MASTER'S THESIS

Inference of Climate Tipping Interactions: From Dynamical Systems to Data-Driven Perspectives

24.05.2024

Author:	*David Leo Strahl*
Born:	*21.02.1999, Berlin*
Student-ID:	*816082*
First Examiner:	*Prof. Dr. Johan Rockström*
Second Examiner:	*Dr. Nico Wunderling*

Abstract

Sustained global warming risks to transgress the critical thresholds of multiple climate tipping elements, possibly driving the Earth system towards qualitatively different conditions. Next to being unprecedented in human history, tipping would have a substantially negative impact on planetary resilience, ecosystems, and human welfare. Despite growing recognition of the destabilizing interplay among tipping elements, details about their interactions remain unsettled. Therefore, this thesis explores whether causal inference methods could be used to reconstruct interactions between tipping elements. To do so, the causal discovery algorithm PCMCI+ is benchmarked on synthetic data generated by a dynamical model of interacting tipping elements (PyCascades). Results indicate that inference of tipping interactions from observational data is in principle possible, although challenging due to stationarity concerns and limited detected power in the presence of overlying feedback loops. As a proof of concept, the algorithm is applied to three paleoclimate proxies reflecting Dansgaard-Oeschger variability, large-scale abrupt transitions that have previously been attributed to frequently studied tipping elements (Greenland and West Antarctic Ice Sheets, Atlantic Meridional Overturning Circulation). Due to inherent data quality and representational issues, an application to paleoclimate proxies is however severely limited and no interactions beyond the bipolar seesaw, also evident through traditional correlation analysis, are found. Nevertheless, results highlight the potential of causal discovery algorithms for further analysis of interacting tipping elements. As a promising application for continued work, more recent observational data or Earth system models constrained by paleoclimate proxies are identified.

Contents

1. Introduction

1.1. Dangerous Climate Change

Since the onset of the Holocene around 12,000 years ago, Earth's climate has displayed an extraordinary level of stability.[1,2] Compared to pre-Holocene conditions, this period provided relatively constant temperatures, consistent precipitation patterns, and overall environmental conditions that were favorable to human evolution.[2,3] Modern society as we know was only possible due to this stability and has not experienced any conditions that altered significantly from it.

However, the advent of the industrial revolution marked a significant change. Human influences on the Earth system, the entirety of environmental processes, have now become the primary driver of change.[4,5] Rate and scale of the alterations made are unparalleled in the Holocene.[4] In just two centuries, human emissions have increased the global mean surface temperature by 1.48°C in 2023 compared to the 1800–1950 pre-industrial average.[6] It is almost certain that global warming will continue in the coming decades, likely surpassing the 1.5°C limit set in the Paris Agreement by the end of the century[7] and potentially overshooting 2.0°C.[8]

As a consequence, the Earth system is currently transitioning away from the stability of the Holocene, possibly towards a new state, the Anthropocene.[5,9] It is characterized by the predominance of human influence over natural processes that persist to stabilize the environmental conditions of the Holocene. However, if anthropogenic pressure ceases today, the Earth system would eventually revert towards the pre-industrial conditions of the Holocene,[5] although residual warming after cessation of emissions may be nonzero.[10,11] This reversion may take thousands of years due to the scale of alterations already made. For example, changes in the energy balance are already likely sufficient to offset the coming ice age cycle by at least 100 kya.[12] It remains an open question whether continued disturbances could be sufficient to disrupt these natural stabilization mechanisms, and propel the Earth system irrevocably towards unprecedented conditions.[5] As a result, the Anthropocene could become fundamentally different from what modern society has adapted to.

Climate change does and will continue to pose significant environmental, social and

economic impacts.[13] This includes, but is not limited to, glacier loss up to widespread deglaciation in mid-latitudes by 2100,[14] increased number of extreme heat events and droughts,[13] amplification of precipitation extremes[15] or accelerated loss of biodiversity.[16] All of these impacts are deeply interconnected with planetary resilience[3] and human welfare. To address them, a thorough understanding of the mechanisms that drive climate change and its implications is required.[13] The Anthropocene however poses a profound challenge for mitigation and adaptation strategies in the long term. We are interested in predicting a system that has been tampered with, based on observations from a past where alterations were absent.

1.2. Tipping Elements in the Earth's Climate System

In 2007, the Intergovernmental Panel on Climate Change (IPCC) acknowledged the possibility of abrupt changes within the Earth system at rates that are substantially faster than seen in recent history.[17] In the subsequent reports, this concept of tipping as a non-smooth response to global warming gained prominence and is now understood to endanger the stability of the Holocene.

Tipping elements are Earth subsystems that show threshold behavior in response to small-scale perturbations, which leads to abrupt and large-scale changes in their properties.[18,19] Even under perturbations, they remain in a stable state until a critical threshold, the tipping point, is crossed. This is because a negative feedback mechanisms acts stabilizing and suppresses perturbations. However, once perturbations become sufficiently large, a self-perpetuating positive feedback mechanism takes over and fundamentally changes the dynamics of the system, driving it towards a qualitatively different state.[1,19] This process often shows hysteresis behavior, where abrupt changes are irreversible in the sense that back-tipping does not simply happen in reverse on human timescales and can often only be enabled by dramatically different initial conditions.[1] Such perturbations that bring a system towards its tipping point can come in many forms,[20,21] including natural variability, abrupt shocks or relatively gradual changes in environmental parameters like the global mean temperature.

Evidence of such tipping elements can be found in Earth's history,[22,23] as well as in our process understanding of their complex dynamics from Earth system models of varying complexity.[19,24] An example for a tipping event would be the shutdown of a major ocean

current like the Atlantic Meridional Overturning Circulation (AMOC) due to perturbations of the salt-advection feedback.[25,26] A dilution of cold and salty waters in the North Atlantic decreases their density and inhibits downwelling that is needed to sustain the circulation. Another example is the Greenland Ice Sheet (GIS), that may tip due to the melt-elevation feedback.[27,28] Sufficient ice loss lowers the top of the ice-sheet and exposes it to temperatures above the melting point that then again accelerate melting. Other tipping systems can be found in the biosphere, where for example the Amazon rainforest[29] or tropical coral reefs[30] are in danger of being rapidly lost.

Recent evidence suggests that multiple low-threshold systems transgress a 50% risk of being tipped under the 2.0°C target, endangering the stability of the Earth system at global warming levels that could be reached in the near future.[1,31,32] Estimated thresholds for the GIS (0.8 – 3°C global mean warming above pre-industrial), the West-Antarctic Ice Sheet (1.0 – 3.0°C) or tropical coral reefs (1.0 – 2.0°C) may lie well within the possible warming range by the end of this century.[8] A tipping point in the AMOC (1.4 – 8.0°C) is less certain but may become relevant under sustained warming.[1]

1.3. Interactions Between Climate Tipping Elements

However, tipping elements are not isolated but rather interconnected entities.[22,33] The state of one tipping element affects another, for example through atmospheric or oceanic processes that couple their dynamics.[1] Such a tipping interaction would be the intrusion of GIS meltwater into the North Atlantic which prevents deep-water formation and could potentially lead to a partial shutdown of the AMOC when the salt-advection feedback can not be sustained.[22,34] This mechanism is established in model simulations and recent research suggests that it may already be ongoing.[35–37] The resulting weakening of the AMOC would itself reduce the energy transport between hemispheres, raising southern temperatures and affecting the stability of the Amazon rainforest and the West-Antarctic Ice Sheet.[22,38] However, tipping interactions can also stabilize a system: a weaker AMOC results in less energy transport to the North Atlantic and thus lower temperatures that stabilize the GIS.[39]

Hence, the state of one tipping element influences others, and can be stabilizing or destabilizing. Overall, model simulations show that the interactions of tipping elements that risk to tip under sustained global warning tend to be destabilizing.[22,40] While the

physical mechanisms that enable these interactions are better understood, details about their relevance and relative strength are less well understood.[22]

A sequential destabilization of tipping elements has been termed a tipping cascade.[40] Although tipping of one element primarily affects a specific region, interactions between elements can amplify their effect to have global-scale implications.[19] Additionally, coupled tipping systems exhibit much more complex threshold behavior that complicates their analysis.[5] Not all mediators of cascades need to be tipping elements themselves:[22] Arctic Sea Ice has not been found to be prone to abrupt transitions,[41] but its absence would warm North Atlantic waters and potentially reduce AMOC strength.[22]

Tipping cascades are hypothesized to enable global tipping of the Earth system towards a hothouse state.[5] While it remains speculative whether such a state exists, cascading transitions become more likely with global warming. Further research into destabilizing tipping interactions is therefore necessary improve understanding of the dangers posed to Earth system stability.

1.4. Interactions Between Tipping Elements in Paleoclimate Data

Research into tipping interactions faces an epistemological challenge. Currently, the conditions of the Holocene are still stabilized and tipping has so far not been observed, although recent evidence indicates that it may be already underway for some tipping elements.

To address this issue, paleoclimate data could provide us with the best analog of observing tipping in the climate system. Multiple instances of abrupt transitions are seen in paleoclimate records for the Quaternary and are frequently attributed to interactions between tipping elements.[23,24] Next to model simulations, such reconstructions may be only real observations from which tipping interactions can be inferred.[31]

Over the last 115k years, against the backdrop of the last glacial maximum, the Earth system quasi-periodically switched from cold glacial states to relatively warm interglacials.[24,42] Abrupt warming of the Northern Hemisphere by 5°C within one century was followed by prolonged phases of cooling that lasted thousands of years.[38,42] The cause of these so-called Dansgaard-Oeschger (DO) events is disputed but has been attributed to shifts between stable states of the AMOC,[43,44] which could have been enabled by decreases

in Arctic Sea Ice cover.[45]

Similarly to DO events, Heinrich events (H) also show abrupt warming across the Northern Hemisphere. They are however accompanied by a high amount of ice rafted debris, e.g. rocky material that is carried into the ocean by icebergs and dropped once they melt.[38] They are equally attributed to changes in AMOC strength, which in this case are due to meltwater input from large breakups of the Laurentide ice sheet across North America.[42,46]

1.5. Scope of This Thesis

This thesis addresses whether novel causal discovery methods[47,48] (PCMCI+) can be used to infer interactions between climate tipping elements from observational, and specifically paleoclimate data. In the context of climate sciences, causal discovery has hitherto been used to study atmospheric teleconnections,[49–52] as a tool for model evaluation[53,54] and in loosely related fields like climate-economics.[55] An analysis of the available literature suggests that no application to interacting tipping elements in general or paleoclimate proxies in specific has so far been published.

Firstly, synthetic data generated by a conceptual model of interacting tipping elements (`PyCascades`) is used to verify that the causal discovery method is applicable to tipping interactions. Section 3.1 takes a statistical physics perspective and presents some insights about `PyCascades` that substantiate this analysis. In section 3.2 the capability of PCMCI at reconstructing interaction networks is systematically benchmarked for increasingly complex setups to identify potential methodical limitations and violations of the algorithms assumptions.

Secondly, in section 3.3 an exemplary application to three paleoclimate proxies reflecting Dansgaard-Oeschger variability is presented as a proof of concept. These proxies have previously been attributed to frequently studies tipping elements and are understood to represent the states of the GIS, AMOC and WAIS elements respectively.[56] This allows an assessment of the benefits and limitations of causal inference algorithms within the context of paleoclimate data.

2. Methods

2.1. PyCascades: Conceptual Tipping Models

Although abrupt transitions have been observed in some complex Earth system models (EMS),[57] they often do not properly represent the nonlinear behavior of potential tipping elements.[58] This is the case because some tipping elements are simply not regarded in the model or are prescribed, i.e. not dynamically simulated. For example, ESMs that contributed to the IPCC AR6 report used prescribed ice sheets, making dynamic cryosphere tipping impossible[59] and several of them lack a dynamical biosphere.[60]

Nonetheless, the physical mechanisms that couple tipping elements are in principle represented in these ESMs. However, when the possibility of tipping is excluded by design, any interactions relevant to the stability of tipping elements can not be resolved. Even when it is included and can be resolved, e.g. in some coupled ocean-cryosphere models,[61] simulations are often expensive to compute, making large-scale ensemble studies difficult to undertake.

Due to computational constraints and the lack of important tipping elements or their feedback mechanisms in several ESMs, additional approaches are required. The upcoming tipping point model intercomparison project (TIPMIP) aims to address these issues and fill critical knowledge gaps. However, it entails high turnaround times and first results can only be expected in 2026.[62]

One approach to resolve this issue is `PyCascades`, a software package designed to build computationally efficient conceptual models of interacting tipping elements.[40,63,64] At its core, it employs a network approach and represents tipping elements using coupled nonlinear differential equations that allow for bifurcations, i.e. qualitative changes of state. Each equation describes one tipping element that can exist in two stable states: a baseline state and a transitioned one. The dynamics are conceptual in the sense that they model the process of tipping and not the physical mechanisms that enable it. Transitions between these states are governed by critical parameters that can be related to environmental factors, such as the global mean surface temperature, coupling between tipping elements and stochasticity. In this way, the susceptibility of these tipping elements to different drivers is reflected.

2.1.1. Fundamental Equations

`PyCascades` describes interacting tipping elements using a system of ordinary coupled differential equations. This leverages the assumption that the dynamics of a tipping element are separable into self-dynamics that enable transitions between states and interactions with other tipping elements.

Each tipping element is represented by a conceptual coordinate that reflects its state. We treat the element as being tipped when the coordinate is larger than zero, and stable otherwise. Given N tipping elements with coordinates $x_i, x_2, \cdots, x_N$ the implementation used in this thesis models them by integrating the equation

$$f(x_i) = \tau_i \frac{\mathrm{d}x_i}{\mathrm{d}t} = \underbrace{-ax_i^3 + bx_i + c_i}_{\text{self-dynamics}} + \underbrace{\sum_{j \neq i} d_{ij}(x_j + 1)}_{\text{interactions}} \tag{2.1}$$

The first three terms stem from the bifurcation model that enables abrupt transitions between states.[63] Here, we use the widely-used fold bifurcation that is implemented using a third-order polynomial, with standard parameters $a = -1$ and $b = 1$. The critical parameter c controls the number of states and is analogous to an external forcing due to environmental conditions. The parameter τ_i represents the characteristic timescale of the tipping element x_i.

To better understand how multistability arises from the equation, we focus solely on the self-dynamics. Furthermore, we assume that there is one uncoupled tipping element $x_0 = x$. Equation (2.1) then simplifies to the third-order polynomial $f(x) = -x^3 + x + c$. The integrated solution is steady, when the state coincides with any non-trivial zero of the polynomial with negative derivative.

As shown in Fig. 2.1, the exact value and number of stable states depends on c, where bistability occurs in a limited range.[64] One can easily derive the limits for which the number of stable states changes.[65] Given $c = 0$, the system has two stable states at

$$\hat{x} = \pm\sqrt{\frac{b}{a}} = \pm 1 \tag{2.2}$$

One of them vanishes when the critical parameter is sufficient to shift both of the polynomial's extrema above or below zero.[65] In the standard form, this is the case for

$$c^* = \pm f(\hat{x}) = \pm 2\sqrt{\frac{1}{a}\left(\frac{b}{3}\right)^3} = \pm\sqrt{\frac{4}{27}} \tag{2.3}$$

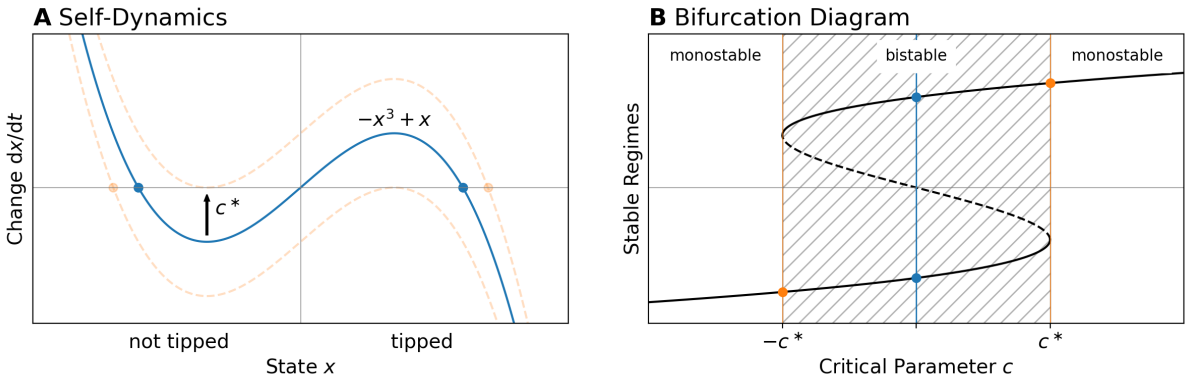


Figure 2.1. | **Fold Bifurcation**. **A** Self-dynamics of a conceptualized tipping element described by a third-order polynomial with (orange) and without (blue) forcing. At the critical value $c = \pm c^*$, the number of roots and thus stable states changes. **B** Bifurcation diagram showing for which values of c the system exhibits bistability.

The sum in equation (2.1) represents interactions between pairs of tipping elements. In this implementation, the state $(x_j + 1)$ is linearly coupled to x_i with coupling strength d_{ij}. The additional sum is included because we treat negative values as stable and want to limit interactions to processes driven by tipping.

Environmental forcing (c) and coupling (d_{ij}) together control what state the system can be found in. Generally, real-world tipping elements can of course have more than two stable states at any time, can be spatially resolved and may not interact linearly. Moisture recycling in the Amazon rainforest can for example be understood as a spatially resolved tipping element and other ecosystems are equally known to exhibit spatial patterns in state transitions.[66]

2.1.2. Stochasticity

Equation (2.1) models tipping elements to be entirely deterministic. However, they really exhibit natural variability and are subject to random perturbations.[20] The complex dynamics of the Earth system inherently result in chaotic behavior that in our model manifests as random fluctuations of the state variable. Daansgard-Oeschger or Heinrich events (section 1.4) may be examples for these noise-induced transitions,[67] although this is still debated[44]

To accommodate for this variability, we can extend the deterministic tipping equation

into a stochastic differential equation (SDE) as follows:[40]

$$\mathrm{d}x_i = f(x_i)\,\mathrm{d}t + \sigma\,\mathrm{d}W_i \tag{2.4}$$

where $\mathrm{d}W$ are Wiener increments with variance $\mathrm{d}t$. The total variance of the noise term is given by $\mathbb{V}[\sigma\,\mathrm{d}W] = \sigma^2\,\mathrm{d}t$.

For the purposes of this thesis, we employ noise sampled from a Gaussian distribution. However, the `PyCascades` framework also supports Lévy and Cauchy distributions that are particularly relevant for representing tipping processes influenced by extreme events.[63]

2.1.3. Earth System Tipping Network

A network of interacting tipping elements that has previously been examined using `PyCascades` includes the Atlantic Meridional Overturning Circulation (AMOC), Greenland Ice Shield (GIS) and West Antarctic Ice Sheet (WAIS).[63] These three tipping elements are especially relevant in the context of this thesis, as their interactions have previously been linked to Dansgaard-Oeschger variability that can be seen paleoclimate proxies.

With three tipping elements, the network can have up to six interactions that are described in Wunderling.[22] Kriegler presented an expert elicitation that tries to constrain their relative strength and sign.[33] These estimates, whose values have been rescaled, are shown in Table 2.1.

2.2. Tigramite: Causal Discovery for Time Series

The principal objective of this thesis is to assess whether interactions between tipping elements can be inferred from observational data using causal discovery algorithms. Unraveling these mechanisms is complicated due to the non-linear, chaotic, and multistable nature of the relationships involved.[23,68] However, a vast array of data describing the system is now available and can be employed for analysis. This includes paleoclimate reconstructions, direct observations from remote sensing technologies, and model simulations of varying complexities.[23,68]

Inference of these interactions however represents a fundamental epistemological challenge. Pearl has shown that from observational data alone, only statements about association can be made and one rather needs to resort to interventional studies to assess causal

Table 2.1. | Earth System Interactions Rescaled range of coupling coefficients d_{ij} based on an expert elicitation by KRIEGLER.[33] Values taken from WUNDERLING.[22] The right column gives the physical mechanism which enables the coupling.

Interaction	Coupling d_{ij}		Physical Process
	from	to	
AMOC → GIS	-1	0	Reduction in northward heat transport
AMOC → WAIS	0	0.15	Heat accumulation in Southern Ocean
GIS → AMOC	0	1	Freshwater influx
GIS → WAIS	0	1	Sea-level rise
WAIS → AMOC	-0.3	0.3	Increase in meridional salinity gradient (-) and advection of freshwater into North Atlantic (+)
WAIS → GIS	0	0.2	Sea level rise

relationships.[69] It is however neither feasible nor justifiable to dump sufficient freshwater into the North Atlantic to validate a potential tipping point. Practical limitations, as well as ethical and social considerations, preclude any interventional experiments on a global or regional scale. As a result, establishing causal relationships from observational data alone becomes a significant challenge[70]

Causal discovery algorithms may bridge the gap between association and causality.[47,70] Here, we use the `tigramite` framework for causal discovery in timeseries, due to its applicability to datasets featuring both linear and non-linear, lagged dependencies, good scalability, high detection power and controlled false positive rates.[71] Causal discovery algorithms from `tigramite` can be used to learn association networks from time series data alone. Furthermore, under specific, well-considered assumptions, these association networks can be interpreted as representing causal processes.

Other methods for causal inference are available, but are limited to specific applications. For instance, the established framework of Granger causality is build on the idea of using a variables history to predict another ones future.[72] To do so, one constructs a linear autoregressive model on the shared history of all variables to predict a target variable. Causation is said to exist if presumed causes appear to be statistically significant for

predicting the target. The linear premise of Granger causality and the computational demands for high-dimensional autoregressive models however typically limit its application to bivariate analyses.[70] It should be noted that the term *Granger causality* is somewhat misleading as it primarily addresses predictability rather than causality per se.[73]

2.2.1. Structural Causal Models

Prior to delving into `tigramite`[47] algorithms for causal discovery, we need to establish the theoretical foundation of structural causal models (SCMs). In this discussion, we predominantly adhere to the notation introduced by RUNGE to facilitate comparability.

We consider a time-dependent system $\mathbf{X}_t = (X_t^1, X_t^2, \cdots, X_t^N)$ comprising N variables. The relationships between the variables are modeled using the following assignments:[69,71]

$$X_t^i \coloneqq f^i(\mathcal{P}[X_t^i], \eta_t^i) \tag{2.5}$$

Here, the causal assignment operator $\coloneqq$ represents a directional, cause-and-effect relationship and should not be misunderstood as an equivalence or mathematical equation that can be inverted. The functions f^i are arbitrary (possibly nonlinear and time dependent) and η_t^i symbolizes mutually independent and dynamic noise terms. The term $\mathcal{P}[X_t^i]$ refers to the causal parents of X^i at time t, encompassing all variables that have a causal effect on X_t^i. Because cause always precedes effect, any causal parent of a variable can only come from the common past, $P[X^i] \subset \mathbf{x}^-$.

SCMs are aptly visualized as directed acyclic graphs (DAGs).[69,71] Each node in the graph corresponds to one variable in the model. The noise terms, though not directly represented, are implicitly included in this structure. A directed link $X^i \to X^j$ is included if X^i is in the set of causal parents $\mathcal{P}[X^j]$. An illustration of a SCM and its corresponding DAG is presented in Fig. 2.2.

Runge extends causal graphs to encapsulate temporal relationships, coining them *time-series graphs*.[47] In this framework, the value of a variable at each time is treated as a separate variable. This effectively discretizes time and in the case of linear f^i coincides with vector autoregressive models. A link $X_{t-2}^i \to X_t^j$ implies that the state of X^i at time $t-2$ influences X^j at time t, establishing a time-lagged relationship. In many scenarios it is assumed, that such a lagged link is consistent across all-time points. The time-series graph can then be simplified by omitting explicit time ordering and indicating the lags

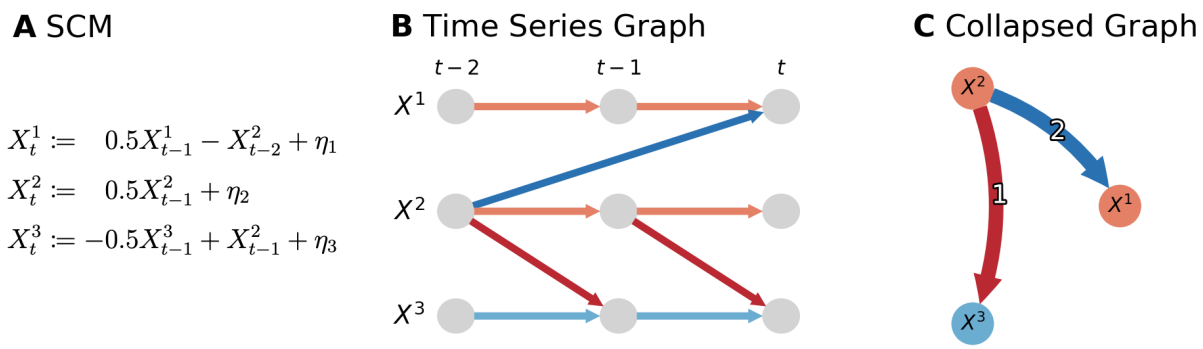


Figure 2.2. | Different ways to represent a structural causal model. **A** As a causal assignment equation. **B** As a time series graph. **C** As a collapsed graph with link lags. Node and link colors represent sign and magnitude of the relationships.

alongside the links, as demonstrated in Fig.2.2

2.2.2. Conditional Independence and Causality

A causal relationships $A \to B$ in a causal graphs implies a correlation between the variables A and B.[69] While such a correlations can aid in unveiling causal relationships, they also present practical challenges. This is, because they can also arise for combinations of variables that are not causally linked. Therefore, the existence of a correlation does not generally imply causation. The converse is however true: the absence of a correlation strongly indicates the lack of a causal relationship.[69] Consequently, relying solely on correlation analysis can become misleading.

This is, because correlations are a necessary but not sufficient condition for causality. A correlation can emerge from a common cause $A \leftarrow B \to C$ affecting two variables, termed a spurious correlation. In such instances, controlling for the common cause by conditioning on it, i.e. fixing its value, can eliminate this misleading correlation, i.e. $A \perp\!\!\!\perp C \mid B$. Conversely, a collider $A \to B \leftarrow C$, where one variable is influenced by two others, does not lead to a correlation between the drivers by default. A spurious correlation arises when one conditions on the central variable, i.e. $A \not\perp\!\!\!\perp C \mid B$, as then one driver explains away the effect of the second.

The challenge of assessing the true causal effect between two variables X^i and X^j therefore lies in identifying a conditioning set that effectively removes all spurious correlations

between them, without introducing new ones. Causal theory has shown that the joint set of their causal parents $\mathcal{P}[X^i] \cup \mathcal{P}[X^j]$ is sufficient for this purpose.[69] Of course, the causal parents are not known a priori. How this set can be nonetheless found, is described in the following section, but to do so, it is necessary to robustly detect conditional independence between variables.

In scenarios with linear relationships, i.e. where the f_i in equation (2.5) are linear functions, and the X^i have Gaussian marginal distributions, testing for independence can be as straightforward as calculating the Pearson correlation coefficient and its statistical significance.[47] This concept is extended to conditional independence by regressing on the conditioning set, i.e. implementing a linear model to remove its effect. Within `tigramite`, this test is called `ParCorr`. Then distributions are not Gaussian, transformations can be applied beforehand to ensure that variables follow a normal distribution. This is done by the conditional independence test `RobustParCorr`.[47]

However, for nonlinear relationships, establishing conditional independence is more challenging. Tests based on Gaussian Process Regression are available, where the model used for conditioning is nonlinear and the conditional independence is measured as the correlation between model residuals.[47] However, this only works for additive dependencies and not all kinds of nonlinearities. More generally, permutation-based estimates of mutual information are used, albeit being less efficient to compute.[74] The corresponding test in `tigramite` is called `CMIknn` (conditional mutual information based on k-nearest neighbours).

Selection of an appropriate conditional independence test is a tradeoff between capturing complexity while avoiding the pitfalls of overfitting. Linear measures, while straightforward, are incapable of detecting nonlinear interactions. Conversely, employing mutual information to test linear relationships may not offer the same statistical robustness as a direct application of the Pearson correlation would.[74] This dilemma highlights the importance of ensuring that the chosen test adequately reflects the underlying dynamics without producing false correlations or overlooking significant causal links.

2.2.3. Assumptions

Learning causal graphs from observational data necessitates several assumptions,[71] some of which are illustrated in Fig. 2.3. With these assumptions in place, the graphical causal

model framework ensures that the inferred links can be interpreted as representing causal relationships.[69] However, in practical scenarios, these assumptions are often not fully met.[47] Despite this, causal discovery algorithms can still effectively identify relationships between variables with a high degree of confidence in their causal interpretation. While some of these assumptions can be relaxed, doing so leads to modifications of the causal discovery algorithms and that are not relevant here. They are:

1. *Causal Markov Condition*: Separation in the causal graph implies independencies in the data. In other words, if two variables are dependent, they must be adjacent in the graph. This simply assumes that we can use the SCM framework to model the interactions between variables.

2. *Faithfulness*: This assumption is the inverse of the Causal Markov Condition. It postulates that data independencies imply separation in the causal graph. The combination of these two conditions establishes a link between structural causal models and causal graphs. Faithfulness may be violated in scenarios in which independencies arise although variables are causally linked, such as when the effects of two causal links exactly cancel.

3. *Causal Sufficiency*: This assumption asserts the absence of unobserved confounders in the graph. Fulfilling this assumption is challenging, as it requires prior knowledge of the very causal structure we aim to discover. Typically, it is assumed that any confounding is minimal enough not to significantly distort the graph's causal structure. A violation of the Causal Sufficiency assumption is shown in Fig. 2.3A.

4. *Causal Stationarity*: The causal mechanism is consistent over time. This suggests that the time series causal graph is translation symmetric, with consistent lagged links between variables at all times. This assumption is needed to increase the sample size for conditional independence tests, as we can interpret all lagged pairs of points to come from the same distribution. A violation of the Causal Stationarity assumption is shown in Fig. 2.3B.

5. *Contemporaneous Acyclicity*: The graph is free from instantaneous feedback loops. Similar to parametric independence, unresolved feedback loops could result in the

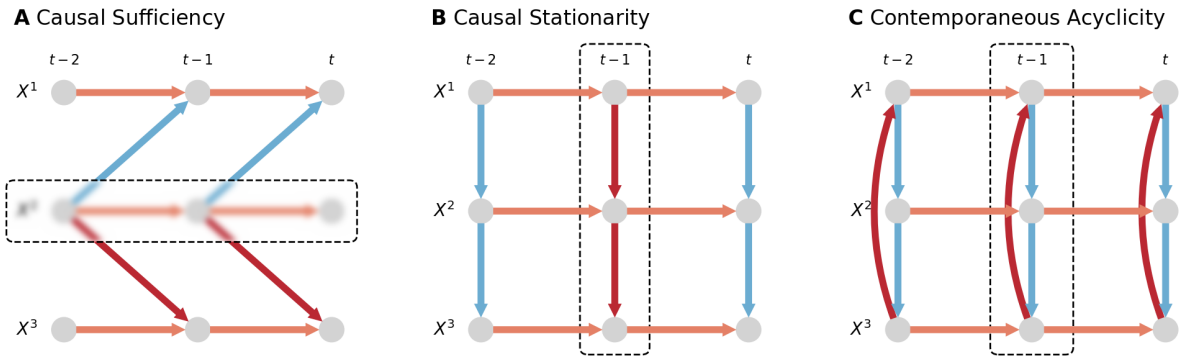


Figure 2.3. | **Violations of Some Assumptions A** Due to the unobserved confounder X^2, a spurious correlation between the other variables would be detected. The set of observed variables is not sufficient for causal inference. **B** Some links change sign at $t-1$. The causal mechanisms is not stationary and therefore proper detection is not possible. **C** A lag zero feedback loop violates the assumption of contemporaneous acyclicity. These links can not be detected by the algorithm.

negation of causal effects. A violation of the Contemporaneous Acyclicity assumption is shown in Fig. 2.3C.

2.2.4. PCMCI(+) Algorithm

The PCMCI algorithm leverages the above assumptions to reconstruct the causal network for the dynamical system described by the SCM in equation (2.5)[70] by finding the optimal set of causal parents for each variable. Let N be the number of variables and $\tau_{\max}$ be the highest lag of all relationships involved.

Causal theory has shown that for each two variables the set of their combined causal parents is sufficient to establish conditional independence between them.[69] However, in the case of time series data, an approach that simply tests all possible combinations of conditioning sets is unpractical, due to the problem's high dimensionality ($N\tau_{\max} - 1$) and reduced effect sizes in the independence tests because of autocorrelation.[75]

PCMCI overcomes this problem by implementing two complementary stages. The PC_1 stage is designed to find a set of potential causal parents for each variable that includes the true parents with high probability. It starts with a fully connected time series graph and iteratively removes links using independence tests with conditioning sets of increasing cardinality. In Fig.2.4A, the PC_1 stage tries to determine the candidate causal parents of

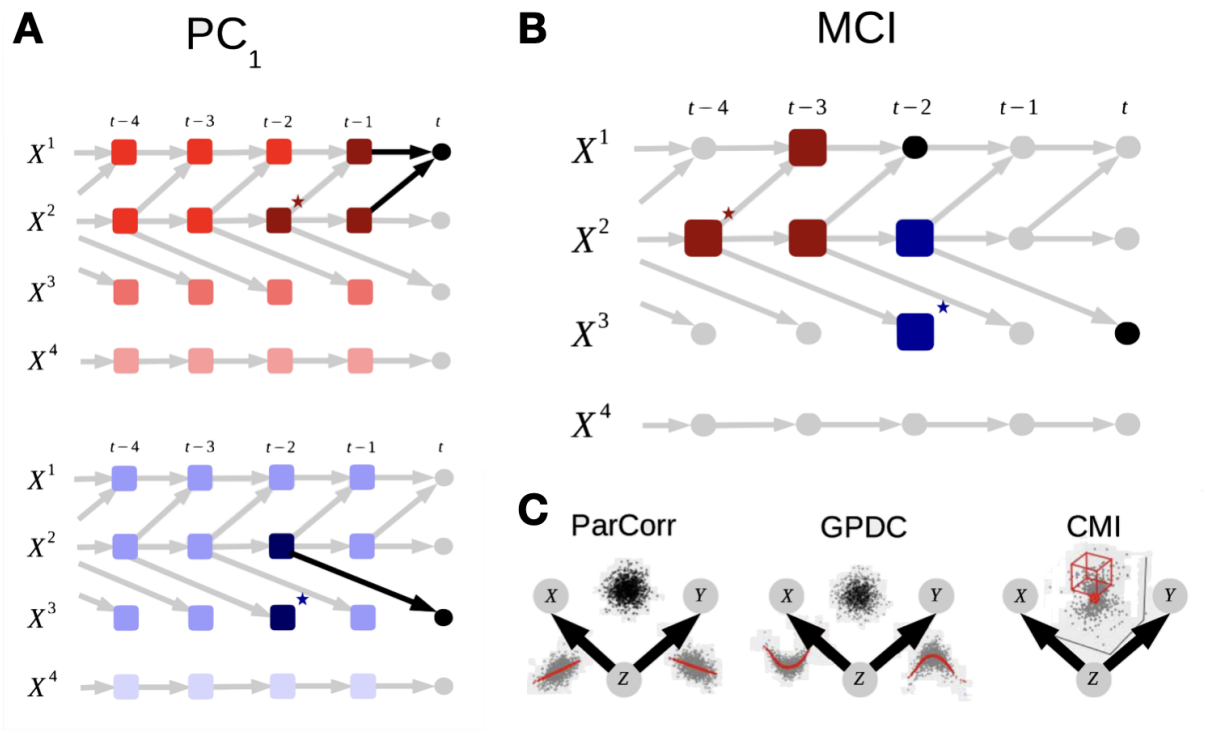


Figure 2.4. | **PCMCI Algorithm**. To robustly detect the true causal parents, the algorithm is split into a preliminary PC_1 stage (**A**) that finds the likely parents and a complementary MCI stage (**B**) that controls for false positives. The figure is taken from RUNGE[70] and slightly adapted. Different conditional independence tests are shown in **C**. Consult the section on PCMCI(+) for further description of the algorithm.

X_t^1 and X_t^3 (black circles). The true causal graph is shown in light gray.

Starting with an empty separation set $\mathcal{S} = \emptyset$, each link $X_{t-\tau}^i \to X_t^j$ for which the conditional independence test $X_{t-\tau}^i \perp\!\!\!\perp X_t^j \mid \mathcal{S}$ succeeds with significance p, is removed. In the example, we test all variables from Fig.2.4A that are marked with colored boxes and reject X^4 (lowest opacity) in the first iteration. For the remaining links, the new parents $\mathcal{P}[X_t^j] \setminus \{X_{t-\tau}^i\}$ are sorted based on their absolute test statistic and the largest s are included in the next iteration set $\mathcal{S}$. In this way, s is iteratively increased and the algorithm converges when no further conditions can be tested. In Fig. 2.4A, the algorithm iteratively removes variables until only the ones marked by the darkest colors remain. Some false positives, which are marked with an asteriks are not properly removed.

The resulting graph now includes the potential causal parents for each variable, although

others variables can be included in this set. The subsequent *momentary conditional information* (MCI) stage is designed to remove these false positives. To correct the true causal effect of the link $X^i_{t-\tau} \to X^j_t$, one tests

$$X^i_{t-\tau} \perp\!\!\!\perp X^j_t \mid \mathcal{P}[X^i_{t-\tau}] \cup \mathcal{P}[X^j_t] \setminus \{X^i_{t-\tau}\} \tag{2.6}$$

and removes the corresponding link when the test succeeds. The restriction of the conditioning set to the potential causal parents found by the PC_1 stage addresses the autocorrelation problem and increases detection power of the algorithm. In Fig. 2.4B, we condition on the potential causal parents of both variables found in the PC_1 stage to assess whether a link $X^1_{t-2} \to X^3_t$ should be included in the graph.

PCMCI is not designed to detect contemporaneous links and suffers from high false-positive rates when the data is strongly autocorrelated. The PCMCI+ algorithm tries to overcome these issues by optimizing the choice of conditioning sets S. Compared to PCMCI, it has higher detection power and controls false-positives better.[48] Furthermore, PCMCI+ is able to resolve contemporaneous links and feedback loops, as long as contemporaneous acyclicity is fulfilled.[48]

2.3. Graph Similarity Measures

In cases in which we know the true causal graph of a mechanism, either due to expert knowledge or because the data is generated by a model, we want to compare the output of a causal discovery algorithm to this ground truth. To measure similarity between two graphs, we use the accuracy metrics *true-positive rate* (TPR) and *false-negative rate* (FPR).

Firstly, we need to convert both graphs into adjacency tensors. Given a graph $\mathcal{G}$ with nodes $X^1, \ldots, X^n$ the adjacency tensor is defined as

$$A_{ij\tau}(\mathcal{G}) = \begin{cases} 1 & X^i_{t-\tau} \to X^j_t \text{ in } \mathcal{G} \\ 0 & \text{otherwise} \end{cases} \tag{2.7}$$

When we want to focus solely on the adjacency structure and not the exact time lags, we can drop the time dimension by looking whether we see an adjacency $A_{ij\tau}$ for any τ. Similarly, when the causal direction is not so important, we can symmetrize the adjacency tensor so that $A_{ij} = A_{ji}$.

To define the TPR and FPR, we build a 2×2 confusion matrix for the adjacency tensors of the true and predicted graphs. Links that are present in the predicted graph either count to the true-positives (TP) or false-positives (FP) depending on whether they are present in the reference graph. Contrary, the absence of a link in the predicted graph can either be a false-negative (FN) or true-positive (FP).

The TPR (sensitivity) is now defined as the ratio of reference links that are correctly identified in the predicted graph, e.g.

$$\mathrm{TPR} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}} \tag{2.8}$$

It can be treated as a measure of detection power of the causal inference algorithm.

The FPR (1 - specificity) is defined as the ratio of predicted links that are not present in the reference graph, e.g.

$$\mathrm{FPR} = \frac{\mathrm{FP}}{\mathrm{FP} + \mathrm{TN}} \tag{2.9}$$

It can be treated as a measure of susceptibility to errors. Note, that both TPR and FPR depend on which graph is treated as the "true" reference and which as the prediction; they are not symmetric measures.

3. Results

In the following section, we will give some theoretical insights into the conceptual `PyCascades` model that explain why causal analysis is possible in the context of tipping interactions. In section 3.2, we will use synthetic data to systematically benchmark the capabilities of PCMCI(+) at reconstructing these interactions from timeseries for increasingly complex networks of interacting tipping elements. With this analysis, we want to address whether causal discovery algorithms can in principle be applied in the context of tipping and assert what practical limitations arise. In the final section 3.3, we will apply PCMCI+ to three paleoclimate proxies from the late Quaternary and attempt to infer the strength of their interactions. These proxies show abrupt transitions that have previously been attributed to frequently studied tipping elements like the Greenland and West Antarctic Ice Sheets (GIS and WAIS) or the Atlantic Meridional Overturning Circulation (AMOC).

3.1. A Statistical Physics Perspective on Interacting Tipping Elements

In this section, we adopt a statistical physics perspective to gain further insights into the dynamical system underlying `PyCascades`. The following results substantiate the causal discovery approach adapted in later sections and add explain why causal analysis is at all possible in the context of tipping.

Firstly, we will show how the stationary tipping equation (2.1) is equivalent to an Ornstein-Uhlenbeck process for a single tipping element. Secondly, we reformulate the problem in terms of a state probability density that can be solved using the Fokker-Planck equation. This allows us to gain insights into how couplings between tipping elements affect the stability landscape, i.e. the number and locations of stable states the system can be found in.

3.1.1. Equivalence to Ornstein-Uhlenbeck Processes

We approximate equation (2.1) around a jointly steady solution, in which all tipping elements are in one state and their coordinates have expected values $\mathbb{E}[x_i] = \hat{x}_i$. Without abrupt transitions between states, variations around this equilibrium are assumed to be small. We decompose the coupling term into one dependent only on the expectation value and another one dependent on the deviation from it. Upon inserting $x_i = \hat{x}_i + (x_i - \hat{x}_i)$ into the SDE (2.4) we get

$$\mathrm{d}x_i = \underbrace{\left(-x_i^3 + x_i + c + \sum_{j \neq i} d_{ji}(\hat{x}_j + 1)\right)}_{g(x_i)} \mathrm{d}t + \mathrm{d}t \sum_{j \neq i} d_{ji}(x_j - \hat{x}_j) + \sigma \, \mathrm{d}W_i \tag{3.1}$$

where we have dropped the timescales τ_i for simplicity. The time-independent coupling term in $g(x_i)$ can be absorbed into the critical parameter c describing the self-mechanics. It can be understood as a joint potential that affects all interacting tipping elements and determines where the stable states can be found.

We can now linearize the term $g(x_i)$ around the steady state $\hat{x}_i$ up to first order to get

$$g(x_i) \approx g(\hat{x}_i) + \left.\frac{\mathrm{d}g}{\mathrm{d}x_i}\right|_{\hat{x}_i} (x_i - \hat{x}_i) + \cdots \approx \beta_i (x_i - \hat{x}_i) \tag{3.2}$$

where, based on the definition of the steady state, $g(\hat{x}_i) = 0$ and β_i is the slope of the polynomial (that is generally negative).

Finally, we can fully express the linearized SDE only in terms of the anomaly:

$$\mathrm{d}x_i = \beta_i (x_i - \hat{x}_i) \, \mathrm{d}t + \mathrm{d}t \sum_{j \neq i} d_{ji}(x_j - \hat{x}_j) + \sigma \, \mathrm{d}W_i \tag{3.3}$$

In equilibrium, a `PyCascades` system is thus fully equivalent to coupled Ornstein-Uhlbeck processes. In the absence of coupling, an analytical solution can be found for each tipping element using Itô's isometry[76]

$$x_i(t) = \hat{x}_i + \sigma \int_0^t \mathrm{d}W_i' \, e^{\beta_i (t' - t)} \tag{3.4}$$

Naturally, the mean value is $\mathbb{E}[x_i] = \hat{x}_i$. In the limit of infinite simulation length, the variance over all possible noise realisations asymptotically approaches

$$\mathbb{V}[x_i] = \frac{\sigma^2}{2\beta_i}\left(1 - e^{-2\beta_i t}\right) \tag{3.5}$$

3.1.2. Reformulation using State Probability Density

Following Livina,[77] we now introduce a state probability density $P(x,t)$ describing the probability of observing a `PyCascades` tipping element in the state x at any given time. We will show that this probability density obeys the Fokker-Planck equation (FPE).

Note, that the evolution of a single tipping element modeled by a fold-bifurcation is analogous to the dynamics of a one-dimensional particle in a double potential well. We introduce the potential $\phi(x) = x^4/4 - x^2/2 - cx$ whose associated gradient force is equal to the fold bifurcation used in the tipping equation (2.1)

$$f(x) = -\frac{\partial}{\partial x}\phi(x) = -x^3 + x + c \tag{3.6}$$

Without coupling, the stochastic tipping equation (2.4) is then equivalent to the Langevin equation of a massless ($m = 0$) particle in the potential $\phi(x)$ affected by uncorrelated noise terms η:

$$m\frac{\mathrm{d}^2x}{\mathrm{d}t^2} + \frac{\mathrm{d}x}{\mathrm{d}t} = -\frac{\partial}{\partial x}\phi(x) + \eta(t) \tag{3.7}$$

The FPE associated with this Langevin equation describes the evolution of the state probability density $P(x,t)$. The drift term is given by the gradient force of the potential and the diffusion constant depends on the variance of the Wiener increments, $D = \sigma^2/2$. The FPE is:

$$\frac{\partial P}{\partial t} = \frac{\partial}{\partial x}\left(\frac{\partial \phi}{\partial x}\,P\right) + \frac{\sigma^2}{2}\frac{\partial^2 P}{\partial x^2} \tag{3.8}$$

We are only interested in the steady-state probability density and thus set $\partial_t P = 0$. We can integrate the equation twice to find the general solution:

$$P(x) = Ae^{-\phi/D} \tag{3.9}$$

where the constant A needs to be determined to normalize the integral of $P(x)$. This equation can be inverted to reconstruct the potential and thus states the system can be in.[77]

Fig. 3.1 shows the potential ϕ, the resulting probability density $P(x)$ and an empirical histogram for the state variable simulated in `PyCascades`. Only the analytical solution to the FPE properly reflects the true shape of its probability distribution. The linearized approach that resulted in the Ornstein-Uhlenbeck equations underestimates the likelihood of finding the tipping element in an intermediate state between the two potential wells.

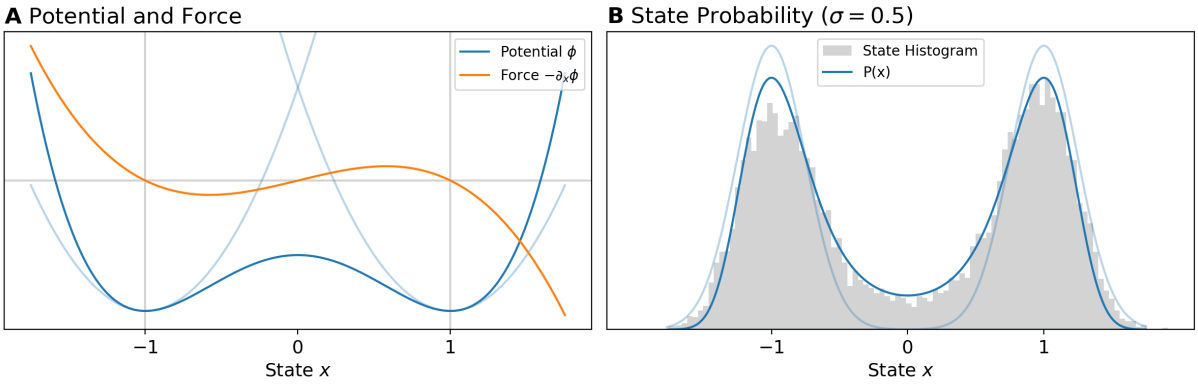


Figure 3.1. | State Probability Distribution A: Potential ϕ as well as its gradient force that leads to the fold-bifurcation tipping equation. Faint curves are quadratic approximations to the potential which result in the Ornstein-Uhlenbeck equations. **B:** Histogram of the state variable for a single tipping element with noise level $\sigma = 0.5$. The probability density $P(x)$ is derived from the Fokker-Planck equation. The faint curves show the probability density of the approximations. Note, how the true probability is skewed towards the center.

Note, that for low noise values, the histogram only approaches the analytic curve in the limit of very long timeseries. This is because the residence time within one potential well is antiproportional to the noise level. In the absence of external forcing, e.g. $c = 0$, tipping may only happen due to noise induced perturbations that move the state across the potential barrier.

3.1.3. Numerical Simulations

In the case of two or more coupled tipping elements, the FPE associated with the setup can no longer be solved analytically. To see why this is the case, we will again consider a binary feedback system. For ease of notation, we call the state coordinates x, y. Without external forcing, i.e. $c = 0$, they obey the tipping equations

$$\frac{\mathrm{d}x}{\mathrm{d}t} = -x^3 + x + d_{xy}(y+1) + \sigma\frac{\mathrm{d}W_x}{\mathrm{d}t} \tag{3.10}$$

$$\frac{\mathrm{d}y}{\mathrm{d}t} = -y^3 + y + d_{yx}(x+1) + \sigma\frac{\mathrm{d}W_y}{\mathrm{d}t} \tag{3.11}$$

where d_{xy} is the coupling from y to x and vice versa. As described above, the straightforward approach would be to introduce a potential $\phi(x, y)$ and then use equation (3.9) as the steady-state solution.

However, this is no longer possible for all combinations of d_{xy} and d_{yx}. To understand why, we can interpret the FPE as an advection-diffusion equation for the probability density. Its velocity field is described by the above equations, and has two terms: (1) the fold bifurcation for which a potential can always be introduced and (2) coupling interactions where a potential can only be introduced when $d_{xy} = d_{yx}$ or at least one the couplings vanishes. For other values, the velocity field picks up a rotational component and the equation must be solved numerically. This can be seen from cross differentiating equations (3.10) and (3.11):

$$\frac{\partial}{\partial x}\frac{\mathrm{d}y}{\mathrm{d}t} - \frac{\partial}{\partial y}\frac{\mathrm{d}x}{\mathrm{d}t} = d_{yx} - d_{xy} \stackrel{!}{=} 0 \tag{3.12}$$

We can however integrate the FPE numerically to learn how interactions determine the stability landscape, i.e. the states in which the tipping elements can be found in. We do so by evolving the FPE from initial conditions until the probability density estimate P (PDE) converges. Although this approach is computationally much more exhaustive than simply solving the tipping equation for many different initial conditions and parameters but guarantees that unlikely or non-attractive states are found, given that they can be resolved by the grid used for numerical integration.

For our analysis of this system of two interacting tipping elements, we simulated the corresponding FPE for roughly 2500 coupling coefficients $d_{ij} \in [-2.5, 2.5]$ and $\sigma = 0.2$. To ensure uniform density in parameter space, we employed latin-hypercube sampling. The numerical integration was done using the Python package `py-pde`[78] with a cartesian grid of extent $[-2.5, 2.5]$ and 150 cells in each dimension. Because the PDE needs to be normalized, we enforced a vanishing partial derivative in all directions at the perimeter of the grid (Dirichlet boundary conditions). Grid extent and resolution were selected after verifying that the PDE was well contained. To find all possible states, we integrated from uniform initial conditions. The integration was terminated, when (1) a simulation time of 20 model time units was reached, (2) a runtime of two minutes was reached or (3) the PDE converged.

Because every tipping element can be in at most two discrete states (not tipped vs. tipped), we have an upper bound on the number of joint states, i.e. 2^n with n being the number of tipping elements. As can be seen from the equivalence to the Ornstein-Uhlenbeck process, each of the states maps to one quadrant in state space, with the exact number and location determined by the zeros of $g(x_i)$. As before, we treat negative

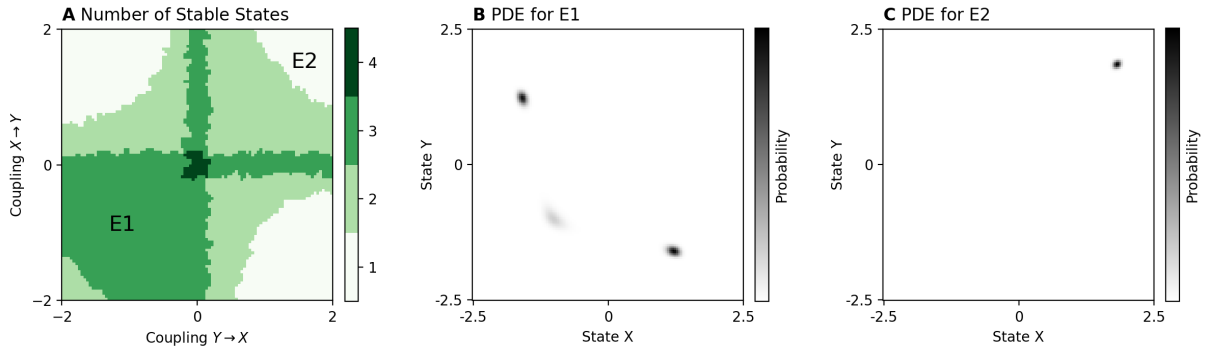


Figure 3.2. | **State Number for a Binary Feedback System. A** shows the number of steady-states found by the quadrant integration approach for a binary feedback system. **B** and **C** show exemplary probability density estimates (PDE) for the marked locations in coupling space. Due to the low noise level of $\sigma = 0.1$, the states are well-separated.

coordinates as indicative of stability and positive ones as being tipped. Assuming that the states are well separated by the quadrant boundaries, we integrated the PDE for each quadrant to find the probability of being in the corresponding state. Due to the finite integration time, quadrant probabilities are never truly zero and we therefore thresholded them by $p = 0.05$. Highly unlikely states may thus be excluded from our analysis. This quadrant integration approach could also be generalized to multiple dimensions.

Alternatively, and to verify the quadrant probabilities, we used a contour detection algorithm to count the number of states. For this, we again thresholded the PDE to get a binary image, now with $p = 0.05 P_{\max}$ where $P_{\max}$ is the highest probability found across all states. Using the python package `opencv`[79] we found contours in the PDE that are mapped to the corresponding states based on their average location in phase space. This approach also works when the states overlap with quadrant boundaries but fails when they begin to merge due to blurring at higher noise levels.

For the following analysis we excluded runs for which the solver terminated prematurely or for which the contour detection algorithm found more than four states. This happened occasionally when the numerical integration became unstable, which resulted in artifacts and rippling close to the grid perimeter that was wrongly identified as a contour.

3.1.4. State Probabilities of the Binary Feedback System

Fig. 3.2A shows the number of stable states a system of two interacting tipping elements can be found in, as counted by the quadrant integration approach. Additionally, some exemplary PDE for different coupling configurations are shown in Fig. 3.2B-C. Fig. 3.3 gives the probabilities for each state, as well as the most probable one. We can broadly characterize the following regions:

1. For destabilizing feedbacks with negative links (lower left in Fig. 3.2A and Fig. 3.3), we observe three stable states in which the system of interacting tipping elements can be in. The states in which one element is tipped and the other one remains stable (green and orange, depending on the direction of the coupling) are always possible. Furthermore, a co-stable state (blue) is seen, although its probability gradually declines for higher absolute coupling values.

2. For destabilizing feedbacks with positive links (upper right in Fig. 3.3), two states are found when coupling is weak. Again, the co-stable state (blue) vanishes for higher coupling values while the state in which both elements are tipped (red) is consistently the most probable.

3. Stabilizing feedback loops show a similar behavior regardless which element tips (upper left and lower right in Fig. 3.3). For low coupling values, either one of the elements can be found in the tipped state (green, orange) while the other one is stable. It is the element from which the positive link emerges that is free to tip. A co-stable state can be observed for higher couplings (blue) and there is a clear boundary to the region in which one element can tip.

4. Close to the origin or near quadrant boundaries ($d_{xy} = 0$ and/or $d_{yx} = 0$), coupling is weak. Because of this, at least one of the tipping elements is unconstrained and can be found in both states. Therefore, the number of states increases. Close to the origin, all four state combinations can be observed.

Both state counting strategies (quadrant and contour) give very similar results for low noise levels, i.e. when states are well separated. For higher noise levels, the PDE became increasingly blurred and the contour detection algorithm failed. Nonetheless, the quadrant approach yields consistent state numbers for $\sigma = 0.1$ and $\sigma = 0.5$ (not shown here).

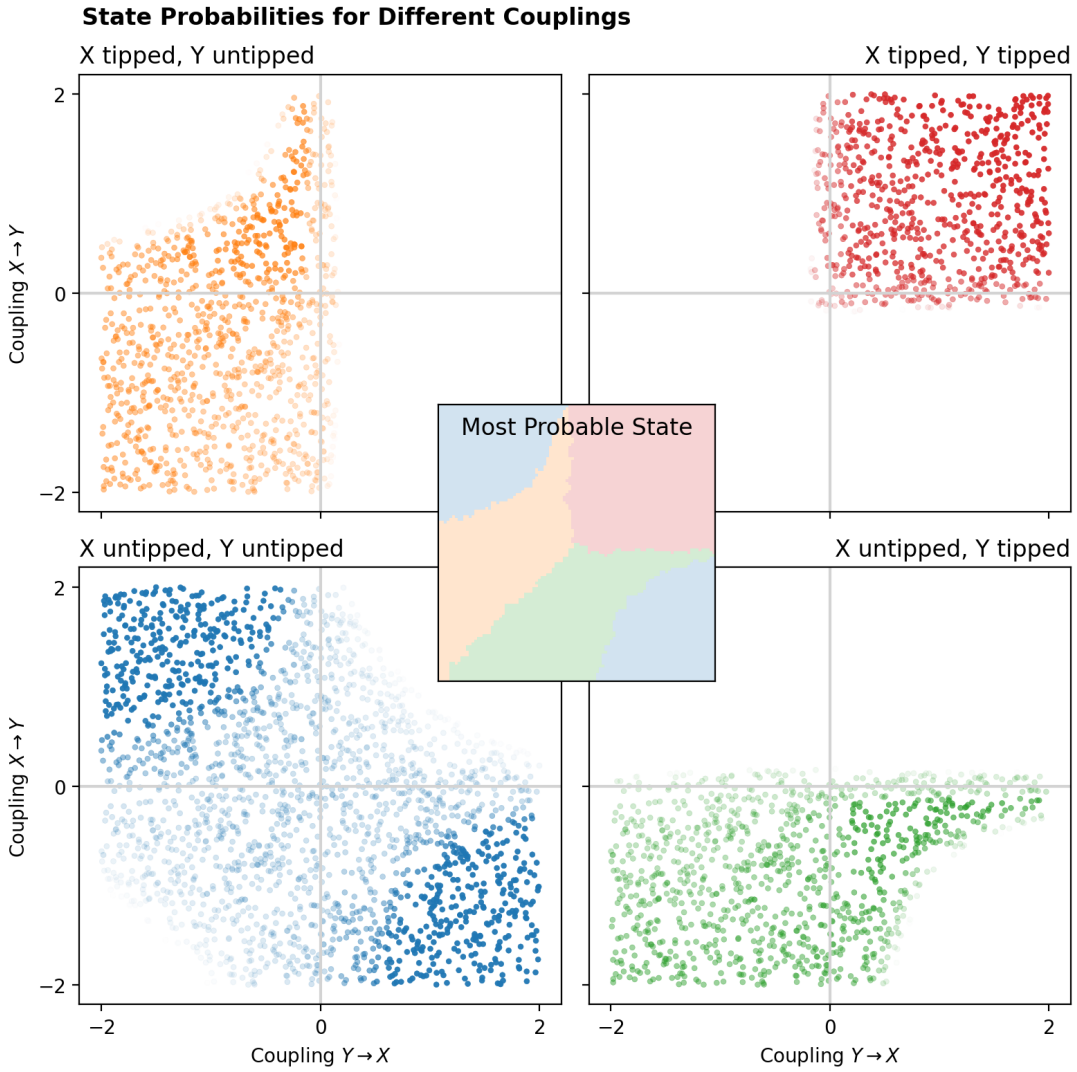


Figure 3.3. | **State Probabilities and Most Probable State**. For a binary feedback system, we can find at most four discrete states. Each subplot corresponds to one of those and their location in the plot as well as color matches the quadrant the state is found in. Every dot is one simulation run and the probability of seeing a state is encoded in the opacity. Evidently, the states in which the system can be found changes with the coupling values and some states are only possible for constrained regions. The **inset** shows the most probable state for each coupling combination, with axis coordinates matching the other subplots.

From these results, we can not infer whether the number of states is discrete as Fig. 3.2 would imply or if they emerge due to the thresholding process and state probabilities really change gradually. It may simply be the case, that a state probability drops below $p = 0.05$ at the boundaries. Regardless, the transition regions seems to follow isocontours of d_{ij}/d_{ji}.

To summarize, results show that the coupling configuration determines the number, location and probability of states in a conceptual tipping model. As an approach to infer interactions between tipping elements, we could reverse this analysis and use the state-space stability landscape to gain insights into the coupling configuration that enabled them. However, this approach is likely not feasible for observational data. Observations are constrained to a limited set of states and a full characterisation of the stability landscape is not possible. However, based on the equivalence to Ornstein-Uhlenbeck processes, all information about interactions in the conceptual model is already contained in the variability around a single state, even when no transitions are observed. Causal discovery algorithms like PCMCI(+) could now leverage this information to reconstruct the tipping networks from timeseries alone, with incomplete knowledge about the complete state-space.

3.2. Causal Discovery for Interacting Tipping Elements

This section explores under which conditions PCMCI(+) can be applied to infer tipping interactions from timeseries data. To do so, it benchmarks the algorithms ability to reconstruct the coupling network in a conceptual model of interacting tipping elements. Furthermore, it described how hyperparameters and conditional independence tests need to be tuned to optimize results and whether any methodical adjustments need to be made to ensure applicability in this context. In section 3.2.1, we describe how the presence of abrupt transitions poses a challenge for conditional independence tests and present a workaround to resolve this issue. In section 3.2.2, we continue to analyse the algorithm for increasingly complex interaction networks. Finally, section 3.2.3 describes how asymmetric feedbacks present an additional challenge for PCMCI(+).

3.2.1. Limitations of Conditional Independence Tests in the Presence of Abrupt Transitions

Before we can systematically benchmark PCMCI(+) on synthetic data, we anticipate a result that is relevant to all later applications: Abrupt transitions present a profound challenge for causal discovery in timeseries, as conditional independence tests struggle to identify causal links when the data exhibits multistability. A simple example is helpful to show why this is the case, and at the same time illustrates how we generated the synthetic data.

We use PyCascades to simulate an unforced ($c = 0$) system of three interacting tipping elements (X^1, X^2, X^3) for various configurations of coupling coefficients d_{ij} (see section 2.1.1). We interpret the coupling matrix from equation (2.1) as the adjacency matrix for the true causal graph, meaning that each entry is treated as a link with lag one. Self links are excluded. From the adjacency matrix, we generate a reference graph to which we can compare the inferred causal mechanism. Note, that both the inferred and reference graph may include links up to the hyperparameter $\tau_{\max}$. It tunes the highest lag that the algorithm should search for (see section 2.2.4). For simulated data, we know that this values is $\tau_{\max} = 1$.

Fig. 3.4A shows simulated timeseries for three interacting tipping elements X^1, X^2 and X^3 with coupling coefficients $d_{12} = 1, d_{21} = -1, d_{31} = -1$ and $d_{32} = -1$. The

negative feedback loop between X^1 and X^2 suppresses tipping in the second element, while the prognostic variable X^3 synchronously tips in-line with X^1. We infer causal graphs from this timeseries using PCMCI(+) and the `RobustParCorr` independence test (see section 2.2.2), with hyperparameters $\tau_{\max} = 5$ and significance level $p = 0.05$. An exemplary slice through state space, on which the conditional independence test acts, as well as the true and inferred causal graphs are also shown in Fig. 3.4B-C.

To illustrate this problem, we note that a false-positive link is found by the PCMCI algorithm between the variables X^3 and X^2 with lag of $\tau = 5$. This can be seen in the second graph in Fig. 3.4C. Tipping of the individual elements allows multistability that is reflected in the marginal distributions of the variables, which poses a challenge for asserting conditional independence. In principle, PCMCI(+) is designed to remove this effect by conditioning on a sufficient set of variables from the common past of X^3 and X^2. However, the algorithm apparently fails to reliably do so and a linear conditional independence test would wrongly assert that the variables are linked (see Fig. 3.4B).

As can be seen in Fig.3.4A, way to mitigate this is masking of the piecewise stationary sections in the timeseries. This means that the causal discovery algorithm is only run on sections for which the system is in one state. Technically, the implementation used in `tigramite`[47] restricts both causal drivers and targets to lie within the mask, while conditioning variables are unrestricted. This leads to a separation of the individual states in state-space and removes any correlation that is due to their shape and locations. For the $X^3_{t-5} \to X^2_t$ link, the marginal distributions of the masked state become uncorrelated (red points in 3.4B) and the link would be removed from the inferred graph. Evidently, this reduces the number of false-positives for both PCMCI and PCMCI+ in Fig.3.4C.

Of course, the inferred graph for a single simulation with `PyCascades` is partly random due to the stochastic nature of the tipping equation (2.4). Fig. 3.5 presents a systematic assessment of the differences between masked and unrestricted usage of the timeseries. All coupling coefficients d_{ij} in this exemplary setup were linearly scaled by a coefficient between $d = 0$ and $d = 5$. For each scaling we inferred a causal graph using both masked and unrestricted timeseries and aggregated the true-positive rate (TPR) and false-positive rate (FPR) over 20 simulations to get robust estimates of their expected values and their confidence intervals at $p = 0.95$ significance.

When comparing PCMCI and PCMCI+ we observe overall improved detection power

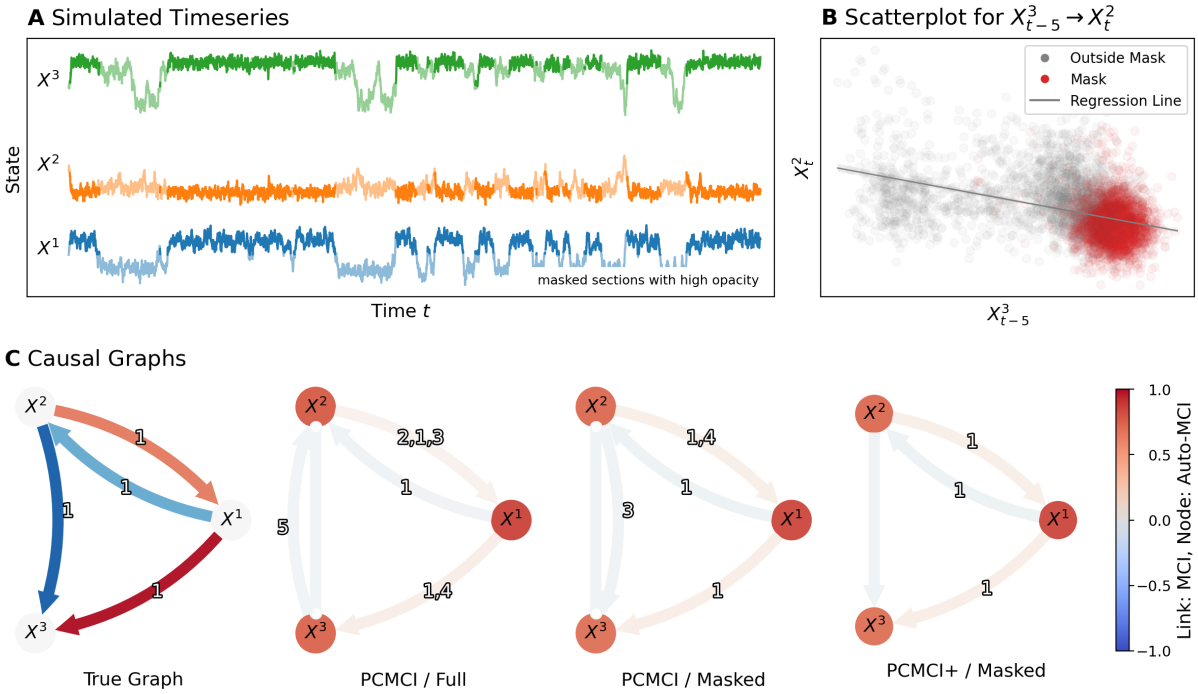
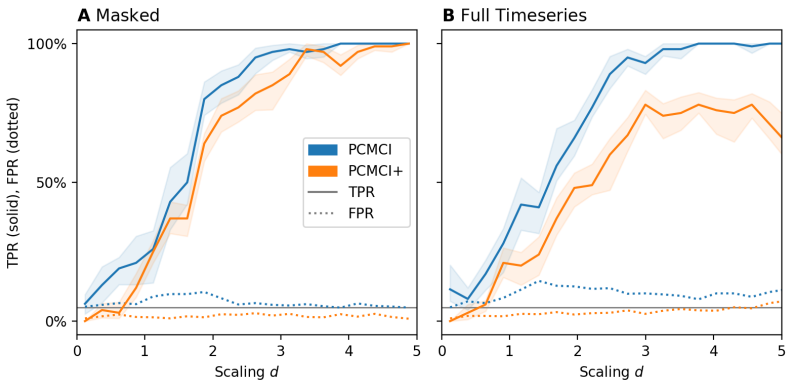


Figure 3.4. | Issues with Causal Discovery for Tipping Systems A: Exemplary timeseries for a system of three tipping elements (X^1, X^2, X^3) whose interactions include one negative feedback loop. Sections in which both X^1 and X^3 are tipped are masked. **B:** PCMCI+ tries to determine their interactions by performing conditional independence tests in state-space. Due to tipping in X^3, a spurious correlation is observed for $X^3_{t-1} \to X^2_t$ (gray line) when the full timeseries is used. A dependency is thus wrongly detected. For the masked section (red dots), the variables are properly uncorrelated. **C** shows the true causal graph for this system of interacting tipping elements (left), as well as discovered graphs using different inference setups. Best results are seen when the PCMCI+ algorithm is run on masked data.

Figure 3.5 | Detection Power and Masking The coupling configuration in Fig. 3.4 is linearly scaled between $d = 0$ and $d = 5$. For each scaling, expected values and confidence intervals are estimated for the TPR and FPR. **A** shows the results for masked sections and **B** when the full timeseries is used.



when the masked timeseries is used. As shown in Fig. 3.5A, PCMCI exhibits a slightly increased TPR and performs better at detecting existing links. However, it comes with the disadvantage of an increased FPR due to autocorrelation of the individual tipping elements. At higher coupling coefficients, both algorithms reach near perfect TPR when masking is used.

These differences are exacerbated when the full timeseries is used. Although TPR is generally lower and FPR higher, especially PCMCI struggles with false-positive control. PCMCI+ fails to reach perfect TPR even for high scaling values, corresponding to stronger coupling. Note that a slight increase in FPR may result in many erroneous links in the causal graph, as it is often sparse and the number of false-positives scales with the number of absent links.

3.2.2. Detection Power for Different Tipping Networks

After having established that masking is necessary to improve detection power, we continued to systematically benchmark PCMCI(+) for more complex tipping setups. We again simulated interacting tipping elements using `PyCascades`. Because tipping in reality is expected to happen infrequently and this simplifies the masking process, we chose a noise level of $\sigma = 0.05$ that made transitions infrequent.

We again ran both causal discovery algorithms with the above described hyperparameters and conditional independence tests. TPR and FPR are again aggregated over 20 simulated timeseries and confidence intervals are assessed using bootstrap estimates with significance $p = 0.95$.

Firstly, we simulated three systems of interacting tipping elements with increasingly complex coupling configurations, where the last one including multiple overlying feedback loops. For these systems, all coupling coefficients were set to one, so that the links have equal relative strength and feedbacks are symmetric. Again, the coupling configuration was scaled between $d = 0$ and $d = 5$.

Secondly, we simulated an Earth system like tipping network as described in section 2.1.3. It includes three tipping elements with an interaction network resembling Wunderling[63] and represents the Greenland Ice Sheet (GIS), the Atlantic Meridional Overturning Circulation (AMOC) and the West-Antarctic Ice Sheet (WAIS). Realistic values of the coupling coefficients d_{ij} for this setup have been described in Tab. 2.1 and

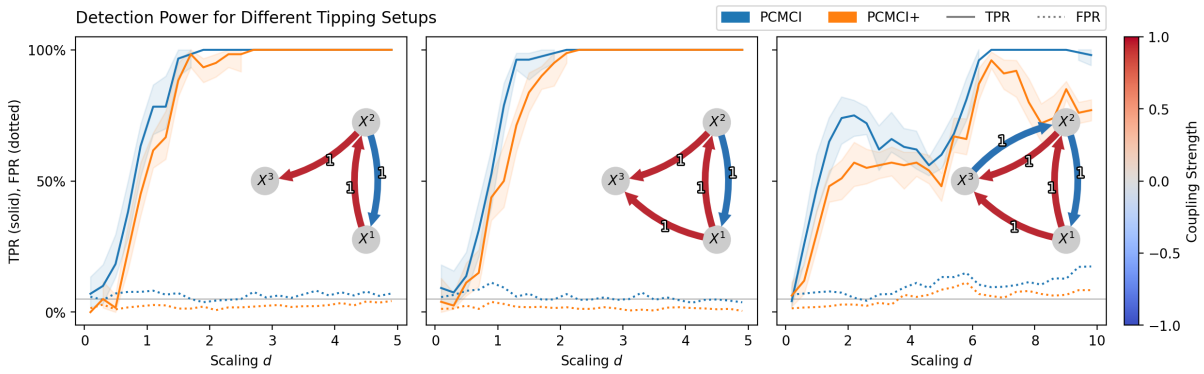


Figure 3.6. | **Detection Power for Different Tipping Setups** Comparison of the true-positive rate (TPR) and false-positive rate (FPR) for inference with PCMCI(+) for three networks of interacting tipping elements. The insets show the true causal mechanisms that were used to simulate the data using `PyCascades`. Curves are aggregated over multiple realisations of the same system, with confidence intervals from bootstrapping at significance level $p = 0.05$. PCMCI cosistently has a slightly better TPR than PCMCI+, but at the same time comes with higher FPR. With sufficient noise and for graphs that contain one feedback loop, both algorithms reach perfect prediction in terms of TPR. For multiple overlaying loops, higher couplings are necessary to reach similar detection power.

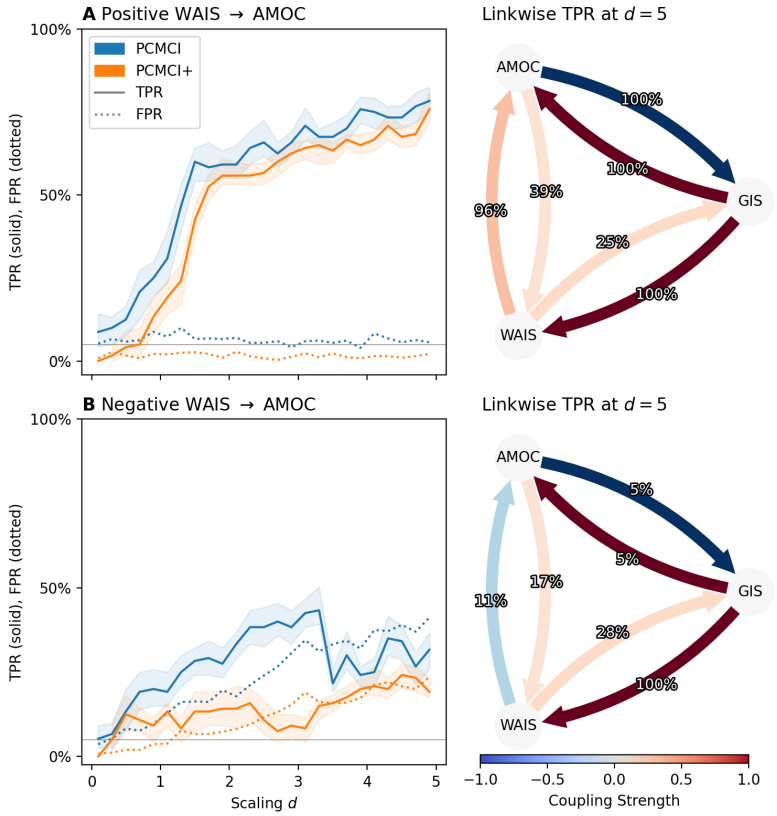


Figure 3.7 | **Detection Power for an Earth System Network** *Left column:* comparison of the true-positive rate (TPR) and false-positive rate (FPR). **A** shows results when WAIS → AMOC is positive and **B** when the link is negative. *Right column:* True causal graphs used to simulate the data. Percentages indicate the link-specific TPR when all coupling values are scaled by $d = 5$.

are based on an expert elicitation by KRIEGLER.[33] Here, we used the upper bound for a conservative estimate, although lower values are implicitly included due to the scaling. We differentiated two setups based on the unclear sign of the WAIS → AMOC. Two competing processes together determine the sign of this interaction that remains unclear:[33,63] tipping of the WAIS increases the meridional salinity gradient which stabilizes the AMOC, but at the same time increases advection of freshwater into the Atlantic that has a destabilizing effect.

Fig. 3.6 shows results for the exemplary setups of interacting tipping elements. Again, PCMCI performs slightly better in terms of TPR, while having a higher FPR. A single feedback mechanism can be resolved regardless of the exact configurations, as long as no second feedback loop is introduced.

However, detection power dramatically decreases for multiple, possibly overlying feedback loops. Note, that it is impossible to have two independent feedback systems with three elements. In Fig. 3.6C, coupling values up to the scaling $d = 5$ result in a TPR below 75% for PCMCI and below 50% for PCMCI+. Detection power for the latter is thus similar to an algorithm that simply assigns links by flipping a coin. PCMCI reaches near perfect TPR at even higher scalings of $d = 10$, although the FPR rate increases significantly.

Fig. 3.7 shows results for the GIS, AMOC and WAIS tipping network, for which the coupling coefficients were set to the upper limits of the estimates described in Tab. 2.1. The system has a total of five overlaying feedback loops (three for each pair of variables and two that include all three). Because these are additional challenges for causal discovery, the detection power is reduced more strongly. With a positive sign for the WAIS → AMOC coupling, both algorithms reach around 75% TPR at a scaling of $d = 5$. Already below $d < 1.5$, above the upper estimate of the realistic coefficients, TPR drops below 50%. Changing the sign of the WAIS → AMOC link results in a consistent reduction of TPR below 50% and increased FPR across both algorithms, as can be seen in Fig. 3.7B.

Additionally, we estimate the TPR individually for each link individually (at $d = 5$ for PCMCI+). The results are shown as the percentages next to each link in the causal graphs in Fig. 3.7. For the case of a positive WAIS → AMOC link, proper detection is seen for all links with $|d_{ij}| > 0.5$, while weaker links are missed more often than not. Apparently, PCMCI(+) struggles to resolve asymmetric feedback systems. For the network where the

WAIS → AMOC link is negative, only the GIS → WAIS link is properly detected. In this case, the stronger links that were reliably detected when the WAIS → AMOC coupling was positive, have now drastically reduced TPR.

Note, that not the whole range of scaling values is realistic. An authentic Earth System probably has coupling values that translates to values below $d_{ij} < 1$.[33,40] For these values, neither PCMCI nor PCMCI+ are able to reliably detect and resolve the interactions between tipping elements.

Furthermore, we assumed that all tipping elements evolve on the same timescales $\tau = 1$. However, oceanic and especially the cryosphere elements evolve on much longer timescales.[40] For the conceptual model, changing the timescales is however equivalent to rescaling the coupling to lower values. This can be seen from equation (2.1), where the interaction only depends on d_{ij}/τ_i and the rescaled bifurcation polynomial only affects the location and number of steady states of the system of interacting tipping elements.

3.2.3. Asymmetric Feedbacks

Finally, we further investigate the applicability of PCMCI+ to asymmetric feedbacks. As was seen in the previous section, this is relevant to more realistic interactions between tipping networks such as the GIS, AMOC and WAIS and may prove problematic due to reduced detection power in the presence of multiple overlying feedback loops.

We simulated two coupled tipping elements X^1 and X^2 with coupling coefficients $d_{21} \in [0, d_{\max}]$ and $d_{12} = d_{21} - d_{\max}$. Let $\alpha = 2d_{21}/d_{\max} - 1$. It indicates the feedback's asymmetry and scales between ± 1, regardless of the value of $d_{\max}$. For $\alpha = -1$ only the coupling $X^1 \to X^2$ exists, while for $\alpha = 1$ only $X^2 \to X^1$ does. For $\alpha = 0$ both links are equal in strength. The resulting feedback loop has a negative sign and acts to supress perturbations.

Fig. 3.8 shows the TPR for the PCMCI+ algorithm in terms of the asymmetry α across multiple $d_{\max}$. Confidence bands were again derived via bootstrap estimation at $p = 0.05$ significance level.

For low $d_{\max}$ (blue curve), which represents an overall weak feedback, detection power is minimized for symmetric feedbacks ($\alpha = 0$) and does not exceed 50% when only one link is included. Apparently, coupling is simply too weak to be picked up and a robust detection of even the stronger link seems challenging.

For higher $d_{\max}$ (green, orange curve), a complex relationship can be observed. The TPR has multiple peaks over the α range, with the most prominent peak occurring near $\alpha = 0$, and smaller local maxima around $\alpha = \pm 0.7$. Here, PCMCI+ is able to discern between links and resolves asymmetries fairly well, although the TPR decreases the weaker any one link is.

Evidently, PCMCI+ resolves symmetric feedback loops better than asymmetric ones and fails to robustly pick up weaker links in both cases. This should be kept in mind when the suspected causal graph exhibits feedbacks.

The multimodal relationship at higher $d_{\max}$ values may be explained by the two stages in which PCMCI+ detects causal relationships between variables. In the initial PC_1 stage, candidates for causal parents are found. As there are only two tipping elements in this system and thus two variables, any conditioning set can only come from their shared past. A causal relationship is rejected when the conditional independence test succeeds with significance p, e.g. when the conditional correlation is sufficiently small. As can be seen in Fig. 3.8 there is a range of α values for which the correlation after the PC_1 stage is below the correlation threshold ρ associated with the significance p (red region). Here, the correlation is reduced due to the autocorrelation of the timeseries. The link is prematurely rejected and not passed to the MCI stage. Therefore, PCMCI+ fails to pick it up.

Around $\alpha = 0.7$ however, the correlation is more likely than not to lie outside the rejection band. This can be seen from the interquartile-range (IQR) exceeding the threshold correlation ρ. Here, the PC_1 stage predominantly retains the link and the MCI stage can address the autocorrelation problem. Here, links are rejected less frequently and a local maximum in the TPR arises.

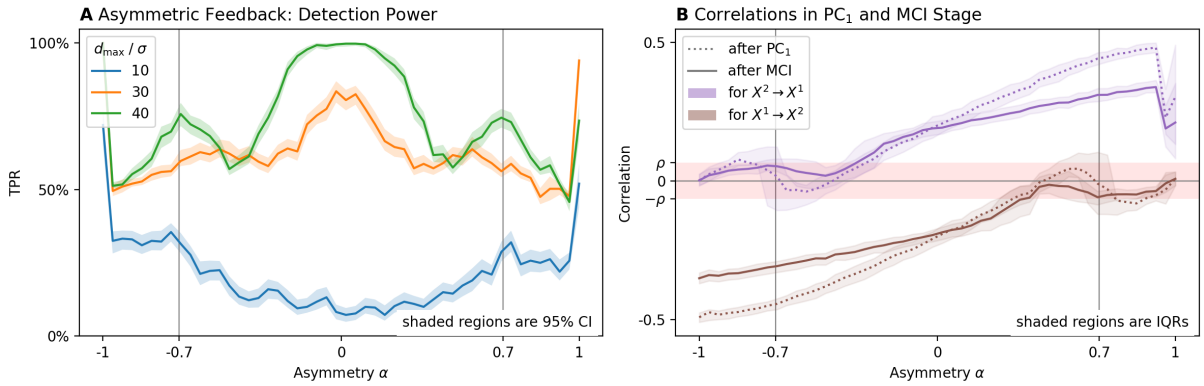


Figure 3.8. | **Detection Power for Asymmetric Feedbacks** Such systems appear, when one link is relatively stronger than another. An asymmetry of $\alpha = \pm 1$ means that only one link is present, while at $\alpha = 0$, both links are equal in strength. **A** A nontrivial relationship between coupling asymmetry and TPR is observed for PCMCI+. **B** Correlations between variables found by PCMCI+ after the preliminary PC_1 stage and after the MCI stage. The red region indicates correlations that are rejected as statistically not significant.

3.3. Detecting Interactions in Paleoclimate Data: The Bipolar-Seesaw

Finally, we applied the `tigramite`[47] framework to paleoclimate proxies and attempted to reconstruct their interactions as a proof of concept. We selected three reconstructions previously used for a correlational analysis by DAVTIAN[56] which reflect the abrupt variability of Dansgaard-Oeschger and Heinrich events during the last glacial period spanning 120 – 10 kyrBP.[24,80,81] They can be broadly linked to the states of the GIS,[82] AMOC[56] and WAIS tipping elements. These proxies were selected, because they have already been synchronized by DAVTIAN to one consistent age model[56] and due to their high resolution which is required for causal inference. The proxies are:

1. *Greenland Air Temperatures* (GAT) reconstructed from d18O isotope records. The data is originally taken from the Northern Greenland Ice Core Project (NGRIP).[83] We treat this record as representing the state of the GIS.

2. *Atlantic Sea-Surface Temperatures* (ASST) reconstructed from a biomarker (ring index of hydroxylated tetraethers, RI-OH') for the MD95-2024 sediment core.[56,84] The core is located at the Iberian Margin southwest of the Iberian Peninsula We understand this record to reflect changes in the AMOC state.

3. *Antarctic Air Temperatur Stack* (ATS) This proxy is prepared by DAVTIAN from the averaged d18O isotope record of three ice cores: Epica Dome C (EDC),[85] West Antarctic Ice Sheet Divide (WD),[38] and Epica Drunning Mode Land (EDML).[86] The stacking procedure is implemented to harmonize contrasting shapes of Antarctic isotope maxima,[56] i.e. temperature curves. We treat this record as representing the state of the WAIS.

3.3.1. Data Preparation

We now describe the preprocessing steps done to the above described proxies, so that they can be used for causal discovery with PCMCI+.

The ice records are densely sampled at 20 yr (GAT) and 10 yr (ATS) resolution but the ASST reconstruction for the Iberian Margin is much coarser at 190 yr. Because uniform sampling is required by PCMCI+, we resampled all three proxies to have 1000 evenly

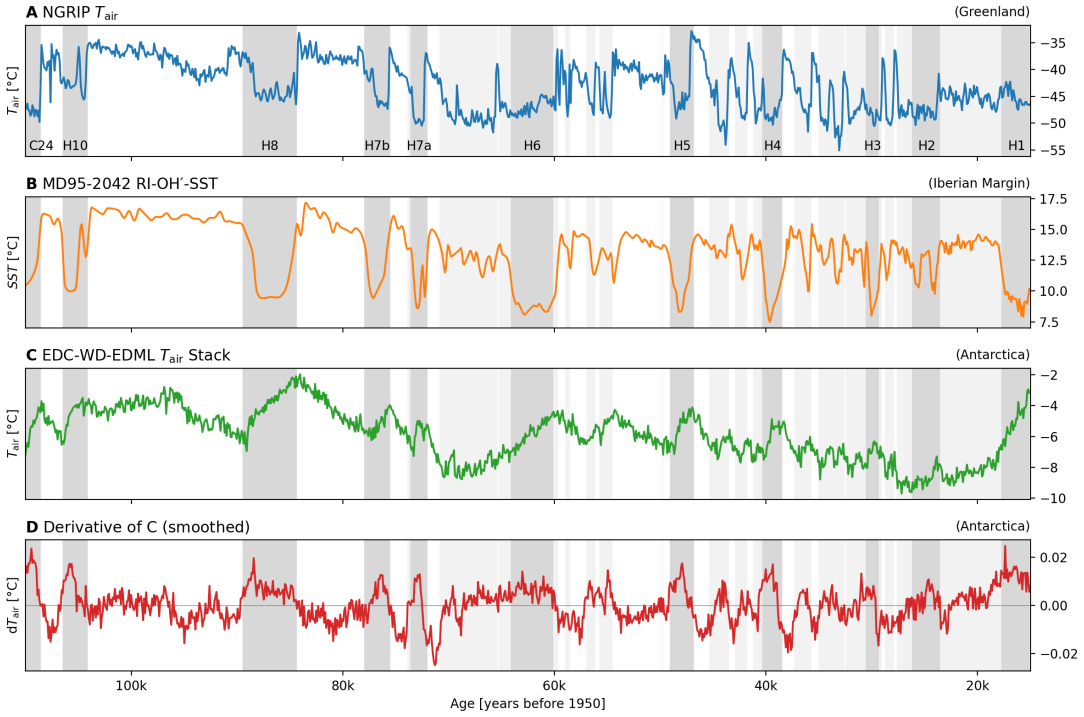


Figure 3.9. | **Paleoclimate Proxies** taken from DAVTIAN.[56] **A** Greenland air temperature record for the NGRIP core derived from d18O isotope data. **B** Iberian margin sea-surface temperature for the MD95-2024 core derived from the ring index of hydroxylated tetrathers (RI-OH'). **C** Antarctic temperature stack as an average of air temperatures for the EDC, WD and EDML cores derived from d18O isotope data. **D** First derivative of the Antarctic temperature stack. Dark gray regions mark Heinrich events, light gray regions NGRIP stadials.

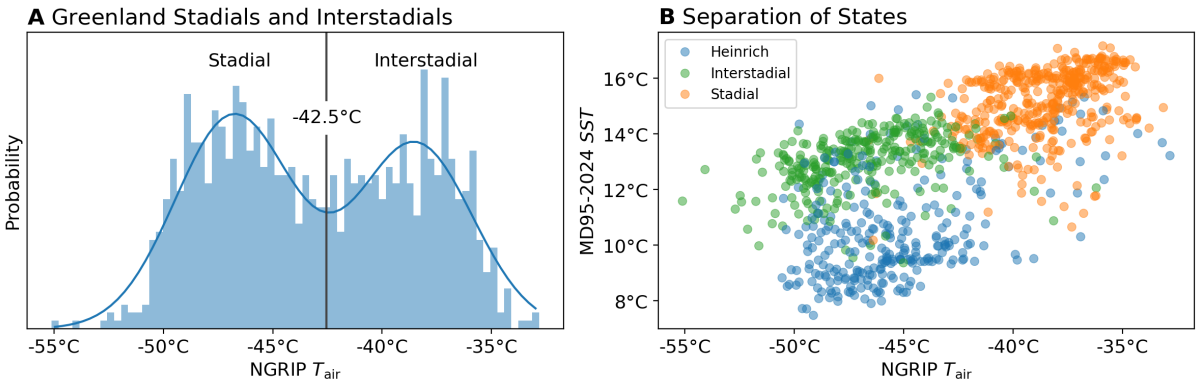


Figure 3.10. | **Classification into Heinrich Events, Stadials and Interstadials A** The bimodal distribution for the NGRIP temperature record is split at the local minimum to classify either stadials or interstadials. **B** In combination with the prescribed Heinrich events, the phase space is clustered into three classes.

spaced points between 110 kyrBP to 15 kyrBP. Though the exact resolution is arbitrary, the sample number was selected to roughly match the ASST to prevent oversampling. The range was selected to limit analysis to the glacial states in which abrupt transitions occurred.

Interpolation was done using an Akima interpolator that fits piecewise cubic polynomials to the data.[87] This strategy was selected over linear interpolation to capitalize on the high resolution of the ice-core records and because of its stability against outliers.

Fig. 3.9 shows the resulting paleoclimate timeseries. Both GAT and the ASST exhibit bistability with synchronized transitions between states. We distinguish three states: Daansgard-Oeschger stadials (cold periods, gray) and interstadials (warm periods, white) in the GAT record and Heinrich events (dark grey).[38,42,56] They have already been described in section 1.4. Note, that GAT and ASST evolve remarkably synchronized, with stadials corresponding to periods of lower ASST. During stadials the ATS tends to grow, while it declines during interstadials.

This relationship between Greenland and Antarctic air temperature reconstructions is well known as the bipolar-seesaw mechanism.[56,86,88,89] A possible explanation for this behavior are changes in the AMOC strength.[43] A weak current coincides with a reduced energy transport between hemispheres, resulting in cold conditions over Greenland and accumulating heat in the Southern Hemisphere.[43] Contrary, a stronger current depletes this energy imbalance and leads to cooling in the Southern Hemisphere.

We implemented two further processing steps to ensure that the proxies are at least piecewise stationary. Firstly, we replaced Arctic Temperatures by their derivative. To do so, we smoothed the original, high-resolution ATS using a rolling average of window length 1.4 kyr. Afterwards, we resampled the derivative to match the resolution of all other variables. This order of operations prevents information loss from premature resampling of the ATS. The window length was selected based on a visual trade-off between reflecting the piecewise linear trends in the original signal and maintaining small scale structures. The exact value is however insignificant, as long as it is greater than the resampling resolution.

Secondly, we masked stadials, interstadials and Heinrich events for causal analysis. For the classification of states, we follow the original paper by Davtian.[56] The GAT record is thresholded into stadials and interstadials, based on the local minimum of the temperature

distribution shown in Fig. 3.10A. The threshold temperature is $T = -42.5\,°\mathrm{C}$. Heinrich event boundaries are directly taken from the paper. In state space, masking leads to three fairly separated clusters with roughly normal marginals, as can be seen in Fig. 3.10B.

3.3.2. Inference Setup and Significance Estimates

For each of these clusters, i.e. the masked timeseries, as well as the complete record, we ran PCMCI+ with three different conditional independence tests and hyperparameters:

Firstly, we used the linear `ParCorr` conditional independence test that assesses independencies using the Pearson correlation coefficient (see section 2.2.2). Secondly, we employed the `RobustParCorr` test that rescales the variables to have a normal distribution beforehand, to see whether this would change results. We used a significance level of $p = 0.05$ for conditional independence. An appropriate $\tau_{\max}$ is found by iteratively increasing its value until the resulting causal graph is stable. We consistently found a maximum lag of $\tau_{\max} = 5$ and thus settled on $\tau_{\max} = 10$ for some margin of error.

Finally, we used the nonlinear `CMIknn`[74] test. Based on Akaike's information criterion, this test automatically selects a p value that should be interpreted as a threshold for the mutual information.[74] Because of its increased runtime, we settled on a shorter $\tau_{\max} = 5$. This value was selected to match the maximum lag seen in the linear tests.

During our analysis we observed that results are fairly sensitive to changes in the input data. Omitting a small fraction of it significantly changed the output causal graph. Furthermore, they seem to be weakly dependent on $\tau_{\max}$.

To address the first issue, we followed Debeire[90] in employing a bootstrap approach to estimate the statistical significance of links for the `ParCorr` and `RobustParCorr` tests. Bootstrap samples are generated as time-ordered and coherent sections of the input data. Because the masked timeseries corresponding to the individual states are already quite short, we used $N = 50$ bootstrap samples to ensure that they remain sufficiently heterogeneous. The algorithm then finds a causal graph for each bootstrap sample and aggregates the results by majority voting, i.e. a link appears in the summary graph when it is present in the majority of bootstrap graphs. The link significance is now expressed as the relative fraction of samples that share this link.

3.3.3. Inferred Causal Graphs

Fig. 3.11 shows the aggregated causal graphs for the whole timeseries (left column) and the masked sections (other columns). Rows correspond to the different independence tests. Inferred causal graphs differ strongly between both tests and masking setup. The following links were found consistently across multiple setups:

1. *Antarctic and Greenland Temperatures:* For most of the setups we find a significant link indicating that changes in Antarctic temperatures have a negative influence on Greenland temperatures. This link is consistently picked up by linear tests in the full timeseries, as well as Heinrich events and with `ParCorr` for interstadials. For stadials, it is not found. The sign of this link can be explained by the above described bipolar-seesaw mechanism, however we would expect that this relationship is due to confounding by the AMOC and should not appear in the causal graph. Because the link is instantaneous, e.g. has lag zero, the link direction should be interpreted with caution, as time ordering can not be exploited to orient the link.

2. *Greenland Temperatures and Atlantic SST:* Here, we find a positive relationship with lag 95 yr from Atlantic SSTs to Greenland temperatures for some of the setups. In the case of the full timeseries, this is probably due to the large scale shape of the data. While the `ParCorr` picks this up, the `RobustParCorr` test may fail to do so, because of the prior variable transformation: rescaling of the bimodal distribution from Fig. 3.10 blurs the boundaries between the two states in phase space. For Heinrich events and stadials, this link is equally picked up by `ParCorr`. After closer inspection, this is however attributed to Heinrich masks including parts of the transition between states. The resultant graph may thus contain artefacts of the independencies found in the whole timeseries. For the stadials, we resolve a negative feedback loop between these variables. However, this should be interpreted with caution, as the feedback is weak and the positive link has low significance.

3. *Atlantic SST and Antarctic Temperatures:* A negative link from the change of Antarctic temperatures to the Atlantic SST is detected in some of the setups. For the full timeseries, the sign is again in-line with the bipolar-seesaw hypothesis. The direction is however reversed.

4. Furthermore, PCMCI+ was not able to find any significant non-linear links using the `CMIknn` test. This indicates, that any relationships between the variables are likely linear according to PCMCI+.

To summarize, this analysis shows that PCMCI+ was not able to pick up a common causal mechanism, neither for all states combined nor between the individual masks. For the full timeseries, detected links seem to match the evident correlations between variables and are probably explained by the large-scale structure of the dataset, which are well known as the bipolar-seesaw mechanism. Within each state, e.g. Heinrich, stadial or interstadial, interactions could not be resolved for all combinations of proxies. Especially for the masked setups, links need to be interpreted with caution, as the mask included parts of the transitions between states.

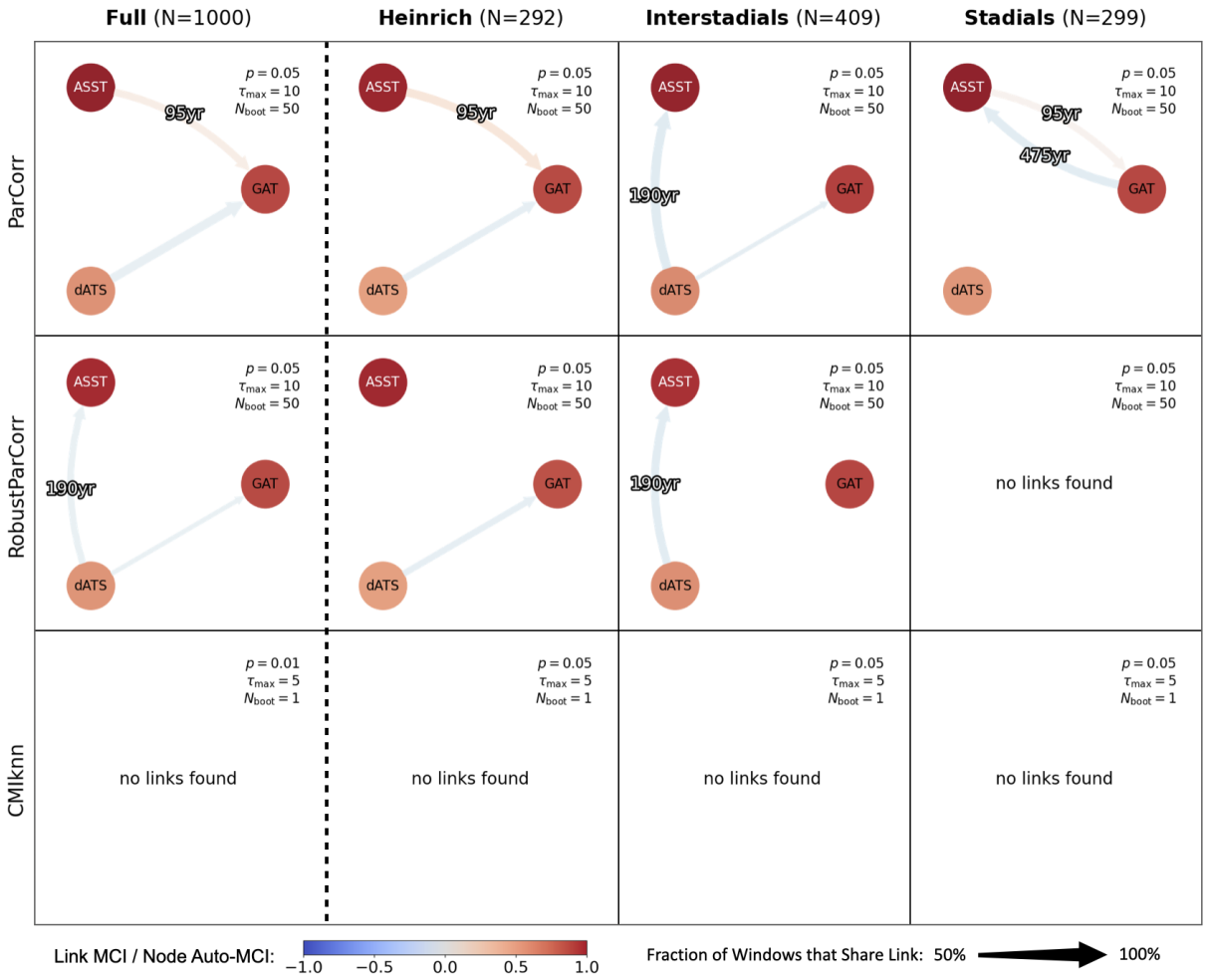


Figure 3.11. | **Causal Analysis of Paleoclimate Proxies** Ressults of the PCMCI+ causal discovery algorithm for the proxy records from Fig. 3.9. Once using the full timeseries and, to enforce piecewise stationarity, on masked sections that correspond to Heinrich events, Greenland stadials and interstadials. We employed three different conditional independence test, two linear onces (`ParCorr` and `RobustParCorr`) and one non-linear one (`CMIknn`). To adress the volatility of results, we implemented a bootstrap approach that determines the statistical significance of links.

4. Discussion

In section 3.1, we applied methods from statistical physics to conceptual tipping networks and expressed the relevant equations in terms of a probability density that obeys the Fokker-Planck equation. Through this approach, we were able to illustrate how the coupling configuration impacts the number and properties of the system's stable states and gained further insights into the emergent dynamics.

In section 3.2, we conducted simulations of conceptual tipping networks to benchmark the capabilities of two causal discovery algorithms (PCMCI and PCMCI+) at reconstructing these interactions from synthetic data. We showed that this is in principle possible, given that coupling is sufficiently strong and masking of the piecewise stationary sections is employed to meet methodical requirements. Out of the two algorithms we assessed, we found that PCMCI was better at detecting existing relationships (TPR), while PCMCI+ showed a lower number of false-positives (FPR). However, this benchmark also revealed several factors that limit an application in the context of interacting tipping elements: both TPR and FPR worsened when the network included multiple overlapping feedback loops or when the relative strength of their links was asymmetric. This especially limits the algorithms ability to detect relationships where coupling is weak. Additionally, we noted that changes in a single link can impair detection for the whole graph, rendering robust inference challenging.

In a conceptual Earth system like network of interacting tipping elements (GIS, AMOC and WAIS), changing the sign of a single link reduced the TPR for most other links. Significant links, which were formerly detected, are then missed. Furthermore, we found that detection power is insufficient for an application to this network when we chose expert-estimated coupling coefficients with realistic relative strength relations. Then, the TPR did not exceed the threshold of 50%.

When we applied PCMCI+ to three paleoclimate proxies that reflect the state of the GIS, AMOC and WAIS tipping elements (section 3.3), causal discovery methods did not yield results beyond a simple correlation analysis and only the well-known bipolar seesaw mechanism was found. Masking of the piecewise stationary sections produced inconclusive causal graphs and no common mechanism could be identified. We attribute some of the links that are shared between sections to improper masking.

In light of these results, causal discovery in the presence of abrupt transitions appears in principle possible, although challenging. Some methodical adaptations need to be made to ensure that requirements of the algorithms are met. Particularly in this context, careful selection of the proper algorithm, conditional independence test and their hyperparameters is crucial and robustness of results hinges strongly on these choices.

In the paleoclimate application explored here, causal discovery did not provide insights beyond a correlation analysis. This suggests that there may be issues with the experimental design.

4.1. Limitations

We identify three main issues with the presented work. Firstly, further unaddressed methodical problems may arise in the context of interacting tipping elements. Secondly, technical difficulties due to the disadvantageous characteristics of paleoclimate data may hinder applicability. Finally, difficulties with the conception and representation of tipping elements complicate a concise interpretation.

4.1.1. Methodical Problems in the Presence of Tipping

We employed masking to ensure stationarity and enable causal discovery in the presence of abrupt transitions. However, this approach of constraining our analysis to sections in which the state is artificially fixed could itself turn out to be problematic. In a setting in which a variable is controlled to be in equilibrium, correlational evidence of interactions with other variables may be suppressed due to a violation of Faithfulness (section 2.2.3).

Friedman presented an analogy that illustrates how this problem arises.[91] Consider a thermostat that controls room temperatures in a house. Clearly, they are influenced by outside temperatures. Radiators and air conditioning are however used to control the room temperature to be constant. This newly introduced artificial link, that is mediated by the thermostat, exactly cancels the former one. By controlling for equilibrium in one variable, temperatures become independent, and no correlation can be observed. A causal discovery algorithm would wrongly assess that inside and outside temperatures are unrelated.

Masking of the stationary states could represent a similar issue. While coupling works between state variables in the conceptual model, the GIS affects the AMOC through

meltwater and therefore depends on the change of ice volume.[22,43] When temperatures and thus ice volume are constrained to be constant, the variability that would represent interactions is suppressed. Causal discovery in the context of tipping is thus a tradeoff between including enough variability to infer interactions and enforcing stationarity to ensure that the assumptions of conditional independence tests are met. We now hypothesize that interactions in the paleoclimate example are likely concealed by large-scale abrupt transitions when the unrestricted data is used and are too weak compared to natural variability to be discernible in the masked sections.

Furthermore, some coupling mechanisms between tipping elements may not be reconstructable at all. Coupling between a state and the derivative or integral of another, can not be properly addressed by the available conditional independence tests. As described above, the GIS influences the AMOC through the intrusion of meltwater. Reversely, due to the heat capacity of ice, the AMOC similarly impacts the GIS through the amount of accumulated heat and not its momentary energy flux.[22] A suitable selection of variables can partly circumvent this problem, as was done in the paleoclimate example by including the change of Antarctic temperatures. When no relationships are hypothesized a priori, an incorrect use of both state and derivative (or integral) is however problematic, as these variables share information and therefore the assumption of independent noise terms is violated (section 2.2.1).

Finally, we have assumed that coupling is constant regardless of the state the system is in. Certainly, physical mechanisms may be suppressed or enabled based on the background state and coupling between tipping elements changes with time.[22] This violates the Causal Stationarity assumption (section 2.2.3) and complicates robust discovery. One method to address this is a windowed approach.[47,92] However, in the paleoclimate example presented here, the resolution is much too coarse to pick up changing points between causal mechanisms and the sample number in the masked sections too small to generate adequate windows.

4.1.2. Disadvantageous Characteristics of Paleoclimate Proxies

Paleoclimate proxies exhibit a multitude of disadvantageous characteristics that hinder an analysis using causal discovery algorithms,[23] regardless of whether they represent tipping elements or not.

Uncertainties in time-ordering due to dating problems or unsynchronized age models between proxies make it impossible to establish the direction of causal relationships.[23,93] While alternative proxies that could reflect tipping elements are provided by Wunderling,[22] their age models are not synchronized. Establishing causality would be impossible. Furthermore, poor resolution can lead to unstable causal graphs, when the sample number for independence tests becomes too small and results are no longer statistically significant.[48] Furthermore, relationships with characteristic timescales below the resolution would be found as contemporaneous by the algorithm.[47]

Proxies reflect a multitude of concurrent processes and may partly share the same signals. Therefore, the algorithmic requirement, that variables do not share information beyond their causal effects on each other, is not met. In combination with the impossibility of including all causal drivers in the dataset to remove confounding, Causal Sufficiency is certainly violated. In the paleoclimate example, all three proxies represent temperatures and therefore share information about general warming trends. Technically, the graph can no longer be interpreted as representing causal relationships. Nonetheless, an application of causal discovery algorithms can still be beneficial, although it should then be understood as a more sophisticated lagged correlation analysis.

4.1.3. Representation of Tipping Elements

The most fundamental issue with an application of causal discovery methods to infer tipping interactions lies however in the conception of tipping elements and their representation in observational data.

In the paleoclimate example, we understood certain proxies to represent the state of tipping elements. We used air temperatures in Greenland and Antarctica as a proxy for ice volume in the GIS and WAIS and sea-surface temperatures close to the Iberian Margin to reflect the strength of the AMOC. This mapping clearly complicates interpretation as these proxies may not be able to reflect such abstract state variables. For example, air temperatures and ice volume may be independent for certain ranges of temperature. Warming does not necessarily translate into ice volume loss when the temperature change is insufficient to kick of the melt-elevation feedback.[28] Conversely, cooling may not lead to increases in ice volume when temperatures are already sufficiently low and the ice sheet is in equilibrium. Additionally, a separation of timescales occurs for these variables. Air

temperatures change much more rapidly than ice volume and fluctuations in the former can not necessarily be interpreted as changes in the latter.

Additionally, it is questionable whether some relevant variables could at all be reconstructed from proxies. Even when tipping of the GIS results only in slow and partial loss of ice volume, information about temperatures and thus ice sheet thickness that would have been preserved in the ice core, is lost. The extreme case of a complete deglaciation, which should be understood as only illustrative, can not be observed in the ice core archives.

Finally, any feedback loops that would have been found in the paleoclimate example are only of limited relevance to potential tipping under sustained global warming. Only limited statements regarding tipping under present conditions can be made based on paleoclimate data from the last glacial. Under sustained global warming, interactions between tipping elements could be enabled, which are not relevant under glacial conditions and therefore can not be seen in the paleoclimate proxies used here. A closer analog to future conditions could be the last interglacial 116 – 129 kyrBP with global mean temperature around 2°C warmer.[94] This period is however too short for causal discovery and does not include abrupt transitions attributed to tipping elements.

4.1.4. Validation of Causal Graphs

Even when all assumptions of causal discovery algorithms could be met, validation of the resulting causal graphs would be difficult. Because interventions into the climate of the past are impossible, verification can only be done by comparison to expert knowledge. However, in the paleoclimate example discussed here, the mechanism that enables abrupt transitions or how proxies interact in detail is still debated.[44,81]

Because validation of causal graphs is evidently hard, we risk confirmation bias: it is easy to tune results until the graph represents the expected relationships. This is especially the case when results are unstable, as is the case here.

One approach to increase certainty about results is based on compatibility of causal graphs learned on subsets of variables.[95] This approach allows to reject erroneous links based on incompatibilities between subset graphs, which arise due to violations of assumptions or limited sample size.[95] In the paleoclimate example presented here, this approach is however not feasible due to the low number of variables.

4.2. Recommendations and Further Work

Overall, causal discovery algorithms are a promising, although limited, tool for climate and Earth system sciences. When assumptions can not be met and results no longer represent causal relationships, causal discovery algorithms like PCMCI(+) still represent a sophisticated lagged correlation analysis. Because of the methodical complexity, it should be clearly stated how and why the causal interpretation is then limited.

An application that analyses interacting tipping elements in models such as PyCascades is possible, although challenging due to reduced detection power in the presence of multiple overlying feedback loops and multistability. Further methodological work on causal discovery in the presence of abrupt transitions is therefore needed.

Applications to paleoclimate datasets, especially targeting interactions of tipping elements, are severely limited due to data quality and representation problems. Here, further work should focus on how a proxy can be used to represent the state of a tipping element.

One approach to circumvent this limited applicability is to use Earth system models which are constrained by paleoclimate reconstructions, such as Climber-X.[61] The simulated variables would better reflect tipping elements and evade problems like shared information between individual proxies. However, this poses the danger of arriving at self affirming results. The causal graphs may just show the relationships that are encoded in the model. Another promising application could focus on remotely sensed data and faster tipping elements.[68] This could either target the Amazon rainforest,[22,29] or atmospheric data to study teleconnections.

5. Conclusion

In conclusion, the use of causal discovery algorithms for climate and Earth system sciences, particularly for an analysis of interacting tipping elements, is promising but comes with significant challenges. We have found that methodological issues arise due to nonstationarity, overlying feedback loops and unmet assumption of the algorithms. An application to paleoclimate data is further complicated by data quality, availability and the problem of properly representing tipping elements. However, when methodical requirements are not met, causal discovery algorithms can still serve as sophisticated lagged correlation analysis tools, though the causal interpretation is lost. When we applied PCMCI+ to three paleoclimate proxies reflecting Dansgaard-Oeschger events, which have previously been linked to frequently studied tipping elements (Greenland and West Antarctic Ice Sheets, Atlantic Meridional Overturning Circulation), the resulting causal graphs reflected the bipolar seesaw, also evident through traditional correlation analysis and no further interactions could be resolved. Future work should focus on adapting algorithms for nonstationarity, better representing tipping elements in paleoclimate proxies and applying the causal inference framework to Earth system models constrained by paleoclimate data.

References

1. Armstrong McKay, D. I. *et al.* Exceeding 1.5°C Global Warming Could Trigger Multiple Climate Tipping Points. *Science* **377,** eabn7950. DOI: 10.1126/science.abn7950 (2022).

2. Burke, K. D. *et al.* Pliocene and Eocene Provide Best Analogs for Near-Future Climates. *Proceedings of the National Academy of Sciences* **115,** 13288–13293. DOI: 10.1073/pnas.1809600115 (2018).

3. Rockström, J. *et al.* A Safe Operating Space for Humanity. *Nature* **461,** 472–475. DOI: 10.1038/461472a (2009).

4. Steffen, W., Broadgate, W., Deutsch, L., Gaffney, O. & Ludwig, C. The Trajectory of the Anthropocene: The Great Acceleration. *The Anthropocene Review* **2,** 81–98. DOI: 10.1177/2053019614564785 (2015).

5. Steffen, W. *et al.* Trajectories of the Earth System in the Anthropocene. *Proceedings of the National Academy of Sciences* **115,** 8252–8259. DOI: 10.1073/pnas.1810141115 (2018).

6. ECMWF. Global Climate Highlights 2023. https://climate.copernicus.eu/copernicus-2023-hottest-year-record (2024).

7. Bustamante, M. *et al.* Ten New Insights in Climate Science 2023/2024. *Global Sustainability,* 1–58. DOI: 10.1017/sus.2023.25 (2023).

8. Boehm, S. *et al.* State of Climate Action 2023. *World Resources Institute.* DOI: 10.46830/wrirpt.23.00010 (2023).

9. Crutzen, P. J. Geology of Mankind. *Nature* **415,** 23–23. DOI: 10.1038/415023a (2002).

10. Jones, C. D. *et al.* The Zero Emissions Commitment Model Intercomparison Project (ZECMIP) Contribution to C4MIP: Quantifying Committed Climate Changes Following Zero Carbon Emissions. *Geoscientific Model Development* **12,** 4375–4385. DOI: 10.5194/gmd-12-4375-2019 (2019).

11. MacDougall, A. H. *et al.* Is There Warming in the Pipeline? A Multi-Model Analysis of the Zero Emissions Commitment from CO₂ *Biogeosciences* **17,** 2987–3016. DOI: 10.5194/bg-17-2987-2020 (2020).

12. Ganopolski, A., Winkelmann, R. & Schellnhuber, H. J. Critical Insolation–CO2 Relation for Diagnosing Past and Future Glacial Inception. *Nature* **529,** 200–203. DOI: 10.1038/nature16494 (2016).

13. Calvin, K. *et al. IPCC, 2023: Climate Change 2023: Synthesis Report. Contribution of Working Groups I, II and III to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change* (Intergovernmental Panel on Climate Change (IPCC), 2023). DOI: 10.59327/IPCC/AR6-9789291691647.

14. Rounce, D. R. *et al.* Global Glacier Change in the 21st Century: Every Increase in Temperature Matters. *Science* **379,** 78–83. DOI: 10.1126/science.abo1324 (2023).

15. Trenberth, K. Changes in Precipitation with Climate Change. *Climate Research* **47,** 123–138. DOI: 10.3354/cr00953 (2011).

16. Bellard, C., Bertelsmeier, C., Leadley, P., Thuiller, W. & Courchamp, F. Impacts of Climate Change on the Future of Biodiversity. *Ecology Letters* **15,** 365–377. DOI: 10.1111/j.1461-0248.2011.01736.x (2012).

17. *Climate Change 2007: Synthesis Report. Contribution of Working Groups I, II and III to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change* (IPCC, Geneva, 2008).

18. Levermann, A. *et al.* Potential Climatic Transitions with Profound Impact on Europe: Review of the Current State of Six 'Tipping Elements of the Climate System'. *Climatic Change* **110,** 845–878. DOI: 10.1007/s10584-011-0126-5 (2012).

19. Lenton, T. M. *et al.* Tipping Elements in the Earth's Climate System. *Proceedings of the National Academy of Sciences* **105,** 1786–1793. DOI: 10.1073/pnas.0705414105 (2008).

20. Feudel, U. Rate-Induced Tipping in Ecosystems and Climate: The Role of Unstable States, Basin Boundaries and Transient Dynamics. *Nonlinear Processes in Geophysics* **30,** 481–502. DOI: 10.5194/npg-30-481-2023 (2023).

21. Ashwin, P., Wieczorek, S., Vitolo, R. & Cox, P. Tipping Points in Open Systems: Bifurcation, Noise-Induced and Rate-Dependent Examples in the Climate System. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* **370,** 1166–1184. DOI: 10.1098/rsta.2011.0306 (2012).

22. Wunderling, N. *et al.* Climate Tipping Point Interactions and Cascades: A Review. *Earth System Dynamics* **15,** 41–74. DOI: 10.5194/esd-15-41-2024 (2024).

23. Brovkin, V. *et al.* Past Abrupt Changes, Tipping Points and Cascading Impacts in the Earth System. *Nature Geoscience* **14,** 550–558. DOI: 10.1038/s41561-021-00790-5 (2021).

24. Boers, N., Ghil, M. & Stocker, T. F. Theoretical and Paleoclimatic Evidence for Abrupt Transitions in the Earth System. *Environmental Research Letters* **17,** 093006. DOI: 10.1088/1748-9326/ac8944 (2022).

25. Lohmann, J., Dijkstra, H. A., Jochum, M., Lucarini, V. & Ditlevsen, P. D. Multistability and Intermediate Tipping of the Atlantic Ocean Circulation. *Science Advances* **10,** eadi4253. DOI: 10.1126/sciadv.adi4253 (2024).

26. Hofmann, M. & Rahmstorf, S. On the Stability of the Atlantic Meridional Overturning Circulation. *Proceedings of the National Academy of Sciences* **106,** 20584–20589. DOI: 10.1073/pnas.0909146106 (2009).

27. Zeitz, M., Haacker, J. M., Donges, J. F., Albrecht, T. & Winkelmann, R. Dynamic Regimes of the Greenland Ice Sheet Emerging from Interacting Melt–Elevation and Glacial Isostatic Adjustment Feedbacks. *Earth System Dynamics* **13,** 1077–1096. DOI: 10.5194/esd-13-1077-2022 (2022).

28. Levermann, A. & Winkelmann, R. A Simple Equation for the Melt Elevation Feedback of Ice Sheets. *The Cryosphere* **10,** 1799–1807. DOI: 10.5194/tc-10-1799-2016 (2016).

29. Flores, B. M. *et al.* Critical Transitions in the Amazon Forest System. *Nature* **626,** 555–564. DOI: 10.1038/s41586-023-06970-0 (2024).

30. Veron, J. *et al.* The Coral Reef Crisis: The Critical Importance Of<350ppm CO2. *Marine Pollution Bulletin* **58,** 1428–1436. DOI: 10.1016/j.marpolbul.2009.09.009 (2009).

31. Lenton, McKay, A., Loriani, Abrams & Lade. *Global Tipping Points Report 2023* (University of Exeter, 2023). https://global-tipping-points.org.

32. Abrams, J. F. *et al.* Committed Global Warming Risks Triggering Multiple Climate Tipping Points. *Earth's Future* **11,** e2022EF003250. DOI: 10.1029/2022EF003250 (2023).

33. Kriegler, E., Hall, J. W., Held, H., Dawson, R. & Schellnhuber, H. J. Imprecise Probability Assessment of Tipping Points in the Climate System. *Proceedings of the National Academy of Sciences* **106,** 5041–5046. DOI: 10.1073/pnas.0809117106 (2009).

34. Lenton, T. M. *et al.* Climate Tipping Points — Too Risky to Bet Against. *Nature* **575,** 592–595. DOI: 10.1038/d41586-019-03595-0 (2019).

35. Van Westen, R. M., Kliphuis, M. & Dijkstra, H. A. Physics-Based Early Warning Signal Shows That AMOC Is on Tipping Course. *Science Advances* **10,** eadk1189. DOI: 10.1126/sciadv.adk1189 (2024).

36. Boers, N. & Rypdal, M. Critical Slowing down Suggests That the Western Greenland Ice Sheet Is Close to a Tipping Point. *Proceedings of the National Academy of Sciences* **118,** e2024192118. DOI: 10.1073/pnas.2024192118 (2021).

37. Caesar, L., McCarthy, G. D., Thornalley, D. J. R., Cahill, N. & Rahmstorf, S. Current Atlantic Meridional Overturning Circulation Weakest in Last Millennium. *Nature Geoscience* **14,** 118–120. DOI: 10.1038/s41561-021-00699-z (2021).

38. WAIS Divide Project Members. Precise Interpolar Phasing of Abrupt Climate Change during the Last Ice Age. *Nature* **520,** 661–665. DOI: 10.1038/nature14401 (2015).

39. Jackson, L. C. *et al.* Global and European Climate Impacts of a Slowdown of the AMOC in a High Resolution GCM. *Climate Dynamics* **45,** 3299–3316. DOI: 10.1007/s00382-015-2540-2 (2015).

40. Wunderling, N., Donges, J. F., Kurths, J. & Winkelmann, R. Interacting Tipping Elements Increase Risk of Climate Domino Effects under Global Warming. *Earth System Dynamics* **12,** 601–619. DOI: 10.5194/esd-12-601-2021 (2021).

41. Notz, D. & Stroeve, J. Observed Arctic Sea-Ice Loss Directly Follows Anthropogenic CO $_2$ Emission. *Science* **354,** 747–750. DOI: 10.1126/science.aag2345 (2016).

42. Bauska, T. K., Marcott, S. A. & Brook, E. J. Abrupt Changes in the Global Carbon Cycle during the Last Glacial Period. *Nature Geoscience* **14,** 91–96. DOI: 10.1038/s41561-020-00680-2 (2021).

43. Ganopolski, A. & Rahmstorf, S. Rapid Changes of Glacial Climate Simulated in a Coupled Climate Model. *Nature* **409,** 153–158. DOI: 10.1038/35051500 (2001).

44. Lohmann, J. Prediction of Dansgaard-Oeschger Events From Greenland Dust Records. *Geophysical Research Letters* **46,** 12427–12434. DOI: 10 . 1029 / 2019GL085133 (2019).

45. Gildor, H. & Tziperman, E. Sea-Ice Switches and Abrupt Climate Change. *Philosophical Transactions of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences* **361** (eds Lawton, J. H., Marotzke, J., Marsh, R. & McCave, I. N.) 1935–1944. DOI: 10.1098/rsta.2003.1244 (2003).

46. Max, L., Nürnberg, D., Chiessi, C. M., Lenz, M. M. & Mulitza, S. Subsurface Ocean Warming Preceded Heinrich Events. *Nature Communications* **13,** 4217. DOI: 10.1038/s41467-022-31754-x (2022).

47. Runge, J., Gerhardus, A., Varando, G., Eyring, V. & Camps-Valls, G. Causal Inference for Time Series. *Nature Reviews Earth & Environment* **4,** 487–505. DOI: 10.1038/s43017-023-00431-y (2023).

48. Runge, J. Discovering Contemporaneous and Lagged Causal Relations in Autocorrelated Nonlinear Time Series Datasets (2020).

49. Merkel, U., Prange, M. & Schulz, M. ENSO Variability and Teleconnections during Glacial Climates. *Quaternary Science Reviews* **29,** 86–100. DOI: 10.1016/j.quascirev.2009.11.006 (2010).

50. Zhuo, W. *et al. Causal Effects of Teleconnection Patterns on Soil Moisture through Different Climate Paths over the Greater Horn of Africa*

51. Kretschmer, M., Coumou, D., Donges, J. F. & Runge, J. Using Causal Effect Networks to Analyze Different Arctic Drivers of Midlatitude Winter Circulation. *Journal of Climate* **29,** 4069–4081. DOI: 10.1175/JCLI-D-15-0654.1 (2016).

52. Bénard, G., Gehlen, M. & Vrac, M. *Multi-Model Comparison of Causal Relationships between Atmospheric and Marine Biogeochemical Variables*

53. Debeire, K., Eyring, V., Nowack, P. & Runge, J. *Causal Discovery as a Novel Approach for CMIP6 Climate Model Evaluation*

54. Docquier, D. *et al.* A Comparison of Two Causal Methods in the Context of Climate Analyses. *Nonlinear Processes in Geophysics* **31,** 115–136. DOI: 10.5194/npg-31-115-2024 (2024).

55. Addo, P. M., Manibialoa, C. & McIsaac, F. Exploring Nonlinearity on the CO 2 Emissions, Economic Production and Energy Use Nexus: A Causal Discovery Approach. *Energy Reports* **7,** 6196–6204. DOI: 10.1016/j.egyr.2021.09.026 (2021).

56. Davtian, N. & Bard, E. A New View on Abrupt Climate Changes and the Bipolar Seesaw Based on Paleotemperatures from Iberian Margin Sediments. *Proceedings of the National Academy of Sciences* **120,** e2209558120. DOI: 10.1073/pnas.2209558120 (2023).

57. Drijfhout, S. *et al.* Catalogue of Abrupt Shifts in Intergovernmental Panel on Climate Change Climate Models. *Proceedings of the National Academy of Sciences* **112.** DOI: 10.1073/pnas.1511451112 (2015).

58. Wang, S. *et al.* Mechanisms and Impacts of Earth System Tipping Elements. *Reviews of Geophysics* **61,** e2021RG000757. DOI: 10.1029/2021RG000757 (2023).

59. Rosser, J., Winkelmann, R. & Wunderling, N. *Cryosphere Tipping Elements Decisive for Tipping Risks and Cascading Effects in the Earth System* preprint.

60. Song, X., Wang, D.-Y., Li, F. & Zeng, X.-D. Evaluating the Performance of CMIP6 Earth System Models in Simulating Global Vegetation Structure and Distribution. *Advances in Climate Change Research* **12,** 584–595. DOI: 10.1016/j.accre.2021.06.008 (2021).

61. Willeit, M., Ganopolski, A., Robinson, A. & Edwards, N. R. The Earth System Model CLIMBER-X v1.0 – Part 1: Climate Model Description and Validation. *Geoscientific Model Development* **15,** 5905–5948. DOI: 10.5194/gmd-15-5905-2022 (2022).

62. Loriani, S. *Tipping Model Intercomparison Project* https://tipmip.pik-potsdam.de/documents/TIPMIP_Flyer.pdf.

63. Wunderling, N. *et al.* Modelling Nonlinear Dynamics of Interacting Tipping Elements on Complex Networks: The PyCascades Package. *The European Physical Journal Special Topics* **230,** 3163–3176. DOI: 10.1140/epjs/s11734-021-00155-4 (2021).

64. Wunderling, N., Gelbrecht, M., Winkelmann, R., Kurths, J. & Donges, J. F. Basin Stability and Limit Cycles in a Conceptual Model for Climate Tipping Cascades. *New Journal of Physics* **22,** 123031. DOI: 10.1088/1367-2630/abc98a (2020).

65. Klose, A. K., Karle, V., Winkelmann, R. & Donges, J. F. Emergence of Cascading Dynamics in Interacting Tipping Elements of Ecology and Climate. *Royal Society Open Science* **7,** 200599. DOI: 10.1098/rsos.200599 (2020).

66. Rietkerk, M., Dekker, S. C., De Ruiter, P. C. & Van De Koppel, J. Self-Organized Patchiness and Catastrophic Shifts in Ecosystems. *Science* **305,** 1926–1929. DOI: 10.1126/science.1101867 (2004).

67. Cimatoribus, A. A., Drijfhout, S. S., Livina, V. & Van Der Schrier, G. Dansgaard–Oeschger Events: Bifurcation Points in the Climate System. *Climate of the Past* **9,** 323–333. DOI: 10.5194/cp-9-323-2013 (2013).

68. Lenton, T. M. *et al.* Remotely Sensing Potential Climate Change Tipping Points across Scales. *Nature Communications* **15,** 343. DOI: 10.1038/s41467-023-44609-w (2024).

69. Pearl, J. *Causality: Models, Reasoning, and Inference* 2nd ed. DOI: 10.1017/CBO9780511803161 (Cambridge University Press, 2009).

70. Runge, J., Nowack, P., Kretschmer, M., Flaxman, S. & Sejdinovic, D. Detecting and Quantifying Causal Associations in Large Nonlinear Time Series Datasets. *Science Advances* **5,** eaau4996. DOI: 10.1126/sciadv.aau4996 (2019).

71. Runge, J. Causal Network Reconstruction from Time Series: From Theoretical Assumptions to Practical Estimation. *Chaos: An Interdisciplinary Journal of Nonlinear Science* **28,** 075310. DOI: 10.1063/1.5025050 (2018).

72. Granger, C. W. J. Investigating Causal Relations by Econometric Models and Cross-spectral Methods. *Econometrica* **37,** 424. DOI: 10.2307/1912791. JSTOR: 1912791 (1969).

73. Leamer, E. E. Vector Autoregressions for Causal Inference? *Carnegie-Rochester Conference Series on Public Policy* **22,** 255–304. DOI: 10.1016/0167-2231(85)90035-1 (1985).

74. Runge, J. Conditional Independence Testing Based on a Nearest-Neighbor Estimator of Conditional Mutual Information (2018).

75. Runge, J., Nowack, P., Kretschmer, M., Flaxman, S. & Sejdinovic, D. Detecting Causal Associations in Large Nonlinear Time Series Datasets. *Science Advances* **5,** eaau4996. DOI: 10.1126/sciadv.aau4996. arXiv: 1702.07007 [physics, stat] (2019).

76. Itô, K. Stochastic Differential Equations in a Differentiable Manifold. *Nagoya Mathematical Journal* **1,** 35–47. DOI: 10.1017/S0027763000022819 (1950).

77. Livina, V. N., Kwasniok, F. & Lenton, T. M. Potential Analysis Reveals Changing Number of Climate States during the Last 60 Kyr. *Climate of the Past* **6,** 77–82. DOI: 10.5194/cp-6-77-2010 (2010).

78. Zwicker, D. *Py-PDE* version 0.38.0. https://github.com/zwicker-group/py-pde/tree/master.

79. *Open Source Computer Vision Library (Python)* version 4.9.0. 2023. https://github.com/opencv/opencv-python.

80. Barker, S. & Knorr, G. Antarctic Climate Signature in the Greenland Ice Core Record. *Proceedings of the National Academy of Sciences* **104,** 17278–17282. DOI: 10.1073/pnas.0708494104 (2007).

81. Menviel, L. C., Skinner, L. C., Tarasov, L. & Tzedakis, P. C. An Ice–Climate Oscillatory Framework for Dansgaard–Oeschger Cycles. *Nature Reviews Earth & Environment* **1,** 677–693. DOI: 10.1038/s43017-020-00106-y (2020).

82. Dutton, A. *et al.* Sea-Level Rise Due to Polar Ice-Sheet Mass Loss during Past Warm Periods. *Science* **349,** aaa4019. DOI: 10.1126/science.aaa4019 (2015).

83. Kindler, P. *et al.* Temperature Reconstruction from 10 to 120 Kyr B2k from the NGRIP Ice Core. *Climate of the Past* **10,** 887–902. DOI: 10.5194/cp-10-887-2014 (2014).

84. Davtian, N., Bard, E., Darfeuil, S., Ménot, G. & Rostek, F. The Novel Hydroxylated Tetraether Index RI-OH as a Sea Surface Temperature Proxy for the 160-45 Ka BP Period Off the Iberian Margin. *Paleoceanography and Paleoclimatology* **36,** e2020PA004077. DOI: 10.1029/2020PA004077 (2021).

85. EPICA community members. Eight Glacial Cycles from an Antarctic Ice Core. *Nature* **429,** 623–628. DOI: 10.1038/nature02599 (2004).

86. EPICA Community Members. One-to-One Coupling of Glacial Climate Variability in Greenland and Antarctica. *Nature* **444,** 195–198. DOI: 10.1038/nature05301 (2006).

87. Akima, H. A New Method of Interpolation and Smooth Curve Fitting Based on Local Procedures. *Journal of the ACM* **17,** 589–602. DOI: 10.1145/321607.321609 (1970).

88. Schmittner, A., Saenko, O. & Weaver, A. Coupling of the Hemispheres in Observations and Simulations of Glacial Climate Change. *Quaternary Science Reviews* **22,** 659–671. DOI: 10.1016/S0277-3791(02)00184-1 (2003).

89. Stocker, T. F. The Seesaw Effect. *Science* **282,** 61–62. DOI: 10.1126/science.282.5386.61 (1998).

90. Debeire, K., Gerhardus, A., Runge, J. & Eyring, V. Bootstrap Aggregation and Confidence Measures to Improve Time Series Causal Discovery. *Proceedings of Machine Learning Research,* 979–1007. https://proceedings.mlr.press/v236/debeire24a/debeire24a.pdf (2024).

91. Friedman, M. The Fed's Thermostat. https://www.wsj.com/articles/SB106125694925954100 (2003).

92. Saggioro, E., de Wiljes, J., Kretschmer, M. & Runge, J. Reconstructing Regime-Dependent Causal Relationships from Observational Time Series. *Chaos: An Interdisciplinary Journal of Nonlinear Science* **30,** 113115. DOI: 10.1063/5.0020538 (2020).

93. Ganopolski, A. & Roche, D. M. On the Nature of Lead–Lag Relationships during Glacial–Interglacial Climate Transitions. *Quaternary Science Reviews* **28,** 3361–3378. DOI: 10.1016/j.quascirev.2009.09.019 (2009).

94. Thomas, Z. A. *et al.* Tipping Elements and Amplified Polar Warming during the Last Interglacial. *Quaternary Science Reviews* **233,** 106222. DOI: 10.1016/j.quascirev.2020.106222 (2020).

95. Faller, P. M., Vankadara, L. C., Mastakouri, A. A., Locatello, F. & Janzing, D. *Self-Compatibility: Evaluating Causal Discovery without Ground Truth* arXiv: 2307.09552 [cs, stat]. http://arxiv.org/abs/2307.09552 (2024). preprint.

Appendix

A. Acknowledgement

First and foremost, I would like to thank my supervisors, Johan Rockström and Nico Wunderling, for their invaluable guidance and support throughout this project. I am also grateful to Georg Feulner, Andrey Ganopolski, and Matteo Willeit for their suggestions and guidance regarding paleoclimate data. Additionally, I thank Urni Ninad, Jonas Wahl, and Alexandrine Lanson for their input on causal inference methods and our productive discussions about their applicability.

B. Code and Data Availability

All code and data is available on the dedicated GitHub repository under https://github.com/david-strahl/master-thesis.

C. Declaration of Authorship

I hereby declare that the thesis is my own unaided work. The use of any direct or indirect sources has been acknowledged as a reference. I am aware, that the thesis can be digitally examined for the use of unauthorized aid and may be deemed as plagiarism if any non-referenced material is incorporated.