Memory Matters: A Case for Granger Causality in Climate Variability Studies

MARIE C. MCGRAW AND ELIZABETH A. BARNES

Department of Atmospheric Science, Colorado State University, Fort Collins, Colorado

(Manuscript received 19 May 2017, in final form 8 January 2018)

ABSTRACT

In climate variability studies, lagged linear regression is frequently used to infer causality. While lagged linear regression analysis can often provide valuable information about causal relationships, lagged regression is also susceptible to overreporting significant relationships when one or more of the variables has substantial memory (autocorrelation). Granger causality analysis takes into account the memory of the data and is therefore not susceptible to this issue. A simple Monte Carlo example highlights the advantages of Granger causality, compared to traditional lagged linear regression analysis in situations with one or more highly autocorrelated variables. Differences between the two approaches are further explored in two illustrative examples applicable to large-scale climate variability studies. Given that Granger causality is straightforward to calculate, Granger causality analysis may be preferable to traditional lagged regression analysis when one or more datasets has large memory.

1. Introduction

The establishment of cause and effect is a fundamental, if elusive, driver of climate science research. While causality is much sought after, it is challenging to establish, especially in observations—recall the adage ''correlation does not equal causation.'' Determining true causality requires not only the establishment of a relationship between two variables, but also the far more difficult task of determining a direction of causality. Although they do not provide information regarding directionality, correlation-based methods, such as lagged linear regression, remain popular and useful tools for identifying lagged relationships between climate variables.

A lagged regression model can provide a straightforward assessment of spatial and temporal variability. Lagged regression analysis has been a popular technique in climate science for nearly 100 years (e.g., Walker 1923, 1924). Since 1988, the phrases ''lagged regression,'' ''lag regression,'' ''lagged correlation,'' and ''lag correlation'' have appeared in a combined total of over 800 manuscripts in the *Journal of Climate* alone. Lagged linear regression analysis has been used in a wide variety of climate science applications, including, but not limited to, stratosphere–troposphere interactions (e.g., Polvani and Waugh 2004); tropical variability patterns, such as the Madden–Julian oscillation and El Niño–Southern Oscillation (e.g., Klein et al. 1999; Hendon et al. 2007); Arctic sea ice extent (e.g., Blanchard-Wrigglesworth et al. 2011); and sea surface temperature variability (e.g., Yu et al. 2010). This is just a small sampling of the hundreds of studies across atmospheric and climate science that utilize linear lagged regression analysis.

While lagged regression can be a straightforward and effective tool for identifying covarying patterns in space and time, lagged regression also has its drawbacks. First, while lagged regression can show the existence of instantaneous and lagged relationships between variables, lagged regression alone cannot indicate the direction of causality. Lagged regression may indicate that two variables are related to each other when in actuality they are linked or driven by a third variable (e.g., Fig. 3 in Kretschmer et al. 2016). Finally, lagged regression can be interpreted to suggest that one variable causes a response in the other when in fact it does not. This can occur when one variable has high memory, or autocorrelation (e.g., Runge et al. 2014; Kretschmer et al. 2016), and this is the scenario that will be explored here.

As an example, consider the relationship between tropical Pacific sea surface temperatures [i.e., El Niño–Southern Oscillation (ENSO)], and surface temperature over North and South America. ENSO is considered to

Corresponding author: Marie C. McGraw, mmcgraw@atmos.colostate.edu

DOI: 10.1175/JCLI-D-17-0334.1

be a primary driver of surface temperature anomalies in these regions (e.g., Ropelewski and Halpert 1986; Gu and Adler 2011). However, on monthly time scales, SST anomalies are quite persistent—the 1-month lag autocorrelation of the Niño-3.4 SST index (anomaly form, with the 1951–2000 mean removed; Rayner et al. 2003) is 0.91, meaning that over 80% of the variability in tropical Pacific SST in the Niño-3.4 region is determined by the previous month. The Niño-3.4 index takes over 6 months to decorrelate (defined using its *e*-folding time). This memory in ENSO can lead to ambiguity when applying lagged linear regression. For example, Fig. 1 shows the lagged relationship between ENSO and land surface temperature T_s [obtained from the NOAA–CIRES Twentieth Century Reanalysis project (Compo et al. 2011), with the mean and second-order trend removed] over the Americas. Figure 1a displays the regression of ENSO on T_s at lags of up to 7 months—that is, the red shading in Fig. 1a indicates grid points for which there is a significant lagged relationship between T_s and ENSO up to 7 months prior (refer to section 2 for details on determining a significant lagged relationship). However, when the regression is performed in the opposite direction—that is, assessing the influence of lagged T_s upon ENSO—Fig. 1b is nearly identical to Fig. 1a. One could interpret Fig. 1b as demonstrating that T_s is driving ENSO up to 7 months in advance, even though it is generally agreed that ENSO drives T_s at these time scales.

Decades of research on ENSO and its impact on surface temperature over the Americas point to ENSO driving surface temperature, not the other way around (e.g., Ropelewski and Halpert 1986; Gu and Adler 2011). However, that conclusion is not clear from Fig. 1—the lagged regression results are ambiguous. One potential cause of this ambiguity could be the high autocorrelation in the Niño-3.4 index. Instead of asking, "Can we use T_s to predict ENSO?" we are better off asking, "Does T_s help us predict ENSO beyond ENSO's ability to predict itself?" We propose the use of Granger causality (Granger 1969) to answer this question and to address the issue of causality in data with nonzero memory. Granger causality analysis consists of a lagged autoregression (e.g., a lagged regression of ENSO on itself), compared to a lagged multiple linear regression (e.g., a lagged regression of T_s and ENSO on ENSO), and is only slightly more challenging to implement than a typical lagged regression analysis. As Granger causality accounts for memory in the data by using a lagged autoregression, it is not susceptible to overreporting of causal relationships with high-memory data, as lagged regression can be. We note that while formal definitions of causality exist as defined by Pearl's causal theory (Pearl 2009) and have been more recently introduced into climate science (e.g., Hannart et al. 2016), here, we loosely define a "causal relationship" as one that shows a significant lagged relationship between variables. The distinction between Pearl causality and Granger causality is discussed further in section 5. It is worth noting that like lagged regression, Granger causality could have difficulty in situations in which there are strong two-way feedbacks occurring on similar time scales; Granger causality is also not applicable in situations in which some additional process not included in the model is driving the modeled processes of interest.

In this paper, we aim to demonstrate the following:

1) Granger causality is typically superior to traditional lagged regression when one or more datasets has substantial memory;
2) Granger causality and lagged regression tend to yield similar results when there is a true causal relationship; and
3) Granger causality is only slightly more difficult to implement than traditional lagged regression.

In section 2, we discuss Granger causality and lagged linear regression. In section 3, we compare the two methods in a Monte Carlo simulation. In section 4, we apply both lagged regression and Granger causality methods to two examples in climate science and show that Granger causality provides additional insight beyond that of a typical lagged linear regression. We note that others have cautioned against lagged regressions with data that are highly autocorrelated and have proposed the use of causal effect networks and graphical models to overcome these issues (e.g., Runge et al. 2014; Kretschmer et al. 2016). While these methods are certainly valuable tools for the climate community (e.g., Ebert-Uphoff and Deng 2012), they are significantly more complex to implement than the lagged regression analysis they may replace. Thus, our goal here is to present a clear, concise, and compelling case for Granger causality analysis in situations when the data are autocorrelated.

2. Statistical model

Granger causality (Granger 1969) was first developed as a predictive model in economics. More recently, Granger causality has found applications in climate science, such as determining the influence of snow cover and vegetation on surface temperature (e.g., Kaufmann et al. 2003), the impact of sea surface temperature on the North Atlantic Oscillation (e.g., Mosedale et al. 2006) or on Atlantic hurricane strength (e.g., Elsner 2006, 2007),

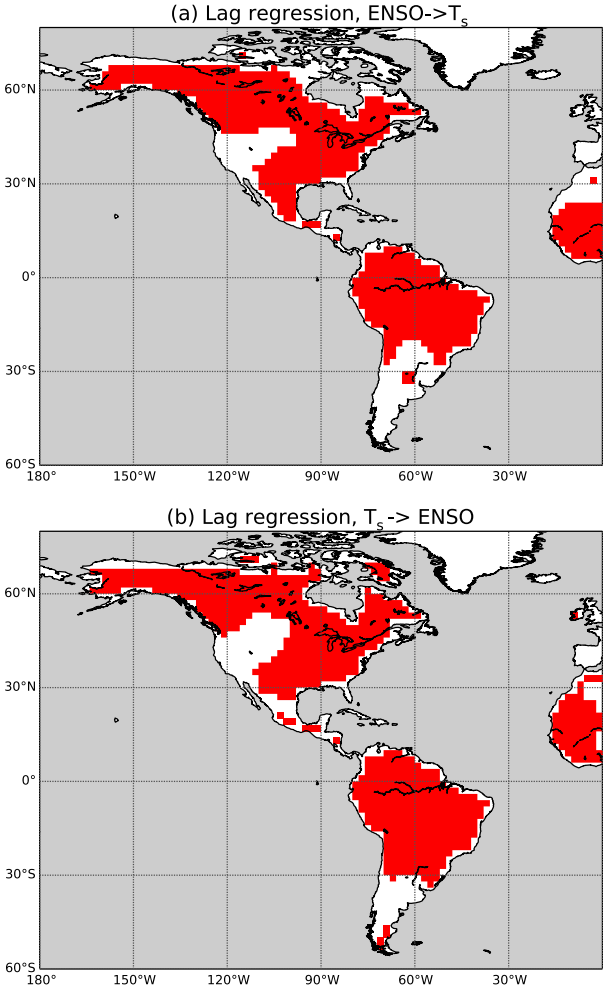

FIG. 1. Using lagged regression to test the hypothesis that (a) ENSO drives T_s and (b) T_s drives ENSO. Red indicates a significant lagged relationship identified at up to 7 months. Significance is assessed at 95% using a two-sided t test.

ENSO's impact on the Indian monsoon (e.g., Mohkov et al. 2011), and attributing global temperature increases to increases in global atmospheric CO_2 (see Attanasio et al. 2013 and references therein). However, the use of Granger causality remains far behind that of lagged regression. We use a Monte Carlo simulation to demonstrate that Granger causality is straightforward and, under specific circumstances, is less likely than lagged regression to lead to the inference of a causal relationship when there is not one.

We start by creating our driver D; D is a first-order autoregressive (AR-1), or red-noise, process, defined as

$$D(t) = \alpha \times D(t-1) + (1-\alpha^2)^{1/2}\varepsilon_D(t), \qquad (1)$$

where α is the lag-1 autocorrelation, ε_D is a random value drawn from a standard normal distribution (i.e., a standard Gaussian random variable), and thus, D has a variance of one. We use D to create a second time series: our response R. By design, R is simply D lagged by some amount of time $\tau > 0$, with added Gaussian noise ε_R:

$$R(t) = D(t-\tau) + \gamma \times \varepsilon_R(t). \qquad (2)$$

When γ is small, there is little additional noise added to R, and the lag-1 autocorrelations of R and D are very similar; large values of γ yield a much noisier R with less memory.

We perform a Monte Carlo simulation in which we vary α, γ, and τ. First, we create a D time series with 550 steps following Eq. (1). After discarding the first 50 values of D, we create R following Eq. (2). We perform our regression analysis (discussed in the next section) and repeat this process 5000 times for each combination of α, γ, and τ. We test 20 values of α, ranging from 0 to 1; 20 values of γ, ranging from 0.005 to 15; and 15 values of τ, ranging from 1 to 15, to ensure that our results are robust.

To evaluate the performance of the statistical model, we first perform a traditional lagged regression, where we use our driver D to predict our response R:

$$R(t) = c_0 + c_1 \times D(t-1) + c_2 \times D(t-2) + \cdots + c_k \times D(t-k), \qquad (3)$$

where k is the maximum lag. The significance of the full model is assessed using a two-sided t test. In all situations, significance is assessed at 95% confidence.

As an alternative to lagged regression, we use Granger causality. Mathematically, establishing Granger causality consists of two regressions: a lagged autoregression of the predictand R,

$$R(t) = c_0 + c_1 \times R(t-1) + c_2 \times R(t-2) + \cdots + c_k \times R(t-k), \qquad (4)$$

and a multiple linear lagged regression including information about both the predictand R, and the predictor (hereafter, D),

$$R(t) = a_0 + a_1 \times R(t-1) + \cdots + a_k \times R(t-k) + b_1 \times D(t-1) + \cdots + b_k \times D(t-k). \qquad (5)$$

The variance explained of R as determined by Eq. (4) is compared to the variance explained of R as determined by Eq. (5). If the multiple linear lagged regression [Eq. (5)] explains significantly more variance in R than the autoregression [Eq. (4)], it is said that D Granger-causes R. Significance is assessed using a two-step process:

1) At least one value of b must be significant according to a two-sided t test.
2) All values of b collectively must increase the variance explained by the regression according to an F test.

For both the standard lagged regression and the Granger causality analysis, we perform the regressions in both directions: the direction we know to be correct (D driving R) and the direction we know to be incorrect (R driving D). In this way, we can evaluate whether or not Granger causality outperforms standard lagged regression, as defined by a lower risk of false detection, given the same identification rate of correct relationships. It is also worth noting that selecting the maximum lag k is an important and potentially challenging part of Granger causality analysis. Typically, k is selected based on a common metric for model selection, such as the Akaike information criterion or the Bayesian, or Schwarz, information criterion (e.g., Mosedale et al. 2006). In both cases, the preferred model is the one with the k value that minimizes the selection criteria and thus limits the model from becoming overfitted. Finally, the approach that we detail here is a relatively straightforward approach to Granger causality that has been used in climate sciences in recent years to great success; it is worth noting, however, that there are alternative ways of calculating Granger causality, many of which have been developed in neuroscience (e.g., Barnett and Seth 2014; Stokes and Purdon 2017).

3. Monte Carlo results

First, we compare the performance of lagged regression and Granger causality by evaluating the ability of D to predict R. Recall that R was created using D, so our models should suggest a causal relationship. Figure 2 shows the percentage of significant results (e.g., the model reports a significant causal relationship for the hypothesis that D drives R at 95%) as a function of memory α (y axis) and noise γ in R (x axis) for the lagged regression model (Fig. 2a) and the Granger causality model (Fig. 2b). Darker colors imply that the model indicated a causal result (in this case, D causes R) more often. Both panels of Fig. 2 look similar—in this case, lagged regression and Granger causality yield comparable results. Both methods show a dependence on γ—that is, as R becomes noisier, both models are less able to predict R from D. Both methods also exhibit minimal dependence on α, demonstrating that in general, both models are quite capable of predicting R, even when D has a very high memory. Here, we note that this lack of dependence on α is specific to the AR-1 process modeled in Eq. (1), where the variance of the noise [the $\varepsilon_D(t)$ term] is standardized. For the more general case where the variance of D is not equal to one, the ability of D to predict R does show a dependence on α, with larger values of α showing an increased ability to correctly identify that D drives R at a given value of γ. This effect occurs for both the lagged linear regression and Granger causality approaches. Thus, even for a more general model of red noise, both methods—lagged regression and Granger causality—yield results that are similar to each other, and either could be used in analyzing the hypothesis that D drives R.

While Fig. 2 demonstrates that lagged regression and Granger causality generally yield similar results in the case of D driving R, there is one notable exception: when memory is very high ($\alpha \geq 0.8$) and noise is moderate ($\gamma > 2$). In this small region, Granger causality exhibits a slightly higher failure rate than lagged regression, as seen by the slight curve near the top of Fig. 2b. This difference between Fig. 2a and Fig. 2b can be explained by the fact that Granger causality evaluates added variance explained—that is, the variance explained beyond what is explained by the autocorrelation of R. If the autocorrelation of D is very high, then R will have a similar autocorrelation and similar values if the noise is moderate. In this case, R has little to add beyond what is already contained in the past values of D, and thus, Granger causality will not indicate a significant causal relationship between R and D, while lagged regression will. It is worth noting that this effect is only seen for a small subset of the Monte Carlo simulations with large memory and moderate noise, that the Granger causality model still confirms the hypothesis that D drives R at a rate of at least 70%, and that this effect is less severe as the sample size is increased. Outside of this small region, lagged regression and Granger causality perform very similarly.

Next, we evaluate lagged regression and Granger causality by using R to predict D; we compare the outcomes of the two methods when we look for causality in the wrong direction (recall that R was created from D). In this case, we would hope that the models do not suggest a causal relationship between R and D. This hypothesis of R driving D is tested in Fig. 3. Figure 3 is laid out similarly to Fig. 2, with darker colors indicating that the model reported a causal relationship more frequently. In Fig. 3, the advantages of Granger causality become apparent. Figure 3a shows that the lagged regression model exhibits a strong dependence on α—as D's memory increases, the lagged regression model is increasingly more likely to suggest that R drives D, which we know to be incorrect. Even at moderate values of α, the lagged regression model implies that there is a causal relationship in the wrong direction. While

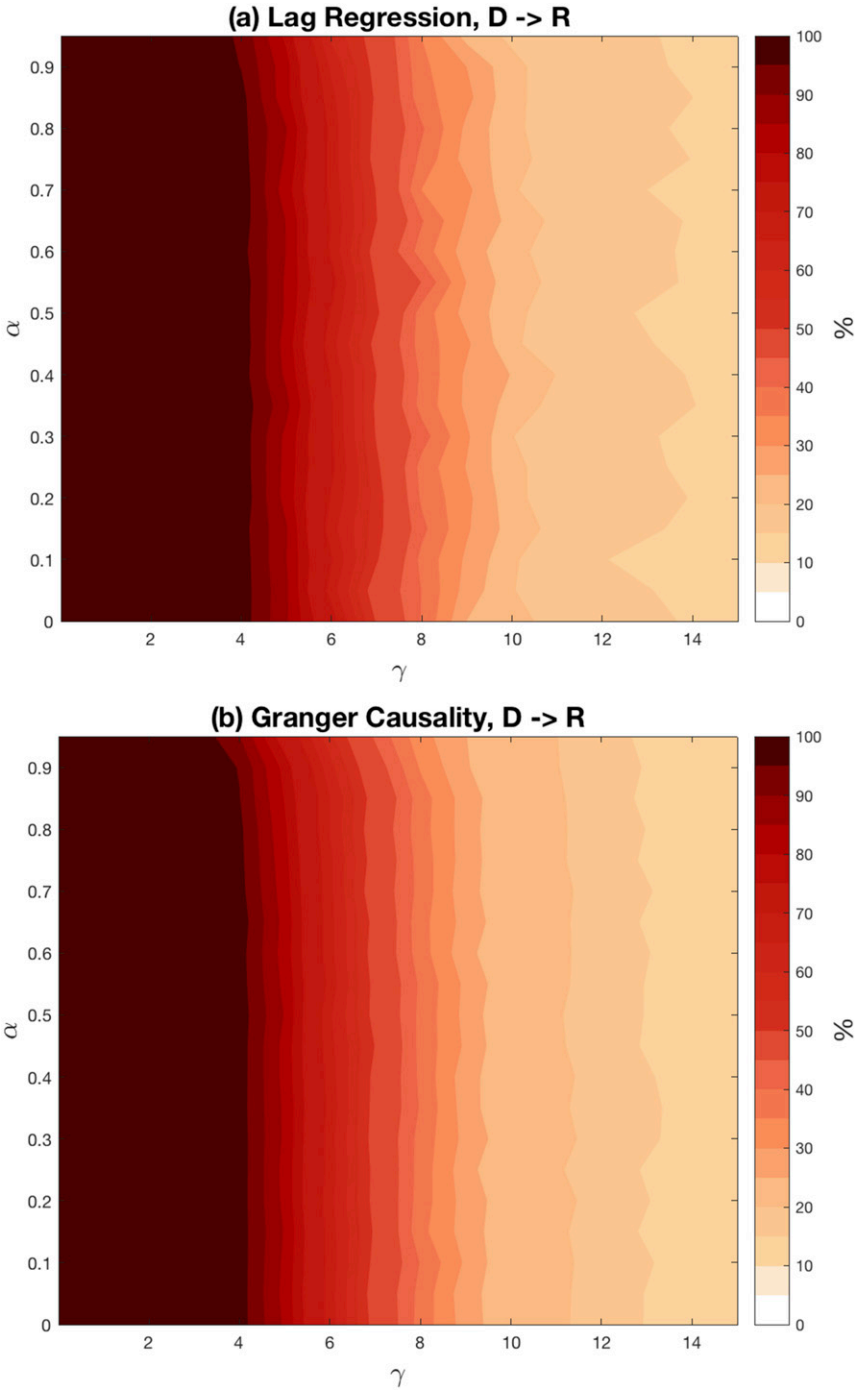

FIG. 2. Testing the hypothesis that D drives R using (a) lagged regression and (b) Granger causality. Shading indicates the percentage of significant results at 95% confidence.

low values of α show a false positive rate between 5% and 10% (recall that significance is assessed at 95% confidence, meaning we would expect a significant result of 5% merely by random chance), at $\alpha = 0.5$, the lagged regression model indicates that R causes D between 10% and 100% of the time, depending on the noisiness of R. For $\alpha \geq 0.8$, this false positive rate is even higher, suggesting 25% of the time that R causes D for even high values of γ. Figure 3a shows only the results for $\tau = 1$ (i.e., a lag of 1 time step), but larger lags are qualitatively similar, although moderate values of α become less sensitive at larger lags (e.g., at $\tau = 3$,

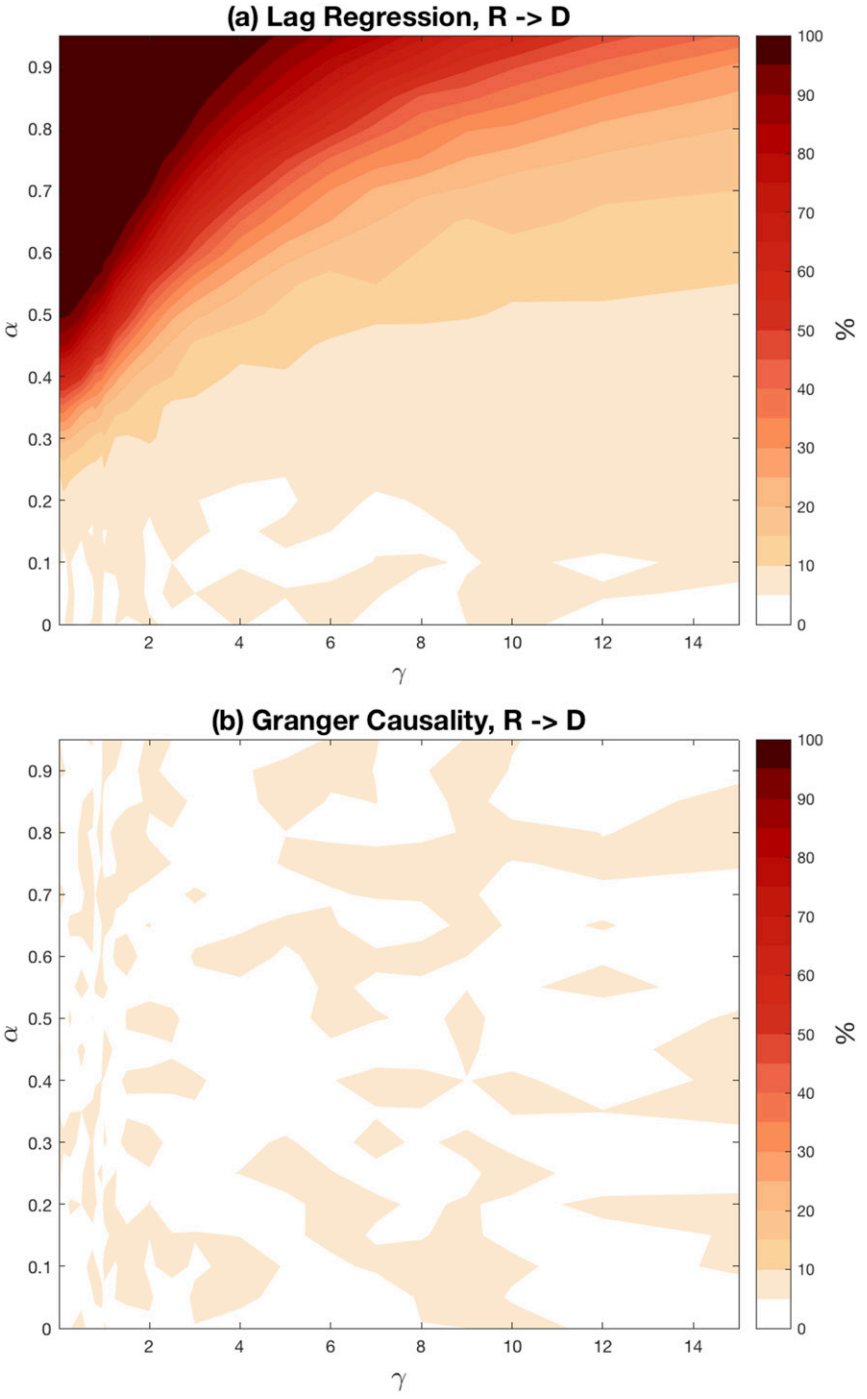

FIG. 3. Testing the hypothesis that R drives D using (a) lagged regression and (b) Granger causality. Shading indicates the percentage of significant results at 95% confidence.

values of $\alpha \leq 0.4$ yield a significant result less than 10% of the time).

There is no such dependence on memory for the Granger causality method, as seen in Fig. 3b. Indeed, Fig. 3b indicates that the results of the Granger causality method are simply noise, with Granger causality yielding a significant result about 5% of the time, consistent with our 95% significance testing. These results are not dependent on lag τ; memory α, or noise γ in R. In this case, Granger causality's insensitivity to α, or memory in D, shows an improvement over a typical lagged regression model for variables with high memory.

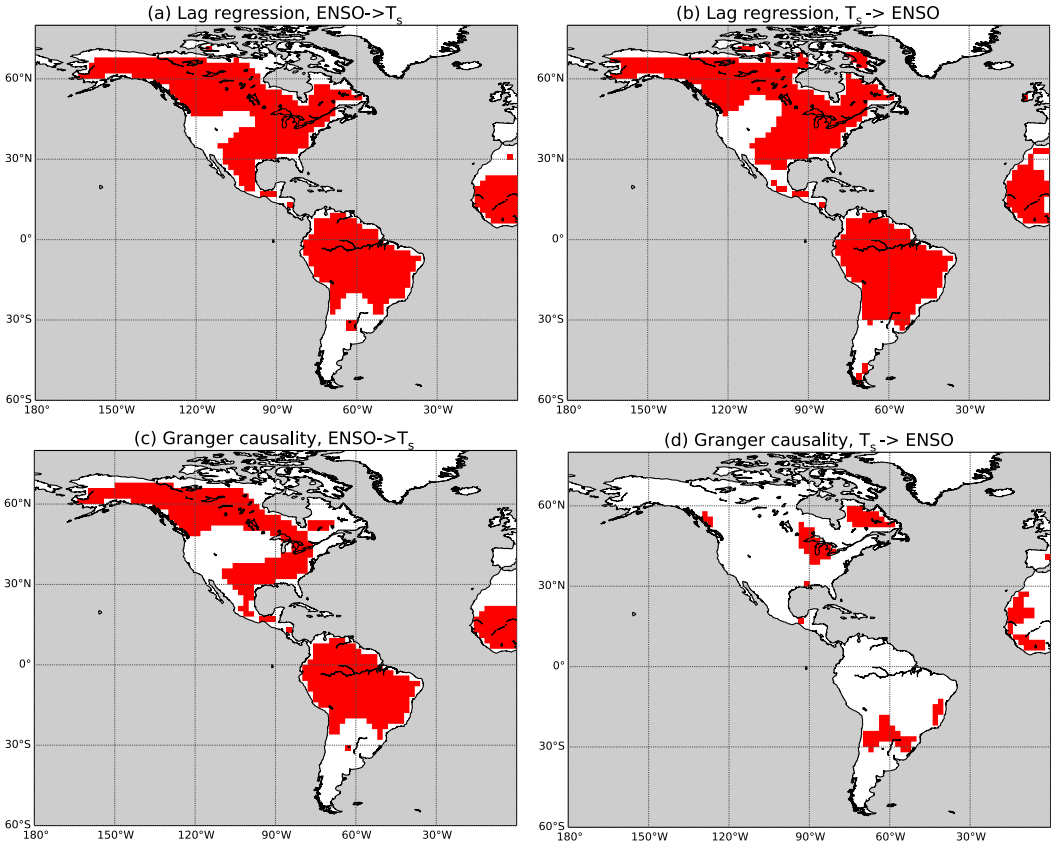

FIG. 4. Using (a),(b) lagged regression and (c),(d) Granger causality to test the hypothesis that (left) ENSO drives T_s and (right) T_s drives ENSO. Red indicates a significant lagged relationship identified at up to 7 months ($k = 7$). Significance is assessed at 95%.

Recall that the 1-month autocorrelation of Niño 3.4 is 0.91. Figure 3a demonstrates that a lagged regression analysis involving Niño 3.4 could be susceptible to reporting a causal relationship when there is none—the lagged regression analysis is simply picking up the memory α in Niño 3.4. Granger causality analysis, on the other hand (as seen in Fig. 3b), would likely not be susceptible to this problem, as the results of the Granger causality analysis do not depend on α, even when α is very high [see Runge et al. (2014) for a more in-depth discussion of this effect]. This will be explored in the following section.

4. Applications in climate variability

a. ENSO and surface temperature

We now apply the results of our statistical model to the apparent paradox of Fig. 1. We know that ENSO's memory is large; do the benefits of Granger causality seen in the statistical model carry over to climate variability problems? This time, we perform lagged regression and Granger causality analysis in both directions—we use ENSO to predict T_s and T_s to predict ENSO. We focus only on temperatures over land.

Figure 4 compares lagged regression (Figs. 4a,b) and Granger causality (Figs. 4c,d) to test the hypothesis that ENSO drives T_s (Figs. 4a,c) and that T_s drives ENSO (Figs. 4b,d). While Fig. 4 shows only the results for a maximum lag of 7 months ($k = 7$), results from maximum lags of 3–9 months (from $k = 3$ to $k = 9$) are comparable. Red in Fig. 4 indicates that a significant lagged relationship is identified for $k = 7$. Red does not convey the magnitude of the relationship; it only indicates whether or not a significant relationship exists at a given grid point at 95% confidence. When testing whether or not ENSO drives T_s, Granger causality (Fig. 4c) and lagged regression (Fig. 4a) perform similarly: both indicate that ENSO from up to 7 months prior drives T_s over much of North and South America. However, when testing the other direction—that T_s from up to 7 months prior drives ENSO—the two methods yield very different results (Figs. 4b,d). In this case, the lagged regression (Fig. 4b) looks quite similar

to the results of the lagged regression testing whether or not ENSO drives T_s (Fig. 4a). Since we know that the autocorrelation of ENSO is very high, it seems unlikely that T_s is exerting such a strong influence on ENSO at lags of 7 months; it is more likely that the results of Fig. 4b are due, at least in part, to the high autocorrelation of ENSO. Granger causality does account for the memory in ENSO and shows that T_s over North and South America up to 7 months prior has little influence on ENSO (Fig. 4d). Put another way, Granger causality asks, ''What is the variance in ENSO due to T_s not already accounted for by ENSO itself?'' Therefore, since most of the variance in ENSO is explained by past values of ENSO, Granger causality does not report that T_s causes ENSO.

Since ENSO dynamics and teleconnections have been well studied and largely understood for decades, climate scientists are unlikely to misinterpret Fig. 4b. The memories of the two variables are vastly different, and the ENSO–T_s relationship is fairly well known. However, in cases where the dynamics are not as well understood, Granger causality analysis could provide valuable insights beyond those of traditional lagged regression.

b. *Arctic–midlatitude connections: Another example*

Finally, we use Granger causality analysis and lagged regression to investigate the relationship between Arctic temperature and low-level winds across the mid-to-high latitudes. The topic of the impact of Arctic warming on midlatitude weather and climate is one of much scientific discussion and debate (e.g., Walsh 2014; Barnes and Screen 2015 and references therein). However, the direction of the causality of this Arctic–midlatitude relationship is not clear—how much does the Arctic temperature drive midlatitude weather, and how much does midlatitude weather drive changes in Arctic temperature? We do not fully address these questions here; we simply seek to point out that Granger causality can provide information about the direction(s) of causal relationships that cannot be determined from traditional lagged regression.

To analyze the relationship between Arctic temperature and low-level winds, we define Arctic temperature T_{pol} as a vertically weighted average of 1000–700-hPa temperature from 70° to 90°N. Low-level zonal winds U700 are evaluated on the 700-hPa surface throughout the Northern Hemisphere. Both T_{pol} and U700 are calculated using daily means of 6-hourly data from the MERRA-2 reanalysis data on a 0.625° by 0.5° spatial grid (GMAO 2015). The seasonal cycle and second-order trends are removed from both T_{pol} and U700. Then T_{pol} and U700 are averaged into 5-day means in order to low-pass filter the data and focus on subseasonal variability rather than on individual synoptic events. Lagged regression and Granger causality analysis are performed for maximum lags spanning from 5 to 30 days ($k = 1, \ldots, 6$); we focus on a subseasonal time scale of 20 days ($k = 4$), but results are similar for maximum lags of 5–30 days. Here, we focus solely on the annual mean; the impacts of seasonality will be discussed in a later study.

Figure 5 displays the results of lagged (Figs. 5a,b) and Granger (Figs. 5c,d) regression analysis for T_{pol} and U700 at a maximum lag of 20 days ($k = 4$). Figures 5a and 5c test the hypothesis that T_{pol} drives U700; Figs. 5b and 5d test the hypothesis that U700 drives T_{pol}. Focusing first on the case of T_{pol} driving U700 (Figs. 5a,c), we see that both Granger causality (Fig. 5c) and lagged regression (Fig. 5a) show large-scale responses across much of Siberia, Alaska, the Canadian Arctic, and northern Europe, as well as signals in interior North America and Asia. Lagged regression, however, shows much larger responses over the ocean basins than Granger causality does. As the autocorrelation of the ocean is larger than that of the land surface (i.e., the ocean has more memory than the land), it is possible that the differences in the response in Fig. 5a, as compared to Fig. 5c, are due to the effect of memory over the oceans. Physically, Figs. 5a and 5c imply that Arctic lower-tropospheric temperatures may drive a response in the low-level zonal winds in the sub-Arctic, particularly over northern Europe, Siberia, and northern Canada. This response is consistent with studies that have reported links between Siberian temperature anomalies and snow cover and between Arctic amplification and sea ice loss (e.g., Inoue et al. 2012; Ghatak et al. 2012; Cohen et al. 2012; Peings et al. 2013).

The case of 700-hPa winds driving Arctic temperatures (Figs. 5b,d) presents a somewhat different picture. Again, the lagged regression (Fig. 5b) shows large-scale responses over much of the Northern Hemisphere—the Atlantic and Pacific storm tracks, much of continental North America, nearly the entire sub-Arctic (poleward of 60°N), most of Europe, and much of Siberia. Granger causality analysis (Fig. 5d) has a more limited large-scale response than that given by lagged regression; notably, Granger causality does not show a significant response over Siberia and shows a weaker, less spatially homogeneous response in the sub-Arctic region when compared to lagged regression. Previous work has linked changes in midlatitude circulation and sea surface temperatures to warmer Arctic temperatures (e.g., Graversen 2006; Screen et al. 2012; Wettstein and Deser 2014; Baggett and Lee 2015); however, as Fig. 5 demonstrates, the

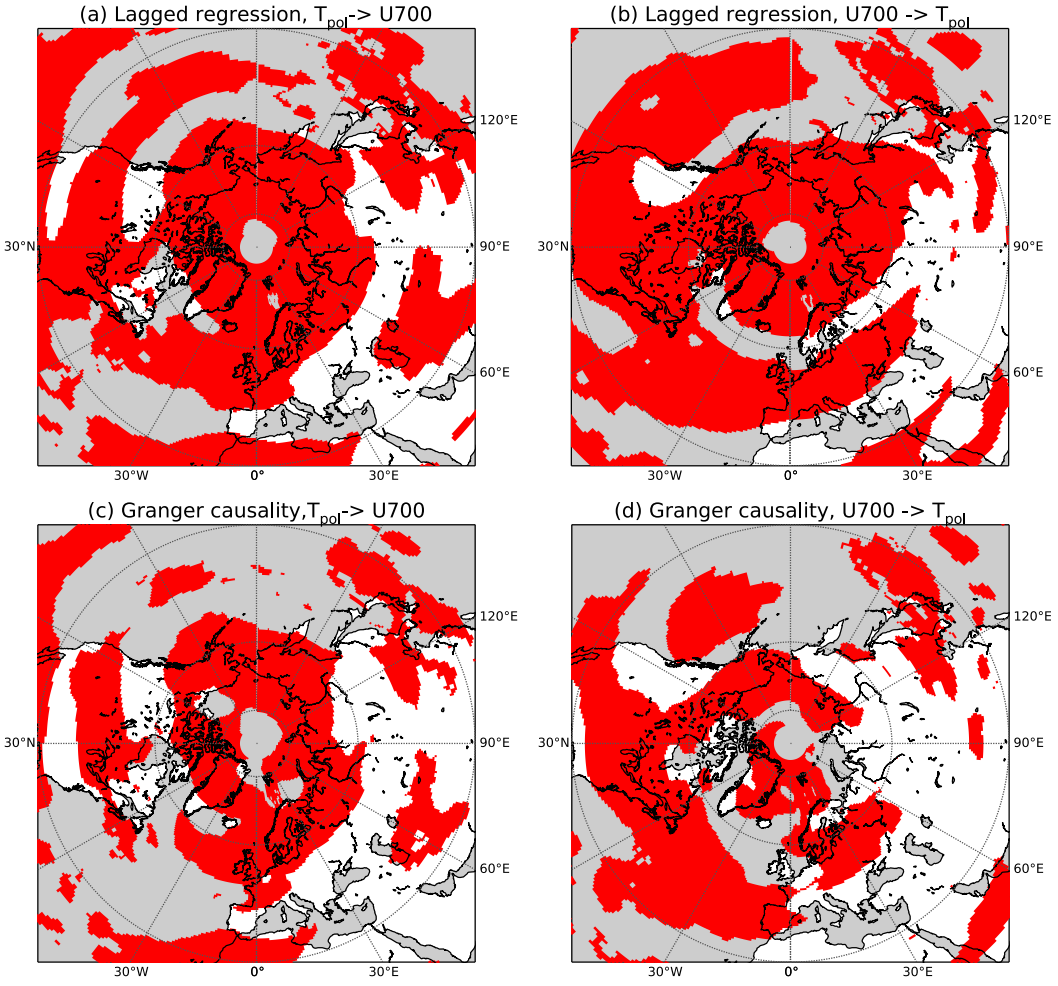

FIG. 5. Using (a),(b) lagged regression and (c),(d) Granger causality to test the hypothesis that (left) polar mean temperature drives 700-hPa zonal winds and (right) 700-hPa zonal winds drive polar mean temperature. Red indicates a significant lagged relationship identified at up to 20 days ($k = 4$). Significance is assessed at 95%.

details of these circulation changes differ with different methodologies.

5. Discussion

In this manuscript, we have tried to present a clear, concise, and compelling argument for an increased use of Granger causality analysis in climate variability studies. We have emphasized Granger causality's superior performance, as compared to lagged regression, in situations in which one or more variables has substantial memory. However, like any approach, Granger causality analysis has its own limitations. One obvious drawback is the possibility of a confounding variable—that is, an additional process or variable could be driving the modeled variables (e.g., in the bivariate case, a third process Z, could influence the independent X and dependent Y variables—$Z \rightarrow X$, $Z \rightarrow Y$). Using the bivariate case as an example, Granger causality may state that X causes Y, even though Z actually drives both X and Y. Similarly, Granger causality does not account for indirect effects or mediating variables. Returning to the bivariate example, a process X may indirectly drive Y via a third process Z ($X \rightarrow Z \rightarrow Y$). Again, Granger causality may state that X drives Y without including the necessary link Z. As discussed here, basic Granger causality analysis also requires assumptions of linear and stationary processes. An out-of-sample approach to Granger causality tests (e.g., Attanasio et al. 2012; Pasini et al. 2012; Attanasio et al. 2013) provides a framework for applying Granger causality to nonstationary processes. Cointegration (e.g., Johansen and Juselius 1990; Kaufmann and Stern 2002) is another approach to analyzing causality in

nonstationary processes that tend to vary together and have stochastic trends.

Moreover, Granger causality is simply one approach to causal analysis. Granger causality provides an opportunity for incremental improvement to the already-extant lagged regression analysis framework that has gained so much traction in climate variability studies. Multiple regression-based approaches, such as vector autoregressive (VAR) models, have built upon this Granger causality approach and have been applied to climate variability studies focused on the influence of sea ice on midlatitude circulation (e.g., Strong et al. 2009; Matthewman and Magnusdottir 2011), intraseasonal variability of sea ice (e.g., Wang et al. 2016), paleoclimate data (Davidson et al. 2016), and the relationship between the North Atlantic Oscillation and North Atlantic sea surface temperatures (e.g., Wang et al. 2004).

Even more recently, probabilistic graphical models based on Pearl causality have been introduced to climate science and represent the current state of the art in causal detection theory [see Ebert-Uphoff and Deng (2012) for a thorough introduction of graphical models in climate research]. This graphical approach to causality was first proposed in the 1980s (e.g., Rebane and Pearl 1987; Pearl 1988) and has since been refined and further developed. (e.g., Spirtes et al. 1991). Granger causality has, in fact, been incorporated into these graphical models, creating an approach known as graphical Granger models (e.g., Arnold et al. 2007). Ebert-Uphoff and Deng (2012) and Runge et al. (2014) have demonstrated the utility of these graphical approaches to causality in climate science, and we encourage readers to refer to these papers for more thorough discussions of these graphical models and their advantages in climate variability studies.

6. Conclusions

While lagged regression is a straightforward, popular, and often effective analysis technique in climate variability studies, it is vulnerable to overstating causal relationships in situations in which one or more datasets has significant memory (e.g., Runge et al. 2014). We use a Monte Carlo model to demonstrate the following:

1) Granger causality outperforms (i.e., lowers the risk of false detection) lagged linear regression when one or more variables has substantial memory;
2) Granger causality and lagged linear regression yield similar results when there is a true causal relationship between the variables (except in the case of very high autocorrelation); and
3) Granger causality analysis is only slightly more challenging to implement than traditional lagged linear regression analysis, as it simply consists of a lagged autoregression and a lagged multiple linear regression.

These general differences between lagged regression and Granger causality are also shown to be relevant for two large-scale climate dynamics examples, demonstrating that Granger causality analysis has useful and viable applications in climate variability studies.

Acknowledgments. Many thanks to Greg Herman, Thomas Birner, Eric Maloney, David Thompson, and Lauren McGough for their suggestions and feedback regarding this work. Many thanks also to the three anonymous reviewers and the editor, whose thorough reviews have greatly strengthened this work. The MERRA-2 data used in this study have been provided by the Global Modeling and Assimilation Office (GMAO) at NASA Goddard Space Flight Center. MCM and EAB are supported by the National Science Foundation under Grant AGS-1545675.

REFERENCES

Arnold, A., Y. Liu, and N. Abe, 2007: Temporal causal modeling with graphical Granger methods. *Proc. 13th ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining*, San Jose, CA, SIGKDD, 10 pp., https://doi.org/10.1145/1281192.1281203.

Attanasio, A., A. Pasini, and U. Triacca, 2012: A contribution to attribution of recent global warming by out-of-sample Granger causality analysis. *Atmos. Sci. Lett.*, **13**, 67–72, https://doi.org/10.1002/asl.365.

——, ——, and ——, 2013: Granger causality analyses for climatic attribution. *Atmos. Climate Sci.*, **3**, 515–522, https://doi.org/10.4236/acs.2013.34054.

Baggett, C. F., and S. Lee, 2015: Arctic warming induced by tropically forced tapping of available potential energy and the role of the planetary-scale waves. *J. Atmos. Sci.*, **72**, 1562–1568, https://doi.org/10.1175/JAS-D-14-0334.1.

Barnes, E. A., and J. A. Screen, 2015: The impact of Arctic warming on the midlatitude jet-stream: Can it? Has it? Will it? *Wiley Interdiscip. Rev.: Climate Change*, **6**, 277–286, https://doi.org/10.1002/wcc.337.

Barnett, L., and A. Seth, 2014: The MVGC multivariate Granger causality toolbox: A new approach to Granger-causal inference. *J. Neurosci. Methods*, **223**, 50–68, https://doi.org/10.1016/j.jneumeth.2013.10.018.

Blanchard-Wrigglesworth, E., K. C. Armour, C. M. Bitz, and E. DeWeaver, 2011: Persistence and inherent predictability of Arctic sea ice in a GCM ensemble and observations. *J. Climate*, **24**, 231–250, https://doi.org/10.1175/2010JCLI3775.1.

Cohen, J. L., J. C. Furtado, M. A. Barlow, V. A. Alexeev, and J. E. Cherry, 2012: Arctic warming, increasing snow cover and widespread boreal winter cooling. *Environ. Res. Lett.*, **7**, 014007, https://doi.org/10.1088/1748-9326/7/1/014007.

Compo, G., and Coauthors, 2011: The Twentieth Century Reanalysis project. *Quart. J. Roy. Meteor. Soc.*, **137**, 1–28, https://doi.org/10.1002/qj.776.

Davidson, J., D. Stephenson, and A. Turasie, 2016: Time series modeling of paleoclimate data. *Environmetrics*, **27**, 55–65, https://doi.org/10.1002/env.2373.

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.

Elsner, J., 2006: Evidence in support of the climate change–Atlantic hurricane hypothesis. *Geophys. Res. Lett.*, **33**, L16705, https://doi.org/10.1029/2006GL026869.

——, 2007: Granger causality and Atlantic hurricanes. *Tellus*, **59A**, 476–485, https://doi.org/10.1111/j.1600-0870.2007.00244.x.

Ghatak, D., C. Deser, A. Frei, G. Gong, A. Phillips, D. A. Robinson, and J. Stroeve, 2012: Simulated Siberian snow cover response to observed Arctic sea ice loss, 1979–2008. *J. Geophys. Res.*, **117**, D23108, https://doi.org/10.1029/2012JD018047.

GMAO, 2015: M2I6NPANA: MERRA-2 inst6_3d_ana_Np: 3d, 6-hourly, instantaneous, pressure-level, analysis, analyzed meteorological fields V5.12.4. Goddard Earth Sciences Data and Information Services Center, accessed 23 March 2017, https://doi.org/10.5067/A7S6XP56VZWS.

Granger, C., 1969: Investigating causal relations by econometric models and cross-spectral methods. *Econometrica*, **37**, 424–438, https://doi.org/10.2307/1912791.

Graversen, R. G., 2006: Do changes in the midlatitude circulation have any impact on the Arctic surface air temperature trend? *J. Climate*, **19**, 5422–5438, https://doi.org/10.1175/JCLI3906.1.

Gu, G., and R. F. Adler, 2011: Precipitation and temperature variations on the interannual time scale: Assessing the impact of ENSO and volcanic eruptions. *J. Climate*, **24**, 2258–2270, https://doi.org/10.1175/2010JCLI3727.1.

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.

Hendon, H. H., M. C. Wheeler, and C. Zhang, 2007: Seasonal dependence of the MJO–ENSO relationship. *J. Climate*, **20**, 531–543, https://doi.org/10.1175/JCLI4003.1.

Inoue, J., M. E. Hori, and K. Takaya, 2012: The role of Barents Sea ice in the wintertime cyclone track and emergence of warm-Arctic cold-Siberian anomaly. *J. Climate*, **25**, 2561–2568, https://doi.org/10.1175/JCLI-D-11-00449.1.

Johansen, S., and K. Juselius, 1990: Maximum likelihood estimation and inference on cointegration—With applications to the demand for money. *Oxford Bull. Econ. Stat.*, **52**, 169–210, https://doi.org/10.1111/j.1468-0084.1990.mp52002003.x.

Kaufmann, R. K., and D. I. Stern, 2002: Cointegration analysis of hemispheric temperature relations. *J. Geophys. Res.*, **107**, 4012, https://doi.org/10.1029/2000JD000174.

——, L. Zhou, R. Myneni, C. Tucker, D. Slayback, N. Shabanov, and J. Pinzon, 2003: The effect of vegetation on surface temperature: A statistical analysis of NDVI and climate data. *Geophys. Res. Lett.*, **30**, 2147, https://doi.org/10.1029/2003GL018251.

Klein, S. A., B. J. Soden, and N.-C. Lau, 1999: Remote sea surface temperature variations during ENSO: Evidence for a tropical atmospheric bridge. *J. Climate*, **12**, 917–932, https://doi.org/10.1175/1520-0442(1999)012<0917:RSSTVD>2.0.CO;2.

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.

Matthewman, N. J., and G. Magnusdottir, 2011: Observed interaction between Pacific sea ice and the western Pacific pattern on intraseasonal time scales. *J. Climate*, **24**, 5031–5042, https://doi.org/10.1175/2011JCLI4216.1.

Mohkov, I., D. Smirnov, P. Nakonechny, S. Kozlenko, E. Seleznev, and J. Kurths, 2011: Alternating mutual influence of El-Niño/Southern Oscillation and Indian monsoon. *Geophys. Res. Lett.*, **38**, L00F04, https://doi.org/10.1029/2010GL045932.

Mosedale, T., D. Stephenson, M. Collins, and T. Mills, 2006: Granger causality of coupled climate processes: Ocean feedback on the North Atlantic Oscillation. *J. Climate*, **19**, 1182–1194, https://doi.org/10.1175/JCLI3653.1.

Pasini, A., U. Triacca, and A. Attanasio, 2012: Evidence of recent causal decoupling between solar radiation and global temperature. *Environ. Res. Lett.*, **7**, 034020, https://doi.org/10.1088/1748-9326/7/3/034020.

Pearl, J., 1988: *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. 2nd ed. Morgan Kaufmann Publishers, 552 pp.

——, 2009: *Causality*. Cambridge University Press, 464 pp.

Peings, Y., E. Brun, V. Mauvais, and H. Douville, 2013: How stationary is the relationship between Siberian snow and Arctic Oscillation over the 20th century? *Geophys. Res. Lett.*, **40**, 183–188, https://doi.org/10.1029/2012GL054083.

Polvani, L. M., and D. W. Waugh, 2004: Upward wave activity flux as a precursor to extreme stratospheric events and subsequent anomalous surface weather regimes. *J. Climate*, **17**, 3548–3554, https://doi.org/10.1175/1520-0442(2004)017<3548:UWAFAA>2.0.CO;2.

Rayner, N., D. Parker, E. Horton, C. Folland, L. Alexander, D. Rowell, E. Kent, and A. Kaplan, 2003: Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century. *J. Geophys. Res.*, **108**, 4407, https://doi.org/10.1029/2002JD002670.

Rebane, G., and J. Pearl, 1987: The recovery of causal poly-trees from statistical data. *Proc. Third Workshop on Uncertainty in AI*, Seattle, WA, AUAI, 222–228.

Ropelewski, C., and M. Halpert, 1986: North American precipitation and temperature patterns associated with the El Niño/Southern Oscillation (ENSO). *Mon. Wea. Rev.*, **114**, 2352–2362, https://doi.org/10.1175/1520-0493(1986)114<2352:NAPATP>2.0.CO;2.

Runge, J., 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.

Screen, J. A., C. Deser, and I. Simmonds, 2012: Local and remote controls on observed Arctic warming. *Geophys. Res. Lett.*, **39**, L10709, https://doi.org/10.1029/2012GL051598.

Spirtes, P., C. Glymour, and R. Scheines, 1991: From probability to causality. *Philos. Stud.*, **64**, 1–36, https://doi.org/10.1007/BF00356088.

Stokes, P. A., and P. Purdon, 2017: A study of problems encountered in Granger causality analysis from a neuroscience perspective. *Proc. Natl. Acad. Sci. USA*, **114**, E7063–E7072, https://doi.org/10.1073/pnas.1704663114.

Strong, C., G. Magnusdottir, and H. Stern, 2009: Observed feedback between winter sea ice and the North Atlantic Oscillation. *J. Climate*, **22**, 6021–6032, https://doi.org/10.1175/2009JCLI3100.1.

Walker, G. T., 1923: Correlations in seasonal variations of weather, VIII: A preliminary study of world weather. *Mem. Indian Meteor. Dept.*, **24**, 75–131.

——, 1924: Correlation in seasonal variations of weather, IX: A further study of world weather. *Mem. Indian Meteor. Dept.*, **24**, 275–333.

Walsh, J., 2014: Intensified warming of the Arctic: Causes and impacts on middle latitudes. *Global Planet. Change*, **117**, 52–63, https://doi.org/10.1016/j.gloplacha.2014.03.003.

Wang, L., X. Yuan, M. Ting, and C. Li, 2016: Predicting summer Arctic sea ice concentration intraseasonal variability using a vector autoregressive model. *J. Climate*, **29**, 1529–1543, https://doi.org/10.1175/JCLI-D-15-0313.1.

Wang, W., B. T. Anderson, R. K. Kaufmann, and R. B. Myneni, 2004: The relation between the North Atlantic Oscillation and SSTs in the North Atlantic basin. *J. Climate*, **17**, 4752–4759, https://doi.org/10.1175/JCLI-3186.1.

Wettstein, J. J., and C. Deser, 2014: Internal variability in projections of twenty-first-century Arctic sea ice loss: Role of the large-scale atmospheric circulation. *J. Climate*, **27**, 527–550, https://doi.org/10.1175/JCLI-D-12-00839.1.

Yu, J.-Y., H.-Y. Kao, and T. Lee, 2010: Subtropics-related interannual sea surface temperature variability in the central equatorial Pacific. *J. Climate*, **23**, 2869–2884, https://doi.org/10.1175/2010JCLI3171.1.