Causal inference for time series

Jakob Runge[1,2*], Andreas Gerhardus[1], Gherardo Varando[3], Veronika Eyring[4,5] and Gustau Camps-Valls[3]

[1*]Deutsches Zentrum für Luft- und Raumfahrt (DLR), Institut für Datenwissenschaften, Jena, Germany.

[2]Technische Universität Berlin, Institute of Computer Engineering and Microelectronics, Berlin, Germany.

[3]Image Processing Laboratory (IPL), Universitat de València, Paterna (València), Spain.

[4]Deutsches Zentrum für Luft- und Raumfahrt (DLR), Institut für Physik der Atmosphäre, Weßling, Germany.

[5]University of Bremen, Institute of Environmental Physics (IUP), Bremen, Germany.

*Corresponding author(s). E-mail(s): jakob.runge@dlr.de;

Abstract

Many research questions in the Earth and environmental sciences are inherently causal, requiring robust analyses to establish whether and how changes in one variable cause changes in another. Causal inference provides the theoretical foundations to use data and qualitative domain knowledge to quantitatively answer these questions, complementing statistics and machine learning techniques. However, there is still a broad language gap between the methodological and domain science communities. In this Technical Review, we explain the use of causal inference frameworks with a focus on the challenges of time series data. Domain-adapted explanations, method guidance, and practical case studies provide an accessible summary of methods for causal discovery and causal effect estimation. Examples from climate and biogeosciences illustrate typical challenges, such as contemporaneous causation, hidden confounding and non-stationarity, and some strategies to address these challenges. Integrating causal thinking into data-driven science will facilitate process understanding and more robust machine learning and statistical models for Earth and environmental sciences, allowing to tackle many open problems with relevant environmental, economic, and societal implications.

Key points

1. Causal inference provides a framework that integrates statistical and machine learning methods to answer causal questions from data.
2. A causal inference analysis enables research questions to be framed as causal questions and transparently lay out the underlying assumptions used to answer these.
3. Causal discovery can be used to learn causal graphs from data to explore and cross-check qualitative causal knowledge.
4. Causal effect estimation allows quantitative causal questions to be answered using a combination of qualitative causal knowledge, statistical or machine learning models, and data.
5. Case studies on climate and biogeosciences exemplify the use of causal inference methods and illustrate typical challenges, such as contemporaneous causation, hidden confounding and non-stationarity.
6. Future developments require more interactions between method and domain sciences to better integrate causal thinking into data-driven science.

Introduction

The Earth is a highly interconnected dynamical system. Some of the most substantial challenges in Earth System Sciences are related to achieving a causal understanding of the internal processes and anthropogenic impacts on the system, with far-reaching environmental, economic, and societal implications [1]. However, an experimental approach to learning causal relations in Earth sciences, at least on a large scale, is infeasible owing to ethical, practical, and/or economic reasons. Instead, observational data and climate model simulations [2] provide a wealth of Earth System data. These (often heterogeneous) data have led to an increasing application of machine learning (ML) and statistical methods [3–6] for finding patterns and associations in Earth and environmental systems [7, 8].

Despite the outstanding predictive capabilities of the current ML and deep learning methods, they are limited. Process understanding and more robust machine learning models that generalize outside of the learning distribution require a clear understanding of the underlying causal relationships between variables [9–11]. Indeed, an ML model might predict ice cream sales from sunburn rates, but this is not a causal relationship: the model will not help creamery owners determine how to boost ice cream sales, and the predictions will fail when better sunscreens become available. Many Earth science questions are inherently causal [10, 12–14]; for example, causality is pivotal for a better understanding of physical processes like the microphysics of cloud formation [15] and large-scale atmospheric teleconnections [16–22], and for causally robust forecasting [23, 24], evaluating and improving the physics embedded in climate models [10, 25–27], or determining effective land-use strategies and attributing climate change and extreme events through counterfactuals [28].

Causal inference strives to discover the system's causal structure and quantify causal effects by combining domain knowledge, ML models, and observational or interventional data [9, 29–32]. Starting out a few decades ago from research questions in social sciences and medicine, today many causal inference methods and tools are available to also address challenges in dynamical systems as in Earth and Environmental sciences [10]. But there is some confusion about how and when to apply them, and how to interpret their results.

In this Technical Review, we provide an accessible overview of causal inference, with the goal of guiding domain scientists to frame and understand their problems from a causal perspective and find the appropriate methods to tackle them. We begin by overviewing the foundations of causal inference, then discuss causal discovery, where the task is to learn the qualitative causal graph, and causal effect estimation, where qualitative knowledge in the form of a causal graph is already assumed and the task is to quantify causal effects. These frameworks are illustrated in two case studies that discuss fundamental challenges such as contemporaneous causation, mediation, feedbacks, hidden confounding, nonlinearity, regime dependence, and non-stationarity. The Review also provides an online supplement with Python code to work through the examples and a guide to further tools and software packages.

Foundations of causal inference

This section introduces the basic terminology of the structural causal model (SCM) and causal-graphical-model frameworks for causal inference [31–33]. Another approach to causal inference is the potential outcome framework, which is popular in the social sciences [34–37]. Both frameworks are equivalent from a theoretical point of view, but the graphical approach can be easier to apply because assumptions are conveniently explicated in the form of causal graphs.

Structural causal models. In the SCM framework the basic underlying assumption about causal relationships within the system of study is that they are described by SCMs [32, 33, 36, 38, 39]. An SCM describes how each variable of the system is caused by other variables of the system and an exogeneous noise term. To represent the dynamic process variables underlying a multivariate time series $\mathbf{V}_t = (V_t^1, \ldots, V_t^N)$, this Review considers temporal SCMs (see upper left part of Fig. 1**a**) consisting of the structural assignments:

$$V_t^j := f^j(pa(V_t^j), \eta_t^j) \quad \text{for all } V_t^j \in \mathbf{V}_t \text{ and } t \in \mathbb{Z} \tag{1}$$

together with jointly independent random variables η_t^j. Here, each f^j is a potentially nonlinear functional causal mechanism that determines the value of the variable V_t^j based on the values of the variable's causal parents $pa(V_t^j)$ and the associated noise variable η_t^j. The noise variable η_t^j subsumes the effects

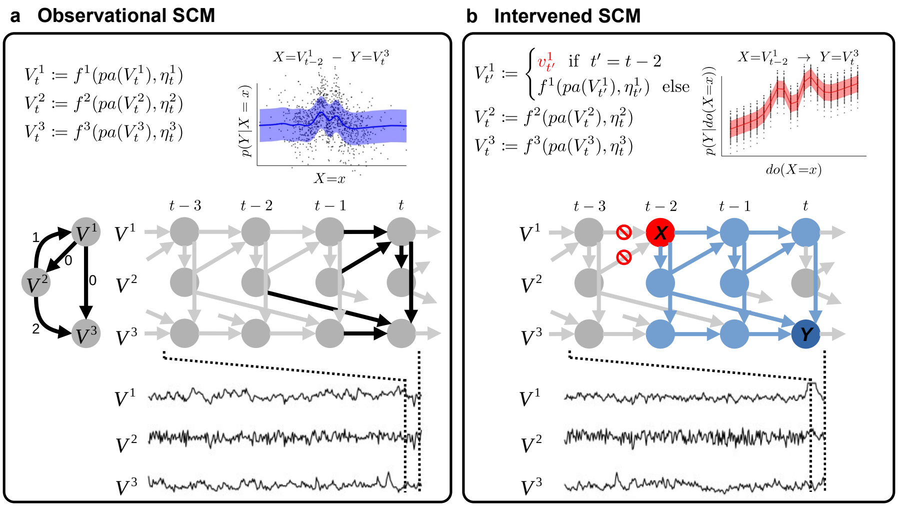


Fig. 1 Foundations of causal inference. a | Observational structural causal model (SCM, top left) for time series variables V^1, V^2, V^3 (bottom) with illustrative observational scatterplot (black dots) and the mean (blue curve) and 90% confidence interval (shaded blue area) of the conditional probability distribution $p(y \mid X = x)$ for $X = V_{t-2}^1$ and $Y = V_t^3$ (top right) as well as a graphical representation of the SCM as a time series graph (middle right). In an SCM, the variables' values are determined by functions $f(\cdot)$ of their causal parents $pa(\cdot)$ and noise variables η. In the time series graph, the causal relationships of the variables at time t with their parents are highlighted in black; these relationships repeat at every timestep and are presented in grey for other time steps. Due to this repeating structure, the graph can be presented in a simplified form as a summary graph (middle left), which contains an edge corresponding to every black edge in the time-series graph that connects two distinct variables. Simultaneous dependencies are presented in straight arrows, and time-lagged dependencies are presented with curved arrows labeled with the length of the time lag. **b** | Intervened SCM (top left) when (experimentally) setting $do(X = x)$ for $X = V_{t-2}^1$. This results in an interventional distribution of Y for $Y = V_t^3$ (top right) as well as a manipulated time series graph where here the parents of V_{t-2}^1 are removed. Light blue edges illustrate causal paths toward V_t^3. Causal inference is about learning the effect of an intervention (panel **b**) from observational data (panel **a**).

of all factors that are not part of the model and are unique to V_t^j; for example, the local weather noise in an SCM that models remote climate variables. The subscript t is used to explicitly indicate the time-dependence and $V_{t-\tau}^j$ refers to a lagged variable. Importantly, the structural assignments describe the system's causal mechanisms; this causal meaning is emphasized by using the symbol ':=' instead of '='.

The causal parents $pa(V_t^j)$, also called direct causes, are a subset of $\{\mathbf{V}_t, \ldots, \mathbf{V}_{t-\tau_{\max}}\} \setminus \{V_t^j\}$ with $\tau_{\max} \geq 0$ a non-negative integer. Restricting the lag τ in $V_{t-\tau}^i \in pa(V_t^j)$ to satisfy $0 \leq \tau \leq \tau_{\max}$ ensures that there are no causal influences backwards in time and that all direct causal influences occur with a finite time lag of at most $\tau_{\max}$ time steps. Zero is included because, due to the finite-time resolution, there can also be contemporaneous causal influences $V_t^i \to V_t^j$; for example, a one-week time lag in monthly-aggregated data.

This Review further assumes the SCM in eq. (1) to be causally stationary; that is, the causal relationships and noise distributions are assumed to be invariant in time. Ensuring this assumption is met might require data preprocessing. Alternatively, different parts of a time series can be represented by different SCMs to model non-stationarity.

Causal graphs. The causal graph of an SCM represents the causal parentships (direct causes) specified by the SCM, thereby retaining the information about qualitative cause-and-effect relationships but abstracting away the functional form of these relationships. The causal graph of the temporal SCM in eq. (1) is the directed graph with vertices V_t^j for $V_t^j \in \mathbf{V}_t$ for all $t \in \mathbb{Z}$ and with edges $V_{t-\tau}^i \to V_t^j$ if and only if $V_{t-\tau}^i \in pa(V_t^j)$. This time series graph [40, 41] is of infinite extension with a repetitive structure in time due to causal stationarity (middle right-hand side of Fig. 1**a**, the black edges repeat as grey edges at different time steps).

Only Directed Acyclic Graphs (DAGs) are considered here. Acyclicity only restricts the contemporaneous slices of the graph at a fixed time step because there cannot exist causal influences and hence no edges backward in time. Despite acyclicity it thus is possible to model a feedback cycle of two processes V_t^1 and V_t^2 that causally influence each other (Fig. 1**a**). However, at least one of the causal relationships must then be lagged. There is (technically more involved) work on causal inference with cyclic contemporaneous slices of causal graphs [42–44].

Often, some of the variables V_t^j are unobserved (either entirely or only at selected time points), even though their causal relationships are known. For example, aerosols can affect cloud formation, but aerosols are often not observed. Whenever some variables are not observed, it is convenient to work with a projected causal graph [45, 46] on the observed variables only to avoid explicitely representing a possibly large number of unobserved variables and their causal relationships. In these projected graphs, an edge $V_{t-\tau}^i \to V_t^j$ signifies a direct or indirect path (through unobserved variables) and a bidirected edge $V_{t-\tau}^i \leftrightarrow V_t^j$ signifies hidden (sometimes also referred to as latent or unobserved) confounding by one or more variables. That is, a bidirected edge $V_{t-\tau}^i \leftrightarrow V_t^j$ means there is at least one unobserved variable L^{ij} that causally influences both $V_{t-\tau}^i$ and V_t^j (for example, $V_{t-\tau}^i \leftarrow L_{t-\tau_L}^{ij} \to V_t^j$). The projected causal graphs are generally Acyclic Directed Mixed Graphs (ADMGs). SCMs without unobserved confounders are also called Markovian models because the variables' noise terms are independent. The observed variables in SCMs with unobserved confounding can have dependent noise terms and are called semi-Markovian models.

Interventions. SCMs allow to specify how systems react to interventions–idealizations of experiments that deactivate the original causal mechanism of a variable $V_{t'}^j$ and instead force the variable to take a prescribed value $v_{t'}^j$, while the system remains unaltered otherwise [33]. An intervention on $V_{t'}^j$

amounts to replacing $V^j_{t'} \coloneqq f^j(pa(V^j_{t'}), \eta^j_{t'})$ with $V^j_{t'} \coloneqq v^j_{t'}$ and is conventionally denoted by $do(V^j_{t'} = v^j_{t'})$. In climate modeling, interventions are routinely done, for example, by prescribing sea-surface temperatures to study their effect on atmospheric dynamics. Interventions can also only modify the causal mechanism, but keep the causal parents intact (so-called parametric interventions), and interventions on multiple variables or at multiple time points are possible. With $\mathbf{X}$ a set of variables, a joint intervention on all variables in $\mathbf{X}$ is denoted by $do(\mathbf{X} = \mathbf{x})$.

On the level of causal graphs, an intervention $do(\mathbf{X} = \mathbf{x})$ removes all edges that point into any of the $X \in \mathbf{X}$ (Fig. 1**b**), and the intervention's effect propagates along proper causal paths (light blue) starting at $\mathbf{X}$. It is important to keep in mind that doing is fundamentally distinct from observing, that is, $do(\mathbf{X} = \mathbf{x})$ is fundamentally distinct from $\mathbf{X} = \mathbf{x}$ (see the difference in distributions between Fig. 1**a** and **b**).

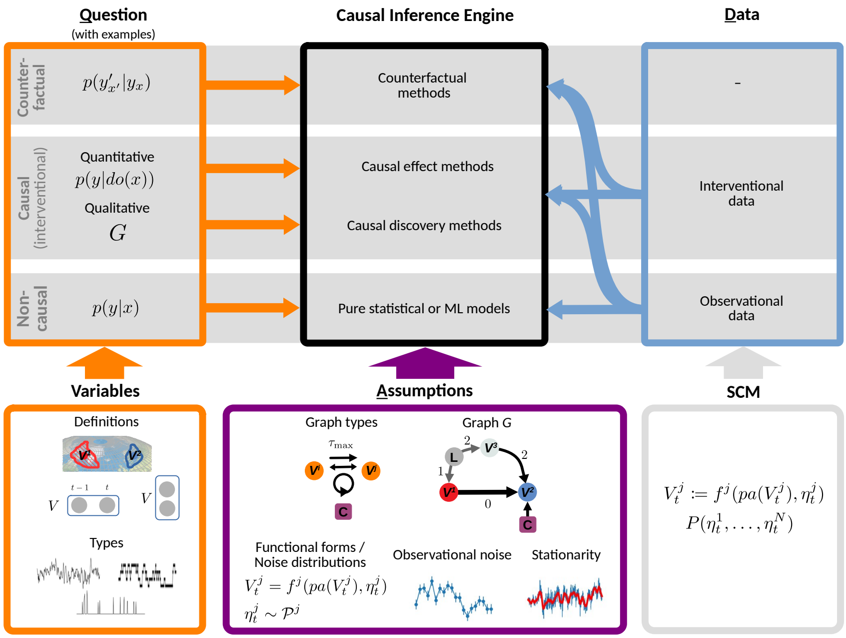
Induced distributions. Due to acyclicity, the SCM variables V^j_t can be re-expressed solely in terms of the noises $\eta^i_{t'}$. Assuming the dynamics described by eq. (1) to be stationary, the noise distributions thus uniquely induce a distribution $p(\cdot)$ over the V^j_t. This distribution $p(\cdot)$, referred to as observational distribution or pre-interventional distribution, generates the observed data (see the scatter plot in black and the associated conditional distributions in blue in Fig. 1**a**).

As an intervention $do(\mathbf{X} = \mathbf{x})$ alters the structural assignments, an intervention also changes the distribution, thereby giving rise to the (post-)interventional distribution denoted $p(\cdot \mid do(\mathbf{X} = \mathbf{x}))$ [33] (in red in Fig. 1**b**). Following the idea that correlation does not imply causation, $p(\cdot \mid do(\mathbf{X} = \mathbf{x}))$ is not, in general, equal to $p(\cdot \mid \mathbf{X} = \mathbf{x})$ (see the vertical cross-sections of the red and blue in distributions in panels **b** and **a** of Fig. 1, respectively).

Next to observational and interventional distributions, one can also consider counterfactual distributions. One example of a counterfactual distribution is $p(y'_{x'} \mid y_x)$ which specifies the probability of observing $Y = y' \neq y$ under the hypothetical past intervention $do(X = x')$ when, in fact, $Y = y$ was observed under the intervention $do(X = x)$.

In Pearl's Three Layer Causal Hierarchy [30, 47] of association (layer 1), intervention (layer 2), and counterfactuals (layer 3), the observational distributions belongs to layer 1, the interventional distribution belongs to layer 2 and the counterfactual distributions belongs to layer 3. In causal inference, the fundamental task is to investigate questions and quantities concerning interventions or counterfactuals with data from distributions at the same or lower layers instead.

Box 1: Question-Assumptions-Data template.



The QAD-template, based on Pearl's Causal Hierarchy [30] and the Causal Inference Engine [48], guides scientists to phrase and solve their problem in the framework of causal inference. Each SCM (grey box in the lower right) induces a hierarchy of observational, interventional, and counterfactual distributions and their associated datasets (blue box with gray layers). Given causal variables and a question (orange boxes) about the distribution of a target SCM, the causal inference engine (black box) utilizes assumptions (purple box) and data from this SCM (and potentially others, blue box) to either return that the question cannot be answered or an estimand, which is an estimation recipe based on an appropriate method.

Variables. The first step is to specify the causal variables of the SCM, which can be defined based on spatial patterns, temporal aggregation, or grouped variables, and whose type can be continuous, discrete or categorical, and event-like, to name a few. The QAD-template then proceeds from questions, assumptions, and available data through the Causal Inference Engine as follows.

Questions. Questions can come from different layers of the causal hierarchy (bottom to top). Non-causal questions are about the observational distribution. Causal questions are about the interventional distribution and can be subdivided into qualitative questions about links or ancestral relations in the causal graph G and quantitative questions about the distributions of Y for different interventional values $X = x$. Counterfactual questions are, for example, about the probability of observing $Y = y' \neq y$ under the hypothetical past intervention $do(X = x')$ when, in fact, $Y = y$ was observed under the intervention $do(X = x)$.

Assumptions. Methods to answer these questions can utilize a plethora of different assumptions. Structural assumptions (top panels) can be about graph types for causal discovery methods [10, 31, 32], including whether the graphs can permit only (uni-)directed links, or also latent confounding represented by bi-directed arrows and contemporaneous feedback cycles represented by circle arrows, up to some maximal time lag $\tau_{\max}$, and additional assumptions encoded in context nodes C [48, 61, 62], such as whether the data is interventional. Structural assumptions can also be about fully specified causal graphs for causal effect estimation methods [9] and can include hidden variables, potential context nodes, and time lags. Parametric assumptions about the SCM can also be made (bottom panels), such as the functional form of the dependency of a variable V_t^j on its parents $pa(V_t^j)$ and noise η_t^j as well as the type of noise distribution, or about observational noise characteristics and smoothness of the influence of a non-stationary confounder (red).

Data. Finally, the causal inference engine requires (potentially multiple) datasets from observational and/or interventional distributions of the target SCM, or also other SCMs if context knowledge about their differences exists [48]. Counterfactual data is generally not available if the SCM is not at hand.

Causal Inference Engine. Given this QAD-input the causal inference engine either returns “Not identifiable” or a (causal) method with a recipe to compute the answer to the question from data. For non-causal questions, pure statistical or ML models or association measures would be suggested, for qualitative causal questions causal discovery methods returning a (partially or fully specified) causal graph $\widehat{G}$, for quantitative causal questions this would be an estimand of an interventional question, and for counterfactual questions there exist several estimation strategies [9, 59].

Framing causal research questions

This section introduces a taxonomy of research questions, types of expert assumptions, and properties of the available time series data to provide a causal language in which researchers can phrase their questions. This Review uses the term QAD to represent the main aspects of translating research questions into actionable causal inference tasks: Questions (Q) - Assumptions (A) - Data (D) (Box 1). The QAD-template is embedded in Pearl's Causal Hierarchy [30] and the Causal Inference Engine [48] that returns whether or not a QAD-problem can be answered and, if so, provides a suitable solution. The following sections are accompanied by a questionnaire (Table 1).

Defining the variables of interest. Causal inference starts with defining the variables of interest, which will become the nodes of the graph (see illustrations in Box 1 "Variables"). This step is critical and requires a concrete research question, as causal inference analyses are typically very sensitive to different variable definitions. A variable can be constructed using dimension-reduction [49] or spatio-temporal aggregation, for instance, by taking monthly or daily averages of a time series over some region. Causal inference analyses are sensitive to these aspects because, for example, a too-coarse temporal aggregation can introduce cycles or confounding. Variables can also be vector-valued to describe the phase and amplitude of oscillatory phenomena or sets of physical variables, for example.

Variable types can be continuously valued (such as physical variables like temperature), discretely valued (for example, above or below a threshold), or purely categorical (like weather regimes). Further, one can also represent variables continuous in time with specific events (like extreme events) indicated as point processes [50]. The variables' types affect the choice of statistical models (such as using linear versus logistic regression) in a causal inference analysis.

A separate topic in causal machine learning is that of causal representation learning [51, 52], which attempts to simultaneously learn the variables of interest and their causal relations from raw, for example, spatiotemporally gridded, data.

The following sections introduce the different types of questions a researcher might ask and associated assumptions, the "Q" and "A" in the QAD-template, together with solutions returned by the causal inference engine, ordered from layer 1 to 3 in the causal hierarchy (bottom to top grey layers in Box 1).

Non-causal questions. Examples of such questions include characterizing precipitation patterns or (time-lagged) statistical associations between two or more variables, including association networks [53, 54] for prediction or forecasting [55]. However, such a forecast is valid only in the limited case that the training and target distribution are the same, for example, in short-term weather forecasts. To make such estimates consistent, assumptions have to be made about the general type of pairwise dependencies—for example, linear

Table 1 Questionnaire for QAD-template.

Aspect	**Examples**
Variables (=nodes): Causal variables and their characteristics	
Dimension	Scalar (potentially aggregated by averaging) or vector-valued (spatial region, time-embedded, multiple variables)
Time resolution	Such as days, months, years, or also at irregular time points
Type of variable	Such as continuously-valued, categorical, ordinal, event-like
Q: Causal question among variables	
Non-causal	Such as purely statistical associations or in-distribution prediction
Causal discovery	Reconstructing the full causal graph or just pairwise causation
Causal effects	Estimating causal effects (of hypothetical interventions) in terms of expectations or full interventional distributions
Counterfactuals	Such as attribution of observed events or mediation analysis
A1: Structural assumptions on the graph-level	
Graph type	Graphs allowing for only unidirected links, or also latent confounding and contemporaneous feedback cycles; maximal time lag $\tau_{\max}$; (different types of) faithfulness
Graph	Partial or full knowledge of graph, including adjacencies and orientations, confounding, time lags, and cyclicity
Contexts	Include further nodes to represent interventional data, missing samples, non-stationary or, in the case of multiple datasets, dataset-specific static causes such as the elevation of a site
A2: Parametric assumptions	
Functional forms	Linear or different types of nonlinearity such as additive models
Distributions	Such as Gaussian, uniform, laplacian, or extremal noise distributions
Obs. errors	Distributions of measurement errors, outliers
Stationarity	Over which part of the dataset (or multiple datasets) stationarity holds (potentially after pre-processing) or smoothness and periodicity of latent confounding
D: Data characteristics	
Dataset type	For example, a single dataset of multivariate time series, or multiple datasets such as at sites or ensemble members of climate models
Interventional	For example, through climate model experiments, real experiments, or pseudo-experiments such as volcanic eruptions
Dimensions	Number of variables and time lags, sample size
Missing samples	For example, due to device failures
Potential pre-processing and further analysis choices	
Detrending	Subtraction of linear or nonlinear trends
Masking	For example, to constrain the analysis to summer/winter or day/night
Transformations	For example, to normal marginals to better match statistical assumptions
Denoising	If knowledge of characteristics of observational noise exists
Sliding windows	To account for slowly varying hidden confounders or to assess robustness

or nonlinear. Nevertheless, causal information can help make models more parsimonious and interpretable [10, 24].

Causal discovery. Causal discovery, or causal structure learning [31, 32], qualitatively reconstructs the links in either the complete causal graph or just the causes of a particular target variable, potentially including time lags and/or bi-directed links for representing latent confounding. Causal discovery methods make use of many assumptions as discussed below in the dedicated section (see also ref. [10]), and ideally provide so-called identifiability theorems [32] that guarantee that the method will result in the correct graph under these assumptions, up to errors due to a finite sample size.

Causal effects. The goal of causal effect estimation [33] is to quantify the total causal effect of one variable or set of variables on another (Fig. 1**b**). These questions require assumed or learned qualitative knowledge in the form of a causal (time series) graph that could include hidden confounding. For example, temperature is known to influence ecosystem respiration but causal effect estimation could be used to quantify by how much, when given a graph of other observed and unobserved confounding variables. The graph encodes assumptions about the absence and presence of causal relations. To quantify a causal effect, the distribution of Y given an intervention in X, $p(y \mid do(\mathbf{X} = \mathbf{x}))$, is expressed as a function of the observational distribution. If such a re-expression is possible, then the causal effect is identifiable and a causal estimand can be obtained, which is an expression involving only the observational distribution. An estimate then involves further parametric assumptions.

Counterfactual questions. Counterfactual questions ask what would have happened to Y if X had been set to a different value [33, 56]. An example of a counterfactual distribution is $p(y'_{x'} \mid y_x)$, which specifies the probability of observing $Y = y' \neq y$ under the hypothetical past intervention $do(X = x')$ when $Y = y$ was actually observed under the intervention $do(X = x)$. In other words, counterfactual questions use present observations, interventions, and outcomes to reason about past alternative interventions. Examples of counterfactual questions in climate research include the causal attribution of extreme events, like determining the probability of an individual extreme event if no anthropogenic climate forcing occurred in the past [57].

Counterfactuals also occur for simpler questions of causal mediation, where X has both direct and indirect effects on Y. For example, in Fig. 1**a**, V^2_{t-2} affects V^3_t directly and indirectly through the mediators $\mathbf{M} = \{V^1_{t-1}, V^1_t, V^2_{t-1}, V^2_t, V^3_{t-1}\}$. Here, disentangling the direct from the indirect effect amounts to determining the effect of V^2_{t-2} on V^3_t in the counterfactual scenario where $\mathbf{M}$ (or a subset thereof) are held fixed [58]. Counterfactual questions generally require more assumptions about the underlying SCM than interventional questions, but even these can sometimes be answered from purely observational data [59]. In this article, we only consider linear mediation for graphs without unobserved confounding and do not further consider counterfactuals in general [33, 56].

Dataset properties. The availability and properties of the datasets affect the methodological choices to implement causal inference analyses. In investigating the causal relationships underlying global teleconnections from observational data, for example, there is only one natural climate system, so in principle there is only a single multivariate time series dataset. Depending on the dataset dimensions (time series length and the number of variables), different methodological choices can be made to balance model complexity, computational cost, and statistical performance. If available, these teleconnections can also be analyzed in historical simulations of climate model ensembles [2], which provide multiple datasets all with the same physical model, but different initial conditions. Similarly, multiple datasets can emerge from the same set of variables at different measurement locations. To estimate a single causal graph from this information, the data must be ensured to come from the same distribution.

However, causal inference can also jointly utilize multiple datasets from different distributions, for example, in fixed-effects panel regression [60], causal effect transportability [48], or (joint) causal inference [61, 62]. If simulations from different climate models or the physical processes in different locations have slightly different underlying distributions and associated causal graphs (such as under different interventions or SCM changes over time), then they can be analyzed separately. Alternatively, if the differences between the causal relationships in the climate models or local physical processes can be explained by so-called context variables (such as climate sensitivity to greenhouse gases or the elevation of the measurement stations), then causal inference can leverage these by including such dataset-specific variables in the analysis. In the Assumptions panel of Box 5, these dataset-specific variables are denoted as C. Dataset- and time-dummy variables can also be added to allow for context-related deconfounding as typically done in econometrics [60].

Another type of data are interventional datasets. These can emerge from targeted experiments with climate models, pseudo-real experiments such as volcanic eruptions, or actual small-scale real experiments. Context nodes can be used to represent interventions.

Real data is often plagued with missing samples, due to problems such as sensor malfunction or the presence of clouds in satellite remote sensing images. If other variables in the graph are responsible for the missing samples, there can be selection biases that can partially be treated in a causal framework [63].

Pre-processing and analysis choices. Pre-processing can often help a dataset meet the parametric and causal assumptions required for different causal analyses. Machine learning models often require pre-processing steps such as standardization or normalization to bring different variables into comparable ranges. Statistical models can be partially adapted to observational errors using assumptions about the error distributions. Removing a trend or seasonal cycle caused by a exogenous latent variable that is not affected by any other variable in the analysis might be necessary and in some cases sufficient to use causal discovery methods that do not allow for latent confounding. Finally, non-stationarity due to regime-like or slowly-varying latent confounders (illustrations in the Assumptions panel of Box 1) can be overcome by restricting the time series via masking or conducting a sliding window analysis. If such pre-processing does not work, frameworks for non-stationary time series can help [62, 64].

An analysis that first performs causal discovery and then causal effect quantification can be considered post-selection inference [65] and can lead to overly optimistic (that is, too narrow) confidence intervals in the effect estimation step. Post-selection inference refers to performing statistical inference (hypothesis testing, estimation of confidence intervals or estimating causal effects) for a model that has been chosen with a data-driven selection procedure as opposed to an a priori chosen model as in the classical statistical theory. Some advances have been proposed to solve the post-selection inference problem [65, 66]. A simple but usually statistically inefficient solution is sample-splitting: two independent subsets of the data are used to perform causal discovery (or model selection) and causal inference.

Causal discovery

The goal of causal discovery is to infer the causal (time series) graph, or at least partial information about it, between a set of variables from samples of the observational distribution $p(\cdot)$ and/or from interventional data by utilizing general assumptions about the graph type and further parametric assumptions (Box 5)[10, 31, 32, 67]. Such qualitative causal knowledge can be used in a subsequent analysis to evaluate causal questions quantitatively or might be the object of interest itself.

The following sections briefly describe the three prevalent general approaches to causal discovery, focusing on the case of only observational data. Runge et al. [10] cover further concepts such as invariant causal prediction [68–70], which utilizes assumptions about multiple datasets coming from different contexts, information-geometric causal inference [71], which assumes deterministic relations, or convergent cross mapping [11, 72, 73], a state-space method that assumes an underlying and reconstructable deterministic nonlinear dynamical system.

Constraint-based approach. The causal inference framework considered here assumes that all data is generated by an SCM. Then the connectivity pattern of the causal graph imprints corresponding marginal and conditional (in-)dependencies into the observational distribution [74, 75]. This property is known as the causal Markov condition [31]. The basic idea of constraint-based causal discovery is then to do reverse engineering, that is, to perform a sequence of statistical tests of independence:

$$H_0: \; X \perp\!\!\!\perp Y \mid \mathbf{Z} \qquad \text{vs.} \qquad H_1: \; X \not\perp\!\!\!\perp Y \mid \mathbf{Z}\,, \tag{2}$$

where $\perp\!\!\!\perp$ denotes independence, for all pairs of variables (X, Y) and appropriately chosen sets of conditioning variables $\mathbf{Z}$. Then, based on the results of these tests, the structure of the causal graph is constrained.

The key underlying assumptions utilized here are the causal Markov condition and causal faithfulness [31] that allow to relate d-separations [76] in the graph to conditional independence in the distribution: X d-sep $Y \mid \mathbf{Z} \iff X \perp\!\!\!\perp Y \mid \mathbf{Z}$, where the causal Markov condition asserts "$\Rightarrow$" and the causal faithfulness assumption "$\Leftarrow$". Two nodes are d-separated [76] given $\mathbf{Z}$ if and only if all paths are blocked given $\mathbf{Z}$. Generally, given a conditioning set $\mathbf{Z}$, a path between two nodes is blocked if and only if it contains at least one non-collider ($\rightarrow\bullet\rightarrow$ or $\leftarrow\bullet\rightarrow$) in $\mathbf{Z}$ or at least one collider ($\rightarrow\bullet\leftarrow$) that is neither in $\mathbf{Z}$ nor in the ancestors of $\mathbf{Z}$ (Box 2). The causal faithfulness assumption excludes, for example, the non-generic situation in which X influences Y via two pathways such that in total the effect cancels out exactly, among other cases [77].

For computational feasibility, most methods choose the conditioning sets $\mathbf{Z}$ in eq. (2) in an adaptive, non-brute force way. A notable example is the PC algorithm [31, 78], which, in addition to an underlying SCM and causal faithfulness, assumes acyclicity and the absence of unobserved variables. The latter assumption is also known as causal sufficiency and is dropped by the FCI algorithm [31, 79–81]. Some methods do not require the acyclicity assumption [42, 44]. PCMCI [21] and its extensions PCMCI$^+$ [82] and LPCMCI [83] are time-series adaptions of PC and FCI that are specifically designed to address the challenges of time series.

The constraint-based approach is modular in the sense that it can be combined with any method for conditional independence testing—for example, a test based on partial correlation for linear Gaussian data or on conditional mutual information [84] for completely general functional relationships—and in this sense it is non-parametric.

With constraint-based causal discovery, it is generally not possible to discover the unique causal graph, but instead only a set of possible graphs that includes the true one in the large-sample limit. This lack of learning a unique causal graph is because different graphs can give rise to the same set of d-separations, thereby defining a Markov equivalence class of graphs (Box 2).

Constraint-based methods have been applied in numerous Earth sciences problems, such as the reconstruction of atmospheric pathways [16–18], climate model evaluation [27], or teleconnection analysis [19]. The prediction-based framework of Granger causality [85–89] and information-theoretic extensions [90–92] also essentially utilize independence relations and have partially been used also in the Earth sciences [93–96].

Asymmetry-based approach. Asymmetry-based methods attempt to solve the fundamental bivariate case, that is, to learn the causal relationships between a pair of variables X and Y. As the causal graphs $X \rightarrow Y$ and $X \leftarrow Y$ belong to the same Markov equivalence class (see the link $V^3_t \circ\!\!-\!\!\circ V^4_t$ in

Box 5**a**), distinguishing between these two options requires stronger assumptions. These stronger assumptions restrict the functional relationships or noise distributions in the data-generating SCM (Assumptions panel in Box 5).

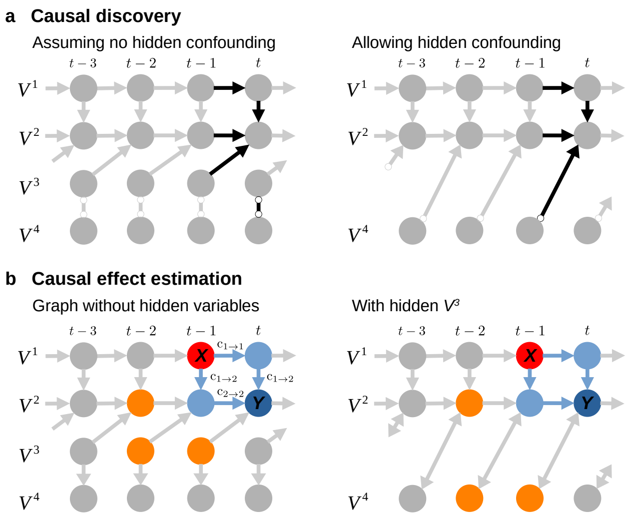
The prototypical example is the LiNGAM (Linear Non-Gaussian Acyclic Model), or VARLiNGAM for time series, which assumes that $Y \coloneqq a \cdot X + \eta^Y$ with $X \perp\!\!\!\perp \eta^Y$ and non-Gaussian noise term(s) is the true SCM. It is then not possible to find a backward model within the same restricted functional model class [97–99]. That is, it is not possible to find a coefficient b and a random variable η^X with $Y \perp\!\!\!\perp \eta^X$ such that $X \coloneqq b \cdot Y + \eta^X$. Hence, the causal $(X \to Y)$ and anti-causal $(X \leftarrow Y)$ directions can be distinguished. The rationale behind this approach is the expectation that the causal model generically is simpler than the anti-causal model [32].

Other examples of identifiable functional model classes are the nonlinear additive noise model [100, 101], the post-nonlinear model [102], and heavy-tailed extremal models [103]. Methods for bivariate causal discovery can be lifted to methods for causal discovery among multiple variables [104]. Asymmetry-based methods have been used in remote sensing and geoscience problems to evaluate emulators and identify appropriate vegetation models [105, 106].

Score-based approach. Score-based methods define a scoring function S, which assigns a score $S(\mathcal{D}, G)$ to the given dataset $\mathcal{D}$ and each possible causal graph G, and then picks the best scoring graph [32]. A typical scoring function is the complexity-penalized log-likelihood for an assumed parametric statistical model. As the number of DAGs grows super-exponentially with the number of variables, the search for the best scoring graph is often carried out greedily. If the assumed statistical model does not yield identifiability, then the search can be reorganized as a search over Markov equivalence classes. A notable representative of this class of methods is the GES (Greedy Equivalence Search) algorithm [107–109], and there are also highly efficient hybrid methods [110]. There are some applications of score-based approaches in the Earth sciences [111, 112].

Score-based causal discovery has also been framed as a continuous optimization problem [113–115] that avoids the infeasible search through a discrete number of graphs, but requires additional assumptions such as availability of interventional data or equal-variance noise terms. These methods are still in their infancy for time series (for example, DYNOTEARS [116]).

Box 2: Causal discovery and causal effect estimation



A constraint-based causal discovery algorithm that assumes no hidden confounding, such as PCMCI$^+$ [82], can be applied to observational data of an SCM with structure as shown in panel (**b**, left) and would here result in the graph shown in the left panel of the figure part **a**. Here V_t^1 and V_{t-1}^3 are d-separated given an empty conditioning set because the only path between the two, $V_t^1 \to V_t^2 \leftarrow V_{t-1}^3$, contains the collider V_t^2. V_{t-1}^1 and V_t^2 are d-connected because at least one path between the two, such as $V_{t-1}^1 \to V_t^1 \to V_t^2$ or $V_{t-1}^1 \to V_{t-1}^2 \to V_t^2$, contains only non-colliders. PCMCI$^+$ would detect, among others, the independence $V_t^1 \perp\!\!\!\perp V_{t-1}^3$ and utilize the Markov, faithfulness, and sufficiency assumptions to conclude that, firstly, there is no link between the two, and, secondly, that $V_t^1 \to V_t^2 \leftarrow V_{t-1}^3$. The other potential motifs either contradict the order of time ($V_t^1 \to V_t^2 \to V_{t-1}^3$ and $V_t^1 \leftarrow V_t^2 \to V_{t-1}^3$) or would imply that V_t^1 and V_{t-1}^3 are dependent ($V_t^1 \leftarrow V_t^2 \leftarrow V_{t-1}^3$) contrary to their observed independence. No such conclusion can be drawn for the link $V_t^3 \circ\!\!-\!\!\circ V_t^4$ and both possible orientations yield the two graphs in the Markov equivalence class as the output of PCMCI$^+$.

The result of a causal discovery algorithm that allows for arbitrary hidden confounders, such as LPCMCI [83], can be used on a dataset where V^3 is hidden (right panel of the figure part **a**). Then $V_t^1 \to V_t^2$ can still be inferred and is interpreted as directed causation (potentially through hidden variables). At the same time, $V_{t-1}^4 \circ\!\!\rightarrow V_t^2$ denotes the presence of either $V^4_{t-1} \to V^2_t$ or hidden confounding $V^4_{t-1} \leftrightarrow V^2_t$ as the two orientations of the Markov equivalence class here.

Causal effect estimation (as in panel **b**) generally requires a fully oriented graph and aims to quantify causal relations. In the directed acyclic time series graph without hidden confounding (left panel), the causal effect of V^1_{t-1} (red) on V^2_t (blue) can, in the general non-parametric case, be estimated by adjustment sets that block all non-causal paths. Here $\mathbf{Z} = \{V^1_{t-2}\}$ would be a valid set, but the statistically optimal **O**-set (orange nodes) is given by the parents of both the effect variable (blue) and the mediators (light blue). Under the assumption of a linear model, the path method allows decomposing this causal effect as the sum-product of the link coefficients $c_{1\to1}c_{1\to2}+c_{1\to2}c_{2\to2}$, which can each be estimated by linear regression of every node on its parents. If V^3 is unobserved (right), an ADMG is obtained through a latent projection operation and the path method is no longer applicable. Adjustment is still possible, though, and in this case the **O**-set comprises not only the parents, but also further nodes [130].

Causal effect estimation

This section explains several concepts and methods for estimating causal effects from observational data, that is, concepts and methods for answering quantitative causal questions. The discussion is restricted to the setting where the qualitative causal relationships in the form of the causal graph are given, either by domain knowledge or from a prior causal discovery analysis.

Problem formulation. A main goal of causal effect estimation is to estimate the total causal effect (often just referred to as causal effect) of a set of variables $\mathbf{X} = \{X_1, \ldots, X_{N_X}\}$ on another variable Y. Causal effects can be defined as the quantity [33]

$$\Delta_{\mathbf{X}\to Y}(\mathbf{x}', \mathbf{x}) \;=\; \mathbb{E}\left[Y \mid do(\mathbf{X}=\mathbf{x}')\right] - \mathbb{E}\left[Y \mid do(\mathbf{X}=\mathbf{x})\right] , \tag{3}$$

that is, as the difference in the expected value of Y when setting $\mathbf{X}$ by intervention to $\mathbf{x}'$ as opposed to $\mathbf{x}$. Other definitions are possible, for example, as the post-interventional mean $\mathbb{E}\left[Y \mid do(\mathbf{X}=\mathbf{x})\right]$ or the post-interventional distribution $p(y \mid do(\mathbf{X} = \mathbf{x}))$ for a whole range of intervention values (Fig. 1**b**). Contrary to a multivariate cause-variable $\mathbf{X}$, for non-singleton effect-variables $\mathbf{Y} = \{Y_1, \ldots, Y_{N_Y}\}$ the individual components $\Delta_{\mathbf{X}\to Y_k}(\mathbf{x}', \mathbf{x})$ decouple entirely and can hence be considered in isolation.

The choice of values $\mathbf{x}$ and $\mathbf{x}'$ is up to the user and part of specifying the question to be answered. For example, for $\mathbf{X} = \{X\}$ denoting sea-surface temperatures in the tropical Pacific, $x = \mathbb{E}[X]$ and $x' = \mathbb{E}[X] + \sqrt{Var(X)}$

need to be chosen to answer by how much the mean of atmospheric pressure Y changes when X is by intervention set to one standard deviation above the mean of X as opposed to the mean of X. If the components of the difference vector $\Delta\mathbf{x} = \mathbf{x}' - \mathbf{x}$ associated to a subset $\tilde{\mathbf{X}} \subset \mathbf{X}$ are equal to zero, then $\Delta_{\mathbf{X}\to Y}(\mathbf{x}', \mathbf{x})$ is also referred to as the controlled direct effect of $\mathbf{X} \setminus \tilde{\mathbf{X}}$ on Y relative to $\tilde{\mathbf{X}}$ [117]. Without a parametric assumption $\Delta_{\mathbf{X}\to Y}(\mathbf{x}', \mathbf{x})$ can be a general non-linear function.

The causal effect $\Delta_{\mathbf{X}\to Y}(\mathbf{x}', \mathbf{x})$ is defined in terms of $p(y \mid do(\mathbf{X} = \mathbf{x}))$. This distribution is, in general, not equal to $p(y \mid \mathbf{X} = \mathbf{x})$ due to the potential existence of confounders that influence both $\mathbf{X}$ and Y. Confounders can introduce a non-causal association between $\mathbf{X}$ and Y. For example, the distributions of aerosols and clouds are both affected by various atmospheric factors. Randomized experiments offer direct access to $p(y \mid do(\mathbf{X} = \mathbf{x}))$ by cutting off the dependence of $\mathbf{X}$ on the causes of $\mathbf{X}$ and hence eliminating the unwanted non-causal associations [33, 36, 118]. In causal inference, however, the goal is to estimate the causal effect from observational data by leveraging appropriate assumptions about the causal graph and, for some methods, parametric assumptions on the structural assignments (Assumptions in Box 5). For example, a graph with confounders of aerosols and clouds can be qualitatively specified [119].

Using the do-calculus [120–123], it is possible to determine whether a causal effect is in principle identifiable from observational data or not. There are cases in which the causal effect is fundamentally unidentifiable without stronger assumptions, for example, in the graph in the Assumptions panel of Box 5 when both L and V^3 are unobserved.

Covariate adjustment. The most well-known method for causal effect estimation from observational data without parametric assumptions is covariate adjustment [33]. Covariate adjustment refers to de-confounding the causal relationship by adjusting for (controlling for) a set of variables $\mathbf{Z}$. In the general case, the adjustment formula reads

$$p(y \mid do(\mathbf{X} = \mathbf{x})) = \int p(y \mid \mathbf{x}, \mathbf{z}) p(\mathbf{z}) d\mathbf{z}\,. \tag{4}$$

For the post-interventional means considered in eq. (3) for each fixed value $\mathbf{z}$ of $\mathbf{Z}$ the $\mathbf{z}$-specific causal association $\mathbb{E}\left[Y \mid do(\mathbf{X} = \mathbf{x}), \mathbf{Z} = \mathbf{z}\right]$ equals the $\mathbf{z}$-specific observed association $\mathbb{E}\left[Y \mid \mathbf{X} = \mathbf{x}, \mathbf{Z} = \mathbf{z}\right]$. Aggregating these values according to the probability of observing $\mathbf{Z} = \mathbf{z}$ then gives

$$\mathbb{E}\left[Y \mid do(\mathbf{X} = \mathbf{x})\right] = \mathbb{E}_{\mathbf{z}}\left[\mathbb{E}\left[Y \mid \mathbf{X} = \mathbf{x}, \mathbf{Z} = \mathbf{z}\right]\right]\,. \tag{5}$$

A set $\mathbf{Z}$ as in eq. (4) is known as an adjustment set [124], and inserting eq. (5) into eq. (3) yields the causal estimand. To use an adjustment set, all its elements must be observed.

Because the right-hand-side of eq. (5) does not contain any do-expression, it can be estimated from observational data $\mathcal{D} = \{(\mathbf{x}_i, y_i, \mathbf{z}_i)\}_{i=1}^n$. This estimation proceeds as follows: first, regress y_i on $(\mathbf{x}_i, \mathbf{z}_i)$ to get estimates $\hat{y}(\mathbf{x}, \mathbf{z})$ of the inner expectation and, second, take the sample average of $\hat{y}(\mathbf{x}, \mathbf{z} = \mathbf{z}_i)$ to approximate the outer expectation. If $\mathbf{Z}$ is an adjustment set, then eq. (5) holds independent of any parametric assumption and can be implemented with a powerful ML model able to learn complex functional (causal) relationships [3–6]. The choice of a ML model depends on the assumed (non-)parametric relations relevant for the causal effect under study, partially linear models can make use of double machine learning approaches [125, 126]. To determine which sets of variables $\mathbf{Z}$ are adjustment sets, the causal graph is used. The qualitative cause-and-effect relationships conveyed by the causal graph imply which variables must be adjusted for in order to disentangle the causal and non-causal associations. For example, assuming the graph in the Assumptions box (Box 1) to hold for the causal effect of aerosols ($X = V^1$) on clouds ($Y = V^2$) including unobserved confounders (L), but with relative humidity (V^3) being observed, then would use $\mathbf{Z} = \{V^3\}$ as an adjustment set. Several graphical conditions exist for reading off adjustment sets from more complex causal graphs. The following sections discuss a general adjustment criterion and a specific adjustment set with favorable statistical properties.

Adjustment criterion. The adjustment criterion derived in ref. [127] allows reading of all adjustment sets from a given causal graph, even if there are unobserved variables. This criterion states that $\mathbf{Z}$ is an adjustment set if and only if $\mathbf{Z}$ satisfies two conditions. First, $\mathbf{Z}$ must not contain any variable in $med \cup de(med)$, where the mediators med is the set of variables, excluding $\mathbf{X}$, on all such directed paths from $X_k \in \mathbf{X}$ to Y that do not intersect $\mathbf{X} \setminus \{X_k\}$ ((light) blue in Box 5**b**) and $de(med)$ denotes the descendants of med in the causal graph. If this first condition were not met, then a part of the causal effect of interest would be blocked off or collider-bias would be introduced [33]. Second, $\mathbf{Z}$ must block all proper non-causal paths from $\mathbf{X}$ to Y. If this second condition were not met, then a non-causal association would be contained in the conditional expectation $\mathbb{E}[Y \mid \mathbf{X} = \mathbf{x}, \mathbf{Z} = \mathbf{z}]$.

Optimal adjustment. Often, there will be more than one adjustment set. Then the question of which of the multiple adjustment sets should be used arises. Under the assumptions of linearity and no unobserved variables, then for a given causal graph, there is a unique adjustment set that (for all possible link coefficients) yields causal effect estimates with the smallest asymptotic variance compared to all other adjustment sets [128]. This so-called optimal set is $\mathbf{O} = pa(med) \setminus med$ with med as in the previous paragraph and $pa(med)$ the parents of med in the causal graph. For certain classes of nonlinear models [125, 126] this property was also shown to hold [129] and in ref. [130] a generalization of the optimal set to causal graphs with unobserved variables (ADMGs) is given (Box 2).

There are cases in which the adjustment method does not allow for identification, but the front-door criterion [120] or, more generally, do-calculus [120] can be used to obtain an estimand, if it is identifiable at all [123] without further assumptions, such as in the instrumental variable setting [131]. A further obstacle to identifiability that can partially be overcome is selection bias [132–135] or missing data [63, 136, 137].

Linear causal effect estimation and the path method. For linear models $\Delta_{\mathbf{X}\to Y}(\mathbf{x}', \mathbf{x})$ in eq. (3) reduces to

$$\Delta_{\mathbf{X}\to Y}(\mathbf{x}', \mathbf{x}) \;=\; \Delta\mathbf{x} \cdot \vec{\delta}\,, \tag{6}$$

where $\vec{\delta} = (\delta_1, \ldots, \delta_{N_X})$ and δ_k is the controlled direct effect of $X_k \in \mathbf{X}$ on Y relative to $\mathbf{X} \setminus \{X_k\}$ and can be estimated as the regression parameter of $X_k \in \mathbf{X}$ in the linear regression of Y on $\mathbf{X} \cup \mathbf{Z}$.

For linear models and graphs without hidden variables, the path method [138, 139] has even lower estimation variance than optimal adjustment [140]. If the causal mechanisms f^j in eq. (1) are linear functions, each edge $V^i_{t-\tau} \to V^j_t$ is parameterized by a single number. This number, referred to as edge weight or link coefficient $c_{\cdot\to\cdot}$, is the coefficient that multiplies $V^i_{t-\tau}$ in the structural assignment of V^j_t. The product of all edge weights along a path assigns a weight to a path, and the controlled effect δ_k in eq. (6) equals the sum of path weights taken over all proper causal paths.

This sum-product path-tracing rule (Box 5**b**) yields estimands for the δ_k and hence, by means of eq. (6), an estimand for the causal effect $\Delta_{\mathbf{X}\to Y}(\mathbf{x}', \mathbf{x})$. In the path method, the weights for $V^i_{t-\tau} \to V^j_t$ are estimated as the regression coefficient of V^j_t in the linear regression of V^j_t on the parents of V^j_t (which include $V^i_{t-\tau}$). For the path method to apply, all variables on the relevant paths and all parents of all these variables must be observed.

Generally, causal effect estimates are in units of the respective variables. Standardizing the time series beforehand leads to an interpretation in terms of standard deviations. As a real data example [20], if a one-standard deviation (std) increase in East Pacific sea-surface temperatures (ENSO) leads to a 0.37-std change in the jet stream position and that in turn causes a 0.81-std change in California winter precipitation, then the (indirect) effect of ENSO on California precipitation is $0.37 \times 0.81 \approx 0.3$.

Linear mediation analysis. Apart from quantifying the total causal effect of $\mathbf{X}$ on Y, causal pathways (the mechanisms by which this effect is transmitted) might also be analyzed. The analysis of causal pathways leads to questions such as how large is the part of the causal effect of $\mathbf{X}$ on Y that passes (or does not pass) through $\mathbf{M}$, where $\mathbf{M} = \{M_1, \ldots, M_m\}$ is a given set of mediators. The formalization of these questions within the structural causal model framework is in terms of counterfactual quantities [58, 117, 141]. To avoid the subtleties and controversies associated with counterfactuals, this Review restricts the discussion of mediation analysis to linear causal models.

For linear models, the following simplifications occur [58, 117, 142]. The effect that passes through $\mathbf{M}$ is the sum-product of edge weights on proper causal paths from $\mathbf{X}$ to Y that pass through $\mathbf{M}$. For example, in Box 5**b** the effect of V^1_{t-1} on V^2_t through $\mathbf{M} = V^1_t$ is just $c_{1\to 1}c_{1\to 2}$. Similarly, the effect that does not pass through $\mathbf{M}$ can be estimated.

Example case studies

The following two case studies from climate science and biogeosciences follow the QAD-questionnaire (Tab. 1) and method-selection flow chart (Fig. 2).

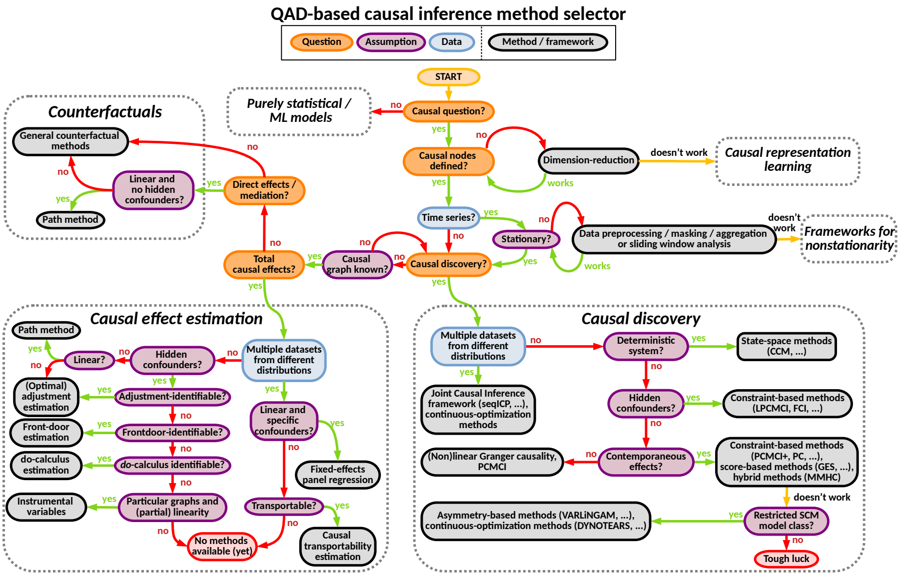


Fig. 2 **QAD-based causal inference method selection flow chart.** General guidance for using the QAD-approach. The individual methods and their hyperparameter choices typically require further assumptions. For example, the PC algorithm and PCMCI$^+$ make the Markov and Faithfulness assumptions and the choice of a suitable conditional independence test needs to be adapted to the data types, distributions, and dependencies. Data characteristics such as the number of variables and sample size are also relevant for choosing among methods.

Climate example. The physical mechanisms of the Walker Circulation [143, 144] are reasonably well understood in terms of climate teleconnections [145]. Such teleconnections are typically analyzed using regression and correlation-based techniques, but additional insights can be gained from employing causal inference [20].

Here, causal inference is used to quantify the causal effects among the processes involved in the equatorial Western-Central Pacific branch of the Walker Circulation. This branch of the circulation describes a clock-wise large-scale time-mean circulation of falling air masses in the Central Pacific (CPAC),

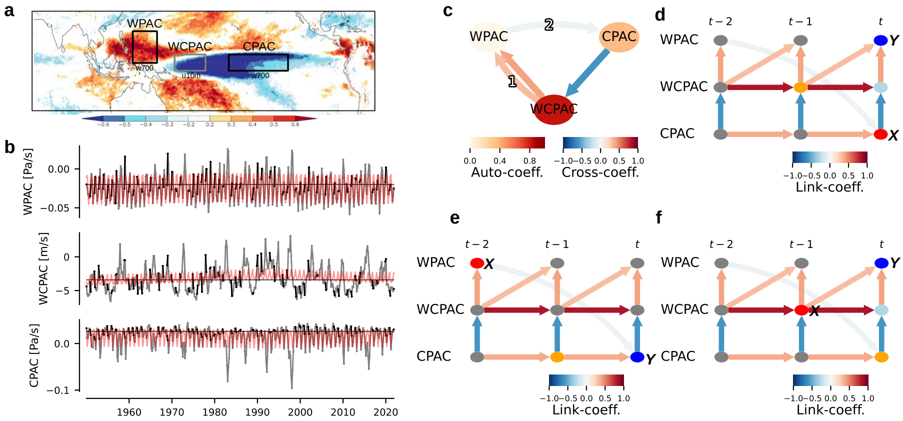


Fig. 3 Causal inference analysis of the Walker circulation. a | WPAC and CPAC are spatial averages of the 700mb-height level vertical velocity averaged across the West Pacific [130°E-150°E, 20°-0°N] and Central Pacific [160°W-120°W, 5°N-5°S], respectively, and WCPAC is the 10m zonal wind averaged across the West-Central Pacific [160°E-180°E, 5°N-5°S]. The underlying contours depict the correlations of the Nino3.4 index with the vertical anomalous Jan-Feb velocity field. **b** | Time series (black, masked samples in grey) extracted from ERA5 reanalysis data [146, 147]. The red lines represent the 15-year Gaussian kernel smoothed long-term trend plus the seasonal cycle. **c** | Assumed causal graph, collapsed in time, where straight edges are contemporaneous causal links and curved edges are lagged links with the corresponding lags shown as numbers on the curved edges. The edge colors show the link coefficients at the lag with maximum absolute link coefficient (lag label), and the node colors depict the link coefficient of the lag-1 autocorrelation links. **d** | Time series graph for causal effect estimate of X (red) on Y (blue), with mediators in light blue and optimal adjustment set in orange. Edge colors depict link coefficients. **e** | As in (**d**), but for different X and Y. **f** | As in (**d**), but for different X and Y. This analysis illustrates the many (pre-)processing steps that are necessary to fulfill the assumptions of causal inference methods and to transparently discuss causal conclusions.

westward surface tradewinds in the Western-central pacific (WCPAC), and rising air masses in the western Pacific (WPAC). This circulation exhibits a regime-dependence, as it strengthens during its La Niña phases, defined by anomalously cool ocean temperatures in the central-eastern Pacific, and weakens (or even reverses and shifts eastward) during El Niño phases, defined by anomalously warm ocean temperatures. In addition, the circulation also exhibits a seasonal regime-dependence with the circulation strength peaking in winter. The following analysis aims to gain a causal understanding of the anomalous circulation during winter-time neutral or La Niña phases.

Closely following the QAD-questionnaire (Tab. 1, Box 5, Fig. 2), the first step is to define and construct the variables of interest. Here the underlying ERA5 reanalysis data [146, 147] comes on a latitude-longitude grid and dimension-reduction needs to be employed. The WPAC and CPAC indices are chosen as variables because they correspond to major regions of, respectively, ascending and descending air masses as measured by their correlation with the Nino3.4 index (Fig. 3**a**). The WCPAC index, which measures the surface zonal wind in between the two regions, is chosen as a third variable. These three variables are all continuously-valued and are, as the next step, also temporally aggregated to a 2-monthly time resolution in order to average-out noisy monthly variability. This procedure defines a single observational dataset of three bi-monthly time series (Figure 3**b**).

The next question is about stationarity (Fig. 2). A trend and seasonal cycle can introduce non-stationary dependencies that can act as confounders. As pre-processing steps, the trend and seasonal cycle are removed by first subtracting both the trend (assumed to be caused by, for example, greenhouse gas forcing) and the seasonal cycle and then dividing by the seasonal variance. The length scale of the Gaussian kernel used for finding the long-term trend is assumed to be decadal (15 years). Further, considering the regime behavior of the system as explained above, only the periods November to February during the neutral or La Niña phases (defined by the 5-month running mean of the Nino3.4 index below 0.5°K) are considered. On a technical level, this selection of considered time periods is achieved by applying a so-called mask. More specifically, here the mask is such that samples for WPAC, WCPAC, and CPAC at time t can only come from Nov-Feb in neutral or La Niña conditions, while samples for the times $t-1$ and $t-2$ can also come from outside this mask. This masking procedure results in a small sample size of $n = 110$ during the years 1950-2021 (there are even fewer samples for El Niño phases), but at least these samples can reasonably be assumed to come from the same stationary distribution.

The causal question is now about causal effect estimation and mediation (Fig. 2). This quantification requires qualitative knowledge about cause-and-effect relationships in form of the causal time series graph. Causal discovery is challenging for such short sample sizes, so domain knowledge is used to come up with the qualitative feedback graph (Fig. 3**c**) with the time lags roughly based on the wind velocities in ref. [148]. The time-dependent structure is explicitly represented in time series graphs (Figs. **d-f**). As atmospheric processes are fast, contemporaneous causal effects are assumed (causal influences on a time scale below the data's time resolution of 2 months). However, for simplicity, it is assumed that there are no contemporaneous causal cycles (that is, if WPAC_t had a causal influence on CPAC_t, then CPAC_t could not have a causal influence on WPAC_t, and vice versa). Different graphs might be assumed here. For example, hidden confounders could be integrated, but due to the pre-processing steps, it is reasonable to assume there is no hidden confounding. However, a more thorough analysis could consider interdependencies with sea-surface temperature as a possible source of hidden confounding. The strength of causal inference lies in transparently laying out such assumptions.

With the causal graph chosen, the causal question is, firstly, on total causal effect estimation and, secondly, on linear mediation analysis among the three variables. According to the QAD flow chart (Fig. 2), here there is only a single dataset. Starting with $X = \text{CPAC}_t$ on $Y = \text{WPAC}_t$, the total causal effect along the causal chain $\text{CPAC}_t \to \text{WCPAC}_t \to \text{WPAC}_t$ in the graph (Fig. 3**d**) has no hidden confounders. Further, given the short sample size, a linearity assumption is reasonable, which suggests the path method (Fig. 2). However, it is prudent to also evaluate alternative methods, here by means of covariate adjustment using the optimal adjustment set $\mathbf{Z} = \{\text{WCPAC}_{t-1}\}$ (for theory see the paragraphs on covariate adjustment, optimal adjustment and linear causal effect estimation and the path method in the section on causal effects). The conditioning set $\mathbf{Z}$ blocks non-causal confounding of X and Y by CPAC_{t-1} as well as by further lagged confounders, and the estimation gives $\Delta_{X\to Y}(x' = 1, x = 0) = -0.31$ $[-0.38, -0.22]$ with a 90% confidence interval obtained via a bootstrap. Similarly, using the path-method instead of causal adjustment yields the causal effect estimate -0.24 $[-0.45, -0.09]$ as the product of the link coefficients of the two links on the causal chain $\text{CPAC}_t \to \text{WCPAC}_t \to \text{WPAC}_t$. Compared to these estimates, the non-causal (and in this case confounded) univariate regression of Y on X yields the value -0.99. The circulation is closed by the lagged causal effect of $X = \text{WPAC}_{t-2}$ on $Y = \text{CPAC}_t$ (Fig. 3**e**). Using covariate adjustment with $\mathbf{Z} = \{\text{CPAC}_{t-1}\}$ yields the relatively weak causal effect -0.05 $[-0.08, -0.02]$.

Finally, a linear mediation analysis assuming no latent confounders (for theory see the paragraph linear mediation analysis in the section on causal effects) can also be conducted by the path method (Fig. 2). The linear mediated effect of $X = \text{WCPAC}_{t-1}$ on $Y = \text{WPAC}_t$ through $M = \text{WCPAC}_t$ (Fig. 3**f**) is 0.33 [0.18, 0.49] whereas the direct effect that is not mediated through M is 0.31 [0.12, 0.53]. The mediated and non-mediated effect sum up to the causal effect 0.64 [0.50, 0.79]. All these effects are in units of the seasonally-standardized data.

This example demonstrates how domain knowledge, such as variable definitions and an assumed qualitative graph, and pre-processing steps to assure certain assumptions are leveraged in a causal inference analysis to quantify causal effects. There are many ways to modify the setup of this analysis: from choosing different variables, over accounting for confounding by sea-surface temperature, to different pre-processing steps and alternative assumptions about the graph. A more comprehensive analysis can and should transparently report how these different assumptions change the conclusions.

Biogeoscience example. Correlation-based analyses still dominate the literature in the biogeosciences, but this example will demonstrate the use of causal inference based techniques to investigate the causal effect of air temperature (Tair) on ecosystem respiration (Reco) from data that also includes gross primary production (GPP) and shortwave radiation (Rg). To better illustrate non-parametric causal effect estimation, this case study considers a synthetic system with known quantitative ground truth [149] akin to daily time series

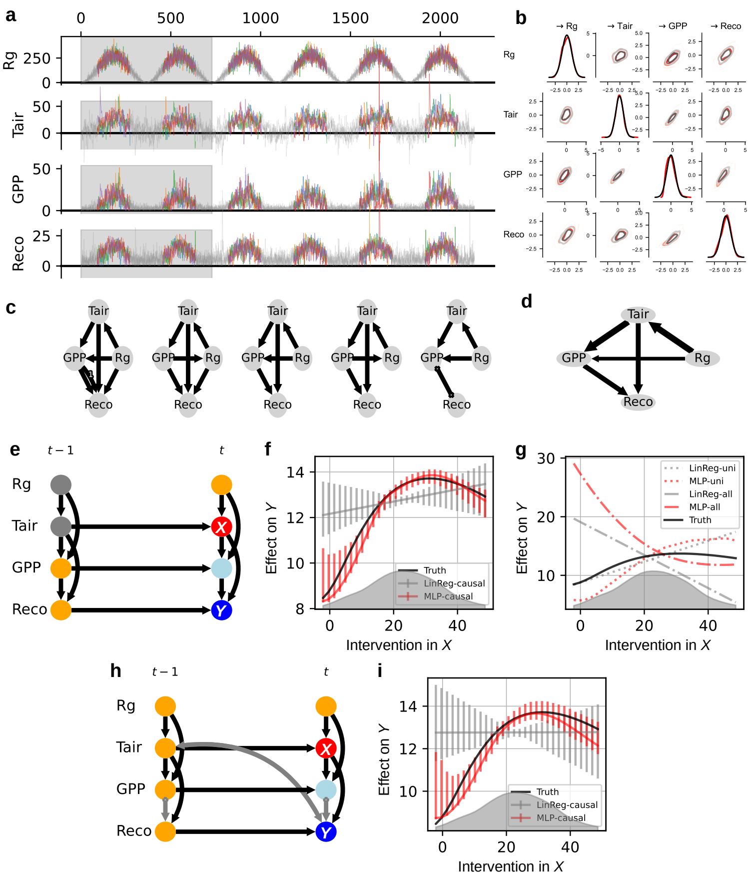


Fig. 4 Nonlinear causal effects of temperature on ecosystem respiration. a | Daily shortwave radiation (Rg), air temperature (Tair), gross primary production (GPP), and ecosystem respiration (Reco) at five synthetically modeled sites (line colors, masked winter samples in grey). **b** | Marginal and joint density plots, at lag zero, of data prior (red) and after (black) a transformation to normally distributed marginals. **c** | Resulting causal graphs of a 2-year sliding window causal discovery analysis (indicated by the grey band in panel **a**, step-size one year) depicted as summary graphs where straight edges are contemporaneous links and curved edges lagged links (label indicates time lag in days). **d** | Aggregated graph consisting of the most frequent link types (including absent links) across the five sliding window graphs. Edge width indicates the frequency of the link type. **e** | Assumed time series graph for causal effect estimation of X (red) on Y (blue) using the adjustment set in orange. **f** | Causal effect estimates for different intervention values (x-axis) with linear regression (LinReg) and a multilayer perceptron (MLP), including 90% bootstrap-based confidence intervals. The grey filled plot depicts the kernel-estimated density of Tair. The black line shows the interventional ground truth. **g** | Non-causal estimates by models without covariates ('-uni') and with all other variables as covariates ('-all'). **h** | As in panel **e**, but two bidirected arrows indicating hidden confounders are added (grey arrows, "$\leftrightarrow\!\!\!\!\!+$" indicates the existence of both a causal effect "$\rightarrow$" and confounding "$\leftrightarrow$"). **i** | Causal effect estimates for adjustment set shown in panel (**h**). This example illustrates that (non-parametric) causal effect estimation for two different assumed graphs yields robust results, whereas non-causal estimates are biased.

of eddy-covariance tower sites in the FLUXNET database [150]:

$$\begin{aligned}
\mathrm{Rg}_t &= |280\ \sin(t\pi/365)^2 + 50\ |\sin(t\pi/365)|\eta_t^{\mathrm{Rg}}| \\
\mathrm{Tair}_t &= 0.8\ \mathrm{Tair}_{t-1} + 0.02\ \mathrm{Rg}_t + 5\eta_t^{\mathrm{Tair}} \\
\mathrm{GPP}_t &= |0.2\ \mathrm{GPP}_{t-1} + 0.002\ \mathrm{Tair}_t\ \mathrm{Rg}_t + 3\eta_t^{\mathrm{GPP}}| \\
\mathrm{Reco}_t &= |0.3\ \mathrm{Reco}_{t-1} + 0.9\ \mathrm{GPP}_t\ 0.8^{0.12(\mathrm{Tair}_t-15)} + 2\eta_t^{\mathrm{Reco}}|
\end{aligned} \tag{7}$$

In these equations, which are interpreted as an SCM, the $\eta_t^{\cdot}$ are mutually independent standard normal noise terms, except for Tair where $\eta_t^{\mathrm{Tair}} = \eta_t + \frac{1}{4}\epsilon_t^3$ (standard normal noises η and ϵ) with a cubic exponent to represent more extreme temperatures. The SCM exhibits a unimodal relationship between Reco and Tair (see the interventional ground truth in Fig. 4**e**), which has also been found in real data [151, 152]. The analysis will first illustrate causal discovery and then causal effect estimation.

The steps again closely follow the QAD-template (Tab. 1 and Fig. 2). As opposed to the climate example, here all variables (nodes) are already defined as daily continuously-valued time series. The next question is about creating a stationary dataset (Fig. 2). Unlike in the climate example, here a setting with multiple datasets (multiple sites) is considered. In the considered synthetic example, stationarity is fulfilled by construction (apart from seasonality shared by all sites), as the sites are just different realizations of the same SCM. Hence, the time series from the different sites can be simply aggregated (pooled). To alleviate the seasonal non-stationarity, only the period April-September (model months) is considered (Figure 4**a**).

Given this stationary dataset, the first causal question is about causal discovery. To choose the appropriate causal discovery method, assumptions that can reasonably be made must be determined. The data here comes from multiple datasets (blue box in the causal discovery frame, Fig. 2), however, the datasets share the same underlying distribution and the next question is whether this system is deterministic. Given the complexity of the dynamics at this scale, it can be assumed to be a non-deterministic system. The next assumption to be made is whether or not there might be hidden confounders, that is, unobserved variables that causally influence two or more observed variables. Here, due to restricting the analysis to seasons across which stationarity can be expected, it is reasonable to assume the absence of hidden confounding, which is true in the underlying SCM.

The structural assumption of the graph type then needs to be made. As the processes here are fast, contemporaneous causal effects (that is, causal influences on a time scale below the data's time resolution of 1 day) might occur. Further, here the domain knowledge that Rg is exogeneous can be enforced by not allowing any parents of Rg in the graph. These assumptions suggest using the constraint-based causal discovery algorithm PCMCI$^+$ [82] (or other similar options, Fig. 2). As a constraint-based method, PCMCI$^+$ also requires that the causal Markov condition and faithfulness are assumed. The causal

Markov condition is the basis of causal inference and, in non-microscopic systems, reasonable to assume. The causal faithfulness assumption is less obvious as it, for example, excludes the situation that two causal influences along different pathways cancel each other out exactly. However, this assumption can still reasonably be made here.

To make an assumption on the maximal time lag in the causal time series graph estimated by PCMCI$^+$ (that is, the maximum over all τ such that $X^i_{t-\tau} \to X^j_t$ is in the graph), data can be used to investigate the lagged dependency functions [17, 21], or, as done here, domain knowledge can be used to justify $\tau_{\max} = 1$ (in units of days). The next hyperparameter choice for PCMCI$^+$ is about the conditional independence test, which requires a parametric assumption (Fig. 4**b**, red contours). The marginal densities are slightly non-normal, but by transforming the variables such that their marginals are normally distributed, the resulting joint densities are reasonably linear (Fig. 4**b**, black contours). Therefore, conditional independence tests in PCMCI$^+$ should be undertaken with a robust partial correlation (RobustParCorr), which is a variant of partial correlation where the variables are first transformed to normally-distributed marginals [153–155].

The data from all sites and all of the six years can now be used to estimate a single causal graph. Alternatively, a sliding window analysis should be considered if non-stationarity of the causal graph could be present. More for illustrative purposes, a sliding window analysis (grey window in Fig. 4**a**) is applied here that learns five causal graphs (Figure 4**c**), each by combing data from two successive of the six years and across all sites (hence $5 = 6 - 1$ graphs, each estimated from $n = 1840$ samples due to masking, significance level of PCMCI$^+$ set to $\alpha_{\mathrm{PC}} = 0.05$). Although there are errors in the individual sliding window graphs, the aggregated graph (Fig. 4**d**) agrees with the ground-truth, thus showing that after the normal-marginals transformation linearity is still a good approximation in this example. Alternatively, nonparametric independence tests can be utilized [21, 84], but there is a trade-off between more sophisticated parametric models and their computational and sample size demands. Finally, in a more realistic scenario, there might be some unobserved confounders. In such cases, a method for causal discovery in the presence of hidden confounders can be used, for example, LPCMCI [83]. Such more generally applicable algorithms will, however, generally come at the cost of a lower detection power.

The second causal question is about total causal effect estimation. According to the QAD flow chart (Fig. 2), this estimation requires knowledge of the causal time series graph. Using the causally discovered time series graph (Fig. 4**e**) the goal is to estimate the causal effect of Tair (red) on Reco (blue) at lag zero. Now all of the data ($n = 1840 \times 3$ samples) is utilized, which is stationary (from the same distribution) because the winter season is masked away.

The graph has no hidden confounders, but the effect is nonlinear, which suggests the optimal adjustment method (Fig. 2). A nonlinear adjustment (through eq. (5)) requires the choice of a statistical or machine learning model and here two approaches are compared: First, covariate adjustment with a linear regression (labeled LinReg, colored grey) and, secondly, covariate adjustment with a multilayer perceptron [156, default parameters] (abbreviated and labeled MLP, colored red), also known as feedforward neural network (Figure 4**f**). The linear regression fails while the MLP better estimates the nonlinear effect. The larger (bootstrap) confidence intervals at the left and right ends illustrate that the estimation worsens in regions with little training data. As a side note: the graph was estimated via causal discovery from the same data as the effect was estimated and, therefore, the confidence intervals could be too small (see comment on post-selection inference above). A compromise would be to split the sample.

An alternative graph can be assumed where hidden confounding between Tair_{t-1} and Reco_t as well as between GPP_t and Reco_t is present (Figure 4**h**). This hidden confounding is graphically represented by the bidirected edges $\mathrm{Tair}_{t-1} \leftrightarrow \mathrm{Reco}_t$ and $\mathrm{GPP}_t \leftrightarrow \mathrm{Reco}_t$, and it can emerge from other atmospheric and hydrological drivers. The causal effect in this modified hidden-variable graph can still be identified (Fig. 2), but the optimal adjustment set changes [130] (orange nodes in Fig. 4**h**). Still the causal effect estimates are almost the same (Fig. 4**i**). If, however, a confounding $\mathrm{Tair}_t \leftrightarrow \mathrm{Reco}_t$ were to be assumed, then the effect would not anymore be identifiable by adjustment nor by any other method, if no further assumptions are made. Estimation results for non-causal models without covariates ('-uni') and with all other variables as covariates ('-all') (Figure 4**g**) suffer from confounding or block the mediator GPP_t, respectively, both leading to biased estimates.

Recommendations and future perspectives

This Technical Review overviews causal inference for time series data with illustrations in Earth and environmental sciences. Our QAD-questionnaire is intended to guide domain scientists on how to understand and frame their questions in the language of causal inference and find the tools to solve them (Box 5, Tab. 1, Fig. 2), including weblinks to software packages (Tab. 2). Case study examples illustrate some typical challenges encountered: defining and pre-processing causal variables, dealing with non-stationarity, contemporaneous causation, and hidden confounding, choosing parametric models for nonlinear dependencies and non-Gaussian distributions, as well as practical data challenges of masking and how to utilise multiple datasets. While our application cases and discussions were focused on Earth sciences, these underlying challenges also apply to many other fields, from neurosciences to economics and social sciences.

With the pioneering work of the past decades [31–33], there is a sound mathematical basis to tackle causal questions across the sciences. However, there are still many open methodological problems, such as better dealing with uncertain expert knowledge and the spatio-temporal complexity of the underlying dynamical phenomena, as well as more robust and statistically efficient algorithms.

An important challenge for this endeavour is the broad language gap between the methodological and domain science communities with their individual jargon. The QAD-template is intended to bridge this gap by translating domain questions into actionable and precisely stated causal inference tasks. Similarly, much of the hesitation in adopting causal inference, next to missing solutions, also comes from a lack of benchmarks to help choose a suitable method, such as causeme.net [10]. More databases with benchmarks, based on real data with known ground truth or synthetic data with similar characteristics as in the domain under study, are needed, which will require more interactions between communities. Addressing a causal inference problem calls for constant interaction between the domain scientist and the method developer; assumptions must be discussed and formalized, data characteristics must be jointly analyzed, and conclusions must be assessed under both lenses.

Machine learning and deep learning are beginning to permeate all scientific fields and there is a need for explanation and interpretation of these often opaque methods. Explainable AI approaches [157] help in explaining the features used by a model, however, these explanations generally carry no causal meaning about the underlying system. We envision, and urge the respective communities to utilize causal inference as the vital link between the data-driven machine learning powerhouse and the ample knowledge prevalent in domain sciences. Together, both can help tackle pressing scientific problems, in Earth sciences from process understanding and causally robust forecasting [23, 24] to evaluating and improving the physics embedded in climate models [10, 25, 27] and attributing climate change and extreme events through counterfactual analyses [28, 57].

Arguably, most of the research questions in Earth sciences can be framed as causal inference problems, with many studies hoping to understand how one factor impacts another. However, conclusions are often either void of explicit causal language or must be more explicit about the assumptions needed to arrive at causal conclusions. Our suggestion is to integrate causal thinking into data-driven science. We encourage scientists to be more explicit and transparent in laying out assumptions that enable more robust conclusions and in assessing and discussing conclusions under alternative sets of assumptions. Furthermore, we argue that this is not only sensible but necessary when dealing with observational data to analyze systems, like the Earth, that bring many challenges and where conclusions have relevant environmental, economic, and societal implications. Causal inference offers tools rooted in sound mathematical grounds and a vibrant, interdisciplinary community to take on the challenge.

Table 2 Open-source software packages for selected causal inference methods. Some methods are specifically adapted to time series data. See the section "Data availability and links to software packages" for links and brief descriptions of the software packages.

Method	Software
Causal discovery	
PC [31, 78], FCI [81]	Tetrad, causal-learn, pcalg, Tigramite, MXM, bnlearn/dbnlearn, PyWhy
PCMCI [21], PCMCI$^+$ [82], LPCMCI [83]	Tigramite
GES [107–109]	Tetrad, causal-learn, pcalg
MMHC [110]	MXM, bnlearn/dbnlearn
VARLiNGAM [99]	causal-learn
Joint Causal Inference [61]	pcalg
seqICP [68–70]	InvariantCausalPrediction
CCM [72, 73]	rEDM
Granger causality [85, 87–89]	causal-learn, statsmodels
DYNOTEARS [116]	Causalnex
Causal effect estimation	
Path method [138–140]	Tigramite
Optimal adjustment [130]	Tigramite, pcalg
Adjustment [33]	pcalg, Tigramite, dagitty, CausalFusion, PyWhy
Front-door adjustment [120], do-calculus [120–123]	CausalFusion
Instrumental variables [60, 131]	dagitty, CausalFusion, Linearmodels
Double machine learning [125, 126]	econML
Causal transportability [48]	CausalFusion
Fixed-effects panel regression [60]	Linearmodels

References

[1] Masson-Delmotte, V. *et al.* *Climate change 2021: the physical science basis* Vol. 2 (Cambridge University Press Cambridge, UK, 2021).

[2] Eyring, V. *et al.* Overview of the coupled model intercomparison project phase 6 (cmip6) experimental design and organization. *Geoscientific Model Development* **9** (5), 1937–1958 (2016) .

[3] Hastie, T., Tibshirani, R., Friedman, J. H. & Friedman, J. H. *The elements of statistical learning: data mining, inference, and prediction* Vol. 2 (Springer, 2009).

[4] Goodfellow, I., Bengio, Y. & Courville, A. *Deep learning* (MIT press, 2016).

[5] Ghahramani, Z. Probabilistic machine learning and artificial intelligence. *Nature* **521** (7553), 452–459 (2015) .

[6] Murphy, K. P. *Probabilistic machine learning: an introduction* (MIT press, 2022).

[7] Reichstein, M. *et al.* Deep learning and process understanding for data-driven Earth system science. *Nature* **566**, 195–204 (2019) .

[8] Camps-Valls, G., Tuia, D., Zhu, X. X. & (Editors), M. R. *Deep learning for the Earth Sciences: A comprehensive approach to remote sensing, climate science and geosciences* (Wiley & Sons, 2021).

[9] Pearl, J. Causal inference in statistics: An overview. *Stat. Surv.* **3**, 96–146 (2009) .

[10] Runge, J. *et al.* Inferring causation from time series in earth system sciences. *Nature communications* **10** (1), 1–13 (2019) .

[11] Diaz, E., Adsuara, J., Moreno-Martinez, A., Piles, M. & Camps-Valls, G. Inferring causal relations from observational long-term carbon and water fluxes records. *Scientific Reports* **12**, 1610 (2022) .

[12] Ebert-Uphoff, I. & Deng, Y. Causal discovery in the geosciences—using synthetic data to learn how to interpret results. *Computers & Geosciences* **99**, 50–60 (2017) .

[13] Niemeijer, D. & de Groot, R. S. Framing environmental indicators: moving from causal chains to causal networks. *Environment, development and sustainability* **10** (1), 89–106 (2008) .

[14] Shepherd, T. G. *et al.* Storylines: an alternative approach to representing uncertainty in physical aspects of climate change. *Climatic change* **151** (3), 555–571 (2018) .

[15] Fan, J., Wang, Y., Rosenfeld, D. & Liu, X. Review of aerosol–cloud interactions: Mechanisms, significance, and challenges. *Journal of the Atmospheric Sciences* **73** (11), 4221–4252 (2016) .

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

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

[18] Kretschmer, M., Coumou, D., Donges, J. F. & Runge, J. Using causal effect networks to analyze different arctic drivers of midlatitude winter circulation. *Journal of climate* **29** (11), 4069–4081 (2016) .

[19] Di Capua, G. *et al.* Tropical and mid-latitude teleconnections interacting with the indian summer monsoon rainfall: a theory-guided causal effect network approach. *Earth System Dynamics* **11** (1), 17–34 (2020) .

[20] Kretschmer, M. *et al.* Quantifying causal pathways of teleconnections. *Bulletin of the American Meteorological Society* **102** (12), E2247–E2263 (2021) .

[21] Runge, J., Nowack, P., Kretschmer, M., Flaxman, S. & Sejdinovic, D. Detecting and quantifying causal associations in large nonlinear time series datasets. *Science advances* **5** (11), eaau4996 (2019) .

[22] Karmouche, S. *et al.* Regime-oriented causal model evaluation of atlantic-pacific teleconnections in cmip6. *EGUsphere* 1–42 (2022) .

[23] Kretschmer, M., Runge, J. & Coumou, D. Early prediction of extreme stratospheric polar vortex states based on causal precursors. *Geophysical research letters* **44** (16), 8592–8600 (2017) .

[24] Runge, J., Donner, R. V. & Kurths, J. Optimal model-free prediction from multivariate time series. *Physical Review E* **91** (5), 052909 (2015) .

[25] Eyring, V. *et al.* Taking climate model evaluation to the next level. *Nature Climate Change* **9** (2), 102–110 (2019) .

[26] Eyring, V. *et al.* Earth system model evaluation tool (esmvaltool) v2. 0–an extended set of large-scale diagnostics for quasi-operational and comprehensive evaluation of earth system models in cmip. *Geoscientific Model Development* **13** (7), 3383–3438 (2020) .

[27] Nowack, P., Runge, J., Eyring, V. & Haigh, J. D. Causal networks for climate model evaluation and constrained projections. *Nature communications* **11** (1), 1–11 (2020) .

[28] Bindoff, N. L. *et al.* *Detection and attribution of climate change: from global to regional* (Cambridge University Press, 2013).

[29] Pearl, J., Glymour, M. & Jewell, N. P. *Causal inference in statistics: A Primer* (John Wiley & Sons, 2016).

[30] Pearl, J. & Mackenzie, D. *The Book of Why: The New Science of Cause and Effect* (Basic books, New York, 2018).

[31] Spirtes, P., Glymour, C. & Scheines, R. *Causation, Prediction, and Search* (MIT Press, Boston, 2000).

[32] Peters, J., Janzing, D. & Schölkopf, B. *Elements of causal inference: foundations and learning algorithms* (The MIT Press, 2017).

[33] Pearl, J. *Causality: Models, Reasoning, and Inference* 2nd edn (Cambridge University Press, Cambridge, UK, 2009).

[34] Rubin, D. B. Bayesian inference for causal effects: The role of randomization. *The Annals of statistics* 34–58 (1978) .

[35] Rubin, D. B. Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of educational Psychology* **66** (5), 688 (1974) .

[36] Imbens, G. W. & Rubin, D. B. *Causal inference in statistics, social, and biomedical sciences* (Cambridge University Press, 2015).

[37] Hernan, M. & Robins, J. *Causal Inference: What if.* (Chapman & Hill/CRC, 2020).

[38] Bollen, K. A. *Structural Equations with Latent Variables* (John Wiley & Sons, New York, NY, USA, 1989).

[39] Bareinboim, E., Correa, J. D., Ibeling, D. & Icard, T. *On Pearl's Hierarchy and the Foundations of Causal Inference*, 507–556 (Association for Computing Machinery, New York, NY, USA, 2022), 1 edn.

[40] Dahlhaus, R. & Eichler, M. Causality and graphical models in time series analysis. *Oxford Stat. Sci. Ser* **27** (2003) .

[41] Runge, J., Heitzig, J., Petoukhov, V. & Kurths, J. Escaping the curse of dimensionality in estimating multivariate transfer entropy. *Physical Review Letters* **108**, 258701 (2012) .

[42] Bongers, S., Forré, P., Peters, J. & Mooij, J. M. Foundations of structural causal models with cycles and latent variables. *The Annals of Statistics* **49** (5), 2885 – 2915 (2021) .

[43] Forré, P. & Mooij, J. M. Adams, R. P. & Gogate, V. (eds) *Causal calculus in the presence of cycles, latent confounders and selection bias.* (eds Adams, R. P. & Gogate, V.) *Proceedings of the 35th Conference on Uncertainty in Artificial Intelligence (UAI-18)* (AUAI Press, 2019).

[44] Forré, P. & Mooij, J. M. Globerson, A. & Silva, R. (eds) *Constraint-based causal discovery for non-linear structural causal models with cycles and latent confounders.* (eds Globerson, A. & Silva, R.) *Proceedings of*

the 34th Conference on Uncertainty in Artificial Intelligence (UAI-18) (AUAI Press, 2018).

[45] Verma, T. & Pearl, J. Bonissone, P. P., Henrion, M., Kanal, L. N. & Lemmer, J. F. (eds) *Equivalence and synthesis of causal models.* (eds Bonissone, P. P., Henrion, M., Kanal, L. N. & Lemmer, J. F.) *Proceedings of the Sixth Annual Conference on Uncertainty in Artificial Intelligence*, UAI '90, 255–270 (Elsevier Science Inc., New York, NY, USA, 1990).

[46] Verma, T. Graphical aspects of causal models technical report no. r-191. Tech. Rep. Technical Report (R-191), UCLA Cognitive Systems Laboratory (1993). Draft Copy.

[47] Bareinboim, E., Correa, J. D., Ibeling, D. & Icard, T. in *On pearl's hierarchy and the foundations of causal inference* (eds Geffner, H., Dechter, R. & Halpern, J. Y.) *Probabilistic and Causal Inference: The Works of Judea Pearl* Ch. VI, 507–556 (ACM books, 2022).

[48] Bareinboim, E. & Pearl, J. Causal inference and the data-fusion problem. *Proceedings of the National Academy of Sciences* **113** (27), 7345–7352 (2016) .

[49] Runge, J. *et al.* Identifying causal gateways and mediators in complex spatio-temporal systems. *Nature communications* **6** (1), 1–10 (2015) .

[50] Mogensen, S. W. Schölkopf, B., Uhler, C. & Zhang, K. (eds) *Equality constraints in linear hawkes processes.* (eds Schölkopf, B., Uhler, C. & Zhang, K.) *Proceedings of the First Conference on Causal Learning and Reasoning*, Vol. 177 of *Proceedings of Machine Learning Research*, 576–593 (PMLR, 2022).

[51] Schölkopf, B. *et al.* Toward causal representation learning. *Proceedings of the IEEE* **109** (5), 612–634 (2021) .

[52] Varando, G., Fernández-Torres, M.-Á., Muñoz-Marí, J. & Camps-Valls, G. Bengio, Y. & von Kügelen, J. (eds) *Learning causal representations with granger pca.* (eds Bengio, Y. & von Kügelen, J.) *UAI 2022 Workshop on Causal Representation Learning* (openreview.net, 2022).

[53] Donges, J. F., Zou, Y., Marwan, N. & Kurths, J. Complex networks in climate dynamics. *The European Physical Journal Special Topics* **174** (1), 157–179 (2009) .

[54] Boers, N. *et al.* Complex networks reveal global pattern of extreme-rainfall teleconnections. *Nature* **566** (7744), 373–377 (2019) .

[55] Ludescher, J. *et al.* Network-based forecasting of climate phenomena. *Proceedings of the National Academy of Sciences* **118** (47), e1922872118 (2021) .

[56] Halpern, J. Y. *Actual causality* (MiT Press, 2016).

[57] Hannart, A., Pearl, J., Otto, F. E., Naveau, P. & Ghil, M. Causal counterfactual theory for the attribution of weather and climate-related events. *Bull. Am. Meteorol. Soc.* **97** (1), 99–110 (2016) .

[58] VanderWeele, T. *Explanation in causal inference: methods for mediation and interaction* (Oxford University Press, 2015).

[59] Correa, J., Lee, S. & Bareinboim, E. Ranzato, M., Beygelzimer, A., Dauphin, Y., Liang, P. & Vaughan, J. W. (eds) *Nested counterfactual identification from arbitrary surrogate experiments.* (eds Ranzato, M., Beygelzimer, A., Dauphin, Y., Liang, P. & Vaughan, J. W.) *Advances in Neural Information Processing Systems 34 (NeurIPS 2021)*, Vol. 34, 6856–6867 (2021).

[60] Angrist, J. D. & Pischke, J.-S. *Mostly harmless econometrics: An empiricist's companion* (Princeton university press, 2009).

[61] Mooij, J. M., Magliacane, S. & Claassen, T. Joint causal inference from multiple contexts. *Journal of Machine Learning Research* **21** (99), 1–108 (2020) .

[62] Huang, B. *et al.* Causal discovery from heterogeneous/nonstationary data. *Journal of Machine Learning Research* **21** (89), 1–53 (2020) .

[63] Mohan, K. & Pearl, J. Graphical models for processing missing data. *Journal of the American Statistical Association* **116** (534), 1023–1037 (2021) .

[64] Hamilton, J. D. State-space models. *Handbook of econometrics* **4**, 3039–3080 (1994) .

[65] Berk, R., Brown, L., Buja, A., Zhang, K. & Zhao, L. Valid post-selection inference. *The Annals of Statistics* **41** (2), 802 – 837 (2013). URL https://doi.org/10.1214/12-AOS1077. https://doi.org/10.1214/12-AOS1077 .

[66] Rinaldo, A., Wasserman, L. & G'Sell, M. Bootstrapping and sample splitting for high-dimensional, assumption-lean inference. *The Annals of Statistics* **47** (6), 3438–3469 (2019) .

[67] Heinze-Deml, C., Maathuis, M. H. & Meinshausen, N. Causal structure learning. *Annual Review of Statistics and Its Application* **5**, 371–391

(2018) .

[68] Peters, J., Bühlmann, P. & Meinshausen, N. Causal inference by using invariant prediction: identification and confidence intervals. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **78** (5), 947–1012 (2016) .

[69] Heinze-Deml, C., Peters, J. & Meinshausen, N. Invariant causal prediction for nonlinear models. *Journal of Causal Inference* **6** (2) (2018) .

[70] Pfister, N., Bühlmann, P. & Peters, J. Invariant causal prediction for sequential data. *Journal of the American Statistical Association* **114** (527), 1264–1276 (2019) .

[71] Daniušis, P. *et al.* Grunwald, P. & Spirtes, P. (eds) *Inferring deterministic causal relations.* (eds Grunwald, P. & Spirtes, P.) *Proceedings of the Twenty-Sixth Conference on Uncertainty in Artificial Intelligence*, 143–150 (2010).

[72] Sugihara, G. *et al.* Detecting causality in complex ecosystems. *science* **338** (6106), 496–500 (2012) .

[73] Ye, H., Deyle, E. R., Gilarranz, L. J. & Sugihara, G. Distinguishing time-delayed causal interactions using convergent cross mapping. *Scientific reports* **5** (1), 1–9 (2015) .

[74] Verma, T. & Pearl, J. in *Causal networks: Semantics and expressiveness* (eds Shachter, R. D., Levitt, T. S., Kanal, L. N. & Lemmer, J. F.) *Uncertainty in Artificial Intelligence*, Vol. 9 of *Machine Intelligence and Pattern Recognition* 69–76 (North-Holland, 1990).

[75] Geiger, D., Verma, T. & Pearl, J. Identifying independence in bayesian networks. *Networks* **20** (5), 507–534 (1990) .

[76] Pearl, J. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference* (Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 1988).

[77] Runge, J. Causal network reconstruction from time series: From theoretical assumptions to practical estimation. *Chaos: An Interdisciplinary Journal of Nonlinear Science* **28** (7), 075310 (2018) .

[78] Spirtes, P. & Glymour, C. An algorithm for fast recovery of sparse causal graphs. *Social science computer review* **9** (1), 62–72 (1991) .

[79] Spirtes, P., Meek, C. & Richardson, T. Besnard, P. & Hanks, S. (eds) *Causal inference in the presence of latent variables and selection bias.* (eds Besnard, P. & Hanks, S.) *Proceedings of the Eleventh Conference on Uncertainty in Artificial Intelligence*, UAI'95, 499–506 (Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 1995).

[80] Zhang, J. On the completeness of orientation rules for causal discovery in the presence of latent confounders and selection bias. *Artificial Intelligence* **172** (16), 1873–1896 (2008) .

[81] Entner, D. & Hoyer, P. O. Myllymäki, P., Roos, T. & Jaakkola, T. (eds) *On causal discovery from time series data using fci.* (eds Myllymäki, P., Roos, T. & Jaakkola, T.) *Proceedings of the 5th European Workshop on Probabilistic Graphical Models*, 121–128 (Helsinki Institute for Information Technology HIIT, Helsinki, FI, 2010).

[82] Runge, J. Peters, J. & Sontag, D. (eds) *Discovering contemporaneous and lagged causal relations in autocorrelated nonlinear time series datasets.* (eds Peters, J. & Sontag, D.) *Proceedings of the 36th Conference on Uncertainty in Artificial Intelligence (UAI)*, Vol. 124 of *Proceedings of Machine Learning Research*, 1388–1397 (PMLR, 2020).

[83] Gerhardus, A. & Runge, J. Larochelle, H., Ranzato, M., Hadsell, R., Balcan, M. & Lin, H. (eds) *High-recall causal discovery for autocorrelated time series with latent confounders.* (eds Larochelle, H., Ranzato, M., Hadsell, R., Balcan, M. & Lin, H.) *Advances in Neural Information Processing Systems 34 (NeurIPS 2021)*, Vol. 33, 12615–12625 (2020).

[84] Runge, J. Storkey, A. & Perez-Cruz, F. (eds) *Conditional independence testing based on a nearest-neighbor estimator of conditional mutual information.* (eds Storkey, A. & Perez-Cruz, F.) *International Conference on Artificial Intelligence and Statistics*, Vol. 84 of *Proceedings of Machine Learning Research*, 938–947 (PMLR, 2018).

[85] Granger, C. W. Investigating causal relations by econometric models and cross-spectral methods. *Econometrica: journal of the Econometric Society* 424–438 (1969) .

[86] Barrett, A. B., Barnett, L. & Seth, A. K. Multivariate Granger causality and generalized variance. *Phys. Rev. E* **81** (4), 41907 (2010) .

[87] Baek, E. & Brock, W. A general test for nonlinear granger causality: Bivariate model. *Iowa State University and University of Wisconsin at Madison Working Paper* (137-156) (1992) .

[88] Hiemstra, C. & Jones, J. D. Testing for linear and nonlinear granger causality in the stock price-volume relation. *The Journal of Finance*

49 (5), 1639–1664 (1994) .

[89] Diks, C. & Panchenko, V. A new statistic and practical guidelines for nonparametric granger causality testing. *Journal of Economic Dynamics and Control* **30** (9-10), 1647–1669 (2006) .

[90] Schreiber, T. Measuring information transfer. *Physical review letters* **85** (2), 461 (2000) .

[91] Smirnov, D. A. Generative formalism of causality quantifiers for processes. *Physical Review E* **105** (3), 034209 (2022) .

[92] Sun, J., Taylor, D. & Bollt, E. M. Causal network inference by optimal causation entropy. *SIAM Journal on Applied Dynamical Systems* **14** (1), 73–106 (2015) .

[93] Triacca, U. Is granger causality analysis appropriate to investigate the relationship between atmospheric concentration of carbon dioxide and global surface air temperature? *Theoretical and applied climatology* **81** (3), 133–135 (2005) .

[94] Diego Bueso, G. C.-V., Maria Piles. Explicit granger causality in kernel hilbert spaces. *Physical Review E* **102**, 062201 (2020) .

[95] McGraw, M. C. & Barnes, E. A. Memory matters: A case for granger causality in climate variability studies. *Journal of climate* **31** (8), 3289–3300 (2018) .

[96] Papagiannopoulou, C. *et al.* A non-linear granger-causality framework to investigate climate–vegetation dynamics. *Geoscientific Model Development* **10** (5), 1945–1960 (2017) .

[97] Shimizu, S., Hoyer, P. O., Hyväinen, A. & Kerminen, A. A linear non-gaussian acyclic model for causal discovery. *Journal of Machine Learning Research* **7** (72), 2003–2030 (2006) .

[98] Hoyer, P. O. *et al.* McAllester, D. & Myllymaki, P. (eds) *Causal discovery of linear acyclic models with arbitrary distributions.* (eds McAllester, D. & Myllymaki, P.) *Proceedings of the Twenty-Fourth Conference on Uncertainty in Artificial Intelligence*, UAI'08, 282–289 (AUAI Press, Arlington, Virginia, USA, 2008).

[99] Hyvärinen, A., Zhang, K., Shimizu, S. & Hoyer, P. O. Estimation of a structural vector autoregression model using non-gaussianity. *Journal of Machine Learning Research* **11** (May), 1709–1731 (2010) .

[100] Hoyer, P., Janzing, D., Mooij, J. M., Peters, J. & Schölkopf, B. Koller, D., Schuurmans, D., Bengio, Y. & Bottou, L. (eds) *Nonlinear causal discovery with additive noise models.* (eds Koller, D., Schuurmans, D., Bengio, Y. & Bottou, L.) *Advances in Neural Information Processing Systems*, Vol. 21 (Curran Associates, Inc., 2008).

[101] Peters, J., Janzing, D. & Schölkopf, B. Burges, C., Bottou, L., Welling, M., Ghahramani, Z. & Weinberger, K. (eds) *Causal inference on time series using restricted structural equation models.* (eds Burges, C., Bottou, L., Welling, M., Ghahramani, Z. & Weinberger, K.) *Advances in Neural Information Processing Systems*, Vol. 26 (Curran Associates, Inc., 2013).

[102] Zhang, K. & Hyvärinen, A. McAllester, D. (ed.) *On the identifiability of the post-nonlinear causal model.* (ed.McAllester, D.) *Proceedings of the Twenty-Fifth Conference on Uncertainty in Artificial Intelligence*, UAI '09, 647–655 (AUAI Press, Arlington, Virginia, USA, 2009).

[103] Gnecco, N., Meinshausen, N., Peters, J. & Engelke, S. Causal discovery in heavy-tailed models. *The Annals of Statistics* **49** (3), 1755–1778 (2021) .

[104] Peters, J., Mooij, J. M., Janzing, D. & Schölkopf, B. Causal discovery with continuous additive noise models. *Journal of Machine Learning Research* **15** (58), 2009–2053 (2014) .

[105] Pérez-Suay, A. & Camps-Valls, G. Sensitivity maps of the hilbert–schmidt independence criterion. *Applied Soft Computing* **70**, 1054–1063 (2018) .

[106] Pérez-Suay, A. & Camps-Valls, G. Causal inference in geoscience and remote sensing from observational data. *IEEE Transactions on Geoscience and Remote Sensing* **57** (3), 1502–1513 (2019) .

[107] Chickering, D. M. Learning equivalence classes of bayesian-network structures. *Journal of Machine Learning Research* **2**, 445–498 (2002) .

[108] Chickering, D. M. Optimal structure identification with greedy search. *Journal of machine learning research* **3** (Nov), 507–554 (2002) .

[109] Chickering, M. Peters, J. & Sontag, D. (eds) *Statistically efficient greedy equivalence search.* (eds Peters, J. & Sontag, D.) *Proceedings of the 36th Conference on Uncertainty in Artificial Intelligence (UAI)*, Vol. 124 of *Proceedings of Machine Learning Research*, 241–249 (PMLR, 2020).

[110] Tsamardinos, I., Brown, L. E. & Aliferis, C. F. The max-min hill-climbing bayesian network structure learning algorithm. *Machine*

learning **65** (1), 31–78 (2006) .

[111] Liu, J. & Niyogi, D. Identification of linkages between urban heat island magnitude and urban rainfall modification by use of causal discovery algorithms. *Urban Climate* **33**, 100659 (2020) .

[112] Mäkelä, J. *et al.* Incorporating expert domain knowledge into causal structure discovery workflows. *Biogeosciences* **19** (8), 2095–2099 (2022) .

[113] Zheng, X., Aragam, B., Ravikumar, P. K. & Xing, E. P. Bengio, S. *et al.* (eds) *DAGs with NO TEARS: Continuous optimization for structure learning.* (eds Bengio, S. *et al.*) *Advances in Neural Information Processing Systems*, Vol. 31 (Curran Associates, Inc., 2018).

[114] Zheng, X., Dan, C., Aragam, B., Ravikumar, P. & Xing, E. Chiappa, S. & Calandra, R. (eds) *Learning sparse nonparametric DAGs.* (eds Chiappa, S. & Calandra, R.) *Proceedings of the Twenty Third International Conference on Artificial Intelligence and Statistics*, Vol. 108 of *Proceedings of Machine Learning Research*, 3414–3425 (PMLR, 2020).

[115] Lorch, L., Rothfuss, J., Schölkopf, B. & Krause, A. Ranzato, M., Beygelzimer, A., Dauphin, Y., Liang, P. & Vaughan, J. W. (eds) *Dibs: Differentiable bayesian structure learning.* (eds Ranzato, M., Beygelzimer, A., Dauphin, Y., Liang, P. & Vaughan, J. W.) *Advances in Neural Information Processing Systems*, Vol. 34, 24111–24123 (Curran Associates, Inc., 2021).

[116] Pamfil, R. *et al.* Chiappa, S. & Calandra, R. (eds) *DYNOTEARS: Structure learning from time-series data.* (eds Chiappa, S. & Calandra, R.) *Proceedings of the Twenty Third International Conference on Artificial Intelligence and Statistics*, Vol. 108 of *Proceedings of Machine Learning Research*, 1595–1605 (PMLR, 2020).

[117] Pearl, J. Breese, J. & Koller, D. (eds) *Direct and indirect effects.* (eds Breese, J. & Koller, D.) *UAI'01: Proceedings of the Seventeenth conference on Uncertainty in artificial intelligence* (Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 2001).

[118] Fisher, R. A. *The Design of Experiments* (Hafner Press, 1935).

[119] Gryspeerdt, E., Quaas, J. & Bellouin, N. Constraining the aerosol influence on cloud fraction. *Journal of Geophysical Research: Atmospheres* **121** (7), 3566–3583 (2016) .

[120] Pearl, J. Causal diagrams for empirical research. *Biometrika* **82** (4), 669–688 (1995) .

[121] Huang, Y. & Valtorta, M. Dechter, R. & Richardson, T. (eds) *Pearl's calculus of intervention is complete.* (eds Dechter, R. & Richardson, T.) *Proceedings of the Twenty-Second Conference on Uncertainty in Artificial Intelligence*, UAI'06, 217–224 (AUAI Press, Arlington, Virginia, USA, 2006).

[122] Shpitser, I. & Pearl, J. Dechter, R. & Richardson, T. (eds) *Identification of conditional interventional distributions.* (eds Dechter, R. & Richardson, T.) *Proceedings of the Twenty-Second Conference on Uncertainty in Artificial Intelligence*, UAI'06, 437–444 (AUAI Press, Arlington, Virginia, USA, 2006).

[123] Shpitser, I. & Pearl, J. Complete identification methods for the causal hierarchy. *Journal of Machine Learning Research* **9**, 1941–1979 (2008) .

[124] Maathuis, M. H. & Colombo, D. A generalized back-door criterion. *The Annals of Statistics* **43** (3), 1060–1088 (2015) .

[125] Chernozhukov, V. *et al.* Double/debiased machine learning for treatment and structural parameters: Double/debiased machine learning. *The Econometrics Journal* **21** (1) (2018) .

[126] Jung, Y., Tian, J. & Bareinboim, E. Honavar, V. & Spaan, M. (eds) *Estimating identifiable causal effects through double machine learning.* (eds Honavar, V. & Spaan, M.) *Proceedings of the Thirty-Fifth AAAI Conference on Artificial Intelligence*, Vol. 35, 12113–12122 (2021).

[127] Shpitser, I., VanderWeel, T. & Robins, J. M. Grunwald, P. & Spirtes, P. (eds) *On the validity of covariate adjustment for estimating causal effects.* (eds Grunwald, P. & Spirtes, P.) *UAI'10: Proceedings of the Twenty-Sixth Conference on Uncertainty in Artificial Intelligence* (AUAI Press, 2010).

[128] Henckel, L., Perković, E., Maathuis, M. H. *et al.* Graphical criteria for efficient total effect estimation via adjustment in causal linear models. *Journal of the Royal Statistical Society Series B* **84** (2), 579–599 (2022) .

[129] Rotnitzky, A. & Smucler, E. Efficient adjustment sets for population average causal treatment effect estimation in graphical models. *J. Mach. Learn. Res.* **21** (188), 1–86 (2020) .

[130] Runge, J. Ranzato, M., Beygelzimer, A., Dauphin, Y., Liang, P. & Vaughan, J. W. (eds) *Necessary and sufficient graphical conditions for optimal adjustment sets in causal graphical models with hidden variables.* (eds Ranzato, M., Beygelzimer, A., Dauphin, Y., Liang, P. & Vaughan, J. W.) *Advances in Neural Information Processing Systems 34 (NeurIPS 2021)* (2021).

[131] Sargan, J. D. The estimation of economic relationships using instrumental variables. *Econometrica: Journal of the econometric society* 393–415 (1958) .

[132] Bareinboim, E. & Pearl, J. Lawrence, N. D. & Girolami, M. (eds) *Controlling selection bias in causal inference.* (eds Lawrence, N. D. & Girolami, M.) *Proceedings of the Fifteenth International Conference on Artificial Intelligence and Statistics*, Vol. 22 of *Proceedings of Machine Learning Research*, 100–108 (PMLR, La Palma, Canary Islands, 2012).

[133] Bareinboim, E., Tian, J. & Pearl, J. Brodley, C. E. & Stone, P. (eds) *Recovering from selection bias in causal and statistical inference.* (eds Brodley, C. E. & Stone, P.) , Vol. 28 (2014).

[134] Correa, J., Tian, J. & Bareinboim, E. McIlraith, S. & Weinberger, K. (eds) *Generalized adjustment under confounding and selection biases.* (eds McIlraith, S. & Weinberger, K.) *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 32 (2018).

[135] Correa, J. D., Tian, J. & Bareinboim, E. Van Hentenryck, P. & Zhou, Z.-H. (eds) *Identification of causal effects in the presence of selection bias.* (eds Van Hentenryck, P. & Zhou, Z.-H.) *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 33, 2744–2751 (2019).

[136] Mohan, K., Pearl, J. & Tian, J. Burges, C., Bottou, L., Welling, M., Ghahramani, Z. & Weinberger, K. (eds) *Graphical models for inference with missing data.* (eds Burges, C., Bottou, L., Welling, M., Ghahramani, Z. & Weinberger, K.) *Advances in Neural Information Processing Systems*, Vol. 26 (Curran Associates, Inc., 2013).

[137] Shpitser, I., Mohan, K. & Pearl, J. Meila, M. & Heskes, T. (eds) *Missing data as a causal and probabilistic problem.* (eds Meila, M. & Heskes, T.) *Proceedings of the Thirty-First Conference on Uncertainty in Artificial Intelligence*, UAI'15, 802–811 (AUAI Press, Arlington, Virginia, USA, 2015).

[138] Wright, S. Correlation and causation. *J. Agric. Res.* **20** (7), 557–585 (1921) .

[139] Nandy, P., Maathuis, M. H. & Richardson, T. S. Estimating the effect of joint interventions from observational data in sparse high-dimensional settings. *The Annals of Statistics* **45** (2), 647–674 (2017) .

[140] Guo, F. R. Efficient least squares for estimating total effects under linearity and causal sufficiency. *Journal of Machine Learning Research* **23** (104), 1–41 (2022) .

[141] Robins, J. M. & Greenland, S. Identifiability and exchangeability for direct and indirect effects. *Epidemiology* 143–155 (1992) .

[142] VanderWeele, T. & Vansteelandt, S. Mediation analysis with multiple mediators. *Epidemiologic methods* **2** (1), 95–115 (2014) .

[143] Walker, G. T. Correlation in seasonal variations of weather, VIII: A Preliminary Study of World Weather. *Mem. Indian Meteorol. Dep.* **24** (4), 75–131 (1923) .

[144] Bjerknes, J. Atmospheric teleconnections from the equatorial pacific. *Monthly weather review* **97** (3), 163–172 (1969) .

[145] Lau, K.-M. & Yang, S. in *Walker Circulation* (ed.Holton, C. J. A. . P. J. A., J. R.) *Encycl. Atmos. Sci.* Ch. X, 2505–2510 (Academic Press, Cambridge, MA, USA, 2003).

[146] Hersbach, H. *et al.* The era5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society* **146** (730), 1999–2049 (2020) .

[147] Bell, B. *et al.* The era5 global reanalysis: Preliminary extension to 1950. *Quarterly Journal of the Royal Meteorological Society* **147** (741), 4186–4227 (2021) .

[148] Gushchina, D., Zheleznova, I., Osipov, A. & Olchev, A. Effect of various types of enso events on moisture conditions in the humid and subhumid tropics. *Atmosphere* **11** (12), 1354 (2020) .

[149] Krich, C. *et al.* Estimating causal networks in biosphere–atmosphere interaction with the pcmci approach. *Biogeosciences* **17** (4), 1033–1061 (2020) .

[150] Pastorello, G. *et al.* The fluxnet2015 dataset and the oneflux processing pipeline for eddy covariance data. *Scientific data* **7** (1), 1–27 (2020) .

[151] Atkin, O. K. & Tjoelker, M. G. Thermal acclimation and the dynamic response of plant respiration to temperature. *Trends in plant science* **8** (7), 343–351 (2003) .

[152] Jia, X. *et al.* Seasonal and interannual variations in ecosystem respiration in relation to temperature, moisture, and productivity in a temperate semi-arid shrubland. *Science of The Total Environment* **709**, 136210 (2020) .

[153] Harris, N. & Drton, M. Pc algorithm for nonparanormal graphical models. *Journal of Machine Learning Research* **14** (11) (2013) .

[154] Liu, H., Lafferty, J. & Wasserman, L. The nonparanormal: Semiparametric estimation of high dimensional undirected graphs. *Journal of Machine Learning Research* **10** (10) (2009) .

[155] Liu, H., Han, F., Yuan, M., Lafferty, J. & Wasserman, L. High-dimensional semiparametric gaussian copula graphical models. *The Annals of Statistics* **40** (4), 2293–2326 (2012) .

[156] Pedregosa, F. *et al.* Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research* **12**, 2825–2830 (2011) .

[157] Montavon, G., Samek, W. & Müller, K.-R. Methods for interpreting and understanding deep neural networks. *Digital signal processing* **73**, 1–15 (2018) .

Acknowledgements. J.R. has received funding from the European Research Council (ERC) Starting Grant CausalEarth under the European Union's Horizon 2020 research and innovation program (Grant Agreement No. 948112), from the European Union's Horizon 2020 research and innovation programme under grant agreement No 101003469 (XAIDA), from the European Union's Horizon 2020 research and innovation programme under Marie Skłodowska-Curie grant agreement No 860100 (IMIRACLI), and by the Helmholtz AI project CausalFlood (grant no. ZT-I-PF-5-11). G.C.V. and V.E. acknowledge funding from the European Research Council (ERC) Synergy Grant USMILE under the European Union's Horizon 2020 research and innovation program (grant agreement 855187), and G.C.V. additionally thanks the support of the Fundación BBVA with the project 'Causal inference in the human-biosphere coupled system (SCALE)'. The authors thank R. Herman, M. Reichstein, E. Bareinboim, E. Galytska, and S. Karmouche for helpful comments.

Author contributions

J.R. envisaged and drafted the outline of the paper. J.R and G.C.-V. wrote the Introduction, Section "Foundations of causal inference" was written by A.G and G.C.-V., Section "Framing causal research questions" by J.R., Section "Causal discovery" by A.G., J.R., and G.V., Section "Causal effect estimation" by A.G., Section "Example case studies" by J.R., and the Outlook by J.R., G.C.-V., and V.E. All authors discussed the content and contributed to editing the manuscript. All data analyses, Python scripts, Table 1, and all figures were created by J.R., Tab. 2 by J.R., A.G., and G.V.

Competing interests

The authors declare no competing interests.

Publisher's note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

This version of the article has been accepted for publication, after peer review but is not the Version of Record and does not reflect post-acceptance improvements, or any corrections. The Version of Record is available online at: http://dx.doi.org/10.1038/s43017-023-00431-y. Use of this Accepted Version is subject to the publisher's Accepted Manuscript terms of use https://www.springernature.com/gp/open-research/policies/accepted-manuscript-terms.

Data availability and links to software packages

The original data ERA5 reanalysis for the climate example can be downloaded from https://www.ecmwf.int/en/forecasts/datasets/

reanalysis-datasets/era5/ or through the KNMI Climate Explorer (https://climexp.knmi.nl) that we also used to extract regional averages. The case studies can be reproduced with Python code provided at https://github.com/jakobrunge/tigramite/blob/master/tutorials/case_studies/, which also includes the data.

The causal inference software packages mentioned in Tab. 2 are: **TETRAD:** Java package covering constraint-based, score-based, and asymmetry-based causal discovery, as well as causal effect estimation, also for time series, GUI available (https://github.com/cmu-phil/tetrad); **causal-learn:** Python reimplementation and extension of TETRAD, see above (https://github.com/cmu-phil/causal-learn); **pcalg:** R package covering constraint-based, score-based, and asymmetry-based causal discovery, as well as causal effect estimation (https://CRAN.R-project.org/package=pcalg); **Tigramite:** Python package covering constraint-based causal discovery and causal effect estimation methods specifically adapted to time series (https://github.com/jakobrunge/tigramite); **MXM:** R package covering constraint-based causal discovery (https://CRAN.R-project.org/package=MXM); **bnlearn/dbn-learn:** R package covering constraint and score-based causal discovery (https://CRAN.R-project.org/package=bnlearn), also for time series (https://CRAN.R-project.org/package=dbnlearn); **InvariantCausalPrediction:** R package covering (sequential) invariant causal prediction (https://cran.r-project.org/package=seqICP); **rEDM:** R-package for convergent cross mapping (https://cran.r-project.org/web/packages/rEDM/index.html); **Causal-nex:** Python package for continuous-optimization structure learning (https://github.com/quantumblacklabs/causalnex); **Statsmodels:** Python time series modeling package (https://www.statsmodels.org/stable/index.html); **CausalFusion:** Web-based tool with focus on causal effect estimation and full coverage of do-calculus (https://www.causalfusion.net); **dagitty:** Web-based tool for causal effect estimation via adjustment (http://dagitty.net); **econML:** Python package for machine learning-based estimation of causal effects (https://github.com/microsoft/EconML); **PyWhy:** Python package for causal machine learning (https://github.com/py-why); **Linearmodels:** Python modeling package (https://github.com/bashtage/linearmodels).