ELSEVIER

Redundancies in the Earth's climatological time series

Cees Diks [a,*], Manfred Mudelsee [b]

[a] *CeNDEF, Department of Economics, Roetersstraat 11, 1018 WB, Amsterdam, The Netherlands*
[b] *Institute of Meteorology, University of Leipzig, Stephanstr. 3, D–04103 Leipzig, Germany*

Abstract

We examine nonlinear coupling among a number of important climatological variables, $\delta^{13}C$, $\delta^{18}O$ and insolation, using mutual information and redundancy. The coupling among these variables is found to vary over different climatological eras. The dependence between $\delta^{13}C$ and $\delta^{18}O$ is of particular strength in more recent samples. Tests for Granger causality suggest that $\delta^{18}O$ has an effect on $\delta^{13}C$ whereas the reverse is not the case. © 2000 Published by Elsevier Science B.V.

1. Introduction

In the last decades several nonlinear methods for the analysis of time series have been developed, including methods for estimating dynamical invariants such as correlation dimension, correlation entropy and Lyapunov exponents (for an overview see e.g. Diks [1]). The application of the Grassberger–Procaccia algorithm to palaeoclimatic time series led to controversies about the existence of a low-dimensional "climatic attractor" (see Ruelle [2] for an early review). Analysis of one of the longest and most reliable records (ODP 659, see Section 3 for a description of the data) yielded the somewhat unsatisfactory conclusion that at least five climate variables are acting [3]. Investigating higher values would

necessitate time series lengths of at least one order of magnitude above the current capacity.

One of the disadvantages of chaos analysis is that the assumptions underlying it (deterministic time series without observational noise) are not realistic for measured palaeoclimatic data. Therefore we analyze the ODP 659 record by means of mutual information and redundancy, which do not require those assumptions, and have already been applied successfully to time series from very different origins, such as meteorology and physiology [4].

This Letter is organized as follows. In the next section two information theoretical concepts, mutual information and its generalization, redundancy, are introduced. In Section 3 we present our analysis of the ODP 659 data and describe the stationary bootstrap used to estimate confidence intervals. In Section 4 we introduce an information theoretical test for Granger causality and apply it to the data to obtain some insights into the causal nature of the

---

* Corresponding author.
E-mail address: diks@fee.uva.nl (C. Diks).

dependence between the different variables. A short geological interpretation is included.

2. Mutual information and redundancy

Consider two stochastic variables X and Y with joint probability density function (pdf) $f_{X,Y}(x,y)$, and marginal pdf $f_X(x)$ and $f_Y(y)$ respectively. The mutual information of X and Y is given by

$$I(X,Y) = \int \int f_{X,Y}(x,y) \log \frac{f_{X,Y}(x,y)}{f_X(x)f_Y(y)} \, dx \, dy. \tag{1}$$

The mutual information can be considered a nonlinear analogue of the correlation between X and Y. Intuitively, the mutual information is the amount of information contained in one of the two variables about the other variable. The mutual information is symmetric, nonnegative and equal to zero if and only if X and Y are independent [5].

Redundancy is the extension of mutual information to higher-dimensional cases. For the random variable $\mathbf{X} = (X_1, \ldots, X_M)$, with joint pdf $f_{\mathbf{X}}(\mathbf{x})$ and marginal pdfs $f_{X_i}(x_i)$, for $i = 1, \ldots, M$, the redundancy is defined as

$$R(\mathbf{X}) = \int_{\mathbf{R}^M} f_{\mathbf{X}}(\mathbf{x}) \log \frac{f_{\mathbf{X}}(\mathbf{x})}{\prod\limits_{i=1}^{M} f_{X_i}(x_i)} \, d\mathbf{x}. \tag{2}$$

This concept of redundancy was used by Paluš [6] in a time series context. Like the mutual information, the redundancy is nonnegative, and zero if and only if the X_i are independent.

Paluš, Albrecht and Dvořák [7] determined the redundancy of multivariate Gaussian random variables, which enables one to define "linear redundancies" in terms of the covariance matrix of the random variables. The linear redundancy function is given by

$$L(\mathbf{X}) = \frac{1}{2} \sum_{i=1}^{M} \log c_{ii} - \frac{1}{2} \sum_{i=1}^{M} \log \sigma_i, \tag{3}$$

where the c_{ii} are the diagonal elements (variances) and σ_i the eigenvalues of the $M \times M$ covariance matrix $\mathbf{C}$ of $\mathbf{X}$. For $M = 2$ this can be written [5] as

$$L(X,Y) = -\tfrac{1}{2} \log(1 - \rho^2), \tag{4}$$

where ρ is the correlation between X and Y.

Prichard and Theiler [8] proposed generalizations of entropy, mutual information and redundancy, by expressing them in terms of generalized correlation integrals $C_q(\epsilon)$. For the mutual information and the redundancy, they give

$$I_q(X,Y;\epsilon) = \log C_q(X,Y;\epsilon) - \log C_q(X;\epsilon)$$
$$- \log C_q(Y;\epsilon), \tag{5}$$

$$R_q(\mathbf{X};\epsilon) = \log C_q(\mathbf{X};\epsilon) - \sum_{i=1}^{M} \log C_q(X_i;\epsilon) \tag{6}$$

respectively. These forms enable convenient estimation by means of plug-in estimates of C_q. The choice $q = 2$, the value which will be used throughout this Letter, is particularly convenient since the estimation of the correlation integral for $q = 2$ is straightforward. The correlation integral of a random variable $\mathbf{X}$ with associated probability measure $\mu_{\mathbf{X}}$ is

$$C_2(\mathbf{X};\epsilon) = \int \int \Theta(\epsilon - \|\mathbf{x} - \mathbf{y}\|) d\mu_{\mathbf{X}}(\mathbf{x}) d\mu_{\mathbf{X}}(\mathbf{y}), \tag{7}$$

where $\Theta(\cdot)$ is the Heaviside function,

$$\Theta(s) = \begin{cases} 0 & \text{if } s < 0 \\ 1 & \text{if } s \geq 0. \end{cases} \tag{8}$$

The correlation integral is usually estimated from a sampled M-dimensional time series $\{\mathbf{x}(j)\}_{j=1}^{N}$ as the fraction of distances smaller than ϵ:

$$\hat{C}_2(\mathbf{x};\epsilon) = \frac{2}{N(N-1)}$$
$$\times \sum_{j=1}^{N-1} \sum_{k=j+1}^{N} \Theta(\epsilon - \|\mathbf{x}(j) - \mathbf{x}(k)\|). \tag{9}$$

3. Data analysis

The data we consider consist of the oxygen isotope ($\delta^{18}O$), carbon isotope ($\delta^{13}C$) and dust flux records from the Ocean Drilling Program (ODP) Site 659, located on the Cape Verde Plateau west of Africa [9]. The covered time interval is the past 5000 ka (1 ka = 1000 years). The delta notation refers to the relative deviation of isotope ratios from a reference standard (PDB). For example,

$$\delta^{18}O[\text{‰ versus PDB}]$$

$$= \frac{\left(^{18}O/^{16}O\right)_{\text{sample}} - \left(^{18}O/^{16}O\right)_{\text{PDB}}}{\left(^{18}O/^{16}O\right)_{\text{PDB}}}$$

$$\times 1000, \tag{10}$$

and an analogous definition gives $\delta^{13}C$ in terms of ^{13}C and ^{12}C.

The $\delta^{18}O$ values are reliable recorders of global ice volume; $\delta^{13}C$ values reflect mainly the strength of formation of North Atlantic Deep Water (NADW) —also an important climate variable but with a slightly less global influence [10]. The ODP 659 dust flux records changes in Sahelian aridity [9]. During our data analysis it became clear that the results involving the dust flux variable were difficult to interpret. For clarity of the presentation of the results, the dust flux is no longer considered in this Letter. Instead, in the analysis we include the time series of solar insolation at 60°N [11] which is regarded as the major external forcing of Plio-/Pleistocene climate.

We divide the ODP 659 time interval into four distinctive climatic periods: The oldest, IV, from 5000 to 3585 ka ago, saw already climate variability, but not as strong as the glaciation of the Northern Hemisphere which came in the late Pliocene (period III), from 3585 to 2625 ka. Period II covers 2470–937 ka, a time of gradual cooling and additional build-up of ice in early Pleistocene. Then, the Mid-Pleistocene Climate Transition occurred, a relatively abrupt increase of global ice volume [12], which led to late Pleistocene ice ages (period I), since 892 ka, with large glacial–interglacial amplitudes. The number N of observations is 216, 301, 179 and 266 for periods I, II, III and IV, respectively.

3.1. Redundancies

For each period, the time series were linearly detrended and then rescaled to zero mean and unit variance. The original time series had only small trends within each of the selected climatic regimes. Note that the multivariate measurements on the sediment core were made at equally spaced depths. Because sedimentation is not a constant process, the ODP 659 time series are not equally spaced. However, due to the respective climatic regime, within each of the four periods the coefficient of variation of the time spacing, that is, the standard deviation of the time spacing divided by the average time spacing, is always less than about 50%. Because of the unequal sampling, the construction of delay vectors from the scalar time series, as is common practice in chaos analysis, can no longer be justified. Hence, we avoid the calculation of redundancies my means of delay vectors as in Paluš [6], and examine the system using the multivariate data instead.

Table 1 shows the estimated redundancy of the full joint 3-variate time series as well as for all pairs in the four different geological periods. For the scale parameter we chose $\epsilon = 1$ (which amounts to one standard deviation of the time series considering the rescaling) throughout. The increase in redundancy among the three climate variables suggests that the dependence between them has increased slowly over geologic time. From the mutual information calculated for the pairs of variables, it appears that the increase in coupling in the full 3-variate time series

Table 1

Estimated redundancies ($\epsilon = 1$) for the triple and the three pairs of variables in the four different climatic periods. The increase in redundancy for the triple appears to be attributable to the increase in redundancy between $\delta^{13}C$ and $\delta^{18}O$.

	$\delta^{13}C$, $\delta^{18}O$, insol.	$\delta^{13}C$, $\delta^{18}O$	$\delta^{13}C$, insol.	$\delta^{18}O$, insol.
I	0.26	0.16	0.00	0.09
II	0.10	0.08	0.01	0.00
III	0.10	0.08	0.02	−0.01
IV	0.05	0.01	0.02	0.01

is mainly due to the increase in coupling between δ^{13}C and δ^{18}O. In period I, some additional cou-

pling comes from the relation between δ^{18}O and insolation.

δ^{18}O lagged, δ^{13}C   insol. lagged, δ^{13}C   insol. lagged, δ^{18}O
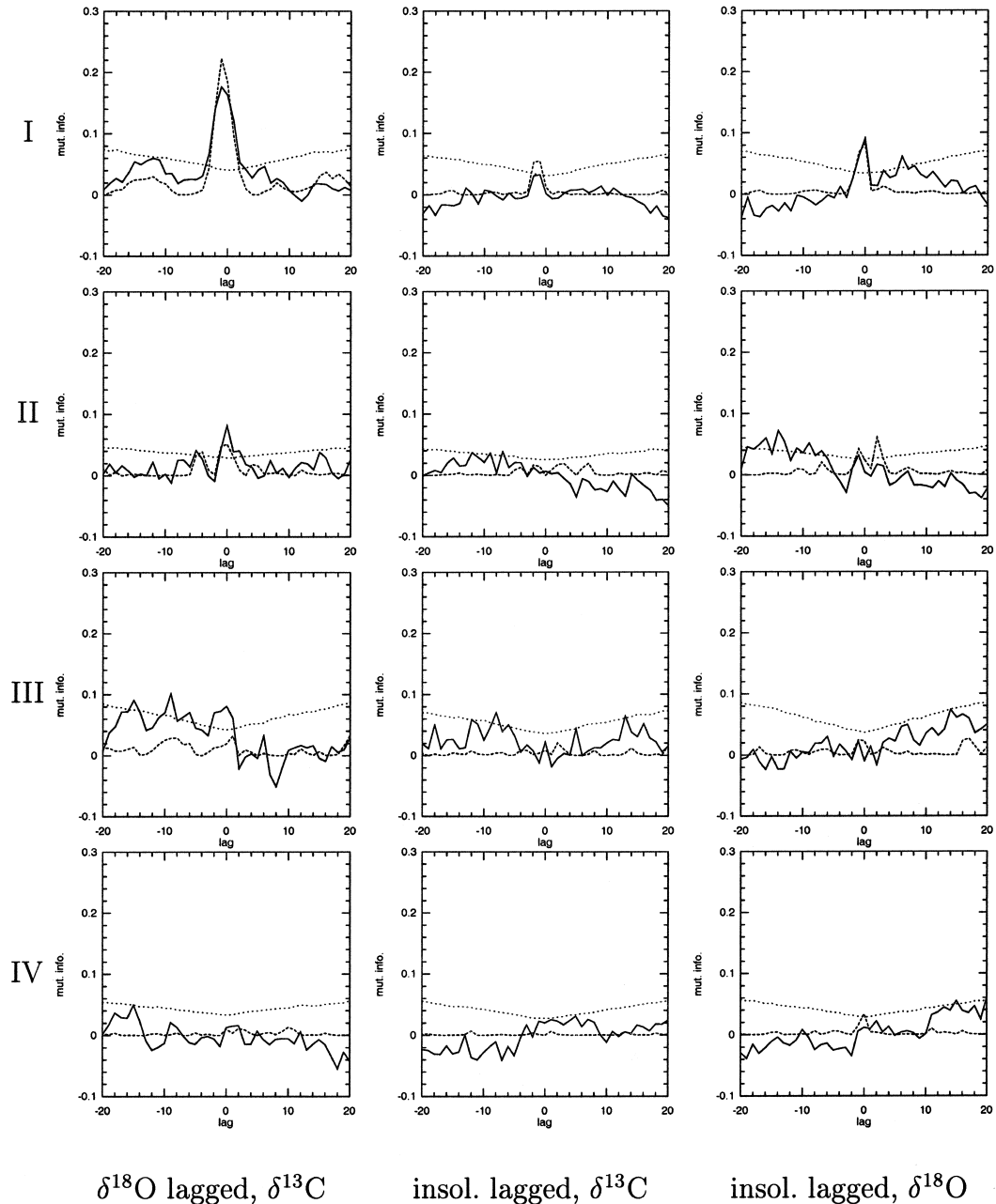
Fig. 1. Estimated cross-redundancies (solid lines) and linear cross-redundancies (dashed lines) as a function of the lag (in sample intervals) in the four different periods. The dotted lines indicate the 95% quantiles for the cross-redundancy in the absence of dependence. The cross-redundancies were estimated with $\epsilon = 1$.

3.2. Cross-redundancies

To examine the relation between $\delta^{13}C$ and $\delta^{18}O$ in more detail, we determined the mutual information of both time series as a function of the delay (termed cross-redundancy in Prichard and Theiler [8]). The cross-redundancy at lag l of two simultaneously measured time series $\{x(j)\}_{j=1}^{N}$ and $\{y(j)\}_{j=1}^{N}$ is

$$I_2(x(j), y(j+l)). \tag{11}$$

Prichard and Theiler [8] showed that the linear redundancy is the same for all q (see Eq. (3)), which suggests that a qualitative comparison of the cross-redundancy and the linear cross-redundancy can be used as a qualitative test for linearity.

Fig. 1 shows the estimated cross-redundancy function for each pair of variables in periods I, . . . , IV. The dotted lines indicate the 97.5% quantiles obtained using 4000 stationary bootstrap replications, under the assumption that no causal relations exist between the two time series (both time series are bootstrapped independently).

The stationary bootstrap [13] is a resampling scheme developed for weakly dependent stationary time series. Given an observed time series, $\{x(j)\}_{j=1}^{N}$, a bootstrap time series $\{x^*(j)\}_{j=1}^{N}$ is constructed as follows. The first element $x^*(1)$ is randomly selected from $x(1), \ldots, x(N)$. The consecutive points are generated as follows: Let $x^*(j) = x(k)$, say, then with probability $(1 - P)$, with P small, the successor of $x^*(j)$ is taken to be the successor of $x(k)$ in the original time series, that is, $x^*(j+1) = x(k+1)$, and with probability P, $x^*(j+1)$ is a randomly chosen element of the original time series. Periodic boundary conditions are imposed to ensure that $x(N)$ has a successor, $x(1)$. Note that this resampling scheme is equivalent to a bootstrap in which random blocks from the original time series are concatenated, with geometrically distributed block lengths. Intuitively, for small P, the bootstrap preserves dependence in the time series up to lags of the order of the mean block length, $1/P$. For the pairs of time series studied here, we generated bootstrap time series independently. As this destroys any existing dependence among the two time series, it can be considered as generating new realizations under the null hypothesis that the two time series are independent. Therefore, the obtained confidence interval can be interpreted as the confidence interval under the null hypothesis of independence of the respective time series (while the dependence structure in the time series themselves is preserved). We used $P = 0.05$ throughout. This choice appears to be a reasonable trade-off between the mean block length ($1/P = 20$) and the expected number of blocks ($PN \simeq 10$).

Considering the uncertainty introduced by the uneven time spacing, the estimated quantiles should be regarded as a rough indication of the true quantiles. The probability that the estimated mutual information exceeds this level is about 2.5% in the case when there is no dependence between the random variables. Note that the quantiles have a minimum for lag zero, and increase for lags further from zero. This results from the decrease of the overlapping part of two time series when the lag is increased, which increases the statistical fluctuations.

For comparison, Fig. 1 also shows the estimated linear redundancy function (dotted lines, see Eq. (3)), which is the mutual information based on the assumption that the time series are jointly Gaussian. Many of the statistically significant peaks in the redundancy functions are strongly reduced or absent in the linear redundancy function, suggesting that these peaks are the result of nonlinear dependence. We regard this as an *a posteriori* justification for the use of nonlinear dependence measures rather than linear correlation.

In several panels, for example for $\delta^{13}C$ and $\delta^{18}O$ in period III and somewhat in periods I and II, an asymmetry can be observed in the mutual information with respect to changing the sign of the lag. For positive lags there are practically no statistically significant estimates of the mutual information, whereas the mutual information is significantly larger than zero for several negative lags. Because the mutual information is merely a (nonlinear) measure for correlation, these asymmetries do not contain sufficient information to decide which variable is driving which. The next section deals with this question in more detail.

4. Granger causality

So far, by using nonlinear correlation measures, we have only examined dependence between the

climatological variables. However, we would like to obtain some insights into the nature of the couplings that bring about the observed dependence. For example, if there is nonlinear dependence between two variables, this might be because the first variable is driving the second, or the second is driving the first, or both (feedback).

The aim of this section is to examine the causal relationships between the variables by testing for Granger causality [14]. We design a test for the following null hypothesis:

$$H_0 : Y \text{ is not a Granger cause of } X. \qquad (12)$$

This well-known concept in econometric theory is based on predictability. Intuitively, Y is a Granger cause of X if past values of Y can improve predictions of future values of X. More formally, Y is a Granger cause of X if future values of X, conditionally on past values of X *and* Y, are distributed differently than future values, conditionally on past values of X *only*. In this case X is conditionally (on past values of X) dependent on Y. This definition of causality is only operational and leaves open the possibility that causality is found between X and Y when they are in fact uncoupled. This can be the case, for example, if both X and Y are driven by a third variable, Z say.

As the test statistic we used the conditional marginal redundancy (see Ref. [15]), that is,

$$T_q(X, m, Y, l) = r_q\big(Y(j - l), \ldots, Y(j); X_{j+1} \,|$$
$$X(j - m), \ldots, X(j)\big) \qquad (13)$$

which is the average amount of information contained in $Y(j - l), \ldots, Y(j)$, given that we know $X(j - m), \ldots, X(j)$. We can write the conditional marginal redundancy as

$$T_q(X, m, Y, l) = r_q(X(j - m), \ldots, X(j),$$
$$Y(j - l), \ldots, Y(j); X(j + 1))$$
$$- r_q(X(j - m), \ldots, X(j); X(j + 1)), \qquad (14)$$

where m and l are time lags and the marginal redundancy $r(\cdot)$ is given by [8]

$$r_q(\mathbf{Z}(j); X(j + 1)) = R_q(\mathbf{Z}(j),$$
$$X(j + 1)) - R_q(\mathbf{Z}(j)), \qquad (15)$$

where $\mathbf{Z}(j)$ stands for $(X(j - m), \ldots, X(j))$ or $(X(j - m), \ldots, X(j), Y(j - l), \ldots, Y(j))$ respectively. The marginal redundancy quantifies the amount of information about the variable $X(j + 1)$ contained in $\mathbf{Z}(j)$. When the distribution of X is independent of past values of Y, T_q is zero, whereas T_q is positive when Y drives X. Therefore, a one-sided test is appropriate, in which it is established whether or not $\hat{T}_q(x, m, y, l)$, that is, the estimated value of T_q using the measured time series x and y, is significantly larger than zero. $\hat{T}_q$ is obtained by plugging in estimated correlation integrals, that is,

$$\hat{R}_q(\mathbf{z}; \epsilon) = \log \hat{C}_q(\mathbf{z}; \epsilon) - \sum_{i=1}^{\dim \mathbf{z}} \log \hat{C}_q(\mathbf{z}_{(i)}; \epsilon), \quad (16)$$

where $\mathbf{z}_{(i)}$ denotes the ith component of $\mathbf{z}$. Again, we use $q = 2$.

The stationary bootstrap is used to estimate the probability (p-value) that $\hat{T}_q(x^*, m, y^*, l) \geq \hat{T}_q(x, m, y, l)$ by the number of such simulations divided by the total number of simulations (1000). A small p-value suggests that it is unlikely that the observed value $\hat{T}_q(x, m, y, l)$ had occurred by chance (under the "no coupling" hypothesis), that means that "H_0 = true" is unlikely. Testing at a confidence level of $1 - \alpha$, one would reject H_0 whenever the p-value is smaller than α.

The tests are performed for $l \leq m$ with m fixed (see Eq. (14)). The value of m was chosen to be as large as possible without leading to estimates of C_2 equal to zero, in which case the estimated marginal redundancies would become infinite. For our data, this requirement leads to $m = 4$.

We test for Granger causality between all pairs of variables in the four periods. Table 2 shows the obtained p-values for each pair of variables in each of the four climatological periods. We find no evidence for Granger causality in the oldest period. For the three most recent periods (I–III), there is strong evidence for $\delta^{18}O$ being a Granger cause of $\delta^{13}C$.

Table 2
Obtained *p*-values for the test for Granger causality among $\delta^{13}C$, $\delta^{18}O$ and insolation. 1000 bootstrap replications. *l*, lag. $\epsilon = 1$, $m = 4$.

	$\delta^{18}O \to \delta^{13}C$				insolation $\to \delta^{13}C$				insolation $\to \delta^{18}O$			
l	I	II	III	IV	I	II	III	IV	I	II	III	IV
1	0.01	0.09	0.02	0.62	0.01	0.24	0.98	0.11	0.07	0.03	0.38	0.47
2	0.02	0.04	0.00	0.47	0.03	0.48	0.84	0.30	0.05	0.30	0.58	0.88
3	0.06	0.02	0.01	0.57	0.05	0.17	0.81	0.18	0.04	0.48	0.55	0.90
4	0.07	0.02	0.02	0.90	0.10	0.30	0.91	0.19	0.07	0.67	0.57	0.95

	$\delta^{13}C \to \delta^{18}O$				$\delta^{13}C \to$ insolation				$\delta^{18}O \to$ insolation			
l	I	II	III	IV	I	II	III	IV	I	II	III	IV
1	0.17	0.26	0.26	0.10	0.50	0.24	0.46	0.22	0.32	0.47	0.08	0.36
2	0.47	0.36	0.37	0.19	0.71	0.20	0.14	0.08	0.47	0.56	0.25	0.33
3	0.40	0.32	0.37	0.33	0.31	0.38	0.27	0.13	0.59	0.55	0.46	0.28
4	0.51	0.39	0.53	0.33	0.17	0.63	0.35	0.05	0.36	0.60	0.63	0.32

We find good evidence suggesting that insolation is a Granger cause of $\delta^{13}C$ during period I and of $\delta^{18}O$ from II to the present.

In geological terms, the increase in coupling reflects the growing influence of Northern Hemisphere ice sheets ($\delta^{18}O$) on global climate (cf. [10]). In period IV, NADW formation ($\delta^{13}C$) was little influenced by ice volume. During the Northern Hemisphere glaciation (period III) the ice sheets grew southwards, reaching the North Atlantic. This influenced NADW formation through various climatological links (e.g., movement of the polar front), leading to nonlinearity and delay behavior. In period II, the grip of ice volume on NADW formation was rather stable. The grip was increased in late Pleistocene ice ages (period I) after marine-based ice sheets had developed [12].

Because $\delta^{13}C$ is not found to be a Granger cause of $\delta^{18}O$, we have found no evidence for feedback between these two variables in our dataset. No evidence is found for $\delta^{18}O$ (or $\delta^{13}C$) being a Granger cause of insolation. This suggests that the effect of possible physical mechanisms by which the climate variables can influence insolation (through changes in the Earth's moment of inertia [16]) is not very strong. Some evidence suggests that insolation has an impact on long-term climate dynamics via $\delta^{18}O$, and possibly $\delta^{13}C$, from periods II and I, respectively, to the present. This is in line with the leading role which climatic Milankovitch theory (see Refs. [9,10] and references therein) ascribes to insolation.

When comparing the *p*-values one should bear in mind that the power of the test in the various periods depends on the number of observations in each period.

5. Concluding remarks

Using recently developed, information theoretical, nonlinear methods for the analysis of time series we found that the coupling in the global climate system is essentially nonlinear and has increased over time. Upon examining the redundancies in subsets of variables, it was found that the increase in coupling could be ascribed mainly to an increased coupling between $\delta^{18}O$ (an indicator of global ice volume) and $\delta^{13}C$ (an indicator of NADW formation). Tests for Granger causality suggest that $\delta^{18}O$ is a Granger cause of $\delta^{13}C$, whereas the reverse is not the case.

References

[1] C. Diks, Nonlinear Time Series Analysis: Methods and Applications, vol. 4 of Nonlinear Time Series and Chaos, World Scientific, Singapore, 1999.

[2] D. Ruelle, Proc. R. Soc. (London) A 427 (1990) 241.

[3] M. Mudelsee, K. Stattegger, Math. Geol. 26 (1994) 799.

[4] M. Paluš, Phys. Lett. A 213 (1996) 138.

[5] C. Granger, J.-L. Lin, J. Time Ser. Anal. 15 (1994) 371.

[6] M. Paluš, Physica D 80 (1995) 186.

[7] M. Paluš, V. Albrecht, I. Dvořák, Phys. Lett. A 175 (1993) 203.

[8] D. Prichard, J. Theiler, Physica D 84 (1995) 476.

[9] R. Tiedemann, M. Sarnthein, N.J. Shackleton, Paleoceanography 9 (1994) 619.

[10] M.E. Raymo, D. Hodell, E. Jansen, Paleoceanography 7 (1992) 645.

[11] A. Berger, M.F. Loutre, Quat. Sci. Rev. 10 (1991) 297.

[12] M. Mudelsee, M. Schulz, Earth Plan. Sci. Lett. 151 (1997) 117.

[13] D.N. Politis, J.P. Romano, J. Am. Stat. Assoc. 89 (1994) 1303.

[14] C.W.J. Granger, Econometrica 37 (1969) 424.

[15] R.M. Gray, Entropy and Information Theory, Springer, Berlin, 1990.

[16] J.X. Mitrovica, A.M. Forte, R. Pan, Geophys. J. Int. 128 (1997) 270.