Quantifying causal pathways of teleconnections

CentAUR

Central Archive at the University of Reading

Reading's research outputs online

Quantifying Causal Pathways of Teleconnections

Marlene Kretschmer, Samantha V. Adams, Alberto Arribas, Rachel Prudden, Niall Robinson, Elena Saggioro, and Theodore G. Shepherd

ABSTRACT: Teleconnections are sources of predictability for regional weather and climate, but the relative contributions of different teleconnections to regional anomalies are usually not understood. While physical knowledge about the involved mechanisms is often available, how to quantify a particular causal pathway from data are usually unclear. Here, we argue for adopting a causal inference-based framework in the statistical analysis of teleconnections to overcome this challenge. A causal approach requires explicitly including expert knowledge in the statistical analysis, which allows one to draw quantitative conclusions. We illustrate some of the key concepts of this theory with concrete examples of well-known atmospheric teleconnections. We further discuss the particular challenges and advantages these imply for climate science and argue that a systematic causal approach to statistical inference should become standard practice in the study of teleconnections.

KEYWORDS: Atmospheric circulation; Teleconnections; Statistical techniques; Time series; Interannual variability; Regional effects

AFFILIATIONS: Kretschmer and Shepherd—Department of Meteorology, University of Reading, Reading, United Kingdom; **Adams**—Informatics Lab, Met Office, Exeter, United Kingdom; **Arribas**—Department of Meteorology, University of Reading, and Microsoft, Reading, United Kingdom; **Prudden**—Informatics Lab, Met Office, and University of Exeter, Exeter, United Kingdom; **Robinson**—Met Office, Exeter, United Kingdom; **Saggioro**—Department of Mathematics and Statistics, University of Reading, Reading, United Kingdom

The term "teleconnection" is used to refer to a recurrent climatic effect resulting from a spatially distant forcing (Wallace and Gutzler 1981). For instance, the phase of El Niño–Southern Oscillation (ENSO) in the tropical Pacific impacts precipitation in California (Swain 2015; Chang et al. 2015), parts of Australia, South Africa, and South America (Iizumi et al. 2014; Dai and Wigley 2000). Several other climate modes such as the Madden–Julian oscillation (MJO), the North Atlantic Oscillation (NAO), the quasi-biennial oscillation (QBO), the Indian Ocean dipole (IOD), and the Pacific decadal oscillation (PDO) have been described, and their interconnections as well as their remote impacts have been extensively studied using observations, climate models and physical theory (Trenberth et al. 1998; Hoskins and Karoly 1981; Wang et al. 2017; Bjerknes 1969; Walker 1925).

Due to their relevance for regional weather and climate, teleconnections remain an extremely active area of research. One key task is to quantify teleconnection strength in both models and observations. For example, understanding potential biases in the strength of teleconnection signals is important to improve their representation in numerical models (Vitart 2017), which is key to improving forecasts on time scales ranging from subseasonal to multidecadal (Mariotti et al. 2020; Lang et al. 2020; López-Parages and Rodríguez-Fonseca 2012). Moreover, given that much of the uncertainty in regional climate projections under global warming is associated with teleconnections (Shepherd 2014), attributing ensemble spreads to changes in large-scale drivers can help to understand and constrain the projected changes (Zappa and Shepherd 2017; Kretschmer et al. 2020; Maraun et al. 2017; Mindlin et al. 2020).

However, robustly estimating the effects of a teleconnection from data remains a challenging task due to the often simultaneous influences of multiple climate modes. For instance, quantifying the causal influence of the stratospheric polar vortex (SPV) on the NAO is difficult, as both the SPV and the NAO are known to be influenced by the MJO, and this influence is likely modulated by the phases of ENSO and the QBO (Barnes et al. 2019; Cassou 2008; Lee et al. 2019). Analyzing a teleconnection pathway in isolation, for example, using pairwise correlation, can therefore lead to wrong inferences about its causal effect (Runge et al. 2014; Kretschmer et al. 2016).

One key problem is that there is no established framework to systematically account for the influence of other teleconnections affecting the pathway in question. While knowledge about the causal mechanisms at play is often available, how to isolate and quantify a particular effect from data are usually unclear. This disconnect between physics and statistics is exemplified by the American Meteorological Society's definition of a teleconnection as a "correlation in the fluctuations of a field at widely separated points," further stating that "such correlations suggest that information is propagating between the distant points through the atmosphere."[1] Climate scientists are well aware that correlation does not necessarily imply causation, but how to overcome this mantra in statistical practice and connect the two perspectives in a quantitative manner is unclear. The difference between correlation and causation becomes crucial when one considers out-of-sample use of the statistical relationships, such as understanding the influences of model biases (Bracegirdle and Stephenson 2012; Kretschmer et al. 2020), storylines of regional climate

[1] https://glossary.ametsoc.org/wiki/Teleconnection

change (Shepherd 2019), or unprecedented events (Diffenbaugh et al. 2017). Moreover, it is important just for understanding the relative role of different causal factors, which is a typical goal in teleconnection studies (Junge and Stephenson 2003; Jiménez-Esteve and Domeisen 2018; Barnes et al. 2019). In short, a process-based framework to quantify the causal teleconnection pathways apparent in correlations is sorely needed.

Here we advocate for a formal causal framework in the statistical analysis of teleconnections, which can be obtained by grounding it in causal inference theory (see sidebar). Statistical analysis in weather and climate science is usually done in the context of physical, hence causal, reasoning, but this reasoning is often only informal. The reasoning can be formalized by expressing expert knowledge about physical mechanisms in the form of a causal network. This has several advantages. First, it forces the researcher to be explicit about their

Causal inference theory in a nutshell

Despite a clear physical conception of causality, a mathematical formalization has long been missing, and only emerged in the last few decades (Pearl and Mackenzie 2018). These major methodological advances are already successfully applied in many research disciplines such as epidemiology (Greenland et al. 1999), psychology (Rohrer 2018), and medical research (Richens et al. 2020), but there are only a few examples from climate science in the context of weather attribution (Hannart et al. 2016) and model assessment (Hirt et al. 2020).

In causal statistics, a causal influence from a process, represented by the random variable X, to another process, represented by the random variable Y, means that intervening in X while keeping everything else fixed changes the probability distribution of Y (Pearl 2000; Pearl et al. 2016). Mathematically, such (usually only hypothetical) interventions are described with the so-called do-operator. The *interventional* conditional probability, denoted by $P(Y|do(X))$, generally does not coincide with the *observational* conditional probability of Y given X, denoted by $P(Y|X)$. For example, the measured pressure by a barometer (X) and the actual pressure (Y) have a strong statistical association and observed values of X will also be good predictors of Y. However, as X does not cause Y, intervening in X, e.g., by moving the needle of the barometer by hand to $X = x$, will not change the surrounding pressure, and thus $P(Y|do(X) = x) \neq P(Y|X = x)$. In contrast, interventions in the pressure will lead to a change in the barometer needle.

Causal inference theory shows that quantifying causal effects to predict the effects of interventions purely based on observed data and without doing any actual experiments or interventions is sometimes possible. In other words, it can be possible to extract the desired interventional probability from the observed probabilities. The underlying idea is that past (naturally occurring) interventions in X that led to changes in Y are present in the data but are biased by other processes that affect both X and Y. To isolate the causal effect from X to Y, one thus has to account for the influence of such confounders.

A necessary requirement for causal inference is to first define a plausible causal model of the hypothesized data-generating mechanisms, usually expressed graphically in the form of a causal network. Note that it is not necessary (nor would it be possible) to represent the full climate system in such a network. Instead, the network represents a reduced model of the truth, tailored to the purpose at hand. If one is for instance interested in the causal effect of X on Y, only those processes that could confound the analysis, i.e., common drivers of X and Y, have to be included. Representing assumed physical relationships in the form of a causal network has several advantages. First, it makes it easier for others to grasp the assumptions of the analyst and follow their argument. Second, such a network transforms expert knowledge into mathematical objects to which the established rules of probability theory apply. This makes it easy to understand how causal information flows along the links in the network. In particular, identifying the confounding factors that one needs to control for to extract a particular causal effect from data follows directly from the network structure.

For some graphical intuition, one can think of the links in a network as pipes which allow the flow of information between the nodes. Each causal network consists of combinations of "chains" ($X \rightarrow Z \rightarrow Y$), "forks" ($Y \leftarrow Z \rightarrow X$), and "collider" structures ($X \rightarrow Z \leftarrow Y$). While the information flows along the links of chains and forks, which is to say that statistical association (i.e., correlation) of X and Y is present, it is "blocked" by the common effect Z in a collider structure, implying statistical independence of X and Y (i.e., no correlation). Once one controls for the variable Z in the first two cases (i.e., the mediator in a chain or the common driver in a fork), which is the same as blocking the information flow, X and Y become independent conditional on Z. In contrast, controlling for the common effect Z in a collider structure "opens" the otherwise blocked path from X to Y and introduces a statistical association between X and Y conditional on Z.

Thus, to quantify a particular causal pathway in the network one has to control for the *correct* processes. While it is necessary to block the effect of a common driver, it can lead to a bias if done for a common effect or an indirect pathway. In many cases, one can identify the correct adjustment set at a glance or by following relatively simple rules [see, e.g., Cinelli et al. (2020) for a summary overview of good and bad adjustment sets in networks]. For more complex setups, one can draw on a comprehensive mathematical theory, providing rules of when and how it is possible to extract a causal effect from data (Pearl 2000; Pearl et al. 2016).

assumptions, which makes it easier for others to follow their argument. Second, it allows to understand how causal information is propagating and where noncausal correlation is expected. Finally, how to remove the influence of common drivers to extract a particular causal effect from data follows directly from the network structure. These advantages can be gained by moving from reasoning informally with causal narratives, to reasoning formally with causal networks. While this only involves small changes in statistical practice, it can lead to significant differences in the framing of the problem and in the interpretation and useability of the results. As Harold Jeffreys noted in his seminal work on the theory of probability, "It is sometimes considered a paradox that the answer depends not only on the observations but on the question; it should be a platitude" (Jeffreys 1961).

The purpose of this paper is not to *detect* causal relationships (which is a different issue) but to show how existing knowledge about physical mechanisms can be used to *quantify* teleconnection pathways. We illustrate this with a number of well-known teleconnection examples. Mostly, we do this using multiple linear regression (MLR), but all the concepts extend naturally into the nonlinear context, as we illustrate with our final example. Finally, we discuss particular opportunities and some practical challenges for the use of a causal framework in climate science.

Data

In all examples, we use monthly NCEP reanalyses[2] covering 1949–2019 (Kalnay et al. 1996). Time series are constructed by area averaging over different regions, variables, and time bins. All time series are standardized by removing the multiyear seasonal mean and dividing by the multiyear standard deviation, and detrended by removing the multiyear linear fit slope. We recognize that NCEP reanalysis data have their limitations, but they serve our purpose of illustrating some of the key principles and methods of causal inference theory with well-known examples of teleconnections. For the same reason, we do not clutter the text with confidence intervals and *p* values, and rely on previous literature for establishing the physical relevance of our examples.

[2] https://psl.noaa.gov/data/gridded/data.ncep.reanalysis.html

In example 1, we consider summer-mean data (June–August) of precipitation in Denmark (DK; 50°–60°N, 2°–15°E) and the Mediterranean (MED; 36°–41°N, 10°–30°E), and an index of the NAO provided by NOAA.[3] In example 2, we consider winter-mean data (December–February) and follow Chang et al. (2015) to calculate an index of precipitation over California (CA) and of the North Pacific jet stream (Jet), the latter based on daily differences of regionally averaged sea level pressure. The El Niño 3.4 index (ENSO) is from NOAA. In example 3, austral spring/summer data (October–December) of ENSO is considered, and we follow Byrne et al. (2019) to calculate a Southern Hemisphere (SH) jet index based on zonal-mean zonal wind data (Jet; 55°–65°S) at 850 hPa, and an SH stratospheric polar vortex breakdown index (SPV), expressed in the number of days following 1 October of each season (Black and McDaniel 2007). In example 4, late-autumn averages (October–December) of Barents and Kara sea ice concentrations (BK; 65°–85°N, 10°–100°E), together with sea level pressure over the North Pacific (NP; 30°–65°N, 160°–220°E) and the Ural Mountains region (URAL; 45°–70°N, 40°–85°E) are considered. Zonal-mean wind velocities at 10 hPa during winter (January–March) are averaged to construct an index of the Northern Hemisphere SPV (60°–75°N). In example 5, we consider ENSO and precipitation in Australia (AU; 12°–38°S, 117°–152°E) during austral spring (September–November). The IOD is calculated as the difference between sea surface temperature (SST) anomalies in the western (10°S–10°N, 50°–70°E) and eastern (10°S –0°, 90°–110°E) tropical Indian Ocean.

[3] https://psl.noaa.gov/data/climateindices/list/

Causal reasoning in the statistical analysis of teleconnections

Teleconnections are commonly analyzed using MLR, a simple and powerful tool to quantify linear dependencies. However, regression coefficients can easily be misinterpreted, or not fully exploited, if the underlying data-generating mechanisms are not taken into account. In contrast, when MLR is combined with physical reasoning, causal conclusions are possible, as we discuss below in more detail.

In this context, causal networks[4] provide a simple graphical tool to facilitate the process-based analysis of teleconnections (Pearl 2009, 2013). Causal networks consist of nodes, representing the physical variables involved in teleconnections (e.g., ENSO, the NAO), and links, indicating the presence and direction of the assumed causal relationship between these variables. As a starting point, causal networks can simply be thought of as a very intuitive way to qualitatively outline a set of physical hypotheses, something that is already widely used in climate science in an informal manner. Schematic overviews summarizing the key findings of a paper are, for instance, often presented in this way (Jiménez-Esteve and Domeisen 2018; Lee et al. 2019).

[4] There are different ways to refer to causal networks in the literature, but the most common is directed acyclic graph (DAG). "Graph" is the mathematical term for network, "directed" means the links between nodes have a direction, and "acyclic" means that no causal loops are permitted (also see "Particular challenges" section). Here, we stick to the more physical term of a causal network.

Throughout this paper, we use causal networks to present the teleconnections we aim to quantify. We first consider a few simple examples involving three variables.

Example 1: Common drivers. Precipitation in Denmark and in the Mediterranean region in summer are negatively correlated ($r = -0.24$). Climate scientists would generally agree that this observed association does not reflect a direct causal relationship between the two features (and thus would not be called a teleconnection) but rather stems from both regions being affected by the position of the North Atlantic storm track as, e.g., described by the summertime NAO index (Folland et al. 2009; Bladé et al. 2012). In other words, the NAO is a common driver of summer precipitation in DK and MED. This assumption, as illustrated in Fig. 1, has implications for understanding the origin and strength of the correlation between DK and MED.

First, it can be tested if the hypothesis of no direct causal relationship between DK and MED is consistent with the observed data. To test this, assuming linear dependence, we calculate the partial correlation of DK and MED conditioned on the common driver NAO. That is, we regress out the effect of the NAO from DK and MED separately and then correlate the residuals. Indeed, the partial correlation of DK and MED conditioned on NAO reduces to 0.01, which we interpret as indistinguishable from zero, meaning that their correlation is entirely explained by the effect of the NAO, consistent with and validating our assumption.

Second, our physical knowledge (or hypotheses) can be harnessed to quantify the causal effects of the NAO on DK and MED. Here we make the further (reasonable) assumption that our system is complete for this purpose, i.e., that there are no further common drivers acting on both NAO and MED or NAO and DK. We can then estimate the causal effect by regressing DK on NAO, giving a standardized causal effect of −0.58; i.e.,

$$DK = -0.58NAO + \varepsilon,$$

Fig. 1. Causal network showing the hypothesized causal influence of the summertime North Atlantic Oscillation (NAO) on precipitation in Denmark (DK) and the Mediterranean (MED).

with ε denoting noise. This means that a change in NAO by one standard deviation implies a decrease in DK by 0.58 standard deviations. In the same way, we can estimate the effect of NAO on MED, which is found to be 0.42. These numbers can be interpreted as the average *causal* effect of the NAO on summer precipitation in these regions, under the assumption of the causal network structure and linear dependence between variables.

Note that the product of the causal links ($-0.58 \times 0.42 = -0.25$) approximately coincides with the correlation of -0.24 between DK and MED. This is a practical property of linear models (with standardized variables) called the path-tracing rule[5] (Pearl 2013). This rule follows from simple algebra and expresses how statistical associations (correlations) reflect the underlying causal effects.

While the above example may seem oversimplistic, it shows how scientific knowledge guides the data analysis. A causal interpretation of regression and correlation coefficients is only justified if one has hypotheses of the underlying causal mechanisms. These physical hypotheses can be tested explicitly (as we did above using partial correlation), where possible, and should be updated in case they are *not* supported by the data.

[5] It says that the correlation of any two variables in a network is the sum of products of the causal effects along all "open" paths connecting the two. The term "open path" is explained in the sidebar on causal inference theory.

Example 2: Mediating pathways. Figure 2 schematically summarizes a large body of evidence of how a Rossby wave train excited by SST anomalies in the tropical Pacific (ENSO) affects the position of the North Pacific jet stream (Jet) and thereby wintertime precipitation in CA (Chang et al. 2015; Horel and Wallace 1981; Trenberth et al. 1998; Hoskins and Karoly 1981).

Suppose we are interested in quantifying the effects of ENSO on CA and of Jet on CA. One naive way to do so would be to use MLR to quantify their contributions, giving

$$CA = 0.05ENSO + 0.79Jet + \varepsilon.$$

However, the very small regression coefficient of ENSO cannot be interpreted as its causal effect on CA. Rather, it shows that the effect of ENSO on CA is almost fully mediated via Jet. In the linear case, controlling for a variable is the same as including it in the regression model. Thus, by including Jet in the model above, we have blocked (or regressed out) the very causal pathway from ENSO to CA (via Jet) that we intended to measure. In contrast, the regression coefficient of Jet of 0.79 can be interpreted as the causal effect of Jet on CA, provided there are no additional common drivers of Jet and CA (as assumed in our causal model). Including ENSO in the regression was not needed to quantify the contribution of Jet on CA, but neither did it do any harm (Cinelli et al. 2020), as ENSO does not confound the relationship between Jet and CA.

Correctly quantifying the causal effect of ENSO on CA could be done by simply regressing CA on ENSO, giving a causal effect strength of 0.34 (i.e., CA = 0.34ENSO + ε'). Alternatively, and perhaps more informatively, one can separately estimate the strength of the individual links along the pathway, and then multiply them to obtain the strength of the full pathway. Here we have an effect of ENSO on Jet of 0.37 (i.e., Jet = 0.37ENSO + ε'') and of Jet on CA of 0.81 (i.e., CA = 0.81Jet + ε'''). The product along the pathway (0.37 $\times$

Fig. 2. Causal network showing the hypothesized mediated causal influence of El Niño–Southern Oscillation (ENSO) on winter precipitation in California (CA) via the position of the North Pacific jet stream (Jet).

0.81 = 0.30) is approximately the same as estimating the effect of ENSO on CA directly (0.34), as expected from the path-tracing rule.

This simple example illustrates that conditioning on a "mediator," here Jet, controls away the effect that one might actually aim to measure. It is often the practice in climate science, especially in the context of statistical predictions, to include different climate indices in a regression model to predict some regional target variable. While this may be unproblematic for purely predictive, within-sample purposes, it can lead to spurious interpretations regarding the individual contributions of different drivers, as illustrated here.

It is worth noting that from a statistical perspective, examples 1 and 2 are indistinguishable. They both involve a target variable Y (DK, say, and CA), and two potential explanatory variables X (MED and ENSO) and Z (NAO and Jet), each of which are correlated with Y. Regressing Y on both X and Z indicates a conditional independence between Y and X, showing that the information pathway between X and Y (reflected in their correlation) is indirect, passing through Z. However, the physical interpretation of the pathway depends entirely on the assumed direction of the causal relation between X and Z, which is opposite in the two examples ($X \leftarrow Z \rightarrow Y$ in example 1 and $X \rightarrow Z \rightarrow Y$ in example 2). Yet this crucial feature of the analysis is completely invisible in the purely statistical formulation of MLR.

Example 3: Direct and indirect pathways. The following example is a combination of the previously discussed mediator and common driver effects. We again consider ENSO which is known to influence the SH jet (Jet) in early austral summer (L'Heureux and Thompson 2006). However, ENSO also affects the timing of the breakdown of the Southern Hemispheric polar vortex (SPV) which in turn impacts the jet position (Domeisen et al. 2019; Saggioro and Shepherd 2019; Byrne et al. 2019). In summary, ENSO influences Jet directly, via a tropospheric pathway, and indirectly via SPV, also called the stratospheric pathway (see Fig. 3).

As there are no assumed common drivers of ENSO and Jet, the *total* effect of ENSO on Jet can be quantified by regressing Jet on ENSO, giving a standardized effect strength of −0.14. The individual strengths of the two causal pathways can be estimated separately. To quantify the *direct* (tropospheric) pathway, one has to block the effect mediated via SPV; i.e., one needs to control for SPV. Expressed in a linear regression model, this gives

$$\text{Jet} = -0.04\text{ENSO} + 0.39\text{SPV} + \varepsilon,$$

meaning that the direct effect of ENSO is estimated to be only −0.04. For the *indirect* (stratospheric) pathway, the link from ENSO to SPV can be calculated without further conditioning, i.e., by regressing SPV on ENSO directly, which gives an average effect strength of −0.26. However, to quantify the contribution from SPV to Jet, one has to control for the common driver ENSO, thus include ENSO in the regression; this yields a causal effect of SPV on Jet of 0.39, as already found above.

Put together, it follows that the strength of the indirect, stratospheric pathway is −0.10 (=−0.26 × 0.39) while that of the direct, tropospheric pathway is only −0.04; i.e., the effect of ENSO on the Jet via the stratosphere
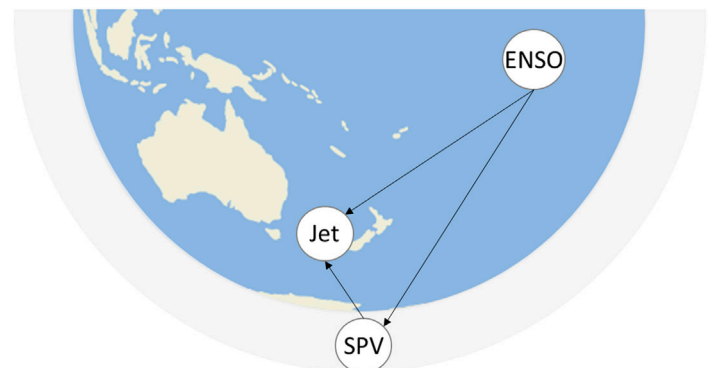

Fig. 3. Causal network showing the hypothesized direct effect of ENSO on the position of the Southern Hemisphere jet stream (Jet), and its indirect causal influence mediated via the late-spring breakdown of the stratospheric polar vortex (SPV).

is (for the data and time period analyzed here) twice as large as the tropospheric link, which is consistent with recent findings (Byrne et al. 2017, 2019). We again note that the sum of the tropospheric and stratospheric pathways (−0.10 − 0.04 = −0.14) is, as expected from the path-tracing rule, approximately equal to the total effect of ENSO on Jet as calculated above.

While quantifying direct and indirect effects between correlated processes through the path-tracing rule is widely appreciated in other fields of science, e.g., psychology (Montoya and Hayes 2017), it has not yet permeated climate science (an exception is Junge and Stephenson 2003) where correlation between explanatory variables is generally regarded as a problem for the application of MLR (von Storch and Zwiers 1999; Wilks 2011). Making a causal hypothesis on the physical basis for the correlation resolves this statistical indeterminism. As with examples 1 and 2, here the physical interpretation of the correlation between ENSO and Jet depends on the assumed direction of causal influence between ENSO and SPV. Thus, instead of having to add a caveat that "correlation does not imply causation," which makes the result ambiguous, we can embed the statistical analysis within an expert-based causal framing and thereby make the numbers useable. For example, Saggioro and Shepherd (2019) showed that the observed delay in SPV breakdown in the last decades of the twentieth century, which is attributed to the development of the ozone hole, well predicts the observed poleward shift in Jet. This makes sense as a nonstationary forcing of SPV, which is unrelated to ENSO, can be expected to induce a nonstationarity in Jet through the same causal effect of SPV on Jet found in the stationary regression model.

Example 4: Blocking the correct paths in the network. The advantages of explicitly defining a causal network become most apparent when there are more than just a few variables and processes involved, such that it quickly becomes confusing and difficult to understand how statistical association is inherited from the causal relations. This is illustrated in the next example.

In recent years, sea ice loss in the Barents and Kara Sea region in autumn (BK) has been suggested to cause a weakening of the wintertime Northern Hemisphere SPV (Kim et al. 2014). This remains a controversial hypothesis, partly due to inconsistent model results (Screen et al. 2018; Cohen et al. 2020; Kretschmer et al. 2020). Quantifying the causal effect from BK on SPV from observed data is, however, challenging, as several potential common drivers have to be taken into account (see Fig. 4). For example, both ENSO and the MJO can affect the SPV via altered pressure anomalies over the NP, which can also lead to changes in BK via a Rossby wave train connecting the Pacific with the Atlantic (Jiménez-Esteve and Domeisen 2018; Warner et al. 2020). Moreover, sea level pressure variability over URAL has been shown to affect BK (Tyrlis et al. 2019) but is also a well-documented driver of SPV variability (Kretschmer et al. 2018b). The role of URAL is particularly tricky to disentangle as the hypothesized teleconnection from BK to SPV is proposed to be mediated via URAL (Kim et al. 2014; Kretschmer et al. 2016).

Representing these different mechanisms in a causal network not only considerably simplifies the description of the data analysis but also helps to understand which factors
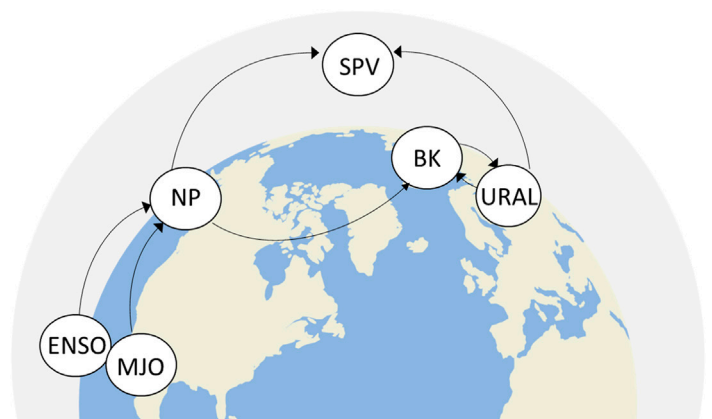

Fig. 4. Causal network showing the hypothesized pathway from Barents and Kara sea ice in autumn (BK) to SPV in winter, through effects on sea level pressure over the Ural Mountains region (URAL). The latter is also assumed to affect BK. Moreover, ENSO and the Madden–Julian oscillation (MJO) influence North Pacific sea level pressure (NP), and thereby both the SPV and BK.

need to be controlled for in order to isolate the causal effect from BK to SPV. Importantly, it is not necessary to control for all common drivers. Instead, according to causal inference theory (see sidebar), one needs only to "block" the confounding pathways in the network linking BK and SPV other than the *causal* pathway (linking BK to SPV via URAL) that we aim to measure. From the network it is clear that the common drivers of BK and SPV, namely, ENSO and MJO, are both mediated via NP. Thus, conditioning on NP is sufficient to block not only the direct influence of NP, but also the indirect influence of ENSO and MJO via NP. Accounting for the influence of URAL is more difficult, as it is assumed to be both a common driver (i.e., we need to block its influence), *and* a mediator (i.e., blocking its path would regress out the effect we aim to measure). Here we make use of the causal assumption that the effect comes after the cause. To then block the confounding role of URAL without blocking its role as mediator, we condition on URAL during the same autumn months as BK, assuming that its mediating role involves some longer time lag. In summary, and assuming linear dependencies, our regression model to quantify the causal effect of autumn BK on winter SPV is

$$SPV_{JFM} = aBK_{OND} + bNP_{OND} + cURAL_{OND} + \varepsilon,$$

giving an effect of $a = 0.21$. Importantly, the regression coefficients of NP ($b = 0.27$) and URAL ($c = 0.00$) cannot be interpreted as their causal effects on SPV. Instead, their causal effects would have to be estimated in different (appropriate) regression models, which would follow from the network structure. This again emphasizes the importance of a causal framing of MLR, as otherwise the regression coefficients are physically uninterpretable [see also Kretschmer et al. (2020), which addresses the causal role of BK in future SPV changes].

We stress once more that the correctness of the estimate of the causal effect of BK on SPV is (apart from sampling uncertainty) conditional on the causal network, our assumption of linear dependence, and the choice of data. For example, researchers could disagree with the assumed links in the network or they might consider different relevant time periods and lags. While all of these aspects are nontrivial and could be addressed with different sensitivity tests, it is important to note that the same concerns hold for *any* type of statistical analysis. The central point here is that if one wishes to have plausible estimates of *causal* effects, then one needs to account for the *correct* confounding processes and this implies making the causal assumptions explicit. Indeed, unless this is done, differences between studies can arise from differences in assumptions that are only implicit and invisible to a purely statistical analysis.

Example 5: Measuring nonlinear dependencies. So far, we have only quantified causal links using MLR, which is adequate when describing linear relationships. A common approach to estimate nonlinear dependencies is through calculating conditional probabilities of different subcategories of the involved variables, often shown in contingency tables (Barnes et al. 2019; Lee et al. 2019). However, just as for MLR, the purely statistical relationships are not physically interpretable unless the conditional probabilities are embedded within a causal framework. We illustrate this through our final example.

ENSO is a key driver of Australian precipitation (AU) in austral spring (Maher and Sherwood 2014; Cai et al. 2011) and exhibits its influence on AU through various pathways (see also example 3). In particular, ENSO influences the IOD, another important driver of AU (Black et al. 2003). Figure 5 summarizes these causal assumptions.

The influence of ENSO on the IOD, and thereby on AU, has been suggested to exhibit asymmetries in strength, implying that the relationship is nonlinear (Cai and van Rensch 2013; Cai et al. 2012). To quantify the influence of ENSO and IOD on AU, we first stratify ENSO into terciles describing La Niña, neutral, and El Niño states. Similarly, we stratify the IOD into terciles describing its negative, neutral, and positive phases. For AU we differentiate

between the two categories of above and below average values, AU+ and AU−. This choice is convenient because the effect on AU can then be represented by a single number, since the two probabilities add up to unity. The resulting conditional probabilities for above-average AU, i.e., P(AU+|IOD, ENSO), are shown in Table 1a. The strength of the conditional dependence can be read as the deviation of the conditional probabilities from their default value of 50%.
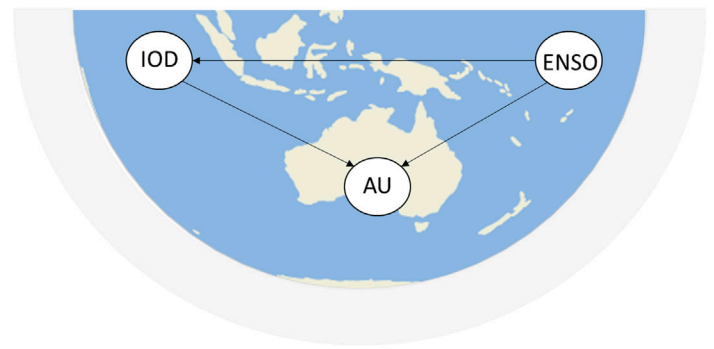

Fig. 5. Causal network showing the hypothesized direct effect of ENSO on spring precipitation in Australia (AU) and its indirect causal influence mediated via the Indian Ocean dipole (IOD).

The joint probabilities of IOD and ENSO, denoted P(IOD, ENSO), are shown through the contingency table in Table 1b. This shows the strong association between the two drivers, with no or almost no occurrences of the combinations El Niño–IOD negative and La Niña–IOD positive, and an enhanced likelihood of the opposite combinations. Together, these two tables provide the complete statistical information of any particular combination of the three indices, written P(AU, IOD, ENSO). For AU+ events it is the product of Tables 1a and 1b, i.e., P(AU+|IOD, ENSO) × P(IOD,ENSO). For AU− events, it is 1 minus the value in Table 1a times the entry in Table 1b, i.e., [1 − P(AU+|IOD, ENSO)] × P(IOD,ENSO). However, these joint probabilities do not themselves provide any causal information. Instead, causality enters in how we read the tables, which depends on our underlying assumptions.

From the marginals in Table 1a we can see an association of both ENSO and IOD with AU, with the negative phases of both indices increasing the probability of above-average rainfall, and the positive phases decreasing it. Since we consider ENSO as a common driver of IOD and AU, we need to control for it in order to isolate the causal IOD–AU relationship. We do this by conditioning the IOD–AU association on the phase of ENSO, which is represented by the columns in Table 1a. The added information provided by IOD, given ENSO, is represented by the *Bayes factor* P(AU|IOD, ENSO)/P(AU|ENSO), which for AU+ is the ratio of the conditional probability to the marginal probability in the bottom entry of the same column. We can see from this that the phase of IOD has barely any effect on AU for either La Niña or El Niño. For example, the Bayes factor for IOD+ during El Niño is 0.24/0.22 = 1.09, meaning that the probability of AU+ during El Niño phases is increased by only 9% if the IOD is positive. When ENSO is neutral there is a suggestion that IOD+ decreases the probability of AU+, while IOD− has a weaker positive effect.

The dependencies in the network further allow us to interpret and decompose the causal effects. The causal effect of ENSO on AU, for instance, can be found by marginalizing over IOD:

$$P\big(\mathrm{AU|ENSO}\big)=\sum_{i=1}^{3}P\big(\mathrm{AU|ENSO,IOD}_i\big)P\big(\mathrm{IOD}_i|\mathrm{ENSO}\big).$$

Here i = 1, 2, 3 represent the three phases of IOD, and the equation applies for any combination of AU and ENSO phase. In the linear case the calculation reduces to the path-tracing rule, with regression coefficients replacing Bayes factors. The first factor in the product can be read off Table 1a, while the second factor is the ratio of the corresponding entry in Table 1b divided by its marginal, which is 0.33. For example, the effect of La Niña on AU+ is quantified as

$$(0.83)\frac{(0.17)}{(0.33)}+(0.80)\frac{(0.14)}{(0.33)}+(1.0)\frac{(0.03)}{(0.33)},$$

which recovers the value of 0.83 in the marginal dependence in Table 1a.

Note that if we reversed the causal assumption between ENSO and IOD, Table 1 would be entirely unchanged. Yet our reading of Table 1 would be completely different, as we would then infer that IOD affects AU through both a direct and an indirect pathway, and that the Bayes factor from ENSO, given IOD (determined by reading Table 1 row-wise rather than column-wise) is quite considerable. In this alternative reading of Table 1, the causal attribution of AU would be more equally shared between IOD and ENSO.

Table 1. (a) Conditional probabilities for above-average AU, i.e., P(AU+|IOD, ENSO). The rightmost column and bottom row represent the marginal probabilities, P(AU+|IOD) and P(AU+|ENSO), respectively. The default probability for above-average AU is P(AU+) = 1/2. (b) Joint probabilities for IOD and ENSO phases, P(IOD, ENSO). The default probability is 1/9.

(a)	La Niña	Neutral	El Niño	Marginal
IOD negative	0.83	0.50	—	0.67
Neutral	0.80	0.43	0.17	0.52
IOD positive	1.0	0.25	0.24	0.30
Marginal	0.83	0.43	0.22	0.50
(b)				
IOD negative	0.17	0.17	0.00	
Neutral	0.14	0.10	0.09	
IOD positive	0.03	0.06	0.24	

In this simple case, the difference in interpretation of the conditional probabilities can be explained in words. For more complex sets of interacting teleconnections, as in example 4, where two-dimensional contingency tables are no longer feasible, it is necessary to be more explicit and rigorous about how the causal assumptions enter into the interpretation of the numbers, otherwise fallacious inferences may result.

The general point we wish to make with this example is that a causal network framework is a fully nonparametric approach. The choice of method to quantify the causal effect of interest depends entirely on the researcher, bearing in mind the purpose of the calculation and the limitations of the available data. While defining meaningful subcategories of a variable can be a challenge on its own (Falkena et al. 2020), there exist a plethora of methods to quantify nonlinear dependencies. For example, next to transfer entropy (Runge et al. 2012), different machine learning methods have been used to estimate nonlinear causal dependencies (Chernozhukov et al. 2018; Blakely et al. 2021). However, nonlinear methods generally require large datasets, often justifying a linear approach, especially when using the short observational record. The relative benefit of using a nonlinear over a linear approach can be quantified using various metrics such as the Bayesian information criterion.

Opportunities for climate science

Once a climate scientist has developed a theory of the causal relationships of the processes they are considering, by drawing the network explicitly they can use the rules of causal inference to determine which covariates to include and which to exclude for their specific analysis. There are several fundamental aspects where such formal causal reasoning could help to make progress.

First, using causal networks and following the rules of causal inference provides an easy and transparent way to quantify teleconnection pathways, and different hypotheses could be tested in this way. Such a diagnostic approach has immediate benefits for analyzing teleconnections in the observational record and is also particularly suitable for evaluating their representation in climate models. For example, huge ensembles can be needed to detect (often only small) teleconnection signals (Smith et al. 2020). Comparing the causal effects is more efficient (Kretschmer et al. 2020); it can moreover shed light on the dynamical sources of model differences and should help in understanding potential signal-to-noise issues (Scaife and Smith 2018).

Quantifying teleconnections in climate model ensembles can also help to reduce uncertainties of regional weather and climate predictions. On both subseasonal-to-seasonal (S2S)

and multidecadal time scales, for instance, it can enable process-informed bias adjustments (Specq and Batté 2020; Mariotti et al. 2020). In the context of climate projections, causal estimates are, for example, expected to provide more robust emergent constraints on regional weather (Maraun et al. 2017). The causal flow in a network further provides a built-in narrative to communicate the sources of uncertainty. More precisely, uncertainties in the response of a regional climate hazard to anthropogenic global warming can be decomposed into different physically self-consistent probabilistic storylines (Shepherd 2019), providing important information for decision makers. In this context, networks could also easily be expanded by including impact variables, such as wildfires or crop yield in a specific region (Lloyd and Shepherd 2020; Lehmann et al. 2020; Guimarães Nobre et al. 2017). In this way, one can quantify the contribution of climatic effects to these impacts in conjunction or in combination with nonclimatic drivers such as land use, which could also be included in the network.

A causal network approach can also be useful to guide model interventions to test the influence of regional processes. Such interventions are currently used both in S2S predictions (MacLeod et al. 2021) and climate projections (Ceppi and Shepherd 2019) to determine the causality of modeled teleconnections. Defining a causal network in these cases, too, is beneficial in order to check whether the experimental setup is consistent with observational knowledge about causality. In other words, comparing the effects derived from the observations (using causal inference rules) with those obtained by climate model experiments allows for a causality-based cross validation of model and observational data.

Finally, we see large potential for a causal inference-based approach to connect classical atmospheric dynamics with novel machine learning techniques (Bareinboim and Pearl 2016; Hernán et al. 2019; Pearl 2019). Both communities often lack a common language, making successful collaborations difficult (Knüsel et al. 2019). To allow for a better physical interpretability of findings from deep learning (e.g., novel extracted climate features), causality-based approaches could, for example, be used to evaluate them against expert knowledge. In turn, deep learning methods could also be used to quantify known causal pathways, potentially providing a more powerful way to estimate nonlinear dependencies from large spatiotemporal climate data (Luo et al. 2020; Ham et al. 2019). Overall, causal reasoning can help build trust in purely data-based findings and is key to physics-guided machine learning (Reichstein et al. 2019; Knüsel et al. 2019).

Particular challenges

The various temporal scales of dependencies in the climate system can be difficult to address. Note that loops and cycles are generally not permitted in a causal network. This might seem contradictory at first given a fully coupled climate system including strong autodependencies. However, depending on context, time lags and different time scales of expected cause–effect relationships can resolve this issue to a reasonable degree (Kretschmer et al. 2020, 2016). In our examples, we restricted ourselves to seasonal-mean data to avoid statistical issues arising from autodependence (McGraw and Barnes 2018; Runge et al. 2014), and also ignored potential effects of interannual memory. Note, however, that time information can also be incorporated in the network where necessary. For example, the "past" of a process can be included as an extra node, with autodependence reflected by a link from the past to present.

The question arises of how one sets up a causal network in the first place, as detecting relevant causal pathways is arguably a key challenge. Causal discovery algorithms (Runge et al. 2019a), such as the PCMCI algorithm (Runge et al. 2019b), which aim at reconstructing the causal interconnections from a given set of time series, can be useful in this context and can complement the data analysis. For example, they can be used in determining the relevant time lags of interactions (Saggioro and Shepherd 2019; Barnes et al. 2019; Kretschmer et al. 2016).

Several successful applications exist in the context of teleconnections (Kretschmer et al. 2016; Di Capua et al. 2020; Siew et al. 2020; Kretschmer et al. 2018a; Samarasinghe et al. 2021; Barnes et al. 2019; Ebert-Uphoff and Deng 2012) and the methodologies, based on principles of causal inference, are being further advanced to deal with more complex settings such as regime-dependent causal dependencies (Saggioro et al. 2020), contemporaneous causal relations (Runge 2020), and latent confounders (Gerhardus and Runge 2021).

Another challenge is that relevant processes should be included in the network, meaning that there exist no confounders. This can never be fulfilled for an object as complex as the climate system. (Of course, the same criticism applies to any statistical analysis.) However, processes not explicitly included can be represented as noise. Consequently, one self-consistency test is to see whether the residuals really look like noise, or instead appear to contain some kind of structure which might suggest the need for another explanatory variable. Also note that it is possible to include unknown drivers in the network and to treat this type of uncertainty explicitly within causal inference theory (Pearl 2009).

While the above concerns have to be carefully taken into account, in the context of teleconnections there is usually enough expertise available to draw a plausible causal model to articulate physical hypotheses. Every climate science paper invokes, in some way, physical hypotheses, as any statistical study design (including the choices of variables, time scales, data preprocessing, and the applied methods) depends on causal assumptions which are often not stated explicitly. Causal networks, in contrast, turn these claims into testable objects, making research conclusions more transparent and traceable.

Summary and conclusions

An improved understanding of teleconnection pathways is crucial to reduce uncertainties about regional climate projections under global warming as well as to improve forecasts beyond the weather time scale. Part of the challenge lies in closing the gap between the physical concept of a teleconnection, which is causal, and its mathematical description, which is correlational.

Here we have argued for adopting causal networks and causal inference theory as a general framework to articulate and quantify teleconnection pathways. While data analyses in climate science generally include the mantra-like statement that correlation does not imply causation, this is usually simply a caveat. However, while correlation may not *imply* causation, it does *reflect* causation; it is just that the causation may be indirect. Extracting the causal effects "hidden" in correlations is possible, but it requires systematic causal reasoning based on the physical knowledge about the involved processes.

Causal inference requires that hypotheses on the physical mechanisms that generated the data are made first, before any conclusions are drawn. Based on the assumptions encoded in the form of a network, the causal effects can then be estimated from the observations by following relatively simple rules. Here we discussed some examples from climate science and quantified the relevant teleconnection pathways in reanalysis data. The provided cases show how causal reasoning should guide the data analysis to obtain more reliable estimates of causal effects. In our view, seeking "objectivity" in data-driven approaches is not necessarily worthwhile, as it requires ignoring physical knowledge, which is usually crucial to achieve meaningful results. While practical challenges of data analysis remain, such as choices of the optimal climate indices, time scales, and data products, this applies to any statistical analysis. We argue that the transparent and deductive nature of causal network analysis can help in overcoming many of the limitations faced in current studies and in reconciling differences between the conclusions of different studies.

Importantly, a causal approach is not meant to compete with traditional climate model experiments or physical theory. Instead, it serves as a scaffold to build scientific intuition

into the statistical analysis of the data. We argue that both basic physics and data science are needed to make progress in climate science and that causal theory is a framework for better reconciling the two.

Data availability statement. The NCEP reanalysis data are made publicly available by NOAA/OAR/ESRL PSL on their website at https://psl.noaa.gov/data/gridded/data.ncep.reanalysis.html and the climate indices of NAO and ENSO can be found at https://psl.noaa.gov/data/climateindices/list/. The Python scripts of the discussed examples are available at https://github.com/informatics-lab/causality/tree/master/causality_paper.

References

Bareinboim, E., and J. Pearl, 2016: Causal inference and the data-fusion problem. *Proc. Natl. Acad. Sci. USA*, **113**, 7345–7352, https://doi.org/10.1073/pnas.1510507113.

Barnes, E. A., S. M. Samarasinghe, I. Ebert-Uphoff, and J. C. Furtado, 2019: Tropospheric and stratospheric causal pathways between the MJO and NAO. *J. Geophys. Res. Atmos.*, **124**, 9356–9371, https://doi.org/10.1029/2019JD031024.

Bjerknes, J., 1969: Atmospheric teleconnections from the equatorial Pacific. *Mon. Wea. Rev.*, **97**, 163–172, https://doi.org/10.1175/1520-0493(1969)097<0163:ATFTEP>2.3.CO;2.

Black, E., J. M. Slingo, and K. R. Sperber, 2003: An observational study of the relationship between excessively strong short rains in coastal East Africa and Indian Ocean SST. *Mon. Wea. Rev.*, **131**, 74–94, https://doi.org/10.1175/1520-0493(2003)131%3C0074:AOSOTR%3E2.0.CO;2.

Black, R. X., and B. A. McDaniel, 2007: Interannual variability in the Southern Hemisphere circulation organized by stratospheric final warming events. *J. Atmos. Sci.*, **64**, 2968–2974, https://doi.org/10.1175/JAS3979.1.

Bladé, I., B. Liebmann, D. Fortuny, and G. J. van Oldenborgh, 2012: Observed and simulated impacts of the summer NAO in Europe: Implications for projected drying in the Mediterranean region. *Climate Dyn.*, **39**, 709–727, https://doi.org/10.1007/s00382-011-1195-x.

Blakely, T., J. Lynch, K. Simons, R. Bentley, and S. Rose, 2021: Reflection on modern methods: When worlds collide—Prediction, machine learning and causal inference. *Int. J. Epidemiol.*, **49**, 2058–2064, https://doi.org/10.1093/ije/dyz132.

Bracegirdle, T. J., and D. B. Stephenson, 2012: On the robustness of emergent constraints used in multimodel climate change projections of Arctic warming. *J. Climate*, **26**, 669–678, https://doi.org/10.1175/JCLI-D-12-00537.1.

Byrne, N. J., T. G. Shepherd, T. Woollings, and R. A. Plumb, 2017: Nonstationarity in Southern Hemisphere climate variability associated with the seasonal breakdown of the stratospheric polar vortex. *J. Climate*, **30**, 7125–7139, https://doi.org/10.1175/JCLI-D-17-0097.1.

——, ——, and I. Polichtchouk, 2019: Subseasonal-to-seasonal predictability of the Southern Hemisphere eddy-driven jet during austral spring and early summer. J. *Geophys. Res. Atmos.*, **124**, 2018JD030173, https://doi.org/10.1029/2018JD030173.

Cai, W., and P. van Rensch, 2013: Austral summer teleconnections of Indo-Pacific variability: Their nonlinearity and impacts on Australian climate. *J. Climate*, **26**, 2796–2810, https://doi.org/10.1175/JCLI-D-12-00458.1.

——, ——, T. Cowan, and H. H. Hendon, 2011: Teleconnection pathways of ENSO and the IOD and the mechanisms for impacts on Australian rainfall. *J. Climate*, **24**, 3910–3923, https://doi.org/10.1175/2011JCLI4129.1.

——, ——, ——, and ——, 2012: An asymmetry in the IOD and ENSO teleconnection pathway and its impact on Australian climate. *J. Climate*, **25**, 6318–6329, https://doi.org/10.1175/JCLI-D-11-00501.1.

Cassou, C., 2008: Intraseasonal interaction between the Madden–Julian oscillation and the North Atlantic Oscillation. *Nature*, **455**, 523–527, https://doi.org/10.1038/nature07286.

Ceppi, P., and T. G. Shepherd, 2019: The role of the stratospheric polar vortex for the austral jet response to greenhouse gas forcing. *Geophys. Res. Lett.*, **46**, 6972–6979, https://doi.org/10.1029/2019GL082883.

Chang, E. K. M., C. Zheng, P. Lanigan, A. M. W. Yau, and J. D. Neelin, 2015: Significant modulation of variability and projected change in California winter precipitation by extratropical cyclone activity. *Geophys. Res. Lett.*, **42**, 5983–5991, https://doi.org/10.1002/2015GL064424.

Chernozhukov, V., D. Chetverikov, M. Demirer, E. Duflo, C. Hansen, W. Newey, and J. Robins, 2018: Double/debiased machine learning for treatment and structural parameters. *Econom. J.*, **21**, C1–C68, https://doi.org/10.1111/ectj.12097.

Cinelli, C., A. Forney, and J. Pearl, 2020: A crash course in good and bad controls. Social Science Research Network, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3689437.

Cohen, J., and Coauthors, 2020: Divergent consensuses on Arctic amplification influence on midlatitude severe winter weather. *Nat. Climate Change*, **10**, 20–29, https://doi.org/10.1038/s41558-019-0662-y.

Dai, A., and T. M. L. Wigley, 2000: Global patterns of ENSO-induced precipitation. *Geophys. Res. Lett.*, **27**, 1283–1286, https://doi.org/10.1029/1999GL011140.

Di Capua, G., M. Kretschmer, R. V. Donner, B. van den Hurk, R. Vellore, R. Krishnan, and D. Coumou, 2020: Tropical and mid-latitude teleconnections interacting with the Indian summer monsoon rainfall: A theory-guided causal effect network approach. *Earth Syst. Dyn.*, **11**, 17–34, https://doi.org/10.5194/esd-11-17-2020.

Diffenbaugh, N. S., and Coauthors, 2017: Quantifying the influence of global warming on unprecedented extreme climate events. *Proc. Natl. Acad. Sci. USA*, **114**, 4881–4886, https://doi.org/10.1073/pnas.1618082114.

Domeisen, D. I. V., C. I. Garfinkel, and A. H. Butler, 2019: The teleconnection of El Niño Southern Oscillation to the stratosphere. *Rev. Geophys.*, **57**, 5–47, https://doi.org/10.1029/2018RG000596.

Ebert-Uphoff, I., and Y. Deng, 2012: Causal discovery for climate research using graphical models. *J. Climate*, **25**, 5648–5665, https://doi.org/10.1175/JCLI-D-11-00387.1.

Falkena, S. K. J., J. Wiljes, A. Weisheimer, and T. G. Shepherd, 2020: Revisiting the identification of wintertime atmospheric circulation regimes in the Euro-Atlantic sector. *Quart. J. Roy. Meteor. Soc.*, **146**, 2801–2814, https://doi.org/10.1002/qj.3818.

Folland, C. K., J. Knight, H. W. Linderholm, D. Fereday, S. Ineson, and J. W. Hurrell, 2009: The summer North Atlantic Oscillation: Past, present, and future. *J. Climate*, **22**, 1082–1103, https://doi.org/10.1175/2008JCLI2459.1.

Gerhardus, A., and J. Runge, 2021: High-recall causal discovery for autocorrelated time series with latent confounders. *34th Conf. on Neural Information Processing Systems*, Virtual, NeurIPS, https://proceedings.neurips.cc/paper/2020/file/94e70705efae423efda1088614128d0b-Paper.pdf.

Greenland, S., J. Pearl, and J. Robins, 1999: Causal diagrams for epidemiologic research. *Epidemiology*, **10**, 37–48, https://doi.org/10.1097/00001648-199901000-00008.

Guimarães Nobre, G., B. Jongman, J. Aerts, and P. J. Ward, 2017: The role of climate variability in extreme floods in Europe. *Environ. Res. Lett.*, **12**, 084012, https://doi.org/10.1088/1748-9326/aa7c22.

Ham, Y.-G., J.-H. Kim, and J.-J. Luo, 2019: Deep learning for multi-year ENSO forecasts. *Nature*, **573**, 568–572, https://doi.org/10.1038/s41586-019-1559-7.

Hannart, A., J. Pearl, F. E. L. Otto, P. Naveau, and M. Ghil, 2016: Causal counterfactual theory for the attribution of weather and climate-related events. *Bull. Amer. Meteor. Soc.*, **97**, 99–110, https://doi.org/10.1175/BAMS-D-14-00034.1.

Hernán, M. A., J. Hsu, and B. Healy, 2019: A second chance to get causal inference right: A classification of data science tasks. *Chance*, **32**, 42–49, https://doi.org/10.1080/09332480.2019.1579578.

Hirt, M., G. C. Craig, S. A. K. Schäfer, J. Savre, and R. Heinze, 2020: Cold-pool-driven convective initiation: Using causal graph analysis to determine what convection-permitting models are missing. *Quart. J. Roy. Meteor. Soc.*, **146**, 2205–2227, https://doi.org/10.1002/qj.3788.

Horel, J. D., and J. M. Wallace, 1981: Planetary-scale atmospheric phenomena associated with the Southern Oscillation. *Mon. Wea. Rev.*, **109**, 813–829, https://doi.org/10.1175/1520-0493(1981)109<0813:PSAPAW>2.0.CO;2.

Hoskins, B. J., and D. J. Karoly, 1981: The steady linear response of a spherical atmosphere to thermal and orographic forcing. *J. Atmos. Sci.*, **38**, 1179–1196, https://doi.org/10.1175/1520-0469(1981)038<1179:TSLROA>2.0.CO;2.

Iizumi, T., J. J. Luo, A. J. Challinor, G. Sakurai, M. Yokozawa, H. Sakuma, M. E. Brown, and T. Yamagata, 2014: Impacts of El Niño Southern Oscillation on the global yields of major crops. *Nat. Commun.*, **5**, 3712, https://doi.org/10.1038/ncomms4712.

Jeffreys, H., 1961: *Theory of Probability.* 3rd ed. Oxford University Press, 447 pp.

Jiménez-Esteve, B., and D. I. V. Domeisen, 2018: The tropospheric pathway of the ENSO–North Atlantic teleconnection. *J. Climate*, **31**, 4563–4584, https://doi.org/10.1175/JCLI-D-17-0716.1.

Junge, M. M., and D. B. Stephenson, 2003: Mediated and direct effects of the North Atlantic Ocean on winter temperatures in northwest Europe. *Int. J. Climatol.*, **23**, 245–261, https://doi.org/10.1002/joc.867.

Kalnay, E., and Coauthors, 1996: The NCEP/NCAR 40-Year Reanalysis Project. *Bull. Amer. Meteor. Soc.*, **77**, 437–471, https://doi.org/10.1175/1520-0477(1996)077<0437:TNYRP>2.0.CO;2.

Kim, B.-M., S.-W. Son, S.-K. Min, J.-H. Jeong, S.-J. Kim, X. Zhang, T. Shim, and J.-H. Yoon, 2014: Weakening of the stratospheric polar vortex by Arctic sea-ice loss. *Nat. Commun.*, **5**, 4646, https://doi.org/10.1038/ncomms5646.

Knüsel, B., M. Zumwald, C. Baumberger, G. Hirsch Hadorn, E. M. Fischer, D. N. Bresch, and R. Knutti, 2019: Applying big data beyond small problems in climate research. *Nat. Climate Change*, **9**, 196–202, https://doi.org/10.1038/s41558-019-0404-1.

Kretschmer, M., D. Coumou, J. F. Donges, and J. Runge, 2016: Using causal effect networks to analyze different Arctic drivers of midlatitude winter circulation. *J. Climate*, **29**, 4069–4081, https://doi.org/10.1175/JCLI-D-15-0654.1.

——, J. Cohen, V. Matthias, J. Runge, and D. Coumou, 2018a: The different stratospheric influence on cold-extremes in Eurasia and North America. *npj Climate Atmos. Sci.*, **1**, 44, https://doi.org/10.1038/s41612-018-0054-4.

——, D. Coumou, L. Agel, M. Barlow, E. Tziperman, and J. Cohen, 2018b: More-persistent weak stratospheric polar vortex states linked to cold extremes. *Bull. Amer. Meteor. Soc.*, **99**, 49–60, https://doi.org/10.1175/BAMS-D-16-0259.1.

——, G. Zappa, and T. G. Shepherd, 2020: The role of Barents and Kara sea ice loss in projected polar vortex changes. *Wea. Climate Dyn.*, **1**, 715–730, https://doi.org/10.5194/wcd-1-715-2020.

Lang, A. L., K. Pegion, and E. A. Barnes, 2020: Introduction to special collection: "Bridging weather and climate: Subseasonal-to-seasonal (S2S) prediction." *J. Geophys. Res. Atmos.*, **125**, e2019JD031833, https://doi.org/10.1029/2019JD031833.

Lee, R. W., S. J. Woolnough, A. J. Charlton-Perez, and F. Vitart, 2019: ENSO modulation of MJO teleconnections to the North Atlantic and Europe. *Geophys. Res. Lett.*, **46**, 13 535–13 545, https://doi.org/10.1029/2019GL084683.

Lehmann, J., M. Kretschmer, B. Schauberger, and F. Wechsung, 2020: Seasonal forecast of Moroccan wheat yields based on causal interactions in the climate system. *Geophys. Res. Lett.*, **47**, e2020GL087516, https://doi.org/10.1029/2020GL087516.

L'Heureux, M. L., and D. W. J. Thompson, 2006: Observed relationships between the El Niño–Southern Oscillation and the extratropical zonal-mean circulation. *J. Climate*, **19**, 276–287, https://doi.org/10.1175/JCLI3617.1.

Lloyd, E. A., and T. G. Shepherd, 2020: Environmental catastrophes, climate change, and attribution. *Ann. N. Y. Acad. Sci.*, **1469**, 105–124, https://doi.org/10.1111/nyas.14308.

López-Parages, J., and B. Rodríguez-Fonseca, 2012: Multidecadal modulation of El Niño influence on the Euro-Mediterranean rainfall. *Geophys. Res. Lett.*, **39**, L02704, https://doi.org/10.1029/2011GL050049.

Luo, Y., J. Peng, and J. Ma, 2020: When causal inference meets deep learning. *Nat. Mach. Intell.*, **2**, 426–427, https://doi.org/10.1038/s42256-020-0218-x.

MacLeod, D., R. Graham, C. O'Reilly, G. Otieno, and M. Todd, 2021: Causal pathways linking different flavours of ENSO with the Greater Horn of Africa short rains. *Atmos. Sci. Lett.*, **22**, e1015, https://doi.org/10.1002/asl.1015.

Maher, P., and S. C. Sherwood, 2014: Disentangling the multiple sources of large-scale variability in Australian wintertime precipitation. *J. Climate,* **27**, 6377–6392, https://doi.org/10.1175/JCLI-D-13-00659.1

Maraun, D., and Coauthors, 2017: Towards process-informed bias correction of climate change simulations. *Nat. Climate Change*, **7**, 764–773, https://doi.org/10.1038/nclimate3418.

Mariotti, A., and Coauthors, 2020: Windows of opportunity for skillful forecasts subseasonal to seasonal and beyond. *Bull. Amer. Meteor. Soc.*, **101**, E608–E625, https://doi.org/10.1175/BAMS-D-18-0326.1.

M. C. McGraw, and E. A. Barnes, 2018: Memory matters: A case for Granger causality in climate variability studies. *J. Climate*, **31**, 3289–3300, https://doi.org/10.1175/JCLI-D-17-0334.1.

Mindlin, J., T. G. Shepherd, C. S. Vera, M. Osman, G. Zappa, R. W. Lee, and K. I. Hodges, 2020: Storyline description of Southern Hemisphere midlatitude circulation and precipitation response to greenhouse gas forcing. *Climate Dyn.*, **54**, 4399–4421, https://doi.org/10.1007/s00382-020-05234-1.

Montoya, A. K., and A. F. Hayes, 2017: Two-condition within-participant statistical mediation analysis: A path-analytic framework. *Psychol. Methods*, **22**, 6–27, https://doi.org/10.1037/met0000086.

Pearl, J., 2000: *Causality: Models, Reasoning and Inference.* 2nd ed. Cambridge University Press, 478 pp.

——, 2009: Causal inference in statistics: An overview. *Stat. Surv.*, **3**, 96–146, https://doi.org/10.1214/09-SS057.

——, 2013: Linear models: A useful "microscope" for causal analysis. *J. Causal Inference*, **1**, 155–170, https://doi.org/10.1515/jci-2013-0003.

——, 2019: The seven tools of causal inference, with reflections on machine learning. *Commun. ACM*, **62**, 54–60, https://doi.org/10.1145/3241036.

——, and D. Mackenzie, 2018: *The Book of Why: The New Science of Cause and Effect*. Basic Books, 432 pp.

——, M. Glymour, and N. P. Jewell, 2016: *Causal Inference in Statistics: A Primer.* Wiley, 160 pp.

Reichstein, M., G. Camps-Valls, B. Stevens, M. Jung, J. Denzler, and N. Carvalhais, 2019: Deep learning and process understanding for data-driven Earth system science. *Nature*, **566**, 195–204, https://doi.org/10.1038/s41586-019-0912-1.

Richens, J. G., C. M. Lee, and S. Johri, 2020: Improving the accuracy of medical diagnosis with causal machine learning. *Nat. Commun.*, **11**, 3923, https://doi.org/10.1038/s41467-020-17419-7.

Rohrer, J. M., 2018: Thinking clearly about correlations and causation: Graphical causal models for observational data. *Adv. Methods Pract. Psychol. Sci.*, **1**, 27–42, https://doi.org/10.1177/2515245917745629.

Runge, J., 2020: Discovering contemporaneous and lagged causal relations in autocorrelated nonlinear time series datasets. *Proc. 36th Conf. on Uncertainty in Artificial Intelligence*, Virtual, PMLR, 1388–1397, http://proceedings.mlr.press/v124/runge20a.html.

——, J. Heitzig, V. Petoukhov, and J. Kurths, 2012: Escaping the curse of dimensionality in estimating multivariate transfer entropy. *Phys. Rev. Lett.*, **108**, 258701, https://doi.org/10.1103/PhysRevLett.108.258701.

——, V. Petoukhov, and J. Kurths, 2014: Quantifying the strength and delay of climatic interactions: The ambiguities of cross correlation and a novel measure based on graphical models. *J. Climate*, **27**, 720–739, https://doi.org/10.1175/JCLI-D-13-00159.1.

——, and Coauthors, 2019a: Inferring causation from time series in Earth system sciences. *Nat. Commun.*, **10**, 2553, https://doi.org/10.1038/s41467-019-10105-3.

——, P. Nowack, M. Kretschmer, S. Flaxman, and D. Sejdinovic, 2019b: Detecting and quantifying causal associations in large nonlinear time series datasets. *Sci. Adv.*, **5**, eaau4996, https://doi.org/10.1126/sciadv.aau4996.

Saggioro, E., and T. G. Shepherd, 2019: Quantifying the timescale and strength of Southern Hemisphere intraseasonal stratosphere-troposphere coupling. *Geophys. Res. Lett.*, **46**, 13 479–13 487, https://doi.org/10.1029/2019GL084763.

——, J. de Wiljes, M. Kretschmer, and J. Runge, 2020: Reconstructing regime-dependent causal relationships from observational time series. *Chaos*, **30**, 113115, https://doi.org/10.1063/5.0020538.

Samarasinghe, S. M., C. Connolly, E. A. Barnes, I. Ebert-Uphoff, and L. Sun, 2021: Strengthened causal connections Between the MJO and the North Atlantic with climate warming. *Geophys. Res. Lett.*, **48**, e2020GL091168, https://doi.org/10.1029/2020GL091168.

Scaife, A. A., and D. Smith, 2018: A signal-to-noise paradox in climate science. *npj Climate Atmos. Sci.*, **1**, 28, https://doi.org/10.1038/s41612-018-0038-4.

Screen, J. A., and Coauthors, 2018: Consistency and discrepancy in the atmospheric response to Arctic sea-ice loss across climate models. *Nat. Geosci.*, **11**, 155–163, https://doi.org/10.1038/s41561-018-0059-y.

Shepherd, T. G., 2014: Atmospheric circulation as a source of uncertainty in climate change projections. *Nat. Geosci.*, **7**, 6–11, https://doi.org/10.1038/ngeo2253.

——, 2019: Storyline approach to the construction of regional climate change information. *Proc. Roy. Soc.*, **475A**, 20190013, https://doi.org/10.1098/rspa.2019.0013.

Siew, P. Y. F., C. Li, S. P. Sobolowski, and M. P. King, 2020: Intermittency of Arctic–mid-latitude teleconnections: Stratospheric pathway between autumn sea ice and the winter North Atlantic Oscillation. *Wea. Climate Dyn.*, **1**, 261–275, https://doi.org/10.5194/wcd-1-261-2020.

Smith, D. M., and Coauthors, 2020: North Atlantic climate far more predictable than models imply. *Nature*, **583**, 796–800, https://doi.org/10.1038/s41586-020-2525-0.

Specq, D., and L. Batté, 2020: Improving subseasonal precipitation forecasts through a statistical–dynamical approach: Application to the southwest tropical Pacific. *Climate Dyn.*, **55**, 1913–1927, https://doi.org/10.1007/s00382-020-05355-7.

Swain, D. L., 2015: A tale of two California droughts: Lessons amidst record warmth and dryness in a region of complex physical and human geography. *Geophys. Res. Lett.*, **42**, 9999–10 003, https://doi.org/10.1002/2015GL066628.

Trenberth, K. E., G. W. Branstator, D. Karoly, A. Kumar, N.-C. Lau, and C. Ropelewski, 1998: Progress during TOGA in understanding and modeling global teleconnections associated with tropical sea surface temperatures. *J. Geophys. Res.*, **103**, 14 291–14 324, https://doi.org/10.1029/97JC01444.

Tyrlis, E., E. Manzini, J. Bader, J. Ukita, H. Nakamura, and D. Matei, 2019: Ural blocking driving extreme Arctic sea ice loss, cold Eurasia, and stratospheric vortex weakening in autumn and early winter 2016–2017. *J. Geophys. Res. Atmos.*, **124**, 11 313–11 329, https://doi.org/10.1029/2019JD031085.

Vitart, F., 2017: Madden–Julian oscillation prediction and teleconnections in the S2S database. *Quart. J. Roy. Meteor. Soc.*, **143**, 2210–2220, https://doi.org/10.1002/qj.3079.

von Storch, H., and F. W. Zwiers, 1999: *Statistical Analysis in Climate Research.* Cambridge University Press, 499 pp.

Walker, G. T., 1925: Correlation in seasonal variations of weather—A further study of world weather. *Mon. Wea. Rev.*, **53**, 252–254, https://doi.org/10.1175/1520-0493(1925)53<252:CISVOW>2.0.CO;2.

Wallace, J. M., and D. S. Gutzler, 1981: Teleconnections in the geopotential height field during the Northern Hemisphere winter. *Mon. Wea. Rev.*, **109**, 784–812, https://doi.org/10.1175/1520-0493(1981)109<0784:TITGHF>2.0.CO;2.

Wang, C., C. Deser, J.-Y. Yu, P. DiNezio, and A. Clement, 2017: El Niño and Southern Oscillation (ENSO): A review. *Coral Reefs of the Eastern Tropical Pacific*, P. W. Glynn, D. P. Manzello, and I. C. Enochs, Eds., Springer, 85–106.

Warner, J. L., J. A. Screen, and A. A. Scaife, 2020: Links between Barents-Kara sea ice and the extratropical atmospheric circulation explained by internal variability and tropical forcing. *Geophys. Res. Lett.*, **47**, e2019GL085679, https://doi.org/10.1029/2019GL085679.

Wilks, D., 2011: *Statistical Methods in the Atmospheric Sciences.* 3rd ed. Elsevier, 676 pp.

Zappa, G., and T. G. Shepherd, 2017: Storylines of atmospheric circulation change for European regional climate impact assessment. *J. Climate*, **30**, 6561–6577, https://doi.org/10.1175/JCLI-D-16-0807.1.