

Technische Universität Berlin

Big Data Engineering (DAMS)

Fakultät IV

Ernst-Reuter-Platz 7

10587 Berlin

<https://www.tu.berlin/dams>



Master's Thesis

Multivariate Causal Inference Methods Comparison and Application to Climate Tipping Point Interactions

Niklas Lohmann

Matriculation Number: 487733

16.05.2025

Supervised by
Prof. Dr. Matthias Boehm
Prof. Dr. Nico Wunderling
Christina Dionysio

Acknowledgements

I would like to thank my supervisors Prof. Matthias Böhm and Christina Dionysio at TU Berlin for their valuable and constructive feedback, and the structure they brought into my work.

Further, I'm very grateful to my supervisor Nico Wunderling at PIK for introducing me to climate tipping point science, for his continuous encouragement and the opportunities he offered me.

Thank you to the ERSU team at PIK for the warm welcome, many inspiring coffee breaks and conversations. Special thanks to Annika Högner for the help with getting started and other valuable tips, to Moritz Kreuzer for sharing his model insights and data, to David Strahl, Jonathan Donges, Willem Huiskamp and Jonathan Krönke for all their valuable feedback and ideas.

Many thanks to Karoline Ramin and Kobe de Maeyer for supporting and motivating me throughout the highs and lows of this work.

Finally, a warm thank you my parents for their support in all the different ways, and to Sara, Misha, Paula, Leon, Lotte, Thimo, Martin, Estelle, Pauler and Max for the fun times, long conversations, and for reminding me what's important.

Eigenständigkeitserklärung

Hiermit versichere ich, dass ich die vorliegende Arbeit eigenständig ohne Hilfe Dritter und ausschließlich unter Verwendung der aufgeführten Quellen und Hilfsmittel angefertigt habe. Alle Stellen die den benutzten Quellen und Hilfsmitteln unverändert oder sinngemäß entnommen sind, habe ich als solche kenntlich gemacht.

Sofern generische KI-Tools verwendet wurden, habe ich Produktnamen, Hersteller, die jeweils verwendete Softwareversion und die jeweiligen Einsatzzwecke (z.B. sprachliche Überprüfung und Verbesserung der Texte, systematische Recherche) benannt. Ich verantworte die Auswahl, die Übernahme und sämtliche Ergebnisse des von mir verwendeten KI-generierten Outputs vollumfänglich selbst.

Die Satzung zur Sicherung guter wissenschaftlicher Praxis an der TU Berlin vom 8. März 2017. https://www.static.tu.berlin/fileadmin/www/10000060/FSC/Promotion__Habilitation/Dokumente/Grundsätze_gute_wissenschaftliche_Praxis_2017.pdf habe ich zur Kenntnis genommen. Ich erkläre weiterhin, dass ich die Arbeit in gleicher oder ähnlicher Form noch keiner anderen Prüfungsbehörde vorgelegt habe.

Berlin, 16.5.2025

.....
(Niklas Lohmann)

Abstract

Climate tipping elements have the potential to alter some of the core mechanisms of the climate system of Earth irreversibly on human time scales, resulting in e.g., sea level rise or biosphere degradation. A tipping event may also (de-)stabilize other tipping elements, potentially leading to a "domino effect" of tipping events. Despite these grave potential consequences, knowledge on climate tipping point interactions is sparse and difficult to obtain from observations while tipping has not yet set in. In complex climate models, computational and conceptual constraints have so far largely prevented a comprehensive analysis of such interactions. Causal inference presents an approach to detect and quantify tipping element interactions from observational time series through different statistical methods that extend beyond correlation analysis. We conduct a systematic evaluation and comparison of three causal inference algorithms under conditions expected for real-world application: The Peter-Clark Momentary Conditional Independence (PCMCI), the Liang-Kleeman Information Flow (LKIF) and Granger Causality for State Space Models (GCSS). Further, we apply these methods to observational and reanalysis data at different spatial resolutions on Arctic sea ice measures (albedo, concentration, extent and thickness), the overturning strength of the Atlantic Meridional Overturning Circulation (AMOC) and the precipitation over the Amazon rainforest to detect an interaction network between tipping elements. According to our findings, melting of Arctic sea ice leads to a short-term strengthening and a long-term weakening of the AMOC on time scales of several months, the latter of which can be physically explained by the increased influx of meltwater with low salinity. Additionally, our results indicate that a weakening of the AMOC would result in small sea ice gains, which is in line with physical expectations of lower heat transport into the Arctic by the AMOC. The destabilizing effects of Arctic sea ice melting on the AMOC underscore the threat posed by interacting tipping elements already in the coming decades.

Zusammenfassung

Klimakipppunkte haben das Potential, wesentliche Mechanismen des Klimasystems der Erde auf menschlichen Zeitskalen irreversibel zu verändern und so zum Beispiel zu Meeresspiegelanstieg oder Schäden der Biosphäre zu führen. Ein einzelnes Kippereignis kann zudem andere Kippelemente (de-)stabilisieren, was schlimmstenfalls zu einem "Domino-Effekt" an Kippereignissen führen könnte. Trotz dieser schwerwiegenden möglichen Konsequenzen ist es schwierig, Wissen über Klimakipppunkte aus Beobachtungsdaten zu ziehen, solange keine Kippdynamik eingesetzt hat. In komplexen Klimamodellen verhindern bisher Grenzen der Rechenleistung und des Modellumfangs eine umfassende Analyse dieser Interaktionen. Kausale Inferenz bietet einen Ansatz, Interaktionen von Kippelementen in Beobachtungsdaten zu detektieren und zu quantifizieren, verschiedene statistische Methoden stehen zur Verfügung, die über bloße Korrelationen hinausgehen. Wir führen eine systematische Evaluierung und einen Vergleich dreier Algorithmen zu kausaler Inferenz unter Bedingungen durch, die für die Anwendung auf reale Daten erwartet werden: Die Methode "Peter-Clark Momentary Conditional Independence" (PCMCI), der "Liang-Kleeman Information Flow" (LKIF) und die Methode "Granger Causality for State Space Models" (GCSS). Darüber hinaus wenden wir diese Methoden auf Beobachtungs- und Reanalysedaten unterschiedlicher räumlicher Auflösung zum arktischen Meereis (Albedo, Konzentration, Ausdehnung und Dicke), zur Atlantischen Umwälzströmung (AMOC) und zum Niederschlag über dem südlichen Amazonas-Regenwald an, und identifizieren so ein Netzwerk an Interaktionen von Kippelementen. Unsere Ergebnisse zeigen, dass ein Schmelzen des Arktischen Meereises zu einer kurzfristigen Stärkung und einer langfristigen Schwächung der AMOC auf Zeitskalen von mehreren Monaten führt, wobei letzteres auf physikalischer Ebene durch den Zufluss an Schmelzwasser mit niedrigem Salzgehalt erklärt werden kann. Zusätzlich zeigen unsere Resultate, dass eine Schwächung der AMOC zu kleineren Zuwächsen an Meereis führt, was den physikalischen Erwartungen eines niedrigeren Wärmetransports in die Arktis durch die AMOC entspricht. Der destabilisierende Effekt der Schmelze des Arktischen Meereises auf die AMOC unterstreicht die Gefahr, die von interagierenden Kippelementen schon in den kommenden Dekaden ausgeht.

Contents

1	Introduction	1
2	Background	3
2.1	Tipping Point Interactions	3
2.1.1	Overview of Tipping Point Interactions	3
2.1.2	Previous Results	5
2.1.3	Tipping Point Models	8
2.2	Causal Inference	11
2.2.1	Assumptions	12
2.2.2	Peter-Clark Momentary Conditional Independence (PCMCI)	14
2.2.3	Liang-Kleeman Information Flow (LKIF)	15
2.2.4	Granger Causality for State Space Models (GCSS)	16
2.2.5	Neural Network Methods	18
3	Data-driven Causal Inference Pipeline	21
3.1	Data Selection	21
3.2	Preprocessing	21
3.3	Method Configuration	23
3.4	Experiment Types	24
3.5	Visualization	24
4	Experiments	27
4.1	Experiment Setup	27
4.1.1	Multivariate Experiment Setup	27
4.1.2	Grid Experiment Setup	33
4.1.3	Applied Experiment Setup	35
4.2	Results for Multivariate Systems	39
4.3	Results for Grid Systems	47
4.4	Application to Tipping Point Interactions	52
4.4.1	Results	52
4.4.2	Discussion	58
5	Related Work	63
6	Conclusions	67
List of Acronyms		69

Bibliography	71
Appendix	81
A1 Method Comparison for Slow Bidirectional Grid Systems	81
A2 Robustness Tests for Interaction of Arctic Sea Ice and AMOC	82

1 Introduction

As global temperatures reach a record high in 2024, with most measurements estimating average temperatures at more than 1.5°C above pre-industrial levels [1, 2], the urgency of reliable risk assessments of the consequences of climate change rises ever further. Over recent decades, climate research has identified several climate tipping points – elements of the Earth system that react to global warming in a nonlinear way, exhibiting a sudden shift to an entirely different state once a threshold is crossed [3, 4, 5]. The tipping process is driven by positive feedback loops that could remain active even under a cooling climate, making it practically irreversible on human time scales.

Climate tipping elements include cryosphere elements like the polar ice sheets that could melt irreversibly [6, 7], oceanic circulations like the Atlantic Meridional Overturning Circulation (AMOC) that could weaken severely [8, 9] and biosphere elements like the Amazon rainforest that could experience a dieback [10]. These examples illustrate both the scale of impacts of potential tipping events and the urgency of research towards a quantification of their thresholds and consequences. Current estimates of thresholds are still subject to large uncertainties, but show several tipping points in the range of current warming levels [4].

Climate tipping elements can also interact with each other through teleconnections, i.e., connections over large spatial scales. Through such teleconnections, a tipping event could stabilize or destabilize other tipping elements by changing the thresholds at which tipping processes set in, or could even lead to a "domino effect" of multiple tipping events [11]. Tipping point interactions therefore further increase the potential impacts of tipping events and add another layer of uncertainty.

Sources of knowledge about tipping point interactions are limited as Earth system models would have to accurately include processes from different fields like oceanic and atmospheric circulations, ice sheets and the biosphere in order to conduct experiments. High complexity models are often limited to atmospheric and oceanic circulations, although comprehensive models of lower complexity could help to fill this gap [12]. Paleoclimatic evidence and expert assessments are other main contributors to the current state of knowledge about tipping point interactions [13].

Causal inference methods can detect and quantify teleconnections from observational data, adding a statistical and data-driven perspective to existing results [14, 15]. Assuming that teleconnections remain similar in scale and direction upon tipping, we can derive tipping point interactions from teleconnections between tipping elements. While the work with observational data avoids the uncertainty introduced by model assumptions and constraints, the use of statistical tools also introduces uncertainty.

To constrain and estimate aleatoric uncertainty (i.e., one that is introduced by data sampling), we conduct a systematic evaluation of three causal inference algorithms un-

1 Introduction

der conditions expected for the application to climate elements and create a data-driven causal inference pipeline. We compare the Peter-Clark Momentary Conditional Independence (PCMCI) method [16], which has frequently been used in climate analysis [15, 17], the Liang-Kleeman Information Flow (LKIF) [18] and the Granger Causality for State Space Models (GCSS) method [19], the latter two have seen fewer applications in climate research so far [20].

We then apply these causal inference algorithms to observational and reanalysis data at different spatial resolutions on Arctic sea ice (regarding its extent, concentration, albedo and thickness), the overturning strength of the AMOC and precipitation over the Southern Amazon Rainforest (SAR) to detect and quantify a tipping point interaction network and discuss the potential physical mechanisms of the identified teleconnections.

This work is structured as follows:

- Chapter 2 provides more context to climate tipping elements, tipping point interactions and how they are conceptualized for causal inference, and explains the algorithms analyzed in this thesis.
- In Chapter 3, we describe the full pipeline from observational time series data to a detailed causal analysis with multiple robustness checks.
- Chapter 4 describes the experiments conducted in this thesis, where we first compare the performance of the causal methods on different types of synthetic data and then apply them to a network of climate tipping elements.
- Chapter 5 presents related work on tipping point interactions and causal methods.
- Finally, Chapter 6 draws conclusions on the presented work and gives an overview of further research directions.

2 Background

This thesis deals with causal inference methods from a theoretical and an applied perspective. The application to climate tipping point interactions offers insights into requirements and assumptions for causal methods and therefore informs our choice of algorithms and the experiments conducted to compare the methods.

We first introduce an overview of tipping point interactions, previous results on the tipping elements analyzed in this work, and some common assumptions in simple tipping point models.

In Section 2.2, the theoretical background of causal discovery methods is explained and formalized, and several causal methods are introduced with their corresponding models and concepts of causality.

2.1 Tipping Point Interactions

In the following sections, we give an explanation of tipping points and their interactions, and describe obstacles and pathways for further research on tipping point interactions. Further, we give an overview over the current state of research on dynamics of the AMOC and Arctic sea ice, on their tipping points and their interactions, based largely on model studies and observational knowledge. Lastly, we introduce simple tipping point models that have been utilized for a low-level description of tipping dynamics, and that allow for a derivation of tipping point interactions from causal methods.

2.1.1 Overview of Tipping Point Interactions

Tipping points are usually defined by abrupt transitions in the state of a system due to a small change in some driving parameter [3] and can occur in several types of systems, including social and ecological ones [21, 22, 23, 24, 25]. Once a tipping point is crossed, positive feedback loops self-amplify the development of the system away from the previous equilibrium, for physical systems this process usually results in a second stable equilibrium, making the transition difficult to reverse.

Over the last two decades, the concept of tipping points has been applied to large-scale elements of the climate system, identifying a number of systems that would likely cross a tipping point under continued global warming with severe impacts on ecosystems and the global climate system [3, 4]. Efforts to determine the sensitivity of tipping elements to climate change, the time scales of tipping points and potential consequences of tipping events have since confirmed the severe risks posed by climate tipping elements within this century [5, 26].

2 Background

There are two main reasons to hypothesize that climate tipping elements can have an impact on each other and that the crossing of one tipping point may set other tipping processes in motion or make climate elements more likely to tip under climate change:

Firstly, previous research in Earth system science has established that elements of the climate system are connected on large scales by atmospheric and oceanic circulations, which can store and transport precipitation and energy over different time scales, including oscillations on decadal to multi-decadal time scales [27]. When these connections span continental scales, we refer to them as *teleconnections*, the most famous example being the El Niño Southern Oscillation (ENSO). It is observed as multi-year phases of anomalies to the sea surface temperatures (SSTs) in the East Pacific Ocean, called El Niño for phases of warm anomalies and El Niña for cold ones. It has been established that these events are causally responsible for shifts in precipitation patterns in Africa, Asia and the Americas, i.e., teleconnections were determined through analysis of observational data and model experiments [28, 29].

Secondly, tipping points have been identified in some of the core mechanisms of the climate system like the ice sheets and the AMOC, making it plausible that tipping elements would also have significant and measurable impacts on each other through teleconnections.

The body of research on such climate tipping point interactions has grown from expert assessments [31, 32] to a variety of model-based studies [30], and has informed policy advice [33]. Figure 2.1 shows an overview of climate tipping point interactions according to a recent review, which draws from all types of sources mentioned above and finds mostly destabilizing teleconnections between established tipping elements. Uncertainties on climate tipping elements in general and on their interactions especially remain relatively high for a number of reasons:

1. Time scales: Many tipping processes operate on larger time scales (decades to centuries, some up to millennia) than the weather (on a scale of days) or changes to the climate (years), which means the period of modern observations does not cover a sufficiently long evolution of these systems and we can not easily test predictions or fit detailed models as can be done with e.g., weather forecasts.
2. Limited knowledge on tipped states: Although some tipping elements like the Greenland ice sheet or the AMOC might already approach a tipping point [6, 9, 34], our knowledge on their tipped states mainly stems from paleo records. These were produced under different climatic conditions than today, which makes it difficult to reliably determine thresholds under current circumstances.
3. Model limitations: State-of-the-art climate models as used in the Coupled Model Intercomparison Project (CMIP, used e.g., in the assessment reports by the Intergovernmental Panel on Climate Change (IPCC)) do not reproduce the observed trends in tipping elements like the AMOC [35] and do not include cryosphere and vegetation tipping elements like the polar ice sheets and the Amazon rainforest. Although efforts of coupling different model types to research tipping phenomena

2.1 Tipping Point Interactions

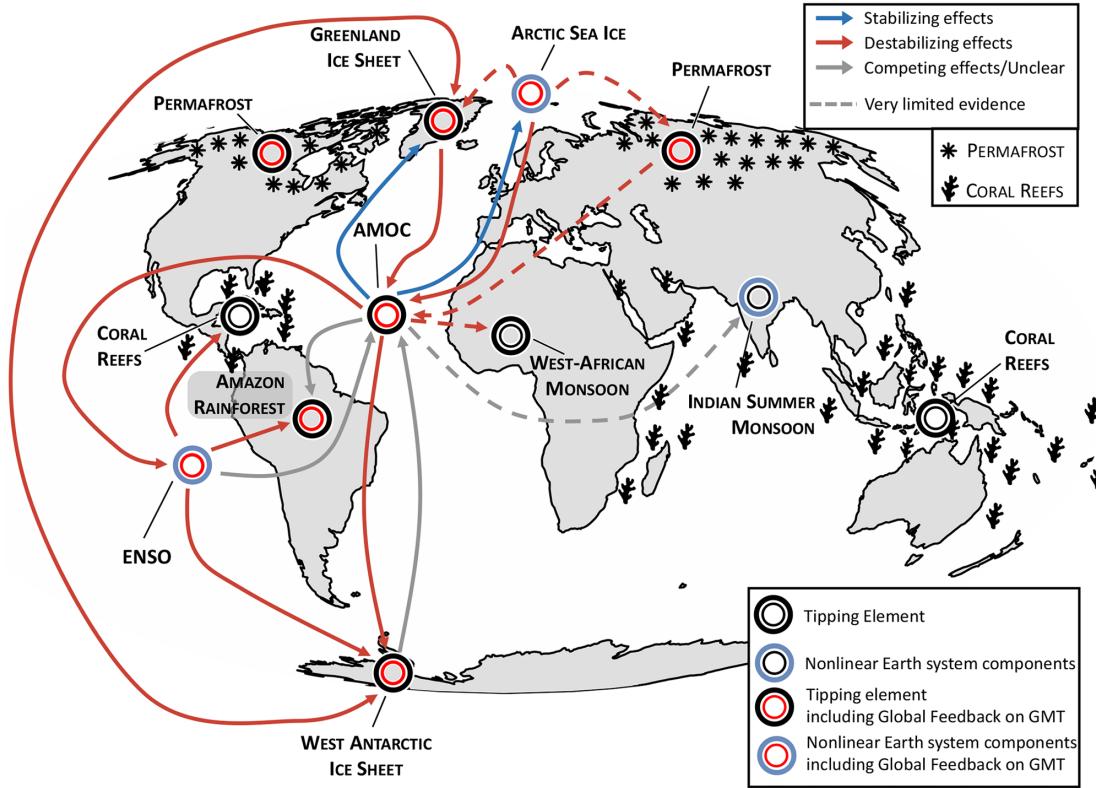


Figure 2.1: Climate tipping elements and nonlinear Earth system components with stabilizing and destabilizing interactions according to a recent review [30]. Figure from Wunderling/von der Heydt et al. [30].

have increased over the past decades, uncertainties remain large with respect to tipping thresholds [4] and even more so for interactions [30, 31].

In this work, we approach the interactions of climate tipping elements through the teleconnections that connect them today and over the past decades of observations. We derive the direction and an estimate of the magnitude of teleconnections from observational and reanalysis data of the past decades. Assuming that teleconnections are not changed qualitatively due to a tipping event, we can estimate the impact of a tipping event on other climate tipping elements.

2.1.2 Previous Results

Looking at tipping point interactions more concretely, in this work we focus on interactions between the AMOC and Arctic sea ice and extend previous results with causal methods that include the Amazon rainforest in the analysis [36].

Arctic sea ice is typically not considered a tipping element in the strict sense [5], because it lacks feedback mechanisms that would prevent regrowth of sea ice in winter or under a

2 Background

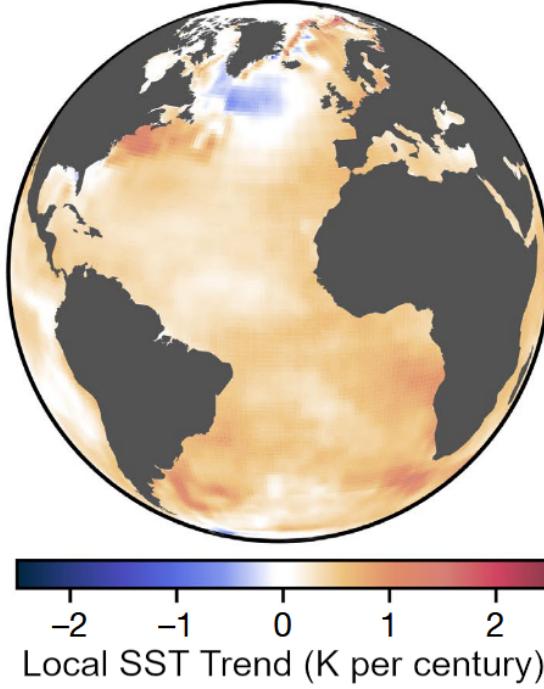


Figure 2.2: The cold anomaly of sea surface temperature in the North Atlantic in the trend from 1870 to 2016. Figure from Caesar et al. [8].

cooling global climate. However, there do exist mechanisms that introduce nonlinearity to sea ice melting processes, such as the ice-albedo feedback by which melting of sea ice reduces surface albedo (i.e., sunlight reflectivity) and thereby increases warming. An ice-free state of the Arctic Ocean in summer months can already be expected within the coming years to decades, with impacts on ecosystems [37], geopolitics [38], and shifts in climate patterns and interactions [39].

The AMOC refers to a circulation system in the Atlantic Ocean that carries warmer surface water northwards, where it cools down, sinks and circles back southwards. The northward heat transport of the AMOC is largely responsible for the relatively warm temperatures over Europe and the Northern hemisphere overall [40]. The AMOC is considered a tipping element that can transition into a second stable state with weak or no overturning, which would lead to a reorganization of climate patterns, especially in Europe [5]. Evidence is now pointing towards an AMOC weakening observed over the past decades, with an anomaly of lower surface temperatures in the North Atlantic in a warming climate as a main indicator [8, 9, 41], shown in Figure 2.2, although it remains uncertain whether tipping is imminent in this century [42, 43].

The North Atlantic Ocean and sub-arctic seas, namely the Greenland, Barents, Irminger and Labrador Seas, are crucial to the downwelling mechanism of the AMOC. Warm and salty surface water from lower latitudes is transported into these colder regions where it loses heat to the atmosphere, which reduces its buoyancy. These denser water masses

2.1 Tipping Point Interactions

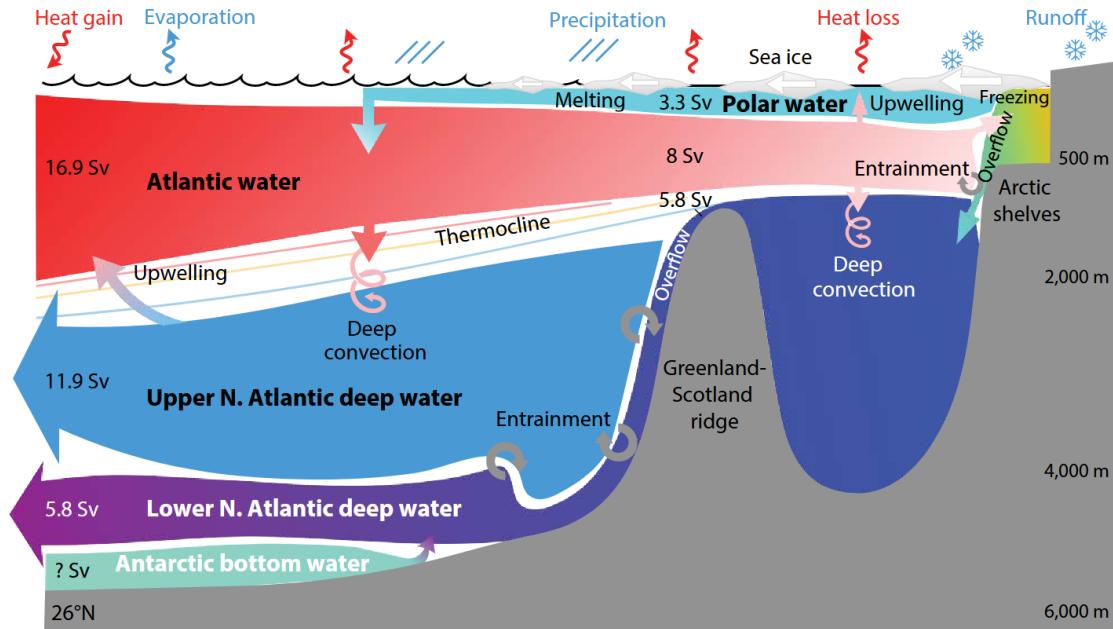


Figure 2.3: Physical dynamics in the North Atlantic and Arctic Ocean are responsible for the overturning mechanism of the AMOC, with a central role for Arctic sea ice in the Arctic Ocean and adjacent subpolar seas. Figure from Weijer et al. [44].

then sink to lower depths of 1,000 – 3,000 m and are transported southwards as North Atlantic Deep Water, driving the AMOC [44]. In the summer, sea ice melting in the Arctic Ocean contributes an inflow of surface freshwater especially into the Greenland Sea, which reduces salinity and thereby density. In the winter, sea ice formation increases surface layer salinity through brine rejection (i.e., salt remains in water when sea ice is formed) [45], which may contribute to the downwelling mechanism.

The effects of a weakening AMOC on Arctic sea ice have been estimated through model simulations and statistical analysis, with some evidence pointing towards a reduction in heat transport into the Arctic and subarctic seas [44, 46, 47], while other authors find a decoupling of this heat transport from the AMOC [48]. A reduced heat transport could be expected to increase sea ice production or reduce melting rates on annual [47] to decadal [44] time scales. However, AMOC weakening would likely be outweighed by a global warming-induced increase in heat transport to the Arctic [44, 49].

The impact of Arctic sea ice on the AMOC is subject to a number of competing effects and explanations. Figure 2.3 provides a conceptual overview of dynamics contributing to the AMOC convection in the North Atlantic and Arctic regions, with estimates for the volume transport given in Sverdrup ($1 \text{ Sv} = 10^6 \frac{\text{m}^3}{\text{s}}$).

- A retreat in sea ice as currently observed could move the deep convection zones of the AMOC northward (i.e., areas where water masses sink and produce North

2 Background

Atlantic Deep Water), which could change the overturning mechanism in complex and unforeseen ways [44]. Wu et al. [50] find that convection would be increased (i.e., the AMOC strengthened) in regions of retreating sea ice edges, because more water is exposed to the atmosphere and the resulting heat loss, but that global warming likely outweighs this effect.

- In contrast to this mechanism, other model and observational studies find that a retreat of sea ice would lead to increased radiative warming (instead of heat loss) in the Arctic Ocean, which in turn would reduce sea ice and freshen the North Atlantic on a decadal to multi-decadal time scale [51, 52]. A number of studies find a weakening of the AMOC at different magnitudes and time scales due to surface freshening induced by sea ice melt [49, 53, 54].

In this thesis, we build on existing work by Högner et al. [36] who identify a robust causal network that indicates a stabilizing interaction from the AMOC to the SAR, by which a weakened AMOC would increase precipitation over the SAR in its dry season. They use monthly observational data to determine proxy indicators of the corresponding climate tipping elements, like sea surface temperature fingerprints of the AMOC by Caesar et al. [8] and Ditlevsen & Ditlevsen [9], and vegetation indices for the Amazon rainforest. With the Caribbean Low Level Jet (CLLJ) as a mediator, they apply a causal method to detect a link from the AMOC to the CLLJ, and from there to the precipitation levels over the SAR, which in turn lead to vegetation density and greenness.

They hypothesize that the North Atlantic Oscillation serves as a mediator bringing increased moisture to Central America, and that the Intertropical Convergence Zone (ITCZ) and the corresponding precipitation are shifted southward by a stronger CLLJ. Overall, the physical explanations and experimental results imply an inversely proportional causal effect, i.e., a weakening of the AMOC would increase precipitation, with a hypothesized pathway through atmospheric teleconnections. Other recent model studies agree with this southward shift of precipitation and derive stabilizing and destabilizing effects of an AMOC collapse on different regions of the Amazon rainforest [55, 56].

We extend this causal network with data on Arctic sea ice and follow similar steps to test the robustness of our findings. A detailed overview of our approach to the application of causal methods is found in Section 4.1.3.

2.1.3 Tipping Point Models

The study of tipping phenomena from a mathematical perspective is usually conducted on simple differential equations that can exhibit bifurcations, i.e., they can lead to qualitatively different states for some parameter λ , where tipping between the states may be induced by the bifurcation point, by noise or by the rate of change of λ [57]. This branch of research on conceptual equations was applied more concretely to a parameterized ensemble of coupled cusp bifurcations to determine potential tipping cascades [11]. The advantage of these simplified model equations lies in their low number of free parameters and thereby a high level of control. Therefore, findings from models, paleo data and expert assessments can be integrated directly by changing the expected range of tipping

2.1 Tipping Point Interactions

thresholds and new stable states in the model (instead of parameterizing the myriad of underlying physical phenomena in a complex model to fit the existing knowledge).

The simplest differential equation of order one exhibiting two stable equilibria is a cubic one, where two parameters a, b control the shape of the curve, a parameter λ introduces external forcing and a Wiener process dW applies noise to the system:

$$\frac{dx}{dt} = ax^3 + bx + \lambda + dW. \quad (2.1)$$

For a deterministic system with $a < 0, b > 0$ and $\lambda = 0$, the equilibria $f(x) = ax^3 + bx = 0$ are at:

$$x_{e1} = 0, \quad x_{e2} = \sqrt{\frac{-b}{a}}, \quad x_{e3} = -\sqrt{\frac{-b}{a}}.$$

By linearizing around the equilibria, one can determine that x_{e2}, x_{e3} are asymptotically stable, while x_{e1} is not. If the forcing crosses a critical threshold $\lambda_{crit} = \pm 2\sqrt{\frac{1}{b}(\frac{a}{3})^3}$, one of the stable equilibria disappears and the system tips into the remaining equilibrium state [58].

The coupling with other tipping elements is usually added as a linear term for simplicity. Additionally, most underlying physical mediators can be expected to follow a linear trend, e.g., the fraction of freshwater from melting processes that reaches the AMOC convection sites likely remains similar regardless of the concrete amount. We therefore extend Equation (2.1) with coupling terms, assuming that the model is in equilibrium x_{e2} ,

$$\frac{dx_i}{dt} = ax_i^3 + bx_i + \lambda + dW_i + \sum_{j \neq i} d_{i,j}(x_j - x_{e2}). \quad (2.2)$$

The goal of the applied part of this work can then be viewed as the detection and estimation of a linear coupling parameter $d_{i,j}$ for the AMOC and Arctic sea ice, in either direction. For this estimation, one could attempt to conduct a model fit of the coupled cubic equations to observational data, but this introduces additional uncertainty and the strong assumption that the cubic equation describes the current behavior to the necessary extent.

We instead add the assumption that the analyzed tipping elements remain around their initial stable equilibrium. Around this equilibrium, we can linearize the differential equation accurately. As the coupling coefficients are already assumed linear, this results in a linear stochastic differential equation.

In climate tipping elements, the forcing λ includes impacts of climate change, which have likely shifted the equilibria over the past decades. As long as the forcing changes linearly and this shift of equilibria is sufficiently small, we can approximate the shift to be linear as well. One can then linearly detrend observational data to account for the changing external forcing. This way, we only need to assume that the tipping element does not exhibit a strongly nonlinear response to forcing yet. We therefore neglect the external forcing λ in the following derivations.

2 Background

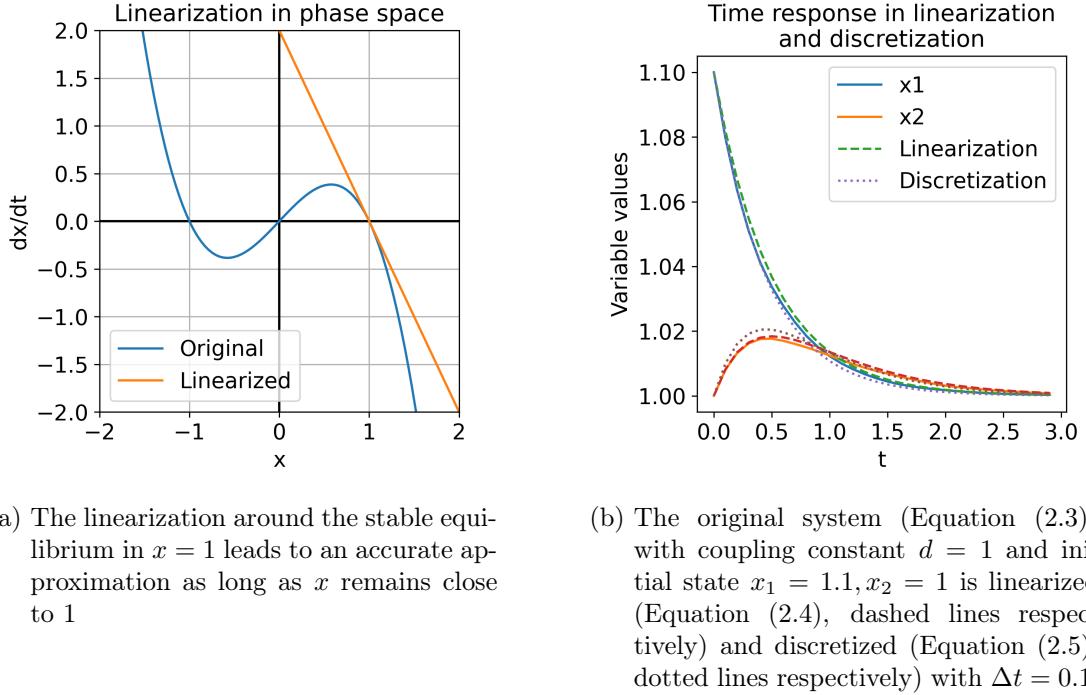


Figure 2.4: Linearization and discretization of cubic differential equations can approximate system behavior around stable equilibria

Suppose a simple coupled system with two unidirectionally coupled variables:

$$\begin{aligned} \frac{dx_1}{dt} &= -x_1^3 + x_1 + dW_1 \\ \frac{dx_2}{dt} &= -x_2^3 + x_2 + d(x_1 - 1) + dW_2. \end{aligned} \quad (2.3)$$

We linearize it around the equilibrium $[1, 1]$, with variables $\Delta x_1, \Delta x_2$ denoting the deviation from the equilibrium,

$$\begin{aligned} \frac{d\Delta x_1}{dt} &= -2\Delta x_1 + dW_1 \\ \frac{d\Delta x_2}{dt} &= -2\Delta x_2 + d\Delta x_1 + dW_2. \end{aligned} \quad (2.4)$$

Stability analysis following Lyapunov [59] shows that this unidirectionally coupled system remains asymptotically stable, in this case with two eigenvalues at -2 . Figure 2.4a visualizes the linearization in phase space.

From these linear differential equations, we can construct a time-discrete vector autoregressive (VAR) system, which takes a variable one time step forward by a weighted addition of its previous value and its coupled variables. For the sake of simplicity, we

here write this discretization with the Euler forward integration method and leave out the stochastic terms. With ever smaller time steps Δt , this discretization can estimate the original continuous system with arbitrary accuracy:

$$\begin{aligned}\Delta x_{1,t+\Delta t} &= \Delta x_{1,t} + \Delta t(-2\Delta x_{1,t}) \\ \Delta x_{2,t+\Delta t} &= \Delta x_{2,t} + \Delta t(-2\Delta x_{2,t} + d\Delta x_{1,t}) \\ \Delta X_{t+\Delta t} &= \begin{bmatrix} 1 - 2\Delta t & 0 \\ d\Delta t & 1 - 2\Delta t \end{bmatrix} \Delta X_t.\end{aligned}\tag{2.5}$$

Figure 2.4b shows that a discretization with $\Delta t = 0.1$ and $d = 1$ can approximate the deterministic part of the cubic and linearized equations well when variables remain around their equilibria. Assuming that our observations form a time discretization roughly on the time scale of the interactions, a linear model fit to observations can estimate the interaction coefficients d . The causal methods presented in the next section require less strict assumptions, but build on some of the described linearization and discretization concepts.

2.2 Causal Inference

Causal discovery algorithms use statistical methods to infer causal networks in multivariate settings from time series data. The broader term of causal inference can also refer to bivariate settings, intervention models, or samples without spatio-temporal connections.

Our definition of a causal discovery algorithm is a method that determines causal effects between variables (each of which provides a time series of its values) with significance testing, such that all edges of a resulting network are significant at some user-chosen error rate α . A second related area is that of causal effect estimation. Here, a causal network and corresponding time series data is given, and the effect strength of each edge in the network is estimated. Some of the presented algorithms solve these two problems in a combined manner, while others keep them separated or only fulfill parts of the mentioned conditions.

We provide a formal definition of a causal discovery method, but not of causal effect estimation methods. The different underlying concepts of causality do not agree on what a "correct" causal effect strength would be. Therefore, we include the effect strength in our definitions only for the sake of completeness and do not set any restrictions on its choice.

Given k time series, each of length n , such that

$$\begin{aligned}X &= \{X^1, X^2, \dots, X^k\} \\ X^i &= \{x_1^i, x_2^i, \dots, x_n^i\} \\ x_j^i &\in \mathbb{R},\end{aligned}$$

an error rate α , and a maximal time lag τ_{max} , we define a group of functions called *Causal Discovery Methods* to contain any function f that from these inputs determines a graph

2 Background

G . This graph encodes the causal effect network consisting of vertices $V = \{v_1 \dots v_k\}$ with v_i corresponding to the time series X^i and directed edges E encoding the causal parent and child node, the time lag l of the causal effect, the causal effect strength s and the p-value of statistical significance p .

$$\begin{aligned} f(X, \alpha, \tau_{max}) &= G \\ G &= (V, E) \\ V &= \{v_1 \dots v_k\} \\ E &= (i, j, l, s, p \mid i, j \in V; l \in \{0 \dots \tau_{max}\}; s, p \in \mathbb{R}) \end{aligned} \tag{2.6}$$

A ground truth graph G' for a dynamic system encodes the known causal effects between variables through its binary adjacency matrix A' . The definition of A' in our experiments is described in more detail in Section 4.1. We evaluate the prediction skill of a *Causal Discovery Method* f by comparing the adjacency matrix A of the graph G to A' :

$$A_{i,j} = \begin{cases} 1 & \text{if } \exists e \in E : e = (i, j, l, s, p \mid l \in \{0 \dots \tau_{max}\}; s, p \in \mathbb{R}) \\ 0 & \text{else.} \end{cases}$$

For our following experiments, different evaluations of the adjacency matrices are used, see Section 4.1 for these metrics and the reasoning behind them.

2.2.1 Assumptions

We refer to causal graphs as Structural Causal Models (SCMs), where directed edges encode a causality direction from cause to effect. A fundamental assumption of the problem statement in Equation (2.6) is that a graph may encode the causal relationships of variables and that the observations allow for a reconstruction of those relationships.

We define and formalize several assumptions:

- *Causal Stationarity*: The causal structure is time-invariant, so that one graph may encode it over the entire duration of the time series.
- *Causal Sufficiency*: There are no hidden confounders or other missing observations such that the observed graph misrepresents the real causal effects.
- The *Causal Markov Condition*: The real causal effects can be described by a graph, where separation in graph structure implies independence of the data.
- *Faithfulness*: Statistically independent variables must be separated in the graph.

For a more formal description, we assume a real process X with a theoretically unlimited amount of variables X^i and arbitrarily high temporal resolution, but discrete time steps. This process is assumed to operate forward in time, where causes precede effects:

$$X_t^i = f^i(X_{t-}^1, X_{t-}^2, \dots, X_{t-}^n, t) \tag{2.7}$$

with $X_{t-}^j = X_0^j, \dots, X_{t-1}^j$ denoting all past information of a variable and arbitrarily defined functions f^i (which may include stochastic effects).

We define causality as a change in probabilities of function outcomes. While holding all other variables in X constant, a change in the causal parent must influence the probability distribution of the causal child. We denote a causal relationship with a double arrow:

$$X_{t-\tau}^i \Rightarrow X_t^j \leftrightarrow \exists X, a, b, c : p(f^j(X_{t-\tau}^i = a, X, t) = c) \neq p(f^j(X_{t-\tau}^i = b, X, t) = c) \quad (2.8)$$

with p denoting the probability of an event.

For notation simplicity, we also define a more general causal effect between entire variables, which is used when the time lag is not decisive:

$$X^i \Rightarrow X^j \leftrightarrow \exists \tau : X_{t-\tau}^i \Rightarrow X_t^j \quad (2.9)$$

We firstly formalize *Causal Stationarity*, i.e., an existing causal effect operating with a time lag τ remains with the same lag at all time steps (though differing effect strength is possible).

$$\exists t \in \mathbb{N} : X_{t-\tau}^i \Rightarrow X_t^j \rightarrow \forall t' \in \mathbb{N} : X_{t'-\tau}^i \Rightarrow X_{t'}^j \quad (2.10)$$

To determine causality, we make observations which we assume to capture single variables of the underlying process, but potentially at a lower time resolution: $Y = \{Y^1, \dots, Y^k\}, Y^i \subseteq X^j, k \leq n$. We may have fewer observation variables than process variables. With this definition, we can formalize *Causal Sufficiency*, i.e., the observed variables must be sufficient to explain the underlying causal effects.

$$\begin{aligned} \#Y^i \in Y : Y^i \subseteq X^i \rightarrow |\{X^j \mid X^i \Rightarrow X^j\}| \leq 1 \\ (Y^i \subset X^i \wedge X_t^i \notin Y^i \wedge \exists t' < t : X_{t'}^j \Rightarrow X_t^i \wedge \exists t'' > t : X_t^i \Rightarrow X_{t''}^m) \\ \rightarrow t'' > \max(u, v) \end{aligned} \quad (2.11)$$

$$\text{where } u = \min_{v>t} (X_v^i \in Y^i), w = \min_{v>t} (X_v^m \in Y^m)$$

The first line in Equation (2.11) means that any unobserved information X^i may only be causal parent to one observed variable (otherwise it is referred to as a *Confounder*). The second line describes the case where X_t^i is a variable's unobserved information (through time sub-sampling). In this case, X_t^i may not propagate a causal effect before observations happen, otherwise X^i would not be detectable as a mediator (implying that observations were not sufficient).

We are left with the assumptions of the *Causal Markov Condition* and *Faithfulness*. These assumptions clarify the informal assumption that "a graph may encode the causal relationships". We first define some notation required for their formalization.

- We denote d-separation of nodes in a graph as $A \bowtie B|C$, where all paths between A and B are blocked under the conditioning set C , i.e., they do not have any other common ancestors or transitive connections beyond nodes in C – see Runge [60] for a more detailed explanation of d-separation.

2 Background

- We denote conditional independence of two variables as:

$$Y^i \perp\!\!\!\perp Y^j \mid Z \leftrightarrow \forall y_i, y_j, z : \\ p(Y^i = y_i, Y^j = y_j \mid Z = z) = p(Y^i = y_i \mid Z = z) \cdot p(Y^j = y_j \mid Z = z) \quad (2.12)$$

Equation (2.12) resembles the absence of a causal effect as defined in Equation (2.8), but with probability distributions derived from observational data, i.e., independence can be detected with some error rate α . This probabilistic independence formalization assumes some restrictions on the functions f^i of the process, e.g., they may not generate synchronized noise (which could make them conditionally dependent without any causal link in the dynamic system) [60].

The *Causal Markov Condition* requires that the process X follows a true causal graph such that a separation of two variables in the graph implies independence in the data. We formulate this condition in a time-dependent manner (but for arbitrary t) following Runge [60] to stress the role of the causal parents $P(X_t^j)$ of a node.

$$X_{t-} \setminus P(X_t^j) \bowtie X_t^j \mid P(X_t^j) \rightarrow X_{t-} \setminus P(X_t^j) \perp\!\!\!\perp X_t^j \mid P(X_t^j) \quad (2.13)$$

The *Faithfulness* assumption declares that the inverse must also be true: Independence in the distributions must be due to separation in the graph.

$$X_{t-} \setminus P(X_t^j) \perp\!\!\!\perp X_t^j \mid P(X_t^j) \rightarrow X_{t-} \setminus P(X_t^j) \bowtie X_t^j \mid P(X_t^j) \quad (2.14)$$

Note that this assumption puts significant limits on the process functions f^i , because deterministic variables and synergistic causal effects (where only the combination of variables leads to dependence) may violate faithfulness [60].

Combining all of these assumptions, we may infer the causal graph entirely from independence or dependence in the data, i.e., from observations.

2.2.2 Peter-Clark Momentary Conditional Independence (PCMCI)

The Peter-Clark Momentary Conditional Independence (PCMCI) method provides a framework around conditional independence tests [61] and requires all of the described assumptions for causal discovery. It operates on a temporally "unfolded" SCM where each time shift of a variable is a separate node up to a maximal time shift τ_{max} . An error rate α is provided for the independence tests, which functions as a significance level indicator. The framework consists of two phases for each variable:

- The PC phase starts with the full set of potential causal parents of a variable, $P(X_t^i) = \{X_{t-1} \dots X_{t-\tau_{max}}\}$, i.e., all observed variables in the time steps $\{t-1 \dots t-\tau_{max}\}$. In a first iteration, unconditional independence tests between X_t^i and each of its parents remove any independent nodes from P . In the following iterations, incrementing the iteration counter c , the c highest scoring (i.e., "most

dependent") parents of X_t^i are conditioned on, removing nodes that are independent at error rate α .

$$\begin{aligned} \forall X_{t-u}^j \in P(X_t^i) : & \text{if } X_t^i \perp\!\!\!\perp_{\alpha} X_{t-u}^j \mid \max_c(P(X_t^i)) \\ & \text{then remove } X_{t-u}^j \text{ from } P(X_t^i) \end{aligned} \quad (2.15)$$

Repeating this procedure for every variable at time step t leads to a set of parents for each variable. According to Causal Stationarity in Equation (2.8), we can assume that this set is time-invariant, i.e., $X_{t-u}^i \in P(X_t^j) \leftrightarrow X_{t-u-v}^i \in P(X_{t-v}^j)$

- The MCI phase addresses false positive control by conducting further conditional independence tests conditioned on the parents of both involved variables.

$$\begin{aligned} \forall X_{t-u}^j \in P(X_t^i) : & \text{if } X_t^i \perp\!\!\!\perp_{\alpha} X_{t-u}^j \mid P(X_t^i) \setminus \{X_{t-u}^j\}, P(X_{t-u}^j) \\ & \text{then remove } X_{t-u}^j \text{ from } P(X_t^i) \end{aligned} \quad (2.16)$$

There are several independence tests available for use with the PCMCI framework, some with further assumptions about the generating process. The choice of an independence test largely depends on expert knowledge and data availability, where model-free approaches are more data-intensive [60].

The simplest independence test is a partial correlation test, which is used throughout this work. For some test $X \perp\!\!\!\perp_{\alpha} Y \mid Z$, the partial correlation test first conducts a linear ordinary least squares regression of X and Y on the conditions Z and returns the corresponding residuals e_X, e_Y .

The Pearson correlation coefficient $\rho(e_X, e_Y)$ of the residuals is considered the effect strength. It is tested against a Student's t-distribution to determine a p-value indicating the likelihood that independent variables would correlate to this degree. The variables X, Y are considered conditionally independent if the p-value is above the error rate α .

If two variables remain conditionally dependent throughout the PC phase and their p-value in the MCI phase is smaller than α , their nodes are connected by a graph edge corresponding to Equation 2.6, with the Pearson correlation coefficient as the effect strength s and the corresponding p-value p derived from the t-distribution.

The PCMCI package also includes an effect strength estimation tool called Linear Mediation, which uses linear regression to estimate the causal effect strength of interactions [62]. This results in an interpretable effect strength x (rather than the Pearson correlation coefficient used in PCMCI), by which an impulse of one standard deviation in the causal parent triggers a response of x standard deviations in the causal child.

2.2.3 Liang-Kleeman Information Flow (LKIF)

The Liang-Kleeman Information Flow (LKIF) method is an information-theoretic approach based on entropy transfer, where the contribution of a variable to the entropy of another variable is considered to be the causal effect strength of that link.

2 Background

As a general approach, the LKIF method performs a model fit to a stochastic linear differential equation as the underlying system. For some vector of variables X , this system is defined as

$$\dot{X} = AX + b + c \cdot dW. \quad (2.17)$$

It is clear that this form can fit exactly to the linearization discussed for the tipping point model in Equation (2.4). Equation (2.17) implicitly fixes the form of the data generating function f^i as defined in Equation (2.7), and data generated from the model would fulfill the *Causal Stationarity*, *Causal Markov* and *Faithfulness* conditions.

To determine entropy transfer as a measure of causality, the change in entropy $\frac{dH_1}{dt}$ is split up into influences from the (potential) causal parent $T_{2 \rightarrow 1}$ and other remaining factors [63],

$$T_{2 \rightarrow 1} = \frac{dH_1}{dt} - \frac{dH_{1 \setminus 2}}{dt}. \quad (2.18)$$

From the model definition in Equation (2.17), one can derive maximum likelihood estimators based on observed data [18], and solve the entropy terms in a closed form:

$$T_{j \rightarrow i} = a_{ij} \frac{C_{ij}}{C_{ii}} = \frac{1}{\det C} \cdot \sum_{k=1}^d \Delta_{jk} C_{k,di} \cdot \frac{C_{ij}}{C_{ii}}, \quad (2.19)$$

with a as entries of matrix A , C as the covariance matrix as observed from data, d the number of variables and $C_{k,di}$ the covariance between x_k and the Euler forward difference approximation of dx_i/dt .

The effect strength s as defined in Equation 2.6 is determined by a normalization of $T_{j \rightarrow i}$ with regard to the entropy H_i . LKIF does not estimate a p-value, but a confidence interval around the effect strength for some given error rate, which can be used for significance thresholding.

The assumption of an underlying linear differential model for several climate phenomena can appear rather far-reaching. However, as described in Section 2.1.3, any system that can be described by a first-order differential equation and remains in a stable equilibrium can be approximated by this linear differential model. In model comparisons the linearity assumption also performed well for simple nonlinear models [64].

It is important to note that the resulting causal effect strength is not as interpretable as with PCMCI, as LKIF only provides absolute measures of entropy transfer. Whether an increase in the causal parent leads to an increase or decrease in the causal child variable is not determined. The differential equation model also does not allow for a determination of the time lags of causal effects.

2.2.4 Granger Causality for State Space Models (GCSS)

Barnett and Seth conduct a model fit of a vector autoregressive moving average (VARMA) model from observational data [19]. VARMA models are time-discrete and include not only autoregressive effects, by which a variable can receive information from other variables with some time lag, but the moving average component also allows for a weighed impact of past noise on a current state.

One can formalize a VARMA model in state space, which we refer to as a state space model, hence the name of this causal method.

This VARMA state space model consists of a hidden state and an observation equation, each with noise. Such a state space model can generally be estimated from observational data y_t .

$$\begin{aligned} x_{t+1} &= Ax_t + u_t && \text{Hidden state transition} \\ y_t &= Cx_t + v_t && \text{Observation} \end{aligned} \quad (2.20)$$

It is easy to see how this hidden state could implement a moving average component simply by extending its dimension such that every noise value is shifted through variables in x up to the maximum lag at which it is relevant. Thus, the past noise values can influence results at multiple time steps as required by the VARMA model.

Granger causality is based on the notion that the causal effect of a variable on another is determined by the difference it would make to the affected variable if the past information of the parent variable was unknown. To achieve conditional Granger causality for some time series X, Y, Z (with arbitrary dimensionality), Geweke [65] proposes a comparison of variances:

$$\begin{aligned} F_{Y \rightarrow X|Z} &= \ln \left(\frac{|var(x_t | X_{t-}, Z_{t-})|}{|var(x_t | X_{t-}, Y_{t-}, Z_{t-})|} \right) \\ &= \ln \left(\frac{|\mathbb{E}\{(x_t - \mathbb{E}\{x_t | X_{t-}, Z_{t-}\})^2\}|}{|\mathbb{E}\{x_t - \mathbb{E}\{(x_t | X_{t-}, Y_{t-}, Z_{t-})^2\}\}|} \right) \end{aligned} \quad (2.21)$$

To determine Granger causality from Equation (2.21), Barnett and Seth transform the previous state space model into an *Innovations Form State Space Model* (IFSS) that aligns closely with the variances required for Granger causality in Equation (2.21), and they provide a way to determine a reduced model that does not contain the information of one of the observed variables [19].

We transform Equation 2.20 into the IFSS form.

$$\begin{aligned} z_{t+1} &= Az_t + K\epsilon_t && \text{Hidden state transition} \\ y_t &= Cz_t + \epsilon_t && \text{Reduced observation} \end{aligned} \quad (2.22)$$

Note that we can conceptualize the parameter ϵ_t in Equation (2.22) as an observation error deviating from the expectation, i.e., $\epsilon_t = y_t - \mathbb{E}\{y_t | y_{t-}\}$, which lets us derive the denominator in Equation (2.21) for every variable in y separately.

To derive the numerator in Equation (2.21), one needs a way to determine the same measure for a system without observations of our (potential) causal parent variable. A reduced model is constructed (Equation (2.23)), indicated by a superscript R , where that variable is removed from y , the corresponding matrix cells in C , and from the noise ϵ . Note that we do not change the hidden processes z by doing so, and y are otherwise fixed observations.

$$\begin{aligned} z_{t+1} &= Az_t + K\epsilon_t && \text{Hidden state transition} \\ y_t^R &= C^R z_t + \epsilon_t^R && \text{Reduced observation} \end{aligned} \quad (2.23)$$

2 Background

This set of equations is not an IFSS because of the reduced noise in the second term and needs to be transformed to IFSS again, where one can derive an ϵ^R shared by hidden and observed processes. Any changes in ϵ^R compared to the full model's ϵ are solely due to the removed information in observations – which may lead to a decrease in prediction ability (and therefore an increase in the variance of the residual ϵ^R). We can now derive the numerator of Equation (2.21) and thereby the measure of causality for a removed causal parent with respect to any other variable. Removing every variable separately then gives a full picture of all directional causalities.

The result of $F_{Y \rightarrow X|Z}$ is taken as the effect strength s in the resulting graph edge of Equation 2.6. The p-value is determined by a χ^2 -test [66], similarly to the t-test conducted for PCMCI significance.

The GCSS approach relies on modeling assumptions that might make it more data-intensive than simpler methods. Its effect strength estimates do not indicate a positive or negative influence of a variable, a time lag analysis is also not possible with the chosen implementation, although the method has been extended in this direction [67].

2.2.5 Neural Network Methods

Several neural network approaches are theoretically suited for causal graph identification. One can broadly categorize them into methods that learn a causal matrix from data (DAG-GNN, SAM), ones based on interventions and Granger causality (TCI-Net), and those reconstructing underlying differential equations, usually referred to as symbolic regression (SINDy).

DAG-GNN (Directed Acyclic Graph - Graph Neural Network) [68] takes an autoencoder approach by which a latent representation of a sample can be determined and the sample reconstructed. The latent representation is generated from a linear combination of nonlinear transformations of every variable. As both the nonlinear transformations and the linear combination matrix are learned from data, an informal intuition of causality is given by the linear combination as information transfer between some "ideally" transformed variables, so that this combination matrix encodes the causal graph. The method also enforces an acyclicity constraint which is not desirable for the SCMs we aim to discover.

Structural Agnostic Modelling (SAM) [69] employs adversarial learning where a generator module learns an underlying causal matrix to generate data that is as indistinguishable as possible from the real time series data. If the real data generation process broadly matches the one employed by the model, this matrix should indicate the causal graph edges after training. Assumptions for data generation match closely with those in Section 2.2.1, yet the inherent limitation to acyclic graphs makes it difficult to use in a causal analysis that may contain bidirectional causal effects at different time lags.

TCI-Net (Time Series Causal Inference Network) [70] uses an intervention-based approach in which a relatively simple neural network is trained to predict future time series values from a window of past values and covariates. The method then measures a treatment effect of one variable on others, which could be interpreted as a form of Granger causality. However, this method has to our knowledge not been extended to detect causal

2.2 Causal Inference

networks of direct and indirect links, and might encounter difficulties with the "removal" of a variable that Granger causality usually requires.

SINDy (Sparse Identification of Nonlinear Dynamics) [71] models the underlying dynamical system more explicitly by sparsely populating a large space of possible parameter combinations that make up the governing equations. It is trained by deriving (usually first-order) derivatives of variables from the variable values at the same time step. Here, the choice of available operators depends on expert knowledge of the system.

In our tests, none of the above approaches worked reliably even for simple graphs in the required ranges of samples and thus we excluded these methods from the experiments.

2 Background

3 Data-driven Causal Inference Pipeline

In the following, we present a pipeline for the application of causal inference algorithms that offers data preprocessing options, recommends and executes fitting algorithms (out of LKIF, PCMCI and GCSS) for some problem setting given by the user, and visualizes results. Figure 3.1 shows a conceptual overview of the different steps. Our implementation of this pipeline is publicly available along with the implementation of our experiments [72].

3.1 Data Selection

In a first step, the user selects suitable data sources from the file system, which contain the time series for the analyzed variables. The choice of variables is crucial to a successful application: One should include relevant variables such that the underlying system is described comprehensively with as few variables as possible. Additionally, the time span of observations needs to contain enough information to describe the underlying dynamic system (at least in some current equilibrium), with samples resolved on the time scale of interactions or higher resolution. All time series need to cover the same time span at an identical temporal resolution. For incomplete time series, one can either fill in values using interpolation methods or hide missing samples through masking (see Section 3.3).

In climate science, the selection of observation variables and regions that represent a proxy for some climatic process (e.g., sea surface temperatures in the subpolar gyre region as a proxy for the AMOC, see Section 4.1.3) is crucial for successful analysis of causal relationships. This region can be derived from expert knowledge on the physical system, or from data-driven dimension reduction of high-dimensional data [62].

3.2 Preprocessing

Several preprocessing steps are available in the pipeline to fix common obstacles to causal discovery in observational data or to approximate the derivative of a time series where it is assumed to be more causally relevant to the physical processes.

Linear detrending can be applied to a time series to remove a linear trend in the variable itself or in the mean of seasonal noise. In climate data, detrending is especially useful in variables that contain a climate change signal. As the climate forcing changes the stable equilibrium of the assumed differential model, detrending also implicitly removes the trend in the equilibrium position, therefore the linearization as conducted in Section 2.1.3 may still hold. However, for strong nonlinearities in the trend, more advanced approaches are required.

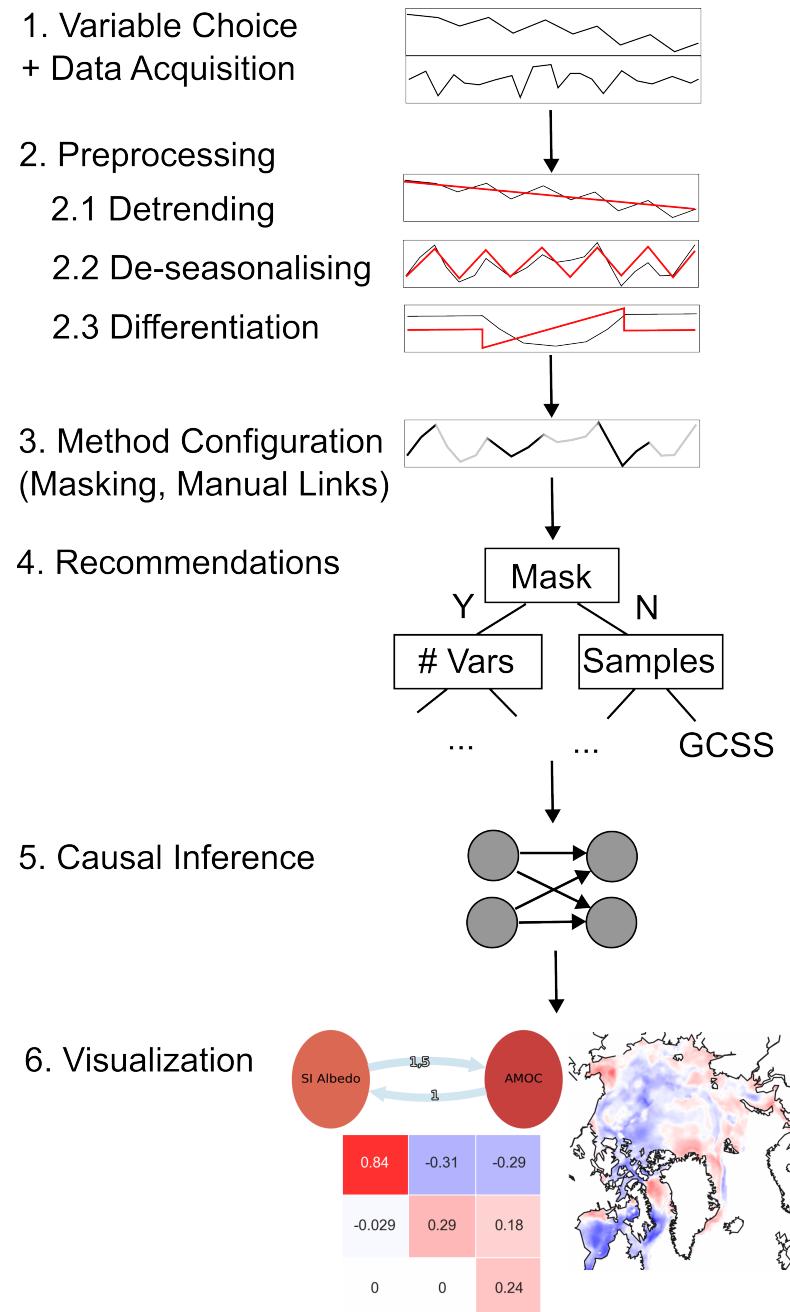


Figure 3.1: The causal analysis pipeline takes several configuration options from the user to run the different causal methods and creates visualizations from the results

De-seasonalising: Many climate variables from direct observations contain seasonal trends, e.g., summer and winter temperatures, precipitation, etc. One may treat this seasonal effect as noise, but it is firstly not white noise, but highly autocorrelated noise, and secondly a potential confounder if other variables also show seasonal effects. De-seasonalising here refers to the process of subtracting the average of each month’s measured values from the time series of that month respectively. Some climate data may already be available as anomaly values from some reference period, or may not experience seasonal effects, in which case the user should disable this option. Note that de-seasonalising also assumes stationarity in the seasonal noise statistics, i.e., that the mean and variance of seasonal effects remain constant.

Differentiation: Differentiation of time series data can be applied to receive an approximation of the first order derivative of some variable. In our implementation, we take the difference of a variable between its current and previous time step to approximate the local rate of change per time step. Differentiation also removes linear trends from the data, as linear trends only remain as a non-zero mean in the derivative. Previous work by Strahl [73] stresses the importance of domain knowledge in the decision whether to differentiate data. For example, the stability of the AMOC is usually dependent on freshwater influx in simple [74] and more complex models [42], i.e., the change in volume of ice masses is the cause rather than the volume itself. In this case, a differentiation of ice volume based on knowledge about the process would be beneficial for causal analysis. However, differentiation amplifies high-frequency noise, e.g., an impulse for one time step in the original signal is differentiated into a sharp peak and valley in two time steps, so it should only be applied where such phenomena are rare or low in amplitude.

3.3 Method Configuration

Masking: The user can apply masks to data to limit analysis to certain time periods, e.g., certain seasons, or to drop intervals with missing or erroneous samples. If the underlying nonlinearity is well known (e.g., in model studies), one can even choose to hide periods that exhibit nonlinearity or resemble a tipped system state in order to improve accuracy on the periods of system stability [73]. Masks are provided as separate time series of binary values. The GCSS algorithm does not support this functionality. In the LKIF algorithm, masked samples are dropped from the analysis entirely and only one mask can be applied to all variables. The PCMCI algorithm allows for a fine-grained specification of one mask per variable, and of the type of masking that is applied, where time points are only masked when analyzed as causal parents, children, or conditional variables. This is a useful technique for cases where an interaction may only occur in some season, but the causal parent or child show effects before or after the season as well.

Manual Link Prescription: The user might have previous knowledge about the underlying processes that guarantee or prohibit the existence of a causal link. In these cases, the user can choose to enforce or disable certain links manually. This option is only available in the PCMCI algorithm. The causal effect estimation with Linear Mediation

3 Data-driven Causal Inference Pipeline

also offers the option to be run on a manually prescribed causal network rather than on the result of PCMCI.

3.4 Experiment Types

Standard Execution: In the normal case, each process is described by one variable. The user is offered to run the analysis with all algorithms that fulfill the selected requirements, and is given a recommendation of the choice of algorithm depending on the method configuration choices and on the number of provided samples and variables. These recommendations are explained in more detail at the end of Section 4.2. For the PCMCI algorithm, the user can also choose between different conditional independence tests.

Spatial Resolution: In many applications, variables are available at some spatial resolution and we are interested in the distribution of some causal effect over this area. Users can add spatial data to the analysis by providing the time series values of cells and projection information for visualization. This analysis should only be conducted when there is sufficient evidence of an existing link to or from an aggregation of this variable and can also serve as a robustness test for previous findings. Results are only recorded between the spatially resolved variable and one other variable, while the remaining variables are included in the causal analysis but their effects are not recorded. For PCMCI, spatial analysis may derive effect strengths from a user-prescribed network using Linear Mediation while statistical significance is determined by whether PCMCI finds the link without the prescriptions.

Causal Stationarity: To test a network for robustness, it can be useful to analyze the strength of a particular causal effect over time, i.e., with a sliding window over the time series. This is referred to as *causal stationarity analysis*. It should only be applied to a link that is the result of previous analysis and can give insight into the robustness and the temporal trend of that causal effect. For PCMCI, one can choose whether a causal network should be prescribed (i.e., only Linear Mediation is applied), or PCMCI should first determine the network, as LKIF and GCSS would. The size of the sliding window and the step size can be determined by the user, e.g., for monthly data analyzed in a 20-year window, steps of one year or twelve samples could give a good overview of trends and significance of the causal link.

3.5 Visualization

For the standard analysis (i.e., not spatially resolved and no sliding windows), results are given as matrices with entries indicating the statistical significance (any link with strength $\neq 0$) and the estimated effect strength of interactions. These matrices are visualized directly as heatmaps, with colors encoding effect strength and variable labels for the corresponding causal parents and children. The results are also transformed into network graphs for easy readability, where edges point from causal parent nodes to children nodes. Edges are colored according to their effect strength and nodes according to their

3.5 Visualization

autocorrelative effect strength (as estimated by the causal method rather than statistical autocorrelation). In PCMCI analysis, edges are also labelled with their estimated time lag, corresponding to the time lag by which the causal parent influences the causal child in an (assumed) VAR model.

Spatial analysis plots the projection area of the spatially resolved variable, colors the spatial cells according to the interaction strength with the selected other variable, and indicates statistical significance by hatched areas.

In causal stationarity analysis, the effect strengths of the links selected by the user are plotted over time, with the step index of the sliding window as the label on the x-axis.

3 Data-driven Causal Inference Pipeline

4 Experiments

We conduct a series of experiments on synthetic and climate data. We describe the setup of several model systems and the experiments with which we evaluate the performance of the LKIF, PCMCI and GCSS causal methods. From these results, we derive recommendations on the optimal method under given parameter configurations and application cases, including spatially resolved data. Further, we apply suitable causal methods to time series of climate tipping elements to detect teleconnections. We conduct several robustness checks and discuss the results with regard to their significance and underlying mechanisms in the Earth system.

4.1 Experiment Setup

All experiments are conducted on a personal computer with an Intel i5-11400 CPU, 16GB of RAM, running Windows 11. We use Python 3.10 and Octave 9.2 for our implementation, with several libraries for the causal methods: We use PCMCI+ from the *Tigramite* package by Runge [75], the Python package *LK_Info_flow* by Rong following an implementation by Liang [18, 76], and the state space Granger causality toolbox by Barnett & Seth [19, 77] originally written in MATLAB is called from Python using an Octave interface. The significance test for GCSS is not contained in this toolbox and was implemented manually [72].

As described in Chapter 3, our pipeline should perform multivariate analysis with a given error rate and allow for spatially resolved variables. In this section, we explain the model systems we use for synthetic data generation and the concrete configuration of parameters in the experiments. We use one time-discrete and one time-continuous model for multivariate analysis and a time-discrete spatial grid model with a global coupling variable. With this data we test the causal methods under varying parameter configurations and compare their performance to determine recommendations for the choice of causal methods in our pipeline. We also describe the data sources and methods we use in the applied experiments on tipping point interactions between the AMOC and Arctic sea ice.

4.1.1 Multivariate Experiment Setup

This section describes the synthetic data generation models and the types of experiments conducted to evaluate the causal methods on systems where every process is described by one variable.

Data Generation: We generate multivariate synthetic data from two types of model systems. The first model is the coupled cubic differential equation model introduced in

4 Experiments

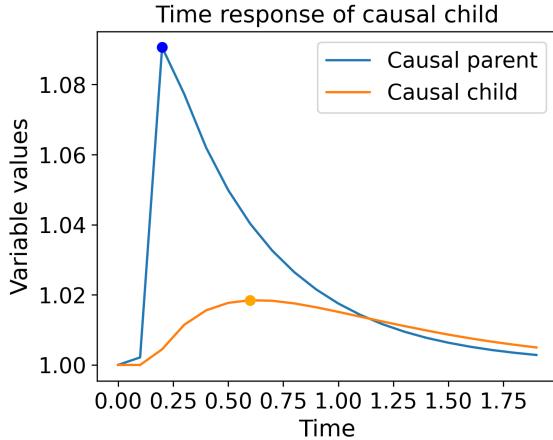


Figure 4.1: A noise of 1 is added to one variable for the time between 0.1 and 0.2, which leads to a peak at time step 0.2 with exponential falloff. The other variable follows more smoothly to a peak at time 0.6, then also falls off. Some small deviations e.g., at the time 0.1 are inaccuracies due to local non-differentiability induced by the unit pulse.

Section 2.1.3. In the multivariate setting with interacting variables used in the experiments, we extend the formalization with a *coupling matrix* D and fix the other parameters for simplicity.

With the simple parametrization from Equation 2.3, and given the vector of variables $X \in \mathbb{R}^n$, a vector of n independent Wiener processes dW , and variables starting in the equilibrium state at $x_i = 1$,

$$\frac{dX}{dt} = -X^3 + X + dW + D(X - 1), \text{ with } D \in \mathbb{R}^{n \times n}. \quad (4.1)$$

The Wiener noise vector is only updated at every time step for which we record the variable states, it is implemented as white noise. In our experiment, every recorded time step spans 0.1 time units. Since the system is integrated over time, this noise-induced constant disturbance for 0.1 time units forms a discrete Dirac impulse in the derivative. This impulse also causes immediate changes in the variable itself, which also means interactions can start without delay. However, interactions are based on the deviation of a variable from its stable state at value 1, so the maximum effect of an interaction is at the point when the disturbance ends. Figure 4.1 shows these steps in more detail.

To avoid additional autocorrelation effects, the diagonal entries of D are set to zero, while an entry $D_{i,j}$ describes the interaction of X_j to X_i . Since the magnitude of entries of D is not relative to the variable's magnitude, the matrix D does not contain the effect strength of an interaction as formalized by Granger causality or other forms of causality employed by the utilized causal methods.

The second model is a VAR system, a multivariate time-discrete linear model, in our case with additive white noise at every time step. It therefore resembles the discretized

4.1 Experiment Setup

model around a stable equilibrium in Equation 2.5. Although it is usually formalized with several orders of time lags up to some maximum lag, we restrict it here to effect delays of one time step:

Given a vector of variables $X_t \in \mathbb{R}^n$ and a white noise vector $\varepsilon \in \mathbb{R}^n$, the vector of variables is iteratively moved through time steps by

$$X_{t+1} = AX_t + b + c\varepsilon, \text{ with } b, c \in \mathbb{R}^n, A \in \mathbb{R}^{n \times n}. \quad (4.2)$$

The matrix A contains the autocorrelation component in the entries $A_{i,i}$, which specify the factor at which a variable's values are translated to the next time step. In the experiments, these diagonal entries are fixed to 0.5.

Non-diagonal entries of A determine the direct influence of a variable on another, i.e., a causal effect. These entries should quantitatively match the causal strength estimated by the Linear Mediation method, given a correct causal network.

For a qualitative analysis of causal effect detection, the matrices A and D are taken as a ground truth of interactions for VAR and cubic systems respectively. A non-zero entry of A or D implies the existence of a direct causal interaction irrespective of the effect's magnitude. Therefore, we create a binary ground truth matrix A' , as introduced in Section 2.2, where the entry 1 indicates a causal effect:

$$\forall i \neq j, A'_{i,j} = \begin{cases} 1 & \text{if } A_{i,j} \neq 0 \\ 0 & \text{else} \end{cases}$$

$$A'_{i,i} = 0.$$

We transform the coupling matrix D of cubic systems and the output matrices of causal methods in the same way (since all causal methods include statistical significance tests, insignificant entries remain at 0). The diagonal entries do not resemble interactions, so we exclude them from the evaluations. We can then compare the binary truth and output matrices with usual statistical metrics.

Metrics: A typical metric to determine a method's performance over different levels of statistical significance is the Receiver Operating Characteristics (ROC) curve, where the false positive rate and true positive rate are mapped to the X and Y coordinates respectively for every parameter setting for α . The area under this curve visualizes the performance of a method, i.e., a large area means there are better tradeoffs between true and false positive rates than for a small area.

For data parameters, we would like to receive a single score that contains information on all four basic metrics of true and false positives and true and false negatives, so that a method is rated not only on detecting true causal links but also on correctly identifying where links do not exist. A single score makes it much easier to compare the performance of the methods with the drawback that a custom weighing of false negatives against false positives is not possible anymore. The Matthews Correlation Coefficient (MCC) is considered a good candidate for such a combined metric in comparison to other frequently used metrics in binary classification [78].

4 Experiments

Given the four basic metrics, we determine the true positive rate, the false positive rate and the MCC in the following way:

$$TPR = \frac{TP}{TP + FN}$$

$$FPR = \frac{FP}{FP + TN}$$

$$MCC = \frac{TP \cdot TN - FP \cdot FN}{\sqrt{(TP + FP) \cdot (TP + FN) \cdot (TN + FP) \cdot (TN + FN)}}.$$

Concrete Model Systems: In our synthetic data generation process, we want to match systems we would expect for real analysis, and create model systems with three, six and twelve variables and varying numbers of interactions between them. Analogous to the graph setting with nodes for variables, we refer to the fraction of edges over nodes as the system’s *density*. For the experiment systems, we create models with the densities < 1 , 1 and 1.5 for each size of a system (in terms of variables). These systems are displayed in Figure 4.2.

Previous analysis on system density showed a loss in accuracy for a growing number of feedback loops for the PCMCI method [73], so we ensure feedback loops of different sizes in our denser model systems while mostly avoiding self-amplification effects. We construct such stable feedback loops through an uneven number of negative weights on the edges that form the loop.

Experiments: The strength of interactions is determined by the matrix entries of the corresponding model system. We refer to this factor as the *coupling strength* and set all interactions to some identical coupling strength for each experiment. In VAR systems, this factor determines a discrete addition of a variable’s value to another variable in the next time step, but in the differential equations, interactions are continuous, and so the coupling strength may be measured better in terms of how much a disturbance peak in one variable is translated to a causal child variable.

In a simple interaction with one noise impulse and coupling strengths set to 1, we can observe a weakening of the signal every time it is transmitted further, see Figure 4.3. As long as variable values stay reasonably close to the equilibrium, we can estimate that a coupling strength that prevents self-amplification in this simple example also ensures stability when noise is added every 0.1 time units. In the example, the relationship between the second and third variable of Figure 4.3 is the stricter criterion for stability rather than that between the first and second variable. The second variable deviates about 2.8 times further from 1 than the third variable, therefore our coupling strength should also stay below this value to ensure stability.

To estimate the isolated impact of some parameter like the coupling strength on the detection skill of causal methods, we keep other data parameters fixed to a default value. Besides the coupling strength and the underlying model system, we also vary the scale of the additive white noise and the number of analyzed data samples.

The only method parameter (in contrast to parameters changing the data generation) that all causal methods provide is a significance level, here conceptualized as an error level α .

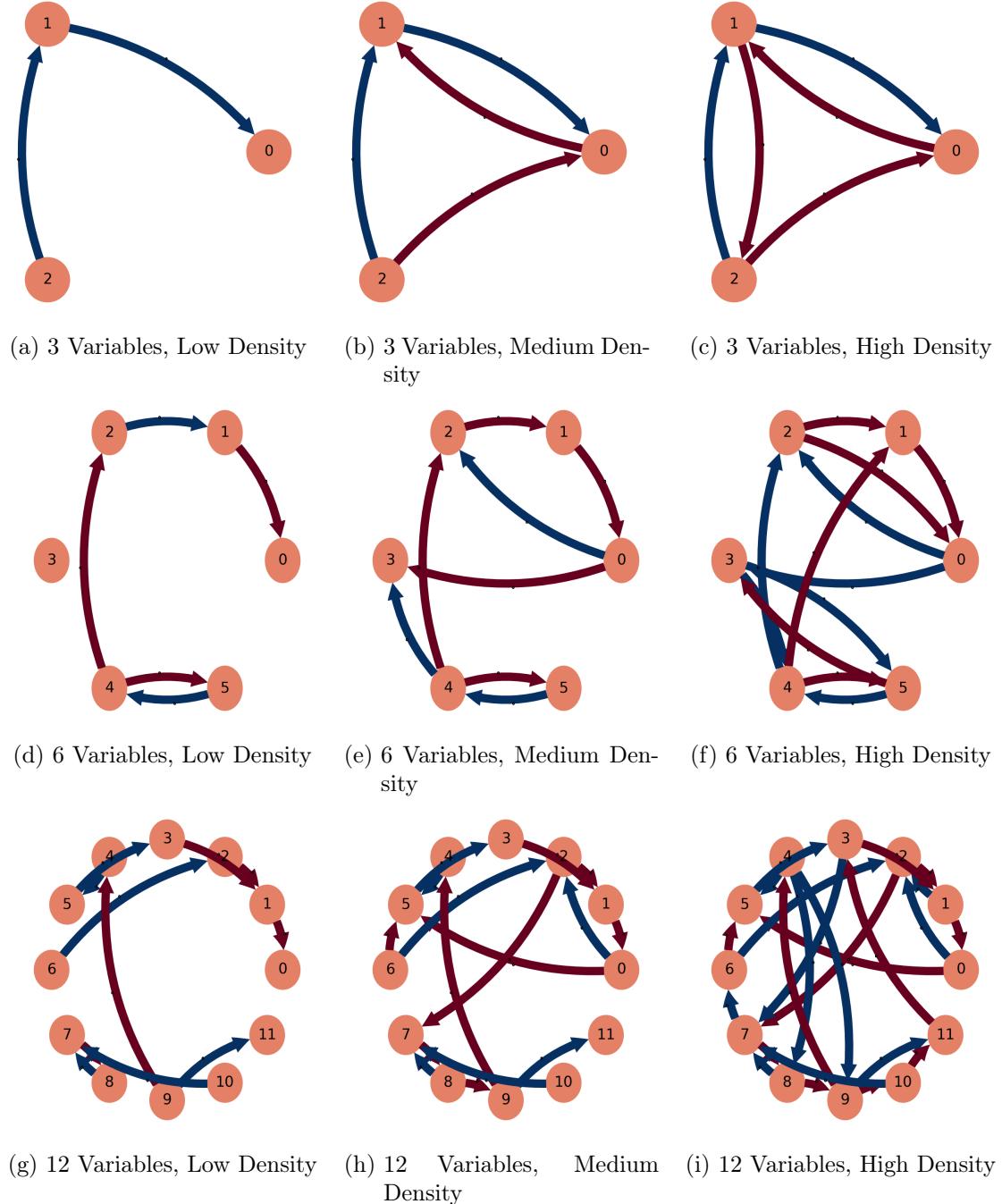


Figure 4.2: The different systems used for synthetic data generation, with varying numbers of variables and different density levels. Red edges indicate positive interaction strength, blue edges indicate negative interaction strength.

4 Experiments

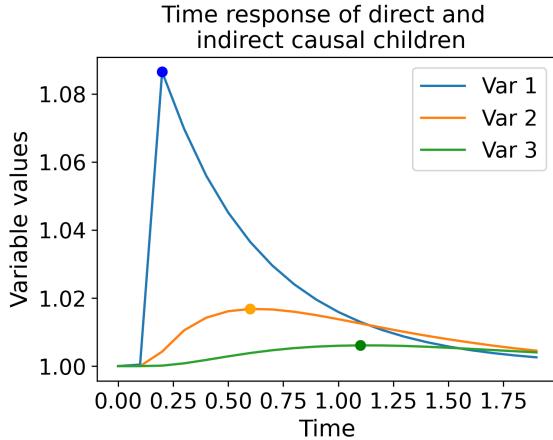


Figure 4.3: Variable 1 receives a unit impulse from time 0.1 to 0.2, it interacts with variable 2 directly with coupling strength 1, and variable 2 affects variable 3 also with coupling strength 1. As a result, the signal weakens and smoothenes with every transmission and the corresponding peaks are increasingly delayed, with variable 2 peaking at time 0.6 and variable 3 at time 1.1.

To determine the performance of the causal methods throughout the entire search space of data parameters, we also conduct experiments where parameter combinations are explored. In these experiments we use the reduced parameter range given in Table 4.1 for reasons of computational performance.

We use a decision tree to determine recommendations for a causal method based on the parameter region. The decision tree optimizes data splits such that the tree leaves make as clear of a distinction as possible between the regions where each method is the highest scoring one in terms of the MCC. To train this classifier, we structure the results of the experiments into pairs of input and labels of the form $([param1, param2..], method)$, where the method is an index for the method that achieved the highest MCC for a run with the listed parameters. Due to the underlying randomization, we can expect different results for the same parameter set over multiple runs, which can indicate that multiple algorithms are similarly well-suited for some parameter set. For the implementation of the decision tree, we use the *DecisionTreeClassifier* from *sklearn*.

All experiments are conducted multiple times with different seeds for random white noise generation, 20 times for experiments on the error rate and for the isolated data parameters, 10 times for combined search space experiments for computational performance reasons.

Finally, we also conduct an experiment on the computation time of the three utilized causal methods. We use a non-coupled cubic model system with a varying number of variables (2, 3, 5, 7, 10) and samples (the reduced range in Table 4.1). Every configuration is run either for 100 datasets or for 100 seconds, whichever condition is fulfilled first. In the experiment run, all but the slowest configuration ran for 100 datasets.

Table 4.1: Parameters used in the experiments with their corresponding ranges

Parameter Name	Full Test Values	Reduced Range	Default Value
Error Rate α	0.001, 0.005, 0.01, 0.02, 0.05, 0.07, 0.1, 0.15, 0.2, 0.25, 0.3, 0.5	0.001, 0.01, 0.05, 0.1, 0.2, 0.5	0.05
Number of Samples	20, 50, 100, 200, 500, 1000, 2000, 5000, 10000	100, 500, 1000, 10000	500
Noise Scaling	0.01, 0.02, 0.05, 0.1, 0.2, 0.5	0.01, 0.05, 0.1, 0.5	0.1
Coupling Strength for VAR systems	0.01, 0.02, 0.05, 0.07, 0.1, 0.15, 0.2, 0.25, 0.3	0.01, 0.05, 0.1, 0.3	0.1
Coupling Strength for cubic systems	0.1, 0.2, 0.5, 0.7, 1, 1.5, 2, 2.5, 3	0.1, 0.5, 1, 3	1

4.1.2 Grid Experiment Setup

We conduct a second experiment on synthetic data to gain more insight into the spatio-temporal conditions under which causal methods can be expected to produce reliable results. In this section, we describe the underlying model system and the approaches we take in the application of causal methods.

Data Generation: We construct a VAR system from a 2D grid of nodes that are coupled to their horizontal and vertical neighbors as a model for a spatially extended physical system with some mechanism of energy transport. Examples for such systems include oceanic heat transport [79], sea ice melting (through ice-albedo feedback, among others [52]) or rainforests where moisture is transported between regions through evaporation and precipitation [80].

A quadratic grid is defined where a node takes a time step by receiving independent and identically distributed (i.i.d.) white noise, maintaining a fraction of its previous value and adding a fraction of the average of the four spatially surrounding nodes. These factors weighing a node's own and surrounding values add up to 1 in order to ensure energy conservation in the network. We also wrap the grid on the borders, i.e., cells in the rightmost column also communicate with those in the leftmost column, etc.

The corresponding VAR system for an $n \times n$ grid maps every node to a variable and every edge between nodes to some constant factor b in the transition matrix $A \in \mathbb{R}^{n^2 \times n^2}$, which also has an autocorrelation factor on the diagonal. As our application area concerns interactions of climate tipping elements, i.e., physically separate systems, a global variable $Y \in \mathbb{R}$ is added to the model that may influence all grid cells and/or receive feedback from them. Figure 4.4 shows a conceptual overview of this grid model. The VAR model is formally defined as follows:

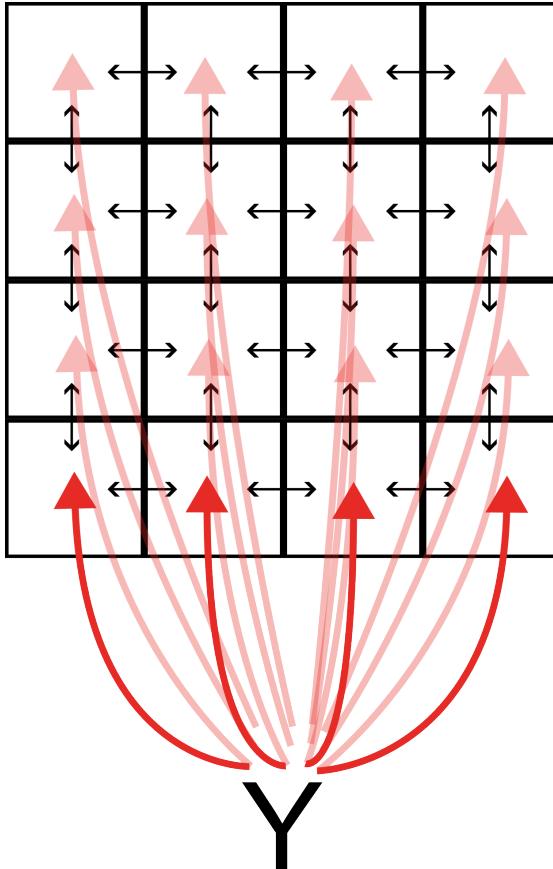


Figure 4.4: Conceptual grid model with interactions between neighboring cells and a global variable that influences all cells

With $X \in \mathbb{R}^{n^2}$, $Y \in \mathbb{R}$, $a, b \in \mathbb{R}^+$, $a + 4b = 1$:

$$A_{i,j} = \begin{cases} a & \text{if } j = i \\ b & \text{if } j = i + 1 \vee j = i - 1 \vee j = i + n \vee j = i - n \\ 0 & \text{else} \end{cases} \quad (4.3)$$

$$X_{t+1} = AX_t + cY_t + d\varepsilon \text{ with } c, d \in \mathbb{R}, \varepsilon \in \mathbb{R}^{n^2}$$

$$Y_{t+1} = uY_t + \frac{v}{n^2} \sum_{i=0}^{n^2} X_{t_i} + w\varepsilon_Y \text{ with } u, v, w, \varepsilon_Y \in \mathbb{R}.$$

Note that the scalar cY_t is implicitly extended to dimension n^2 with identical values and that the border wrapping of the grid is implemented using a modulo operation in the cases for A , which is left out of the notation here for readability.

Disturbances can occur due to noise in cells or due to noise transferred from Y , in the latter case energy conservation keeps the grid system at its changed mean value

4.1 Experiment Setup

permanently (i.e., until noise from Y changes it again). Similarly, a strong disturbance in a single cell is "flattened out" over time, producing ripples as the disturbance reaches cells further away in each time step. Over time, this mechanism leads to a rebalancing of the system towards a mean state, and we refer to it as a *diffusion* process. While this process is clearly a strong simplification of physical processes, we can parameterize it through choice of b to determine its time scale. Since b is in a range of $[0, 0.25]$, we construct an inverted measure which is referred to as the *Diffusion Factor* for easier analysis towards both ends of the range, such that $b = \frac{0.25}{2DF}$.

Experiments: When applying causal methods to a grid to determine spatial patterns of intra-grid interactions, we need to run the algorithms with all variables at once, which could only be computationally handled by the LKIF method in our tests. For that reason, our experiment to test how well the given grid structure is detected is only conducted with the LKIF method. For even larger grids, one could divide spatially resolved variables into regions to determine intra-regional interaction patterns.

The interactions between the grid and the global variable can be determined either with the entire grid included in one run of the causal methods, or by applying the methods pairwise to the global variable and each separate grid cell. The latter approach only requires analysis with two variables, but can lead to distortions when true causal links are fine-grained. For example, if there exists a causal link between the global variable and only one spatial cell, neighboring cells also show a causal relationship to the global variable in pairwise analysis due to the indirect link through the true causal grid cell. We conduct another experiment with LKIF to compare these two approaches qualitatively.

While spatial analysis can give insight into the patterns of interactions, heterogeneous results can make it difficult to draw conclusions on the general existence of a causal link between the grid and the driver. We can instead aggregate the spatial grid into one variable by taking the mean of cell values, which is by definition more closely related to Y – the grid average has an impact on Y and Y impacts the grid uniformly. We conduct experiments to test the prediction skill of PCMCI, LKIF and GCSS under these idealized assumptions with varying diffusion factors and for different directions of the true underlying causality.

A list of default parameters for the experiments is displayed in Table 4.2. For experiments with a simultaneous causality in both directions, coupling parameters were lowered to ensure system stability and only varied inversely proportional. A factor of 1 would here imply coupling from the global variable to the grid at strength 0.2, and from each grid cell to the variable at $\frac{0.5}{n^2}$. In all of the aggregation experiments, we run the causal methods ten times for each configuration to obtain uncertainty estimations.

4.1.3 Applied Experiment Setup

In this section, we describe data sources, preprocessing steps, the causal method configurations and the types of experiments we conduct to detect teleconnections (and therefore tipping point interactions) between the AMOC and Arctic sea ice.

Data Sources: Data on sea ice extent was taken from the *Sea Ice Index* by the National Snow and Ice Data Center (NSIDC) [81] as a monthly aggregate of sea ice

4 Experiments

Table 4.2: Parameters used in grid experiments with default values

Parameter Name	Default Value
Error Rate α	0.05
Number of Samples	500
Diffusion Factor	0.5
Grid Size $n \times n$	10×10
Cell Noise Scaling	0.5
Global Variable Noise Scaling	0.5
Global Variable Autocorrelation Factor	0.5
Coupling Strength: Global Var. \rightarrow Grid	1
Coupling Strength: Grid \rightarrow Global Var.	$\frac{1}{n^2}$

observations across the Arctic Sea and the Northern Atlantic. It is based on satellite observations and covers the period of 1979 through 2023.

We use spatially resolved data on sea ice concentration, thickness and albedo from the Copernicus Marine Data Store [82] as monthly gridded data with a resolution of 3x3km, covering all sea surface areas above 50°N. It is a reanalysis product based on the neXtSIM sea ice model [83] forced by satellite observations and atmosphere and ocean reanalysis data. The data set covers the period of 1993 to 2023. For a tradeoff between computational feasibility and spatial resolution, grid cells were further aggregated to a resolution of 30x30km.

For some experiments, we aggregate the spatial sea ice data over a specific region, following our results on synthetic grid systems (Section 4.3). As explained in Section 2.1.2, the overturning mechanism of the AMOC relies on sinking water masses in the North Atlantic, including areas experiencing sea ice accumulation and melt every year, like the Labrador, Greenland and Barents Seas [44]. To focus our analysis on the most pronounced sea ice dynamics within these regions, we determine the mean of sea ice albedo, concentration and thickness over regions in the 66th percentile of variance of sea ice concentration between longitudes -90° to 90° E. This region is displayed in Figure 4.5, we also refer to it as the *sea ice edge region*. We also conduct robustness tests with different regional aggregations, see Section A2.

For AMOC indices, SST data was taken from the ECMWF Reanalysis v5 (ERA5) dataset [84], a broadly used global reanalysis product, at monthly resolution and a spatial resolution of $0.25^\circ \times 0.25^\circ$. We conduct an area-weighted aggregation of SSTs according to regions defined by Caesar et al. [8] and by Pontes & Menkveld [85]. The fingerprint indices then subtract the global mean SST with a factor of 1 (Caesar et al. [8]), or a factor of 2 (Pontes & Menkveld [85], Ditlevsen & Ditlevsen [9]), the latter index is otherwise identical to that by Caesar et al. [8]. The global mean SST is taken from the Hadley

4.1 Experiment Setup

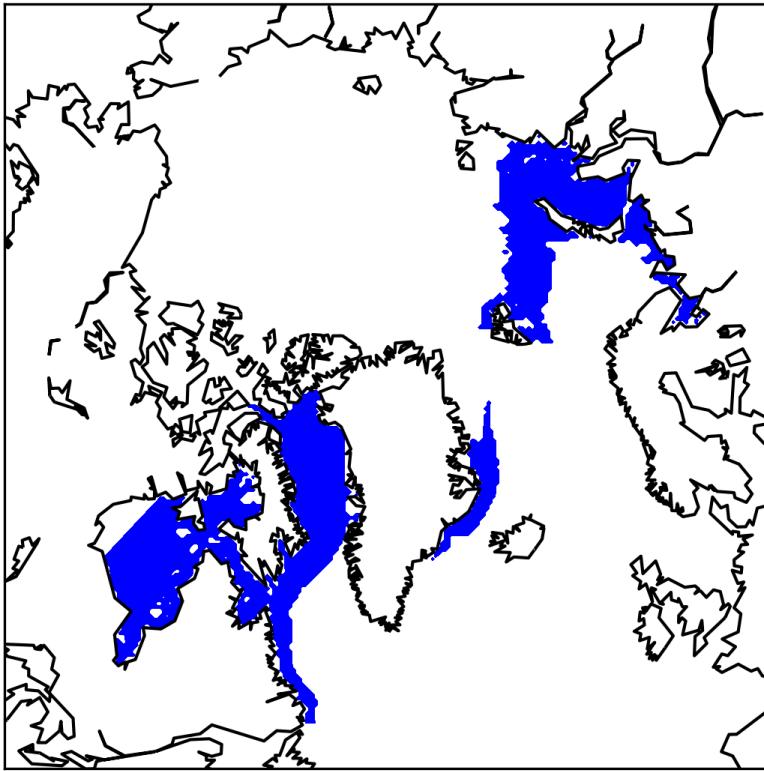


Figure 4.5: Regions between 90°W and 90°E that are in the 66th percentile of sea ice concentration variance are aggregated to focus analysis on expected interaction mechanisms between AMOC downwelling and sea ice melting

Centre Sea Ice and Sea Surface Temperature (HadSST) dataset [86].

Extending the work of Högner et al. [36], we also retain previously used indices of the Southern Amazon Rainforest (SAR) and the Caribbean Low Level Jet (CLLJ), for which precipitation and wind speeds respectively are aggregated from ERA5 data over the corresponding areas.

Preprocessing: We linearly detrend all time series data to remove the climate change signal except for the spatially resolved sea ice measurements. Since separate regions often recover to the same state in winter but lose increasingly more ice in summer, detrending only introduces biases here. We also de-seasonalise all sea ice data against its month-wise mean over the entire respective observation periods to avoid the strong seasonal effect of sea ice accumulation and melting processes. However, due to the significant decrease of summer sea ice over the past decades (while winter sea ice remained relatively stable), the processed data still shows some periodic patterns with decreasing minima in summer. We do not differentiate this data as it results in significant high-frequency artifacts. An overview of variables with their data sources and aggregation steps is given in Table 4.3.

Method Configurations: Where not specified otherwise, we use a maximum time

4 Experiments

Table 4.3: Data used for causal analysis, with details on regions, aggregation methods, and data sources

Variable	Region	Aggregation	Sources
AMOC Fingerprint	Subpolar Gyre, see Caesar et al. [8]	Detrended, De-seasonalised	ERA5 [84], HadSST [86]
Arctic Sea Ice Extent	Sea area North of 31°N	Detrended, De-seasonalised	NSIDC [81]
Arctic Sea Ice Albedo, Concentration, Thickness	Sea area North of 50°N	Deseasonalised, regional aggregate also detrended	neXtSIM [83]
CLLJ	Zonal wind speed at 925 hPa in region 7.5 - 12.5°N, 75 - 85°W	Spatial mean, Detrended, De-seasonalised	ERA5
Southern AR Precipitation	AR basin South of 5°S	Spatial mean, Detrended, De-seasonalised	ERA5

lag of 5 months for PCMCI and an error rate $\alpha = 0.05$ in the following experiments. In a simple analysis of sea ice extent, AMOC strength, the CLLJ and precipitation in the SAR, a link between sea ice and precipitation was identified and disabled in the following analysis, because this direct relationship is considered physically unrealistic without a mediator like the AMOC. In LKIF, manual disabling of links is not possible, we instead leave out the SAR precipitation variable to prevent the false positive from skewing other causal relationships.

We also apply a mask to all variables to focus on causal effects in summer (May to September), which is the time where sea ice melt is most severe, with its minimum usually in September, so a causal effect on the AMOC would be expected to be observed here. This period also coincides with the dry season of the Amazon Rainforest, the period where decreases in precipitation have the largest negative impact on vegetation resilience [87].

The causal methods can handle data masks in different ways:

- In PCMCI, we apply the mask such that only the tested causal parent is restricted to the summer period in order to focus specifically on the impact of Arctic sea ice melting.
- In LKIF, the mask is applied by simply leaving out all masked time steps.
- GCSS can not be used with masks, so we drop the method from most analysis steps.

Experiments: We conduct causal analysis with PCMCI and LKIF for most combinations of variables, and use PCMCI exclusively where the sign of an effect is of higher

4.2 Results for Multivariate Systems

interest. The Linear Mediation effect estimation method is applied to the results of PCMCI to offer physical interpretability.

We also conduct spatially resolved causal analyses using the data on sea ice albedo, concentration and thickness. The causal methods are applied separately to each grid cell to test its causal relationship to the AMOC, thereby avoiding the computational cost of an analysis with all cells included. This procedure corresponds to the pairwise analysis that we find to be advantageous over a full grid analysis in synthetic data (Section 4.3). Spatially resolved analysis can give insights into the spatial distribution and homogeneity of causal effects, including the sign of causal effects when using Linear Mediation. For the effect strength estimation with Linear Mediation, a causal network needs to be prescribed, which we determine from previously identified causal graphs. The CLLJ and SAR variables are mostly included in the causal analysis to reproduce conditions of the aggregated analysis, but their causal links are ignored in our evaluation for simplicity. We apply masking for PCMCI and LKIF, and also conduct our spatial analysis with GCSS over the duration of the entire year, i.e., without masking (Section A2).

A causal stationarity analysis is conducted using a 20 year sliding window over the time series data. We use the PCMCI and Linear Mediation methods for this to test results on sea ice extent and on the aggregated regions. This approach can reveal trends in causal effect strength, or confirm robustness of the identified causal effects as it resembles a bootstrapping confidence test (i.e., we can get a qualitative estimate of the variance of the results).

4.2 Results for Multivariate Systems

In this section, we describe and discuss the results of experiments on multivariate model systems as described in Section 4.1.1.

Error Rate: Figure 4.6 shows ROC curves for the default data generation settings. As we are more interested in trends rather than exact recommendations, we avoid noisy results by keeping the same data for varying error rates, still 20 randomized sets of data are generated from the default configuration. At a growing error rate, more connections are deemed statistically significant, which leads to a growth in true and false positive rates. We also determine an optimal value for α by a minimal Euclidean distance to the point $(0, 1)$ in the ROC diagram. For this default data generation, the optimal α lies between 0.02 and 0.07 for PCMCI, around 0.2 for GCSS and appears highly dependent on the model type for LKIF, where the detection for cubic systems is much more sensitive. Depending on the application, one can adapt the error rate to get a larger/smaller tolerance for false positives in a tradeoff for more/less true positive results.

Samples: Figure 4.7 displays the MCC for different amounts of samples. For data from a VAR model, all methods stay within one standard deviation of each other, with the only notable exception being the GCSS algorithm converging to an MCC of exactly 1 for sample counts of 5,000 and higher. GCSS also shows a slightly worse performance for sample counts of 500 and 1,000 than LKIF and PCMCI. This trend appears similarly

4 Experiments

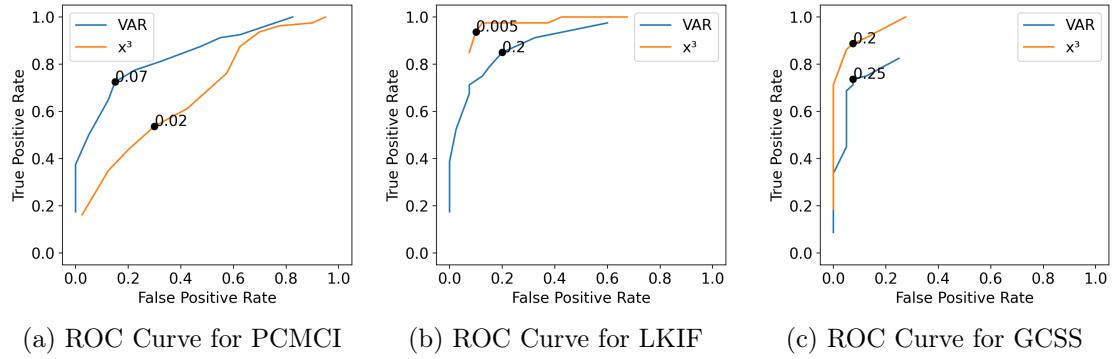


Figure 4.6: ROC Curves for varying error rate α in the default data generation process of VAR and cubic systems with labelled values as the optimal values for α

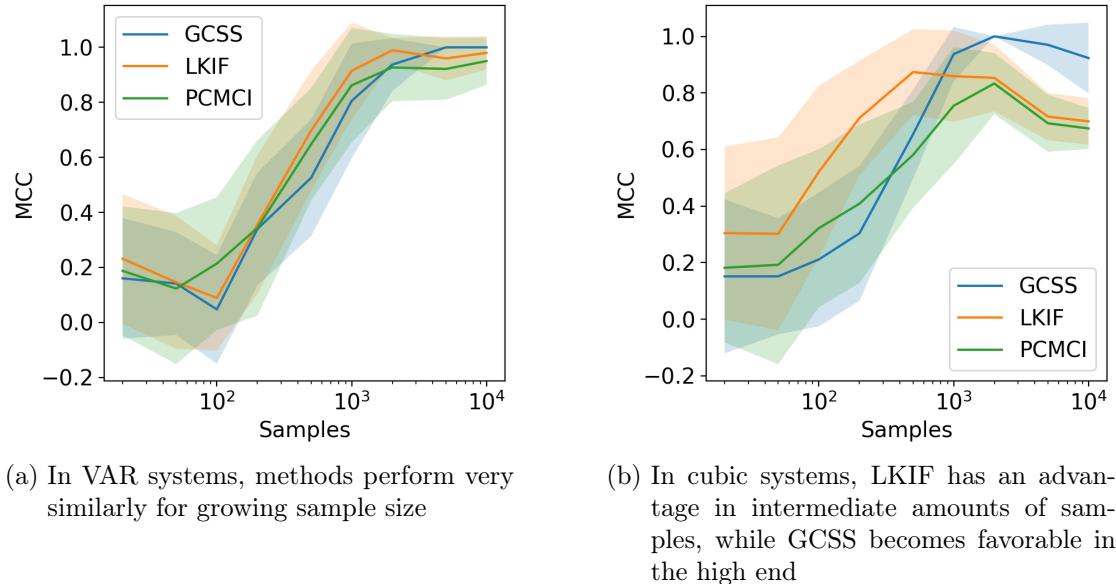


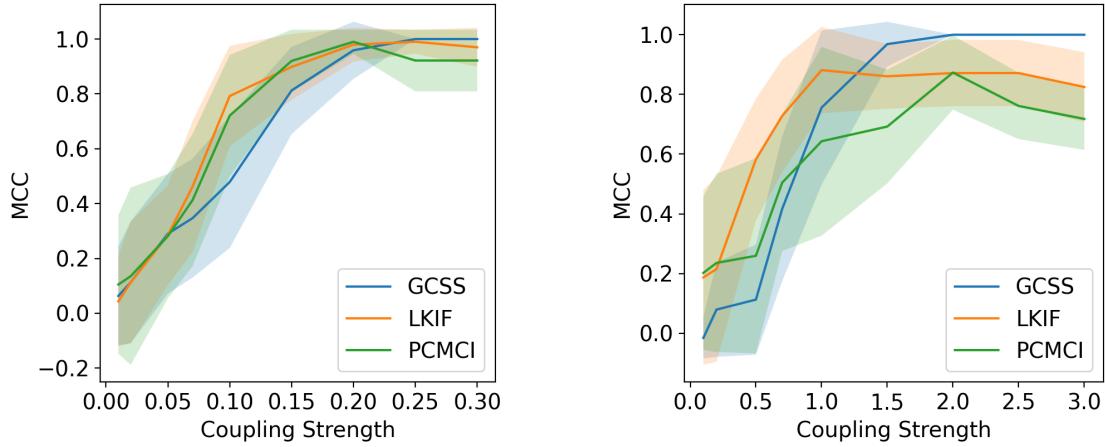
Figure 4.7: MCCs for PCMCI, LKIF and GCSS for different amounts of samples provided in VAR and cubic systems

for cubic systems, where GCSS performs significantly better for 2,000 samples upwards, but LKIF outperforms GCSS and PCMCI at 100 to 500 samples.

Since it is straightforward to determine the sample count in real applications, our results can give some orientation for usage of the different algorithms: GCSS is quite clearly favorable for large numbers of samples, which could occur e.g., when daily data over many years is used, and LKIF for intermediate amounts of samples where nonlinearity is assumed to play a significant role in system dynamics. In all other cases, PCMCI can offer a reliable alternative method with an advantage in flexibility.

Coupling Strength: The coupling strength parameter is also crucial for the detection

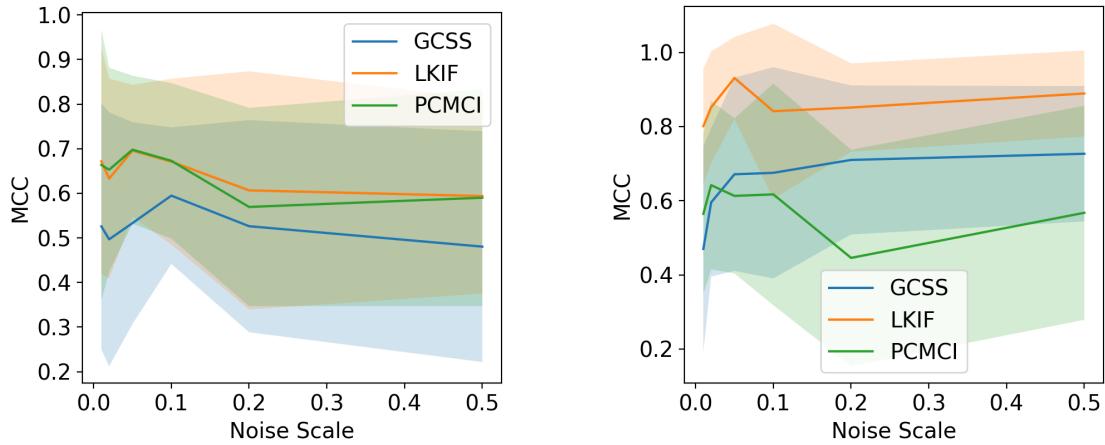
4.2 Results for Multivariate Systems



(a) In VAR systems, methods perform similarly for growing coupling strength, except for GCSS in intermediate coupling strength

(b) Cubic systems favor LKIF for lower coupling strength and GCSS for higher coupling strength

Figure 4.8: MCCs for PCMCI, LKIF and GCSS for different coupling strengths in VAR and cubic systems



(a) No clear trends emerge in VAR systems for varying scale of the white noise

(b) Clear differences in performance between the methods for cubic systems, but no clear trends either for growing white noise scale

Figure 4.9: MCCs for PCMCI, LKIF and GCSS for different scaling of white noise in VAR and cubic systems

skill of all methods, see Figure 4.8. In VAR systems, MCCs grow from just above 0 to 0.8 when the coupling strength is increased from 0.01 to 0.1 for all but the GCSS method, which only reaches an MCC around 0.5 at that point. Raising the coupling strength to

4 Experiments

0.2 increases the performance of all methods to almost perfect detection. This is a very good sign for the detection power of the methods, since a coupling strength of 0.1 in VAR systems (with identical noise strength in all variables) implies that the receiving variable has a signal-to-noise ratio of 1:10 for the transferred signal that is detected by the causal methods. In the multivariate setup with feedback loops and potentially overlaying signals, the signal-to-noise ratio is only further deteriorated, yet methods show quite reliable results already.

In cubic systems, the coupling strengths show a similar trend, where performance rises sharply up to a coupling strength of 1, at this point LKIF demonstrates the highest average MCC of more than 0.8. An increase in coupling strength from here only benefits the GCSS method, which again converges to perfect detection from coupling strength of 2 or higher.

Note that these coupling strengths cannot be compared directly to the linear ones, as we are dealing with integration of differential equations rather than discrete time steps, as demonstrated in Figure 4.3. Linearization and discretization around the equilibrium for a coupling strength of 1 and a time step of 0.1 resulted in a coupling strength of 0.1 in the VAR system for the specific underlying system of the synthetic data, see Section 2.1.3, although this relationship is not generalizable to other parameter configurations and time scales. Overall, these findings strongly suggest that the coupling strength is decisive for causal method performance in a similar manner for nonlinear and linear systems.

Noise Scale: The scale of white noise in the synthetic systems does not display a clear trend for any of the methods, see Figure 4.9. This result is unsurprising for linear systems, where the scale of disturbances is irrelevant to the qualitative behavior. However, cubic systems should react nonlinearly, i.e., they should display stronger or weaker reactions to a disturbance than suggested by the local linearization, with larger deviations from the linearization for stronger noise. All causal methods exhibit tolerance towards these deviations from linearity, which supports their usage in cases without exact fulfillment of their underlying assumptions.

Model Systems: We discuss the following results for different model system configurations (with respect to size and interaction density) for each algorithm separately. As we do not know the true density of a system before causal analysis, we are more interested in relative trends of algorithms rather than the best causal method to use for some configuration.

PCMCI shows no clear trend with respect to the number of variables, which supports its usage in the applications to tipping element networks, where one would rarely expect to encounter more than 10 interacting variables.

In both linear and cubic systems, PCMCI shows better performance the lower the density of a system, which supports the findings by Strahl [73] claiming that feedback loops deteriorate PCMCI performance (Figure 4.10a and 4.10d).

The LKIF algorithm exhibits slight increases in MCCs for increasing system size, and while system density is not relevant to its performance on linear systems, on cubic systems it achieves higher scores on lower density systems (Figure 4.10b and 4.10e).

4.2 Results for Multivariate Systems

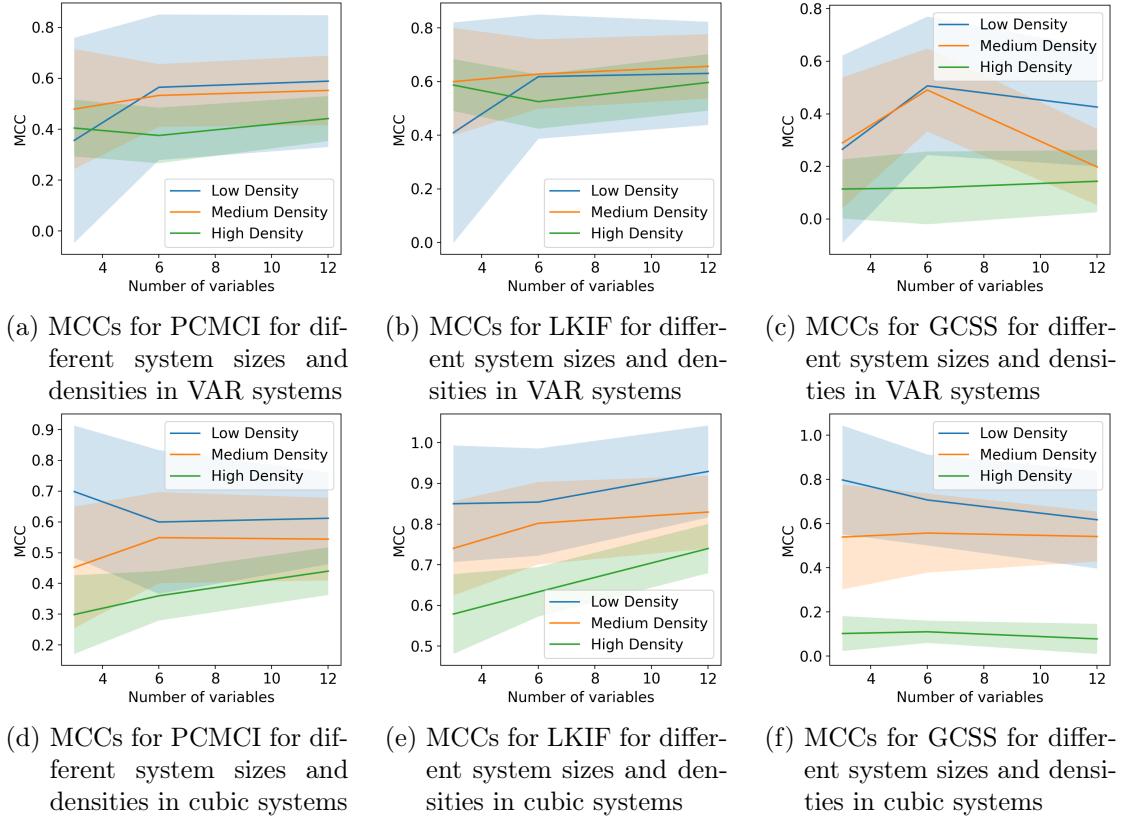


Figure 4.10: MCCs for all algorithms for different system sizes and densities, refer to Figure 4.2 for underlying systems

The GCSS algorithm does not show a clear trend with system size, but a very weak performance in high density systems. It appears to handle cubic systems better than linear ones at lower system sizes, which might be explained by its underlying moving average component that could be an overfit to linear systems but gives additional degrees of freedom to describe nonlinearities and differential system integration effects (Figure 4.10c and 4.10f).

Error Rate, Samples and Coupling Strength: Given the significantly varying optimal values of α in the default configuration (Figure 4.6), one may have doubts about the choice of $\alpha = 0.05$ as the default value in the previous experiments. We therefore conduct a comparison on the combined search space of α , the coupling strength and the number of samples (as the most influential parameters) and determine the optimal choice of α for every configuration according to the MCC.

Figure 4.11 shows the resulting plots which indicate the optimal choices of α in labels as well as the MCC value by color. There is a clear trend in all algorithms towards lower values of α for higher coupling strength and sample count. The maximum MCC also increases monotonically in these directions for VAR systems, but drops off for PCMCI

4 Experiments

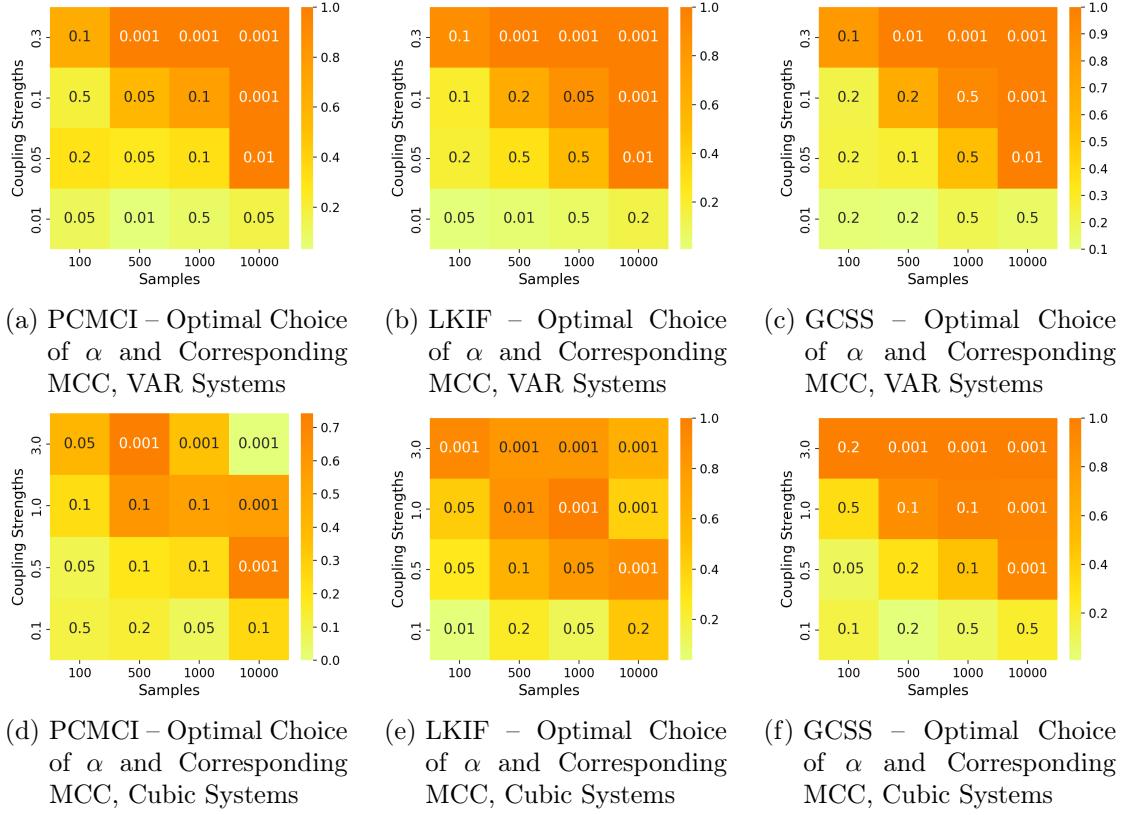


Figure 4.11: Plots for varying sample count and coupling strength, with cell entries indicating the optimal choice of α and cell colors indicating the MCC achieved with that α

and LKIF towards very high samples and coupling strength in cubic systems. Detailed analysis reveals this dropoff is due to a large fraction of false positives, i.e., a smaller value of α would likely increase the scores.

Although this confirms that the recommended choice of α heavily depends on the underlying data, for 1,000 or less samples and coupling strengths of 0.1 or lower (and 1 or lower for cubic systems) as expected in applications to climate data, error rate values of 0.05 to 0.1 seem to offer a good balance.

Combination of Data Parameters: For the following experiment, we keep the error rate fixed to 0.05 and explore the entire parameter space of data generation configurations. We generate two decision trees (for VAR and cubic systems), that differentiate regions in the parameter space that favor one method over the others. Where two (or all three) methods perform best with the same MCC, this is indicated by the label "Equal".

In the VAR systems, the system size appears as the most important parameter, i.e., at the first node of the tree, see Figure 4.12. The notation "@0.50" refers to the index of the system sizes (3, 6 or 12 variables) where the tree is split into a side with smaller

4.2 Results for Multivariate Systems

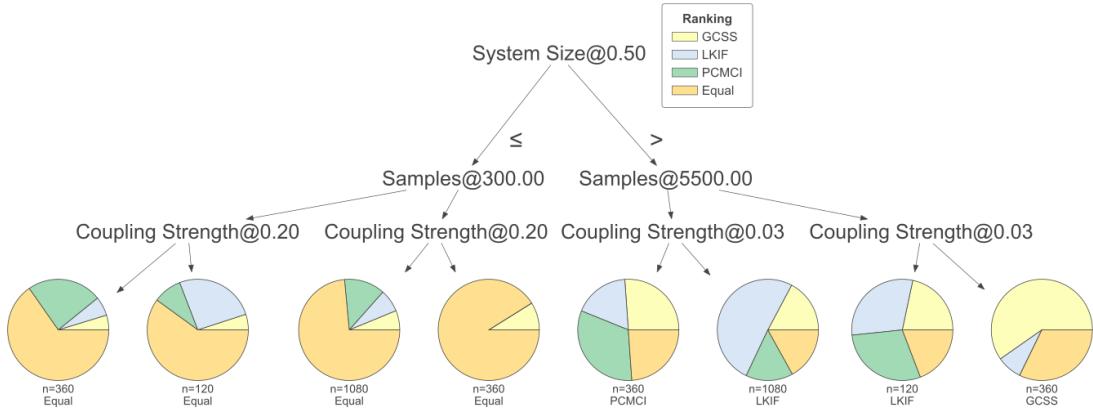


Figure 4.12: Decision Tree for VAR systems and parameters as listed in Table 4.1

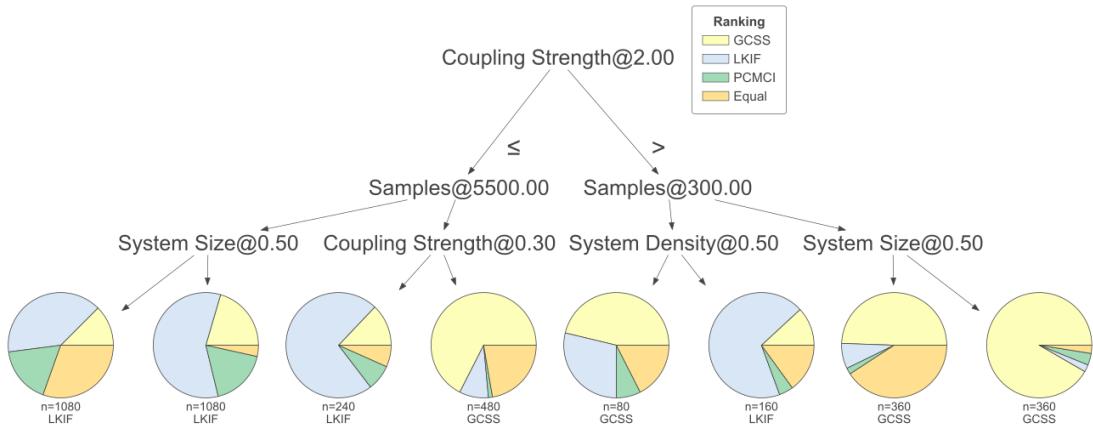


Figure 4.13: Decision Tree for cubic systems and parameters as listed in Table 4.1

and larger parameter settings, i.e., all nodes to the left contain runs with 3 variables, the nodes to the right contain runs with 6 or 12 variables. The further decisions are made based on sample count, then coupling strength. The resulting sets in the leaves of the tree show that methods perform similarly well for small systems, while GCSS is favored for larger systems with many samples and high coupling strength. Both PCMCI and LKIF appear to be reliable choices for all other configurations.

For cubic systems (Figure 4.13) the coupling strength is the most crucial parameter, with the sample count at the second level nodes. Compared to VAR systems, the system size and density (with densities low (index 0), medium (index 1) and high (index 2)) play a larger role in differentiating results. Overall, the GCSS method is advantageous in several configurations, especially for higher samples and coupling strengths, but struggles with higher system density. LKIF turns out to be the best method for all other scenarios,

4 Experiments

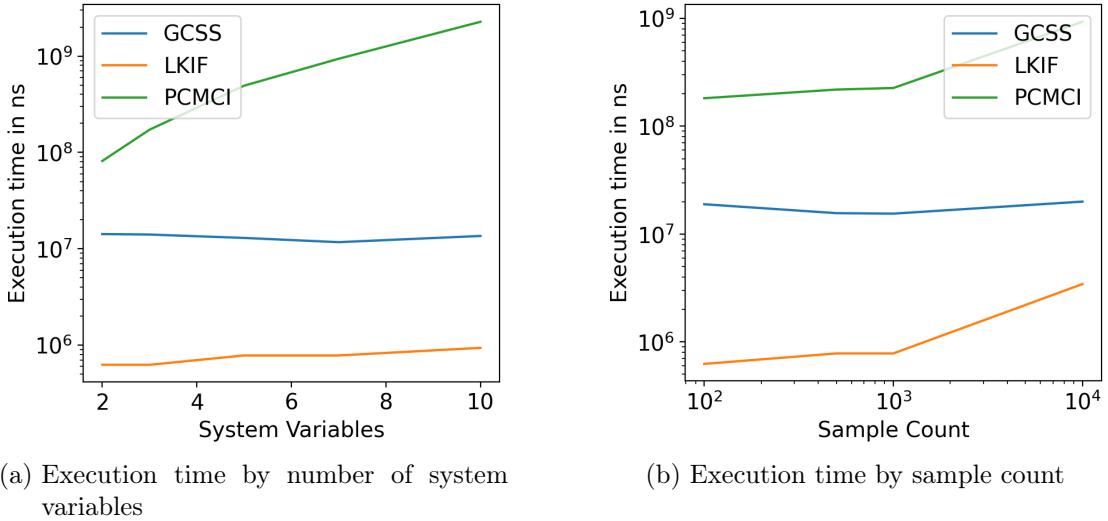


Figure 4.14: Execution time of the three causal methods by number of variables and samples

and PCMCI is favored rarely.

Performance: We find clear differences in the execution time of the causal methods, where LKIF is an order of magnitude faster than GCSS, and GCSS is one to two orders of magnitudes faster than PCMCI. Figure 4.14a shows that the execution time of LKIF and GCSS is independent of the number of variables for the chosen range, while PCMCI becomes roughly quadratically slower for increasing numbers of variables. This slowdown is most likely caused by PCMCI’s pruning mechanism, which starts at a full graph and removes edges using conditional independence tests, and the number of edges in a full graph scales quadratically with the number of nodes. The asymptotic time complexity of LKIF and GCSS with regard to the number of variables is likely hidden by constant overhead here. Under increasing samples, LKIF and PCMCI both become slower, but appear to scale less than linearly, see Figure 4.14b. The underlying calculation of correlation and covariance in the algorithms should result in linear time complexity, we again consider it likely that some constant overhead raises the computation time for low numbers of samples here. Overall, LKIF remains the fastest method in all tested scenarios, which makes it preferable in large datasets (e.g., due to high spatial or temporal resolution).

Summary: To summarize, our findings show a relatively detailed picture of the advantages and shortcomings of each causal method:

- The LKIF method performs best in lower sample counts and lower coupling strengths, especially in nonlinear settings, which is likely due to its underlying model fit to a differential equation. These factors make it a strong candidate for the detection of tipping point interactions, where some degree of nonlinearity can be assumed in the internal dynamics of tipping elements, and samples may be relatively sparse. It also appears to be quite robust to denser systems, which can be expected in tipping

element networks that often exhibit bidirectional interactions.

- The PCMCI method provides reliable performance for most parameter settings, but falls somewhat short of the other two methods in nonlinear systems. It also runs into performance problems in systems with many variables. However, its flexibility with regards to data processing (e.g., masking, prohibiting variable interactions) can make it advantageous in expert knowledge integration and for a relatively low sample count and coupling strength.
- The GCSS method makes a clear tradeoff in comparison to the other two methods, where it detects statistically significant causal effects perfectly under ideal settings, like high sample counts or strong coupling. In some cases of very long time series, e.g., daily observation data of basic climate variables, the method can be expected to be the most reliable out of the three. However, it struggles severely when fewer samples are available, couplings are weak or when systems exhibit a high density of connections.

Analysis on tipping elements usually contains imperfect proxies for tipping elements (thereby weakening the interaction signal), operate on monthly to yearly data with lower numbers of samples and potentially higher system densities, therefore GCSS seems least suitable to our purposes. Where applicable, one may want to use all three algorithms to provide higher robustness to the results.

From our results, we derive broader recommendations for the pipeline described in Chapter 3, based on information available at the start of causal analysis (i.e., without specific knowledge about the underlying model or interaction density):

- If more than 10 variables are included in the analysis, PCMCI is warned against for performance reasons.
- If less than 200 samples are available, low confidence in results is indicated.
- If the user prescribes some links manually, only PCMCI can fulfill this requirement. If masking is applied, only LKIF and PCMCI can be used.
- If more than 1000 samples are available, GCSS is recommended, ideally to be combined with the other methods.
- In all other cases, LKIF and PCMCI are recommended.

4.3 Results for Grid Systems

We describe and discuss the results of several experiments on grid systems as introduced in Section 4.1.2 and summarize our recommendations.

Grid Network Detection: In a first experiment, we test the detection ability of the spatial interactions of the LKIF method, where all variables are included in one run of

4 Experiments

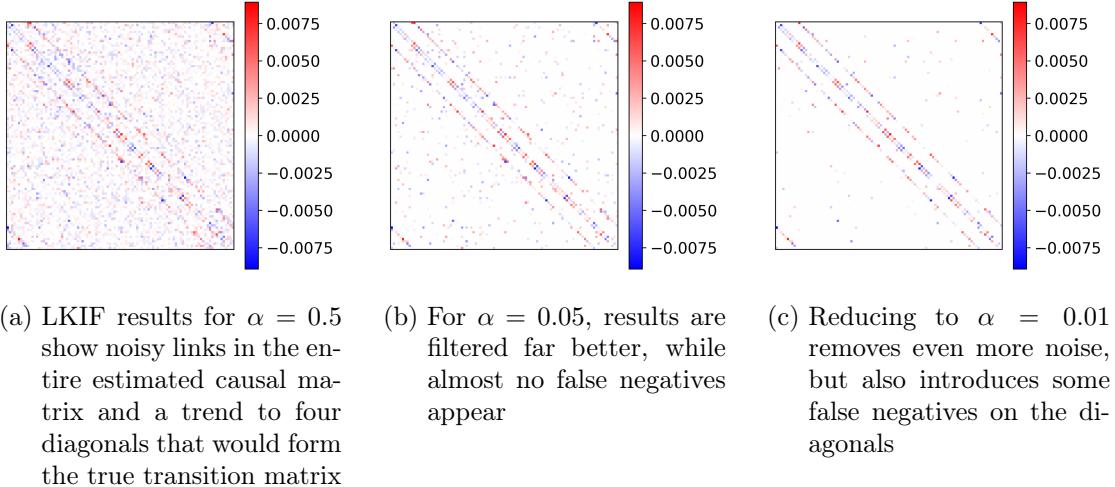


Figure 4.15: Causal matrices determined by LKIF for varying α values for one run with default grid settings

the causal discovery algorithm. The true matrix would contain four edges for each node as described in Section 4.1.2, i.e., four diagonals would form.

The results in Figure 4.15 demonstrate the prediction skill of LKIF, where even very low values for the error rate clearly show the four diagonals that make up the spatial interaction pattern. The noisier results for $\alpha = 0.5$ do not show any discernible "smearing" of the diagonals, i.e., indirect links from nodes towards their diagonal neighbors (and further away) are not biased towards false positives. However, the causal effect strength is not estimated consistently by LKIF, only qualitative analysis can be recommended for intra-grid causal analysis. Also note that the diffusion factor and the homogeneous transition matrix are idealizing assumptions that may not transfer to observational data.

External Interactions: In the next experiment, we add the global variable (often shortened to "Global" in the following figures) to our analysis with either direction of causality between the grid and the variable. We determine the spatial accuracy of the causal analysis of LKIF in the cases where all grid cells are kept in the analysis and where causalities are determined pairwise, i.e., with one grid cell at a time.

When there is a true causality from the global variable to each grid cell, LKIF can identify it well both in full and pairwise analysis as seen from Figure 4.16, with pairwise analysis giving more accurate homogeneous results in terms of effect strength. When cell noise is significantly stronger than the noise received from the global variable, the spatial distribution of effect strength becomes more noisy, but remains qualitatively correct for both analysis approaches.

The opposite direction of a causality from grid cells to the global variable is not detected consistently. Figure 4.17 shows the sparse causal links detected in the full and pairwise analysis approaches. This direction of causality is naturally more difficult to discover as noise is uncorrelated between cells, and 100 instances of white noise can be expected

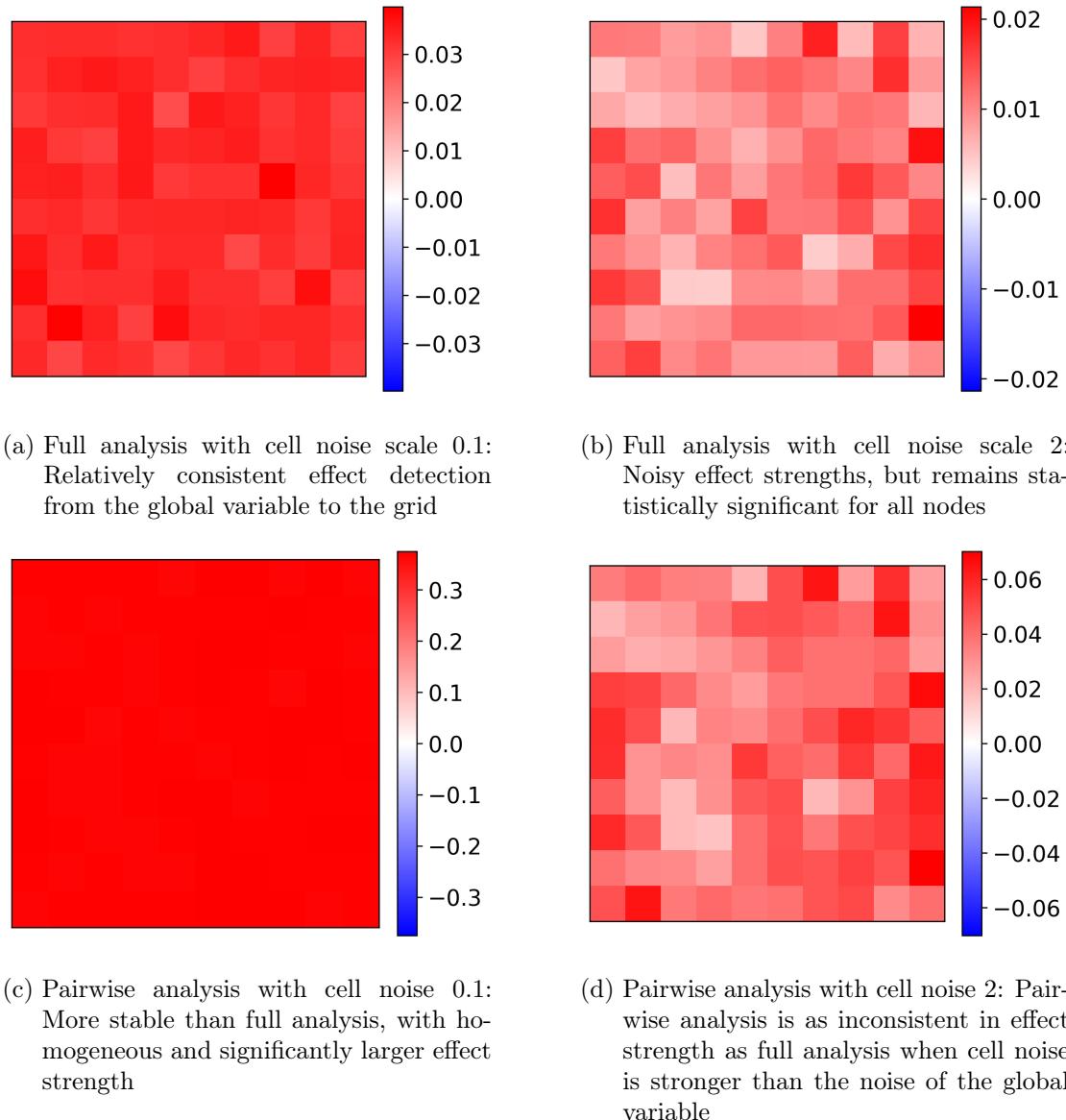


Figure 4.16: Causal effect from the global variable to each grid cell determined by full or pairwise analysis for different magnitudes of cell noise

4 Experiments

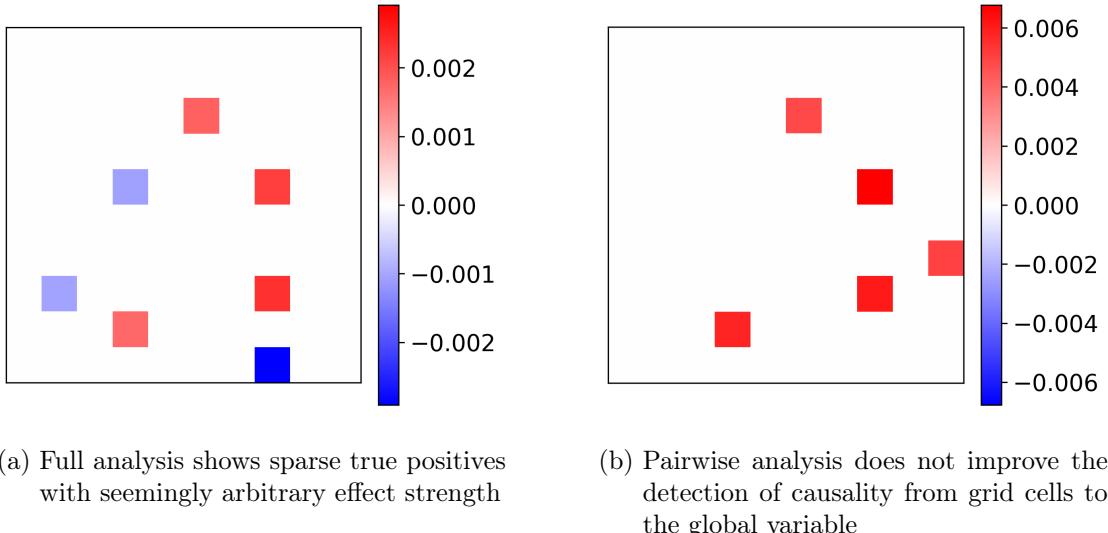


Figure 4.17: Causal effect from the grid to the global variable determined by full or pairwise analysis for default cell noise level of 0.5

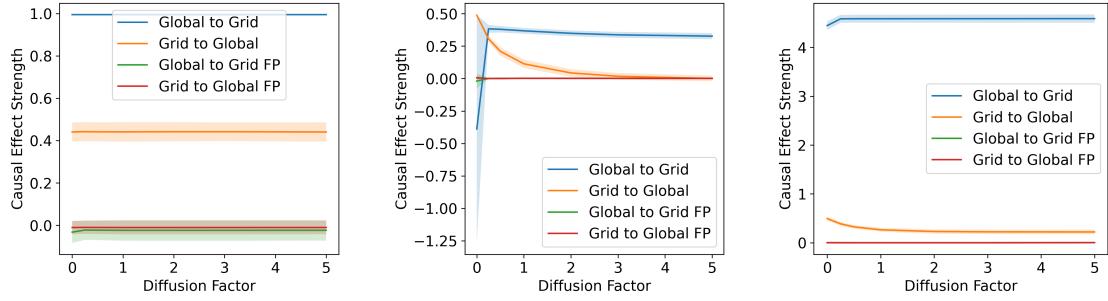
to cancel out in their average, i.e., the impact of a single cell on the global variable is too weak for detection under the chosen conditions. In real scenarios with spatially correlated noise as introduced by most climate phenomena like insolation, cloud cover or precipitation, pairwise analysis can be expected to detect statistically significant effects. This improvement would come at the cost of quantitative accuracy, as correlated noise would have a large effect on the global variable and this large effect would be detected for every cell that experiences the correlated noise.

Aggregate Interactions: Finally, we conduct experiments to determine how an aggregation over the grid cells changes causal detection ability for PCMCI, LKIF and GCSS. As before, experiments are separated by the direction of causality, here we also introduce one where both directions of causality are enabled.

In the unidirectional case of causal links (Figure 4.18), PCMCI and GCSS demonstrate good detection abilities for both directions regardless of the diffusion factor. In contrast, LKIF does not handle very low or high diffusion factors well, with the detected causal effect from the grid to the global variable converging towards zero at higher diffusion factors. This is likely due to the underlying differential model of LKIF, while PCMCI and GCSS can model the strong autocorrelative effects that appear at higher diffusion factors quite well. To give some intuition on the physical meaning, a diffusion factor of 2, which is roughly where LKIF becomes unreliable, implies an autocorrelation factor of 0.75 in the transition matrix A .

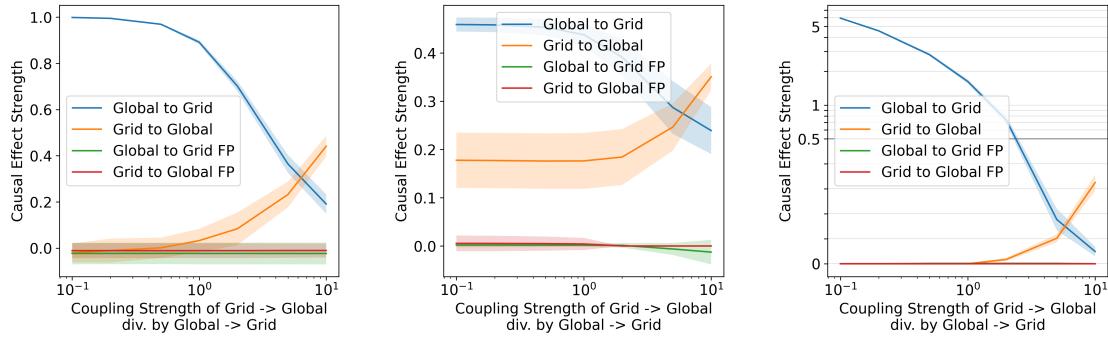
In the case where causality occurs in both directions simultaneously (Figure 4.19), we find contrasting results. PCMCI and GCSS only detect the causal effect from the grid when its has high coupling strength, while LKIF finds both links at all tested coupling

4.3 Results for Grid Systems



- (a) PCMCI obtains constant causal effect strength estimates and remains significantly above one standard deviation of false positives
- (b) LKIF shows inconsistent performance, where a diffusion factor of 0 produces large deviations in the direction from the global variable to the grid, while a large diffusion factor prevents detection of the effect from the grid to the global variable
- (c) GCSS provides consistent causal detection for both directions of causality and avoids false positives entirely

Figure 4.18: Causal effect strength for unidirectional causality estimated by the different causal methods with varying diffusion factor, with a false positive control scenario where no causal link exists



- (a) PCMCI detects causality from the global variable consistently, but from the grid cells only at high coupling strength
- (b) LKIF obtains significant causal effects for both directions at all of the tested coupling strength configurations
- (c) GCSS behaves similarly to PCMCI as it only finds significant effects for both directions when the effect from the grid is stronger than from the global variable

Figure 4.19: Causal effect strength for bidirectional causality estimated by the different causal methods with varying fraction of coupling strengths, with a false positive control scenario where the corresponding causal link is removed

4 Experiments

strengths. Further experiments with a higher diffusion factor show that LKIF profits from its high sensitivity to low interaction strengths even in systems with higher autocorrelation, so that it might be more useful than PCMCI and GCSS when the causality from the grid is relatively small, see Section A1.

Summary: We can conclude that PCMCI and GCSS are preferable in scenarios with high autocorrelative effects and only one direction of causality or bidirectional causality with large effects from the spatially resolved variable. The LKIF method would be recommended in scenarios where a bidirectional link is expected, and it becomes more reliable for low autocorrelations between time steps. LKIF is also able to detect the spatial network structure and the spatial distribution of the effect of a global variable on the grid cells.

4.4 Application to Tipping Point Interactions

In this section, we apply the causal methods LKIF, PCMCI and GCSS to observational and reanalysis data on climate tipping elements to determine a network of interactions between them. Building on previous work by Högner et al. [36] on a stabilizing effect of the AMOC on the SAR, we add Arctic summer sea ice to our analysis to check for causal links between the AMOC and Arctic sea ice and estimate the interaction strength of these tipping elements. This could inform a better estimation of the consequences of a weakening AMOC on sea ice and of a potential ice-free Arctic ocean on the AMOC.

In the results, we present evidence for a bidirectional causal effect between AMOC and Arctic sea ice, in which a strong AMOC leads to melting events in sea ice, and melting sea ice leads to a strengthening of the AMOC in the short term and a weakening in the longer term on scales of several months.

In the discussion of these results, we integrate knowledge of underlying physical processes for the interaction of the AMOC and Arctic sea ice, and find that our results match the competing dynamics in the AMOC downwelling mechanism in the North Atlantic and Arctic Ocean. We hypothesize that what we observe in causal analysis corresponds to these two phenomena, firstly that a decrease in sea ice allows for more heat loss of surface water to the atmosphere (i.e., cooling) that leads to a stronger downwelling and AMOC strength, and secondly a slower freshwater influx from melting sea ice that reduces sea surface salinity, and thereby reduces downwelling and AMOC strength.

4.4.1 Results

For the following results, we use the AMOC index established by Caesar et al. [8], analysis with other indices is found in the appendix A2. We first apply PCMCI and LKIF to sea ice extent data (referred to as "SI extent" in the following figures), then go into more detail on the regional aggregate and spatially resolved data on sea ice albedo, concentration and thickness. Finally, we conduct a causal stationarity analysis.

Analysis of Sea Ice Extent: We detect a causal effect of the AMOC on sea ice extent using PCMCI analysis, where higher AMOC strength lowers the sea ice extent after one

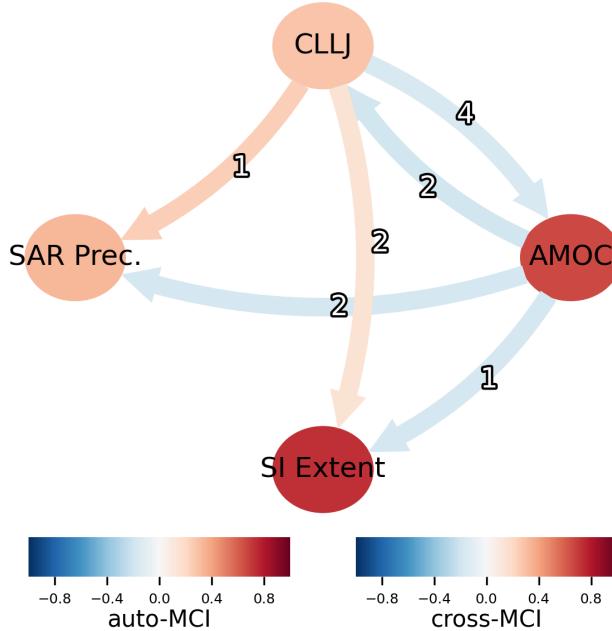


Figure 4.20: Causal network between sea ice extent, AMOC and Amazon rainforest: PCMCI analysis shows a fast negative link from AMOC strength to sea ice extent

month, see Figure 4.20. The network between the AMOC, CLLJ and SAR precipitation identified by Högner et al. [36] also remains intact, although some additional causal relationships appear.

Analysis with the LKIF algorithm confirms this link: The AMOC strength has a significant causal impact on sea ice extent (see Figure 4.21a) with an effect strength about half as large as the effect of the AMOC on the CLLJ in relative terms. This effect strength is shown in the top right cell in the effect strength matrix in Figure 4.21b (although there is no immediate physical interpretation for its value).

The Linear Mediation method estimates an interpretable effect strength for the network identified by PCMCI. We find quite small effect strengths for the causal links influencing sea ice extent: Figure 4.21c shows values of -0.065 and 0.063 for the effects of AMOC and CLLJ on sea ice extent, almost an order of magnitude smaller than the effect from AMOC to the CLLJ at -0.31 . We assume that the impact of the CLLJ on Arctic sea ice is a false positive for lack of a known physical mechanism.

The results of Linear Mediation can be interpreted such that one standard deviation in the causal parent leads to x standard deviations in the causal child, in our case $-0.065 \cdot \sigma_{SI_extent}$. To translate these values to physical units, we apply a conversion of 3.8 Sv K^{-1} for the AMOC fingerprint as determined by Caesar et al. [8]. Our results imply that a weakening of the AMOC by 1 Sv results in $13.6 \cdot 10^3 \text{ km}^2$ larger sea ice extent compared to a month's average. With recent minima of just over $4 \cdot 10^6 \text{ km}^2$ of

4 Experiments

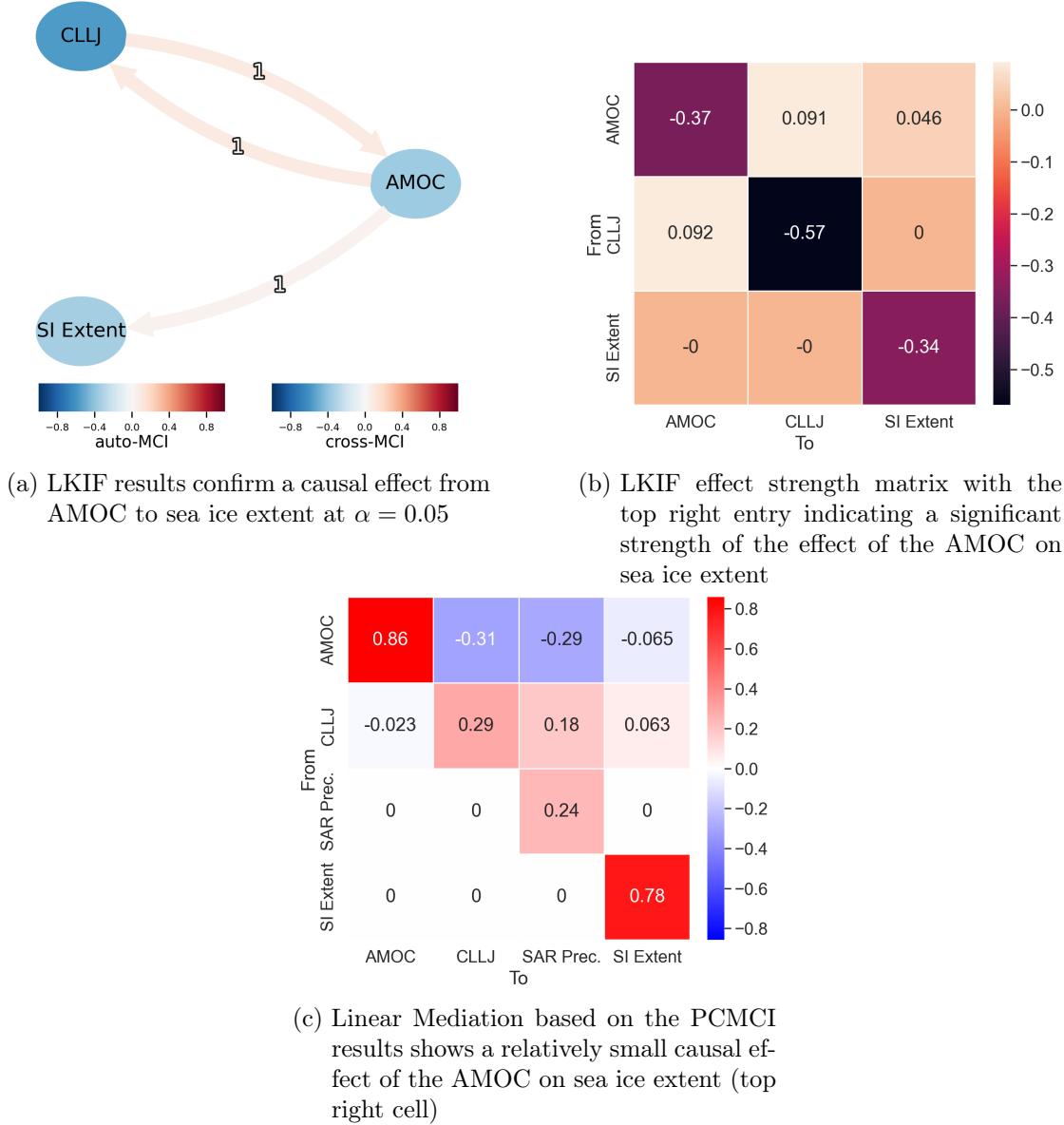


Figure 4.21: Effect strengths estimated by LKIF and Linear Mediation for sea ice extent and AMOC strength

4.4 Application to Tipping Point Interactions

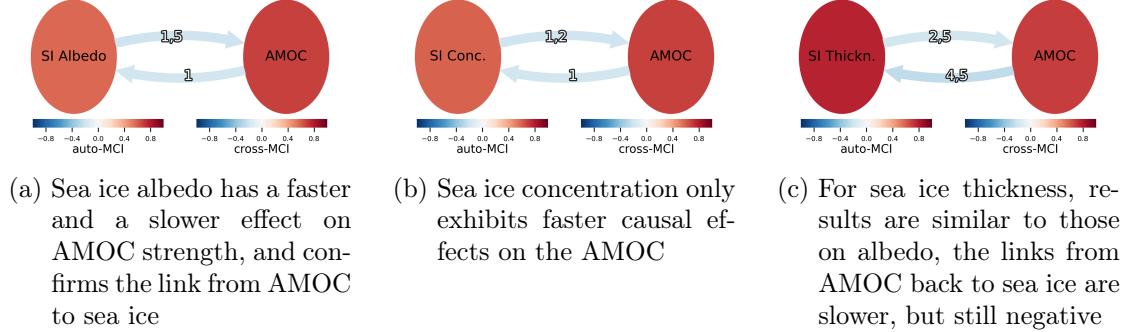


Figure 4.22: Causal graphs for aggregated sea ice variables over the North Atlantic region as specified in Figure 4.5 using the PCMCI method

sea ice extent in September [88], that would equal a gain of 0.34%, which is negligible compared to the effects of climate change on sea ice.

Aggregated Sea Ice Edge Region: We extend our analysis to regions with the highest variance of sea ice concentration on the Atlantic side of the Arctic shown in Figure 4.5. Results on separate aggregations for regions to the West and East of Greenland are found in Section A2. The established causal network including the Amazon rainforest does not remain consistent upon introduction of the aggregated sea ice variables for unknown reasons. To prevent a distortion of results between sea ice and AMOC, we remove the variables CLLJ and SAR precipitation from the following analysis.

Analysis with PCMCI on the aggregated region shows causal effects from sea ice to AMOC at multiple time lags: All three measurements (albedo, concentration and thickness) agree on a fast negative link, i.e., lower values in sea ice measurements strengthen the AMOC on time scales of one to two months. Sea ice albedo and thickness variables also show a competing positive effect at a time lag of 5 months, implying that sea ice melting would lead to a weaker AMOC in the longer term.

All three spatially aggregated variables also experience a negative causal effect from AMOC, confirming the results found for sea ice extent. Figure 4.22 shows the corresponding causal graphs, where the positive links from sea ice to AMOC are hidden due to the faster negative ones.

The links between sea ice albedo and AMOC are estimated to have an effect strength of around 0.1 in Linear Mediation analysis. For sea ice concentration and thickness, values between 0.05 and 0.2 are identified for these links, which overall points towards a significant causal interaction of the AMOC and sea ice in the aggregated region. The effect strength matrix in Figure 4.23b shows the effect strength estimated by Linear Mediation only for the stronger negative link.

In physical units, the positive (negative) link implies that a reduction in surface albedo in the selected sea ice edge regions of 0.01 in comparison to a month's average leads to a decrease (increase) of 0.050 Sv (0.060 Sv) in AMOC strength. Linear regression suggests a reduction of 0.034 in average surface albedo per decade over the past three decades (in the aggregated region), i.e., these decreases in surface albedo could have contributed

4 Experiments

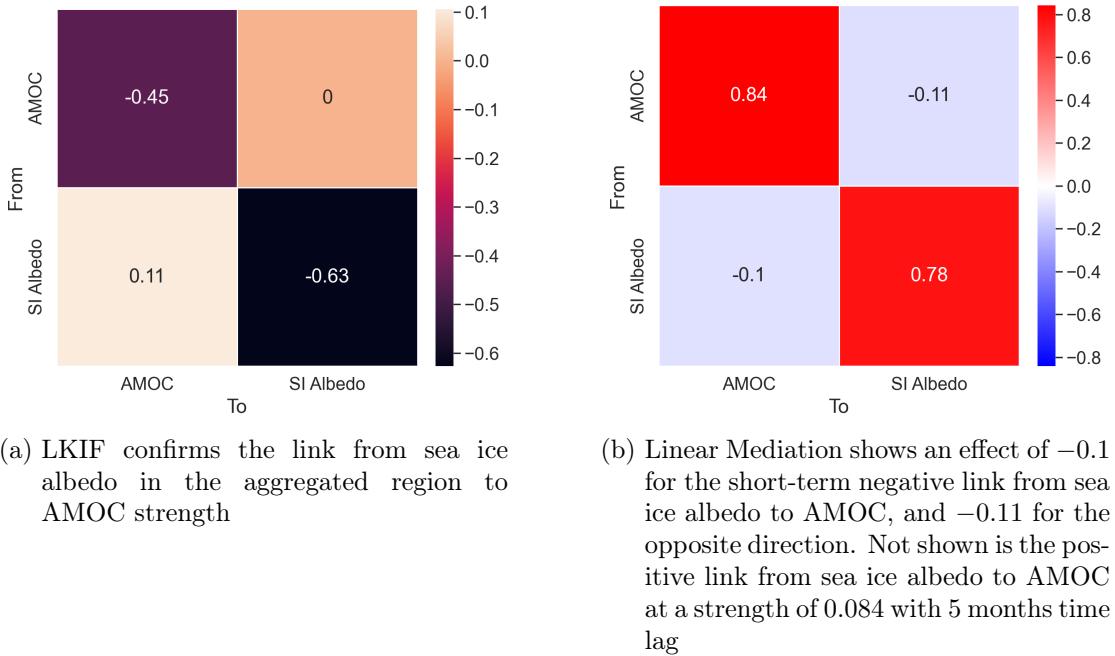


Figure 4.23: Causal effect strengths of the link from sea ice albedo to the AMOC as determined by LKIF and Linear Mediation for the aggregated regions of the Arctic

significantly (with about 0.5 Sv) to a weakening or loss of resilience of the AMOC. In the other direction, a decrease of 1 Sv in AMOC strength would only result in an increase of 0.0018 in surface albedo, which is again a negligible effect like the effect found for sea ice extent.

However, these estimates depend heavily on assumptions of anomaly-free detrending and de-seasonalising, so we advise to not interpret them as accurate physical estimates, but as orientation on the scale of effects.

LKIF analysis confirms these findings, where a causal effect is found from sea ice to AMOC, but not the other way around, see Figure 4.23a. Although LKIF does not specify the time lags of causal effects, this result provides additional robustness to the significant link strength identified by Linear Mediation.

Spatially Resolved Analysis: Spatial analysis on the three sea ice measurements is conducted using PCMCI with Linear Mediation. Here, Linear Mediation is applied with a prescribed causal network that uses the results of Högner et al. [36] for the links of AMOC, CLLJ and SAR, and adds a bidirectional connection of AMOC and Arctic sea ice with lags taken from the previous analysis on sea ice extent and albedo. We only analyze the slower destabilizing link from sea ice to the AMOC. Figure 4.24 shows a visualization of the prescribed graph.

Sea ice albedo and concentration show a relatively homogeneous positive effect on AMOC strength, especially in areas around Greenland. Sea ice thickness shows a mostly

4.4 Application to Tipping Point Interactions

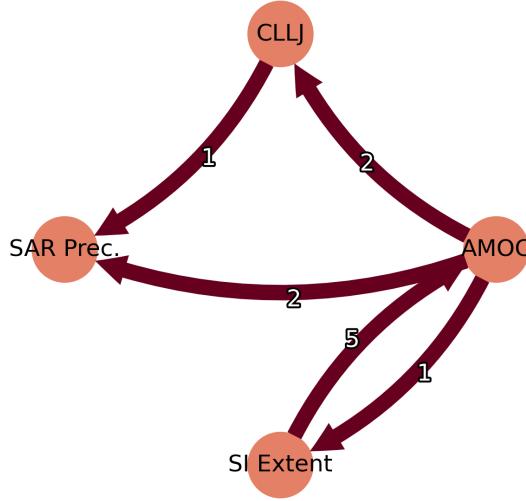


Figure 4.24: Causal graph prescribed for Linear Mediation in spatially resolved causal analysis, with edge weights to be determined

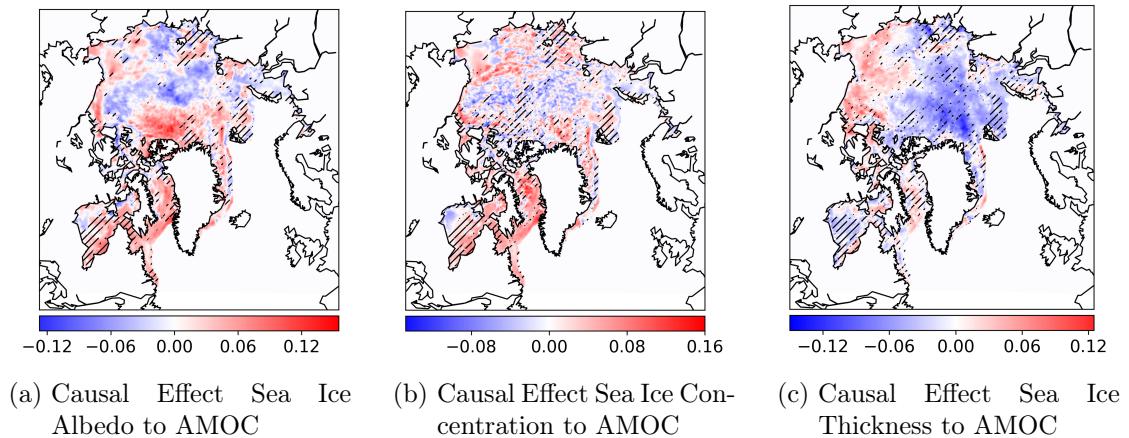


Figure 4.25: Spatially distributed causal effects of sea ice to AMOC according to the Linear Mediation method with hatched areas indicating significance at error rate $\alpha = 0.05$ determined by PCMCI

negative effect in the central Arctic Ocean and towards the Fram Strait, and positive effects in regions around the North of Canada. The regions around Greenland are also more likely to show statistically significant effects, supporting the aggregation of these regions conducted earlier. Figure 4.25 shows a detailed overview of the causal effect distributions with link sign and strength indicated by colors and significance by hatching.

The effect of AMOC on sea ice albedo, concentration and thickness is largely heterogeneous, with mostly positive causal links in the Arctic Ocean and east of Greenland,

4 Experiments

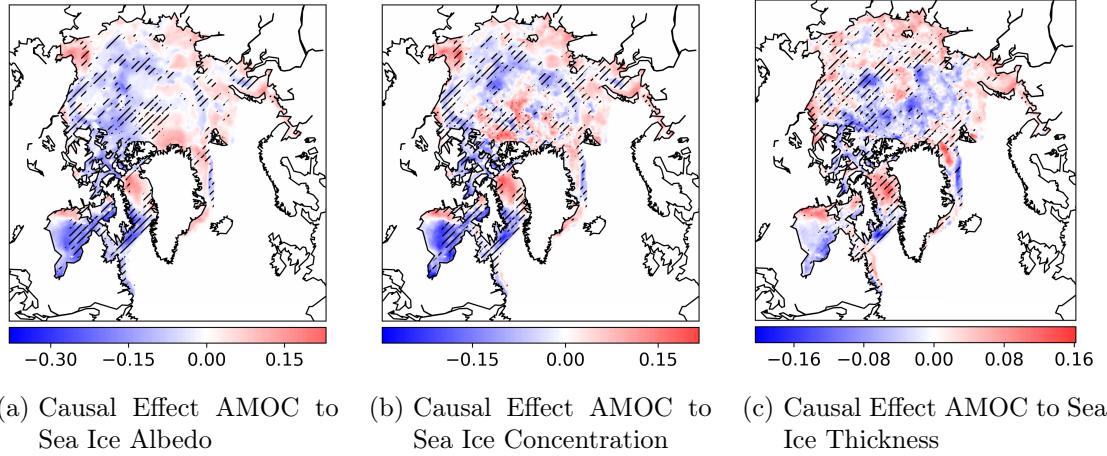


Figure 4.26: Spatially distributed causal effects of AMOC to sea ice according to the Linear Mediation method with hatched areas indicating significance at error rate $\alpha = 0.05$ determined by PCMCI

while the Labrador Sea and adjacent areas exhibit negative causal effects, that would match with the link detected by PCMCI. Figure 4.26 shows the spatial distributions.

Spatial analysis with the LKIF and GCSS methods shows few areas with significant causal effects, indicating low robustness of the previous results, see Section A2.

Causal Stationarity: The causal stationarity analysis exhibits some inconsistencies in causal effects over the analyzed time periods, but appears to support the bidirectional causal effects we identify. Figures 4.27a and 4.27b show the stationarity results for the extended causal network of sea ice extent depicted in Figure 4.20, while Figure 4.27c and 4.27d show the results for aggregated sea ice concentration and AMOC, with previous results in Figure 4.22b. In both constellations, PCMCI detects statistically significant links, but the sign of the causal effects jumps between positive and negative. Linear Mediation mostly confirms the sign of causal effects found for sea ice albedo, where a strong AMOC leads to sea ice melting, and sea ice melting weakens the AMOC.

A reason for the inconsistencies in causal stationarity analysis could be the small amount of samples for time windows of 20 years. As we still mask data in causal parents to only show five summer months, the number of available samples in each time window drops below 200. Our results on model systems imply relatively low confidence in any causal method under such circumstances.

4.4.2 Discussion

We summarize our findings on the interactions of sea ice extent, albedo, concentration and thickness with the AMOC. We identify three causal links and describe potential underlying physical mechanisms, summarized in Table 4.4.

Firstly, we find a robust causal link from the AMOC to Arctic sea ice extent, by which

4.4 Application to Tipping Point Interactions

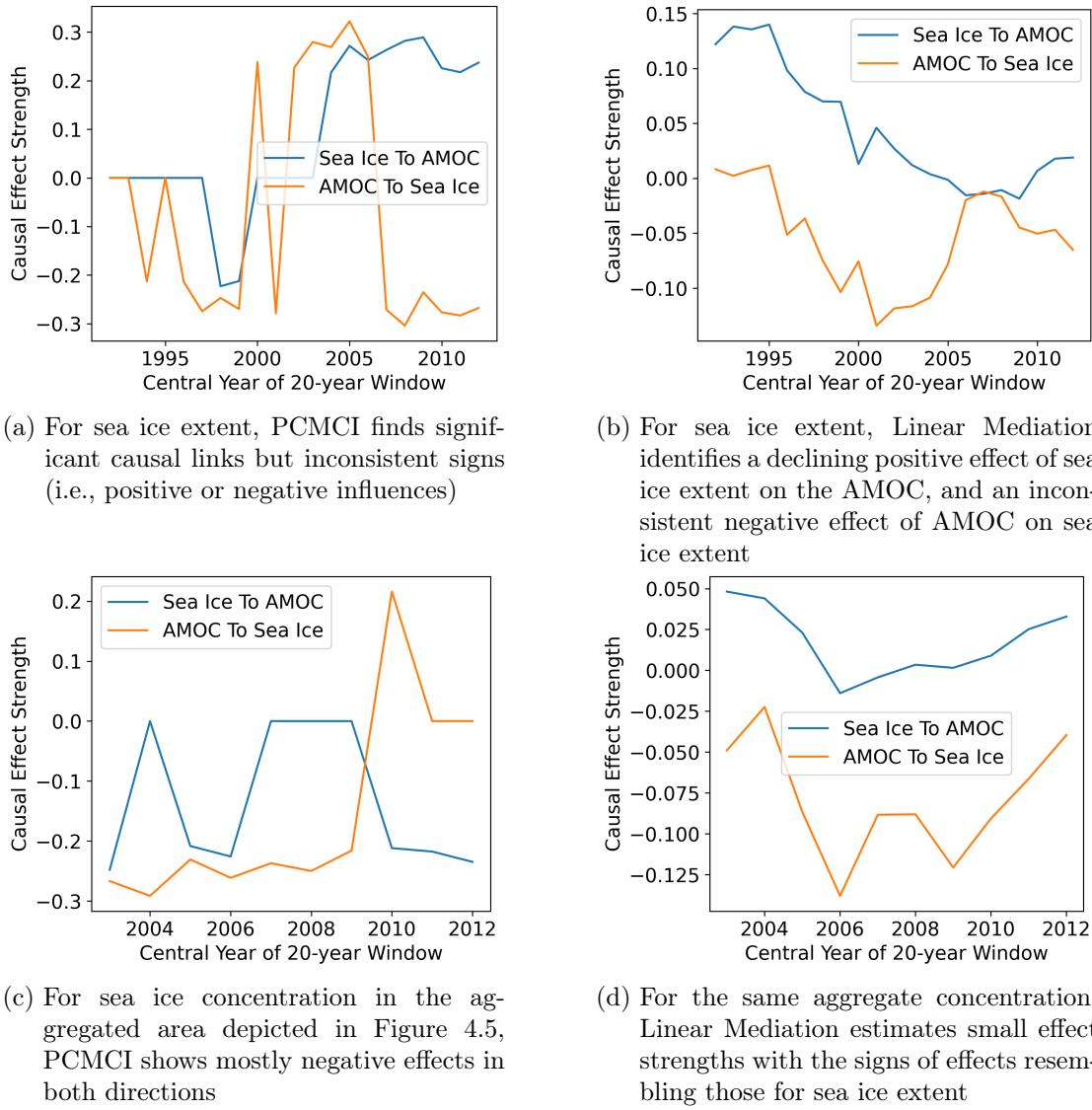


Figure 4.27: Causal stationarity analysis for AMOC interaction with sea ice extent and spatially aggregated sea ice concentration

a weaker AMOC leads to an increase in sea ice extent. This causal network remains intact even when including the CLLJ and SAR precipitation. The effect on sea ice could be explained by the reduced northward heat transport of a weakened AMOC, as suggested in previous studies [44, 47], although the time scale of one to two months identified in our causal analysis is significantly faster than expected from literature.

Secondly, a slow effect from sea ice to AMOC is identified in analysis on sea ice albedo and thickness and implies a common trend, i.e., a decrease in sea ice would lead to a decrease in AMOC strength with a delay of 5 months. Two mechanisms could explain

4 Experiments

Table 4.4: Causal links between AMOC strength and Arctic sea ice

Direction & Effect	Estimated Strength	Physical Explanation
Weaker AMOC → More Sea Ice	$-1 \text{ Sv} \rightarrow +13600 \text{ km}^2$ sea ice extent	Decreased northward oceanic heat transport [44, 46, 47]
Less Sea Ice → Weaker AMOC	$-0.01 \text{ mean surface albedo} \rightarrow -0.05 \text{ Sv}$	Freshwater influx [53, 49, 54] or increased sea surface heating from radiation [51, 52]
Less Sea Ice → Stronger AMOC	$-0.01 \text{ mean surface albedo} \rightarrow +0.06 \text{ Sv}$	Increased heat loss of sea surface [50]

this link:

- Additional freshwater influx from melting sea ice reduces the surface salinity in convection regions of the AMOC, which in turn weakens the overturning mechanism. This mechanism is also supported by several model studies [49, 53, 54].
- Reducing the area covered in sea ice could increase insolation-induced heating of surface water and weaken AMOC convection induced by temperature gradients between surface and bottom water. However, studies only detect this phenomenon on decadal time scales [51, 52].

This link from Arctic sea ice to AMOC is supported further by spatial analysis of sea ice concentration and albedo, which shows a relatively homogeneous positive effect across the Arctic Ocean and around Greenland. However, the results on spatially resolved sea ice thickness are more ambiguous. Causal stationarity analysis largely supports this positive link from sea ice to AMOC strength.

Finally, we also discover a fast negative link from sea ice to AMOC in the sea ice albedo, concentration and thickness variables, by which a decrease in sea ice could lead to a stronger AMOC. This link hints towards a competing physical effect in the overturning dynamic: Wu et al. [50] suggest that with a larger area of open ocean in edge regions of sea ice, more heat can be transported from the ocean to the atmosphere, in turn resulting in cooling of surface water. Since the temperature gradient of surface and bottom water is a major driver of the downwelling mechanism, this phenomenon can increase the formation of bottom water in these regions, and therefore strengthen the AMOC through a decrease in sea ice.

Our causal effect strength estimates using the Linear Mediation method suggest that the impact of sea ice on AMOC is more significant than the other way around. A decrease in average surface albedo of 0.01 (i.e., 1% reflectivity) could already weaken the AMOC by 0.05 Sv. Given the recent decreases in average surface albedo in the sea ice edge

4.4 Application to Tipping Point Interactions

region of about 0.034 per decade, further deterioration of Arctic sea ice due to climate change could significantly reduce the resilience of the AMOC.

However, robustness tests with other indices of the AMOC, spatial analysis and causal stationarity analysis demonstrate limits to our confidence in these results. These inconsistencies call for more detailed research, potentially including models or additional observational data. Further investigation into the temporal and spatial patterns of the two-fold contribution of sea ice to AMOC strength could also explore the risks of decreasing summer sea ice and extreme sea ice melting events to AMOC overturning and bottom water formation.

4 Experiments

5 Related Work

We present work that applies the causal methods used in this thesis to other components of the climate system to demonstrate the widespread use and detailed analysis conducted with these methods. Following that, we show some methodologically related work that discusses causal inference algorithms and extensions of the methods applied in this work. Finally, we give an overview of relevant work on climate tipping point interactions, including other approaches to estimate interaction strength and providing the context of tipping point network analysis that this work fits into.

Causal Methods in Climate System Analysis

Causal inference methods have been applied in numerous areas of climate science: The LKIF method has been applied most prominently by Docquier et al. [20] to determine causal drivers of Arctic sea ice, although they focus on regional factors like oceanic heat transport and near-surface temperature rather than the AMOC. PCMCI is used most commonly in the context of climate system analysis [15, 89, 90, 91], with a notable application to AMOC convection by Falkena et al. [92] who analyze the role of SST and salinity in the subpolar gyre convection with data from CMIP6 models. Their results support the assumption of our work that SST and salinity changes have a significant and measurable impact on convection strength. From a methodological perspective, the application of causal methods to model data is especially useful in scenarios where model interventions are unfeasible.

Spatially resolved causal analysis can reveal detailed physical mechanisms, as shown e.g., by Di Capua et al. [17], who apply PCMCI to spatially resolved atmospheric variables at smaller time scales and identify several physical mechanisms that explain their findings on different spatial and temporal scales. Despite the inconsistent results in our application, causal stationarity analysis can not only serve as a robustness test but also detect trends in causal effects: Tian et al. [93] demonstrate its explanatory potential on the interactions of dry and hot weather extremes, where focus is directed towards the changes of causal effects in a warming climate, i.e., deriving trends in causality rather than a fixed network. Similarly, Du et al. [94] identify changes in causal effects on the Indian Summer Monsoon over the past four decades.

The combination of different causal approaches we use in this work is similarly conducted by Shao et al. [95], who analyze the effect of vegetation cover on regional temperatures and use three different causal methods to gather a fuller picture of the underlying causality. We recommend such combined approaches especially given the results of our synthetic experiments.

5 Related Work

Methodological Extensions and Alternatives

A good overview of causal methods and their suitability to climate science can be found in Runge et al. [14]. However, some of the discussed methods do not fulfill the requirements of a rigorous multivariate analysis, like convergent cross mapping [96] that only provides qualitative indicators of significance rather than testable criteria. Docquier et al. [64] have also compared the LKIF and PCMCI methods on VAR and continuous systems, which finds both algorithms to perform well on simple models and confirms the advantage of LKIF in nonlinear continuous systems, as found in this work.

Further recent developments of causal inference methods include a time-dependent version of the LKIF algorithm, where parameters are estimated in time slices using Kalman filtering, a parallel to the GCSS approach [97]. Iglesias-Suarez et al. [98] integrate causal methods into a deep learning model where uninformative input parameters are pruned using PCMCI to reduce biases due to spurious correlations in input data. Granger causality has also been suggested as a method in climate change attribution, where the adaptation of model parameters to resemble counterfactual conditions without climate change allows for an estimate of the impact of climate change [99].

Estimation of Tipping Point Interactions

Analysis on tipping point interactions as conducted in this thesis can inform tipping point network models qualitatively to confirm causal links and quantitatively to estimate effect strengths. Such networks can capture risks of tipping cascades under different climate change scenarios [11, 100] and identify crucial sources of uncertainty [101, 102]. Tipping points and tipping point interactions can also be applied to social systems or combined into human-Earth or World-Earth interactions in the context of climate change [23], highlighting the bidirectional influences of climate change and human systems like resource allocation, political stability or displacement [103].

Where interactions are qualitatively established and conceptualized with simple differential equations in a network model, one can also use the additional assumptions to fit the model to observational data more directly, given sufficient data availability. A recent framework for parameter estimation of multivariate nonlinear stochastic differential equations [104] has been applied to a conceptual AMOC model [9], but could be extended to a network model as well. Parameter estimation for a fixed structure of differential equations can also be conducted using machine learning, building on the approach of neural ordinary differential equations [105].

An alternative approach to determine tipping point interactions is the use of Earth system models of intermediate complexity like CLIMBER-X [106], which also model systems like vegetation and permafrost that larger models usually neglect, while still maintaining significantly higher computation speeds. These capabilities allow for comprehensive simulations of tipping element interactions and tipping events, as conducted by Sinet et al. [107] for West Antarctica and the AMOC, or by Kaufhold et al. [108], stressing the feedback loops between permafrost and climate change. This approach is especially useful for climate tipping elements where observational data is sparse, like for the polar

ice sheets, but might produce less accurate results than more complex models and is clearly less faithful than observational data.

5 Related Work

6 Conclusions

Climate tipping elements and their interactions are among the biggest factors of uncertainty regarding the future trajectory of climate change and its impacts. We hypothesize that one can estimate such interactions from observation data using causal inference methods. We conduct experiments to evaluate and compare the performance of three causal methods, the Peter-Clark Momentary Conditional Independence (PCMCI), the Liang-Kleeman Information Flow (LKIF) and the Granger Causality for State Space Models (GCSS). The synthetic data generation is performed with two models that are commonly used to approximate properties of physical dynamic systems, a linear vector autoregressive model (VAR) and coupled cubic differential equations, which can exhibit saddle-node bifurcations that resemble tipping points.

We find the detection skill of all three algorithms to depend mainly on the number of samples provided and the underlying interaction strength of variables. For intermediate amounts of samples (100 - 500 samples) in the nonlinear system, the LKIF method performs significantly better than the other two, while the GCSS method benefits from larger amounts of samples and outperforms the other methods from 5,000 samples. For a linear model system and parameter configurations expected in experiments on climate tipping elements, all three methods provide similar results with a Matthews Correlation Coefficient (MCC) of about 0.6 (where 1 indicates perfect predictions and 0 random predictions). Therefore, we have confidence that the methods can detect correct causal links in many cases, but it is advisable to apply multiple methods and conduct robustness tests to increase this confidence further. We also conduct experiments on a spatially resolved model and find encouraging results for the detection of causal parents of spatially resolved variables and their effect distribution. For outgoing effects of a spatially resolved variable, aggregation can lead to successful causal detection, with the LKIF algorithm more suited for systems with low autocorrelation and the other methods performing better for slower systems.

Future work could explore the parameter space further, use different model systems, or improve the modeling of spatially resolved systems and evaluate implications for causal analysis. A growing body of work attempts to derive causality using neural network approaches, which could result in a more model-agnostic analysis that handles nonlinearity more easily than existing methods.

We analyze interactions of Arctic sea ice and the Atlantic Meridional Overturning Circulation (AMOC), building on previous work that connected the AMOC to the Amazon rainforest [36]. We utilize the causal methods discussed before, with a focus on PCMCI due to its higher flexibility. In a series of experiments at different spatio-temporal resolutions and using observational and reanalysis data, we detect multiple causal links between Arctic sea ice and the AMOC.

6 Conclusions

The most robust link implies that a weakening of the AMOC could increase summer sea ice extent, which can be explained by decreased heat transport into the subpolar and Arctic seas. Two competing causal effects are found for the effect of sea ice on the AMOC: Reductions in sea ice albedo and thickness weaken the AMOC significantly on time scales of several months, likely due to freshwater influx into its crucial convection sites. A faster link is also identified, where sea ice melting would strengthen the AMOC with a lag of one month. This effect could be due to increased areas of exposed ocean surface, increased heat loss and therefore sea surface cooling in the AMOC convection sites. However, ambiguous findings on sea ice concentration and causal stationarity tests demonstrate the limited robustness of our results.

As previous model results on causal links between Arctic sea ice and the AMOC are ambiguous in their findings, future efforts could apply the presented causal methods to simulation data to test our identified network, or extend the existing model intervention experiments to a system that models several climate tipping elements. Further knowledge on tipping point interactions in general is crucial for a broader risk assessment of tipping points and cascades and the implications of ongoing climate change.

List of Acronyms

AMOC	Atlantic Meridional Overturning Circulation
CLLJ	Caribbean Low Level Jet
CMIP	Coupled Model Intercomparison Project
DAG-GNN	Directed Acyclic Graph - Graph Neural Network
ENSO	El Niño Southern Oscillation
ERA5	ECMWF Reanalysis v5
GCSS	Granger Causality for State Space Models
HadSST	Hadley Centre Sea Ice and Sea Surface Temperature
IFSS	Innovations Form State Space Model
i.i.d.	independent and identically distributed
IPCC	Intergovernmental Panel on Climate Change
ITCZ	Intertropical Convergence Zone
LKIF	Liang-Kleeman Information Flow
MCC	Matthews Correlation Coefficient
NSIDC	National Snow and Ice Data Center
PCMCI	Peter-Clark Momentary Conditional Independence
ROC	Receiver Operating Characteristics
SAM	Structural Agnostic Modelling
SAR	Southern Amazon Rainforest
SCM	Structural Causal Model
SI	Sea Ice
SINDy	Sparse Identification of Nonlinear Dynamics
SST	Sea Surface Temperature
TCI-Net	Time Series Causal Inference Network
VAR	Vector autoregressive
VARMA	Vector autoregressive moving average

List of Acronyms

Bibliography

- [1] G. Lentze. “2024 was the warmest year on record, Copernicus data show.” (Jan. 2025), [Online]. Available: <https://www.ecmwf.int/en/about/media-centre/news/2025/2024-was-warmest-year-record-copernicus-data-show> (visited on 04/21/2025).
- [2] E. Bevacqua, C.-F. Schleussner, and J. Zscheischler, “A year above 1.5 °C signals that Earth is most probably within the 20-year period that will reach the Paris Agreement limit,” *Nature Climate Change*, pp. 1–4, Feb. 2025. DOI: 10.1038/s41558-025-02246-9.
- [3] T. M. Lenton *et al.*, “Tipping elements in the Earth’s climate system,” *Proceedings of the National Academy of Sciences*, vol. 105, no. 6, pp. 1786–1793, Feb. 2008. DOI: 10.1073/pnas.0705414105.
- [4] D. I. Armstrong McKay *et al.*, “Exceeding 1.5°C global warming could trigger multiple climate tipping points,” *Science*, vol. 377, no. 6611, eabn7950, Sep. 2022. DOI: 10.1126/science.abn7950.
- [5] T. M. Lenton *et al.*, *The global tipping points report 2023*, 2023.
- [6] N. Boers and M. Rypdal, “Critical slowing down suggests that the western Greenland Ice Sheet is close to a tipping point,” *Proceedings of the National Academy of Sciences*, vol. 118, no. 21, e2024192118, May 2021. DOI: 10.1073/pnas.2024192118.
- [7] J. Feldmann and A. Levermann, “Collapse of the West Antarctic Ice Sheet after local destabilization of the Amundsen Basin,” *Proceedings of the National Academy of Sciences*, vol. 112, no. 46, pp. 14 191–14 196, Nov. 2015. DOI: 10.1073/pnas.1512482112.
- [8] L. Caesar, S. Rahmstorf, A. Robinson, G. Feulner, and V. Saba, “Observed fingerprint of a weakening Atlantic Ocean overturning circulation,” *Nature*, vol. 556, no. 7700, pp. 191–196, Apr. 2018. DOI: 10.1038/s41586-018-0006-5.
- [9] P. Ditlevsen and S. Ditlevsen, “Warning of a forthcoming collapse of the Atlantic meridional overturning circulation,” *Nature Communications*, vol. 14, no. 1, p. 4254, Jul. 25, 2023. DOI: 10.1038/s41467-023-39810-w.
- [10] C. A. Boulton, T. M. Lenton, and N. Boers, “Pronounced loss of Amazon rainforest resilience since the early 2000s,” *Nature Climate Change*, vol. 12, no. 3, pp. 271–278, Mar. 2022. DOI: 10.1038/s41558-022-01287-8.

Bibliography

- [11] N. Wunderling, J. F. Donges, J. Kurths, and R. Winkelmann, “Interacting tipping elements increase risk of climate domino effects under global warming,” *Earth System Dynamics*, vol. 12, no. 2, pp. 601–619, Jun. 2021. DOI: [10.5194/esd-12-601-2021](https://doi.org/10.5194/esd-12-601-2021).
- [12] M. Willeit, A. Ganopolski, A. Robinson, and N. R. Edwards, “The earth system model climber-x v1.0 – part 1: Climate model description and validation,” *Geoscientific Model Development*, vol. 15, no. 14, pp. 5905–5948, 2022. DOI: [10.5194/gmd-15-5905-2022](https://doi.org/10.5194/gmd-15-5905-2022).
- [13] N. Wunderling *et al.*, “Climate tipping point interactions and cascades: A review,” *Earth System Dynamics*, vol. 15, no. 1, pp. 41–74, Jan. 2024. DOI: [10.5194/esd-15-41-2024](https://doi.org/10.5194/esd-15-41-2024).
- [14] J. Runge *et al.*, “Inferring causation from time series in Earth system sciences,” *Nature Communications*, vol. 10, no. 1, p. 2553, Jun. 2019. DOI: [10.1038/s41467-019-10105-3](https://doi.org/10.1038/s41467-019-10105-3).
- [15] M. Kretschmer, D. Coumou, J. F. Donges, and J. Runge, “Using Causal Effect Networks to Analyze Different Arctic Drivers of Midlatitude Winter Circulation,” *Journal of Climate*, vol. 29, no. 11, pp. 4069–4081, Jun. 2016. DOI: [10.1175/JCLI-D-15-0654.1](https://doi.org/10.1175/JCLI-D-15-0654.1).
- [16] J. Runge, P. Nowack, M. Kretschmer, S. Flaxman, and D. Sejdinovic, “Detecting and quantifying causal associations in large nonlinear time series datasets,” *Science Advances*, vol. 5, no. 11, eaau4996, Nov. 2019. DOI: [10.1126/sciadv.aau4996](https://doi.org/10.1126/sciadv.aau4996).
- [17] G. Di Capua, E. Tyrlis, D. Matei, and R. V. Donner, “Tropical and mid-latitude causal drivers of the eastern Mediterranean Etesians during boreal summer,” *Climate Dynamics*, vol. 62, no. 10, pp. 9565–9585, Oct. 2024. DOI: [10.1007/s00382-024-07411-y](https://doi.org/10.1007/s00382-024-07411-y).
- [18] X. S. Liang, “Normalized Multivariate Time Series Causality Analysis and Causal Graph Reconstruction,” *Entropy*, vol. 23, no. 6, p. 679, Jun. 2021. DOI: [10.3390/e23060679](https://doi.org/10.3390/e23060679).
- [19] L. Barnett and A. K. Seth, “Granger causality for state-space models,” *Physical Review E*, vol. 91, no. 4, p. 040101, Apr. 23, 2015. DOI: [10.1103/PhysRevE.91.040101](https://doi.org/10.1103/PhysRevE.91.040101).
- [20] D. Docquier, S. Vannitsem, F. Ragone, K. Wyser, and X. S. Liang, “Causal Links Between Arctic Sea Ice and Its Potential Drivers Based on the Rate of Information Transfer,” *Geophysical Research Letters*, vol. 49, no. 9, e2021GL095892, 2022. DOI: [10.1029/2021GL095892](https://doi.org/10.1029/2021GL095892).
- [21] C. D. Brummitt, G. Barnett, and R. M. D’Souza, “Coupled catastrophes: Sudden shifts cascade and hop among interdependent systems,” *Journal of The Royal Society Interface*, vol. 12, no. 112, p. 20150712, Nov. 2015. DOI: [10.1098/rsif.2015.0712](https://doi.org/10.1098/rsif.2015.0712).

- [22] M. Scheffer and S. R. Carpenter, “Catastrophic regime shifts in ecosystems: Linking theory to observation,” *Trends in Ecology & Evolution*, vol. 18, no. 12, pp. 648–656, Dec. 2003. DOI: [10.1016/j.tree.2003.09.002](https://doi.org/10.1016/j.tree.2003.09.002).
- [23] C. L. E. Franzke *et al.*, “Perspectives on tipping points in integrated models of the natural and human Earth system: Cascading effects and telecoupling,” *Environmental Research Letters*, vol. 17, no. 1, p. 015004, Jan. 2022. DOI: [10.1088/1748-9326/ac42fd](https://doi.org/10.1088/1748-9326/ac42fd).
- [24] E. K. Smith, M. Wiedermann, J. F. Donges, J. Heitzig, and R. Winkelmann, “A global threshold model of enabling conditions for social tipping in pro-environmental behaviours – the role of sea level rise anticipation and climate change concern,” *Earth System Dynamics*, vol. 16, no. 2, pp. 545–564, Apr. 2025. DOI: [10.5194/esd-16-545-2025](https://doi.org/10.5194/esd-16-545-2025).
- [25] M. Milkoreit *et al.*, “Defining tipping points for social-ecological systems scholarship—an interdisciplinary literature review,” *Environmental Research Letters*, vol. 13, no. 3, p. 033005, Mar. 2018. DOI: [10.1088/1748-9326/aaaa75](https://doi.org/10.1088/1748-9326/aaaa75).
- [26] T. M. Lenton *et al.*, “Climate tipping points — too risky to bet against,” *Nature*, vol. 575, no. 7784, pp. 592–595, Nov. 2019. DOI: [10.1038/d41586-019-03595-0](https://doi.org/10.1038/d41586-019-03595-0).
- [27] H. A. Dijkstra, *Nonlinear Climate Dynamics*. Cambridge University Press, Jun. 2013.
- [28] H. F. Diaz, M. P. Hoerling, and J. K. Eischeid, “ENSO variability, teleconnections and climate change,” *International Journal of Climatology*, vol. 21, no. 15, pp. 1845–1862, Dec. 2001. DOI: [10.1002/joc.631](https://doi.org/10.1002/joc.631).
- [29] J. Ye, C. Yuan, M. Yang, X. Lu, J.-J. Luo, and T. Yamagata, “Delayed Impacts of ENSO on the Frequency of Summer Extreme Hot Days in the Asian Monsoon Region. Part I: Observation, Historical Simulation, and Future Projection in CMIP6 Models,” *Journal of Climate*, vol. 36, no. 9, pp. 3095–3112, Apr. 2023. DOI: [10.1175/JCLI-D-21-0667.1](https://doi.org/10.1175/JCLI-D-21-0667.1).
- [30] N. Wunderling *et al.*, “Climate tipping point interactions and cascades: A review,” *Earth System Dynamics*, vol. 15, no. 1, pp. 41–74, Jan. 2024. DOI: [10.5194/esd-15-41-2024](https://doi.org/10.5194/esd-15-41-2024).
- [31] E. Kriegler, J. W. Hall, H. Held, R. Dawson, and H. J. Schellnhuber, “Imprecise probability assessment of tipping points in the climate system,” *Proceedings of the National Academy of Sciences*, vol. 106, no. 13, pp. 5041–5046, Mar. 2009. DOI: [10.1073/pnas.0809117106](https://doi.org/10.1073/pnas.0809117106).
- [32] K. Zickfeld, A. Levermann, M. G. Morgan, T. Kuhlbrodt, S. Rahmstorf, and D. W. Keith, “Expert judgements on the response of the Atlantic meridional overturning circulation to climate change,” *Climatic Change*, vol. 82, no. 3, pp. 235–265, Jun. 2007. DOI: [10.1007/s10584-007-9246-3](https://doi.org/10.1007/s10584-007-9246-3).

Bibliography

- [33] Y. Cai, T. M. Lenton, and T. S. Lontzek, “Risk of multiple interacting tipping points should encourage rapid CO₂ emission reduction,” *Nature Climate Change*, vol. 6, no. 5, pp. 520–525, May 2016. DOI: [10.1038/nclimate2964](https://doi.org/10.1038/nclimate2964).
- [34] R. M. van Westen, M. Kliphuis, and H. A. Dijkstra, “Physics-based early warning signal shows that AMOC is on tipping course,” *Science Advances*, vol. 10, no. 6, eadk1189, Feb. 2024. DOI: [10.1126/sciadv.adk1189](https://doi.org/10.1126/sciadv.adk1189).
- [35] G. D. McCarthy and L. Caesar, “Can we trust projections of AMOC weakening based on climate models that cannot reproduce the past?” *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 381, no. 2262, p. 20220193, Dec. 2023. DOI: [10.1098/rsta.2022.0193](https://doi.org/10.1098/rsta.2022.0193).
- [36] A. Högner, G. D. Capua, J. F. Donges, R. V. Donner, G. Feulner, and N. Wunderling, *Causal pathway from AMOC to Southern Amazon rainforest indicates stabilising interaction between two climate tipping elements*, Jan. 24, 2025. DOI: [10.48550/arXiv.2501.14374](https://arxiv.org/abs/2501.14374).
- [37] D. Lannuzel *et al.*, “The future of Arctic sea-ice biogeochemistry and ice-associated ecosystems,” *Nature Climate Change*, vol. 10, no. 11, pp. 983–992, Nov. 2020. DOI: [10.1038/s41558-020-00940-4](https://doi.org/10.1038/s41558-020-00940-4).
- [38] Konsortium Nationale Interdisziplinäre Klimarisiko-Einschätzung, *Nationale Interdisziplinäre Klimarisiko-Einschätzung*, Feb. 2025.
- [39] A. Jahn, M. M. Holland, and J. E. Kay, “Projections of an ice-free Arctic Ocean,” *Nature Reviews Earth & Environment*, vol. 5, no. 3, pp. 164–176, Mar. 2024. DOI: [10.1038/s43017-023-00515-9](https://doi.org/10.1038/s43017-023-00515-9).
- [40] M. W. Buckley and J. Marshall, “Observations, inferences, and mechanisms of the Atlantic Meridional Overturning Circulation: A review,” *Reviews of Geophysics*, vol. 54, no. 1, pp. 5–63, 2016. DOI: [10.1002/2015RG000493](https://doi.org/10.1002/2015RG000493).
- [41] N. Boers, “Observation-based early-warning signals for a collapse of the Atlantic Meridional Overturning Circulation,” *Nature Climate Change*, vol. 11, no. 8, pp. 680–688, Aug. 2021. DOI: [10.1038/s41558-021-01097-4](https://doi.org/10.1038/s41558-021-01097-4).
- [42] D. Swingedouw, M.-N. Houssais, C. Herbaut, A.-C. Blaizot, M. Devilliers, and J. Deshayes, “AMOC Recent and Future Trends: A Crucial Role for Oceanic Resolution and Greenland Melting?” *Frontiers in Climate*, vol. 4, Apr. 4, 2022. DOI: [10.3389/fclim.2022.838310](https://doi.org/10.3389/fclim.2022.838310).
- [43] K. H. Kilbourne *et al.*, “Atlantic circulation change still uncertain,” *Nature Geoscience*, vol. 15, no. 3, pp. 165–167, Mar. 2022. DOI: [10.1038/s41561-022-00896-4](https://doi.org/10.1038/s41561-022-00896-4).
- [44] W. Weijer, A. Siddiqui, W. Cheng, M. Veneziani, and P. Kurtakoti, “Interactions between the Arctic Mediterranean and the Atlantic Meridional Overturning Circulation: A Review,” *Oceanography*, 2022. DOI: [10.5670/oceanog.2022.130](https://doi.org/10.5670/oceanog.2022.130).
- [45] B. Rudels and E. Carmack, “Arctic Ocean Water Mass Structure and Circulation,” *Oceanography*, vol. 35, no. 3/4, pp. 52–65, 2022.

- [46] D. Docquier and T. Koenigk, “A review of interactions between ocean heat transport and Arctic sea ice,” *Environmental Research Letters*, vol. 16, no. 12, p. 123 002, Nov. 2021. DOI: [10.1088/1748-9326/ac30be](https://doi.org/10.1088/1748-9326/ac30be).
- [47] S. Mahajan, R. Zhang, and T. L. Delworth, “Impact of the Atlantic Meridional Overturning Circulation (AMOC) on Arctic Surface Air Temperature and Sea Ice Variability,” *Journal of Climate*, vol. 24, no. 24, pp. 6573–6581, Dec. 15, 2011. DOI: [10.1175/2011JCLI4002.1](https://doi.org/10.1175/2011JCLI4002.1).
- [48] T. Tsubouchi *et al.*, “Increased ocean heat transport into the Nordic Seas and Arctic Ocean over the period 1993–2016,” *Nature Climate Change*, vol. 11, no. 1, pp. 21–26, Jan. 2021. DOI: [10.1038/s41558-020-00941-3](https://doi.org/10.1038/s41558-020-00941-3).
- [49] W. Liu and A. Fedorov, “Interaction between Arctic sea ice and the Atlantic meridional overturning circulation in a warming climate,” *Climate Dynamics*, vol. 58, no. 5–6, pp. 1811–1827, Mar. 2022. DOI: [10.1007/s00382-021-05993-5](https://doi.org/10.1007/s00382-021-05993-5).
- [50] Y. Wu, D. P. Stevens, I. A. Renfrew, and X. Zhai, “The Response of the Nordic Seas to Wintertime Sea Ice Retreat,” *Journal of Climate*, vol. 34, no. 15, pp. 6041–6056, Aug. 1, 2021. DOI: [10.1175/JCLI-D-20-0932.1](https://doi.org/10.1175/JCLI-D-20-0932.1).
- [51] F. Sévellec, A. V. Fedorov, and W. Liu, “Arctic sea-ice decline weakens the Atlantic Meridional Overturning Circulation,” *Nature Climate Change*, vol. 7, no. 8, pp. 604–610, Aug. 2017. DOI: [10.1038/nclimate3353](https://doi.org/10.1038/nclimate3353).
- [52] M. T. Jenkins and A. Dai, “Arctic Climate Feedbacks in ERA5 Reanalysis: Seasonal and Spatial Variations and the Impact of Sea-Ice Loss,” *Geophysical Research Letters*, vol. 49, no. 16, e2022GL099263, 2022. DOI: [10.1029/2022GL099263](https://doi.org/10.1029/2022GL099263).
- [53] H. Li, A. Fedorov, and W. Liu, “AMOC Stability and Diverging Response to Arctic Sea Ice Decline in Two Climate Models,” *Journal of Climate*, vol. 34, no. 13, pp. 5443–5460, Jul. 1, 2021. DOI: [10.1175/JCLI-D-20-0572.1](https://doi.org/10.1175/JCLI-D-20-0572.1).
- [54] R. M. van Westen, V. Jacques-Dumas, A. A. Boot, and H. A. Dijkstra, “The Role of Sea Ice Insulation Effects on the Probability of AMOC Transitions,” *Journal of Climate*, vol. 37, no. 23, pp. 6269–6284, Nov. 8, 2024. DOI: [10.1175/JCLI-D-24-0060.1](https://doi.org/10.1175/JCLI-D-24-0060.1).
- [55] D. Nian *et al.*, “A potential collapse of the Atlantic Meridional Overturning Circulation may stabilise eastern Amazonian rainforests,” *Communications Earth & Environment*, vol. 4, no. 1, pp. 1–11, Dec. 2023. DOI: [10.1038/s43247-023-01123-7](https://doi.org/10.1038/s43247-023-01123-7).
- [56] T. K. Akabane *et al.*, “Weaker Atlantic overturning circulation increases the vulnerability of northern Amazon forests,” *Nature Geoscience*, vol. 17, no. 12, pp. 1284–1290, Dec. 2024. DOI: [10.1038/s41561-024-01578-z](https://doi.org/10.1038/s41561-024-01578-z).
- [57] P. Ashwin, C. Perryman, and S. Wieczorek, “Parameter shifts for nonautonomous systems in low dimension: Bifurcation- and rate-induced tipping,” *Nonlinearity*, vol. 30, no. 6, pp. 2185–2210, Jun. 2017. DOI: [10.1088/1361-6544/aa675b](https://doi.org/10.1088/1361-6544/aa675b).

Bibliography

- [58] A. K. Klose, V. Karle, R. Winkelmann, and J. F. Donges, “Emergence of cascading dynamics in interacting tipping elements of ecology and climate,” *Royal Society Open Science*, vol. 7, no. 6, p. 200599, Jun. 2020. DOI: [10.1098/rsos.200599](https://doi.org/10.1098/rsos.200599).
- [59] A. M. Lyapunov, *General Problem of the Stability Of Motion*. CRC Press, Aug. 1992.
- [60] J. Runge, “Causal network reconstruction from time series: From theoretical assumptions to practical estimation,” *Chaos: An Interdisciplinary Journal of Nonlinear Science*, vol. 28, no. 7, p. 075310, Jul. 23, 2018. DOI: [10.1063/1.5025050](https://doi.org/10.1063/1.5025050).
- [61] J. Runge, P. Nowack, M. Kretschmer, S. Flaxman, and D. Sejdinovic, “Detecting and quantifying causal associations in large nonlinear time series datasets,” *Science Advances*, vol. 5, no. 11, eaau4996, Nov. 27, 2019. DOI: [10.1126/sciadv.aau4996](https://doi.org/10.1126/sciadv.aau4996).
- [62] J. Runge *et al.*, “Identifying causal gateways and mediators in complex spatio-temporal systems,” *Nature Communications*, vol. 6, no. 1, p. 8502, Oct. 2015. DOI: [10.1038/ncomms9502](https://doi.org/10.1038/ncomms9502).
- [63] X. S. Liang, “Information flow and causality as rigorous notions ab initio,” *Physical Review E*, vol. 94, no. 5, p. 052201, Nov. 1, 2016. DOI: [10.1103/PhysRevE.94.052201](https://doi.org/10.1103/PhysRevE.94.052201).
- [64] D. Docquier, G. Di Capua, R. V. Donner, C. A. L. Pires, A. Simon, and S. Vanvitsem, “A comparison of two causal methods in the context of climate analyses,” *Nonlinear Processes in Geophysics*, vol. 31, no. 1, pp. 115–136, Feb. 27, 2024. DOI: [10.5194/npg-31-115-2024](https://doi.org/10.5194/npg-31-115-2024).
- [65] J. F. Geweke, “Measures of Conditional Linear Dependence and Feedback Between Time Series,” *Journal of the American Statistical Association*, vol. 79, no. 388, pp. 907–915, 1984. DOI: [10.2307/2288723](https://doi.org/10.2307/2288723).
- [66] L. Barnett and A. K. Seth, *Detectability of Granger causality for subsampled continuous-time neurophysiological processes*, Sep. 2016.
- [67] L. Barnett and A. K. Seth, “Inferring the temporal structure of directed functional connectivity in neural systems: Some extensions to Granger causality,” in *2019 IEEE International Conference on Systems, Man and Cybernetics (SMC)*, Oct. 2019, pp. 4395–4402. DOI: [10.1109/SMC.2019.8914556](https://doi.org/10.1109/SMC.2019.8914556).
- [68] Y. Yu, J. Chen, T. Gao, and M. Yu, “DAG-GNN: DAG Structure Learning with Graph Neural Networks,” in *Proceedings of the 36th International Conference on Machine Learning*, PMLR, May 24, 2019, pp. 7154–7163.
- [69] D. Kalainathan, O. Goudet, I. Guyon, D. Lopez-Paz, and M. Sebag, *Structural Agnostic Modeling: Adversarial Learning of Causal Graphs*, Jul. 25, 2022. DOI: [10.48550/arXiv.1803.04929](https://doi.org/10.48550/arXiv.1803.04929).
- [70] S. Ali *et al.*, *Quantifying Causes of Arctic Amplification via Deep Learning based Time-series Causal Inference*, Sep. 25, 2023. DOI: [10.48550/arXiv.2303.07122](https://doi.org/10.48550/arXiv.2303.07122).

Bibliography

- [71] S. L. Brunton, J. L. Proctor, and J. N. Kutz, “Discovering governing equations from data by sparse identification of nonlinear dynamical systems,” *Proceedings of the National Academy of Sciences*, vol. 113, no. 15, pp. 3932–3937, Apr. 12, 2016. DOI: 10.1073/pnas.1517384113.
- [72] N. Lohmann. “Causal inference pipeline.” (2025), [Online]. Available: <https://github.com/loh-nik/Causal-Inference-Pipeline> (visited on 05/09/2025).
- [73] D. L. Strahl. “Inference of Climate Tipping Interactions: From Dynamical Systems to Data-Driven Perspectives.” (May 2024), [Online]. Available: <https://github.com/david-strahl/master-thesis> (visited on 03/17/2025).
- [74] A. Neff, A. Keane, H. A. Dijkstra, and B. Krauskopf, “Bifurcation analysis of a North Atlantic Ocean box model with two deep-water formation sites,” *Physica D: Nonlinear Phenomena*, vol. 456, p. 133907, Dec. 15, 2023. DOI: 10.1016/j.physd.2023.133907.
- [75] J. Runge, “Discovering contemporaneous and lagged causal relations in autocorrelated nonlinear time series datasets,” in *Proceedings of the 36th Conference on Uncertainty in Artificial Intelligence (UAI)*, PMLR, Aug. 2020, pp. 1388–1397.
- [76] Y. Rong. “Lkif.” (2024), [Online]. Available: <https://github.com/YinengRong/LKIF> (visited on 04/18/2025).
- [77] L. Barnett. “Lcbarnett/ssgc.” (2020), [Online]. Available: <https://github.com/lcbarnett/ssgc> (visited on 04/18/2025).
- [78] D. Chicco and G. Jurman, “The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation,” *BMC Genomics*, vol. 21, p. 6, Jan. 2, 2020. DOI: 10.1186/s12864-019-6413-7.
- [79] L. Zanna, S. Khatiwala, J. M. Gregory, J. Ison, and P. Heimbach, “Global reconstruction of historical ocean heat storage and transport,” *Proceedings of the National Academy of Sciences*, vol. 116, no. 4, pp. 1126–1131, Jan. 22, 2019. DOI: 10.1073/pnas.1808838115.
- [80] N. Wunderling *et al.*, “Recurrent droughts increase risk of cascading tipping events by outpacing adaptive capacities in the Amazon rainforest,” *Proceedings of the National Academy of Sciences*, vol. 119, no. 32, e2120777119, Aug. 9, 2022. DOI: 10.1073/pnas.2120777119.
- [81] Fetterer, F.; Knowles K.; Meier W.; Savoie M.; Windnagel A., *Sea ice index, version 3*, 2017. DOI: 10.7265/N5K072F8.
- [82] E.U. Copernicus Marine Service Information (CMEMS). Marine Data Store (MDS)., *Arctic ocean sea ice reanalysis*, 2024. DOI: 10.48670/mds-00336.
- [83] T. Williams, A. Korosov, P. Rampal, and E. Ólason, “Presentation and evaluation of the arctic sea ice forecasting system nextsim-f,” *The Cryosphere*, vol. 15, no. 7, pp. 3207–3227, 2021. DOI: 10.5194/tc-15-3207-2021.
- [84] H. Hersbach *et al.*, *Era5 monthly averaged data on single levels from 1940 to present*, 2023. DOI: 10.24381/cds.f17050d7.

Bibliography

- [85] G. M. Pontes and L. Men viel, “Weakening of the Atlantic Meridional Overturning Circulation driven by subarctic freshening since the mid-twentieth century,” *Nature Geoscience*, vol. 17, no. 12, pp. 1291–1298, Dec. 2024. DOI: [10.1038/s41561-024-01568-1](https://doi.org/10.1038/s41561-024-01568-1).
- [86] J. J. Kennedy, N. A. Rayner, C. P. Atkinson, and R. E. Killick, “An Ensemble Data Set of Sea Surface Temperature Change From 1850: The Met Office Hadley Centre HadSST.4.0.0.0 Data Set,” *Journal of Geophysical Research: Atmospheres*, vol. 124, no. 14, pp. 7719–7763, 2019. DOI: [10.1029/2018JD029867](https://doi.org/10.1029/2018JD029867).
- [87] B. M. Flores *et al.*, “Critical transitions in the Amazon forest system,” *Nature*, vol. 626, no. 7999, pp. 555–564, Feb. 2024. DOI: [10.1038/s41586-023-06970-0](https://doi.org/10.1038/s41586-023-06970-0).
- [88] W. N. Meier *et al.*, “NOAA Arctic Report Card 2024 : Sea Ice,” 2024. DOI: [10.25923/aksk-7p66](https://doi.org/10.25923/aksk-7p66).
- [89] J. D. Kromer and L. D. Trusel, “Identifying the Impacts of Sea Ice Variability on the Climate and Surface Mass Balance of West Antarctica,” *Geophysical Research Letters*, vol. 50, no. 18, e2023GL104436, Sep. 2023. DOI: [10.1029/2023GL104436](https://doi.org/10.1029/2023GL104436).
- [90] S. Karmouche, E. Galytska, G. A. Meehl, J. Runge, K. Weigel, and V. Eyring, “Changing effects of external forcing on Atlantic–Pacific interactions,” *Earth System Dynamics*, vol. 15, no. 3, pp. 689–715, Jun. 2024. DOI: [10.5194/esd-15-689-2024](https://doi.org/10.5194/esd-15-689-2024).
- [91] C. Krich *et al.*, “Estimating causal networks in biosphere–atmosphere interaction with the PCMCI approach,” *Biogeosciences*, vol. 17, no. 4, pp. 1033–1061, Feb. 2020. DOI: [10.5194/bg-17-1033-2020](https://doi.org/10.5194/bg-17-1033-2020).
- [92] S. K. J. Falkena, H. A. Dijkstra, and A. S. v. d. Heydt, *Causal Mechanisms of Subpolar Gyre Variability in CMIP6 Models*, Apr. 2025. DOI: [10.48550/arXiv.2408.16541](https://doi.org/10.48550/arXiv.2408.16541).
- [93] Y. Tian *et al.*, “Historical changes in the Causal Effect Networks of compound hot and dry extremes in central Europe,” *Communications Earth & Environment*, vol. 5, no. 1, pp. 1–14, Dec. 2024. DOI: [10.1038/s43247-024-01934-2](https://doi.org/10.1038/s43247-024-01934-2).
- [94] D. Du, A. C. Subramanian, W. Han, U. Ninad, and J. Runge, “Causal Analysis Discovers an Enhanced Impact of Tropical Western Pacific on Indian Summer Monsoon Subseasonal Anomalies,” *Geophysical Research Letters*, vol. 51, no. 10, e2023GL106431, May 2024. DOI: [10.1029/2023GL106431](https://doi.org/10.1029/2023GL106431).
- [95] Y. Shao, D. F. T. Hagan, S. Li, F. Zhou, X. Zou, and P. Cabral, “The Many Shades of the Vegetation–Climate Causality: A Multimodel Causal Appreciation,” *Forests*, vol. 15, no. 8, p. 1430, Aug. 2024. DOI: [10.3390/f15081430](https://doi.org/10.3390/f15081430).
- [96] G. Sugihara *et al.*, “Detecting Causality in Complex Ecosystems,” *Science*, vol. 338, no. 6106, pp. 496–500, Oct. 2012. DOI: [10.1126/science.1227079](https://doi.org/10.1126/science.1227079).
- [97] F. Zhou *et al.*, “Estimating Time-Dependent Structures in a Multivariate Causality for Land–Atmosphere Interactions,” *Journal of Climate*, vol. 37, no. 6, pp. 1853–1876, Feb. 2024. DOI: [10.1175/JCLI-D-23-0207.1](https://doi.org/10.1175/JCLI-D-23-0207.1).

Bibliography

- [98] F. Iglesias-Suarez *et al.*, “Causally-Informed Deep Learning to Improve Climate Models and Projections,” *Journal of Geophysical Research: Atmospheres*, vol. 129, no. 4, e2023JD039202, 2024. DOI: [10.1029/2023JD039202](https://doi.org/10.1029/2023JD039202).
- [99] M. D. Risser, M. Ombadi, and M. F. Wehner, *Granger causal inference for climate change attribution*, Apr. 2025. DOI: [10.48550/arXiv.2408.16004](https://doi.org/10.48550/arXiv.2408.16004).
- [100] S. Sinet, A. S. Von Der Heydt, and H. A. Dijkstra, “AMOC Stabilization Under the Interaction With Tipping Polar Ice Sheets,” *Geophysical Research Letters*, vol. 50, no. 2, e2022GL100305, Jan. 2023. DOI: [10.1029/2022GL100305](https://doi.org/10.1029/2022GL100305).
- [101] J. P. Rosser, R. Winkelmann, and N. Wunderling, “Polar ice sheets are decisive contributors to uncertainty in climate tipping projections,” *Communications Earth & Environment*, vol. 5, no. 1, pp. 1–11, Nov. 2024. DOI: [10.1038/s43247-024-01799-5](https://doi.org/10.1038/s43247-024-01799-5).
- [102] S. Sinet, P. Ashwin, A. S. von der Heydt, and H. A. Dijkstra, “AMOC stability amid tipping ice sheets: The crucial role of rate and noise,” *Earth System Dynamics*, vol. 15, no. 4, pp. 859–873, Jul. 2024. DOI: [10.5194/esd-15-859-2024](https://doi.org/10.5194/esd-15-859-2024).
- [103] V. Spaiser *et al.*, “Negative social tipping dynamics resulting from and reinforcing Earth system destabilization,” *Earth System Dynamics*, vol. 15, no. 5, pp. 1179–1206, Sep. 2024. DOI: [10.5194/esd-15-1179-2024](https://doi.org/10.5194/esd-15-1179-2024).
- [104] P. Pilipovic, A. Samson, and S. Ditlevsen, “Parameter estimation in nonlinear multivariate stochastic differential equations based on splitting schemes,” *The Annals of Statistics*, vol. 52, no. 2, pp. 842–867, Apr. 2024. DOI: [10.1214/24-AOS2371](https://doi.org/10.1214/24-AOS2371).
- [105] H. Aliee, F. J. Theis, and N. Kilbertus, *Beyond Predictions in Neural ODEs: Identification and Interventions*, Jan. 2025. DOI: [10.48550/arXiv.2106.12430](https://doi.org/10.48550/arXiv.2106.12430).
- [106] M. Willeit *et al.*, “The Earth system model CLIMBER-X v1.0 – Part 2: The global carbon cycle,” *Geoscientific Model Development*, vol. 16, no. 12, pp. 3501–3534, Jun. 2023. DOI: [10.5194/gmd-16-3501-2023](https://doi.org/10.5194/gmd-16-3501-2023).
- [107] S. Sinet, A. S. v. d. Heydt, and H. A. Dijkstra, *West Antarctic Meltwater can Prevent an AMOC Collapse*, Feb. 2025. DOI: [10.48550/arXiv.2502.17104](https://doi.org/10.48550/arXiv.2502.17104).
- [108] C. Kaufhold, M. Willeit, S. Talento, A. Ganopolski, and J. Rockström, “Interplay between climate and carbon cycle feedbacks could substantially enhance future warming,” *Environmental Research Letters*, vol. 20, no. 4, p. 044 027, Apr. 2025. DOI: [10.1088/1748-9326/adb6be](https://doi.org/10.1088/1748-9326/adb6be).

Bibliography

Appendix

A1 Method Comparison for Slow Bidirectional Grid Systems

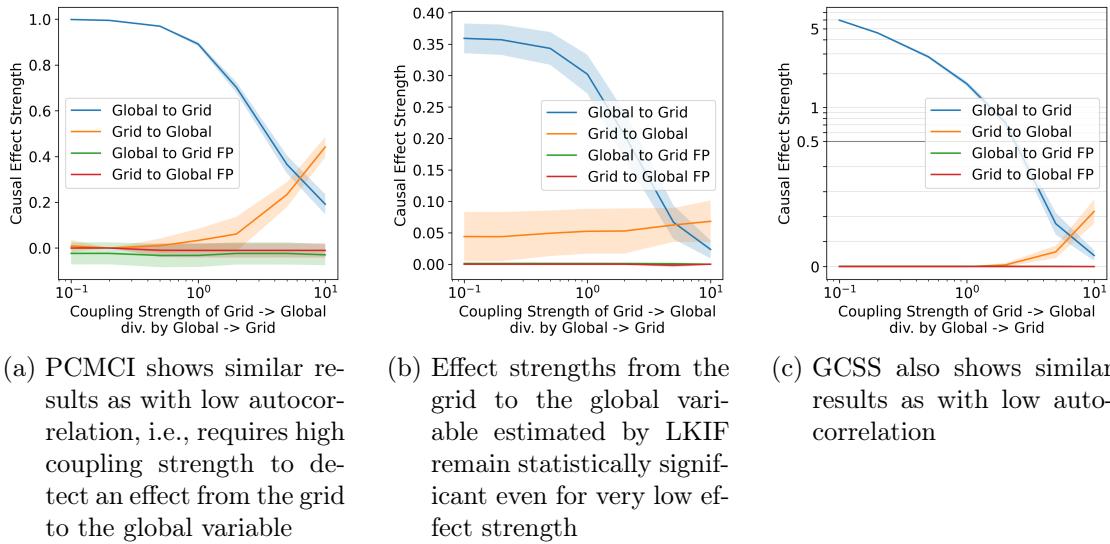


Figure A.1: Bidirectional detection skill at a diffusion factor of 4 (i.e., an autocorrelation factor of 0.94)

We discuss the causality direction from grid cells to the global variable more closely, as it is the most difficult to detect for all methods. We confirm that the detection skill of LKIF deteriorates at higher autocorrelation, i.e., a lower diffusion factor (see Figure A.1). However, considering the previous results on LKIF performance deterioration in Figure 4.18, the bidirectionality of causalities seems to help LKIF find statistically significant results for unknown reasons.

PCMCI and GCSS show similar behavior to the low autocorrelation results in Figure 4.19. For a large effect strength from the grid to the global variable, PCMCI and GCSS do appear more suitable than LKIF for slow bidirectional systems. However, when the effect strength from the grid cells to the global variable is relatively low, LKIF is the only method that can still detect the effect, albeit not as reliably as for faster systems.

Appendix

A2 Robustness Tests for Interaction of Arctic Sea Ice and AMOC

We present further tests for robustness of the applied results of this work.

Figure A.2 displays results for the causal analysis on sea ice extent, albedo, concentration and thickness as presented in Section 4.4.1, but with the AMOC indices by Pontes & Menviel [85] and Ditlevsen & Ditlevsen [9]. We find that the Pontes index does not reproduce any of the earlier results, which could be due to the integration of South Atlantic temperatures, i.e., additional noise. The Ditlevsen index shows very similar results to the Caesar index used in the main analysis. Overall, this demonstrates the limited robustness of our results.

We conduct analysis on two further regional aggregations, one over the region East of Greenland between -45°W and 90°E , where we choose the 75th percentile of the variance of sea ice concentration, the other one over the West Greenland region between -90°W and -45°W at the 66th percentile of the variance of sea ice concentration. These percentile levels are chosen to roughly match the larger aggregated region of Figure 4.5. Figures A.3 and A.4 show the results of causal analysis with PCMCI on the East and West Greenland regions respectively, where the West Greenland region largely confirms the two stabilizing links found for the larger regional aggregation in Section 4.4.1, while the Eastern region only exhibits the slower destabilizing link for sea ice albedo and what we assume to be a false positive for sea ice concentration. This spatial separation of stabilizing and destabilizing effects could motivate further work, e.g., exploring spatial patterns in coupled ocean-ice models.

Next we present results of spatial analysis with LKIF and GCSS, which only includes the AMOC and sea ice variables in the underlying pairwise tests. We determine the causal effect for all areas and indicate significant areas with hatching, see Figure A.5 and A.6. GCSS uses the entire time series as it does not offer masking functionality, and finds almost no significant areas. LKIF results show relatively inconsistent spatial patterns for the direction of sea ice to AMOC with significance mainly in outer regions of the Arctic Ocean, as well as around Greenland and Svalbard. For the causal effects from AMOC to sea ice, almost no significant areas are found in analysis with LKIF. A possible conclusion from the spatial patterns found in LKIF analysis is that it supports the stronger interactions of West Greenland regions with AMOC, see Figure A.4, although the heterogeneity of spatial results do not allow for high confidence in such an interpretation.

A2 Robustness Tests for Interaction of Arctic Sea Ice and AMOC

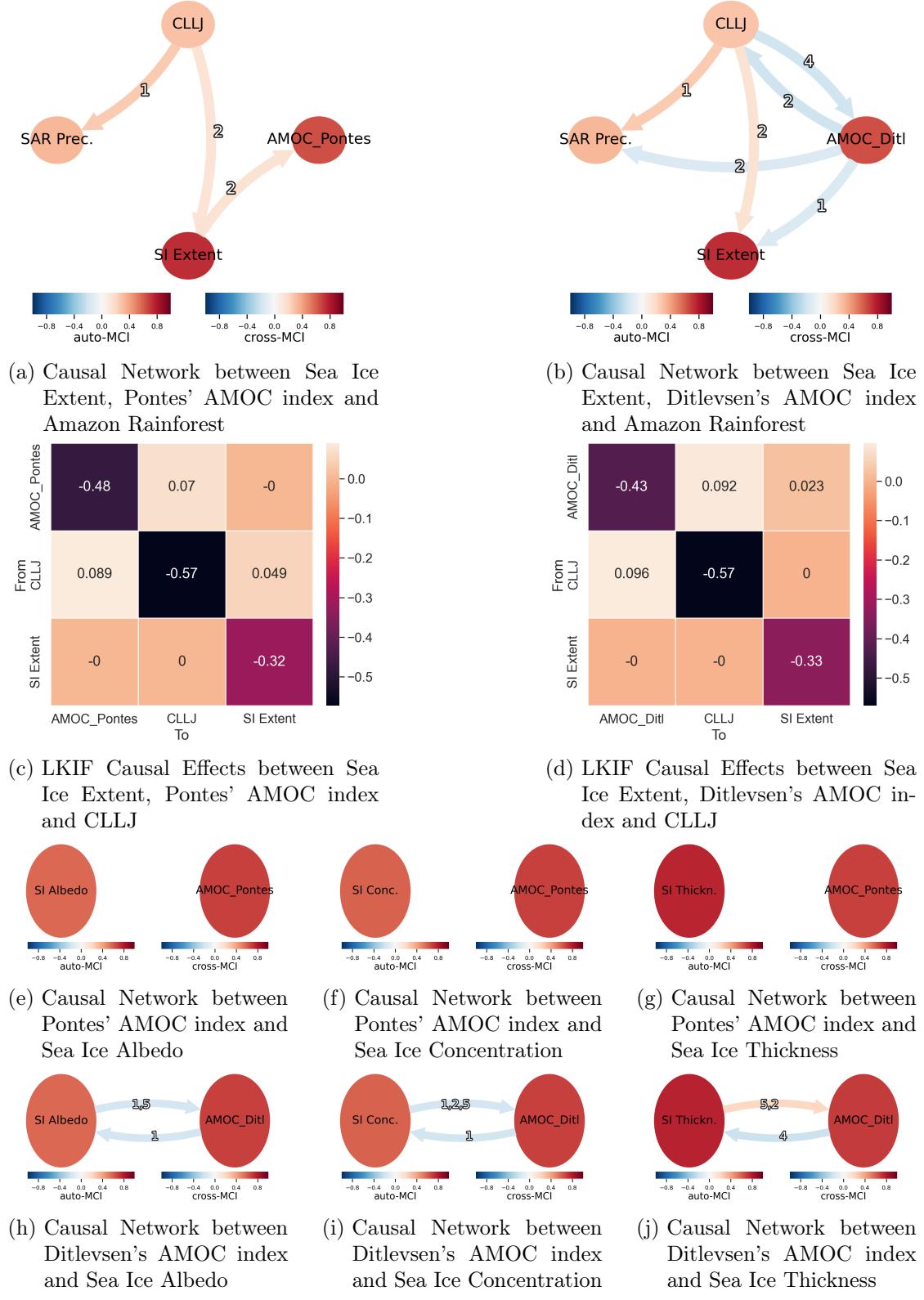
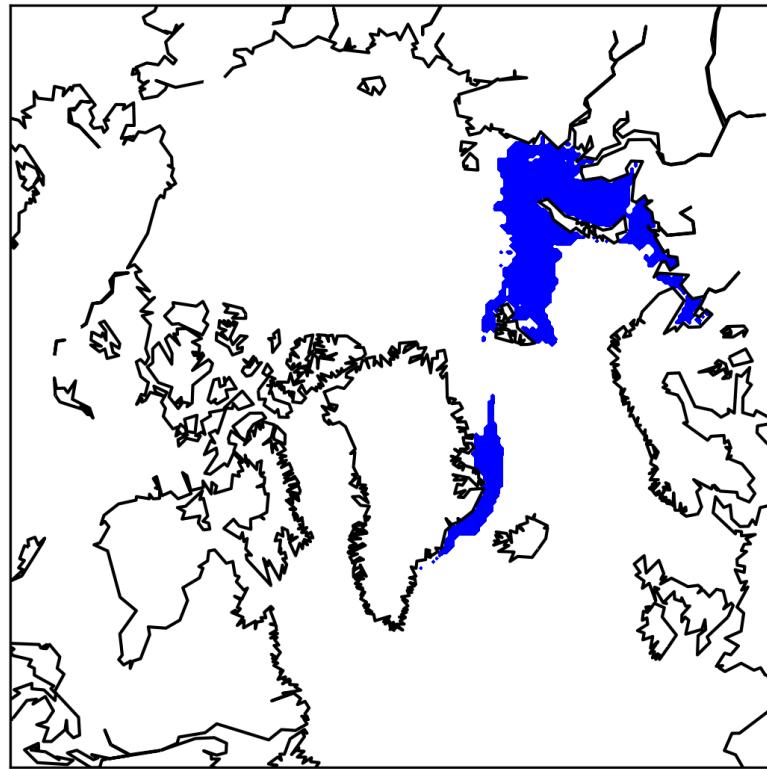
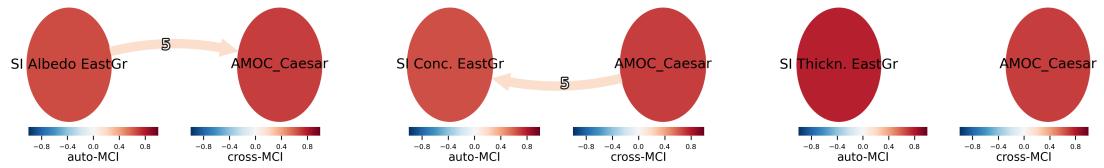


Figure A.2: Analysis with AMOC indices by Pontes & Menviel [85] and Ditlevsen & Caesar [9] on sea ice extent (a-d) and the aggregated region (e-j). Pontes' index does not show any causal effects for regional analysis while Ditlevsen's index largely supports the findings with the index by Caesar et al. [8]

Appendix



(a) Definition of the East Greenland Sea Ice Edge Region



(b) Causal Network between AMOC and Sea Ice Albedo

(c) Causal Network between AMOC and Sea Ice Concentration

(d) Causal Network between AMOC and Sea Ice Thickness

Figure A.3: PCMCI results for the sea ice edge regions East of Greenland show the slow destabilizing link from sea ice albedo to AMOC but otherwise does not reproduce any of the results for the larger region

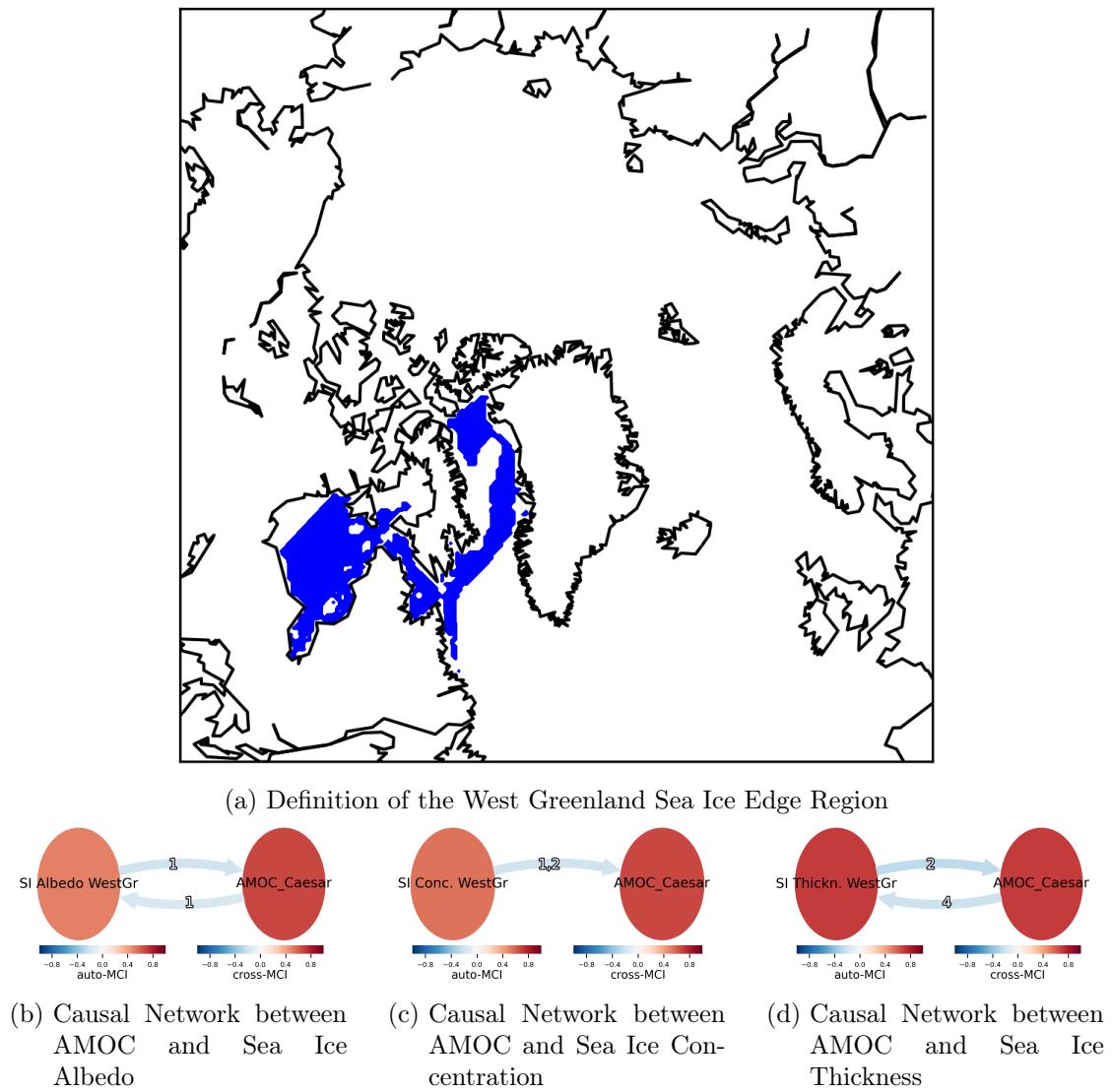


Figure A.4: PCMCI results for the sea ice edge regions West of Greenland reproduce the stabilizing bidirectional links, but not the destabilizing one found for sea ice albedo and thickness for the larger aggregation region

Appendix

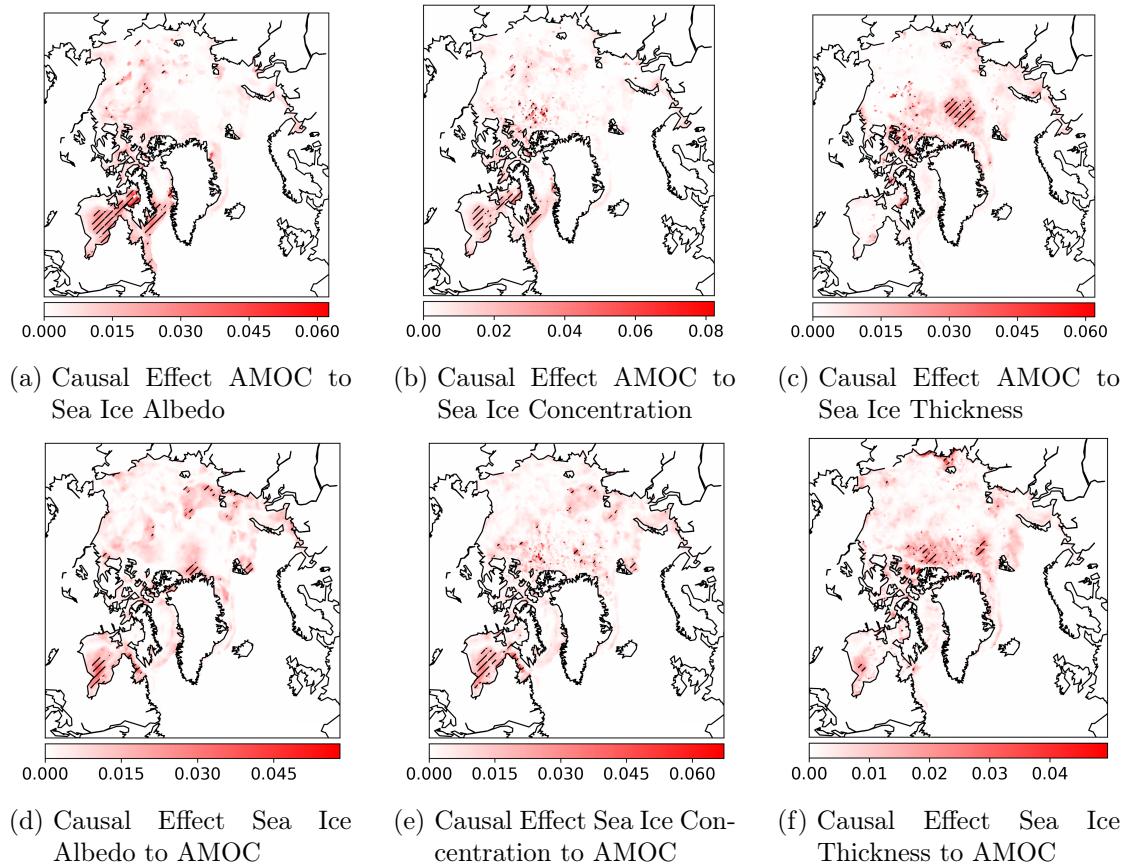


Figure A.5: Spatially distributed causal effects of a-c) AMOC to sea ice and d-f) sea ice to AMOC according to the GCSS method with hatched areas indicating significance at error rate $\alpha = 0.05$

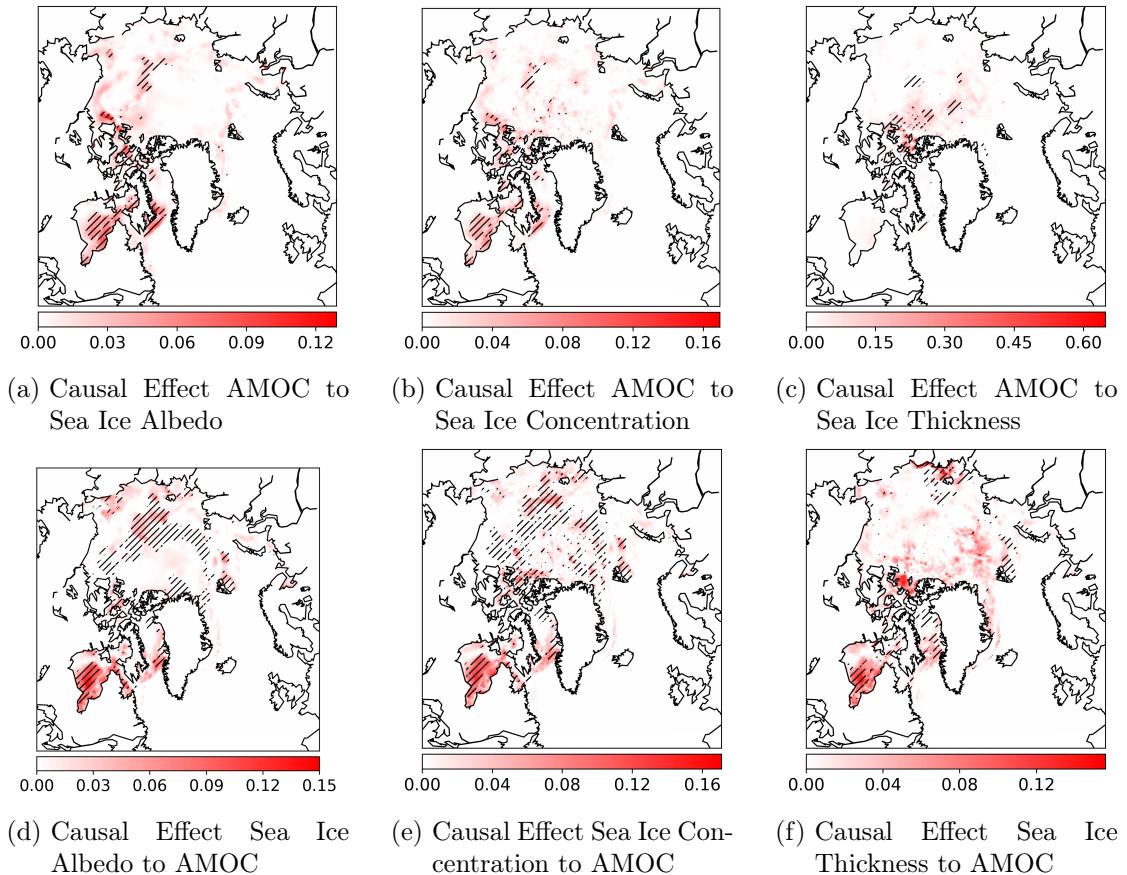


Figure A.6: Spatially distributed causal effects of a-c) AMOC to sea ice and d-f) sea ice to AMOC according to the LKIF method with hatched areas indicating significance at error rate $\alpha = 0.05$