This also implies no contemporaneous causal links.

Definition 1 (Time series graph). Using notation from Runge et al. (2019b), we now define V to be the set of vertices, containing all variables in $\mathcal{G}$, and let $\mathcal{G} = (V, E)$, be the corresponding time series graph. $V = (X^1, X^2, ..., X^N)$ has N variables. τ represents the number of lags, while τ_{max} is the maximum lag in graph $\mathcal{G}$.

Part I: PC

The first part of the PCMCI algorithm is equivalent to the PC-stable algorithm but adapted to time series. Figure 3 exemplifies the search for X^1's parents, showed in teal). The first step is an unconditional independence test, where six test statistics

---

[6]Since we assume temporal precedence, conditioning on their parents is equivalent to conditioning on adjacent nodes, as done in the PC algorithm.

were found to be significant (illustrated in violet boxes), and are recorded in P_0. The second step involves conditional independence tests between X_t^1 and the ith element in P_0, conditional on the jth element in P_0, where $i \neq j$. All variables whose test statistics were significant are recorded in P_1, and are illustrated in the figure with orange hexagons. The next step is conditioning on two variables in P_1, illustrated as green circles. The algorithm continues until it finds all parents of variable X^1, so when $P_n = P_{n+1}$, or when the elements in $P_n \geqslant K$ conditioning elements. This process is repeated for all remaining variables.
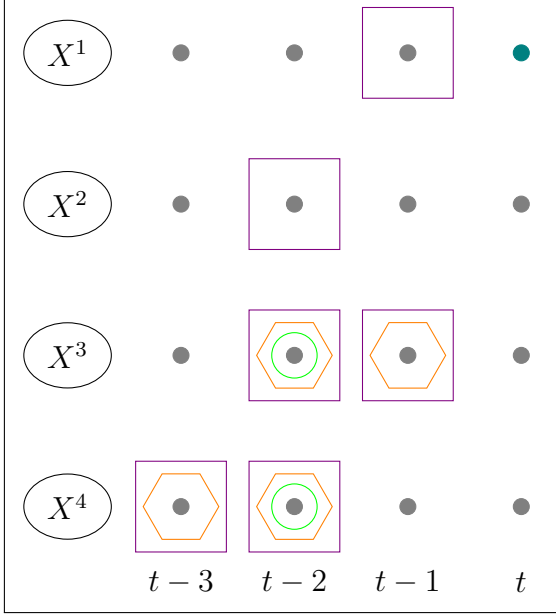
- Number of conditions $K = 0$ □
  $(X_t^i \perp\!\!\!\perp X_{t-\tau}^i) \; \forall \; i \; and \; \tau$
  $P_0 = \{X_{t-1}^1, X_{t-2}^2, X_{t-1}^3, X_{t-2}^3, X_{t-2}^4, X_{t-3}^4\}$
- Number of conditions $K = 1$ ⬡
  $(X_t^1 \perp\!\!\!\perp P_0^i) \mid P_0^j \; \forall \; i \neq j$
  $P_1 = \{X_{t-1}^3, X_{t-2}^3, X_{t-2}^4, X_{t-3}^4\}$
- Number of conditions $K = 2$ ○
  $(X_t^1 \perp\!\!\!\perp P_1^i, \; P_1^j) \mid P_1^k \; \forall \; i \neq j \neq k$
  $P_2 = \{X_{t-2}^3, X_{t-2}^4\}$
- Algorithm converges: $\mathcal{P}_{X^1} = P_2$

Figure 3: Example of part I of the PCMCI algorithm, with $\tau_{max} = 3$ and $N = 4$ variables.

Part II: MCI

The second part of the PCMCI algorithm consists of determining the strength of association between the variables. The MCI part ascribes a p-value and strength to each possible link (Krich et al., 2020). The causal strength is the outcome of the partial correlation tests, and so the number ascribed is between -1 and 1. The MCI step iterates through all pairs $\left(X_{t-\tau}^i, X_t^j\right)$, and performs a conditional independence test of the form $X_{t-\tau}^i \perp\!\!\!\perp X_t^j \mid \mathcal{S}$ where $\mathcal{S}$ is the set of parents $\mathcal{P}\left(X_t^j\right)$ and $\mathcal{P}\left(X_{t-\tau}^i\right)$ obtained from the PC algorithm. The additional conditioning on the parents $\mathcal{P}\left(X_{t-\tau}^i\right)$ in the MCI step accounts for autocorrelation leading to well-controlled false positive rates at the expected level (Runge, 2018).

The main advantage of the PCMCI methodology, is that it is able to take into account non-causal, spurious relations among variables, such as those depicted in figure 4, so that the relationship that is left must have a causal interpretation (when assumptions hold). Standard correlation analysis used in climate science do not remove the effects of spurious correlations.
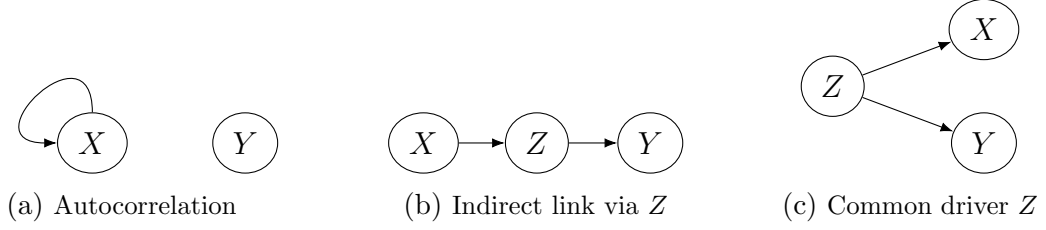
(a) Autocorrelation   (b) Indirect link via Z (c) Common driver Z

Figure 4: Spurious relations among variables.

3.2.5 Simulated example of PCMCI

To end the explanation of the PCMCI algorithm, a simulated example will be provided to fully understand each step in the process [7]. The following stationary VAR(1) is simulated:

$$
\begin{aligned}
X_t &= \theta X_{t-1} + \varepsilon_X \\
Y_t &= \theta Y_{t-1} + \varepsilon_Y
\end{aligned}
\tag{1}
$$

Where ε follows a multivariate normal distribution:

$$
\begin{pmatrix} \varepsilon_X \\ \varepsilon_Y \end{pmatrix} \sim NID \left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma^2 & \theta^2 \\ \theta^2 & \sigma^2 \end{pmatrix} \right]
\tag{2}
$$

Both time series X and Y are simulated for $n = 1000$ periods. Equation 1 and 2 represent the data generating process. The maximum lag in this simulated example is $\tau_{max} = 2$. Therefore, the DGP is a special case of the model, and should correctly estimate the DGP. The variance σ^2 is assumed to be equal to 1, and the AR coefficient $\theta = 0.8$. The covariance between both time series is $Cov(X_t, Y_t) = \theta^2 = 0.64$. Therefore, the two time series are correlated with each other, as well as autocorrelated. A significance level of 5% is used. Here we exemplify the search of X's parents, and purposefully omit the search of Y's parents, in order to simplify the example. As seen in figure 5, the PCMCI algorithm was able to correctly estimate the DGP. Furthermore, the final test statistic in the MCI step for the partial correlation between X_{t-1} and X_t given X_{t-2} is equal to 0.625. This will be the final value ascribed to the link between X_{t-1} and X_t and can be interpreted as a standardized measure of causal strength (Runge et al., 2019b).

---

[7]The code for this simulation can be found in the following GitHub page, in the script named pcmci_simulation_example.py.

- Step I: PC algorithm:
  - Testing condition set of dimension 0:
    * Test $X_{t-1} \perp\!\!\!\perp X_t$: Value: 0.799, p-value = 0.000. Significance detected.
    * Test $X_{t-2} \perp\!\!\!\perp X_t$: Value: 0.637, p-value = 0.000. Significance detected.
    * Test $Y_{t-1} \perp\!\!\!\perp X_t$: Value: 0.669, p-value = 0.000. Significance detected.
    * Test $Y_{t-2} \perp\!\!\!\perp X_t$: Value: 0.545, p-value = 0.000. Significance detected.
  - Sorting parents in decreasing order: $X_{t-1}, Y_{t-1}, X_{t-2}, Y_{t-2}$
  - Testing condition set of dimension 1:
    * Test $X_{t-1} \perp\!\!\!\perp X_t | Y_{t-1}$: Value: 0.588, p-value = 0.000. Significance detected.
    * Test $Y_{t-1} \perp\!\!\!\perp X_t | X_{t-1}$: Value: 0.037, p-value = 0.237. No significance.
    * Test $X_{t-2} \perp\!\!\!\perp X_t | X_{t-1}$: Value: 0.000, p-value = 0.999. No significance.
    * Test $Y_{t-2} \perp\!\!\!\perp X_t | X_{t-1}$: Value: 0.026, p-value = 0.417. No significance.
  - Algorithm converged for X_t. Parents of X_t: X_{t-1}.
  - (Process ommited) Algorithm converged for Y_t. Parents of Y_t: Y_{t-1}.
- Step 2: MCI algorithm:
  - Testing all pairs again, conditioning on X & Y's parents, and their lagged parents.
    * Test $X_{t-1} \perp\!\!\!\perp X_t | X_{t-2}$: Value: 0.625, p-value = 0.000. Significance detected.
    * Test $X_{t-2} \perp\!\!\!\perp X_t | X_{t-1}, X_{t-3}$: Value: 0.011, p-value = 0.740. No significance.
    * Test $Y_{t-1} \perp\!\!\!\perp X_t | X_{t-1}, Y_{t-2}$: Value: 0.027, p-value = 0.386. No significance.
    * Test $Y_{t-2} \perp\!\!\!\perp X_t | X_{t-1}, Y_{t-3}$: Value: 0.045, p-value = 0.152. No significance.
- Conclusion: Variable X_t has one significant link X_{t-1}: Value: 0.625, p-value = 0.000

Figure 5: Example output of the PCMCI algorithm: The search for X's parents

3.3 Graphical models vs. Vector auto-regressive models

In definition 1, we defined V to be the set of all variables in $\mathcal{G}$, E is the set of edges or links, and $\mathcal{G} = (V, E)$ is a time series graph. Why not simply construct a $VAR(p)$ model for all time series in V ? It turns out that there is a one-to-one relationship between the vector auto-regressive representation of V, and its graphical representation. In order for the both representations to be equal, the pairwise Granger-causal Markov property must hold.

Definition 2 (Pairwise Granger-causal Markov property)**.** The stochastic process $\mathbf{X}$ satisfies the pairwise Granger-causal Markov property with respect to graph $\mathcal{G}$ $i, j \in V$ with $i \neq j$:

(i) $i \rightarrow j \notin E \Rightarrow X^i \nrightarrow X^j$

(ii) $i - j \notin E \Rightarrow X^i \nsim X^j$

where the symbol $\nrightarrow$ represents Granger non-causal relation, and $\nsim$ represents contemporaneously conditionally independent relation (Eichler, 2012). Both cases

have the same idea, except that the first condition concerns lagged dependencies, and the second is with respect to contemporaneous dependencies.

Let's consider the following generic $VAR(p)$ representation:

$$X_t = \sum_{\tau=1}^{p} \Phi\left(\tau\right) X_{t-\tau} + \varepsilon_t, \quad \varepsilon_t \sim NID\left(0, \Sigma\right) \tag{3}$$

where $\Phi\left(\tau\right)$ is a $V \times V$ matrix of autoregressive coefficients. The multivariate process X satisfy the Pairwise Granger-causal Markov property with respect to graph $\mathcal{G}$ if the following conditions hold (Eichler, 2007, 2012):

(i) $i \to j \notin E \Rightarrow \Phi_{ij}\left(\tau\right) = 0$

(ii) $i - j \notin E \Rightarrow \Sigma_{ij}^{-1} = \Sigma_{ji}^{-1} = 0$

These conditions indicate that each coefficient in the matrices $\Phi\left(\tau\right)$ correspond to a directed link pointing forward in time and each nonzero value in the inverse covariance matrix Σ^{-1} coincide with an undirected contemporaneous link (Runge et al., 2014). So the value in the jth row and ith column in Σ^{-1} represent the contemporaneous relation between the jth and ith variable.

Graphs offer a convenient way to represent and visualize conditional independencies between random variables (Ebert-Uphoff and Deng, 2012). Furthermore, many causal discovery algorithms attempt to uncover the true causal structure of a set of variables, while a $VAR(p)$ model only uncovers Granger causality. In addition, the PCMCI method defines link strength by partial correlations, and thus are normalized values between -1 and 1. $VAR(p)$ offer regression coefficients, which are easy to interpret. Appendix A offers an in-depth mathematical derivation that relates the partial correlation coefficient with the regression coefficient from multiple regression.

3.4   Conditional independence tests

The choice of conditional independence test is crucial for unraveling the true causal diagram. The test chooses the edges to include or exclude in the graph, and should therefore be an accurate measure of conditional independence among variables. The importance of the conditional independence test is especially highlighted given the fact that incorrectly removing a node between two variables will affect the inclusion or removal of other nodes in the graph (Spirtes et al., 2001).

The simplest choice is a partial correlation test (Baba et al., 2004), which consists of examining whether the partial correlation of X and Y given Z [8] is significantly different from zero. If it is, we can interpret the result as X and Y being conditionally

---

[8]Note that the conditioning set Z can be empty, which reduces the test to an unconditional independence test, or could be multivariate.

dependent given Z. More formally, the testing procedure starts by regressing X on Z and Y on Z, which results in the following linear regression models (Rao and Sievers, 1995):

$$X = \beta_X Z + \varepsilon_X$$
$$Y = \beta_Y Z + \varepsilon_Y \tag{4}$$

Where the errors follow a normal distribution $\varepsilon \sim N\left(0, \sigma^2\right)$. We extract the residuals $\hat{\varepsilon}_X$ and $\hat{\varepsilon}_Y$, and calculate Pearson's correlation coefficient $\hat{\rho}_{XY|Z} = \hat{\rho}\left(\hat{\varepsilon}_X, \hat{\varepsilon}_Y\right)$. The correlation coefficient is bound between $[-1, 1]$, so we apply the following equation:

$$t = \hat{\rho}\left(\hat{\varepsilon}_X, \hat{\varepsilon}_Y\right) \sqrt{\frac{n - 2 - D_Z}{1 - \hat{\rho}\left(\hat{\varepsilon}_X, \hat{\varepsilon}_Y\right)^2}} \tag{5}$$

which follows a t-distribution with $n - 2 - D_Z$ degrees of freedom (Kendall and Stuart, 1983; Runge et al., 2019b), where n is the sample size and D_Z is the number of variables in the conditioning set Z. This can then be used to formally test the null hypothesis $H_0 : \rho_{XY|Z} = 0$ against the null hypothesis $H_A : \rho_{XY|Z} \neq 0$. Appendix B offers a small Monte Carlo simulation study to investigate the expected false positive rate in the partial correlation test.

The partial correlation test relies on the assumptions of normality and linear relations between variables. Nevertheless, Voortman and Druzdzel (2008) have shown that the partial correlation test is very robust against departures from normality, and thus the PC and PCMCI algorithm yield robust results. Furthermore, there is also a robust version of the partial correlation test (Rao and Sievers, 1995) that is more suitable for different marginal distributions. However, partial correlation has low power in detecting conditional dependence in the presence of non-linear relations (Cai et al., 2022). Therefore, the test is likely to be unsuitable in the presence of strong non-linearities.

3.5 False discovery rate

The false discovery rate (FDR) is a statistical concept that quantifies the proportion of significant findings that are false discoveries in a multiple hypothesis testing scenario. In other words, it quantifies the rate at which you expect to find false positives when conducting multiple statistical tests.

The p-values calculating during the PCMCI procedure, are corrected using the Benjamini and Hochberg (1995)'s method. This approach starts by sorting all p-values in ascending order $i = 1, 2, ..., m$, where m is the total number of tests performed. The next step is to calculate each test's BH critical value, which is equal to $i \times \left(\frac{\alpha}{m}\right) \times Q$, where i is each test's rank, α is the significance level, and Q is the

chosen false discovery rate. We only reject those hypotheses whose p-value is less than or equal to the BH critical value. Since the false discovery rate is positive, it usually leads to rejecting more tests than usual, since some of them were rejected by chance. In the context of graphical models, this represents that some links may not be statistically significant anymore.

3.6 Comparison metrics

In order to compare networks among each other, we define several so-called comparison metrics:

1. **Average strength difference:** This metric is calculated by taking only the significant links of both networks, and calculate the average strength of the links. The next step is to calculate the difference between the two networks. This metric is useful when comparing two networks from different time periods. The process is visualized in figure 6b. All significant links in network 1 are summed and divided by the total number of significant links. The same process is applied to network 2, and finally, the difference between the two numbers is taken:

$$\sum_{i=1}^{n} |l_{1i}| - \sum_{i=1}^{m} |l_{2i}| \tag{6}$$

   Where n is the total number of significant cross-links in network 1 (benchmark), and m is the total number of significant cross-links in network 2 (comparison). We do not take autocorrelation links into account. This links are shown in the results as a colored node, and represent a significant link between a variable and its lag. These links are not included in the calculation, because we are interested in assessing whether the causal strength between variables has weakened or strengthened over time.

2. **Coefficient deviation:** This metric takes all statistically significant coefficients (both cross-links and autocorrelation links) of the base network (benchmark) and subtracts the coefficient of the comparison network. The absolute value is taken of all differences, and finally, the average. This metric serves as a measure of how different the comparison network is from the benchmark network, in terms of the estimated coefficients. The equation is:

$$\sum_{i=1}^{n} |\Delta_i| \tag{7}$$

   The difference between the two metrics, is that the average strength difference calculates the difference of the sums, while the coefficient deviation calculated the sum of the differences. The process is visualized in figure 6a.
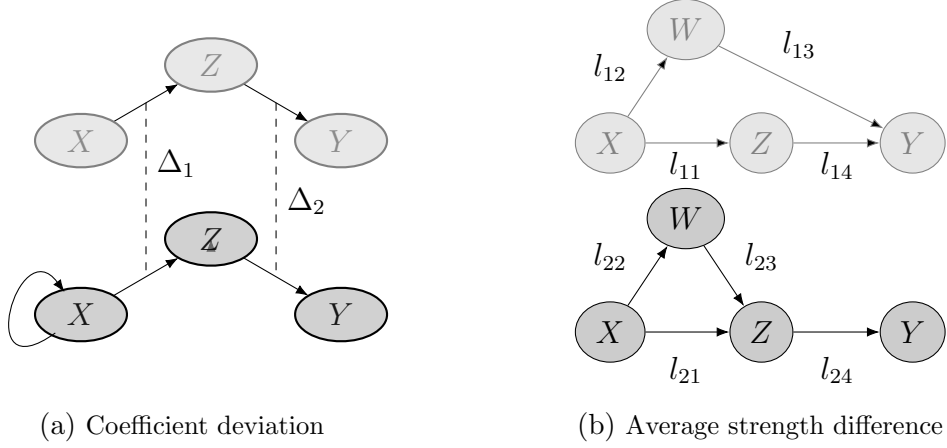
(a) Coefficient deviation

(b) Average strength difference

Figure 6: Comparison metrics visualized

3. **F_1 score:** This metric is based on the concepts of precision and recall. Firstly, we assume that the benchmark network is the actual network that best explains the data generating process, while the comparison network tries to mimic this process, and thus serves as a prediction, or predicted model. If the benchmark contains a link (cross-link or autocorrelation link) while the comparison network does not, it can be interpreted to be a false negative. On the other hand, if the benchmark network does not contain a link while the comparison network does, it can be considered a false positive. This is illustrated in the confusion matrix in figure 7. The precision metric is the percentage of true positives out of the total of predicted positives. The recall metric represents the percentage of predicted positives out of all actual positives. In this case, it represents the share of links the predicted model has been able to correctly identify. Equation 8 presents the definitions of both precision and recall, based on the confusion matrix in figure 7.

$$
\begin{aligned}
Precision &= \frac{TP}{TP + FP} \\
Recall &= \frac{TP}{TP + FN}
\end{aligned}
\tag{8}
$$

Finally, the F_1 score is the harmonic mean of precision and recall, and thus can be defined as:

$$
F_1 = \frac{2 * Precision * Recall}{Precision + Recall}
\tag{9}
$$

Nowack et al. (2020) use a modified asymmetric F_1 score to compare ensemble runs within a CMIP5 model (intra-model variability) and between runs of different CMIP5 models (inter-model variability).

Figure 7: Confusion matrix

4 ENSO Theory

4.1 ENSO phases

The El Niño-Southern Oscillation (ENSO) is an irregular cyclical variation in winds and sea surface temperatures over the tropical eastern Pacific Ocean that affects the climate globally. El Niño is the warming phase of the sea temperature, and La Niña is the cooling phase. The neutral phase of ENSO occurs when sea surface temperatures (SST) in the equatorial Pacific are most similar to the average. During this phase, the trade winds blow consistently from east to west across the tropical Pacific, and the oceanic conditions are relatively stable.
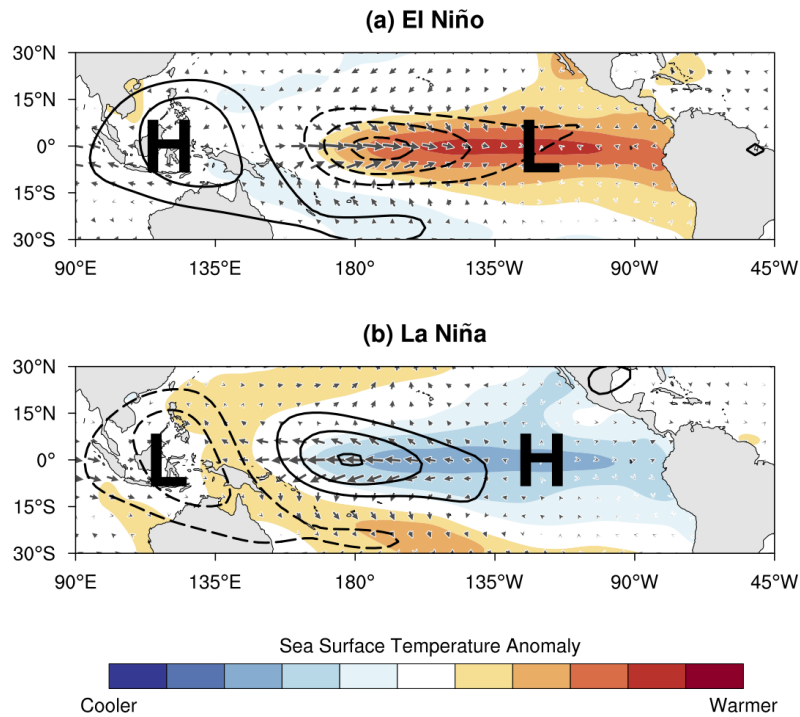
Figure 8: Empirical example of El Niño and La Niña events

The La Niña phase is a strengthening of the neutral phase. Stronger-than-usual trade winds push water from east to west, causing upwelling of deep, cold and nutrient-dense water in the east pacific, while having relatively warm water in the west pacific. This in turn causes high sea-level pressure (SLP) in the west pacific, and low SLP in the east. Some of the impacts of La Niña events include increased rainfall in parts of Indonesia and the Philippines as well as the western equatorial Pacific, drier than usual conditions in western South America. Winters in the north of the United States and Canada are cooler than usual.

El Niño occurs when sea surface temperatures in the eastern and central equatorial Pacific Ocean are warmer than average. It is related with a weakening or even reversal of the trade winds that blow over the tropical Pacific from east to west.

This causes warmer-than-usual SST in the east and central pacific. These warmer conditions create low SLP in the region, while the opposite happens in the west pacific. Some of the impacts of El Niño events include increased rainfall and storm activity over the central and eastern Pacific, including western South America, drought conditions in parts of Southeast Asia and Australia, Warmer temperatures in the northern United States and Canada during the winter.

Figure 8 exemplifies the two main phases of ENSO, with El Niño reversing trade winds, high pressure area in the west pacific, low pressure area and warm SST around central and east pacific. The opposite happens during La Niña: strong trade winds, pressure areas are swapped, and the emergence of the cold tongue.

4.2   Walker Circulation

Lastly, the Walker circulation is an atmospheric process greatly associated to ENSO. When trade winds arrive to the west pacific, the high SST creates convection, which increases rainfall and lowers SLP. The warm air travels eastwards, cooling down in the process, where it finally descends around the central and east pacific, creating high SLP and lower rainfall. This process is strengthened during La Niña, and weakened or even reversed during El Niño. Figure 9 exemplifies how ENSO affects climate in the tropics.
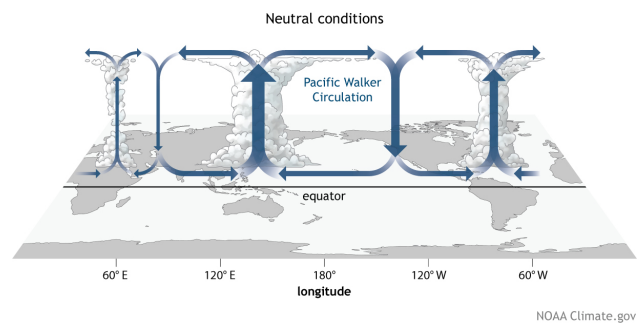

Figure 9: Conceptual model of the Walker circulation during neutral conditions, from NOAA (2023)

4.3   Bjerknes Feedback

One of the first hypotheses to explain the mechanisms behind El Niño and La Niña events was proposed by renowned meteorologist Jacob Bjerknes. At its core, the Bjerknes feedback involves a positive or negative loop of interactions between the ocean and atmosphere that can reinforce or suppress climate anomalies.

Figure 10 exemplifies the positive feedback loop. Bjerknes did not specify which element in the loop changes, which in turn triggers a chain reaction enhancing the feedback loop. Instead, somewhere in the system a change occurs, which propagates through the remaining elements.

The feedback loop in figure 10 starts with an exogenous shock in the trade winds, which causes them to strengthen. Stronger trade winds, pushes water from East to West, creating the west pacific warm pool. The pool of warm water increases
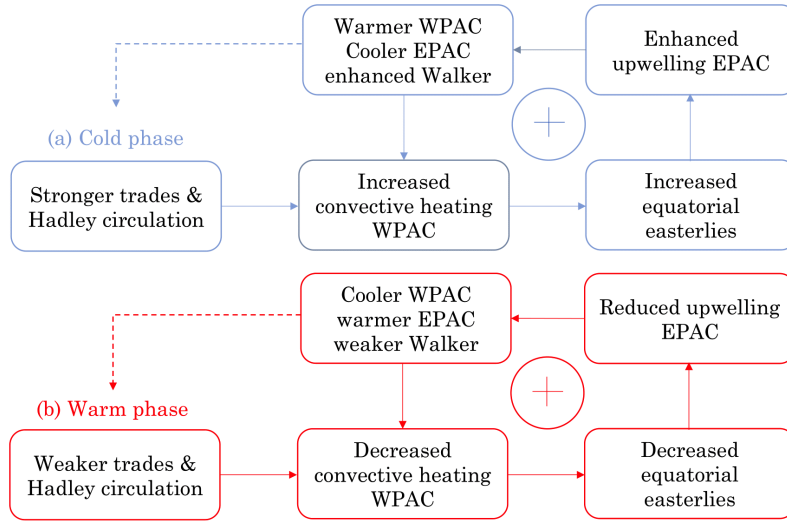
Figure 10: Bjerknes positive feedback loop hypothesis

convective heating and easterlies. This process in turn enhances upwelling in the East pacific and downwelling in the west pacific, which results in a warmer west, colder east, and enhanced Walker circulation. This chain of events is reinforced due to the closed feedback nature of the system, and thus produces La Niña events.

The process can happen the other way around, where weaker trade winds decreases convective heating in the west pacific, and in turn, enhances upwelling in the region, while simultaneously reducing upwelling in the east. This results in a cooler west pacific, a warmer east pacific and a weaker Walker circulation. These events keep enforcing each other, producing El Niño events.

The Bjerknes feedback explain the occurrence of SST anomalies, and ENSO events. Nevertheless, after ENSO event matures, negative feedbacks are needed to cease the ENSO anomaly growth, since these events cannot last forever. There are currently several hypotheses as to how ENSO events dissipate, which are excellently explained in Wang (2018).

5   Data

5.1   Variable choice

The choice of variables is based on the physical knowledge of ENSO dynamics. The relevant regions are:

Figure 11: Regions used in the analysis[9]

- West Pacific: 5°N - 5°S and 130°E - 150°E.

- Central Pacific: 5°N - 5°S and 150°W - 120°W.

- East Pacific : 5°N - 5°S and 100°W - 80°W.

- Niño 3.4 region: 5°N - 5°S and 170°W - 120°W.

- In between West & Central Pacific: 5°N - 5°S and 130°E - 180°E.

The analysis is divided into two parts. Firstly, a simple network of three variables is constructed that attempts to capture the dynamics of the Walker circulation. The second part of the results expands on this analysis, by including two extra variables. This network attempts to capture the dynamics of the Bjerknes feedback. The five variables, including their name as presented in the networks are:

- **WPAC:** Sea-level pressure in West Pacific. Measured in milibars (mb).

- **CPAC:** Near-surface temperature in Central Pacific. Measured in degrees celsius (°C).

- **EPAC:** Near-surface temperature in East Pacific. Measured in degrees celsius (°C).

- **EWW:** Eastward near-surface wind in between West & Central Pacific. This variable is multiplied by -1 in order to measure the strength of the easterlies. Measured in meters per second $(m\,s^{-1})$.

- **THF:**Total surface upward heat flux in between West & Central Pacific. This is the sum of sensible heat flux (energy that increases temperature of matter) and latent heat flux (energy to convert state of matter). Measured in watts per square meter $(w\,m^{-2})$.

---

[9]The EWW & THF and the Niño 3.4 region are slightly lowered for visual clarity. All variables are within the 5°N - 5°S latitude region.

While the first three variables can represent the Walker circulation system, the last two variables are added to better represent ENSO dynamics. Finally, sea-surface temperature (SST) over the El Niño 3.4 region (5°N - 5°S en 170°W - 120°W) is used to determine El Niño, La Niña and neutral conditions, by measuring deviations from the long-run trend. The long-run trend is calculated using a running mean with a window width of 30 years. ENSO usually peaks during winter (November, December and January), so the average over this period is taken to determine whether the year represents a El Niño, La Niña and neutral year. The use of this variable is further explained in section 5.3.2. The regions are shown in figure 11. Averages are taken over the region to extract time series.

5.2 Data Sources

5.2.1 Sources

The data for the first part of the analysis comes from the Climate Explorer (KNMI, 2023), while the second part consists of data from Copernicus Climate Data Store (CDS, 2023). As explained before, the first part consists of a simple three-variable network where Climate Explorer data will be used, since multiple ensemble runs are available for the same model.Constructing additional networks from the ensemble runs and compare them to the original network, serves as a robustness check for intra-model variability. Furthermore, the Climate Explorer offers averages over regions so that time series are given as output. This eases the pre-processing of the data. The Climate Data Store will be used for the second part of the results, since the advantage is that more variables are available.

5.2.2 Time period & Climate scenarios

All variables consists of monthly time series from 1950 until 2100. The time series are split into two periods of equal length: 1950-2025 and 2025-2100, in order to make comparisons between the two. The SSP5-RCP8.5 climate scenario is chosen since it assumes the most anthropogenic warming. The SSP5 is a socio-economic scenario that describes a future world based on fossil-fuel development. Moreover, the RCP8.5 is a climate scenario that assumes a radiative forcing of 8.5 $w\,m^{-2}$ at the end of the century.

5.2.3 CMIP6 model choice

The main CMIP6 model chosen for the analysis is KNMI's EC-Earth3 model, since it performs well in simulating tropical climate compared to observations and reanalysis data (Desmet and Ngo-Duc, 2021). Furthermore, as a robustness check in the second part of the results, we compare the estimated causal networks from EC-Earth3 model, to other CMIP6 models, namely HadGEM3-GC31-LL (UK),

MIROC6 (Japan), ACCESS-CM2 (Australia), GFDL-ESM4 (USA), MPI-ESM1-2-LR (Germany), and FGOALS-f3-L (China).

5.3 Data pre-processing

5.3.1 Trend and Seasonality

A trend and seasonal cycle may introduce non-stationary dependencies that are confounding and cannot be explained causally (Runge et al., 2023). Therefore, as a pre-processing, we estimate the the seasonal and trend components and extract them from the time series. An additive model is first estimated, where we assume that each time series X^i can be decomposed into a seasonal, a trend, and a residual component:

$$X_t^i = S_t^i + T_t^i + R_t^i \qquad (10)$$

The additive model starts by estimating the trend component T_t^i, by using a centered moving average with a window width of 12 months. This trend is subtracted, in order to estimate the seasonal component S_t^i, which is estimated by calculating each month's average over the entire periods. We then subtract only the seasonal component from the time series.

Next, the long-term trend is estimated through a Gaussian kernel smoother. This methods is equivalent to a running weighted mean, where the the observations around time t in the time series, are given a higher weight compared to observations further away. These weights are based on the Gaussian kernel. The window width of the running mean is chosen to be 15 years, in order to capture the long run trend, in accordance with Runge et al. (2023).

5.3.2 Masking procedure

Additionally, a mask is applied to the multivariate time series, in order to exclude some parts of the dataset. This allows to make use of background knowledge about causal stationary. In particular it allows to specify a time range within which a certain causal mechanism is believed to be stationary. For instance, if a one variable is thought to cause another variable during winter, but the regime is switched during summer, the masking procedure allows to isolate winter months from summer months and vice versa. Three masking procedures are applied:

- **Spring barrier**: The spring predictability barrier (SPB) refers to the period February-March-April where it exceptionally difficult to predict ENSO later in the year. This is because SST anomalies usually peak during winter months, and return to average during spring. Therefore, the spring barrier mask takes out the period February-March-April from the data.

- **Spring barrier towards El Niño:** This mask excludes the spring barrier, and the non-El Niño years. Since ENSO peaks in winter, if the average SST anomaly during November-December-January is smaller than +0.5, then we exclude the period May-January only [10]

- **Spring barrier towards La Niña:** This mask excludes the spring barrier, and the non-La Niña years. The procedure from the previous mask is applied, except for cases with SST anomalies during November-December-January greater than −0.5.

5.4 Summary statistics
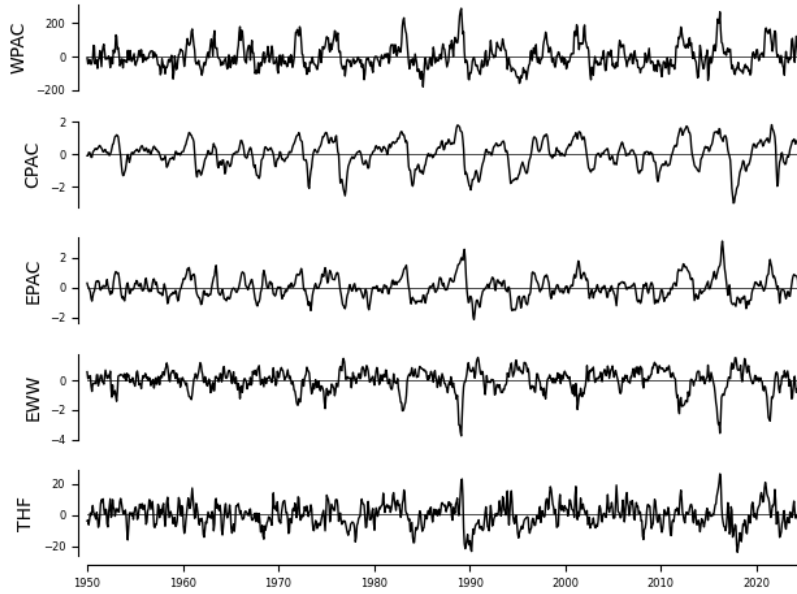
Figure 12: Time series of all variables from 1950 until 2025 (EC-Earth3)

Figure 12 shows the time series used for estimating the causal network during the period 1950-2025. The time series were deseasonalized and detrended, so their seasonal component and long-term trend were removed. We can already observe that the time series show characteristics of stationarity, although this will be formally tested in section 5.5.3.

Table 3 show some summary statistics of the main variables. Since the time series are deseasonalized and detrended, their mean is zero. We see that the median is not exactly zero, indicating that the time series are not exactly normal. Nevertheless, small deviations from normality are allowed for the use of the partial correlation test.

---

[10]As an example, if the average SST anomaly during November 2022, December 2022 and January 2023 is smaller than 0.5, then the period May 2022 - January 2023 is considered an El Niño year.

Table 3: Summary statistics for all five variables

	WPAC (mb)	CPAC $(°C)$	EPAC $(°C)$	EWW (m/s)	THF (w/m^2)
St. dev.	69.47	0.81	0.69	0.73	7.18
Min	-179.64	-3.00	-2.14	-3.76	-23.96
Max	285.52	1.80	3.08	1.55	26.32
Median	-11.71	0.09	-0.03	0.11	0.41

5.5   Assumptions

The time series and the subsequent diagnostic plots for the period 1950-2025 will be shown, while the same plots for the period 2025-2100 will be in appendix C. Since the analysis is performed for the two time periods separately, it is more appropriate to show these plots separately as well. This section will therefore focus on verifying some of the assumptions used in the methodology.

5.5.1   Normality
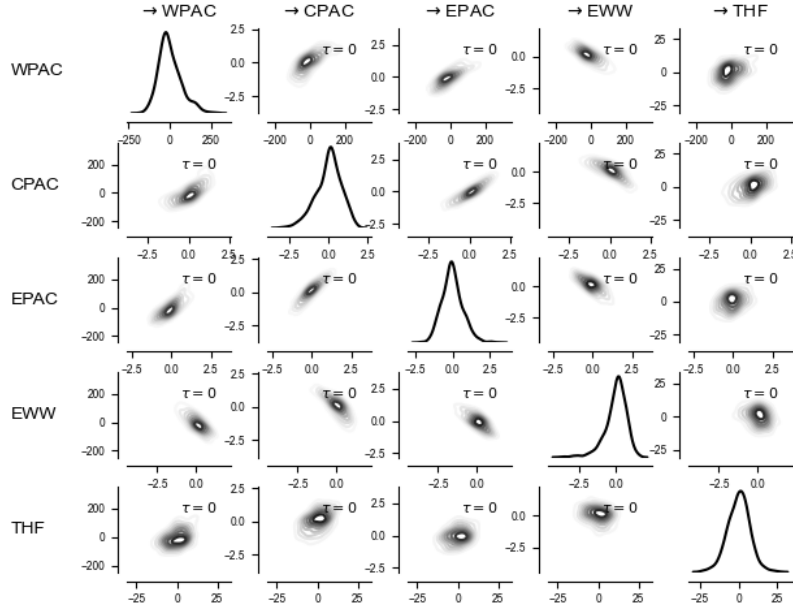
Figure 13: Density plots of all variables from 1950 until 2025

Figure 13 shows marginal densities of all variables in the diagonal, and joint densities in the off-diagonal plots. All variables show characteristics for the normality assumption to hold. However, EPAC and EWW show evidence of fatter tails than usual. Nevertheless, as explained in section 3.4, the partial correlation test is robust against departures from normality.
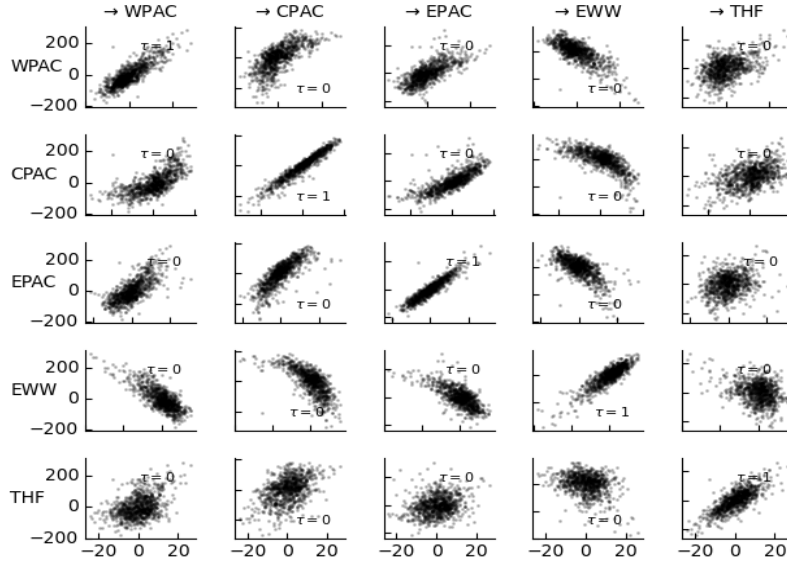
Figure 14: Scatterplots of all variables from 1950 until 2025

5.5.2 Linearity

Figure 14 show scatterplots for all variables. This diagnostic plot is check for linear or non-linear relation among variables. Along the diagonal scatterplots are shown of variables with their first lag. The off-diagonal elements show simple scatterplots among variables. Most relations seem to be linear, although some seem to be slightly non-linear, as is the case with scatterplots with the Eastward wind (EWW). The non-linearities are sightly stronger for 2025-2100 data, as shown in figure 26 in appendix C. Nevertheless, the non-linearity is not very strong, and thus we continue with the assumption of linear relationships in the data, in order to use the partial correlation test.

5.5.3 Stationarity

Table 4: Augmented Dickey-Fuller test results

	WPAC	CPAC	EPAC	EWW	THF
Test-statistic	-7.809	-7.176	-7.313	-7.821	-7.004
p-value	0.000	0.000	0.000	0.000	0.000

Stationarity is formally tested using an Augmented Dickey-Fuller (ADF) test. The maximum number of lags in the Dickey-Fuller regression is chosen based on the Akaike information criterion (AIC), and no constant nor trend are included in the regression. This is because the mean of all variables is equal to zero, and the time series have been detrended. The ADF test results reject the null hypothesis of a unit root (non-stationarity), and thus there is strong evidence of stationarity.

6 Results

6.1 Walker circulation

6.1.1 Causal network results

This section provides the first estimated causal graphs. As explained in section 5, we first provide estimates for the causal graph representing the Walker circulation. This system can be explained by three variables, namely SLP in the Western pacific (WPAC), and near surface temperature in central pacific (CPAC) and eastern pacific (EPAC). For this section, I will sometimes refer as **model 1** to the specification that includes all observations (or all years/data points), **model 2** to the specification that masks the spring barrier, model **model 3** excludes spring barrier, and includes only La Niña conditions, and **model 4** also excludes spring barrier, but includes only El Niño conditions.
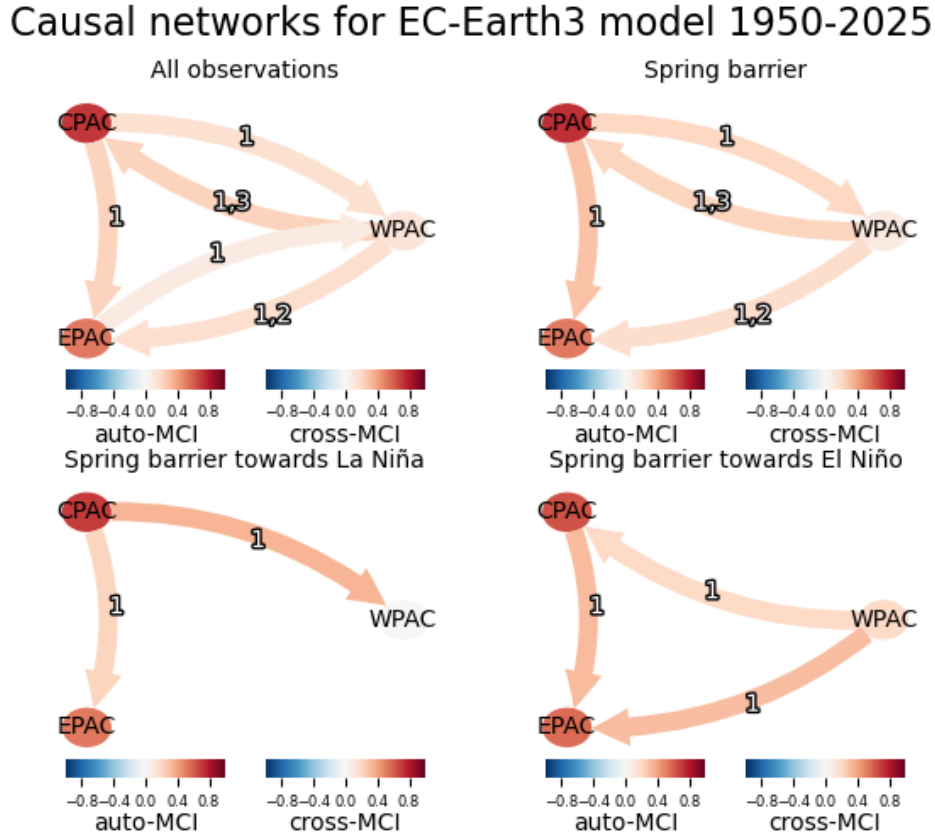
Figure 15: Causal networks for all four specifications, EC-Earth3 model, using data for the period 1950-2025.

The upper left panel of figure 15 shows the causal network for all observations. Most possible links are found to be statistically significant, possibly due to the inclusion of multiple regimes: including spring barrier, El Niño, La Niña and neutral conditions. The fact that many links among variables are bidirectional is also likely to be due

to the inclusion of multiple regimes. Furthermore, model 2's results are similar to those of model 1, except for a strengthening of multiple links and one missing link between EPAC and WPAC. The strengthening can be seen with the fact that some links between variables are darker. This can be explained by the fact that during the spring barrier, ENSO conditions are most uncertain are more difficult to predict, and thus excluding this part of the data strengthens the relation among variables. Lastly, we observe that many links are not statistically significant anymore for the last two specifications. We do however observe the direction change that is expected between CPAC and WPAC. During La Niña conditions, the easterlies strengthen,and thus the directionality goes eastwards. During El Niño conditions, the easterlies are weakened and often reversed, which explain the change in directionality.
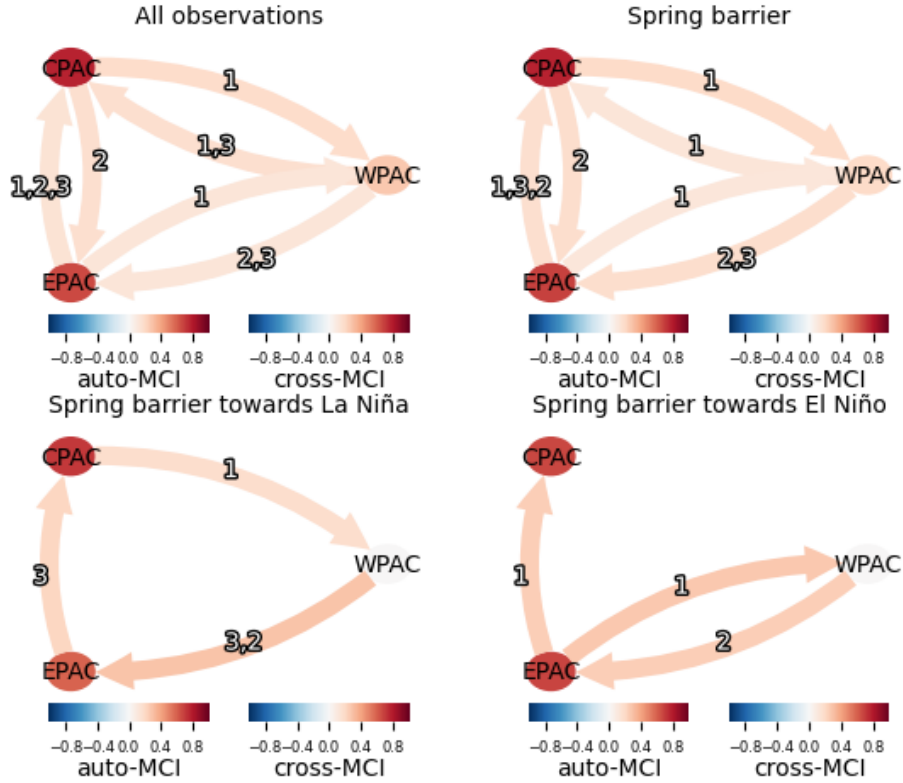

Figure 16: Causal networks for all four specifications, EC-Earth3 model, using data for the period 2025-2100.

Figure 16 shows estimated causal networks for all four specifications, using simulation data from 2025 until 2100. The first two model specification give similar results to those networks where 1950-2025 data is used. For model 3, we observe the feedback loop between the three variables, which is strengthened during La Niña conditions. It is however surprising to see the bidirectional relation between WPAC and EPAC during El Niño conditions.

6.1.2 Robustness check

For the three variables used to model the Walker circulation, alternative ensemble runs were available for the EC-Earth3 model. The variation between ensemble runs arise from slightly different input conditions or simulation parameters, while holding the equations that represent the processes and interactions that drive the Earth's climate, the same (Potter et al., 2009). The results are not expected to change dramatically when using alternative ensemble runs, and thus, we repeat the previous analysis for other ensemble runs and compare the results. This serves to check if the estimated networks are robust. The EC-Earth3 model has a total of seven ensemble runs, and thus, we will compare the original ensemble run used to construct the estimated network in section 6.1.1, to the remaining six ensemble runs.
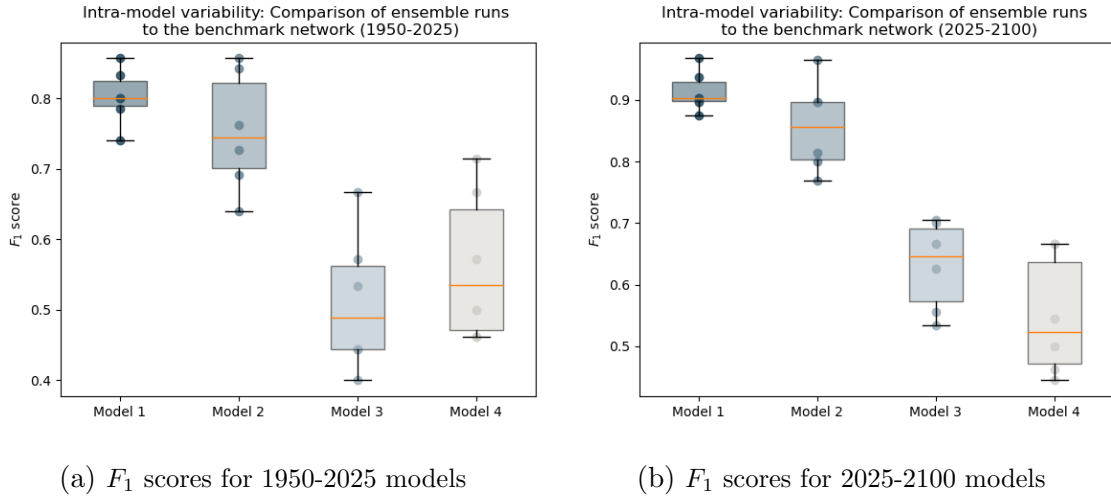

(a) F_1 scores for 1950-2025 models   (b) F_1 scores for 2025-2100 models

Figure 17: Intra-model variability: F_1 scores for ensemble runs compared to benchmark network

We observe in figure 17 the results of comparing different ensemble runs to the benchmark ensemble run used in section 6.1.1. High F_1 scores represent networks that are similar to the benchmark network in terms of the significant links found, while low F_1 scores represent dissimilar networks. We observe that model 1 and 2 (i.e., All observations, and SB masking) are robust to using different ensemble runs of the same CMIP6 model, while model 3 and 4 (i.e., SP towards El Niño and SP towards La Niña) are not robust. It is important to mention that the last specification uses a time-varying masking procedure. The decision on whether a certain time period is masked or not, depends on the simulated atmospheric conditions, that are likely to differ among ensemble runs during specific time periods. This results in comparing networks constructed with different sample sizes, and thus F_1 scores are expected to be lower compared to those comparisons made with fixed sample size (as for model specifications 1 & 2). Nevertheless, the results differ to much between ensemble runs, and thus we conclude that the results are not robust. Estimated networks from alternative ensemble runs can be found in appendix G.

6.1.3 Comparison under strong global warming

Table 5 display the results of the comparison metrics between causal networks for the EC-Earth3 model, for the period 1950-2025 and the period 2025-2100. Because the Average strength difference between the two networks is positive for all four specifications, the average strength of the benchmark network (1950-2025) is stronger compared to the comparison network (2025-2100). This can be seen in the graphs themselves, since the color of the links for the first network are darker compared to the second network. This suggest that the walker circulation weakens under strong global warming, which is in line with the literature that predicts this slowdown using climate model simulations (Vecchi and Soden, 2007; Zhao and Allen, 2019; Heede et al., 2021). Nevertheless, the weakening or strengthening of the relationship among these variables is found to significantly vary among CMIP6 models (Bayr and Latif, 2023).

Table 5: Comparison between causal networks for the period 1950-2025 & 2025-2100.

Specification	Average strength difference	Coefficient deviation	F1-score
All observations	0.0072	0.082	0.666
Spring barrier	0.0100	0.086	0.500
SB towards La Niña	0.0647	0.139	0.545
SB towards El Niño	0.0201	0.159	0.363

The absolute coefficient deviation shows how the two networks differ in terms of the final coefficients that are estimated. We see that the metric increases up to 0.159 in the last model specification. The value represents the average difference in coefficients between the same links. Since the coefficients can only be between $[-1, 1]$, we can say that this difference is relatively high. The F_1 score represents how well the predicted model (i.e. comparison network) is able to replicate the actual network. In this case, it is measure of how the 2025-2100 model is able to find the same significant links as the 1950-2025 model. As we can see, the F_1 is relatively low, and decreases the stricter the specification becomes. Therefore, we see that the network differ substantially.

6.2 ENSO results

6.2.1 Causal network results

This section provides the second set of estimated causal graphs. These results attempt to represent ENSO dynamics, by expanding on the previous results. This system can be explained by five variables, where Eastward winds (EWW) and total heat flux (THF) are added. Firstly, by comparing these results to those in section

[6.1.1](), we see some small changes in the relation between WPAC, CPAC and EPAC. Figures [18a]() and [18b]() estimate a negative WPAC → CPAC link. Furthermore, the WPAC → EPAC link is absent in figure [18a](), and it seems that the connection is mediated through EWW and THF. All in all, the changes between figure [18]() and [15]() are relatively modest.
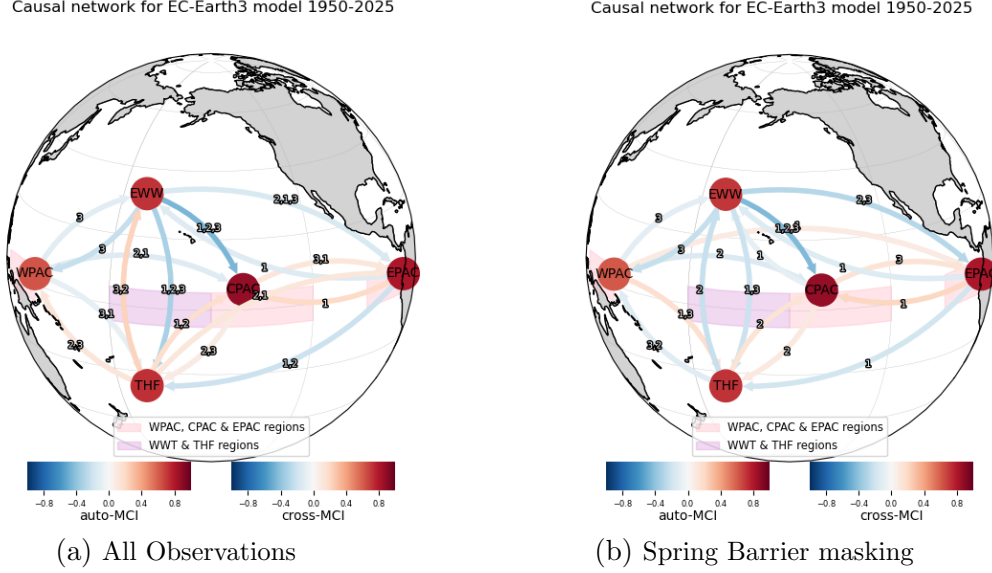

(a) All Observations   (b) Spring Barrier masking

Figure 18: Causal networks for EC-Earth3 model (1950-2025)

Ignoring the directionality of the links, we observe that many links have the expected sign. Since strong easterlies pushes warm water away from the central and East Pacific, we expect a negative relation between EWW and CPAC & EPAC. Furthermore, stronger trade winds are associated with La Niña conditions, and thus low sea surface pressure in the west. This is why we observe a negative EWW ↔ WPAC relation as well. Moreover, Total heat flux (THF) is positively associated with CPAC, as expected. This is because warm temperatures in central Pacific, lead to upward heat fluxes in the form of evaporation (latent heat) and increasing SST (sensible heat). Furthermore, in all model specifications, we observe a negative EWW ↔ THF relation, which is in line with the literature ([Bayr et al., 2017]()). This is because strong winds push during La Niña conditions brings cold water along the equator and thus less outward energy.

Nevertheless, there are some signs that were not expected. The relation between EPAC and THF is expected to be mediated through CPAC, but the algorithm found a negative EPAC → THF link. Because many relationships are bidirectional, it becomes difficult to asses whether the Bjerknes Feedback is well represented in the model. Similarly to the previous section, this might be due to the fact that the data contains multiple regimes.

Figure [19]() show more results. We observe similar results compared to the three-variable network in figure [15](). For La Niña events, we see a clear causal link from

EPAC to CPAC, due to the effect of EWW, which also has a significant negative link with the two variables. Furthermore, for the El Niño specification in figure 19b we see the same bidirectional relation between CPAC and EPAC, possibly due to the fact that during El Niño events, the trade winds either weaken or reverse. Interestingly as well, we see that causal strength is generally higher during La Niña conditions.
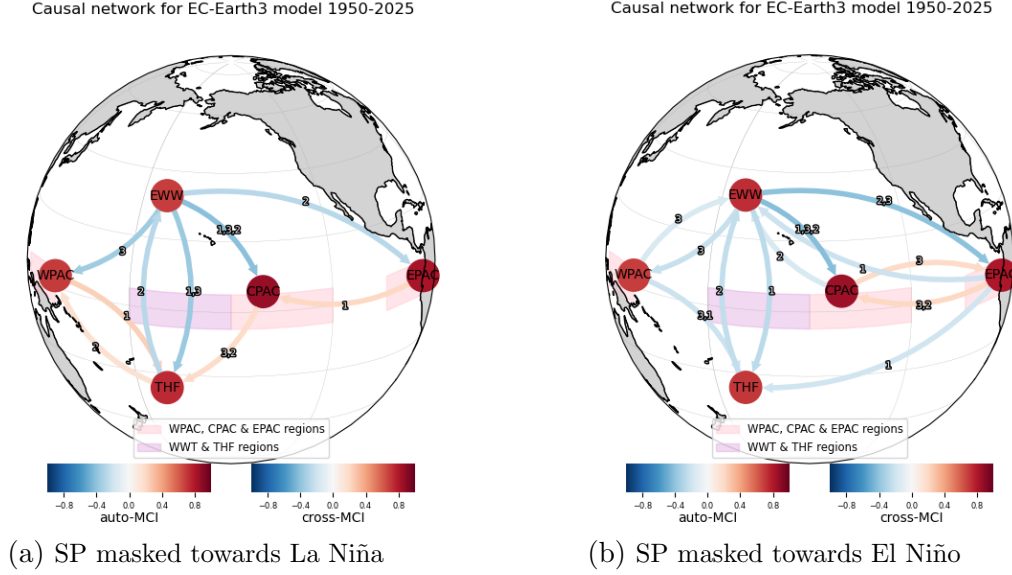

(a) SP masked towards La Niña

(b) SP masked towards El Niño

Figure 19: Causal networks for EC-Earth3 model (1950-2025)

The network in 19a, and to some extent, in the network in figure 19b, are characterized by the variable EWW which strongly affects all remaining variables in the network. This falls in line with the hypothesis reviewed by Wang (2018) and postulated by Gebbie et al. (2007) that ENSO events are triggered by stochastic atmospheric/oceanic forcing, such as westerly wind bursts.

In figure 19a we observe a structure somewhat resembling a feedback loop. EWW affects both EPAC and CPAC, and EPAC affects CPAC, because during La Niña events, strong winds push water eastwards, and thus the cold water upwelled in the east travels to the central pacific affecting air temperature. Furthermore, colder temperatures in the central pacific leads to lower THF, and in turn, lower SLP in the West Pacific (WPAC). In addition, outward heat flux (THF) negatively affects trade winds, which closes the loop.

The El Niño network has several links that are unexpected. For instance, The negative WPAC → THF link, since this relation is expected to be positive. In addition, the EPAC → THF link is negative, even though it is expected to have no relation, or the same type of relation compared to CPAC. This negative link might be due to a delayed response between CPAC and EPAC, since, it takes time for the cold water to travel from east to central pacific. This would be coherent with
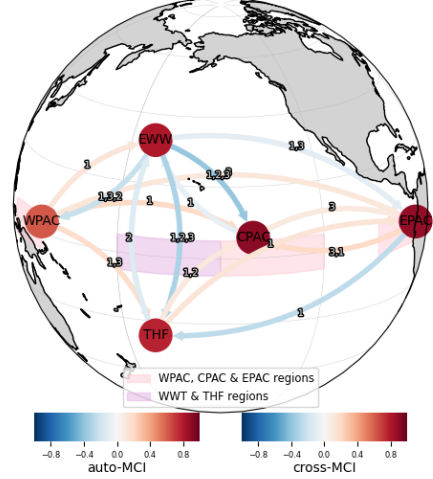
the links between EPAC and CPAC, since they are only significant for lags two and three.

Causal network for EC-Earth3 model 2025-2100
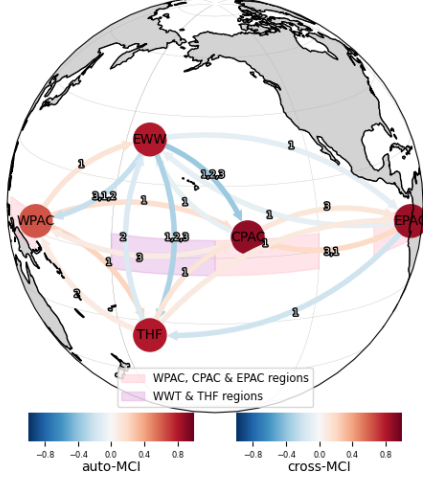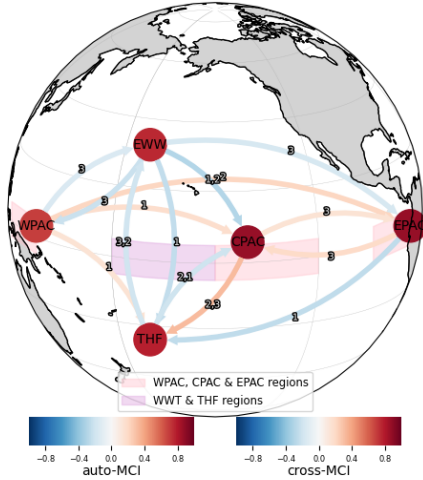

(a) All Observations

Causal network for EC-Earth3 model 2025-2100
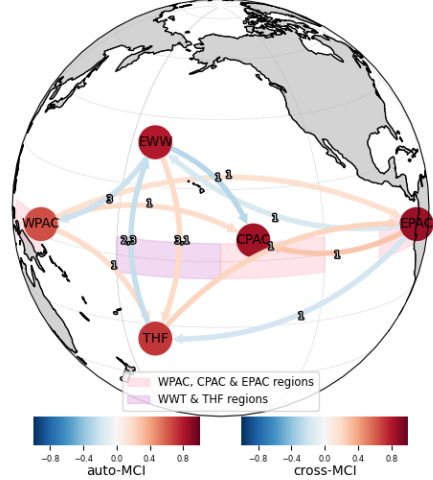

(b) Spring Barrier masking

Figure 20: Causal networks for EC-Earth3 model (2025-2100)

Causal network for EC-Earth3 model 2025-2100

(a) SP masked towards La Niña

Causal network for EC-Earth3 model 2025-2100

(b) SP masked towards El Niño

Figure 21: Causal networks for EC-Earth3 model (2025-2100)

Figure 20 and 21 show the estimated causal networks of all four specification for the period 2025-2100. Generally, the estimated plots are similar to those estimated for the period 1950-2025.

6.2.2 Robustness check for ENSO networks

Similarly to section 6.1.2, F_1 scores are computed for the ENSO networks. Since different ensemble runs for the same CMIP6 model were unavailable, the F_1 scores are constructed by comparing to estimated networks from different CMIP6 models. The results are presented in figure 22 and 23, with respect to each of the four model specifications, as well as each CMIP6 model.

Overall, there is a much smaller difference between model specifications, as wee see that the median F_1 score, denoted by the orange line, varies less in figures 22a and 23a compared to those in figure 17. The minimum F_1 score of all four specifications is relatively low, and could be considered outliers, since they all pertain to the F_1 scores associated to the FGOALS-f3-L model in figures 22b and 23b. Furthermore, there is no clear difference in variability among F_1 scores for the 2025-2100 networks compared to the 1950-2025 networks, since the boxplots are not consistently wider for one set of networks. Surprisingly, inter-model variability, as depicted in figures 22 and 23, is slightly lower compared to intra-model variability in figure 17. This is unusual, since it was expected that different ensemble runs of the same CMIP6 model would reproduce similar networks, compared to networks estimated from different CMIP6 models. The ENSO networks might better represent the underlying process compared to the Walker circulation networks.
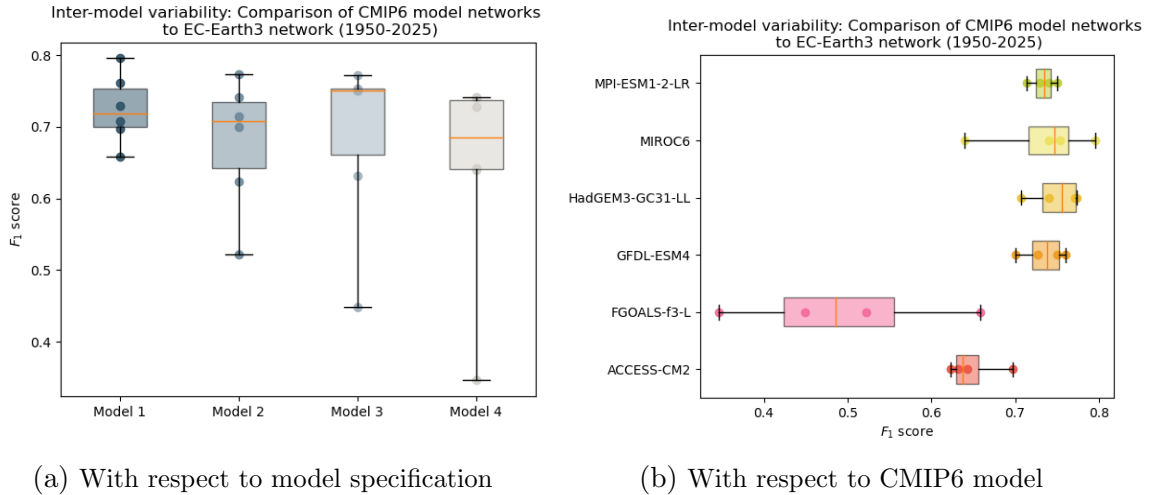

(a) With respect to model specification   (b) With respect to CMIP6 model

Figure 22: Inter-model variability: F_1 scores for ENSO networks (1950-2025)

Comparably to figure 17, moving from the less restrictive model specification (All observations) to the more restrictive specification (SB masked + El Niño), we see that the F_1 scores start to decrease, although a much smaller decrease is seen for ENSO networks. This might simply be because the available sample decreases from the less to more restrictive specification, and thus, identifying the same links becomes more difficult.

In addition, most CMIP6 models have a relatively high F_1 score, indicating that

(a) With respect to model specification

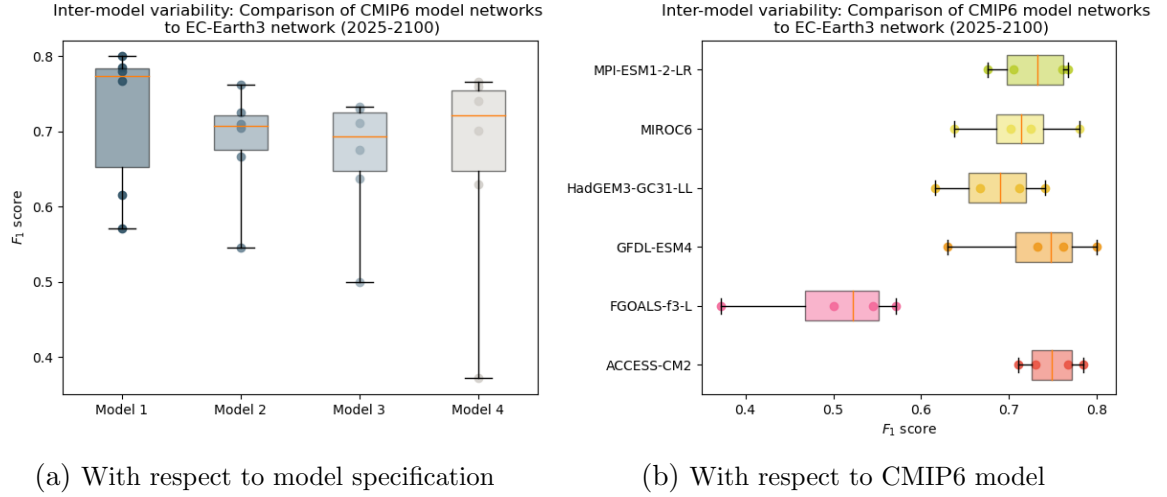(b) With respect to CMIP6 model

Figure 23: Inter-model variability: F_1 scores for ENSO networks (2025-2100)

these models are able to identify many of the same links compared to those identified using EC-Earth3 simulation data. The big exceptions is the FGOALS-f3-L model, which has very low F_1 scores

6.2.3 Comparison under strong global warming

Table 6: Comparison between causal networks for the period 1950-2025 & 2025-2100, for ENSO networks.

Specification	Average strength difference	Coefficient deviation	F1-score
All observations	-0.0010	0.087	0.776
Spring barrier	0.0134	0.086	0.769
SB towards La Niña	0.0425	0.097	0.754
SB towards El Niño	0.0204	0.092	0.733

Table 6 shows some metrics for the comparison of the 1950-2025 ENSO networks, and the 2025-2100 networks in the previous section. We observe again that the Average strength difference is positive for all of them except the specification where all observations are used, which is negative and very close to zero. Since the coefficients from the 2025-2100 networks are subtracted from the 1950-2025 networks, positive average strength difference indicate that the causal strength doe the 1950-2025 networks are higher than the 2025-2100 networks. This is again evidence that the causal strength between variables weakens in the future. The coefficient deviation shows how much deviation there is in the final estimated coefficients. The numbers are much smaller compared to the Walker circulation networks, indicating there is a smaller difference between the networks. The same can be said in terms of the

estimated links, since the F_1 scores are relatively high, which indicates that the 2025-2100 networks are able to correctly estimate the same links most of the times compared to the 1950-2025 networks.

6.2.4   Trade wind trends and causal strength differences

Most CMIP6 models simulate weaker trade winds until the end of the century due to strong global warming. This has a direct effect on the conclusion of a weakened Walker circulation. This section investigates the long-term trend of trade wind strength, and how the estimated coefficients from the networks change.

(a) Time Series Trade winds (1950-2100)   (b) Estimated coefficients for specification 1
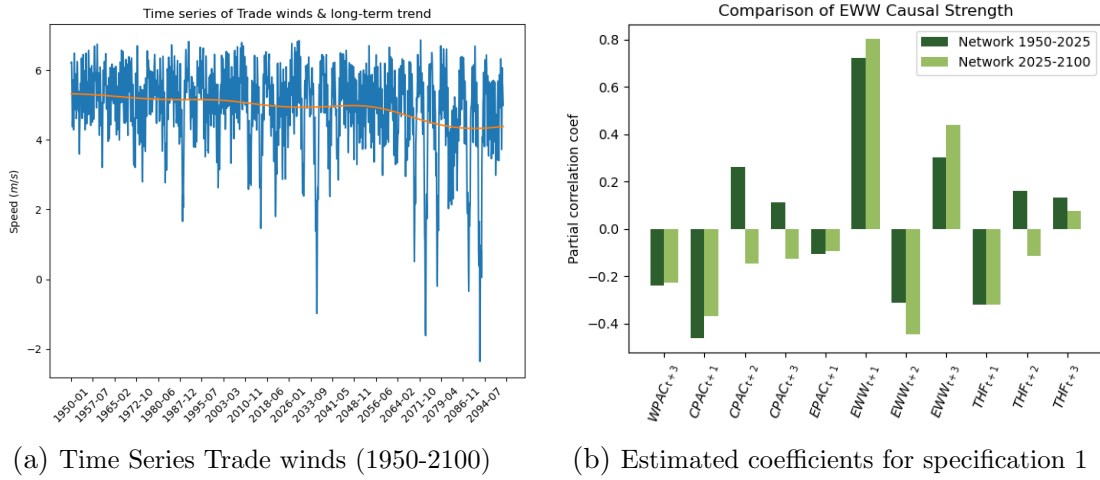
Figure 24: Time series and partial correlation coefficients (EC-Earth3)

Figure 24a shows a clear downward trend, which is a characteristic of most CMIP6 models. Figure 24b show the difference in estimated coefficients between the 1950-2025 network and the 2025-2100 network, for the same link. Therefore, only the common links between the two networks are shown. The first bar shows the estimated partial correlation coefficient of the link $EWW_t \rightarrow WPAC_{t+3}$ which is equivalent to $EWW_{t-3} \rightarrow WPAC_t$. The resulting coefficients come from the first specification, which uses all observations, and data from the EC-Earth3 model.

We see no evidence of a consistent change in coefficients, and thus, no change in the causal strength of EWW to all other variables. The plots for the remaining specification can be found in appendix F, but they also show no consistent difference in partial correlation coefficients. Therefore, we conclude that even though there is a negative long term trend in Trade wind strength, there is no change in causal strength with itself and other variables.

7   Discussion & Conclusion

Throughout this thesis, the central research question guiding the investigation has been whether causal discovery algorithms perform well in describing ENSO dynamics, and how do ENSO dynamics change under strong global warming. The first step was to analyze the Walker circulation through a three-variable network. Reconstructing the networks based on all the available data, and masking only the Spring barrier, yielded many estimated links, while the last two specifications (El Niño and La Niña years) produced sparser results, with only a few estimated links. Using data from 2025-2100 instead of 1950-2025 produced substantially different results, as evident from the Coefficient deviation metric and the F_1-score. We further learn that the results vary greatly depending on the ensemble run chosen, especially for the last two specifications. Remember that ensemble runs of the same model, keep the underlying biophysical processes and interactions the same, while changing only the initial conditions. Therefore, it was unexpected to detect remarkably different network from the same CMIP6 model.

The second part of the results added EWW and THF to the network, to better capture ENSO dynamics and atmosphere-ocean feedbacks. Many of the signs of the relations among variables are as expected, although the bidirectional links hinders interpretation of causality. Comparing 1950-2025 and 2025-2100 networks yield a smaller Coefficient deviation metric and F_1-score, meaning there are less changes in the network over time. Furthermore, the results are much more robust than before, since comparing to networks from other CMIP6 models yield higher F_1-scores compared to the ensemble run comparisons made for the Walker circulation networks.

There has been an increasing concern about the ability to correctly simulate ENSO dynamics in coupled general circulation models, such as CMIP5 & CMIP6 (Guilyardi et al., 2009). For instance, the amplifying zonal wind feedback and the damping heat flux feedback, two relevant processes of ENSO dynamics, are harshly underestimated according to Bayr and Latif (2023). It was also shown that estimated feedback parameters are highly variable among CMIP5 and CMIP6 models (Bayr et al., 2017). They also found that both feedbacks are dependent on sea surface temperatures (SST) suggesting a link to the equatorial cold (tongue) bias. The equatorial cold bias is a common problem in many modern climate simulation models, and it is caused by several factors, including too strong mean easterly zonal winds, which leads to increased ocean upwelling of cold subsurface water, and too weak mean short-wave solar radiation due to overestimated mean cloud cover. Seager et al. (2019) show that the erroneous East Pacific warming in climate simulation models is a consequence of the cold bias of their equatorial cold tongues. The many problems of CMIP6 models to correctly simulate ENSO dynamics explains the varying results of this thesis. For further research, the analysis can be redone for CMIP6 models

that best represent ENSO dynamics and whose biases are the smallest according to the literature.

Furthermore, alternative modeling choices could have been taken during the network reconstruction phase. For instance, certain links among variables could have been fixed to zero based on background knowledge on the physical processes. Nevertheless, this undermines the objective in the first place, which is to see whether the causal network could be reconstructed based on the data alone. Alternatively, Runge et al. (2023) also quantifies the causal effects among the processes involved in the Walker circulation, and as a prepossessing step, they take averages of two-month periods, to convert the data to bimonthly, in order to reduce monthly noise in the data. This significantly reduces the available sample, but is an option to consider for further research.

In addition, it might be the case that some of the more difficult to prove assumptions do not hold. For instance, variables such as the depth of the thermocline, cloud cover, or short-wave radiation might also be important to explain ENSO dynamics, meaning that the causal sufficiency assumption might not hold. The temporal precedence assumptions states that cause must be in the past, but there might be contemporaneous effects at play.

This thesis highlights the challenges of inferring causality from observational data in climate sciences. The different estimated networks give a good picture of the relation among variables in the system, but due to the lack of robustness, they can hardly be interpreted as the true causal graph. Furthermore, the use of CMIP6 model simulation data has also been shown to be a limitation, and thus, observations and reanalysis data might yield robuster results. Nevertheless, the field of causal discovery in climate sciences is still in its infancy, and further research is needed in the development of methods as well as in the application to climate systems that are better understood than ENSO. This thesis may serve as a guide in setting up a causal discovery study, and all the modeling choices that are relevant for the results.

In summary, this thesis has delved into the dynamics of El Niño-Southern Oscillation (ENSO) under strong global warming through a comprehensive causal network reconstruction study using CMIP6 model simulations. It has shown the difficulties of applying causal network reconstruction algorithms to climate data, but also its promising nature. Further research endeavors in this area, building upon the insights gained here, can potentially refine our understanding of ENSO behavior and to inform robust planning and decision-making.

A Relation between correlation coefficient from multiple regression and partial correlation

Consider three random variables Y, X_1, and X_2 that follow a joint standard normal distribution. We examine the regression coefficient β_1 in the following regression:

$$Y = \beta_1 X_1 + \beta_2 X_2 + \varepsilon \tag{11}$$

and thus we want to estimate the effect of X_1 on Y controlling for the effect of X_2. Concurrently, we are also interested in the partial correlation between Y and X_1 given X_2, that is $\rho_{Y,X_1|X_2}$. The standard OLS results in matrix notation for β_1 can be expressed in terms of the variance and covariance:

$$\beta_1 = (\mathbf{XX'})^{-1} \mathbf{XY} = Var(\mathbf{X})^{-1} Cov(\mathbf{X}, \mathbf{Y}) \tag{12}$$

which in non-matrix notation is:

$$\beta_1 = \frac{\sum_{i=1}^{n} Y_i X_{1,i} - \sum_{i=1}^{n} Y_i X_{2,i} \sum_{i=1}^{n} X_{1,i} X_{2,i}}{\sum_{i=1}^{n} X_{1,i}^2 - \left(\sum_{i=1}^{n} X_{1,i} X_{2,i}\right)^2} = \frac{Cov(X_1, Y|X_2)}{Var(X_1|X_2)} \tag{13}$$

The standard formula for the regression coefficient of two variables is:

$$\rho_{Y,X_1} = \frac{Cov(Y, X_1)}{\sqrt{Var(Y)}\sqrt{Var(X_1)}} \tag{14}$$

The partial correlation coefficient is therefore:

$$\rho_{Y,X_1|X_2} = \frac{Cov(Y, X_1|X_2)}{\sqrt{Var(Y|X_2)}\sqrt{Var(X_1|X_2)}} \tag{15}$$

In order to derive these quantities, we use standard results from multivariate regression theory, which is named Lemma I in Durbin and Koopman (2012). A proof of the Lemma is derived in section 4.2 of the book. This result is used to derive:

$$Var(X_1|X_2) = Var(X_1) - Cov(X_1, X_2)^2 Var(X_2)^{-1} = 1 - \rho(X_1, X_2)^2$$
$$Var(Y|X_2) = Var(Y) - Cov(Y, X_2)^2 Var(X_2)^{-1} = 1 - \rho(Y, X_2)^2 \tag{16}$$

The last step occurs because the variances of all three variables are equal to 1, and since correlation coefficient is defined as: $Cov(X_1, X_2)/\sigma_{X_1}\sigma_{X_2}$, the covariance between two variables boils down to the correlation coefficient. Moreover, the Schur complement shows that any partial covariance can be expressed solely in terms of variances and covariances:

$$Cov(X_1, Y | X_2) = Cov(X_1, Y) - Cov(X_1, X_2)Cov(Y, X_2)Var(X_2)^{-1}$$
$$= \rho_{X_1,Y} - \rho_{X_1,X_2}\rho_{Y,X_2} \tag{17}$$

Now we plug in these values in equation 13 and 15:

$$\beta_1 = \frac{\rho_{X_1,Y} - \rho_{X_1,X_2}\rho_{Y,X_2}}{1 - \rho(X_1, X_2)^2}$$
$$\rho_{Y,X_1|X_2} = \frac{\rho_{X_1,Y} - \rho_{X_1,X_2}\rho_{Y,X_2}}{\sqrt{1 - \rho(Y, X_2)^2}\sqrt{1 - \rho(X_1, X_2)^2}} \tag{18}$$

We see that both numerators are the same. Multiplying the partial correlation coefficient by the ratio of standard deviations, as done by Waldorp and Marsman (2021), yields the regression coefficient:

$$\rho_{Y,X_1|X_2} = \frac{\rho_{X_1,Y} - \rho_{X_1,X_2}\rho_{Y,X_2}}{\sqrt{1 - \rho(Y, X_2)^2}\sqrt{1 - \rho(X_1, X_2)^2}} \times \frac{\sqrt{1 - \rho(Y, X_2)^2}}{\sqrt{1 - \rho(X_1, X_2)^2}} = \beta_1 \tag{19}$$

B Monte Carlo simulation on false positives in the partial correlation test

The partial correlation test is used in the PCMCI algorithm, for linear and normally distributed data. The test uses a function to convert the correlation coefficient that is bounded between $[-1, 1]$, into a number with domain $(-\infty, +\infty)$ which follows a t-distribution with $n - 2 - D_Z$ degrees of freedom, as explained in section 3.4. For large samples, relatively small correlation coefficients can already yield statistically significant results. This is crucial, since we want the algorithm to have an expected false positive rate in lines with the chosen critical value. In order to investigate this, a small Monte Carlo study is performed.

Three random samples of size $n = 1000$ are drawn from three different normal distributions, with different means and variances. A partial correlation test is performed on the simulated data, and the test statistic is compared to the 5% critical value. This process is repeated 1000 times, and a false positive rate is returned. The false positive rate is expected to be around the significance level. In order to evaluate whether the generated false positive rate is not close to the significance level just by chance, we repeated this process in increasing order to generate a histogram, which we expect to have a normal distribution with mean around the significance level. In other words, we expect to mistakenly reject the null hypothesis $\left(H_0 : \rho_{XY|Z} = 0\right)$, 5% of the times.
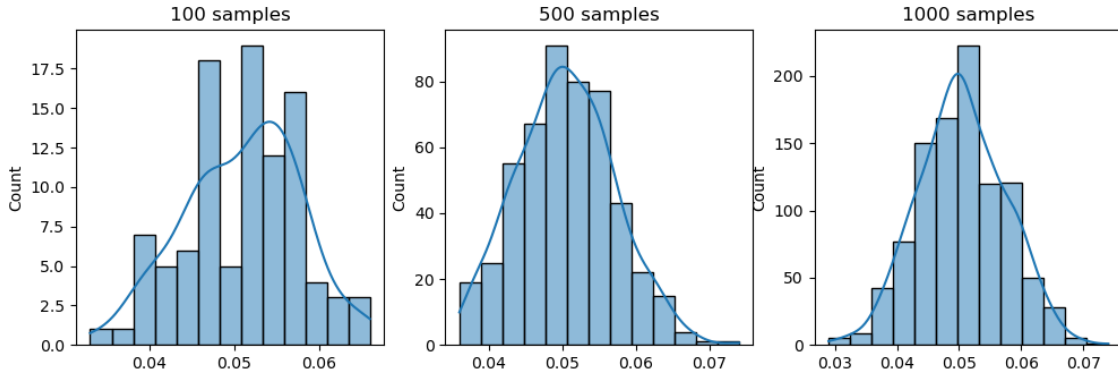
Figure 25: Results of Monte Carlo study, producing 100, 500 and 1000 false positive rates.

We observe that samples have a false positive rate that is normally distributed around 0.05. The results confirm the expected false positive rate from the test and it serves as validation for the use of the partial correlation test. A Python script is available in the GitHub repository, for replication purposes.

C   Additional diagnostic plots
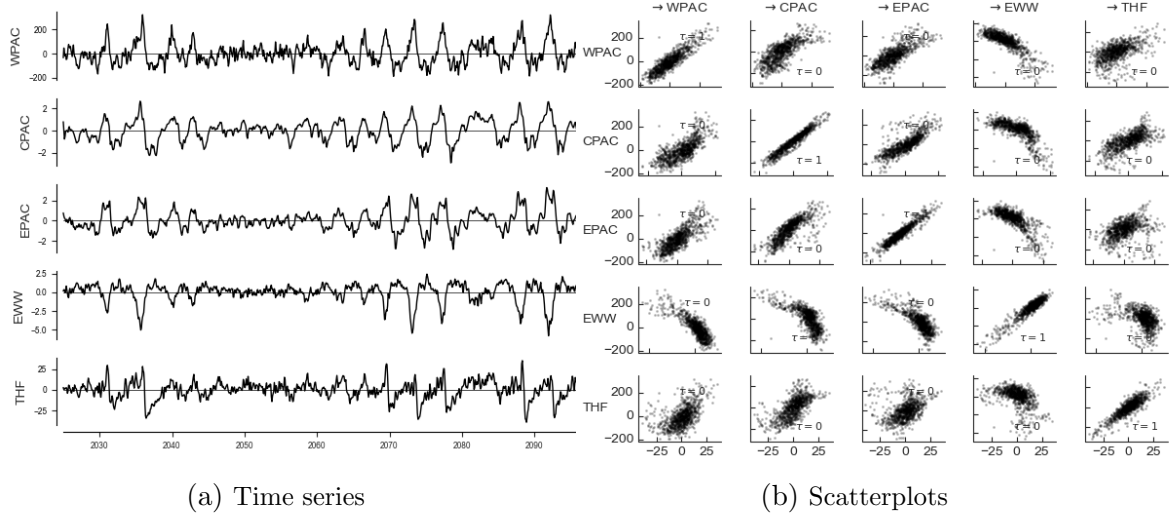

(a) Time series

(b) Scatterplots

Figure 26: Time series and Scatterplots for all variables from 2025 until 2100
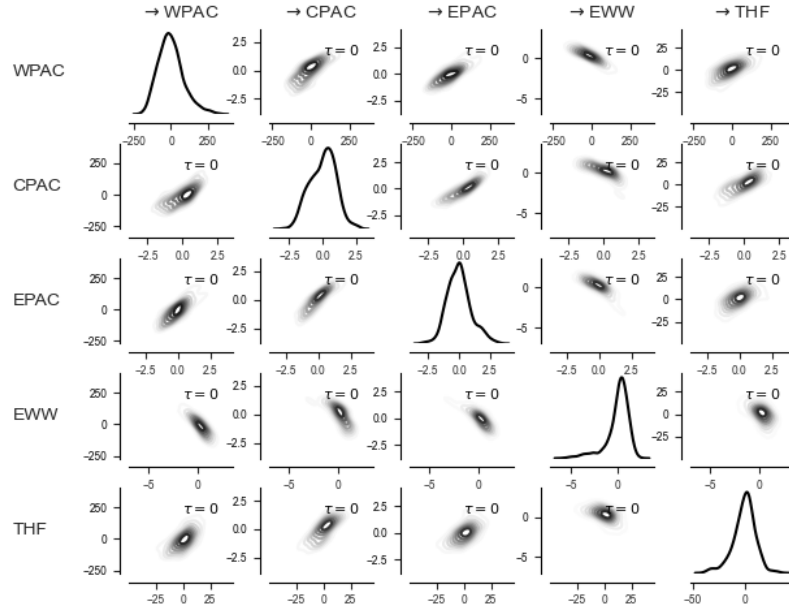
Figure 27: Density plots for all variables from 2025 until 2100.

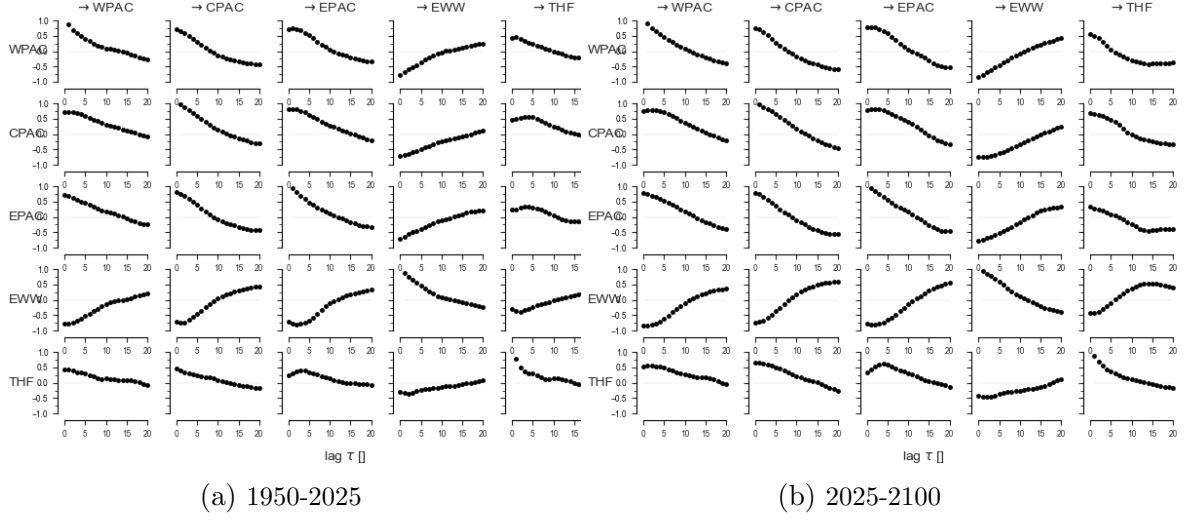# D   Lag plots of variables

(a) 1950-2025

(b) 2025-2100

Figure 28: Lag plots for main variables

Lag plots are used in the PCMCI methodology as a diagnostic plot, to asses the maximum number of lags used in the algorithm. The maximum number of lags should be equal to the maximum correlation between variables, in absolute values. We see that most correlations already peak at the very beginning. Some other correlations peak around a lag of 3, and a few peak around 4-5 lags. In order to not make the analysis more complex, a maximum lag number of 3 is chosen for all estimated networks in this report.

E   Data instructions and versions

The analysis of this report has been performed with Python 3.10.10 and Tigramite 5.2.0.2. Tigramite is a causal inference for time series python package. It allows to efficiently estimate causal graphs from high-dimensional time series datasets (causal discovery). The data has been downloaded from KNMI (2023) and CDS (2023), and here we provide instructions on how to download the data. All scripts and data is available in this GitHub repository. The fields used to download the data are:

CDS:

1. Press on 'Download Data' tab

2. Temporal Resolution: Monthly

3. Experiment: Historical & SSP5-8.5

4. Variable: Eastward near-surface wind, Near-surface air temperature, Sea level pressure, Surface upward latent heat flux, Sea surface temperature and Surface upward sensible heat flux.

5. Model: HadGEM3-GC31-LL (UK), MIROC6 (Japan), ACCESS-CM2 (Australia), GFDL-ESM4 (USA), MPI-ESM1-2-LR (Germany), and FGOALS-f3-L (China).

6. Select 1950-2014 for historical, select 2015-2100 for SSP5-8.5. In the Python script I merge the two to get a time series from 1950 until 2100.

7. Month: all months from January to December.

8. Geographical area: Sub-region extraction. Specified in section 5.1. Note that both Eastern and Southern hemisphere coordinates need to be specified with a minus in front.

Climate explorer:

1. Row:

   - Model:'16 EC-Earth3'
   - Exp: ssp585

2. Column: 'tas' and 'psl'

3. Press on select field to go to next page

4. Scroll down to ensemble member and press on analyse separately.

5. Get grid points: select according to section 5.1.

6. Press on 'make time series'

7. right click 'raw data' and press 'download linked file as...' in order to download in .dat format.

F   Additional Trade winds causal strength plots
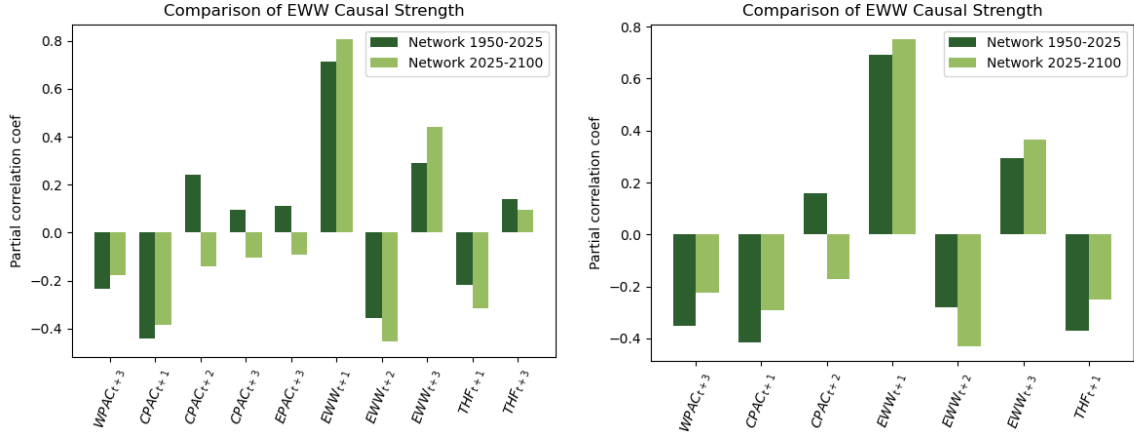
(a) Causal strength difference SP masking specification

(b) Causal strength difference SP masking + La Niña specification
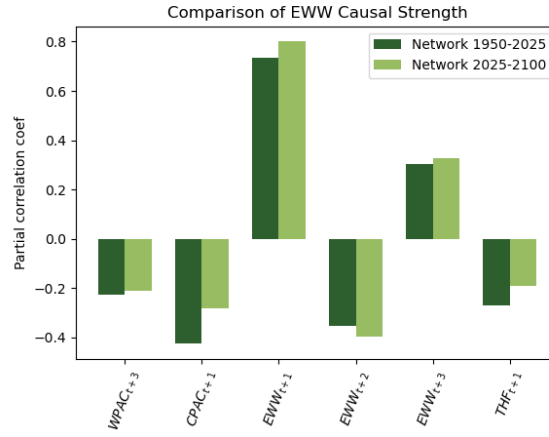
(c) Causal strength difference SP masking + El Niño specification

Figure 29: Differences in estimated coefficients. Remaining specifications (EC-Earth3).
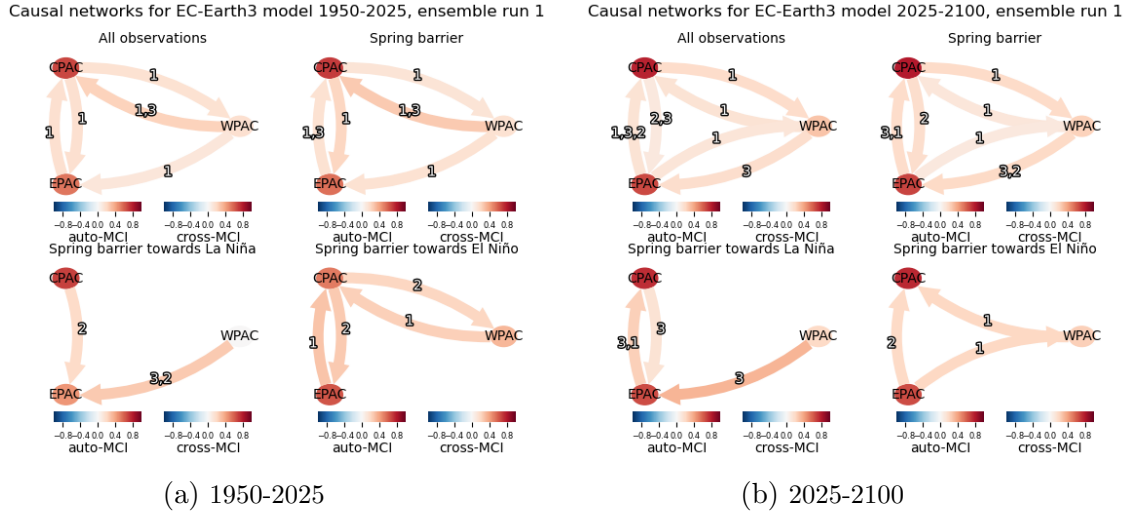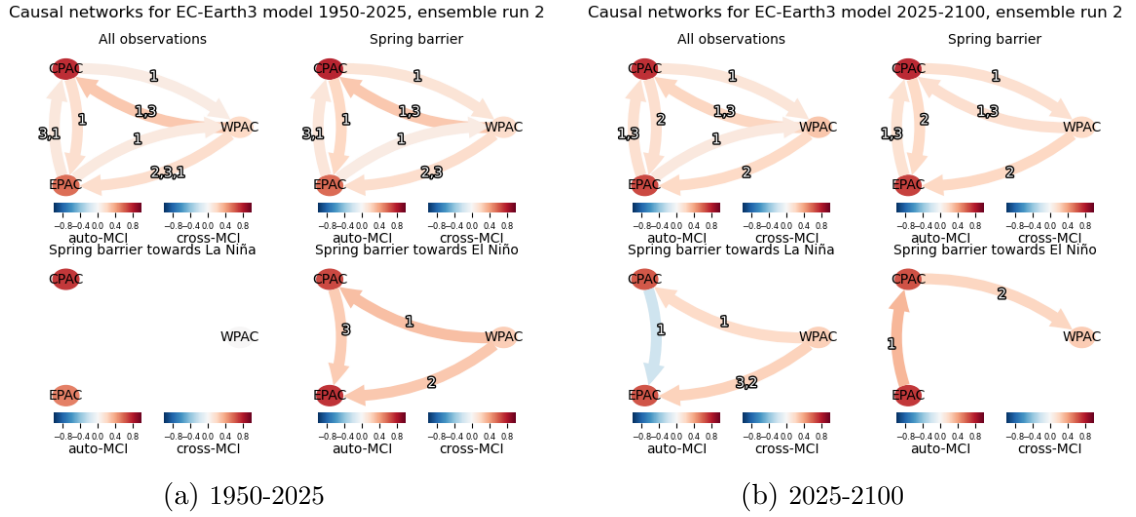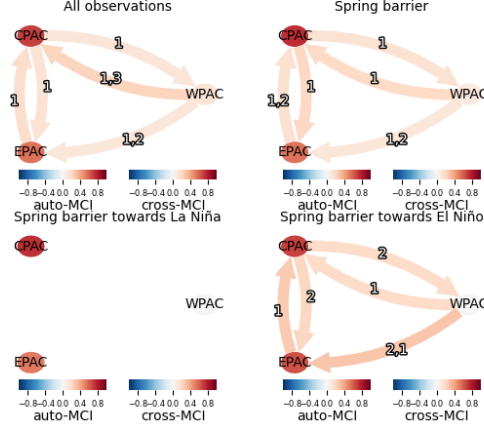
G   Additional empirical results

(a) 1950-2025

(b) 2025-2100

Figure 30: Results ensemble run 1

(a) 1950-2025

(b) 2025-2100

Figure 31: Results ensemble run 2

(a) 1950-2025

(b) 2025-2100

Figure 32: Results ensemble run 3
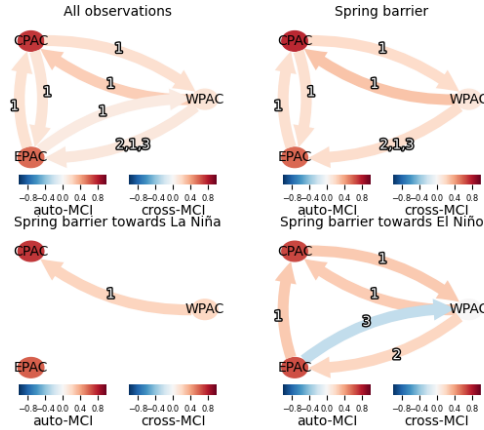
(a) 1950-2025

(b) 2025-2100

Figure 33: Results ensemble run 4
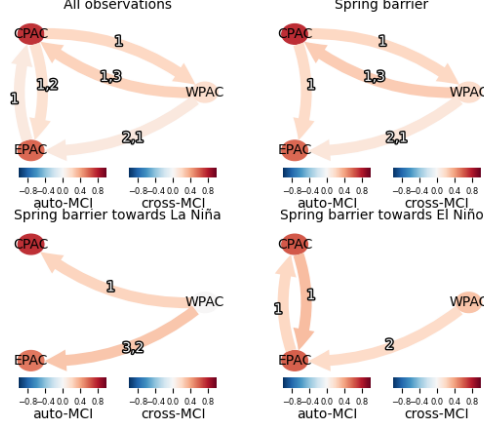
(a) 1950-2025 (b) 2025-2100

Figure 34: Results ensemble run 5
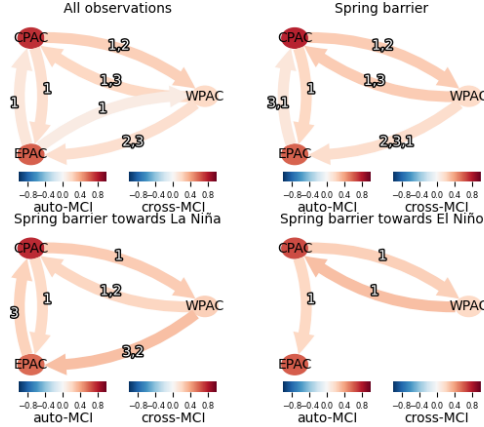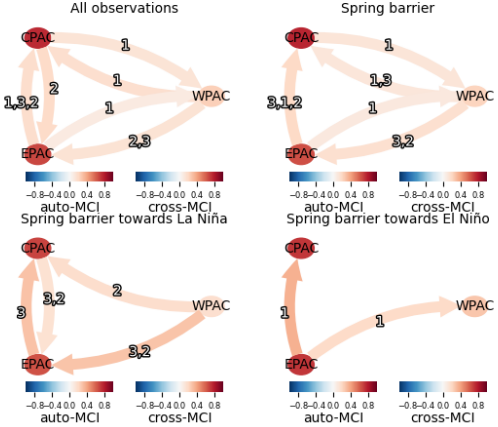
(a) 1950-2025 (b) 2025-2100

Figure 35: Results ensemble run 6

References

Baba, K., Shibata, R., and Sibuya, M. (2004). Partial correlation and conditional correlation as measures of conditional independence. *Australian and New Zealand Journal of Statistics*, 46(4):657–664.

Bayr, T. and Latif, M. (2023). Enso atmospheric feedbacks under global warming and their relation to mean-state changes. *Climate Dynamics*, 60(9–10):2613–2631.

Bayr, T., Latif, M., Dommenget, D., Wengel, C., Harlaß, J., and Park, W. (2017). Mean-state dependence of enso atmospheric feedbacks in climate models. *Climate Dynamics*, 50(9–10):3171–3194.

Beniston, M., Stephenson, D. B., Christensen, O. B., Ferro, C. A. T., Frei, C., Goyette, S., Halsnaes, K., Holt, T., Jylha, K., Koffi, B., Palutikof, J., Scholl, R., Semmler, T., and Woth, K. (2007). Future extreme events in European climate: an exploration of regional climate model projections. *Climatic Change*, 81:71–95.

Benjamini, Y. and Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society*, 57:289–300.

Cai, W., Borlace, S., Lengaigne, M., van Rensch, P., Collins, M., Vecchi, G., Timmermann, A., Santoso, A., McPhaden, M. J., Wu, L., England, M. H., Wang, G., Guilyardi, E., and Jin, F.-F. (2014). Increasing frequency of extreme El Niño events due to greenhouse warming. *Nature Climate Change*, 4:111–116.

Cai, W., Santoso, A., Wang, G., Yeh, S.-W., An, S.-I., Cobb, K. M., Collins, M., Guilyardi, E., Jin, F.-F., Kug, J.-S., Lengaigne, M., McPhaden, M. J., Takahashi, K., Timmermann, A., Vecchi, G., Watanabe, M., and Wu, L. (2015). ENSO and greenhouse warming. *Nature Climate Change*, 5:849–859.

Cai, Z., Li, R., and Zhang, Y. (2022). A distribution free conditional independence test with applications to causal discovery. *Journal of Machine Learning Research*, 23:1–41.

Capua, G. D., Runge, J., Donner, R. V., van den Hurk, B., Turner, A. G., Vellore, R., Krishnan, R., and Coumou, D. (2020). Dominant patterns of interaction between the tropics and mid-latitudes in boreal summer: causal relationships and the role of timescales. *Weather and Climate Dynamics*, 1:519–539.

CDS (2023). https://cds.climate.copernicus.eu/cdsapp#!/dataset/projections-cmip6?tab=overview. Copernicus Climate Data Store. Accessed: 2023-06-01.

Clem, K. R., Renwick, J. A., and McGregor, J. (2018). Autumn cooling of Western East Antarctica linked to the tropical pacific. *Journal of Geophysical Research: Atmospheres*, 123(1):89–107.

Cunningham, S. (2021). *Causal Inference.The Mixtape*. Yale Press.

Desmet, Q. and Ngo-Duc, T. (2021). A novel method for ranking CMIP6 global climate models over the Southeast Asian region. *International Journal of Climatology*, 42(1):97–117.

Durbin, J. and Koopman, S. J. (2012). *Time Series Analysis by State Space Methods*. Oxford Statistical science series. Oxford University Press, 3rd edition.

Ebert-Uphoff, I. and Deng, Y. (2012). Causal discovery for climate research using graphical models. *Journal of Climate*, 25(17):5648–5665.

Eichler, M. (2007). Granger causality and path diagrams for multivariate time series. *Journal of Econometrics*, 137:334–353.

Eichler, M. (2012). Graphical modelling of multivariate time series. *Probability Theory and Related Fields*, 153(1–2):233–268.

Foley, J. A., DeFries, R., Asner, G. P., Barford, C., Bonan, G., Carpenter, S. R., Chapin, F. S., Coe, M. T., Daily, G. C., Gibbs, H. K., Helkowski, J. H., Holloway, T., Howard, E. A., Kucharik, C. J., Monfreda, C., Patz, J. A., Prentice, I. C., Ramankutty, N., and Snyder, P. K. (2005). Global consequences of land use. *Science*, 309:570–574.

Gebbie, G., Eisenman, I., Wittenberg, A., and Tziperman, E. (2007). Modulation of westerly wind bursts by sea surface temperature: A semistochastic feedback for ENSO. *Journal of the Atmospheric Sciences*, 64(9):3281–3295.

Glymour, C., Zhang, K., and Spirtes, P. (2019). Review of causal discovery methods based on graphical models. *Frontiers in Genetics*, 10:1–15.

Guilyardi, E., Wittenberg, A., Fedorov, A., Collins, M., Wang, C., Capotondi, A., van Oldenborgh, G. J., and Stockdale, T. (2009). Understanding el niño in ocean–atmosphere general circulation models: Progress and challenges. *Bulletin of the American Meteorological Society*, 90(3):325–340.

Heede, U. K., Fedorov, A. V., and Burls, N. J. (2021). A stronger versus weaker walker: Understanding model differences in fast and slow tropical pacific responses to global warming. *Climate Dynamics*, 57(9–10):2505–2522.

Huang, B., Zhang, K., Lin, Y., Scholkopf, B., and Glymour, C. (2018). Generalized score functions for causal discovery. *Proceedings of the 24th International Conference on Knowledge Discovery Data Mining*, pages 1551–1560.

Imbens, G. W. and Rubin, D. B. (2015). *Causal Inference for Statistics, Social, and Biomedical Sciences*. Cambridge University Press.

IPCC (2022). *Summary for Policymakers*. Cambridge University Press. In Press.

Kendall, M. G. and Stuart, A. (1983). *The Advanced Theory of Statistics*. Griffin.

KNMI (2023). Climate Explorer. https://climexp.knmi.nl/start.cgi. Royal Netherlands Meteorological Institute.

Kretschmer, M., Coumou, D., Donges, J. F., and Runge, J. (2016). Using causal effect networks to analyze different arctic drivers of midlatitude winter circulation. *Journal of Climate*, 29:4069–4081.

Kretschmer, M., Runge, J., and Coumou, D. (2017). Early prediction of extreme stratospheric polar vortex states based on causal precursors. *Geophysical Research Letters*, 44(16):8592–8600.

Krich, C., Runge, J., Miralles, D. G., Migliavacca, M., Perez-Priego, O., El-Madany, T., Carrara, A., and Mahecha, M. D. (2020). Estimating causal networks in biosphere-atmosphere interaction with the PCMCI approach. *Biogeosciences*, 17:1033–1061.

Kumar, P., Kuttippurath, J., and Mitra, A. (2021). Causal discovery of drivers of surface ozone variability in Antarctica using a deep learning algorithm. *Environmental Science: Processes and Impacts*.

Latif, M. and Keenlyside, N. (2009). El Niño Southern Oscillation response to global warming. *Proceedings of the National Academy of Sciences of the United States of America*, 106.

Lehmann, J., Kretschmer, M., Schauberger, B., and Wechsung, F. (2020). Potential for early forecast of Moroccan wheat yields based on climatic drivers. *Geophysical Research Letters*, 47.

Malloy, K. and Kirtman, B. P. (2023). Subseasonal great plains rainfall via remote extratropical teleconnections: Regional application of theory-guided causal networks. *Journal of Geophysical Research: Atmospheres*, 128.

Marchant, R., Mumbi, C., Behera, S., and Yamagata, T. (2007). The indian ocean dipole. the unsung driver of climatic variability in East Africa. *African Journal of Ecology*, 45:4–16.

Morgan, S. L. and Winship, C. (2014). *Counterfactuals and Causal Inference. Methods and Principles for Social Research. Second Edition.*

Muñoz, E., Poveda, G., Arbeláez, M. P., and Vélez, I. D. (2020). Spatiotemporal dynamics of dengue in Colombia in relation to the combined effects of local climate and enso.

NOAA (2023). El Niño Southern Oscillation (ENSO).

Nowack, P., Runge, J., Eyring, V., and Haigh, J. D. (2020). Causal networks for climate model evaluation and constrained projections. *Nature Communications*, 11.

Pearl, J. (1993). Comment: Graphical models, causality and intervention. *Statistical Science*, 8.

Pearl, J. (1994). Causal diagrams for empirical research. *UCLA: Department of Statistics.*

Pearl, J. (2003). Statistics and causal inference: A review. 12:101–165.

Potter, K., Wilson, A., Bremer, P.-T., Williams, D., Doutriaux, C., Pascucci, V., and Johhson, C. (2009). Visualization of uncertainty and ensemble data: Exploration of climate modeling and weather forecast data with integrated visus-cdat systems. *Journal of Physics: Conference Series*, 180:012089.

Rao, S. and Sievers, G. (1995). A robust partial correlation measure. *Journal of Nonparametric Statistics*, 5:1–20.

Ropelewski, C. F. and Halpert, M. S. (1986). North american precipitation and temperature patterns associated with the El Niño Southern Oscillation (ENSO). *Monthly Weather Review*, 114:2352–2362.

Rubin, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology*, 66:688–701.

Runge, J. (2018). Causal network reconstruction from time series: From theoretical assumptions to practical estimation. *Chaos*, 28.

Runge, J., Bathiany, S., Bollt, E., Camps-Valls, G., Coumou, D., Deyle, E., Glymour, C., Kretschmer, M., Mahecha, M. D., Munoz-Marí, J., van Nes, E. H., Peters, J., Quax, R., Reichstein, M., Scheffer, M., Scholkopf, B., Spirtes, P., Sugihara, G., Sun, J., Zhang, K., and Zscheischler, J. (2019a). Inferring causation from time series in earth system sciences. *Nature Communications*, 10.

Runge, J., Gerhardus, A., Varando, G., Eyring, V., and Camps-Valls, G. (2023). Causal inference for time series. *Nature Reviews Earth & Environment*.

Runge, J., Nowack, P., Kretschmer, M., Flaxman, S., and Sejdinovic, D. (2019b). Detecting and quantifying causal associations in large nonlinear time series datasets. *Science Advances*, 5.

Runge, J., Petoukhov, V., and Kurths, J. (2014). Quantifying the strength and delay of climatic interactions: The ambiguities of cross correlation and a novel measure based on graphical models. *Journal of Climate*, 27(2):720–739.

Saggioro, E. and Shepherd, T. G. (2019). Quantifying the timescale and strength of southern hemisphere intraseasonal stratosphere-troposphere coupling. *Geophysical Research Letters*, 46:13479–13487.

Samarasinghe, S. M., Connolly, C., Barnes, E. A., Ebert-Uphoff, I., and Sun, L. (2021). Strengthened causal connections between the mjo and the north atlantic with climate warming. *Geophysical Research Letters*, 48.

Seager, R., Cane, M., Henderson, N., Lee, D.-E., Abernathey, R., and Zhang, H. (2019). Strengthening tropical pacific zonal sea surface temperature gradient consistent with rising greenhouse gases. *Nature Climate Change*, 9(7):517–522.

Spirtes, P., Glymour, C., and Scheines, R. (2001). Causation, prediction, and search.

Splawa-Neyman, J., Dabrowska, D. M., and Speed, T. P. (1990). On the application of probability theory to agricultural experiments. *Statistical Science*, pages 465–472.

Timmermann, A., Oberhuber, J., Bacher, A., Esch, M., Latif, M., and Roeckner, E. (1999). Increased El Niño frequency in a climate model forced by future greenhouse warming. *Nature*, 398:694–697.

Trenberth, K. E. and Hurrell, J. W. (1994). Decadal atmosphere-ocean variations in the pacific. *Climate Dynamics*, 9:303–319.

Vecchi, G. A. and Soden, B. J. (2007). Global warming and the weakening of the tropical circulation. *Journal of Climate*, 20(17):4316–4340.

Verma, A., Chandel, V., and Ghosh, S. (2022). Climate drivers of the variations of vegetation productivity in india. *Environmental Research Letters*, 17:084023.

Voortman, M. and Druzdzel, M. J. (2008). Insensitivity of constraint-based causal discovery algorithms to violations of the assumption of multivariate normality. *Proceedings of the Twenty-First International FLAIRS Conference*.

Waldorp, L. and Marsman, M. (2021). Relations between networks, regression, partial correlation, and the latent variable model. *Multivariate Behavioral Research*, 57(6):994–1006.

Wang, B., Wu, R., and Fu, X. (2000). Pacific–East Asian teleconnection: How does ENSO affect east asian climate? *Journal of Climate*, 13:1517–1536.

Wang, C. (2018). A review of ENSO theories. *National Science Review*, 5(6):813–825.

Wasserman, L. (2004). All of statistics. a concise course in statistical inference.

Zhao, X. and Allen, R. J. (2019). Strengthening of the walker circulation in recent decades and the role of natural sea surface temperature variability. *Environmental Research Communications*, 1(2):021003.

Zheng, F., Fang, X.-H., Zhu, J., Yu, J.-Y., and Li, X.-C. (2016). Modulation of Bjerknes feedback on the decadal variations in enso predictability. *Geophysical Research letters*, 43:12,560–12,568.