Research Article Environmetrics

Received: 5 May 2015, Revised: 23 October 2015, Accepted: 24 October 2015, Published online in Wiley Online Library: 10 December 2015

(wileyonlinelibrary.com) DOI: 10.1002/env.2373

Time series modeling of paleoclimate data

James E. H. Davidson[a]*, David B. Stephenson[a] and Alemtsehai A. Turasie[b]

This paper applies time series modeling methods to paleoclimate series for temperature, ice volume, and atmospheric concentrations of CO_2 and CH_4. These series, inferred from Antarctic ice and ocean cores, are well known to move together in the transitions between glacial and interglacial periods, but the dynamic relationship between the series is open to question. A further unresolved issue is the role of Milankovitch theory, in which the glacial/interglacial cycles are correlated with orbital variations. We perform tests for Granger causality in the context of a vector autoregression model. Previous work with climate series has assumed nonstationarity and adopted a cointegration approach, but in a range of tests, we find no evidence of integrated behavior. We use conventional autoregressive methodology while allowing for conditional heteroscedasticity in the residuals, associated with the transitional periods.

Keywords: paleoclimate; ice cores; stationarity; Granger causality; Milankovitch cycles; VAR modeling

1. INTRODUCTION

Data gathered from cores drilled in the Antarctic ice cap have played a prominent role in the debate on the causes of climate change. Time series collected by the European Project for Ice Coring in Antarctica (EPICA) (Jouzel *et al.*, 2007, Lüthi *et al.*, 2008) now extend back 800,000 years (800 kyr) BP. Measurements include the concentrations of carbon dioxide and methane in air trapped in bubbles in the ice, and the amount of deuterium in the ice as a proxy for ocean temperature. A series relating to the same paleological period but collected from a different source is the oxygen isotope concentration record from ocean cores (Imbrie, 1984; Bassinot *et al.* 1994, inter alia), which provides a proxy for global ice volume and therefore an independent record of climatic change over the late Pleistocene epoch.

Theoretical work on the interpretation of these observed variations focuses on the affect of the Milankovitch orbital cycles on the intensity of solar radiation reaching the Earth in different periods (Paillard, 2001). Three sources of variation identified are the eccentricity of the Earth's orbit, the obliquity or tilt of the Earth's axis of rotation with respect to the orbital plane, and the precession, or orientation, of the rotational axis. These orbital cycles provide a basis for explaining the glacial cycles, yet there are unresolved puzzles with this hypothesis. The eccentricity and precession cycles have varied substantially in magnitude, reaching a minimum at around 400 kyr BP when the Earth's orbit was nearly circular, but also a time when the glacial cycle was as pronounced as at any period in the record.

The aim of the present paper is to use time series techniques to model the relationships between these series. A leading ambition is to test the timing of glacial changes in a way that may throw light, in particular, on the interactions of CO_2 and CH_4 concentrations and the variations in temperature and glaciations. It has been observed (Caillon *et al.*, 2003) that at least in one specific episode (termination III), the changes in the temperature record appear to lead those in CO_2 concentration by an order of 800 ± 200 years. Fischer *et al.* (1999) estimated that increases in CO_2 lagged temperatures in the Vostok record by 600 ± 400 years at the start of the last three terminations. Mudelsee (2001) undertakes a parabolic regression analysis to determine phase relations in the Vostok data and finds a lag of 1300 years of CO_2 behind the temperature proxy.

Our object is to test causal relations between these series. Causation of a time-series variable Y by another variable X was defined by Granger (1969) as the phenomenon that the information contained in X can be used to improve forecasts of future values of Y. Two-way Granger causation could imply that X and Y are both driven by some possibly unobserved third factor. One-way Granger causation, on the other hand, where the property does not obtain when the roles of the variables are interchanged, is evidence in favor of (although strictly does not imply) a direct causal mechanism. A recent study by Kaufmann and Juselius (2013) reports a cointegration modeling exercise involving 10 series, including the four that we study, for the period from 391 kyr BP. MacMillan and Wohar (2013) report regressions for temperature and CO_2, covering the full 800 kyr period with conclusions comparable with our own. However, a formal statistical test of Granger causality applied to the whole span of available data in a four-equation system has not previously been attempted. In common with MacMillan and Wohar (2013), we assume stationary series (an approach carefully justified in Section 3), and our approach involves constructing a full-rank dynamic model within which the restrictions of interest can be defined and tested.

* *Correspondence to: James E. H. Davidson, Department of Economics, University of Exeter Business School, Rennes Drive, Exeter EX4 4PU, U.K. E-mail: james.davidson@exeter.ac.uk*

a *University of Exeter, Exeter, U.K.*

b *University of Witwatersrand, Johannesburg, South Africa*

The paper is organized as follows. Section 2 describes the data set. Section 3 reports a range of tests to establish the stationarity characteristics of the series. Section 4 then develops the vector-autoregressive modeling framework, Section 5 reports and discusses the results, and Section 6 contains concluding remarks. Some technical details are given in two appendices. Additional details of the estimation procedure and results are not reported here for reasons of space but are available in an online Supporting Information file.

2. THE DATA SET

The series for temperature and gaseous concentrations are taken from the website of the US National Oceanic and Atmospheric Administration National Climatic Data Center.[†] The raw data on temperature take the form of 5800 core samples, for each of which are reported the depth of the sample, the imputed date, the deuterium content of the ice, and the imputed temperature. The temperature proxy is obtained from the assumption that water vapor from evaporation contains less deuterium (the heavier isotope of hydrogen) than the ocean, and this effect is more pronounced in colder climatic periods (Jouzel *et al.*, 2003). The samples are dated by methods such as counting the seasonal snowfall variations, as well as correlation with other dated series (Parrenin *et al.*, 2007). While the mean interval between observations is only 138 years, the older observations are considerably sparser than the recent ones. We should emphasize that equating these measurements with "global" mean temperature, while tempting, is necessarily speculative. Most ice core scientists argue that the deuterium record tells us only about local condensation temperatures at the site during the time when snow is accumulating (personal communication, Louise Sim, British Antarctic Survey).

The CO_2 and CH_4 measurements number 1096 and 2104, respectively, covering the same 800,000-year period. While the mean interval between CO_2 measurements is 729 years, over 40 of the intervals exceed 2000 years. The ice coverage proxy series measures the ratio of $\delta^{18}$to δ^{16} oxygen isotopes obtained from the shells of planktonic foraminifera in a variety of deep sea cores, assembled by Lisiecki and Raymo (2005).[‡] These data are a series of 699 observations covering the period back to 798 kyr BP, reported for 1 kyr intervals except for some of the earliest periods, for which the intervals are 2 kyr. Unlike the EPICA Antarctic temperature record, these measurements are not expressed in units directly related to temperature or ice coverage, but the changes in this series correlate closely with (inverse) temperature and might be treated as an alternative measure of climatic conditions.

A standard requirement for time series modeling is a set of observations regularly spaced in time. A feasible approach in this case, also adopted by Kaufmann and Juselius (2013), is to impute values to a suitable set of equally spaced dates by linear interpolation. The natural interval to match the available EPICA observations is 1 kyr, resulting in 798 imputed data points. The ice volume series has also been interpolated, but this effects only a few of the early observations at 2 kyr intervals.

A feature of the data series so constructed is a quite pronounced asymmetry, with extreme variations more evident in the interglacial than in the glacial periods. Such data features are not appropriate to a linear Gaussian modeling framework, which assumes symmetric shocks. Nonlinear modeling is implemented most simply by forming monotonic transformations of the data, and the transformations applied are listed in Table 1 where in each case, X denotes the original measurement. Gas concentrations are conventionally reported in parts per million, but parts per 10^4 is a choice of unit that gives all the series comparable orders of magnitude. For comparison, the increase in CO_2 concentration over the period 1970–2010 is about 0.6 parts per 10^4. The units of the ice proxy have no natural interpretation, but their sample range is approximately two units, for comparison with the sample range of temperatures of 14 °C.

The object of the logarithmic transformations is twofold, to avoid dependence of the model parameterization on units of measurement, and to render the data distributions roughly symmetric and to achieve best agreement between the observations and our linear parametric model.[§] We emphasize that the model does not attempt to embody the functional form of relations derived from atmospheric physics, which would scarcely be feasible in the present framework, although it is a convenient fact that power law relations are linearized exactly in logarithms. The transformed series on which we conduct all subsequent analyses are shown in Figure 1.[¶] Note that time advances from left to right in our plots and hence the dates, expressed in kyr BP, run in decreasing order.

3. STATIONARITY OF THE TIME SERIES

In time series analysis, inference techniques depend critically on whether the series are judged to be stationary or to feature stochastic trends. In linear modeling, unit autoregressive roots are the usual representation of nonstationary behavior, while stationarity implies mean reversion and, hence, autoregressive roots in the stable region. While relations between stationary variables are measured by correlations, relations in nonstationary data imply common stochastic trends in the series, following from the existence of fewer unit roots than variables. This modeling approach is known as cointegration. Recent econometric work on climatological series from the more recent past has tended to assume nonstationarity, for example, Kaufmann and Juselius (2013), Kaufmann and Stern (2002), Kaufmann *et al.* (2006, 2010), Gay-Garcia *et al.* (2009), and Mills (2010).

However, there is little visual evidence of unit root behavior in the data in Figure 1. Notwithstanding the pronounced alternation of glacial and interglacial episodes, the series are confined within fairly well-defined limits and appear mean reverting. A range of tests for integration order have been performed and are reported in Appendix A. The so-called "I(1)" null hypothesis, of a unit root, is decisively rejected by all the tests. The "I(0)" null hypothesis, that of stationarity and weak dependence, is in general not rejected. Our findings confirm the results reported by MacMillan and Wohar (2013) in respect of the temperature and CO_2 series.

[†] http://www.ncdc.noaa.gov/paleo/pubs/ipcc2007/fig63.html
[‡] http://www.lorraine-lisiecki.com/stack.html
[§] The latter effect is illustrated in Figure S2 of the online Supporting Information.
[¶] The original measurements are shown in Figure S1 of the online Supporting Information.

Table 1. Data transformations

Temperature:	$\log(X + 16)$
CO_2 :	$\log(X/100)$
CH_4 :	$\log(X/100)$
Ice:	$-\log(8 - X)$

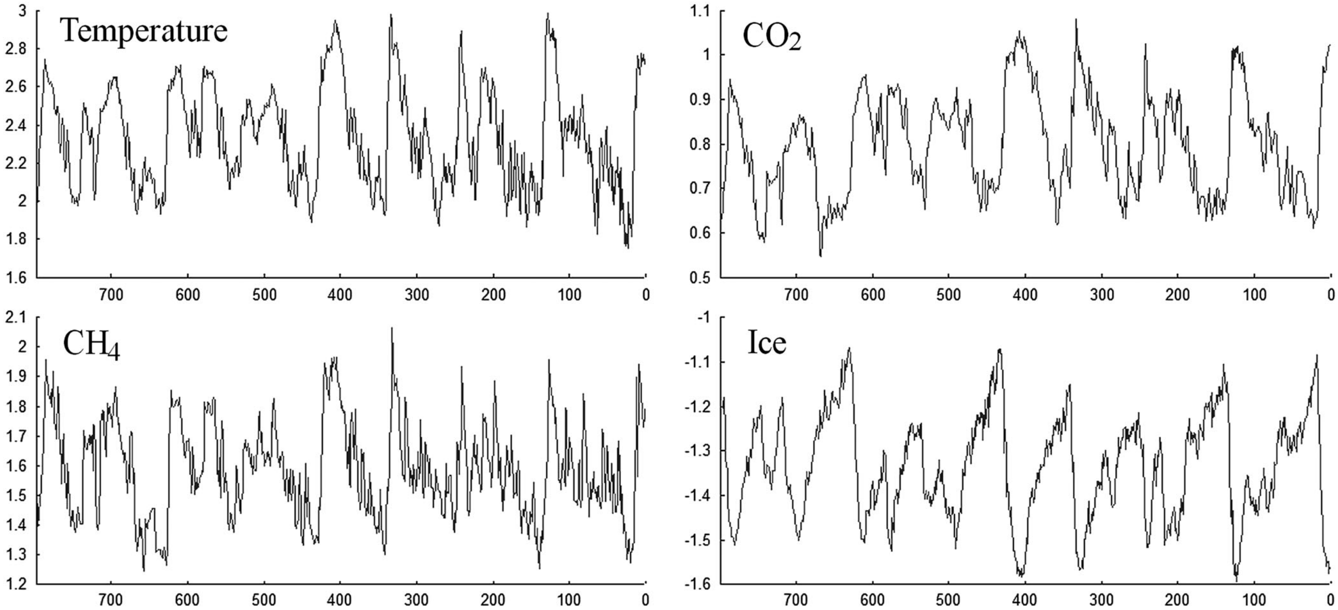

Figure 1. The data set, after transformation

These findings are at first sight difficult to square with the cointegration modeling approach of the previously cited studies. However, the puzzle is resolved by considering the differing time-spans of data analyzed by different authors. This is illustrated rather clearly by Figure 2, which shows the Phillips–Perron statistic, which tests the null hypothesis of a unit root, computed for a range of spans starting with the most recent 50 observations and rolling backwards to include the full sample. What appears as a nonstationary process when only a portion of a single glacial cycle is observed is revealed as stationary as the sample is extended. MacMillan and Wohar (2013) also emphasize this important point, by contrasting with a very much shorter 160-year record, from 1850.

There remains the problem of reconciling stationarity with pronounced cyclical behavior. Because the series appear to move within fixed upper and lower bounds, "I(1)" (a simple unit root) appears an improper null hypothesis. However, such bounds do not rule out the possibility of stochastic trending behavior within their confines. A bounded random walk model with just these features has been studied by Cavaliere (2005) and Granger (2010). Cavaliere (2005) considers the case

$$X_t = \theta + Y_t$$

$$Y_t = Y_{t-1} + \varepsilon_t + \underline{\xi}_t - \overline{\xi}_t$$

where ε_t is a weakly dependent process and $\underline{\xi}_t$ and $\overline{\xi}_t$ are non-negative "regulator processes" such that $\underline{b} - \theta \leqslant Y_t \leqslant \overline{b} - \theta$ and hence X_t is confined within the interval $[\underline{b}, \overline{b}]$. If the bounds $\underline{b}$ and $\overline{b}$ are known, it is possible to consider a test in which the bounded random walk forms the null hypothesis. Cavaliere and Xu (2014) derive a test based on the distribution of the augmented Dickey–Fuller (ADF) statistic under the assumption of a bounded random walk. The problem is to choose values for $\underline{b}$ and $\overline{b}$. The tighter these are chosen, the further the null distribution is shifted to the left and the smaller the critical values, and it therefore makes sense to choose the extreme case, setting the bounds to the recorded maxima and minima of the series. A rejection in this case cannot be contradicted by different choices. Critical values for the ADF statistic are calculated using a Monte Carlo procedure calibrated by the maximum, minimum, and initial observations, so that a different tabulation is required for each series. The calculations are explained in more detail in Section 2 of the Supporting Information. Table 2 shows these critical values, and also the ADF statistics from Table A.1 of Appendix A, for comparison. The bounded unit-root null hypothesis is rejected in each case at the 1% level, in favor of the stationary alternative.

Another view of the evidence is provided by Table 3, which shows the results of Johansen (1991) tests for reduced rank in the context of a vector autoregression (VAR), with lag order 6 selected by the Akaike criterion. These null hypotheses all entail the assumption that the

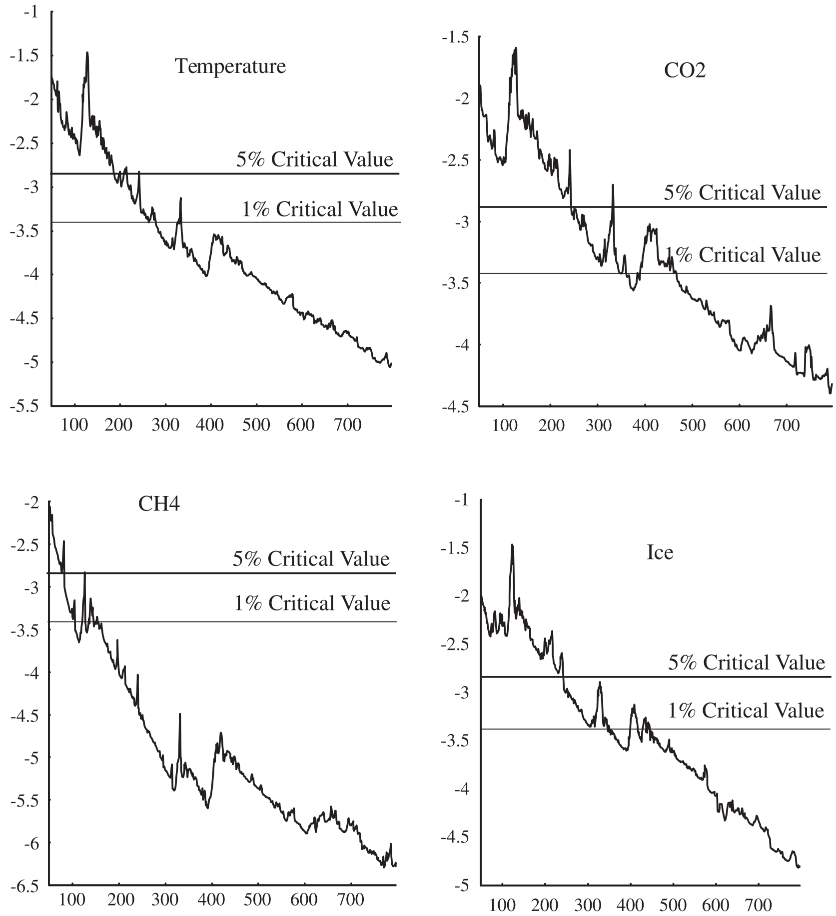

Figure 2. Phillips–Perron statistics for the N most recent observations, for $50 \leqslant N \leqslant 798$

Table 2. Unit root tests subject to barriers

	Augmented Dickey–Fuller statistic	5% critical value	1% critical value
Temperature:	−4.73	−3.97	−4.58
CO_2 :	−4.23	−3.61	−4.31
CH_4 :	−6.08	−4.230	−4.731
Ice:	−5.81	−5.345	−5.775

Table 3. Johansen tests for cointegrating rank, six lags

	Maximum eigenvalue test		Trace test	
Rank	Statistic	p-value	Statistic	p-value
0	75.44	< 0.01	187.6	< 0.01
1	55.08	< 0.01	112.9	< 0.01
2	36.79	< 0.01	57.10	< 0.01
3	20.31	< 0.01	20.31	< 0.01

series are I(1)[**] and the rejections at the 1% significance level further reinforce the conclusion of stationarity. Kaufmann and Juselius (2013) report a similar finding of a full rank system, although in their case of dimension 10.

[**]In terms of the representation in Section 4, "full rank" here means that the matrix **E** in Equation (3) has full rank. Reduced rank of this matrix implies that the model features unit roots and generates nonstationary series.

Table 4. Semiparametric estimates of the long-memory parameter

	$\hat{d}$ (Local Whittle)			$\hat{d}$ (GPH)			Bias test	
Bandwidth	$[T^{0.5}]$	$[T^{0.6}]$	$[T^{0.7}]$	$[T^{0.5}]$	$[T^{0.6}]$	$[T^{0.7}]$	Statistic	p-value
Temp.	0.271	0.741	0.949	0.0213	0.475	0.639	3.26	0.001
CO_2	0.586	0.762	0.913	0.447	0.629	0.800	4.13	0
CH_4	0.184	0.525	0.750	0.036	0.272	0.582	2.02	0.021
Ice	0.260	0.748	1.18	0.059	0.446	0.787	1.96	0.025

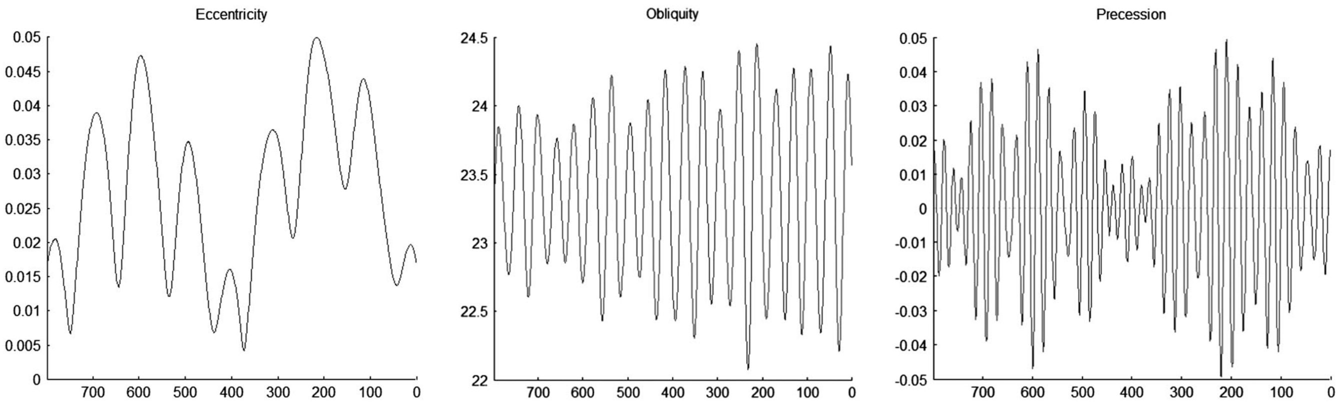

Figure 3. Orbital variables

If the I(1) and I(0) hypotheses were both to be rejected, which is one way to interpret the mixed results in the tests of Appendix A, the data might possibly be characterized as fractionally integrated, or I(d) for $0 < d < 1$. An I(d) process exhibits long-range dependence with nonsummable autocovariance sequence, d being the parameter measuring the rate of divergence of the spectral density at the origin. Under this hypothesis, d may be estimated by a semiparametric method such as the Geweke and Porter-Hudak (1983) log-periodogram regression (GPH), or "local Whittle" maximum likelihood (Kunsch, 1987, Robinson, 1995). Estimates by these methods, with alternative bandwidths, are shown in Table 4. A bandwidth of $O(T^{1/2})$ was proposed by Geweke and Porter-Hudak, whereas broader bandwidth choices are optimal on an mean squared error criterion (Hurvich *et al.*, 1998). The bias test (Davidson and Sibbertsen, 2009) is asymptotically normal under the null hypothesis of a pure fractionally integrated process.

What is notable in the present case is how sensitive these estimates are to both estimator and bandwidth choice. Because bias due to the omission of short-run dynamics is known to exist and is very sensitive to bandwidth choice, it seems certain that these estimates are, in varying degrees, biased upwards. These are predictable results if the autocovariance functions are summable but not absolutely summable, and Figures S4, S5 and S6 of the Supporting Information are of interest in this connection. Stationary fractional processes have a characteristically self-similar appearance that these series lack. Moreover, a linear representation of long-range dependence, as represented by either a fractional integral or fractional Gaussian noise representation, implies that the autocorrelations must decline monotonically beyond a certain point and so cannot provide an adequate description of these data series.

4. VECTOR AUTOREGRESSIVE ANALYSIS

Global stationarity, specifically ruling out unit roots, does not rule out systematic changes in central tendency because of periodic effects. Economic series of monthly or quarterly observations commonly show a seasonal pattern, whereas here, we hypothesize cyclical shifts because of Milankovitch cycles (Berger 1978, Imbrie *et al.*, 1992, Loutre 2002). These variations in the eccentricity, obliquity, and precession of the Earth's orbit can be reconstructed for the relevant epochs using a public domain computer program (Paillard *et al.*, 1996). The basic series so obtained are shown in Figure 3. These series are included in our descriptive model of the paleoclimate as non-stochastic explanatory variables.[††]

The model is formulated as follows. Letting L denote the lag operator, such that $Ly_t = y_{t-1}$ for a dated variable y_t, let $\boldsymbol{A}(L) = \sum_{j=0}^{q} \boldsymbol{A}_j L^j$ $(m \times m)$ be a finite-order matrix polynomial with $\boldsymbol{A}(0) = \boldsymbol{I}_m$. Also, let $\boldsymbol{\Lambda}(L) = \sum_{j=0}^{r} \boldsymbol{\Lambda}_j L^j$ denote a $m \times p$ polynomial, $\boldsymbol{u}_t$ a m-vector of zero-mean random shocks, and $\boldsymbol{d}_t$ $(p \times 1)$ the vector of known orbital variables. Assuming the variables have means of 0, let the evolution of the m-vector $\boldsymbol{y}_t$ of variables generated by the geophysical system be described by the vector autoregressive (VAR) equation

$$\boldsymbol{A}(L)\boldsymbol{y}_t = \boldsymbol{\Lambda}(L)\boldsymbol{d}_t + \boldsymbol{u}_t \tag{1}$$

[††]It is possible using the same software to compute series for imputed insolation at different latitudes. However, there is some uncertainty as to which of these various insolation series is relevant. We choose to let the dynamic model determine directly how orbital variations influence the target variables.

Here, m is unknown, as are the identities of many of the elements of $\boldsymbol{y}_t$ that are in any case unobserved. Let $m = 4+m_Z$ and $\boldsymbol{y}_t = (\boldsymbol{x}_t', z_t')'$ where $\boldsymbol{x}_t$ (4×1) denotes the observables and z_t $(m_Z \times 1)$ the unobserved variables. As to the possible identities of these variables, note that Kaufmann and Juselius (2013) present a 10-equation model comparable with ours and including our four variables, using measurements available for a shorter time span. There are of course many other phenomena that might appear in (1) but evade observation entirely. However, in Appendix B, we show how z_t can be eliminated, and the reduced four-dimensional system for $\boldsymbol{x}_t$ approximated by a VAR of the form

$$\boldsymbol{B}(L)\boldsymbol{x}_t = \boldsymbol{\Upsilon}(L)\boldsymbol{d}_t + \boldsymbol{\varepsilon}_t \tag{2}$$

There is a temptation to see (2) as a "mis-specified" version of (1) with omitted variables, but its parameters are functions of those of the big system and simply require correct interpretation. They represent partial knowledge of the climate system given missing information, but there are some testable restrictions on $\boldsymbol{B}(L)$ and $\boldsymbol{\Upsilon}(L)$ that reflect restrictions of interest on $\boldsymbol{A}(L)$ and $\boldsymbol{\Lambda}(L)$.

It is further shown in Appendix B that (2) can be expressed in the reparameterized form

$$\boldsymbol{C}(L)\Delta\boldsymbol{x}_t = \boldsymbol{D}(L)\Delta\boldsymbol{d}_t + \boldsymbol{E}(\boldsymbol{x}_{t-1} - \boldsymbol{\Pi}\boldsymbol{d}_{t-1}) + \boldsymbol{\varepsilon}_t \tag{3}$$

where $\Delta = 1 - L$ and $\boldsymbol{C}(L)$ and $\boldsymbol{D}(L)$ are derived lag polynomials, with $\boldsymbol{C}(0) = \boldsymbol{I}_4$ if $\boldsymbol{A}(0) = \boldsymbol{I}_m$ and $\boldsymbol{E} = -\boldsymbol{B}(1)$. Assuming $\boldsymbol{B}(1)$ nonsingular, the system possesses fixed relations of central tendency, defined by setting the exogenous variables to constants $\boldsymbol{d}$ and the shock vector to 0, of the form

$$\boldsymbol{x} = \boldsymbol{\Pi}\boldsymbol{d} \tag{4}$$

where $\boldsymbol{\Pi} = \boldsymbol{B}(1)^{-1}\boldsymbol{\Upsilon}(1)$. Equation (3) is the form of the system to be estimated. Relaxing the assumption that the data have means of 0, their locations are accounted for by including intercepts in these dynamic relations. This is observationally equivalent to placing intercepts in the steady-state relations and ensures that the equation residuals have unconstrained means of 0 in the sample.

$\boldsymbol{E}$ is often called the matrix of error correction coefficients, the off-diagonal elements establishing directions of Granger causation in the system. Granger causation of a variable x_{jt} by another variable x_{kt} might be characterized as correlation between x_{jt} and $x_{k,t-l}$ for $l > 0$ after partialing out the effects of other predictors, in particular, $x_{j,t-l}$. The condition would be technically fulfilled by a partial correlation of Δx_{jt} with lagged Δx_{kt}, implied by non-zero off-diagonal elements of the terms in $\boldsymbol{C}(L)$, but such effects are transitory, representing influences on the rate of approach to a new level, but not on the new level itself. The causality relations of interest are those where Δx_{jt} is partially correlated with deviations of $x_{k,t-1}$ from its steady-state relations with the exogenous drivers in (4). It is only the coefficients of $\boldsymbol{E}$ that can tell us how far, for example, shocks to greenhouse gas concentrations have permanent effects on temperature or ice volume.

A useful feature of this model is that the structure is valid whether or not the system contains unit roots and is therefore robust to the assumption of stationarity. In nonstationary systems, there are fewer steady-state relations than variables, and $\boldsymbol{E}$ has reduced rank. In that case, steady-state paths are undefined, but cointegration means that linear combinations of these relations are trend free. However, in our set-up, $\boldsymbol{E}$ has rank 4 allowing (4) to define a steady-state path. The diagonal elements of $\boldsymbol{E}$ must be negative to reflect the mean reversion properties of the system.

By suitable choice of lag length, the model should return uncorrelated residuals. However, we find that clusters of larger than average shocks are observed around the glacial–interglacial transitions, as might be expected during periods of rapid climatic change. Taking account of such behavior is necessary to ensure valid inferences, and a natural model to explain volatility clustering is "generalized autoregressive conditional heteroscedasticity" or GARCH (Bollerslev, 1986). Letting $E_{t-1}\boldsymbol{\varepsilon}_t^2 = \boldsymbol{h}_t$ (4×1) denote the vector of conditional variances of the process $\boldsymbol{\varepsilon}_t$, the first-order multivariate GARCH equation takes the form

$$\boldsymbol{h}_t = \boldsymbol{\omega} + \boldsymbol{\alpha}\boldsymbol{\varepsilon}_{t-1}^2 + \boldsymbol{\beta}\boldsymbol{h}_{t-1} \tag{5}$$

where $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are 4×4 matrices. Letting $\hat{\boldsymbol{h}}_t$ denote the diagonal matrix in the elements $\boldsymbol{h}_t$, the shock process is represented as $\boldsymbol{\varepsilon}_t = \hat{\boldsymbol{h}}_t^{1/2}\boldsymbol{u}_t$ where by hypothesis $\boldsymbol{u}_t \sim$ i.i.d.$(\boldsymbol{0}, \boldsymbol{I}_4)$. The GARCH model requires the distribution of the shocks to be symmetric and, as noted previously, our decision to work with the logarithmically transformed variables ensures symmetry to the best approximation. The so-called "ARMA in squares" form of the GARCH model, with representation

$$\boldsymbol{\varepsilon}_t^2 = \omega + \boldsymbol{\delta}\boldsymbol{\varepsilon}_{t-1}^2 + \boldsymbol{w}_t - \boldsymbol{\beta}\boldsymbol{w}_{t-1} \tag{6}$$

where $\boldsymbol{\delta} = \boldsymbol{\alpha} + \boldsymbol{\beta}$ and $\boldsymbol{w}_t = \boldsymbol{\varepsilon}_t^2 - \boldsymbol{h}_t$, is a convenient form for estimation and has the benefit that the eigenvalues of $\boldsymbol{\delta}$ fix the covariance stationarity conditions for the volatility model. $\boldsymbol{\delta}$ and $\boldsymbol{\beta}$ are the parameters we estimate.

5. ESTIMATION RESULTS

The four-equation model consisting of (3) plus (6) was fitted to the data by Gaussian maximum likelihood.[‡‡] This estimator is robust to the form of the shock distribution and is consistent and asymptotically normally distributed under fairly mild regularity conditions; see, for example, Davidson (2000) for an account of the estimation theory supporting these assertions. The order of the polynomials $\boldsymbol{C}(L)$ and $\boldsymbol{D}(L)$ is in both cases set at two, on the basis of parsimony considerations and evidence from the estimation diagnostics. The coefficients

[‡‡]All calculations reported in this paper were made with the package Time Series Modelling 4 (Davidson, 2015) running under the Ox 7 programming system (Doornik, 2013).

Table 5. Estimated matrix $\boldsymbol{E}$ with robust standard errors in parentheses

	Temp	CO_2	CH_4	Ice
Temp	-0.184^{***} (0.037)	0.090 (0.058)	0.038 (0.038)	-0.138^{**} (0.055)
CO_2	0.047^{***} (0.012)	-0.107^{***} (0.019)	-0.053^{***} (0.014)	-0.048^{**} (0.019)
CH_4	0.127^{***} (0.031)	0.026 (0.058)	-0.316^{***} (0.034)	-0.123^{**} (0.048)
Ice	-0.028^{**} (0.012)	-0.033^{*} (0.020)	-0.010 (0.012)	-0.121^{***} (0.018)

Table 6. 95% confidence intervals for 1-kyr changes (columns) following unit deviations from steady state (rows)

(a)		ΔTemp	ΔIce
	CO_2	[−0.113, 0.974]	[−0.117, 0.004]
	CH_4	[−0.078, 0.243]	[−0.026, 0.010]
(b)		**ΔCO_2**	**ΔCH_4**
	Temp	[0.005, 0.015]	[0.030, 0.087]
	Ice	[−0.050, −0.006]	[−0.281, −0.037]

are otherwise estimated unrestrictedly. However, in the GARCH equation, we constrain the matrices $\boldsymbol{\delta}$ and $\boldsymbol{\beta}$ to be diagonal and hence rule out cross-equation volatility effects. The complete estimation outputs are reported, with commentary, in the Supporting Information. Here, we give the findings pertinent to our causality analysis, and Table 5 shows the estimates of the matrix $\boldsymbol{E}$ with the associated robust standard errors, and one, two, and three stars indicating statistical significance at the 10%, 5% and 1% levels, respectively, under asymptotic criteria.

The diagonal elements of this matrix are all significantly negative, indicating mean reversion as implied by the presumption that the system is globally stationary. Present interest focuses on the off-diagonal elements, which are (with change of sign) the sums of the relevant coefficients of $\boldsymbol{B}(L)$ in (2) and measure the total effect of these. If element e_{jk} (row j, column k) is nonzero, then variable j is changing whenever variable k deviates from the steady-state path defined by (4). The elements relating temperature to future CO_2 and CH_4 concentration, and ice volume to future CO_2, are significantly positive at the 1% level and that relating ice to future CH_4 is significant at the 5% level. On the other hand, the effects of CO_2 and CH_4 on future temperature are statistically insignificant at the 10% level, and the effect of CO_2 on future ice volume is insignificant at the 5% level.

Because the model is logarithmic, the coefficients are unitless, and the results are best evaluated by considering what they imply about actual changes in the specified units of measurement. These can be approximated at the points of sample means, according to the formulations in Table 1. Letting $X_1, \ldots, X_4$ denote measurements in the original units with sample means $\bar{X}_1, \ldots, \bar{X}_4$, the change ΔX_j following a unit shock to X_k can be approximated by

$$\Delta X_j = e_{jk} \frac{\bar{X}_j}{\bar{X}_k}$$

Given the appropriate interval estimate of e_{jk}, approximate Gaussian confidence intervals are constructed by the formula

$$\Pr\left(\Delta X_j \in \frac{\bar{X}_j}{\bar{X}_k}\left[\hat{e}_{jk} - z_{\alpha/2}s_{jk}, \hat{e}_{jk} + z_{\alpha/2}s_{jk}\right]\right) \approx 1 - \alpha \tag{7}$$

where $\hat{e}_{jk}$ is the point estimate from Table 5, s_{jk} its standard error, and z_a represents the usual Gaussian quantile such that $\Pr(z > z_\alpha) = \alpha$.

Table 6 presents scenarios showing the ranges of predicted changes over 1 kyr, with 95% confidence. The bounds in the table are computed from formula (7) setting $z_{0.025} = 1.96$. The sample means used in the formula can be found in Table S1 of the Supporting Information, although note that for temperature and ice, these are modified to incorporate the shifts of origin and sign in Table 1. Table 6(a) shows the ranges of predicted 1 kyr changes in temperature (°C) and ice following unit shocks (1 part per 10^4) to gas concentrations. Table 6(b) shows the corresponding predicted responses of gas concentrations to shocks to the climatic variables. In all the former cases, but none of the latter, 0 is included in the intervals. The results indicate positive causal effects of climate on subsequent atmospheric gas concentrations, but little evidence for the reverse causation. These conclusions may well reflect the evidence cited in the introduction on the timing of glacial transitions.

We next consider evidence on the role of the orbital cycles in driving glacial changes. Estimates of the matrix $\boldsymbol{\Pi}$ are shown in Table 7. The coefficients of precession are significant at the 5% level, but eccentricity plays no statistically significant role. Plots of the estimated

Table 7. Estimated matrix $\mathbf{\Pi}$, robust standard errors in parentheses

	Eccentricity	Obliquity	Precession
Temp	0.334 (3.87)	-0.960^{*} (0.606)	-34.14^{**} (13.95)
CO_2	-0.677 (1.91)	-0.368 (0.286)	-16.81^{**} (6.586)
CH_4	-1.733 (2.49)	-0.745^{*} (0.388)	-22.15^{**} (8.84)
Ice	-0.847 (2.04)	0.542^{*} (0.316)	18.31^{**} (7.376)

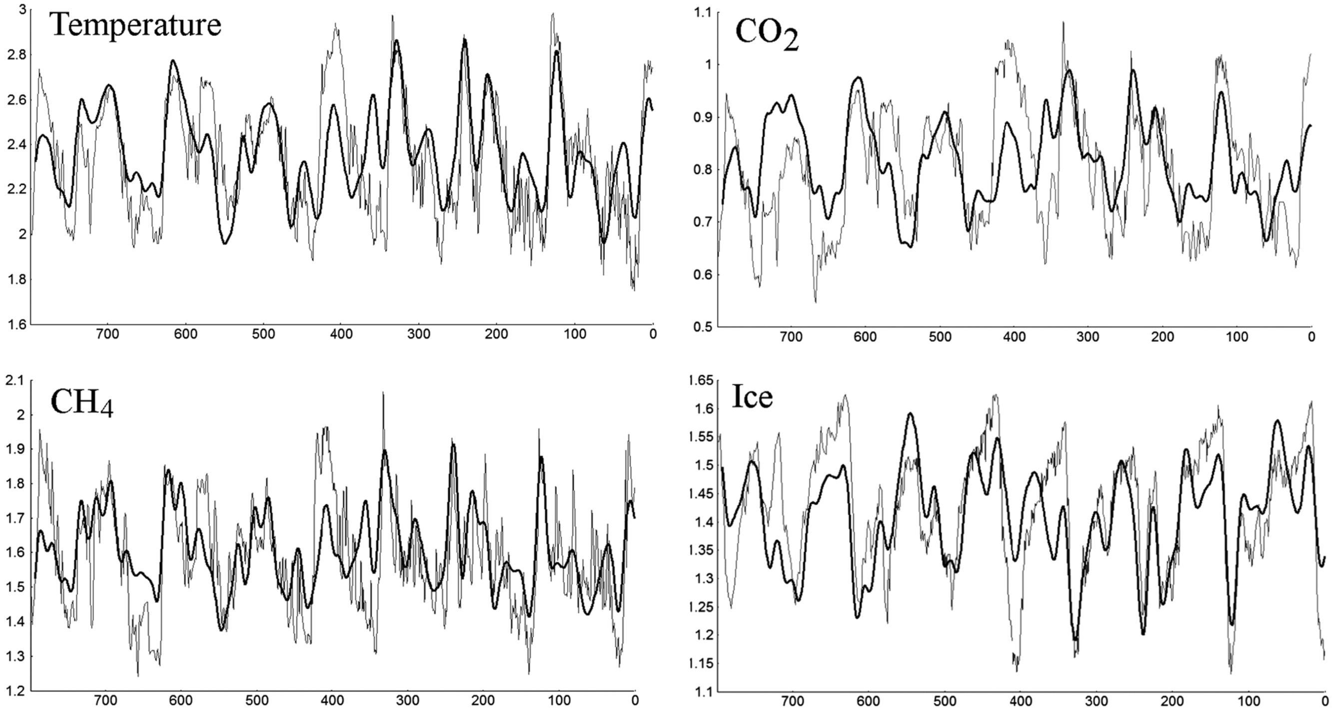

Figure 4. Solved model paths with disturbances set to 0 (thick lines). The observations from Figure 1 are shown for comparison

steady-state deviations, $\boldsymbol{x}_t - \mathbf{\Pi}\boldsymbol{d}_t$, are included in Figure S9 of the Supporting Information. These curves are dominated by the orbital cycles but not by the glacial cycles, showing that the steady-state relation between $\boldsymbol{x}_t$ and $\boldsymbol{d}_t$ does succeed broadly in tracking the ice ages, even though there are substantial higher frequency oscillations that need to be accounted for by the system dynamics.

The role played by the cycles is shown in Figure 4, which plots (broad lines) the solutions of the model when shocks are set to 0 throughout. Hence, it shows what, according to our model, Milankovitch theory predicts the glacial cycles to look like. Note that the plotted curves are the fitted linear combinations of nine periodic series, the current, and once and twice lagged values the three orbital variables in Figure 3. The figure also reproduces the series from Figure 1 for comparison (narrow lines). Over some historical periods, notably, between 300 and 50 kyr BP, the match between the series is close, illustrating clear empirical support for Milankovitch theory. But over other periods, such as that known as "marine isotope stage 11" between 450 and 350 kyr BP, the match is very poor. This is a period in which near-circularity of the Earth's orbit diminished the variations in eccentricity and precession (Figure 3) coincident with one of the longest and most pronounced interglacial episodes. The model can provide no clue as to what might account for this divergence, but it is worth observing that the model residuals (plotted in Figures S7 and S8 of the Supporting Information) show at most a modest structural break, or sustained change in predictability, over that period. The endogenous dynamics of the system appear able to account for much of the observed discrepancies. There is no corresponding call for additional forcing influences needed to make our model "work", in the sense of predicting variations over 1 kyr steps.

6. CONCLUDING REMARKS

In this paper, we have applied some conventional time series inference techniques to a set of paleoclimate series, with a view to a better understanding of the role of Milankovitch theory in glacial cycles, and also to see whether any conclusions can be drawn via these methods about the timing of changes in greenhouse gas concentrations and glacial/interglacial changes. The model shows that the endogenous dynamics of the climate system can be modeled quite effectively by a linear autoregressive scheme applied to the logarithmically transformed series, with allowance for local volatility effects in the transitions. Orbital variables (save eccentricity) do appear to play a statistically significant role in this explanation, according to the estimates of $\mathbf{\Pi}$ in Table 7. Some aspects of the Milankovitch story remain mysterious, however.

We have demonstrated significant Granger causation of greenhouse gas concentrations by prior climatic conditions. The reverse causation, of climate by prior changes in greenhouse gas concentration, remains unproven, reinforcing an inference that has been made previously about the timing of glacial–interglacial transitions, although it is the large uncertainty (reflected in the width of confidence intervals) that matters more to this conclusion than the proximity of point estimates to 0. However, we should draw attention to our earlier caveats about the interpretation of Granger causality. The hypotheses we would really like to test are on the elements of (1), and specifically $\boldsymbol{A}_{xx}(L)$ in (B.1) in Appendix B. Restrictions on $\boldsymbol{E}$ in (3) reflect these only under certain restrictions on the role of the unobservables, specifically, that $\boldsymbol{A}_{xz}(L)$ and/or $\boldsymbol{A}_{zx}(L)$ in (B.1) are restricted in an appropriate way. While a 0 element of $\boldsymbol{E}$ must imply a corresponding 0 in $\boldsymbol{A}_{xx}(1)$, the reverse implication does not hold. Our reported findings need to be interpreted in this light.

REFERENCES

Akaike H. 1969. Fitting autoregressive models for prediction. *Annals of the Institute of Statistical Mathematics* **21**:243–247.

Bassinot FC, *et al.* 1994. The astronomical theory of climate and the age of the Brunhes-Matuyama magnetic reversal. *Earth and Planetary Science Letters* **126**:91-108.

Berger AL. 1978. Long-term variations of daily insolation and quaternary climatic change. *Journal of the Atmospheric Sciences* **35**:2362–2367.

Bollerslev T. 1986. Generalized autoregressive conditional heteroscedasticity. *Journal of Econometrics* **31**:307–327.

Caillon N, Severinghaus JP, Jouzel J, Barnola J-M, Kang J, Lipenkov VY. 2003. Timing of atmospheric CO_2 and antarctic temperature changes across termination III. *Science* **299**:1728–1731.

Cavaliere G. 2005. Limited time series with a unit root. *Econometric Theory* **21**(5):907–945.

Cavaliere G, Xu F. 2014. Testing for unit roots in bounded time series. *Journal of Econometrics* **178**:259–272.

Davidson J. 2000. *Econometric Theory*. Blackwell Publishers: Oxford, UK.

Davidson J, Sibbertsen P. 2009. Tests of bias in log-periodogram regression. *Economics Letters* **102**:83–86.

Davidson J. 2015. Time series modelling 4.47 at http://www.timeseriesmodelling.com [Accessed on May 2015].

Doornik J. 2013. *An Object-Oriented Matrix Programming Language—Ox 7*. Timberlake Consultants Ltd: London.

Elliott G, Rothenberg TJ, Stock JH. 1996. Efficient tests for an autoregressive unit root. *Econometrica* **64**(4):813–836.

Fischer H, Wahlen M, Smith J, Mastroianni D, Deck B. 1999. Ice core records of atmospheric CO_2 around the last three glacial terminations. *Science* **283**:1712–1714.

Geweke J, Porter-Hudak S. 1983. The estimation and application of long-memory time series models. *Journal of Time Series Analysis* **4**:221–237.

Gay-Garcia C, Estrada F, Sanchez A. 2009. Global and hemispheric temperatures revisited. *Climatic Change* **94**:333–349.

Giraitis L, Kokoszka P, Leipus R, Teyssiere G. 2003. Rescaled variance and related tests for long memory in volatility and levels. *Journal of Econometrics* **112**:265–294.

Granger CWJ. 1969. Investigating causal relations by econometric models and cross-spectral methods. *Econometrica* **37**:424–38.

Granger CWJ. 2010. Some thoughts on the development of cointegration. *Journal of Econometrics* **158**:3–6.

Harris D, McCabe B, Leybourne S. 2008. Testing for long memory. *Econometric Theory* **24**(1):143–175.

Hurvich CM, Deo R, Brodsky J. 1998. The mean squared error of Geweke and Porter–Hudak's estimator of a long memory time series. *Journal of Time Series Analysis* **19**:19–46.

Imbrie J, Hays JD, Martinson DG, Mcintyre A, Mix AC, Morley JJ, Pisias NG, Prell WL, Shackleton NJ. 1984. *The orbital theory of Pleistocene climate: support from a revised chronology of the marine δ^{1B}O record. Milankovitch and Climate Part 1,* Berger, A.L, *et al,* (eds). D. Reidel Publishing Company.

Imbrie J, Boyle EA, Clemens SC, Duffy A, Howard WR, Kukla G, Kutzbach J, Martinson DG, Mcintyre A, Mix AC, Molfino B, Morley JJ, Peterson LC, Pisias NG, Prell WL, Raytoo ME, Shackletons NJ, Toggweiler JR. 1992. On the structure and origin of major glaciation cycles. 1. Linear responses to Milankovitch forcing. *Paleoceanography* **7**:701–738.

Johansen S. 1991. Estimation and hypothesis testing of cointegration in Gaussian vector autoregressive models. *Econometrica* **59**:1551–80.

Jouzel J, Vimeux F, Caillon N, Delaygue G, Hoffman G, Masson-Delmotte V, Parrenin F. 2003. Magnitude of isotope/temperature scaling for interpretation of central Antarctic ice cores. *Journal of Geophysical Research* **D12**(4631):108, DOI: 10.1029/2002JD002677.

Jouzel J, *et al.* 2007. Orbital and millennial Antarctic climate variability over the past 800,000 years. *Science* **317**:793–796.

Kaufmann RK, Stern DI. 2002. Cointegration analysis of hemispheric temperature relations. *Journal of Geophysical Research* **107**:ACL8.1–ACL8.10.

Kaufmann RK, Kauppi H, Stock JH. 2006. Emissions, concentrations and temperature: a time series analysis. *Climatic Change* **77**:248–278.

Kaufmann RK, Kauppi H, Stock JH. 2010. Does temperature contain a stochastic trend? *Climatic Change* **101**:395–405.

Kaufmann RK, Juselius K. 2013. Testing hypotheses about glacial cycles against the observational record. *Paleoceanography* **28**:1–11.

Kunsch HR. 1987. Statistical Aspects of Self-similar Processes. In *Proceedings of 1st World Congress of the Bernoulli Soc.,* Prohorov Y, Sazanov VV (eds). VNU Science Press: Utrecht 1, 67–74.

Kwiatkowski D, Phillips PCB, Schmidt P, Shin Y. 1992. Testing the null of stationarity against the alternative of a unit root. *Journal of Econometrics* **54**:159–178.

Lisiecki LE, Raymo ME. 2005. A Pliocene-Pleistocene stack of 57 globally distributed benthic δ^{18}O records. *Paleoceonography* **20**:PAA1003, 1–17.

Lo AW. 1991. Long-term memory in stock market prices. *Econometrica* **59**(5):1279–1313.

Lobato IN, Robinson PM. 1998. A nonparametric test for I(0). *Review of Economic Studies* **65**(3):475–495.

Loutre MF. 2002. Ice Ages (Milankovitch Theory). In *Encyclopedia of Atmospheric Sciences,* Curry JA, Pyle JA (eds). Elsevier Science Ltd, *995–1003*.

Lüthi D, Le Floch M, Bereiter B, Blunier T, Barnola J-M, Siegenthaler U, Raynaud D, Jouzel J, Fischer H, Kawamura K, Stocker T. 2008. High-resolution carbon dioxide concentration record 650,000–800,000 years before present. *Nature* **453**:379–382.

MacMillan DG, Wohar ME. 2013. The relationship between temperature and CO_2 emissions: evidence from a short and very long dataset. *Applied Economics* **45**(26):3683–3690.

Mills TC. 2010. "Skinning a cat": alternative models of representing temperature trends. *Climatic Change* **101**:415–426.

Mudelsee M. 2001. The phase relations among atmospheric CO_2 content, temperature and global ice volume over the past 420 ka. *Quaternary Science Reviews* **20**:583–589.

Newey WK, West KD. 1994. Automatic lag selection in covariance matrix estimation. *Review of Economic Studies* **61**:631–653.

Paillard D, Labeyrie L, Yiou P. 1996. Macintosh program performs time-series analysis. *Eos Transactions American Geophysical Union* **77**:379.

Paillard D. 2001. Glacial cycles: towards a new paradigm. *Reviews of Geophysics* **39**(3):325–346.

Parrenin F, Van Ommen TD, Wolff EW. 2007. Preface, the EPICA (EDC and EDML) ice cores age scales. *Climate of the Past*1–3. Special Issue.

Phillips PCB, Perron P. 1988. Testing for a unit root in time series regression. *Biometrika* **75**:335–346.
Phillips PCB, Solo V. 1992. Asymptotics for linear processes. *Annals of Statistics* **20**:971–1001.
Robinson PM. 1995. Gaussian semiparametric estimation of long-range dependence. *Annals of Statistics* **23**:1630–1661.
Said SE, Dickey DA. 1984. Testing for unit roots in autoregressive-moving average models of unknown order. *Biometrika* **71**:599–607.

APPENDIX A: TESTS OF I(1) AND I(0)

Table A.1 shows a conventional battery of classic and more recently derived tests for unit roots and stationarity/weak dependence, performed on the transformed series shown in Figure 1. The p-values shown are computed either from asymptotically exact tables, where available, or from Monte Carlo simulations. Where p-values are available for a discrete range of significance levels, upper bounds according to the tabulated points are indicated by "<". ADF denotes the Augmented Dickey–Fuller test (Said and Dickey, 1984), the PP test is from Phillips and Perron (1988), and the DF-GLS and P tests are from Elliott, Rothenberg and Stock (1996). The KPSS test is from Kwiatkowski *et al.* (1992), the V/S test from Giraitis *et al.* (2003), and the modified RS test from Lo (1991). These last three tests represent different ways of comparing the cumulated process with a Brownian motion. The single rejection at the 5% level by the V/S test of CO_2, given the consensus of the results, is plausibly a type 1 error. The RL test from Lobato and Robinson (1998) and the HML test from Harris *et al.* (2008) are for the "weak dependence" null hypothesis, that the spectral density is finite at the origin. While the HML test rejects in each case, it is valid only for linear processes and hence likely to be inappropriate here. Compare Figures S3–S6 the Supporting Information.

Table A.1. Tests of unit roots (I(1)) and stationarity/weak dependence (I(0))

	Temperature		CO_2		CH_4		Ice	
	Statistic	p-value	Statistic	p-value	Statistic	p-value	Statistic	p-value
ADF*	−4.73 [3]	< 0.01	−4.23 [3]	< 0.01	−6.08[1]	< 0.01	−5.81[7]	< 0.01
PP**	−4.94 [5]	< 0.01	−4.10 [5]	< 0.01	−6.19[3]	< 0.01	−4.48[10]	< 0.01
DF-GLS*	−4.56 [1]	< 0.01	−3.93 [3]	< 0.01	−5.83[1]	< 0.01	−5.23[7]	< 0.01
P**	0.530 [4]	< 0.01	0.757 [4]	< 0.01	0.399[7]	< 0.01	0.687[10]	< 0.01
KPSS**	0.129 [13]	< 0.46	0.228 [13]	< 0.22	0.132[12]	< 0.44	0.057[13]	< 0.83
V/S**	0.063 [13]	0.563	0.202 [13]	0.037	0.048[12]	0.729	0.055[13]	0.650
RS**	1.072 [13]	< 0.66	1.65 [13]	< 0.06	0.975[12]	< 0.81	1.082[13]	< 0.69
RL***	−1.23 [12]	0.891	−0.50 [12]	0.691	−1.30[12]	0.093	−1.14[12]	0.88
HML †	7.631	0	7.60	0	7.67	0	7.60	0

*Lags (in square brackets) chosen by Akaike (1969) criterion.
**HAC variance computed with Parzen kernel and bandwidth (in square brackets) chosen by the Newey and West (1994) plug-in method.
***Bandwidth (in square brackets) from Lobato and Robinson (1998) formula.
†Setting $c = 1, L = 0.66$. See Harris *et al.* (2008) for details.

APPENDIX B: DERIVATION OF THE VECTOR ERROR-CORRECTION MODEL SYSTEM

Partition system (1) as

$$\begin{bmatrix} \boldsymbol{A}_{xx}(L) & \boldsymbol{A}_{xz}(L) \\ \boldsymbol{A}_{zx}(L) & \boldsymbol{A}_{zz}(L) \end{bmatrix} \begin{bmatrix} \boldsymbol{x}_t \\ z_t \end{bmatrix} = \begin{bmatrix} \boldsymbol{\Lambda}_x(L) \\ \boldsymbol{\Lambda}_z(L) \end{bmatrix} \boldsymbol{d}_t + \begin{bmatrix} \boldsymbol{u}_{xt} \\ \boldsymbol{u}_{zt} \end{bmatrix} \quad \begin{matrix} (4 \times 1) \\ (m_z \times 1) \end{matrix} \tag{B.1}$$

Solving out z_t yields the reduced system

$$\tilde{\boldsymbol{B}}(L)\boldsymbol{x}_t = \tilde{\boldsymbol{\Upsilon}}(L)\boldsymbol{d}_t + \boldsymbol{v}_t \tag{B.2}$$

where

$$\tilde{\boldsymbol{B}}(L) = |\boldsymbol{A}_{zz}(L)|\boldsymbol{A}_{xx}(L) - \boldsymbol{A}_{xz}(L)\text{adj}\boldsymbol{A}_{zz}(L)\boldsymbol{A}_{zx}(L) \tag{B.3a}$$

$$\tilde{\boldsymbol{\Upsilon}}(L) = |\boldsymbol{A}_{zz}(L)|\boldsymbol{\Lambda}_x(L) - \boldsymbol{A}_{xz}(L)\text{adj}\boldsymbol{A}_{zz}(L)\boldsymbol{\Lambda}_z(L) \tag{B.3b}$$

and

$$\boldsymbol{v}_t = |\boldsymbol{A}_{zz}(L)|\boldsymbol{u}_{xt} - \boldsymbol{A}_{xz}(L)\text{adj}\boldsymbol{A}_{zz}(L)\boldsymbol{u}_{zt} \tag{B.4}$$

Note that if $\boldsymbol{A}(0) = \boldsymbol{I}_m$, then $\tilde{\boldsymbol{B}}(0) = \boldsymbol{I}_4$. System (B.2) has a vector ARMA structure, noting that the right-hand side term in (B.4) is a sum of finite-order moving averages. Hence, it has a terminating autocovariance sequence and a representation as a finite-order moving average of white noise elements. There exists a white noise vector $\boldsymbol{\varepsilon}_t$ such that

$$\boldsymbol{v}_t = \boldsymbol{\Theta}(L)\boldsymbol{\varepsilon}_t \tag{B.5}$$

where $\boldsymbol{\Theta}(L)$ (4×4) is a finite-order matrix polynomial and by choice of the covariance matrix of $\boldsymbol{\varepsilon}_t$, we may set $\boldsymbol{\Theta}(0) = \boldsymbol{I}_4$ without loss of generality. Invertibility of this component allows us to write the reduced VAR as (2) where $\boldsymbol{B}(L) = \boldsymbol{\Theta}(L)^{-1}\tilde{\boldsymbol{B}}(L)$ (with $\boldsymbol{B}(0) = \boldsymbol{I}_4$) and $\check{}(L) = \boldsymbol{\Theta}(L)^{-1}\tilde{\boldsymbol{\Upsilon}}(L)$. While these structures are in general of infinite autoregressive order, their coefficients must be at least 1 summable, and we assume they can be truncated to finite order in a useful approximation.

If the system is stationary and exhibits mean reversion, the steady-state relations have the form (4). Re-parameterizing the system, so that the deviations from (4) appear explicitly, is performed using the so-called Beveridge–Nelson decompositions of the matrices (see Phillips and Solo, 1992). For example,

$$\begin{aligned} \boldsymbol{B}(L) &= \boldsymbol{B}(1) + \boldsymbol{B}^*(L)(1-L) \\ &= \boldsymbol{B}(1)L + \boldsymbol{C}(L)(1-L) \end{aligned}$$

where $\boldsymbol{C}(L) = \boldsymbol{B}^*(L) + \boldsymbol{B}(1)$ and $\boldsymbol{B}^*(L)$ is the matrix polynomial with coefficients

$$\boldsymbol{B}_j^* = -\sum_{k=j+1}^{\infty} \boldsymbol{B}_k$$

Note that $\boldsymbol{B}^*(0) = \boldsymbol{B}(0) - \boldsymbol{B}(1)$ and therefore $\boldsymbol{B}(0) = \boldsymbol{I}$ implies $\boldsymbol{C}(0) = \boldsymbol{I}$, by construction. The system (2) can therefore be rewritten in the so-called vector error-correction model form (3) where $\boldsymbol{D}(L) = \boldsymbol{\Upsilon}^*(L) + \boldsymbol{\Upsilon}(1)$ similarly to $\boldsymbol{C}(L)$, and $\boldsymbol{E} = -\boldsymbol{B}(1)$. Notice that replacing $\boldsymbol{x}_{t-1}$ and $\boldsymbol{d}_{t-1}$ by any lags up to the minimum of the lag orders of $\boldsymbol{D}(L)$ and $\boldsymbol{C}(L)$ yields an observationally equivalent system, apart from a reparameterization of the dynamic components. Hence, there is no loss of generality in choosing this form.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of this article at the publisher's web site.

