Exploring Granger causality between global average observed time series of carbon dioxide and temperature

Evan Kodra · Snigdhansu Chatterjee ·
Auroop R. Ganguly

Abstract Detection and attribution methodologies have been developed over the years to delineate anthropogenic from natural drivers of climate change and impacts. A majority of prior attribution studies, which have used climate model simulations and observations or reanalysis datasets, have found evidence for human-induced climate change. This papers tests the hypothesis that Granger causality can be extracted from the bivariate series of globally averaged land surface temperature (GT) observations and observed CO_2 in the atmosphere using a reverse cumulative Granger causality test. This proposed extension of the classic Granger causality test is better suited to handle the multisource nature of the data and provides further statistical rigor. The results from this modified test show evidence for Granger causality from a proxy of total radiative forcing (RC), which in this case is a transformation of atmospheric CO_2, to GT. Prior literature failed to extract these results via the standard Granger causality test. A forecasting test shows that a holdout set of GT can be better predicted with the addition of lagged RC as a predictor, lending further credibility to the Granger test results. However, since second-order-differenced RC is neither normally distributed nor variance stationary, caution should be exercised in the interpretation of our results.

E. Kodra
Statistics, Operations and Management Science,
University of Tennessee,
331 SMC, 916, Volunteer Blvd,
Knoxville, TN 37996-0532, USA

E. Kodra · A. R. Ganguly (✉)
Geographic Information Science and Technology Group,
Computer Science and Engineering Division,
Oak Ridge National Laboratory,
1 Bethel Valley Rd,
Oak Ridge, TN 37831-6017, USA
e-mail: gangulyar@ornl.gov

S. Chatterjee
School of Statistics, University of Minnesota,
313 Ford Hall, 224 Church Street S.E.,
Minneapolis, MN 55455, USA

1 Introduction

The Fourth Assessment Report (AR4) of the Intergovernmental Panel on Climate Change notes that it is "very likely" (90%) that "most" (greater than 50%) of the increase in global average temperature in the second half of the twentieth century can be attributed to anthropogenic greenhouse gases (Solomon et al., IPCC WG1 AR4 Report, Summary for Policymakers 2007). Substantial advances have been made in climate change detection and attribution through the analysis of climate model outputs (e.g., Hasselmann 1979; Santer et al. 1991; Santer et al. 1993; Hegerl et al. 1996; Hegerl et al. 1997; Meehl et al. 2007; Barnett et al. 2008). A few detection and attribution studies have also relied solely on observations (e.g., Tol and De Vos 1993; Sun and Wang 1996; Tol and De Vos 1998; Triacca 2005). Past work attempting to statistically assess Granger causality from observed emissions to globally averaged temperature observations has varied in both method and results (e.g., Sun and Wang 1996; Triacca 2005).

We hypothesize that Granger causality (GC) may serve as a tool for attribution in the multivariate case of anthropogenic emissions, natural cycles, and temperature. As a step in testing this hypothesis, this work introduces a variant of the classic bivariate GC test, reverse cumulative

Granger causality (RCUMGC) testing, applied to two variables, a proxy for radiative forcing (RC), which in this work is a transformation of CO_2, and global land surface temperature anomalies (GT). The results presented in this work seem to indicate GC from RC to GT, at least in relatively large datasets. However, the appropriateness of the classical GC tests we employ are subject to several probabilistic assumptions being valid, many of which can be verified from data. These assumptions are not all met, and so our results have a degree of uncertainty beyond the usual uncertainty quantified in statistical testing. In addition, we find that the effects of RC might be overshadowed by other more statistically significant causal variables, such as the El Niño Southern Oscillation Index (ENSO).

Section 2 briefly discusses GC in its general form and in the specific form employed here, as well as previous applications of GC in climatology, several specifically between CO_2 and temperature. Section 3 describes the specific data used as well as how it is preprocessed. Section 4 describes the testing procedures selected and discusses why they were used. In addition, we discuss the limitations of the chosen methods. Section 5 reveals the results and discusses their implications. Section 6 discusses the limitations of this work and briefly proposes its possible future extensions.

The Electronic supplementary material contains details on selected procedures from this work.

2 Granger causality overview

GC is a technique first developed for use in econometrics. It attempts to identify causal relationships between sets of two or more time series (Granger 1969).

We present GC for two variables, which we employ in this work. To say one variable X Granger causes another variable Y is to say that, by using past values of both variables X and Y, we can better predict future values of Y than by using only past values of Y. That is, past observations of X contain information useful for predicting Y, beyond what is available from past observations of Y itself.

Suppose X and Y form a bivariate time series given by the dynamic relationship:

$$Y_t = \phi_0 + \sum_{j=1}^{n} \alpha_j Y_{t-j} + \sum_{j=1}^{n} \beta_j X_{t-j} + \in_{1t} \quad (1)$$

$$X_t = \lambda_0 + \sum_{j=1}^{n} \delta_j X_{t-j} + \sum_{j=1}^{n} \omega_j Y_{t-j} + \in_{2t}. \quad (2)$$

If $\beta = (\beta_1, \dots \beta_n)^T$ is *not* the zero vector $(0, \dots 0)^T$ and $\omega = (\omega_1, \dots \omega_n)^T$ *is* the zero vector $(0, \dots 0)^T$, then X is said to Granger cause Y. If ω is not the zero vector and β is the zero vector, then Y Granger causes X. If neither β nor ω is the zero vector,

then there is dependence in both directions, i.e., feedback between X and Y. If both β and ω are the zero vectors, there is no GC. The terms represent the white noise innovation at each instance of time t and are assumed to be independently and identically distributed with a bivariate normal law. The terms ϕ_0 and λ_0 represent intercepts for each equation.

GC has been applied numerous times in climate studies. Elsner (2006, 2007) applied a GC analysis to sea surface temperature anomalies and global surface temperature anomalies for Atlantic hurricanes. GC has been used to assess the "feedback of daily sea surface temperatures (SSTs) on daily values of the North Atlantic as simulated by a realistic coupled general circulation model (GCM)" (Mosedale et al. 2006). They find that SST Granger causes the North Atlantic Oscillation. Kaufmann and Stern (1997) use GC tests to show evidence that Southern Hemisphere leads the Northern Hemisphere in regards to temperature, which suggests that humans have contributed to climate change. Salvucci et al. (2002) uses GC to investigate soil moisture feedbacks from precipitation in Illinois. This work attempts to show evidence for causality going from soil moisture to precipitation, using a form of GC involving a Markov model (Salvucci et al. 2002). Sun and Wang (1996), using a classic GC partial F test, suggest that CO_2 Granger causes global temperature. Smirnov and Mokhov (2009) introduce the idea of long-term Granger causality and apply it to temperature and CO_2; their method involves empirical modeling, extending the concept of GC to deal with longer-term behavior. They conclude that the rise in temperature over the last several decades can only be explained with the presence of CO_2. However, Triacca (2005) suggests that there is no significant GC from CO_2 to global temperature, and that GC does not appear to be an appropriate tool for studying the causal relationship between these two variables. This work uses methodology from Toda and Yamamoto (1995), which is robust to the integration/co-integration properties of the data, but might require large sample size to obtain reliable results. Attanasio and Triacca (2010), a follow-up on Triacca (2005), find that a neural network based (and hence non-linear) approach suggests that GC exists from radiative forcing to global temperature.

We follow the classical F test methodology for our analysis in this paper. In case of potential violation of the classical statistical assumptions that are needed to justify the F test, Hacker and Hatemi-J (2006) suggests using a bootstrap distribution instead. Another potential method involves using a non-parametric test for GC (Hiemstra and Jones 1994).

Recently, grouped Granger causality models have appeared in literature. This extends the notion of temporal lags in GC to spatial lags, meaning neighboring spatial points can be tested for GC in addition to past values as discussed earlier. This method has been applied to observations in a study of climate change (Lozano et al. 2009a, b).

3 Data and preliminary data analysis

3.1 Data

The CO_2 atmospheric concentration data in annual parts per million (ppm) are from the Mauna Loa observatory starting at 1959 (C.D. Keeling, T.P. Whorf, and the Carbon Dioxide Research Group Scripps Institute of Oceanography (SIO) University of California La Jolla, CA, USA). The CO_2 concentration from several months in 1964 was originally missing, but the open-source statistical software R has a base package ("datasets") that contains the monthly PPM data from 1959 to 1997, with those missing months in 1964 estimated by linear interpolation, which gives us the 1964 annual average. The annual 1964 value from R is added to the existing data, giving us observations from 1959 to 2008.

For annual CO_2 PPM values from 1860 to 1958, we use the 20-year smoothed values estimated from the Law Dome DE08, DE08-2, and DSS ice cores (Etheridge et al., Division of Atmospheric Research, CSIRO, Aspendale, Victoria, Australia; J-.M. Barnola, Laboratoire de Glaciologie et Geophysique de l'Environnement, Saint Martin d'Heres-Cedex, France and V.I. Morgan Antarctic Division, Hobart, Tasmania, Australia).

GT for 1860–2008 are obtained from the Climate Research Unit at the University of East Anglia. These global land surface temperature values are expressed as anomalies from the average of the base years 1961–1990.

We obtain ENSO annual indices for 1860–2008 from http://jisao.washington.edu/data/globalsstenso/ (Todd Mitchell, Joint Institute for the Study of the Atmosphere and Ocean, University of Washington, Seattle, WA, USA). The December 2008 value was not yet available at the time of analysis, so 2008 is an average of January through November. The anomalies are expressed in hundredths of degrees Celsius as deviations from the period 1950–1979.

To obtain the proxy RC variable, we apply the following transformation to CO_2 as per Myhre et al. (1998)

$$RC = 5.35 \ln(C/C_0). \qquad (3)$$

Here, C represents CO2 and $C_0 = 280$ is the assumed pre-1750 concentration of CO_2 in ppm (Myhre et al. 1998). We assume RC as a proxy for all anthropogenic emissions. A time plot of RC is shown in Fig. 1 along with GT, CO_2, and ENSO.

3.2 Preliminary data analysis

Inference based on ordinary least squares parameter estimates and normality-driven white noise process is justifiable when the bivariate time series is second-order stationary (Kwiatkowski et al. 1992). Consequently, it is necessary to find a level of differencing at which each variable is approximately normal and stationary.

In Table 1, we introduce differencing notation, which is applied throughout this work.

For example, GT-k represents the k-differenced GT series.

Figure 1 suggests that neither the RC nor the GT series are stationary or have normal innovations. To test this formally, we conduct the Shapiro–Wilks test (Shapiro and Wilks 1965) and the KPSS test (Kwiatkowski et al. 1992), respectively.

The Shapiro–Wilks tests reveal that RC is not approximately normal at any level of differencing, likely due to the interpolated values of the ice core CO_2 data. This will have implications which will be addressed later in our analysis. GT is sufficiently normal at GT-1 and GT-2. However, note that the Shapiro–Wilks test relies on the assumption that data are independent. Autocorrelation plots of all six variables reveal that none of the variables may be deemed independent.

The KPSS test is used in many applications to test for stationarity of a series (Kwiatkowski et al. 1992). Based on the result of this test, the RC series may be considered stationary only at RC-2, while the GT series may be considered stationary at both GT-1 and GT-2. These tests partially confirm the results of Stern and Kaufmann (1999). We conduct GC tests, then, with the series GT-2 and RC-2.

Bivariate plots of GT-2 and RC-2 (Fig. 2) suggest an adequate linear relationship for conducting linear GC tests. Figure 2 shows the bivariate plot first fit by an ordinary least squares trend and then with a trend fit by robust maximum likelihood estimation. The robust regression trend line puts less emphasis on outliers that might heavily and wrongly influence the estimation of the slope which could in turn potentially influence the significance of the trend. Both of these trend lines are significant and nearly identical, and thus we deem the relationship between these two variables adequately linear for GC testing.

The ENSO index, which is integrated further into our analysis, is found to be approximately normal and stationary at ENSO-1. Thus, when conducting GC tests on GT and ENSO, we use GT-1 and ENSO-1.

4 Methods

In order to estimate the maximum lags at which our data will be tested for GC, we use vector autoregression (VAR), an extension of univariate autoregression (Zivot and Wang 2002). We look at three different criteria in order to find an optimal maximum lag: the Akaike information criterion (AIC), the Bayesian information criterion (BIC), and the Hannan–Quinn (HQ) statistic. These criteria attempt to find
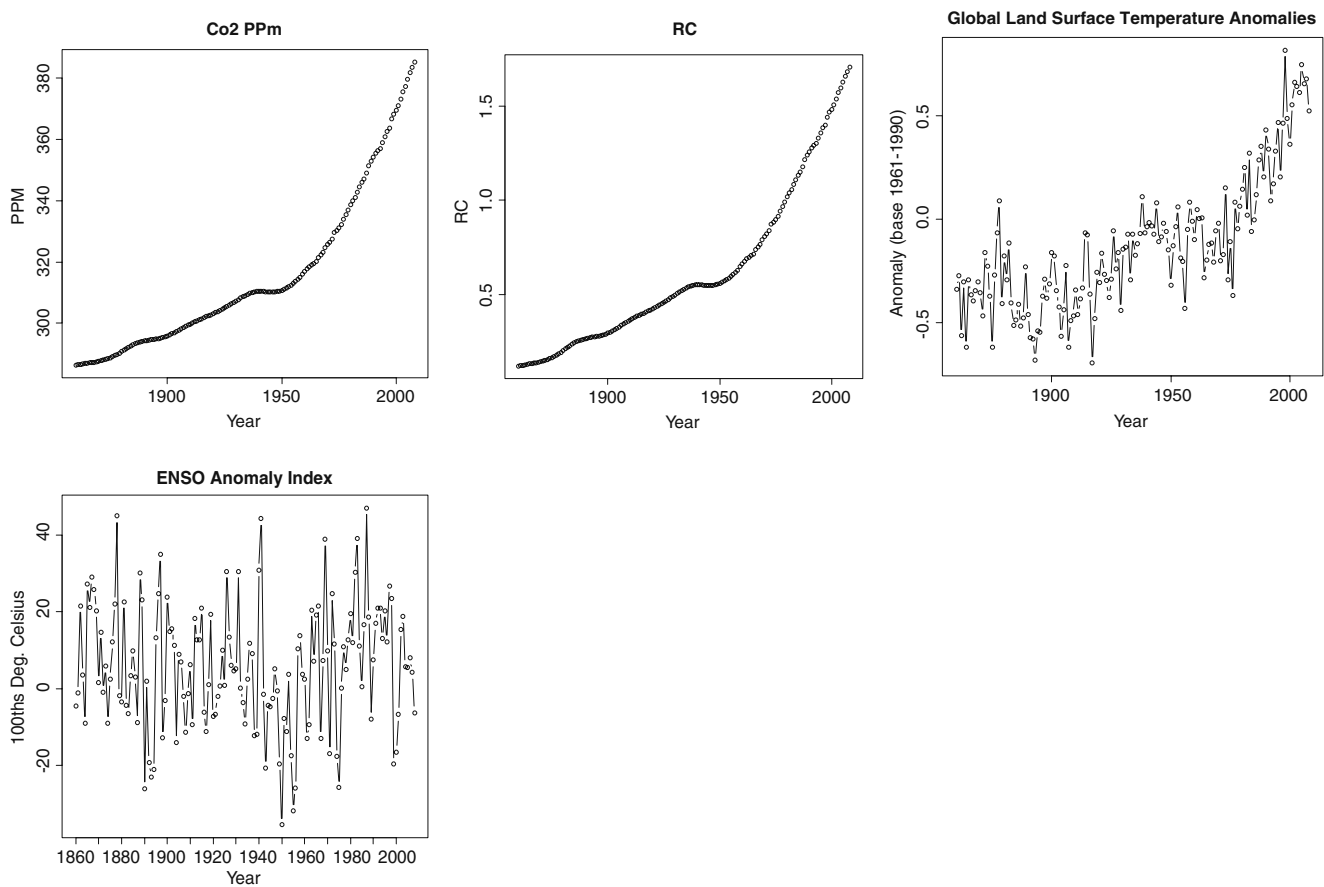
Fig. 1 Time plots of CO_2, RC, GT, and the ENSO index, 1860–2008

a lag with sufficient information content on the two variables without over fitting (Zivot and Wang 2002).

Once we have identified a set of lags at which to test for GC, we conduct the forward cumulative windows GC test (Triacca 2005). Specifically, we perform classic GC partial F tests, in both directions, GT to RC and RC to GT, on range of lags, using forward cumulative windows (years 1860–1900, 1860–1910, 1860–1920... 1860–2008, adding

10 years for each test, except for the last, where only 8 years are added). Section 5.1 discusses the results from this test.

Subsequently, we introduce an alternative test of causality between these two variables: RCUMGC testing. This test is motivated by the results of the forward cumulative tests as well as the interpolation present in the ice core CO_2 data. The motivation from the forward

Table 1 All variables used in this study, shown with respective definitions

Variable name	Variable definition
GT	Raw, undifferenced annual GT series, 1860–2008. GT also refers to GT in general
GT-1	First-differenced GT series
GT-2	Second-differenced GT series
RC	Transformation of CO_2, undifferenced. 1860–2008. RC also refers to RC in general
RC-1	First-differenced RC series
RC-2	Second-differenced RC series
ENSO	Raw, undifferenced ENSO series from 1860 to 2008. ENSO also refers to ENSO in general
ENSO-1	First-differenced ENSO series
ENSO-2	Second-differenced ENSO series

Fig. 2 In order to use the proposed GC F tests, the relationship between RC-2 and GT-2 must be adequately linear. The first plot is an ordinary least squares regression line fit to the relationship, with GT-2 as the dependent variable. The slope of this line is found to be significant, indicating a linear relationship. The second plot is a robust regression trend line. The term $d=2$ simply means that these two variables are RC-2 and GT-2
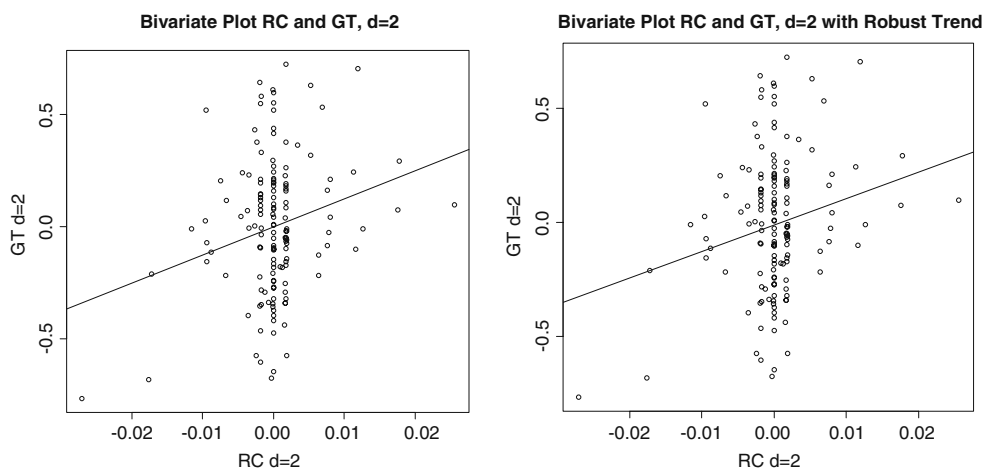
cumulative tests is presented in Section 5.1. The interpolation of many annual ice core CO_2 values may mask some of the dependence structure between RC and GT. Thus, the RCUMGC conducts some tests without the presence of any ice core CO_2 data.

In this test, we apply GC F tests with additive "latest windows," meaning that a GC test is conducted, for example, corresponding to the last 30 years (1979–2008) of RC-2 and GT-2. The test is conducted for the last 15, 20, 25, 30, 40, 50, 60, 70, 80, 90, 100, 110, 120, 130, 140, and 147 years. These tests are conducted for seven lag values, in both RC-2-to-GT-2 and GT-2-to-RC-2 directions, so there are 224 total RCUMGC tests. The years 1959–2008 (50 years) contain no ice core CO_2-derived RC values.

In order to summarize the results of these RCUMGC tests, we construct a plot to show the direction of causality. We define for each latest-window size:

$$H = \frac{F_1}{F_2} \text{ if } F_1 > F_2,$$
$$H = -\frac{F_2}{F_1} \text{ if } F_2 > F_1, \qquad (4)$$
$$H = 0 \text{ if } F_1 = F_2,$$

where F_1 and F_2 are, respectively, the partial F ratios, a measure of significance, of the RC→GT and the GT→RC GC tests. Partial F ratios quantify the amount of additional variance explained by a group of predictor variables (Kutner et al. 2004). Thus, in this case, a partial F ratio quantifies the additional variance explained in the predictand by a subset of predictors; the subset is either lagged RC or lagged GT, whichever happens to be the exogenous group of variables in any given test. Sufficiently, numbers of positive (negative) values of H imply the broad generic pattern of GC from RC→GT (GT→RC). Examining the ratio of two F ratios gives us additional insights, that p values by themselves may not reveal (Stigler 2005). The results of the RCUMGC test are discussed in Section 5.2. Stern and Kaufmann (1999) use a similar procedure, where data are added in reverse in GC testing.

We also conduct a forecasting test. We consider four forecasting scenarios: predicting GT using past values of GT, GT using past values of RC and GT, RC using past values of RC, and RC using past values of RC and GT. All the four forecasting problems are studied using lag values 1, 2, and 3 and using the GT-2 and RC-2 series. We use the first 137 time points as a training set to construct the forecasting procedure, and use the last 10 time points as the test set for evaluating each forecasting scheme. We evaluate the forecasting schemes using the coefficient of determination (R^2), the mean absolute percentage error (MAPE), and the maximum absolute percentage error (Max. APE) of the 10-length validation set. The results of this forecast test are discussed in Section 5.3.

Finally, an exploratory analysis of the dependence structure between RC and GT shows that, at times, RC-2 is negatively correlated with GT-2. One possible explanation for this periodic negative correlation is the ENSO index. Hence, we integrate ENSO into our dependence structure analysis. Specifically, we test for correlation significance between GT-2 and RC-2, between GT-1 and ENSO-1, and between RC-2 and ENSO-2, using 30-year moving windows. We also use the same GC F tests to check for GC between GT-1 and ENSO-1. The results and of this analysis are discussed in Section 5.4.

5 Results

5.1 Forward cumulative GC test results

The three information criterion tests for fitting a VAR model yield mixed results. Lag 7 was indicated as the maximum lag for the AIC and HQ tests, while the BIC test indicated lag 3 as the maximum lag.

We conduct our test at each of these lags (3 and 7), in both RC-2 to GT-2 and GT-2 to RC-2 directions, using forward cumulative windows as described in Section 4.

There are 11 tests in each direction for both lags 3 and 7 (a total of 44 tests), and only two tests have p values less than 0.10. If all of these tests were independent of each other, with a significance level of 0.10, approximately four false positives would be expected if the null hypothesis of no GC is true. Thus, we should not interpret that these two p values of less than 0.10 imply GC. The results for the forward cumulative GC tests are reported in Tables 2 and 3.

Note the sudden dip in the size of RC→GT p values as more data are added to the tests, specifically starting with the 1860–1980 lag 3 test. This, along with the interpolation of the CO_2 ice core values, leads us to the RCUMGC tests, since this dip in p values may indicate that a GC from RC→GT exists in more recent times while being less significant in earlier times, which may have been masked in the forward cumulative window testing procedure. Thus, with the RCUMGC tests, we look backward in time rather than forward.

5.2 RCUMGC results

As the AIC and HQ suggests seven maximum lags and the BIC suggests 3, we conduct RCUMGC tests in both directions at lags 1–7 for a total of 224 tests. Figures 3 and 4 suggest that with progressively larger latest-window sizes, there is increasing evidence for GC in RC-2 to GT-2. With short latest windows, we see evidence in some lags for GC from GT-2 to RC-2. However, tests at the latest 15, 20, and 25 years may not be credible since they suffer from small sample problems. These problems are exaggerated in high lag (i.e., 4, 5, 6, and 7) tests, where even more data are lost.

All points in Fig. 4 are calculated as per Eq. 4. The plot of H shows that as more data are added, the case becomes stronger for GC from RC→GT. We also find that the

Table 2 Causality tests with forward cumulative windows, lag=3

Period	Test statistic (RC→GT)	p value (RC→GT)	Test statistic (GT→RC)	p value (GT→RC)
1860–1900	0.537	0.661	1.023	0.396
1860–1910	1.040	0.385	0.569	0.638
1860–1920	1.037	0.384	0.572	0.636
1860–1930	0.941	0.426	0.603	0.615
1860–1940	0.636	0.595	0.339	0.797
1860–1950	0.709	0.549	0.375	0.771
1860–1960	0.805	0.494	0.255	0.857
1860–1970	0.119	0.948	0.298	0.827
1860–1980	1.329	0.269	0.402	0.752
1860–1990	2.220	0.089[a]	0.239	0.869
1860–2008	1.983	0.119	0.624	0.601

Results from GC test in both possible directions of causality, lag 3, using RC-2 and GT-2

[a] Significant at 0.10

Table 3 Causality tests with forward cumulative windows, lag=7

Period	Test statistic (RC→GT)	p value (RC→GT)	Test statistic (GT→RC)	p value (GT→RC)
1860–1900	0.577	0.766	2.424	0.059[a]
1860–1910	0.637	0.721	0.637	0.721
1860–1920	0.622	0.735	0.777	0.611
1860–1930	0.478	0.846	0.516	0.818
1860–1940	0.252	0.969	0.716	0.659
1860–1950	0.198	0.985	0.447	0.869
1860–1960	0.309	0.948	0.303	0.951
1860–1970	0.583	0.768	0.677	0.691
1860–1980	0.804	0.586	0.469	0.855
1860–1990	0.868	0.535	1.226	0.295
1860–2008	1.482	0.179	1.634	0.132

Results from GC test in both possible directions of causality, lag 3, using RC-2 and GT-2

[a] Significant at 0.10

overwhelming majority of RC-2-to-GT-2 models have approximately normally distributed residuals, while most of the GT-2-to-RC-2 models do not. Since the GT-2-to-RC-2 residuals are not distributed normally, the significance of the vectors $\omega=(\omega_1,...\omega_n)^T$ may actually be over- or understated.

The final set of seven points of Fig. 3 (where all data are used, 147 indices for each variable) show that the ratio (H) of F statistic ranges from slightly less than −1 or more than 1, implying no real difference in F ratios in either direction, to 10, meaning an F ratio for RC→GT is 10 times larger than the one from GT→RC. Five of these final seven points have ratios of more than 2, while the other two have ratios close to |1|. Overall, this shows evidence for RC→GT and no significant evidence for GT→RC.

The residuals from the RC→GT models are all approximately normal, while many of those from the GT→RC models (55/112) are not. Thus, the significance of many of the vectors $\omega=(\omega_1,...\omega_n)^T$ here may be over- or understated here.

These results are not definitive (refer to Section 6.1 for a discussion of limitations), but the RCUMGC procedure gives us a previously unexplored perspective on the relationship between RC and GT.

5.3 Forecast results

The results for the forecasting tests are displayed in Table 4.

The results in Table 4 show no appreciable differences between models 3 and 4. However, model 2 seems to have an appreciably higher R^2 holdout and lower maximum APE than model 1.

The above forecasting test results give more evidence for the RC→GT hypothesis and lend more credibility to the results reported in Section 5.3.
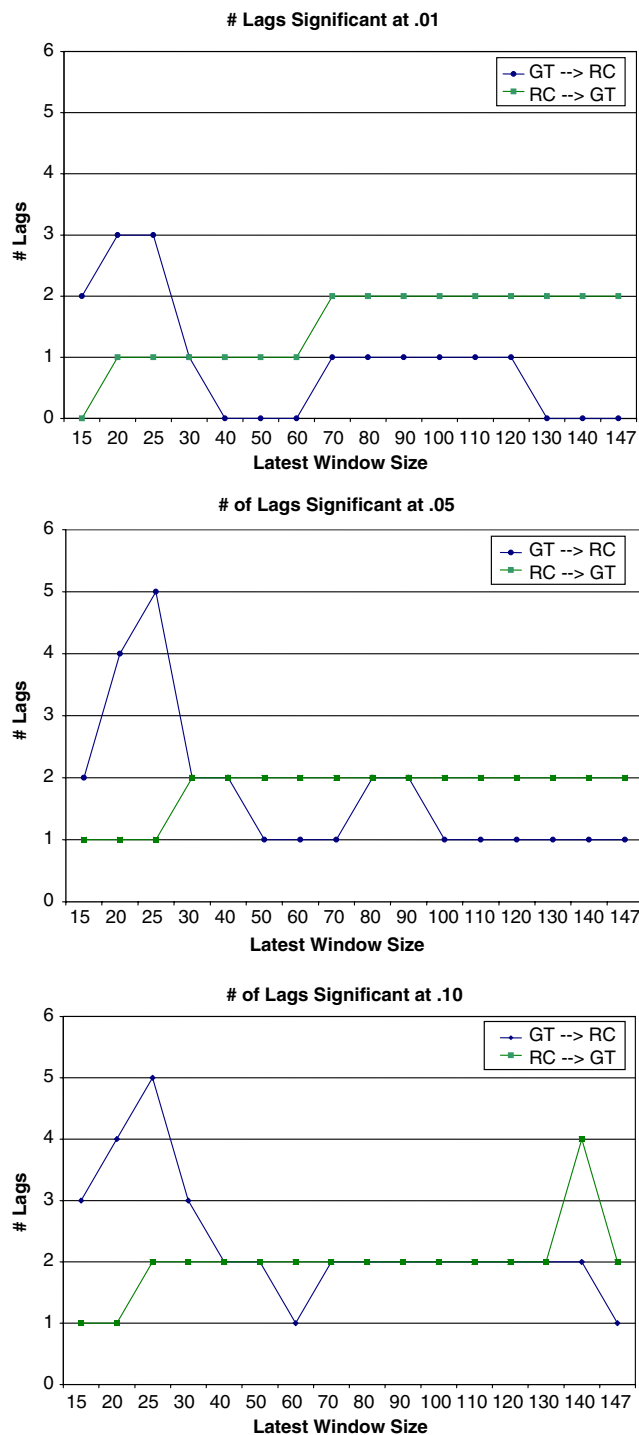
Fig. 3 These graphs represent the evolution of number of lags significant (out of seven for tests at maximum lags 1, 2 …7) at $\alpha =$ 0.01, 0.05, and 0.10 with changing latest-window size. That is, a latest-window size of 15 is for the *last* 15 years of the RC-2 and GT-2 data and so on. As the latest-window size increases, there is a general trend providing evidence that RC Granger causes GT

5.4 Exploratory correlation analysis

We begin by specifying the motivation for integrating the ENSO index into this part of our work. From the forward

cumulative GC tests, we see quite high p values before the 1970s. The ENSO index has a significant relationship with global temperature patterns (Ropelewski and Halpert 1986). We hypothesize that the ENSO index may be one factor that affects GT and also possibly statistically obscures the effects of RC, particularly in earlier years. Thus, perhaps ENSO is, at least partly, a reason why we see no GC in our forward cumulative GC tests in Section 4.

We plot spline curves of the ENSO index and GT along with RC (Fig. 5). This plot indicates a relationship between ENSO and GT, as their temporal patterns appear to be very similar in many aspects. Next, we plot a 30-year forward moving window correlation-significance index between GT-1 and ENSO-1, between GT-2 and RC-2, and finally RC-2 and ENSO-2 (Fig. 6). For example, a p value at year 1991 indicates the significance of the correlation between GT-1 and ENSO-1, GT-2 and RC-2, or RC-2 and ENSO-2 for the years 1962–1991.

From Fig. 6, we see that an abrupt change in significance of the correlation between RC-2 and GT-2 occurs in the 30-year time window ending in 1974. This is the same time frame when there was a sudden drop in the RC→GT p value from the forward cumulative GC tests. We find evidence that this sudden change in significance is not *solely* due to a change in data source, i.e., where in 1959 CO_2 values become Mauna Loa data as opposed to ice core data.

Here, we note that an abrupt climate regime change occurred in the 1970s, and its cause is not well known (Graham 1994). An abrupt climate change is said to occur when a climate system is forced across some threshold (Committee on Abrupt Climate Change, National Research Council 2002). Alley et al. (2005) notes that even a slow forcing could cause an abrupt change. This noted abrupt climate regime change corresponds closely with the time frame in which the RC-2 and GT-2 correlation becomes significant. This is an interesting phenomenon and one that deserves attention; we may report observations on this in a future work.

We have seen that there are interesting temporal correlation patterns between the ENSO and GT and especially between RC and GT. The ENSO index may indeed affect the nature of causality between RC and GT. Several GC tests between GT and ENSO as well as GT and RC show evidence from ENSO→GT only. This GC is statistically more significant than that found from RC→GT, supporting the hypothesis that ENSO could be obscuring our view of some causality or at least correlation between RC-2 and GT-2 in earlier years. Table 5 displays the results for these tests. Note in Table 5 that GC tests indicated feedback between RC-2 and ENSO-2.

With the accumulation of all results, we conceive three competing hypotheses as to why there is a sudden jump in correlation significance between RC-2 and GT-2:

Fig. 4 The *left* graph shows the evolution of an *H* ratio (a ratio of *F* ratios) for each lag. Increasingly *negative numbers* indicate evidence of GT Granger causing RC, and *positive numbers* indicate evidence of RC Granger causing GT. There is an overall trend that as size of latest window increases, causality is increasingly in the RC→GT direction. The *second graph on the right* (*zoomed in*) is simply shown for the reader to more easily see the trend toward RC→GT with larger latest-window sizes
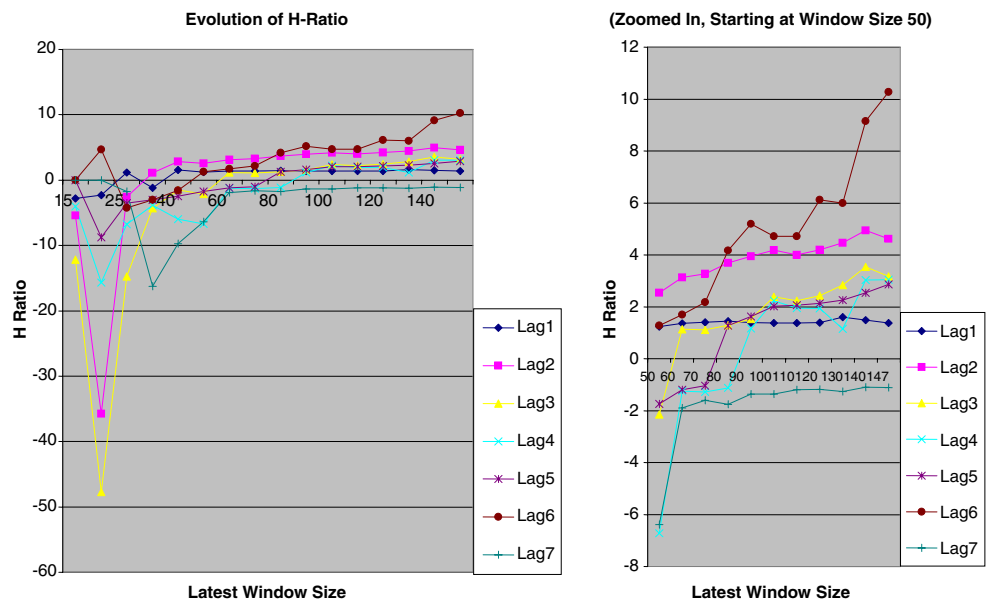
1. The smoothing of the early CO_2 values hides the early dependence structure between the two variables. Thus, we only see it later. We remark several paragraphs earlier that we find evidence (contained in the Electronic supplementary material) that this is at least not *completely* the case, although it may contribute.
2. ENSO is a more statistically significant covariate at times and thus sometimes hides the RC-2 and GT-2 correlation. Here, we are not limited to ENSO; other atmospheric circulation variables could play a masking role as well.
3. The aforementioned abrupt regime change in the 1970s explains this jump (Graham 1994; Alley et al. 2005; Committee on Abrupt Climate Change). This hypothesis is supported by Fig. 6: The change in the slope of the GT regression line coincides remarkably with a sudden significance between RC-2 and GT-2.

Note that in reality, these three hypotheses may not be mutually independent—that is, it could be a combination of more than one that causes this jump in correlation. Note also that we are certainly not limited to these three possible explanations.

We stress the fact that these hypotheses are considerably unrefined and are partially visually derived. Climate oscillators are not the only factors which influence the relationship between CO_2 and temperature. Kaufmann et al.

(1991) and others have discussed the potential warming or cooling effects of tropospheric aerosol activity, which influences the net RC. Future research efforts need to decompose aerosol-related effects from RC to evaluate its significance for GC results. Recent reports (e.g., Schiermeier 2010) highlight our lack of understanding of the impacts of aerosols as one of four major holes in climate science.

6 Concluding remarks

6.1 Limitations

One caveat in general with GC lies in the outside variable factor. If *X* seems to cause *Y* but is simply highly correlated with *Z*, which actually causes *Y*, then it is possible to incorrectly designate *X* as the causal influence. In climate applications, this type of error might be complex and hard to detect.

Another possible limitation in using classic bivariate GC *F* test, more specific to this work, is that it follows the assumption that the predictand should be normally distributed. We do not meet the normality assumption with RC at any level of differencing. This could potentially affect our *p* values, especially those obtained in GT→RC models, where RC is the predictand.

Table 4 Two forecasting models are built in each direction (RC-2→GT-2 and vice versa) using a lag 3 GC model

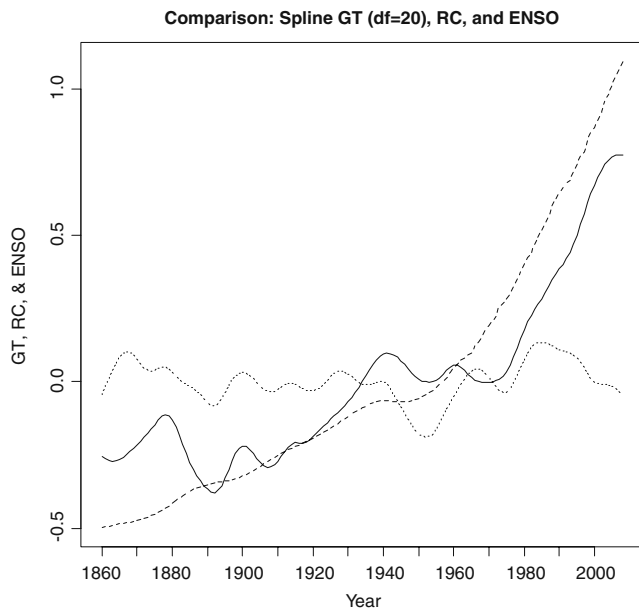| Model	R^2 holdout	MAPE	Maximum APE
1. $\hat{GT}$ = PastGT	0.722	0.215	1.431
2. $\hat{GT}$ = PastGT + PastRC	0.767	0.186	0.987
3. $\hat{RC}$ = PastRC	0.495	0.005	0.012
4. $\hat{RC}$ = PastRC + PastGT	0.514	0.005	0.011

Fig. 5 The *dashed line* represents RC (with its mean subtracted for better comparison), while the *solid line* represents a *df*=20 smoothed spline of GT. The *dotted line* is the ENSO index, smoothed with a spline and scaled for display purpose. This figure shows how the 20-year smoothing (with only some values of CO_2 used) of ice core CO_2 affects the correlation structure with time: It induces an early strong negative correlation between RC and GT
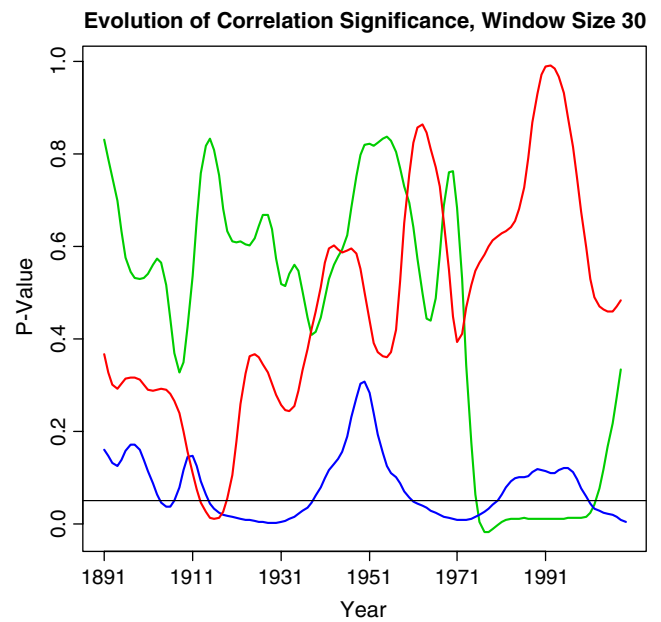
Fig. 6 Forward moving Pearson correlation significance (window size 30) measure for GT-2 and RC-2 (*solid green line*), GT-1 and ENSO-1 (*blue line*), and RC-2 and ENSO-2 (*red line*). The *horizontal line* represents a *p* value of 0.05. All three correlation-significance indices are plotted as splines with 30° of freedom. An index dropping below the horizontal 0.05 threshold indicates a significant correlation

The conditional heteroskedasticity found in RC-2 may also pose a problem to our GC testing. This non-constant variance is a result of smoothing in the early ice core CO_2 data, where many annual values are interpolated. The change in data source may have some effect on our GC tests and/or our results in Fig. 6. We do not study the effect of using smoothed averages between 1860 and 1959. In general, such imputation or smoothing has the effect of reducing noise variance; hence, *p* values of hypothesis tests tend to be recorded as lower than their true values.

In fact, if we look at Fig. 4, we note that after a latest-window size of 50, there is a sudden jump in most of the *H* trend lines. Perhaps not coincidentally, the jump occurs when the latest 10 ice core-derived RC values are added to the test (1949–1958). As implied in the previous paragraph, the interpolation of many of the ice core values could have the effect of artificially increasing the significance of the *F* tests. Thus, perhaps the larger F_1 values at latest-window sizes 60 and above are not a result of GC but a result of the interpolated ice core values. These two effects, then, may be confounded.

One final limitation concerns the number of RCUMGC tests we conduct. Since we are conducting 224 such tests and 57 *p* values significant at 0.05 (39 at 0.01), the standard problems of false positives and false negatives associated with multiple testing will be present. These can be controlled using standard procedures; see for example Benjamini and Hochberg (1995) for a method of controlling the false discovery rate. This same limitation applies to the forward cumulative tests.

Table 5 ENSO, GT, and RC Granger tests

Granger tests on lags 1–7 for entire length indices (all years available)

[a] Indicates significance at 0.01
[b] Indicates significance at 0.05
[c] Indicates significance at 0.10

	GT-1→ENSO-1		ENSO-1→GT-1		RC-2→ENSO-2		ENSO-2→RC-2	
	F test	*p* value	*F* test	*p* value	*F* test	*p* value	*F* test	*p* value
Lag1	2.436	0.121	20.253	1.39e−05[a]	13.224	3.8e−04[a]	18.804	2.71e−05[a]
Lag2	0.600	0.550	10.620	5.06e−05[a]	4.611	0.012[b]	4.560	0.012[b]
Lag3	0.531	0.662	6.167	0.001[a]	2.810	0.042[b]	3.079	0.030[b]
Lag4	0.654	0.625	4.892	0.001[a]	2.195	0.073[c]	4.936	0.001[a]
Lag5	0.543	0.743	3.435	0.006[a]	2.395	0.041[b]	2.763	0.021[b]
Lag6	0.332	0.919	2.735	0.016[b]	1.249	0.286	1.910	0.084[c]
Lag7	0.328	0.940	2.437	0.022[b]	0.932	0.484	1.675	0.121

6.2 Implications

We can develop a list of key implications from this work:

1. RC does seem to Granger cause GT, as demonstrated by the RCUMGC test. The forecast procedure adds support to this.
2. There is a sudden jump in correlation significance between RC and GT beginning in the 1970s—in Section 5.4, we propose several competing hypotheses for this phenomenon. It is possible that no one of these hypotheses stands alone.
3. There may be one or more variables that Granger cause GT besides RC, including but not limited to ENSO. An investigation of such variables and their relation to GT might be pursued in future research.
4. The data (particularly RC-2) do not meet the assumptions necessary to apply the chosen GC F tests, and so to further substantiate our results, we may have to investigate a test that takes our data limitations into consideration.

References

Alley RB, Marotzke J, Nordhaus WD, Overpeck JT, Peteet DM, Pielke RA Jr, Pierrehumbert RT, Rhines PB, Stocker TF, Talley LD, Wallace JM (2005) Abrupt climate change. Science 299 (5615):2005–2010

Attanasio A, Triacca U (2010) Detecting human influence on climate using neural networks based Granger causality. Theor Appl Climatol. doi:10.1007/s00704-010-0285-8

Barnett TP, Pierce DW, Hidalgo HG, Bonfils C, Santer BD, Das T, Bala G, Wood AW, Nozawa T, Mirin AA, Cayan DR, Dettinger MD (2008) Human-induced changes in the hydrology of the Western United States. Science 319(5866):1080–1083

Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Stat Soc B 57(1):289–300

Committee on Abrupt Climate Change, National Research Council (2002) Abrupt climate change. National Academy, Washington

Elsner J (2006) Evidence in support of the climate change— Atlantic hurricane hypothesis. Geophysical Research Letters 33:L16705

Elsner J (2007) Granger causality and Atlantic hurricanes. Tellus 59 (4):476–485

Graham NE (1994) Decadal-scale climate variability in the tropical and North Pacific during the 1970s and 1980s: observations and model results. Clim Dyn 10(3):135–162

Granger CWJ (1969) Investigating causal relations by econometric models and cross-spectral methods. Econometrica 37(3):424–438

Hacker RS, Hatemi-J A (2006) Tests for causality between integrated variables using asymptotic and bootstrap distributions: theory and application. Appl Econ 38(13):1489–1500

Hasselmann K (1979) On the signal-to-noise problem in atmospheric response studies. In: Shaw DB (ed) Meteorology over the Tropical Oceans. Roy al Meteorological Society, London, pp 251–259

Hegerl GC, von Storch H, Hasselmann K, Santer BD, Cubasch U, Jones PD (1996) Detecting greenhouse-gas-induced climate change with an optimal fingerprint method. J Clim 9:2281–2306

Hegerl GC, Hasselmann K, Cubasch U, Mitchell JFB, Roeckner E, Voss R, Waszkewitz J (1997) Multi-fingerprint detection and attribution analysis of greenhouse gas, greenhouse gas-plus-aerosol and solar forced climate change. Climate Dyn 13:613–634

Hiemstra C, Jones JD (1994) Testing for linear and non-linear Granger causality in the stock price-volume relation. J Finance 49 (5):1639–1664

Kaufmann RK, Stern DI (1997) Evidence for human influence on climate from hemispheric temperature relations. Nature 388:39–44

Kaufmann YJ, Fraser RS, Mahoney RL (1991) Fossil fuel and biomass burning effect on climate—heating or cooling? J Clim 4:578–588

Kutner M, Nachtsheim C, Neter J, Li W (2004) Applied linear statistical models. McGraw-Hill, New York

Kwiatkowski D, Phillips PCB, Schmidt P, Shin Y (1992) Testing the null hypothesis of stationarity against the alternative of a unit root. J Econom 54:159–178

Lozano A, Naoki A, Liu Y, Rosset S (2009) Grouped graphical Granger modeling methods for temporal causal modeling. The 15th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, Paris, France.

Lozano A, Li H, Niculescu-Mizil A, Liu Y, Perlich C, Hosking J, Abe N (2009) Spatial-temporal causal modeling for climate change attribution. The 15th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, Paris, France.

Meehl GA, Arblaster JM, Tebaldi C (2007) Contributions to natural and anthropogenic forcing to changes in temperature extremes over the United States. Geophys Res Lett 34:L19709

Mosedale TJ, Stephenson DB, Collins M, Mills TC (2006) Granger causality of coupled climate processes: ocean feedback on the North Atlantic Oscillation. J Climate 19(7):1182–1194

Myhre G, Highwood EJ, Shine K, Stordal F (1998) New estimates of radiative forcing due to well mixed greenhouse-gases. Geophys Res Lett 25(14):2715–2718

Ropelewski CF, Halpert MS (1986) North American precipitation and temperature patterns associated with the El Nino/Southern Oscillation. Mon Weather Rev 114:2352–2362

Salvucci GD, Saleem JA, Kaufmann RK (2002) Investigating soil moisture feedbacks on precipitation with tests of Granger causality. Adv Water Resour 25(8–12):1305–1312

Santer BD, Wigley TML, Jones PD, Schlesinger ME (1991) Multivariate methods for the detection of greenhouse-gas-induced climatic change. In: Schlesinger ME (ed) Greenhouse-gas-induced climatic change: a critical appraisal of simulations and observations. Elsevier, Amsterdam, pp 511–536

Santer BD, Wigley TML, Jones PD (1993) Correlation methods in fingerprint detection studies. Climate Dyn 8:265–276

Schiermeier Q (2010) The real holes in climate science. Nature 463:284–287

Shapiro SS, Wilk MB (1965) An analysis of variance test for normality (complete samples). Biometrika 52(3/4):591–611

Smirnov DA, Mokhov II (2009) From Granger causality to long-term causality: application to climactic data. Phys Rev E. 80(1) doi:10.1103/PhysRevE.80.016208

Solomon S, Qin D, Manning M, Chen Z, Marquis M, Averyt KB, Tignor M, Miller HL, IPCC (2007) Summary for policymakers. In: Climate change 2007: the physical science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.

Stern DI, Kaufmann RK (1999) Econometric analysis of global climate change. Environ Modell Softw 14(6):597–605

Stigler S (2005) Correlation and causation: a comment. Perspect Biol Med 48:S88–S94

Sun L, Wang M (1996) Global warming and global dioxide emission: an empirical study. J Environ Manage 46:327–343

Toda HY, Yamamoto T (1995) Statistical influence in vector autoregressions with possibly integrated processes. J Econom 66(1–2):225–250

Tol RSJ, De Vos AF (1993) Greenhouse statistics—time series analysis. Theor Appl Climatol 48:63–74

Tol RSJ, De Vos AF (1998) A Bayesian statistical analysis of the enhanced greenhouse effect. Clim Change 38:87–112

Triacca U (2005) Is Granger causality analysis appropriate to investigate the relationship between atmospheric concentration of carbon dioxide and global surface air temperature? Theor Appl Climatol 81:133–135

Zivot E, Wang J (2002) Modeling financial time series with S-PLUS. Insightful, Seattle