JOURNAL OF GEOPHYSICAL RESEARCH, VOL. 110, D18103, doi:10.1029/2005JD005895, 2005

Limitations on regression analysis due to serially correlated residuals: Application to climate reconstruction from proxies

Peter Thejll and Torben Schmith
Climate Division, Danish Meteorological Institute, Copenhagen, Denmark

Received 25 February 2005; revised 4 May 2005; accepted 1 July 2005; published 27 September 2005.

[1] The effects of serially correlated residuals on the accuracy of linear regression are considered, and remedies are suggested. The Cochrane-Orcutt method specifically remedies the effects of serially correlated residuals and yields more accurate regression coefficients than does ordinary least squares. We illustrate the effects of serially correlated residuals, explain the application of the CO method, and evaluate the gains to be achieved in its use. We apply the method to an example from climate reconstruction, and we show that the effects of serial correlation in residuals are present and show the significantly improved result.

Citation: Thejll, P., and T. Schmith (2005), Limitations on regression analysis due to serially correlated residuals: Application to climate reconstruction from proxies, *J. Geophys. Res.*, *110*, D18103, doi:10.1029/2005JD005895.

1. Introduction

[2] Studies of past climate can be based on reconstructions that rely on the capture of climate information in various properties of biological remnants and geological formations that remain accessible long after the record was made. *Jones and Mann* [2004] provide an overview of the field. The captured climate information can sometimes be recovered, but the quality of the extracted information depends on the record itself, the proxy, and the type of methods used to extract the information. This paper will look at one specific issue of this problem, related to the use of regression methods in the extraction process.

[3] Typically an instrumental climate record exists for a location or region alongside proxy records, and the analysis method consists of finding the relationship between the two by regression: $y_t = \alpha + \beta x_t + u_t$, where u_t is an error term. The parameters α and β are not known but must be estimated from the predictor and predictand series. This estimation procedure is often considered as being synonymous with the technique of "ordinary least squares" (OLS), where the parameters in the regression model are estimated by minimizing the sum of the squared "residuals" (observed error terms).

[4] However, this is not the best approach under all circumstances. There are conditions to be met, in order for the OLS estimate to be the "best" estimate of the model parameters. This is well known and described in statistical text books [e.g., *von Storch and Zwiers*, 1999], but in applications these conditions are not always met, or explicitly considered.

[5] Central to this is the Gauss-Markov theorem, which states that if the error term time series is stationary and has no serial correlation, then the OLS parameter estimate is the Best Linear Unbiased Estimate (BLUE), meaning that all other linear unbiased estimates will have a larger variance. An estimator having the smallest possible variance is referred to as an "efficient estimator." The Gauss-Markov theorem thus points to the error term and not the time series themselves as being important to consider. Next, it states, that under these conditions, the OLS estimate has two nice properties, namely it is unbiased and has the smallest possible variance among the linear estimates.

[6] The premise in the Gauss-Markov theorem essentially states that the error term must have no structure; for instance, the level of the residuals must not have a trend and the variance must be constant through time. There is no a priori reason to trust that residuals should be without structure; there are at least two ways in which structure in the residuals can occur. First, there is the effect of missing predictors. Any factor that a model fails to incorporate, either by being unrecognized or by being unknown, will turn up in the residuals. Therefore the nature of the residuals depends on the factors omitted. Some of these factors may be serially correlated and thus give rise to serially correlated residuals. Second, there is the effect of mixing variables with different levels of serial correlation. Because the residuals are a linear combination of the predictors and predictand it is possible that the residuals will be serially correlated if one of the dependent or independent variables also is.

[7] When the error term in the regression does not fulfil the premise in the Gauss-Markov theorem, OLS is still unbiased; however, it is not BLUE, i.e., OLS does not exploit the data at hand to give the most efficient estimate of the parameters in the model. In this situation, a strategy would be to transform the problem (i.e., the variables and their regression relationship) so that the error term in the transformed problem has no structure. This strategy will be adopted in the following section ending up in the procedure

0148-0227/05/2005JD005895$09.00

known as the Cochrane-Orcutt algorithm [*Cochrane and Orcutt*, 1949]. Subsequently, we will illustrate this algorithm within the field of climate reconstruction by proxies. In this field multiple regression techniques are widely applied in the "standard OLS form," but the conditions for the Gauss-Markov theorem are usually not tested for. For example, in the past five volumes of *Journal of Climate*, six papers on climate reconstructions were published but the serial correlation of residuals was considered in only two of them.

[8] An alternative way to overcome the problems with estimation of a regression model when the error term has serial correlation was offered by *Guiot* [1985] who separated predictor and predictand series into three spectral band components and estimated regression models in each band, where the residuals were tacitly assumed to be approximately white. The method is applied to the reconstruction of summer temperature in Marseilles from tree rings. Recently, a variant of the method, the "hybrid frequency-domain modification" of the RegEM approach, has been investigated more systematically by *Rutherford et al.* [2005] to reconstruct global temperature fields from proxies. These authors find that this approach does not perform systematically better than the conventional time domain RegEM. A reason for this could be that low-frequency part of the estimation suffers from a small number of statistical degrees of freedom leading to a poor estimation of this part of the model. The strength of the Cochrane-Orcutt procedure in this context is that it introduces one extra parameter only, thus minimizing the risk of overfitting.

[9] Maximum Likelihood estimation is another alternative, offering asymptotic consistent and efficient estimates for any structure of error terms and for a wide range of models, including nonlinear regression models. Among the drawbacks of this method are computational costs due to the necessary iterative procedure and the occasional lack of robustness in comparison with simpler estimators. For a discussion of this estimation method, see *Harvey* [1990].

[10] Serially correlated error terms in climate reconstruction by proxies can have several different causes, two of which were outlined above. It could also be caused by the omission of a lagged copy of an included predictor. In this case the lag lengths to be included can be determined by considering the cross-correlation function (*Fritts et al.* [1971]; for the statistical basis, see, e.g., *Box and Jenkins* [1976]). This procedure will usually introduce several additional parameters into the model, enhancing the risk of overfitting. Another cause of serially correlated errors is "absent proxies," i.e., a proxy we do not have but which would have explained a large fraction of variance in the reconstruction model. In this case we cannot construct a model with unstructured error term. We must identify and estimate a model using the proxies at hand and with a serially correlated error term. That is exactly what the Cochrane-Orcutt procedure offers.

2. Cochrane-Orcutt (CO) Algorithm

[11] Consider a multiple regression model

$$y_t = \alpha + \sum_{k=1}^{K} \beta_k x_t^{(k)} + u_t, \tag{1}$$

where the error term u_t follows an AR(1) process with the autocorrelation at lag 1 being ρ (whose value is unknown at this stage):

$$u_t = \rho u_{t-1} + \epsilon_t, \tag{2}$$

where ϵ is a series of serially independent numbers with mean zero and constant variance. If ρ is not zero then the Gauss-Markov theorem cannot be applied and therefore OLS is generally not an efficient estimator of model parameters, and other methods are called for. One such method is that suggested by *Cochrane and Orcutt* [1949], which modifies equation (1) by rewriting it for $t - 1$ instead of t, multiplying all terms by ρ, subtracting the result from equation (1), using equation (2), and rearranging terms to obtain

$$(y_t - \rho y_{t-1}) = \alpha(1 - \rho) + \sum_{k=1}^{K} \beta_k \left(x_t^{(k)} - \rho x_{t-1}^{(k)} \right) + \epsilon_t \tag{3}$$

for t = 2,..,N.

[12] Equation (3) is a regression equation with modified variables, coefficients, and an error term that satisfies the Gauss-Markov theorem. We have, however, introduced one new parameter, namely ρ, which prevents us from applying OLS directly.

[13] We can solve the problem iteratively by first estimating (using OLS) α and the βs from equation (3) for an initial guess of ρ. Then, using the values of α, β just determined a new value for ρ is found (using equations (1) and (2)), which is then held fixed and used to find new values for the α, βs, and so on, until convergence occurs (see *Ramanathan* [2002, p. 393] for a detailed description of the algorithm). Discussions have appeared as to whether this technique guarantees convergence to a good solution [e.g., *Dufour et al.*, 1980]; the outcome of the discussion seems to be that a grid search through parameter space almost always reveals that the iterated solution is the best one.

[14] The Cochrane-Orcutt method is well known in the econometrics literature, but has, it seems, not been widely appreciated outside this field. In the following sections we will show that there is reason to take notice of the method in geophysics, as it offers advantages in realistic situations where OLS is commonly applied without being wholly appropriate.

3. Illustrating the Cochrane-Orcutt Algorithm by Applying It to Artificial Series With Known Properties: Variables Without Serial Correlation

[15] We next show the results of applying the CO algorithm to artificial problems. We generate suitable regression problems from the following model:

$$y_t = \alpha + \beta_1 x_t^{(1)} + \beta_2 x_t^{(2)} + u_t, \tag{4}$$

where $x_t^{(1,2)}$ are two uncorrelated predictor vectors generated from normally distributed random numbers; furthermore,

$$u_t = \rho u_{t-1} + \epsilon_t, \tag{5}$$

is the noise, where ϵ is normally distributed white noise.

[16] After generating independent random vectors $x^{(1,2)}$ we generate y_t (in equation (4)) by picking values of α, β_1, and β_2, and the auto-correlated series u_t is generated by picking a value for ρ and generating the series iteratively. We estimate α, β_1, β_2 using OLS and the CO method. In Figure 1 we show the results of 1000 simulations of this procedure for $\rho = 0.7$, which is a realistic value as we shall see below.

[17] We note that the disturbance u causes a spread in the estimates of the coefficients, but that these are centered on the correct solutions; hence, there is no evidence of a bias on average, in accordance with theory.

[18] In summary, we have tested the performance of OLS versus the CO method in trials with independent and "white" predictors with auto-correlated additive noise. We have shown that if residuals have structure, in the sense of having a serial correlation different from 0, then the CO method will outperform OLS in determining regression coefficients more the larger ρ is.

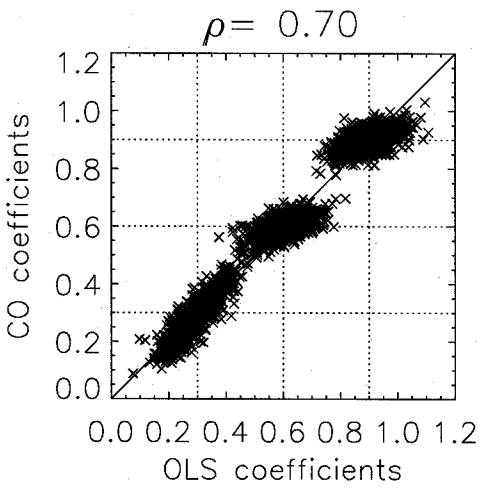

Figure 1. Results of simulations of OLS and CO regression on 1000 simulated data sets. From lower left to upper right, the clouds of crosses are the results for the constant term α and the regression coefficients β_1 and β_2; these were assigned the values 0.3, 0.6, and 0.9 in the model, respectively (the dotted lines). Independent white noise time series of length 100 points were used. The noise added to the model is auto-correlated, with $\rho = 0.7$. There is clearly a larger spread in the OLS regression coefficients. Bias in the estimates seem low: the values are centered on the model values (dotted lines).

4. Application of the CO Method to a Climate Reconstruction Based on Proxies

[19] Instrumental climate series are rarely as long as are the "proxy" series that can be developed from natural records, such as tree ring data, ice cores, lake varves, coral rings, and so on. These natural observables may reflect environmental conditions to some extent, and by calibrating these against instrumental series we can obtain climate information back in time, before instrumental records began. The calibration of the proxy can in its simplest form be performed by a regression, and the present discussion about how well regression methods perform is relevant.

[20] One early attempt to calibrate temperature proxies against instrumental data was that of Landsberg and Groveman [*Groveman and Landsberg*, 1979; *Groveman*, 1979], who utilised a technique whereby supposed proxies for global mean temperatures were related to an instrumental temperature curve, using multiple linear regression. Although the data available to Landsberg and Groveman were limited compared to the much larger data collections now used in climate reconstructions, and the method, in the form chosen by those authors, is not now commonly used, we chose the example in order to show the need for CO instead of OLS for the calibration. The illustrative powers of the example are undiminished by the choice of data and the details of the method of that work.

[21] Very briefly, the Landsberg and Groveman method consists of scaling or calibrating climate proxies against a constructed Northern Hemisphere mean instrumental record [*Borzenkova et al.*, 1976] through multiple regression. The proxies include long instrumental temperature series, tree rings, and winter temperatures estimated from dates of lake freezings [*Groveman*, 1979]. Not all proxies have the same length, so in the application of Landsberg and Groveman proxies were chosen that all ended near the end of the instrumental record but which started at different times, from 1579 AD and forward in time. Sets of proxy time series were chosen on the basis of the years in which they overlapped. In this way a final reconstruction was patched together from many segments, each of which are the results of a calibration during the instrumental period (1881–1954, or 1881–1975 depending on the proxy data and reconstruction period), but being used only for a specific time interval before this era. Twenty-eight such intervals occur.

[22] We first reconstructed the method of Landsberg and Groveman from the data published [*Groveman*, 1979]. The residuals had auto-correlations at lag 1 from 0.1 to 0.42, depending on the choice of proxy data. The residuals' auto-correlation were next tested for significance using the Bartlett cumulated periodogram test [*Bartlett*, 1966]; the residuals are significantly auto-correlated for some choices of proxy data. This conclusion was also obtained using another test for serial correlation in time series: the Durbin Watson test [*Draper and Smith*, 1981]. The Durbin Watson test is a test of a statistic d, which in our case is calculated from the residuals $e_1, .., e_n$:

$$d = \frac{\sum_{t=2}^{n} (e_t - e_{t-1})^2}{\sum_{t=1}^{n} e_t^2}. \quad (6)$$

[23] The relationship between ρ and d is $d \approx 2(1 - \rho)$, so the range for d is between 0 and 4, and d is approximately 2 when $\rho = 0$. The value of d obtained from the residuals is compared to critical values (e.g., the tables of *Draper and Smith* [1981]). If d is less than a lower limit then the null hypothesis of no serial correlation in the residuals can be rejected, whereas if d is above an upper limit the hypothesis cannot be rejected. Additionally, values in between are indeterminate.

[24] The DW test gave similar results to the Bartlett test, namely, that the residuals, in the instrumental calibration range, are serially correlated for those cases when a few

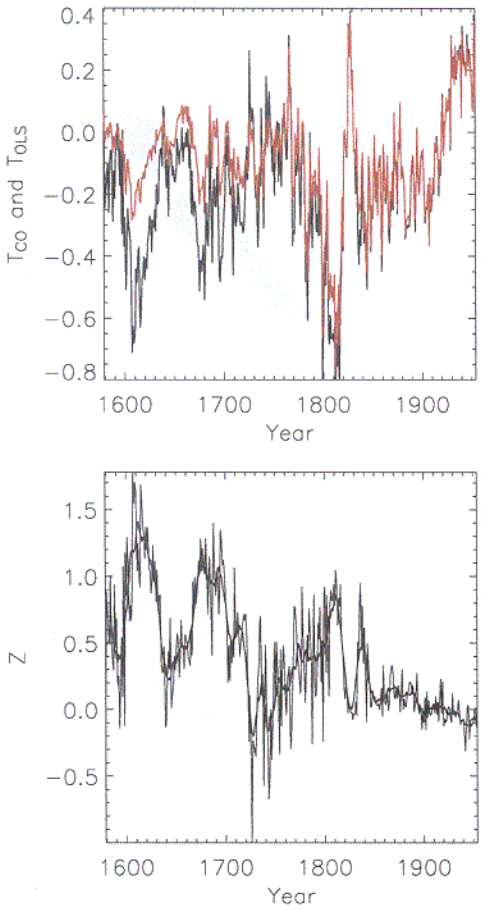

Figure 2. (top) Landsberg-Groveman NH temperature reconstruction reconstructed, using OLS (thin line) and CO (red line). (bottom) Difference between the CO and the OLS reconstructions, and error limits on the difference, generated from an error propagation calculation and the unexplained variances on both curves (see text).

proxy time series from tree ring data are used, notably the first period from 1579–1658 AD (at the 99% significance level). Values of the d statistic close to the lower limit, but inside the "indeterminate range" were obtained for other early intervals, notably 1706–1764 and 1817–1820. There is therefore support for recalculating the temperature reconstruction with CO substituted for OLS, with the expectation that significantly different results could be obtained for the early years of the reconstruction.

[25] We therefore replaced the OLS regressions by the CO algorithm and derived a new reconstructed temperature curve. The original and the new curve are shown in Figure 2 (top). The difference between the two and its uncertainty are shown in Figure 2 (bottom). The uncertainty on the difference is calculated from the errors on each curve, ΔT_{OLS} and ΔT_{CO}, as $\sqrt{\Delta T_{OLS}^2 + \Delta T_{CO}^2}$. The uncertainties on $T_{OLS,CO}$ are estimated from the variance of the residuals in the calibration interval.

[26] We see that there are considerable differences between the reconstructions in the early years (e.g., 0.4°C near 1600 AD), and a tendency for a slope in the difference toward nearly zero difference for the 19th century.

[27] In order to show that this qualitative impression of a significant change corresponds to a quantifiable improvement in reconstruction results, we perform several statistical tests next.

[28] First, we calculate the proportion of unexplained variance during the calibration period, using "out of sample data." The "out of sample data" are constructed in a sliding window that moves through the calibration period. The window is long enough to ensure statistical independence between the middle point and the interval endpoints, and the middle point is chosen as a member of the "out of sample data," progressively. The subinterval length was chosen as twice the longest decorrelation time of any data (proxy or calibration temperature) in the calibration period. The largest serial data correlation was 0.527 which corresponds to a decorrelation time of about 3 years. An interval of 7 years was therefore taken out of the data, progressively, the 3 years on either side of the central point discarded and the central point used to build the set of independent data. This test on independent data shows that the level of uncertainty on the temperature difference ($T_{OLS} - T_{CO}$) is near 0.15°C. Thus parts of the 17th century have a significant difference in reconstructed temperatures, with our reconstruction being warmer for most years in the proxy period.

[29] Second, we apply a standard statistical test to the means of the two temperature reconstructions in sliding subintervals 51 years long, starting in 1579 and moving forward. We find that Student's t-test indicates a significant difference in series means up to the late 1600s. The p values for the period after 1700 are all above 0.1, indicating that it is not possible to reject, at the 90% significance level, the hypothesis that the means are the same after 1700. These two tests thus give us some time-resolved information on when the CO and OLS method give significantly different results, in this particular example.

[30] We finally apply a third test which will give information about the difference between the two methods in the whole interval from 1579 to 1880. This method is based on regressing $T_{OLS} - T_{CO}$ against time and testing whether the regression coefficient is significant. Significance is estab-

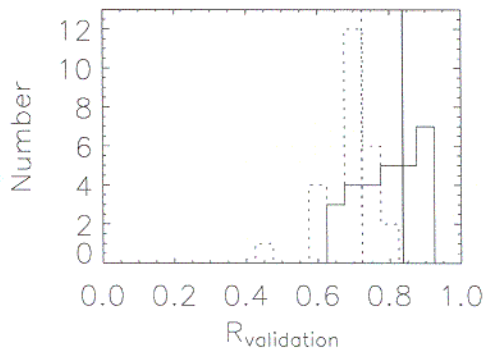

Figure 3. Comparing CO and OLS on independent data. The solid line histogram gives the distribution of R (Pearson's correlation coefficient) for the 28 possible regressions using the CO method, while dotted lines indicate OLS results. Median values are plotted as vertical lines. R_{CO} is about 17% larger than R_{OLS}. See text for explanation of the validation process.

Table 1. Results of Students *t*-Test Applied to Distributions of Correlations for OLS and CO[a]

Quantity	*t*	*p*
R	4.76932	0.0000144

[a]The statistics are calculated for the assumption that variances of the two distributions compared are equal. Allowing unequal variances does not change the result. We tested the hypothesis that the distributions of *R* from CO and OLS had the same mean value. Variable *t* is the T statistic of this test and *p* is the probability that the hypothesis is true.

lished at the 99.55% level by comparing the absolute value of the slope to the absolute value of the slope of pseudo-randomly generated time series of identical length. The surrogate data series are generated by the method of phase scrambling [*Theiler and Prichard*, 1996], which implies that we are generating time series with power spectra identical to the original series. Series generated in that way will all have different appearances when plotted against time but are all picked from a population of series statistically similar to the original series. Of 10,000 trials on surrogate data the observed slope (absolute value) was only exceeded 45 times by chance, implying a significance level of 99.55%, i.e., we can certainly reject the hypothesis that the two series, taken over their whole length, are identical.

[31] These three tests show us a generally significant difference between *T* reconstructed with the OLS and the CO methods, and a particularly large difference in the 1600s.

5. Validation on Independent Data

[32] We have thus shown that a significant difference exists in the results from OLS and CO; let us now consider whether it is possible to show that the results are better with CO than OLS. We do this by a validation method based on the correlation between instrumental data and independent predictions of the instrumental era data, using a sliding window technique. This is applied to both the OLS and the CO methods, and subsequently compared.

[33] We perform the validation by withholding data in successive windows sliding through the instrumental era. Using windows that are twice as wide as the estimated decorrelation time of the data (see above), we regress the proxies against the windowed instrumental data, use the derived regression constant and coefficients to estimate the value of the instrumental data at window midpoint, and thus, by sliding the window and repeating the above procedure, generate independent estimates of the instrumental data, finally calculating the correlation between the predicted instrumental data and the instrumental data. As there are 28 possible choices of regressions to perform we receive 28 values of *R* (Pearson's correlation coefficient) from the CO and OLS methods, and proceed to test these distributions for significant differences.

[34] The distributions of R_{CO} and R_{OLS} were compared (see Figure 3). Median R_{CO} is about 17% larger than R_{OLS}. We tested whether the distributions of *R* are significantly different between CO and OLS, using a standard Student's *t* test. Table 1 shows that the hypothesis that they have the same mean values can be rejected. Thus validation on independent data shows that a significant improvement in correlation is possible when using the CO method instead of the OLS method.

6. Summary

[35] We have by example shown how important the effects of serially correlated residuals can be for the results of regressions, and offered a remedy for situations when regression must be performed in the presence of such autocorrelated residuals, namely the Cochrane-Orcutt method.

[36] In a climatological application we have shown the extent of the effects of using CO instead of OLS in a proxy reconstruction. We have validated the CO and OLS methods on independent data during the instrumental era and shown that the CO reconstruction was significantly better correlated with the target than the OLS reconstruction, and shown that during the proxy-only era significant differences between reconstructed temperatures using CO and OLS exist.

[37] Although the chosen proxy-based temperature reconstruction method may no longer be current, the use of proxy-based reconstruction methods to build a description of past climate is growing. Methods used in recent climate reconstructions include principal component regression and canonical analysis [*Luterbacher et al.*, 2002; *Jones and Mann*, 2004]. While the first is a multiple regression, the second "is at the top of a hierarchy of regression modelling approaches" [*Barnett and Preisendorfer*, 1987, p. 1827]. Therefore both methods potentially suffer from the shortcomings of OLS. A generalized Cochrane-Orcutt algorithm could profitably be applied to these methods, but is beyond the scope of this paper.

[38] **Acknowledgment.** We wish to thank Jörg Clostermann and Richard Reichel for discussions and careful reading of the manuscript and acknowledge support from the Danish Climate Centre.

References

Barnett, T. P., and P. Preisendorfer (1987), Origins and levels of monthly and seasonal forecast skill for United States surface air temperatures determined by canonical correlation analysis, *J. Clim.*, *115*, 1825–1850.

Bartlett, M. (1966), *An Introduction to Stochastic Processes*, 2nd ed., Cambridge Univ. Press, New York.

Borzenkova, I., K. Y. Vinnikov, L. Spirina, and D. Stekhnovskii (1976), Variation of Northern Hemisphere air temperature from 1881 to 1975, *Meteorol. Gidrologiya*, *7*, 27–35.

Box, G., and G. M. Jenkins (1976), *Time Series Analysis: Forecasting and Control*, rev. ed., Holden-Day, San Francisco, Calif.

Cochrane, D., and G. H. Orcutt (1949), Applications of least square regression to relationships containing autocorrelated error term, *J. Am. Stat. Assoc.*, *44*, 32–61.

Draper, N., and H. Smith (1981), *Applied Regression Analysis*, 2nd ed., John Wiley, Hoboken, N. J.

Dufour, J.-M., M. Gaudry, and T. C. Liem (1980), The Cochrane-Orcutt procedure, numerical examples of multiple admissible minima, *Econ. Lett.*, *6*, 43–48.

Fritts, H. C., T. J. Blasing, B. P. Hayden, and J. E. Kutzbach (1971), Multivariate techniques for specifying tree-growth and climate relationships and for reconstructing anomalies in paleoclimate, *J. Appl. Meteorol.*, *10*, 845–864.

Groveman, B. S. (1979), Reconstruction of Northern Hemisphere temperature: 1579–1880, Ph.D. thesis, Univ. of Md., College Park.

Groveman, B. S., and H. E. Landsberg (1979), Simulated Northern Hemisphere temperature departures 1579–1880, *Geophys. Res. Lett.*, *6*, 767–769.

Guiot, J. (1985), The extrapolation of recent climatological series with spectral canonical regression, *J. Climatol.*, *5*, 325–335.

Harvey, A. (1990), *The Econometric Analysis of Time Series*, Philip Alan, New York.

Jones, P., and M. E. Mann (2004), Climate over past millenia, *Rev. Geophys.*, *42*, RG2002, doi:10.1029/2003RG000143.

Luterbacher, J., E. Xoplaki, D. Dietrich, R. Rickli, J. Jacobeit, C. Beck, D. Gyalistras, C. Schmutz, and H. Wanner (2002), Reconstruction of sea level pressure fields over the eastern North Atlantic and Europe back to 1500, *Clim. Dyn.*, *18*, 545–561.

Ramanathan, R. (2002), *Introductory Econometrics With Applications*, 5th ed., Southwest Coll. Publ., Mason, Ohio.

Rutherford, S., M. E. Mann, T. J. Osborn, R. S. Bradley, K. R. Briffa, M. K. Hughes, and P. D. Jones (2005), Proxy-based Northern Hemisphere surface temperature reconstructions: Sensitivity to methodology, predictor network, target season and target domain, *J. Clim.*, *18*, 2308–2329.

Theiler, J., and D. Prichard (1996), Constrained-realization Monte-Carlo method for hypothesis testing, *Physica D*, *94*, 221–235.

von Storch, H., and F. W. Zwiers (Eds.) (1999), *Statistical Analysis in Climate Research*, Cambridge Univ. Press, New York.

T. Schmith and P. Thejll, Danish Meteorological Institute, Climate Research Division, Lyngbyvej 100, DK-2100 Copenhagen Ø, Denmark. (ts@dmi.dk; pth@dmi.dk)