Eur. Phys. J. Special Topics **174**, 49–63 (2009)

DOI: 10.1140/epjst/e2009-01089-3

THE EUROPEAN PHYSICAL JOURNAL SPECIAL TOPICS

Regular Article

Break function regression

A tool for quantifying trend changes in climate time series

M. Mudelsee[1,2,a]

[1] Climate Risk Analysis, Hannover, Germany
[2] AWI, Bremerhaven, Germany

Abstract. The break is a continuous function consisting of two linear parts. It serves as a regression model for trend changes in time series. A typical application field of such a model is climatology. We introduce break-model fitting by combining a weighted least-squares criterion with a brute-force search. We explain how to determine error bars and confidence intervals for the break model parameters by means of autoregressive bootstrap resampling. Our approach takes into account the statistical properties of real-world climatological problems: non-Gaussian distributional shape, serial dependence, uneven time spacing and timescale uncertainties. A Monte Carlo experiment shows the excellent coverage performance of bootstrap bias-corrected and accelerated confidence intervals for data sizes above 100 or 200. An application quantifies trend changes in modelled Arctic river runoff during the interval from 1936 to 2001.

1 Introduction

It has become a commonplace to call the climate system "nonlinear" and to describe a rather fast change as "abrupt". The statistical model suggested here helps to put numbers with error bars to those suggestions and therefore to assess their validity. Evidently, "abrupt" changes seem to occur also in finances and the economy. Also that realm, amenable to a rational, if not scientific treatment [1,2], may benefit from quantifications that are supported by error bars.

We decompose a stochastic, continuous-time process, $X(T)$, in a "classical" manner as follows:

$$X(T) = X_{\mathrm{trend}}(T) + S(T) \cdot X_{\mathrm{noise}}(T). \tag{1}$$

X is a random variable describing, for example, climate; T is time; $X_{\mathrm{trend}}(T)$ is the trend component; $S(T)$ is the standard deviation, denoted in this paper as variability, a function that scales the unit-standard deviation noise component, $X_{\mathrm{noise}}(T)$. A formulation for discrete time, $T(i)$, where i is an integer, is written readily; we use then the abbreviations $X(i) \equiv X(T(i))$, $X_{\mathrm{trend}}(i) \equiv X_{\mathrm{trend}}(T(i))$, and so forth. A time series, $\{t(i), x(i)\}_{i=1}^{n}$ is a sample of size n of the process. Note that the present paper adopts a definition of the expression "noise" that allows for serial dependence. Other definitions, not allowing for serial dependence, exist in the literature.

The trend is the time-dependent mean or first-order moment and estimating this component using time series data can yield first-order insights. This is true also for the climate sciences, where concepts such as "abrupt changes" or "gradual temperature rises" are analysed. For good reasons is the functional form of the trend not specified. This gives the researcher freedom how to describe the long-term, systematic components of the observed climate system.

[a] e-mail: mudelsee@mudelsee.com

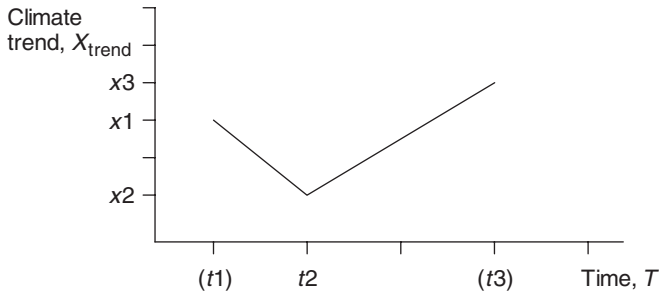

Fig. 1. The break regression model. It has four free parameters: $x1, t2, x2$ and $x3$. (The parameter $t1$ is constrained as the left, $t3$ as the right bound of the time interval.)

That description depends on the data at hand and the research questions of interest. In the present paper, we are interested in changes of the trend. The simplest model for this is the break function (Fig. 1).

The function $S(i)$ is the time-dependent standard deviation (heteroscedasticity) and related to the second-order moment. The formulation of Eq. (1) with time-dependent trend and standard deviation reflects the "classical" definition of climate in terms of mean and variability that was developed in the late 19th and early 20th century [3–5].

What previous work exists on break regression? In statistical science this method is better known as "two-phase regression" [6,7]. It has been applied by Solow [8] to southern hemisphere temperature, 1858–1985. Reinsel and cowokers [9] studied generalized least-squares estimation of the break model. They did so, however, under the often unrealistic assumption that the change-point in time, $t2$, is known. This method was then applied [10,11] to detect trend changes in stratospheric ozone concentrations, 1977–2002, that means in particular, the effects of the Montreal Protocol on Substances that Deplete the Ozone Layer from 1987 and its Amendments. Hinkley [12] mentioned and Julious [13] studied bootstrap resampling for the two-phase regression; the latter paper devised a hypothesis test for the existence of an unknown change-point. Tomé and Miranda [14,15] presented an algorithm for fitting a continuous regression model with several break points to data and applied this method to study changes of Azores temperatures, Northern Hemisphere temperatures, the North Atlantic Oscillation index and Lisbon winter precipitation. The deficit of the presented methodology [14,15] is that no error bars or confidence intervals for the estimated trend parameters are determined.

The present paper gives an estimation algorithm of the break model parameters (Fig. 1) based on weighted least-squares (WLS) combined with a brute-force search for $t2$. WLS takes into account the heteroscedasticity. The paper shows how to construct bootstrap confidence intervals (CIs) for the estimated parameters by means of autoregressive resampling. Such confidence intervals are a type of uncertainty measure robust with respect to (1) the distributional shape and (2) the size of serial dependence of the noise component. This makes the method particularly useful for climate time series analysis, where such difficult noise properties are ubiquitous. Because we employ a first-order autoregressive or AR(1) process that is embedded in continuous time, our method is not restricted to even time spacing. Thus, we avoid interpolation and its distorting influence on estimation results. We explain how the resampling step can be adapted to take timescale errors into account. Errors in $T(i)$ are the rule in paleoclimatology, where one employs climate archives such as sediment cores or speleothems, which have to be dated. A Monte Carlo experiment with a large number of artificially generated time series studies the accuracy of the bootstrap CIs for a realistic setting. Finally, an application to modelled Arctic river runoff explores whether and when anthropogenic forcing (greenhouse gas emissions) lead to a trend change.

The author had previously presented [16] a three-phase regression ("ramp") as a model for climate transitions. The ramp has two unknown change points. As the "break" methodology is rather similar to the "ramp" methodology, readers may find it useful to consult that previous paper. An other piece of own work on climate time series analysis using bootstrap confidence intervals dealt with correlation estimation [17].

2 Method

2.1 Break model estimation

The break regression model, written in continuous time as

$$X_{\text{trend}}(T) = X_{\text{break}}(T) = \begin{cases} x1 + (T - t1)(x2 - x1)/(t2 - t1) & \text{for } T \le t2, \\ x2 + (T - t2)(x3 - x2)/(t3 - t2) & \text{for } T > t2, \end{cases} \tag{2}$$

has four free parameters: $x1$, $t2$, $x2$ and $x3$. An alternative formulation would comprise the four parameters $t2$, $x2$, $\beta_1 = (x2 - x1)/(t2 - t1)$ and $\beta_2 = (x3 - x2)/(t3 - t2)$. The break is useful for describing a change in linear trend at one point $(t2, x2)$, from slope β_1 to β_2.

Assume known variability $S(i)$. The break model can then be fitted by minimizing the weighted least-squares sum,

$$SSQW(x1, t2, x2, x3) = \sum_{i=1}^{n} \left[x(i) - x_{\text{break}}(i)\right]^2 / S(i)^2 \,, \tag{3}$$

where $x_{\text{break}}(i)$ is the discrete-time, sample version of $X_{\text{break}}(T)$ (Eq. (2)).

Because we assume that the break is a suitable description over the whole record length, $t1$ and $t3$ are constrained (Fig. 1) and only one time point, namely $\widetilde{t2} = t(\widetilde{i2})$, where $\widetilde{i2}$ is an integer between 1 and n, needs to be considered as candidate for $t2$. Then the minimizers $\widehat{x1}, \widehat{x2}$ and $\widehat{x3}$ of $SSQW(x1, \widetilde{t2}, x2, x3)$ follow as

$$\begin{aligned} \widehat{x2} &= \big(K_1 K_2/K_3 - K_4 K_5/K_6 - K_7 + K_9\big)/\big(K_1^2/K_3 + K_4^2/K_6 - K_8 - K_{10}\big), \\ \widehat{x1} &= \widehat{x2}\, K_1/K3 - K_2/K_3, \\ \widehat{x3} &= \widehat{x2}\, K_4/K6 + K_5/K_6, \end{aligned} \tag{4}$$

where

$$\begin{aligned} K_1 &= \sum_{i=1}^{\widetilde{i2}} S(i)^{-2}\big[t(i) - t1\big]\big[t(i) - \widetilde{t2}\big]/\big[\widetilde{t2} - t1\big]^2, \\ K_2 &= \sum_{i=1}^{\widetilde{i2}} S(i)^{-2} x(i)\big[t(i) - \widetilde{t2}\big]/\big[\widetilde{t2} - t1\big], \\ K_3 &= \sum_{i=1}^{\widetilde{i2}} S(i)^{-2}\big[t(i) - \widetilde{t2}\big]^2/\big[\widetilde{t2} - t1\big]^2, \\ K_4 &= \sum_{i=\widetilde{i2}+1}^{n} S(i)^{-2}\big[t(i) - t3\big]\big[t(i) - \widetilde{t2}\big]/\big[t3 - \widetilde{t2}\big]^2, \\ K_5 &= \sum_{i=\widetilde{i2}+1}^{n} S(i)^{-2} x(i)\big[t(i) - \widetilde{t2}\big]/\big[t3 - \widetilde{t2}\big], \end{aligned}$$

$$K_6 = \sum_{i=\widetilde{i2}+1}^{n} S(i)^{-2}\left[t(i) - \widetilde{t2}\right]^2 / \left[t3 - \widetilde{t2}\right]^2,$$

$$K_7 = \sum_{i=1}^{\widetilde{i2}} S(i)^{-2} x(i)\left[t(i) - t1\right] / \left[\widetilde{t2} - t1\right],$$

$$K_8 = \sum_{i=1}^{\widetilde{i2}} S(i)^{-2}\left[t(i) - t1\right]^2 / \left[\widetilde{t2} - t1\right]^2,$$

$$K_9 = \sum_{i=\widetilde{i2}+1}^{n} S(i)^{-2} x(i)\left[t(i) - t3\right] / \left[t3 - \widetilde{t2}\right],$$

$$K_{10} = \sum_{i=\widetilde{i2}+1}^{n} S(i)^{-2}\left[t(i) - t3\right]^2 / \left[t3 - \widetilde{t2}\right]^2. \tag{5}$$

To estimate the change-point in time, $t2$, a brute-force search over all candidate points is performed:

$$\left(\widehat{t2}\right) = \text{argmin}\left[SSQW\left(\widehat{x1}, \widetilde{t2}, \widehat{x2}, \widehat{x3}\right)\right]. \tag{6}$$

Computing costs are clearly reduced (by a factor of $\sim n$) compared with estimating the ramp model [16]. The other properties the break shares:

- The solution is a global (parameter hyperspace) optimum.
- The $\widehat{t2}$ estimate is "coarse" because it is from the set $\{t(i)\}_{i=1}^{n}$. However, this is in practice rarely a problem because the error bar for $\widehat{t2}$ is usually larger than the time spacing at around $\widehat{t2}$. See Section 4 for an example.

2.2 Iterative estimation procedure

The following iterative procedure can be applied when $S(i)$ is unknown.

1. Make an initial guess, $\widehat{S}^{(0)}(i)$, of the variability.
2. Estimate the break regression parameters, with the guessed variability used instead of $S(i)$.
3. Calculate $e(i) = x(i) - \widehat{x}_{\text{break}}(i), i = 1, \ldots, n$. The function $\widehat{x}_{\text{break}}(i)$ is the estimated, discrete-time, sample version of $X_{\text{break}}(T)$ (Eq. (2)), obtained by plugging in $\widehat{x1}$ for $x1$, $\widehat{x2}$ for $x2$, $\widehat{x3}$ for $x3$, $\widehat{t2}$ for $t2$ and $t(i)$ for $T(i)$. The $e(i)$ are called the unweighted regression residuals.
4. Obtain a new variability estimate, $\widehat{S}^{(1)}(i)$ from the residuals. This can be done either non-parametrically by smoothing (e.g., running standard deviation of $e(i)$) or fitting a parametric model for $S(i)$ to $\{e(i)\}_{i=1}^{n}$.
5. Go to Step 2 with the new, improved variability estimate until regression estimates converge.

2.3 Confidence intervals

Before explaining bootstrap confidence interval construction for the break model parameters, we review briefly the general theory.

Let θ be the parameter of interest of the climatic process $\{X(T)\}$ and $\widehat{\theta}$ be the estimator. Any meaningful construction lets the estimator be a function of the process, $\widehat{\theta} = g\left(\{X(T)\}\right)$. That means, $\widehat{\theta}$ is a random variable with statistical properties. The standard deviation of $\widehat{\theta}$, denoted as standard error, is the square root of the estimation variance,

$$\text{se}_{\widehat{\theta}} = \left[VAR\left(\widehat{\theta}\right)\right]^{1/2}. \tag{7}$$

The bias of $\widehat{\theta}$ is its expectation minus the true value,

$$\text{bias}_{\widehat{\theta}} = E\left(\widehat{\theta}\right) - \theta. \tag{8}$$

$\text{bias}_{\widehat{\theta}} > 0$ ($\text{bias}_{\widehat{\theta}} < 0$) means a systematic overestimation (underestimation). Desirable estimators have small $\text{se}_{\widehat{\theta}}$ and small $\text{bias}_{\widehat{\theta}}$. In many estimations, a trade-off problem between $\text{se}_{\widehat{\theta}}$ and $\text{bias}_{\widehat{\theta}}$ occurs. A convenient measure is the root mean squared error,

$$\begin{aligned} \text{RMSE}_{\widehat{\theta}} &= \left\{ E\left[\left(\widehat{\theta} - \theta \right)^2 \right] \right\}^{1/2} \\ &= \left(\text{se}_{\widehat{\theta}}{}^2 + \text{bias}_{\widehat{\theta}}{}^2 \right)^{1/2}. \end{aligned} \tag{9}$$

While $\widehat{\theta}$ is a best guess of θ or a point estimate, a CI is an interval estimate that informs how good a guess is. The CI for θ is

$$\text{CI}_{\widehat{\theta},1-2\alpha} = \left[\widehat{\theta}_\text{l}; \widehat{\theta}_\text{u} \right], \tag{10}$$

where $0 \leq 1-2\alpha \leq 1$ is a prescribed value, denoted as confidence level. Of relevance in climatology are the levels 90% ($\alpha = 0.05$) and 95% ($\alpha = 0.025$). $\widehat{\theta}_\text{l}$ is the lower, $\widehat{\theta}_\text{u}$ the upper endpoint of the CI. $\widehat{\theta}_\text{l}$ and $\widehat{\theta}_\text{u}$ are random variables and have statistical properties such as standard error or bias. The properties of interest for CIs are the coverages,

$$\gamma_\text{l} = \text{prob}\left(\theta \leq \widehat{\theta}_\text{l} \right), \tag{11}$$

$$\gamma_\text{u} = \text{prob}\left(\theta \geq \widehat{\theta}_\text{u} \right) \tag{12}$$

and

$$\gamma = \text{prob}\left(\widehat{\theta}_\text{l} < \theta < \widehat{\theta}_\text{u} \right) = 1 - \gamma_\text{l} - \gamma_\text{u}. \tag{13}$$

Exact CIs have coverages, γ, equal to the nominal value $1 - 2\alpha$. Construction of exact CIs requires knowledge of the distribution of $\widehat{\theta}$, which can be achieved only for simple problems. In more complex situations, such as break function regression, only approximate CIs can be constructed. As regards the division of the nominal coverage between the CI endpoints, we adopt a practical approach and consider only equi-tailed CIs, where nominally $\gamma_\text{l} = \gamma_\text{u} = \alpha$. As a second CI property besides coverage, one often studies interval length, $\widehat{\theta}_\text{u} - \widehat{\theta}_\text{l}$, which is ideally small.

2.4 Real world climatological problems

Real problems in the climatologist's world are complex. One source of complexity originates in models and estimation procedures. An example is estimating the break model parameter $t2$, for which an exact CI cannot be calculated because the traditional procedure (via the curvature of $SSQW$) is not applicable as the model (Eq. (2)) is not differentiable with respect to $t2$. The other source of complexity is the noise process with properties such as non-Gaussian distributions and persistence.

Complex problems often defy analytical solutions and require advanced computational methods such as bootstrap resampling (Section 2.7). CIs for complex problems are not exact but only approximate, they exhibit coverage errors.

Coverage error, C, is defined by means of a single-sided CI endpoint [18], for example,

$$C = \gamma_\text{l} - \alpha. \tag{14}$$

If C decreases with sample size as $\mathcal{O}\left(n^{-1/2}\right)$, then the CI is called first-order accurate; if C is of $\mathcal{O}\left(n^{-1}\right)$, then the CI is called second-order accurate; and so forth. The same CI accuracy applies also to two-sided CIs. Desirable approximate CIs have a high-order accuracy.

2.5 AR(1) persistence model

The simplest persistence model is an AR(1) process. Consider an unevenly spaced time series; the spacing, $d(i) = t(i) - t(i-1)$, is not constant. Denote the average of $d(i)$ as $\bar{d}$. The AR(1) process for uneven spacing is given by

$$\begin{aligned} X(1) &= \mathcal{E}_{\mathrm{N}(0,\,1)}(1), \\ X(i) &= \exp\left\{-\left[T(i) - T(i-1)\right]/\tau\right\} \cdot X(i-1) \\ &\quad + \mathcal{E}_{\mathrm{N}(0,\,1-\exp\{-2[T(i)-T(i-1)]/\tau\})}(i), \\ &\quad i = 2, \ldots, n. \end{aligned} \tag{15}$$

We write $\mathcal{E}_{\mathrm{N}(\mu,\,\sigma^2)}(\cdot)$ for a Gaussian random variable with mean μ and variance σ^2.

The parameter τ defines the "equivalent autocorrelation coefficient", a', via $a' = \exp(-\bar{d}/\tau)$ for the case of uneven time spacing. For even spacing, a' equals a, the ordinary AR(1) autocorrelation parameter. Introducing in Eq. (15) a heteroscedastic innovation term, $\mathcal{E}$, makes this process strictly stationary.

The discrete-time version of the AR(1) process (Eq. (15)) under the condition of $a \geq 0$ has a uniquely related counterpart in continuous time [19]. This embedding ensures that no problem of model identification [20] arises.

The estimation of a (even spacing) is not trivial because of estimation bias. For an AR(1) process with unknown mean, which is the usual case in climatology, the estimator of a is

$$\widehat{a} = \frac{\sum_{i=2}^{n} \left[X(i) - \bar{X}\right] \cdot \left[X(i-1) - \bar{X}\right]}{\sum_{i=2}^{n} \left[X(i) - \bar{X}\right]^2}, \tag{16}$$

where $\bar{X} = \sum_{i=1}^{n} X(i)/n$ is the sample mean. The approximate expectation of $\widehat{a}$ is [21]

$$E\left(\widehat{a}\right) \simeq a - (1 + 3a)\,/\,(n-1)\,. \tag{17}$$

That means, $\widehat{a}$ underestimates a. Eq. (17) can be used to correct for the negative bias. However, Eq. (17) is valid only for a not too large [21]. A bias formula applicable also to large values (above, say, 0.9) of a has yet only been derived for the simpler case of known mean [22].

Regarding the estimation of τ (uneven spacing), Monte Carlo simulations (Mudelsee, M.: Climate Time Series Analysis: Classical Statistical and Bootstrap Methods, Springer, Heidelberg, manuscript in preparation) show that the least-squares estimator of $a' = \exp(-\bar{d}/\tau)$ has a bias similar in size to the bias of $\widehat{a}$. Mudelsee [23] presented a least-squares estimation procedure for τ with bias correction and bootstrap percentile confidence intervals.

2.6 Bootstrap principle

The idea of the bootstrap is to use the data to mimic the unknown distribution function, which is now replaced by the empirical distribution function. Mimicking the data generating process is achieved by drawing random samples from the data set. The simplest form is the ordinary bootstrap, that means, drawing one by one with replacement. Preserving the persistence properties of time series data requires adaptions of the ordinary bootstrap, one parametric adaption is given in Section 2.7. Re-applying the estimation procedure to the new random samples, called resamples, yields new estimates, called replications. Section 2.8 explains CI construction using the replications. Fig. 2 shows the bootstrap principle and the workflow.

The bootstrap means that numerical simulation replaces theoretical derivation of the distribution of an estimator. This can be an improvement, especially if the complexity of the problem defies obtaining an exact theoretical result. But also the bootstrap is not free of assumptions. The main requirement is that the properties distributional shape and persistence are preserved by the bootstrap resampling. There is also "simulation noise", but this can be made arbitrarily

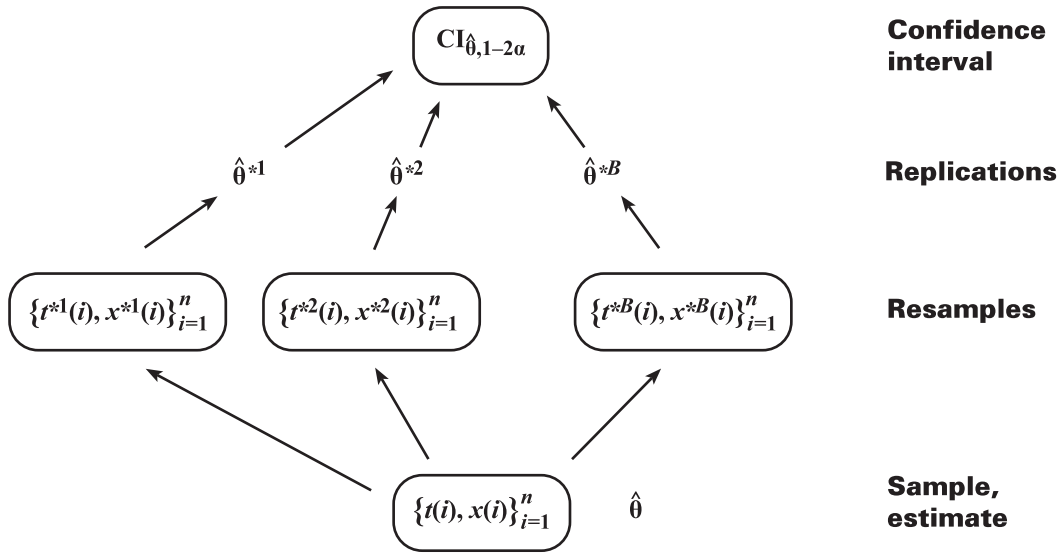

Fig. 2. Bootstrap principle for constructing confidence intervals. Given is a sample of data and an estimate of a parameter of interest. Using bootstrap resampling new data sets – resamples – are formed. The resamples ideally preserve fully the statistical properties of the process that generated the data. In the simple case where $t(i)$ are perfectly known and also persistence is absent, $t^*(i) = t(i), i = 1, \ldots, n$, and $\{x^*(i)\}_{i=1}^n$ is obtained by drawing randomly, one by one and with replacement, n elements from the set of sample values, $\{x(i)\}_{i=1}^n$. The resamples are marked with an asterisk and numbered with an index, $b = 1, \ldots, B$. The number of resamples, B, is typically a few thousand. The estimator is applied to each of the resamples, yielding B new estimates – the replications. The set of replications $\{\widehat{\theta}^{*b}\}_{b=1}^B$ is then used for CI construction (Section 2.8).

small by using a large number of resamples, B. Assumptions made at CI construction add to the fact that in complex situations, bootstrap CIs, like classical CIs, are not exact but approximate. However, bootstrap CIs seem to be more flexible and require less strict assumptions. A word on usage of "simulation": we follow statisticians who reserve this for Monte Carlo experiments, where statistical methods are tested by means of artificial data from models with pre-defined properties. The bootstrap procedure, on the other hand, is referred to as "resampling".

2.7 Autoregressive bootstrap resampling

The autoregressive bootstrap algorithm (ARB), shown in Fig. 3, is the ordinary bootstrap applied to the white-noise residuals, $\epsilon(i)$.

How are the white-noise residuals determined? Consider first the unweighted regression residuals, $e(i)$. Then take the weighted residuals, $r(i) = e(i)/\widehat{S}(i)$. Using the AR(1) persistence model for $r(i)$, the residuals $\epsilon(i)$ are then formed. $\epsilon(i)$ are treated as realizations of a white-noise process. Note that the distributional shape needs not be Gaussian, which is a considerable advantage.

We list the ARB for even spacing. Although the bias correction (Step 7) is only approximate, this is considered an important step because ignoring bias can lead to a bad bootstrap performance [24]. Scaling, as done in Step 8 using a factor $\left[1 - (\widehat{a}')^2\right]^{-1/2}$, is non-standard. It has the computational advantage that no transient behaviour is required in Step 11. Centering (Step 9) achieves that the resample-generating process has expection zero, as the white-noise process is supposed to have. After Step 9, a further scaling with a factor $[(n-1)/(n-2)]^{1/2}$, as has been suggested [24], is omitted. This factor is in the general case only approximate [25] and its effect is considered negligible against the other uncertainties. Lahiri [26] explains the "traditional" method to generate a number of samples that is very much larger than n at

Step 1	Data	$\{t(i), x(i)\}_{i=1}^{n}$
Step 2	Resampled times unchanged	$\{t^*(i)\}_{i=1}^{n} = \{t(i)\}_{i=1}^{n}$
Step 3	Estimated trend and variability	$\{\widehat{x}_{\text{break}}(i)\}_{i=1}^{n}$, $\{\widehat{S}(i)\}_{i=1}^{n}$
Step 4	Weighted residuals	$\{r(i)\}_{i=1}^{n} = \{[x(i) - \widehat{x}_{\text{break}}(i)]/\widehat{S}(i)\}_{i=1}^{n}$
Step 5	Assume $\{r(i)\}_{i=1}^{n}$ to come from	AR(1) model
Step 6	Estimate AR(1) parameter	$\widehat{a}$
Step 7	Bias correction	$\widehat{a}'$
Step 8	White-noise residuals	$\epsilon(i) = [r(i) - \widehat{a}' \cdot r(i-1)] \times \left[1 - (\widehat{a}')^2\right]^{-1/2}, \quad i = 2, \ldots, n$
Step 9	Centering	$\tilde{\epsilon}(i) = \epsilon(i) - \sum_{i=2}^{n} \epsilon(i)/(n-1)$
Step 10	Draw $\tilde{\epsilon}^*(j), j = 2, \ldots, n$, with replacement from	$\{\tilde{\epsilon}(i)\}_{i=2}^{n}$
Step 11	Resampled weighted residuals	$r^*(1)$ drawn from $\{r(i)\}_{i=1}^{n}$, $r^*(i) = \widehat{a}' \cdot r^*(i-1) + \left[1 - (\widehat{a}')^2\right]^{1/2} \cdot \tilde{\epsilon}^*(i), \quad i = 2, \ldots, n$
Step 12	Resampled data	$x^*(i) = \widehat{x}_{\text{break}}(i) + \widehat{S}(i) \cdot r^*(i), \quad i = 1, \ldots, n$

Fig. 3. Autoregressive bootstrap algorithm (ARB), even spacing.

Step 10 and use them at Step 11 for extracting $r^*(i)$ from the transient sequence. The advantage of the non-standard formulation with scaling (Step 8) corresponds to the advantage of strict stationarity of the non-standard formulation of the AR(1) model (Section 2.5).

The ARB for uneven spacing corresponds basically to the ARB for even spacing, where the persistence parameter, a, is replaced by $\exp\{-[t(i) - t(i-1)]/\tau\}$.

A method related to the ARB is the surrogate data approach. However, this is a simulation rather than a resampling method. No residuals are drawn as in the ARB. Instead, climate equation residuals $\{r^*(i)\}_{i=1}^{n}$ are obtained by numerical simulation from the persistence model with estimated parameters. Because also the distributional shape is specified, the surrogate data approach is bounded stronger by parametric restrictions than the ARB. Therein lies its danger: it is more prone than the ARB to systematic errors from violated assumptions. Contrary to the assertion in a review on surrogate time series [27], this approach is *not* the common choice in the bootstrap literature. An other resampling approach to preserve serial dependence is the moving block bootstrap (MBB), which we shall shortly visit in the application example (Section 4).

2.8 Bootstrap bias-corrected and accelerated confidence interval

Estimation of θ is repeated for the resamples, $\{t^{*b}(i), x^{*b}(i)\}_{i=1}^{n}, b = 1, \ldots, B$. This yields the bootstrap replications, $\{\widehat{\theta}^{*b}\}_{b=1}^{B}$. The replications are used to construct equi-tailed $(1 - 2\alpha)$ confidence intervals, $\text{CI}_{\widehat{\theta},1-2\alpha}$.

Two approaches, standard error based and percentile based, dominate theory and practice of bootstrap CI construction. The estimated bootstrap standard error is the sample standard

error of the replications,

$$\widehat{\mathrm{se}}_{\widehat{\theta}^*} = \left\{ \sum_{b=1}^{B} \left[\widehat{\theta}^{*b} - \left\langle \widehat{\theta}^{*b} \right\rangle \right]^2 \Big/ (B-1) \right\}^{1/2}, \tag{18}$$

where $\langle \widehat{\theta}^{*b} \rangle = \sum_{b=1}^{B} \widehat{\theta}^{*b}/B$. The percentiles result from the empirical distribution function of the replications. The accuracy of bootstrap CIs depends critically on the similarity (in terms of standard errors or percentiles) of the distribution of the bootstrap replications and the true distribution, $f(\widehat{\theta})$. Various concepts exist for accounting for the deviations between the two distributions.

Suppressing "simulation noise" requires more resamples for percentile estimation than for bootstrap standard error estimation. We follow the textbook recommendation [18] and set for CI construction $B = 1999$. For a reasonable α value such as 0.025, this means that a number of 50 replications is outside the percentile bound. (For the mere objective of bootstrap standard error determination, using $B = 400$ is sufficient [18].)

The bootstrap bias-corrected and accelerated (BCa) confidence interval [18,28] is

$$\mathrm{CI}_{\widehat{\theta},1-2\alpha} = \left[\widehat{\theta}^*(\alpha 1), \widehat{\theta}^*(\alpha 2) \right], \tag{19}$$

where $\widehat{\theta}^*(\alpha')$ is the $100\alpha'$th percentage point of the empirical distribution of $\{\widehat{\theta}^{*b}\}_{b=1}^{B}$,

$$\alpha 1 = F\left(\widehat{z}_0 + \frac{\widehat{z}_0 + z(\alpha)}{1 - \widehat{a}\,[\widehat{z}_0 + z(\alpha)]} \right) \tag{20}$$

and

$$\alpha 2 = F\left(\widehat{z}_0 + \frac{\widehat{z}_0 + z(1-\alpha)}{1 - \widehat{a}\,[\widehat{z}_0 + z(1-\alpha)]} \right). \tag{21}$$

$F(\cdot)$ is the standard normal distribution function. $z(\alpha')$ is the percentage point of the normal distribution. For example, $z(0.95) \approx 1.644854$. $\widehat{z}_0$, the bias correction, is computed as

$$\widehat{z}_0 = F^{-1}\left(\frac{\#\left\{ \widehat{\theta}^{*b} < \widehat{\theta} \right\}}{B} \right), \tag{22}$$

where $\#\{\widehat{\theta}^{*b} < \widehat{\theta}\}$ means the number of replications where $\widehat{\theta}^{*b} < \widehat{\theta}$ and $F^{-1}(\cdot)$ is the inverse function of $F(\cdot)$. Note that in Eqs. (20), (21) and (23), $\widehat{a}$ denotes a quantity called acceleration [18]. It is computed as

$$\widehat{a} = \frac{\sum_{j=1}^{n} \left[\left\langle \widehat{\theta}_{(j)} \right\rangle - \widehat{\theta}_{(j)} \right]^3}{6 \left\{ \sum_{j=1}^{n} \left[\left\langle \widehat{\theta}_{(j)} \right\rangle - \widehat{\theta}_{(j)} \right]^2 \right\}^{3/2}}, \tag{23}$$

where $\widehat{\theta}_{(j)}$ is the jackknife value of $\widehat{\theta}$. Consider the original sample with the jth point removed, that is, $\{t(i), x(i)\}, i = 1, \ldots, n, i \neq j$. The jackknife value is then the value of $\widehat{\theta}$ calculated using this sample of reduced size. The average, $\langle \widehat{\theta}_{(j)} \rangle$, is given by $[\sum_{j=1}^{n} \widehat{\theta}_{(j)}]/n$.

$\widehat{z}_0$ corrects for the median estimation bias; for example, if just half of the replications have $\widehat{\theta}^{*b} < \widehat{\theta}$, then $\widehat{z}_0 = 0$. The acceleration takes into account scale effects, which arise when the standard error of $\widehat{\theta}$ itself depends on the true parameter value, θ.

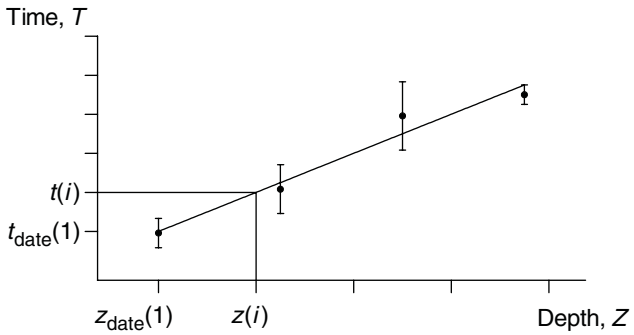

Fig. 4. Linear timescale model. The dating points $\{z_{\mathrm{date}}(j), t_{\mathrm{date}}(j)\}_{j=1}^{n_{\mathrm{date}}}$ are shown as filled symbols, their dating errors $S_{\mathrm{date}}(j)$ as vertical bars ($\pm$). (Strictly speaking, the dating error is an unknown random variable with standard deviation $S_{\mathrm{date}}(j)$.) The linear WLS regression fitted to the dating points (solid straight line) is used to convert the depth value, $z(i)$, of a measurement ($x(i)$) into time, $t(i)$.

2.9 Timescale errors

Up to now we assumed the time values to be exactly known. The ARB algorithm (Section 2.7) employed a parametric AR(1) model for resampling the noise $\{X_{\mathrm{noise}}(i)\}_{i=1}^{n}$ and left the times $\{T(i)\}_{i=1}^{n}$ unchanged.

In the presence of timescale uncertainties the ARB can be adapted by parametric modelling; we introduce this algorithm as timescale-autoregressive bootstrap or timescale-ARB.

1. Data, $\{t(i), x(i)\}_{i=1}^{n}$.
2. Weighted regression residuals, $r(i)$.
3. Resample $\{x^*(i)\}_{i=1}^{n}$ by applying ARB to $r(i)$ under $\{t(i)\}_{i=1}^{n}$.
4. Model parametrically $\{t^*(i)\}_{i=1}^{n}$.

The ARB algorithm is first applied to the regression residuals using the time values $t(i)$ to produce the $x^*(i)$ resamples. Then the $t^*(i)$ are resampled from a parametric model of the accumulation process of the climate archive.

A parametric model for $\{T^*(i)\}_{i=1}^{n}$ comes from a physical description of the accumulation process of the climate archive. Consider as a simple example a linear accumulation and a number n_{date} of dated points (Fig. 4), which is a good approximation for many sedimentary and speleothem time series.

Generating the simulated time points $\{t^*(i)\}_{i=1}^{n}$ is straightforward (Fig. 5). The assumptions made are:

1. linear accumulation process with
2. positive slope and
3. independent, Gaussian distributed dating errors.

The constraint "positive slope" refers to the assumed monotonic growth of an archive; it is taken into account by retaining only those model simulations with a positive slope (Fig. 5).

2.10 Model suitability

We have made a number of assumptions for estimation and confidence interval construction: suitability of the break model, serial dependence of AR(1) form and, in case of timescale errors, suitability of the timescale model. Evidently, it should be tested in practical applications how well the assumptions are fulfilled. Besides quantitative measures of the goodness of fit it is important to evaluate fit quality also using graphical tools of residual analysis [29].

Step 1	Dating points	$\{z_{\mathrm{date}}(j), t_{\mathrm{date}}(j)\}_{j=1}^{n_{\mathrm{date}}}$,
	dating errors	$\{S_{\mathrm{date}}(j)\}_{j=1}^{n_{\mathrm{date}}}$
Step 2	Parameter estimates	$\widehat{\beta}_0$ (intercept), $\widehat{\beta}_1$ (slope)
	of linear WLS regression to	$\{z_{\mathrm{date}}(j), t_{\mathrm{date}}(j)\}_{j=1}^{n_{\mathrm{date}}}$
Step 3	Timescale	$t(i) = \widehat{\beta}_0 + \widehat{\beta}_1 z(i),\ i = 1, \ldots, n$
Step 4	Simulated dating points	$T^*_{\mathrm{date}}(j) = T_{\mathrm{date}}(j) + S_{\mathrm{date}}(j) \cdot \mathcal{E}_{\mathrm{N}(0,\,1)}(j),$
		$j = 1, \ldots, n_{\mathrm{date}}$
Step 5	WLS regression to	$\{z_{\mathrm{date}}(j), t^*_{\mathrm{date}}(j)\}_{j=1}^{n_{\mathrm{date}}}$
Step 6	Parameter estimates, simulation	$\widehat{\beta}^*_0, \widehat{\beta}^*_1$
Step 7	If $\widehat{\beta}^*_1 > 0$, then	
	calculate simulated timescale	$t^*(i) = \widehat{\beta}^*_0 + \widehat{\beta}^*_1 z(i),\ i = 1, \ldots, n$

Fig. 5. Timescale resampling, linear accumulation model.

3 Monte Carlo experiment

A number of $n_{\mathrm{sim}} = 475$ random samples were generated from $X(i) = X_{\mathrm{break}}(i) + X_{\mathrm{noise}}(i)$, where the prescribed break model parameters (Eq. (2)) are $x1 = 2.0, t2 = 0.5n, x2 = 1.0$ and $x3 = 4.0$, the prescribed times are $T_{\mathrm{pre}}(i) = i, i = 1, \ldots, n$ and the noise is a Gaussian AR(1) process with $a = 1/\mathrm{e} \approx 0.37$.

Timescale errors were subsequently introduced as follows. A linear timescale model (Fig. 4) with $n_{\mathrm{date}} = 2$ dating points and independent, Gaussian distributed timescale errors was used to generate the $T(i)$ as $T(i) = T_{\mathrm{pre}}(i) + \mathcal{E}_{\mathrm{N}(0,\,25.0)}(i)$ for $i = 1$, $T(i) = T_{\mathrm{pre}}(i) + \mathcal{E}_{\mathrm{N}(0,\,100.0)}(i)$ for $i = n$, and then by linear interpolation for $i = 2, \ldots, n-1$.

The bootstrap BCa CIs were constructed using timescale-ARB resampling (Fig. 3), $B = 1999$ and $\alpha = 0.025$.

The results (Table 1) reveal that for data sizes above 100 or 200, the time parameter ($t2$), the level parameters ($x1, x2, x3$) and the slopes (β_1, β_2) have excellent coverage performances also for heteroscedastic timescale errors.

We emphasize that for the Monte Carlo experiment we have selected "nice" $t2$ values, namely in the middle of the time interval. Other experiments, carried out for the break and also the ramp regression model, indicate that coverage accuracy may be reduced when the change-point time (like $t2$ for the break model) is not in the middle but closer to one of the bounds of the time interval.

4 Application: Trend change in modelled Arctic river runoff

HadCM3 [30] is a coupled Atmosphere–Ocean General Circulation Model for the global domain. The atmospheric component has a horizontal grid spacing of 2.5° in latitude by 3.75° in longitude and 19 vertical levels. The oceanic component has 20 vertical levels on a 1.25° by 1.25° grid. The timestep used [31] for integrating the differential equations representing the atmospheric component was 30 minutes, for the oceanic component one hour. The total interval simulated (∼140 years) was longer than the data shown (Fig. 6). The time series were constructed from four ensemble runs, that is, runs with identical forcings but different initial conditions. The initial conditions used were taken from states separated by 100 years in a

Table 1. Monte Carlo experiment, break regression with timescale errors and AR(1) noise of normal shape: CI coverage performance. The confidence level is $1 - 2\alpha = 90\%$. The standard error of γ is nominally $[2\alpha(1 - 2\alpha)/n_{\mathrm{sim}}] = 0.01$. (The γ values are rounded.)

n	$\gamma_{\widehat{x1}}$	$\gamma_{\widehat{t2}}$	$\gamma_{\widehat{x2}}$	$\gamma_{\widehat{x3}}$	$\gamma_{\widehat{\beta}_1}$	$\gamma_{\widehat{\beta}_2}$	Nominal
10	0.63	0.92	0.63	0.62	0.64	0.82	0.95
20	0.74	0.90	0.76	0.76	0.75	0.83	0.95
50	0.87	0.89	0.85	0.89	0.87	0.90	0.95
100	0.91	0.88	0.89	0.93	0.90	0.92	0.95
200	0.93	0.95	0.93	0.92	0.92	0.94	0.95
500	0.94	0.95	0.95	0.96	0.95	0.96	0.95
1000	0.95	0.95	0.95	0.95	0.95	0.94	0.95

HadCM3 run, in which external forcings where set to have no year-to-year variations. Plotted are annual-mean ensemble averages, whereby the model year starts on 1 December.

The natural forcing included changes in the amount of stratospheric aerosols stemming from volcanic eruptions and variations in solar irradiation. The anthropogenic forcing included changes in atmospheric concentrations of CO_2, methane, sulfate aerosols and ozone. The river runoff records were generated [31] by summing the precipitation contributions from affected grid cells north of 65° N. Model simulations can be used to analyse past and predict future climate changes. Questions associated with the data are those after the size and the timing of changes in runoff as a result of an intensified hydrological cycle caused by anthropenically induced warming.

For the interval 1936–2001 ($n = 66$), the break model was fitted to the data from the combined natural and anthropogenic forcing using WLS (with $S(i)$ linearly increasing from 70 $\mathrm{km^3a^{-1}}$ to 120 $\mathrm{km^3a^{-1}}$ within the fit interval). The break fit, shown in Fig. 6a as solid line, has following parameter estimates with bootstrap standard errors: change-point, $\widehat{t2} = 1973 \pm 6$, $\widehat{x2} = 3238 \pm 26$ $\mathrm{km^3a^{-1}}$; slopes, $\widehat{\beta}_1 = -1.8 \pm 1.6$ $\mathrm{km^3a^{-2}}$, $\widehat{\beta}_2 = 9.7 \pm 3.6$ $\mathrm{km^3a^{-2}}$.

Note that the error estimates are based not on ARB but on MBB resampling [32], with $B = 400$ and a block length selection oriented on the data size and the AR(1) autocorrelation parameter [33,34]. Own Monte Carlo experiments revealed that in many cases parametric (ARB) and nonparametric (MBB) resampling yield similar results when the assumption of AR(1) dependence is met. It has been shown by climate theory [35] that, in general, the AR(1) dependence structure results from the averaging of short-term weather components in the climate system.

The fitted break regression reveals a change-point at $\widehat{t2} = 1973 \pm 6$ (Fig. 6a). This date is close to the per-eye estimate [31] of 1965. Before the change the trend is downward, however, with large error bars; after the change it is strongly upward (Fig. 6a). Wu and coworkers [31] describe a second model run, namely runoff with natural forcing only. This time series (1936–1996) does not exhibit significant slope changes of the break fit (Fig. 6b): The slope in the earlier part, before the break at $\widehat{t2} = 1989 \pm 20$, is $\widehat{\beta}_1 = -1.3 \pm 42.4$ $\mathrm{km^3a^{-2}}$, the slope in the later part is $\widehat{\beta}_2 = 15.1 \pm 48.6$ $\mathrm{km^3a^{-2}}$.

5 Discussion

5.1 More than one break point

Consider the case that the trend model contains not one single break point (Fig. 1) but several break points. Denote their number as n_{break}. This is an important generalization because it allows an objective analysis of longer climate time series with several changes. An application field for this type of analysis may be the succession of Dansgaard–Oeschger events (rapid warmings) during the last ice-age cycle.

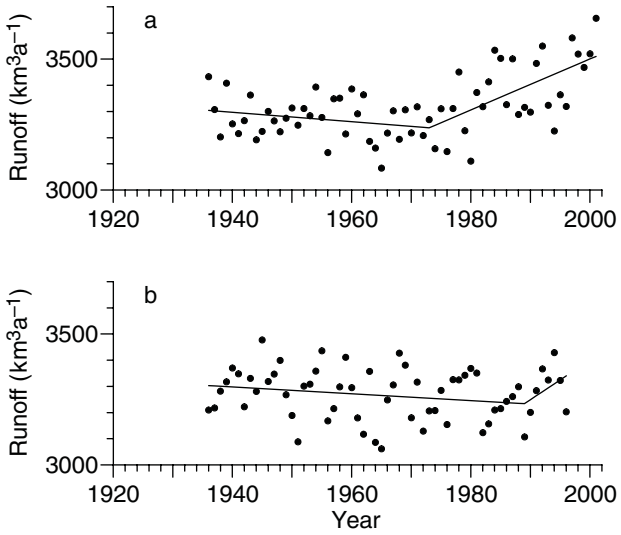

Fig. 6. Break change-point regression fitted to modelled Arctic river runoff. (a) Natural and anthropogenic forcing, (b) natural forcing only.

The practical difficulty of using this model stems from the fact that the number of possible combinations of break points within a time series (data size n) is given by $\binom{n-1}{n_{\text{break}}}$. This is a number that groes rapidly with n_{break}. This means a rapid growth in computing costs required to carry out the estimations.

5.2 Random walk and unit-root processes

The random walk process emerges as a special case of an AR(1) process for $\tau \to \infty$ (uneven spacing and homoscedastic innovation term) or $a \to 1$ (even spacing and homoscedastic innovation term). For this process, the variance increases linearly with time, that means, the random walk is not stationary [20]. In climatology, where the variables are within certain bounds, the random walk has to be modified to serve as a noise model. In that manner, it has been applied to short-term temperature fluctuations [36]. In case of Pleistocene timescales, Wunsch [37] suggested a random walk to explain the 100 ka ice-age cycle. He put bounds to the ice-volume variable $X_{\text{noise}}(i)$; when the system attempts to leave the permitted range, it is thrown back. Mostly other fields than climatology, such as econometrics, apply tests of the hypothesis "$a = 1$" for the autoregressive process, or its generalization ("unit-roots tests") for the ARMA(p, q) process [38]. An exception is a paper by Stern and Kaufmann [39], where unit-roots were identified in tests to hemispheric temperature records, circa 1855–1995. Because these tests generally have poor power (loosely speaking, detection probability) [40], and because of the nonstationarity, we do not consider random-walk models for $X_{\text{noise}}(i)$.

5.3 Generalized least-squares estimation

A generalized least-squares sum, containing the estimated variance–covariance matrix, may replace $SSQW$ and lead to smaller error bars and narrower CIs than from using WLS [9]. This gain may be large in the case of highly resolved time series ($d(i) \ll \tau$) because then the degree of redness (the non-diagonal elements of the variance–covariance matrix) may be high.

6 Concluding remarks

Trend-change regression models are a useful tool for quantifying transitions in the climate or economic systems. A simple model is the break (Fig. 1), which has previously been studied by various researchers in both fields.

The present paper introduces a combination of WLS and a brute-force search to estimate the break regression parameters. It presents, also a novelty, bootstrap BCa CI construction based on timescale-ARB resampling. This method takes into account the following typical properties of climate time series: non-Gaussian distributional shape, serial dependence, uneven time spacing and timescale uncertainties.

The application of the introduced method to modelled Arctic river runoff demonstrates the usefulness for analysing real-world climatological problems.

I thank Peilu Wu for supplying the climate model data. I acknowledge the constructive remarks made by all four reviewers. I thank the DAMES 2008 workshop organizers for providing a scientific avenue with stimulating input from various disciplines on climate data analysis. A version of the break-fit software (without timescale simulation) is freely available via the Climate Risk Analysis website (www.mudelsee.com).

References

1. P. Perron, J. Econom. **80**, 355 (1997)
2. J. Bai, P. Perron, Econometrica **66**, 47 (1998)
3. E. Brückner, Geogr. Abhandl. **4**, 153 (1890)
4. J. Hann, *Lehrbuch der Meteorologie* (Tauchnitz, Leipzig, 1901)
5. W. Köppen, *Die Klimate der Erde: Grundriss der Klimakunde* (de Gruyter, Berlin, 1923)
6. D.V. Hinkley, Biometrika **57**, 1 (1970)
7. D.V. Hinkley, Biometrika **58**, 509 (1971)
8. A.R. Solow, J. Clim. Appl. Meteorol. **26**, 1401 (1987)
9. G.C. Reinsel, E.C. Weatherhead, G.C. Tiao, A.J. Miller, R.M. Nagatani, D.J. Wuebbles, L.E. Flynn, J. Geophys. Res. **107**, 4078 (2002)
10. G.C. Reinsel, Geophys. Res. Lett. **29**, 1451 (2002)
11. G.C. Reinsel, A.J. Miller, E.C. Weatherhead, L.E. Flynn, R.M. Nagatani, G.C. Tiao, D.J. Wuebbles, J. Geophys. Res. **110**, D16306 (2005)
12. D.V. Hinkley, J.R. Statist. Soc. B **50**, 321 (1988)
13. S.A. Julious, J.R. Statist. Soc. D **50**, 51 (2001)
14. A.R. Tomé, P.M.A. Miranda, Geophys. Res. Lett. **31**, L02207 (2004)
15. A.R. Tomé, P.M.A. Miranda, Nonlin. Processes Geophys. **12**, 451 (2005)
16. M. Mudelsee, Comput. Geosci. **26**, 293 (2000)
17. M. Mudelsee, Math. Geol. **35**, 651 (2003)
18. B. Efron, R.J. Tibshirani, *An Introduction to the Bootstrap* (Chapman and Hall, London, 1993)
19. P.M. Robinson, Stoch. Proc. Appl. **6**, 9 (1977)
20. M.B. Priestley, *Spectral Analysis and Time Series* (Academic Press, London, 1981)
21. M.G. Kendall, Biometrika **41**, 403 (1954)
22. M. Mudelsee, Statist. Pap. **42**, 517 (2001)
23. M. Mudelsee, Comput. Geosci. **28**, 69 (2002)
24. R.A. Stine, J. Am. Statist. Assoc. **82**, 1072 (1987)
25. S.C. Peters, D.A. Freedman, J. Bus. Econom. Statist. **2**, 406 (1984)
26. S.N. Lahiri, *Resampling Methods for Dependent Data* (Springer, New York, 2003)
27. T. Schreiber, A. Schmitz, Physica D **142**, 346 (2000)
28. B. Efron, J. Am. Statist. Assoc. **82**, 171 (1987)
29. D.C. Montgomery, E.A. Peck, *Introduction to Linear Regression Analysis*, 2nd edn. (Wiley, New York, 1992)
30. C. Gordon, C. Cooper, C.A. Senior, H. Banks, J.M. Gregory, T.C. Johns, J.F.B. Mitchell, R.A. Wood, Clim. Dynam. **16**, 147 (2000)
31. P. Wu, R. Wood, P. Stott, Geophys. Res. Lett. **32**, L02703 (2005)

32. H.R. Künsch, Ann. Statist. **17**, 1217 (1989)
33. E. Carlstein, Ann. Statist. **14**, 1171 (1986)
34. M. Sherman, F.M. Speed Jr., F.M. Speed, J. Appl. Statist. **25**, 333 (1998)
35. K. Hasselmann, Tellus **28**, 473 (1976)
36. O. Kärner, J. Geophys. Res. **107**, 4415 (2002)
37. C. Wunsch, Clim. Dynam. **20**, 353 (2003)
38. W.A. Fuller, *Introduction to Statistical Time Series*, 2nd edn. (Wiley, New York, 1996)
39. D.I. Stern, R.K. Kaufmann, Clim. Change **47**, 411 (2000)
40. C. Chatfield, *The Analysis of Time Series: An Introduction*, 6th edn. (Chapman and Hall, Boca Raton, 2004)