PERGAMON

Computers & Geosciences 28 (2002) 69–72

COMPUTERS & GEOSCIENCES

Short Note

TAUEST: a computer program for estimating persistence in unevenly spaced weather/climate time series ☆

Manfred Mudelsee*

Institute of Meteorology, University of Leipzig, Stephanstr 3, D-04103 Leipzig, Germany

Received 20 May 2000; received in revised form 20 March 2001; accepted 1 April 2001

1. Introduction

Persistence is characteristic for weather and climate fluctuations (e.g. Wilks, 1995). Quantifying the memorizing ability from proxy records of past climates extends our understanding of natural weather and climate variability. The simplest yet successful persistence model is a first-order autoregressive (AR(1)) process where a fluctuation depends only on its own immediate past plus a random component (Gilman et al., 1963; Mann and Lees, 1996). Because the proxy data (from sediment or ice cores) are usually unevenly spaced in time, a simple estimation of the AR(1) model via the autocorrelation coefficient ρ has to be replaced by fitting the AR(1) model (Robinson, 1977)

$$\begin{aligned} &x(1) \sim N(0;1),\\ &x(i) = x(i-1)\exp\{-[t(i)-t(i-1)]/\tau\} + \varepsilon(i),\\ &\quad i = 2,\dots,n, \end{aligned} \tag{1}$$

to the (scaled) data x. Therein, t is time, $\varepsilon(i) \sim N(0; 1-\exp\{-2[t(i)-t(i-1)]/\tau\})$ is the heteroscedastic random component and $\tau > 0$ is the decay period of the autocorrelation function of model (1). We denote τ as *persistence time*. Various measures of persistence exist in literature (e.g. von Storch and Zwiers, 1999). The advantage of using model Eq. (1) is that τ corresponds directly to the relevant physical time scale.

The Fortran/Gnuplot program TAUEST estimates τ. Since geological interpretation of the result requires knowledge about the statistical accuracy, TAUEST includes Monte Carlo simulations. TAUEST shows scatterplots for comparing data and fit residuals which help to assess the suitability of the fitted model.

The fit subroutine of TAUEST was also combined with the SPECTRUM frequency-analysis package for unevenly spaced data (Schulz and Stattegger, 1997) to quantify the analogue of persistence in the frequency domain—red noise (Schulz and Mudelsee, 2002).

2. Estimation procedure

For x to describe weather/climate fluctuations, deterministic signals have to be removed prior to the estimation. TAUEST allows (1) subtraction of the mean and (2) linear unweighted detrending. For some data it might be advisable to consider more complicated functions such as a seasonal cycle.

Since Eq. (1) can be fitted only numerically, it is advantageous to rewrite it and to introduce scaling. Defining $a := \exp(-1/\tau)$ leads to

$$\begin{aligned} &x(1) \sim N(0;1),\\ &x(i) = x(i-1)a^{[t(i)-t(i-1)]} + \varepsilon(i), \quad i = 2,\dots,n, \end{aligned} \tag{2}$$

and the parameter a is bound between 0 and 1. The least-squares estimator $\hat{a}$ thus minimizes

$$S(a) = \sum_{i=2}^{n} \left\{x(i) - x(i-1)a^{[t(i)-t(i-1)]}\right\}^2. \tag{3}$$

$\hat{\tau} = -1/\ln(\hat{a})$ is then the persistence time estimate (without bias correction). Note that in the situation of equidistance $(t(i) - t(i-1) =: d = \text{constant})$ Eq. (3) leads to $\hat{a} = \hat{\rho}^{1/d}$, where

$$\hat{\rho} = \sum_{i=2}^{n} x(i)x(i-1) \Big/ \sum_{i=2}^{n} x(i-1)^2 \tag{4}$$

☆ Code available from server at http://www.iamg.org/CGEditor/index.htm or http://www.uni-leipzig.de/~meteo/MUDELSEE/.

*Tel.: +49-0341-9732866; fax: +49-0341-9732899.

E-mail addresses: mudelsee@rz.uni-leipzig.de (M. Mudelsee).

PII: S0098-3004(01)00041-3

is the usual autocorrelation coefficient estimator. The choice of time units (t-scaling) adopts the finding of own numerical experiments, namely that $\hat{\tau}$ (units) $= 1$ offers a stable solution and fast convergence. (The final result $\hat{\tau}$ is rescaled to the original time units.) x is scaled to unit variance. $S(a)$ is minimized using Brent's method (Press et al., 1989) with starting value $a_{\text{start}} = 1/\text{e}$.

Monte Carlo simulations (Efron and Tibshirani, 1993) allow the accuracy of the estimate $\hat{\tau}$ to be evaluated. Let $x^*_{j=1}$ be a time series generated after Eq. (1) with a value of τ equal to $\hat{\tau}$. The estimation yields $\hat{\tau}^*_{j=1}$. Simulation and estimation is repeated, B times in total. The median $med(\hat{\tau}^*_{j=1,\ldots,B})$ may be compared with $\hat{\tau}$ to evaluate the estimation bias. The 5% and 95% percentiles constitute an equi-tailed 90% confidence interval for $\hat{\tau}$. To suppress simulation noise, at least 2000 simulations are recommended (Efron and Tibshirani, 1993).

It is well known that in the situation of equidistant data and mean detrending, the estimator $\hat{\rho}$ has a bias of approximately $-(1+3\rho)/(n-1)$ (Kendall and Stuart, 1966). Numerical experiments with artificial time series suggest a similar bias for the analogous parameter in a situation of non-equidistance,

$$\hat{\rho}_{\text{non}} = (\hat{a})^{\bar{d}}, \tag{5}$$

where $\bar{d}$ is the average spacing. In many practical applications (n more than a few hundred), the bias is negligible. However, TAUEST offers a correction, by prescribing the value of τ for the Monte Carlo simulations to a higher value than $\hat{\tau}$. If $med(\hat{\tau}^*_{j=1,\ldots,B}) \approx \hat{\tau}$, the prescribed value is an unbiased estimate.

3. Computer program

TAUEST is a batch program, combining calculation (Fortran 77) and visualization (Gnuplot 3.6 or higher). Use the DOS mode or the DOS window on your PC. (Installation to other systems is straightforward.) Gnuplot executable should be in the path. Start the program with the command "TAUEST". Supply path and data file name. The time series is plotted on the screen. Proceed with "Enter". Supply n. Supply whether t =age or time. $\hat{\tau}$ is not invariant in this respect (although the effect is negligible unless n is very small). Choose the detrending method (no/mean/linear). "No" assumes that the original time series has zero mean and no trend. Select "mean" or "linear" when either the data should be accordingly detrended or have already been detrended.

TAUEST calculates $\hat{\tau}$ and plots $S(a)$. This allows to check whether $S(a)$ is well-behaved or has, for example, several local minima. That situation might occur for extremely irregular data, very small n or an unsuited model. (Press "Enter" key.) The lag-1 scatterplot of the data ($x(i-1)$ vs $x(i)$) and of the residuals ($r(i-1)$ vs $r(i)$) is shown. The residuals are calculated as

$$r(i) = x(i) - x(i-1)\exp\{-[t(i)-t(i-1)]/\hat{\tau}\},\quad i = 2,\ldots,n, \tag{6}$$

and plot—in case of a proper fit and a suited AR(1) model—as a cloud without orientation whereas autocorrelated x produce a cloud oriented along the 1 : 1 line.

TAUEST then displays the setting and the estimation result. Input the number of Monte Carlo simulations and seed the random number generator. The empirical relation, CPU time (s) $\approx 0.00018nB$, was determined for a PC with a 433 MHz Pentium processor. If you choose then to correct for bias, TAUEST gives the value of $\hat{\rho}_{\text{non}}$. After entering the prescribed value ρ_{pre} (you may try $\rho_{\text{pre}} = \hat{\rho}_{\text{non}} + (1+3\hat{\rho}_{\text{non}})/(n-1)$), the value $\tau = -\bar{d}/\ln(\rho_{\text{pre}})$ is then used for the simulations.

After the simulations, TAUEST plots a kernel density estimate (Silverman, 1982, 1986) of $\hat{\tau}^*_{j=1,\ldots,B}$ together with $med(\hat{\tau}^*_{j=1,\ldots,B})$, the 5% and 95% percentiles and the uncorrected estimate $\hat{\tau}$. If the median is sufficiently close to $\hat{\tau}$, you obtain a bias-free persistence estimate with 90% confidence interval.

TAUEST writes the scatterplot data, the simulated persistence times and the kernel density values in output files. The source codes of TAUEST on the IAMG server give further details.

4. Example

The oxygen-isotope (δ^{18}O) record from the GISP2 ice core (Grootes and Stuiver (1997), data from UW Quaternary Isotope Laboratory, 2000) document temperature fluctuations over Greenland. Here, we analyse data from the high-sampling interval AD 818–1800 ($n = 8733$, $\bar{d} = 0.11$ years). Data from after AD 1800 are excluded since they may be affected by anthropogenic climate perturbations.

The data (Fig. 1A) show only a small trend and are approximately Gaussian distributed (Fig. 1B).

The estimation with TAUEST, using mean subtraction, yields $\hat{\tau} = 1.83$ years. The lag-1 scatterplots (Figs. 1C and D) indicate a strong reduction of autocorrelation between the residuals and thus attest that the AR(1) model is suited well for describing the data. With n so large, estimation bias is negligible. The 90% confidence interval for $\hat{\tau}$ is [1.63; 2.05] years, as derived from 2000 Monte Carlo simulations (Fig. 1E).

The persistence estimate proved to be robust against the detrending method: Neither linear detrending nor removing the seasonal cycle (using SPECTRUM's harmonic analysis tool) changed $\hat{\tau}$ significantly (1.83 and 1.86 years, respectively). Also, using the coarser 1 m

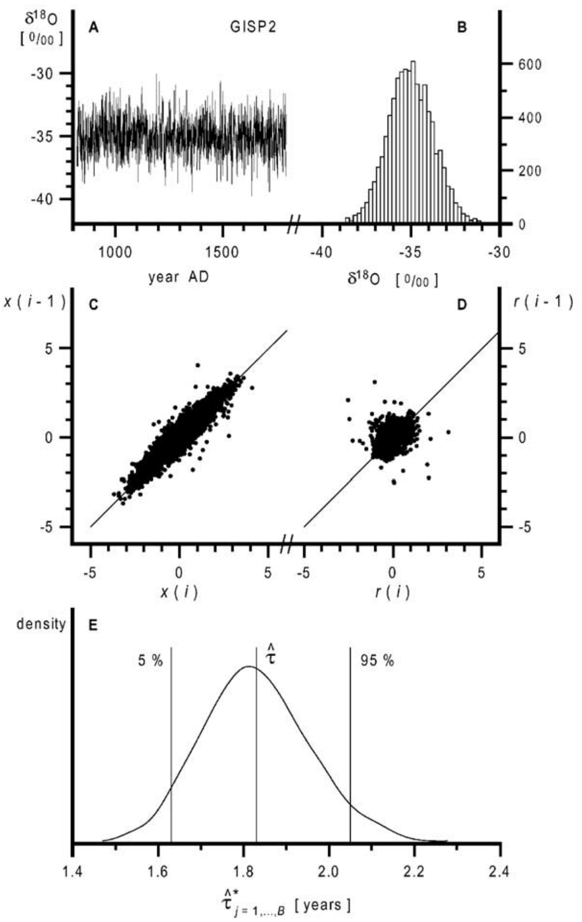

Fig. 1. Persistence estimation for GISP2's δ^{18}O record. (A) Time series. (B) Histogram of data. (C) Lag-1 scatterplot of mean-subtracted and scaled data x, 1 : 1 line. (D) Lag-1 scatterplot of residuals (see Eq. (6)), 1 : 1 line. (E) Kernel density estimate of simulated persistence times. Also shown are 5% and 95% percentiles, and estimate $\hat{\tau}$.

data from GISP2 (Grootes and Stuiver, 1997) gave an indistinguishable result (1.93 years).

The attempt to interpolate the data to equidistance and use $\hat{\rho}$ for persistence estimation would, in general, introduce an artificial statistical dependence and thus lead to an overestimation of τ. A value of 2.15 years would result in case of GISP2's high-sampling data, and a value of 3.98 years in case of the 1 m data.

We note that previous persistence quantification attempts using a power-law model (e.g., Koscielny-Bunde et al., 1998) lack tests of model suitability and also do not provide confidence intervals.

To conclude, using the persistence estimation program TAUEST for unevenly spaced data, evidence is found that natural temperature fluctuations of the past $\sim$1100 years (as recorded in ice from Greenland) have the same memory as was found for recent global temperature (equidistant data), namely about two years (Allen and Smith, 1994).

Acknowledgements

M. Schulz kindly provided a version of SPECTRUM for very large n. The present work was supported by a Marie Curie Research Fellowship from the European Union, a Conference Grant from the Royal Society, London, and a Research Grant from the Deutsche Forschungsgemeinschaft, Bonn. Executables of TAUEST for various systems are available from the author's homepage Manfred Mudelsee, http://www.uni-leipzig.de/~meteo/MUDELSEE/.

References

Allen, M.R., Smith, L.A., 1994. Investigating the origins and significance of low-frequency modes of climate variability. Geophysical Research Letters 21 (10), 883–886.

Efron, B., Tibshirani, R.J., 1993. An Introduction to the Bootstrap. Chapman & Hall, London, 436pp.

Gilman, D.L., Fuglister, F.J., Mitchell, J.M., 1963. On the power spectrum of "red noise". Journal of the Atmospheric Sciences 20 (2), 182–184.

Grootes, P.M., Stuiver, M., 1997. Oxygen 18/16 variability in Greenland snow and ice with 10^{-3}–10^{5} year time resolution. Journal of Geophysical Research 102 (C12), 26455–26470.

Kendall, M.G., Stuart, A., 1966. The Advanced Theory of Statistics, Vol. 3. Griffin, London, 552pp.

Koscielny-Bunde, E., Bunde, A., Havlin, S., Roman, H.E., Goldreich, Y., Schellnhuber, H.-J., 1998. Indication of a universal persistence law governing atmospheric variability. Physical Review Letters 81 (3), 729–732.

Mann, M.E., Lees, J.M., 1996. Robust estimation of background noise and signal detection in climatic time series. Climatic Change 33 (3), 409–445.

Press, W.H., Flannery, B.P., Teukolsky, S.A., Vetterling, W.T., 1989. Numerical Recipes (Fortran). Cambridge University Press, Cambridge, 702pp.

Robinson, P.M., 1977. Estimation of a time series model from unequally spaced data. Stochastic Processes and their Applications 6, 9–24.

Schulz, M., Mudelsee M., 2002. REDFIT: Estimating red-noise spectra directly from unevenly spaced paleoclimatic time series. Computers & Geosciences, in press.

Schulz, M., Stattegger, K., 1997. SPECTRUM: Spectral analysis of unevenly spaced paleoclimatic time series. Computers & Geosciences 23 (9), 929–945.

Silverman, B.W., 1982. Kernel density estimation using the Fast Fourier Transform. Applied Statistics 31 (1), 93–99.

Silverman, B.W., 1986. Density Estimation for Statistics and Data Analysis. Chapman & Hall, London, 175pp.

von Storch, H., Zwiers, F.W., 1999. Statistical Analysis in Climate Research. Cambridge University Press, Cambridge, 484pp.

UW Quaternary Isotope Laboratory, GISP2 Oxygen Isotope Data, 2000. http://www.depts.washington.edu/qil/datasets/gisp2_main.html, March 7, 2000.

Wilks, D.S., 1995. Statistical Methods in the Atmospheric Sciences. Academic Press, San Diego, 467pp.