



Unbiased Proxy Calibration

Manfred Mudelsee^{1,2} 

Received: 10 January 2023 / Accepted: 8 November 2023

© The Author(s) 2023

Abstract

The linear calibration model is a powerful statistical tool that can be utilized to predict an unknown response variable, Y , through observations of a proxy or predictor variable, X . Since calibration involves estimation of regression model parameters on the basis of a limited amount of noisy data, an unbiased calibration slope estimation is of utmost importance. This can be achieved by means of state-of-the-art, data-driven statistical techniques. The present paper shows that weighted least-squares for both variables estimation (WLSXY) is able to deliver unbiased slope estimations under heteroscedasticity. In the case of homoscedasticity, besides WLSXY, ordinary least-squares (OLS) estimation with bias correction (OLSBC) also performs well. For achieving unbiasedness, it is further necessary to take the correct regression direction (i.e., of Y on X) into account. The present paper introduces a pairwise moving block bootstrap resampling approach for obtaining accurate estimation confidence intervals (CIs) under real-world climate conditions (i.e., non-Gaussian distributional shapes and autocorrelations in the noise components). A Monte Carlo simulation experiment confirms the feasibility and validity of this approach. The parameter estimates and bootstrap replications serve to predict the response with CIs. The methodological approach to unbiased calibration is illustrated for a paired time series dataset of sea-surface temperature and coral oxygen isotopic composition. Fortran software with implementation of OLSBC and WLSXY accompanies this paper.

Keywords Proxy calibration · Linear regression · Bootstrap resampling · Paleoclimatology · Corals

This work has been funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation), project number 468589022 (M.M.), within the SPP 2299, project number 441832482.

✉ Manfred Mudelsee
manfred.mudelsee@uni-potsdam.de; mudelsee@climate-risk-analysis.com

¹ Institute of Geosciences, University of Potsdam, Karl-Liebknecht-Str. 24–25, 14476 Potsdam, Germany

² Climate Risk Analysis, Kreuzstr. 27, 37581 Bad Gandersheim, Germany

1 Introduction

A proxy is an observable variable, X , that can be utilized to predict an unobservable variable, Y . An ideal relation for the task of prediction is a mathematical one-to-one relation between X and Y , where one X value corresponds to one unique Y value, and vice versa (Kotz et al. 1986, p. 323 therein). The simplest and easiest to interpret one-to-one relation is a linear model,

$$Y = \beta_0 + \beta_1 \cdot X, \quad (1)$$

where β_0 and β_1 are the intercept and slope parameter, respectively.

Proxy models are widely employed in natural sciences, social sciences, medicine, and engineering (Cronin 2010; Mahnken et al. 2014; Falcy et al. 2016). The application in the present paper is from paleoclimatology. Here, natural archives (e.g., marine sediments or corals) are sampled, and proxy variables (e.g., chemical concentrations or isotopic compositions) are measured on the samples to predict climate variables (e.g., precipitation or temperature) back in time—over thousands and millions of years, when no devices (e.g., rain gauges or thermometers) were available (Bradley 1999). Typically for paleoclimatology, but for other application fields as well, the proxy relation is established for the recent time period, where paired data for X and Y are available. The determined proxy relation is then called a calibration, and the estimated calibration parameters (β_0 and β_1 in the linear case) are taken to predict the response variable, Y , through observations of the predictor variable, X .

In the real world, however, a perfect calibration model is never realized, since the calibration data are affected by observational errors (McClelland et al. 2021). In the case of the paleoclimatic application (Sect. 4), the Y data (temperature) of the calibration show errors stemming from the imperfect device (thermometer, which also may not always be available on-site but only somewhere nearby) and the X data (oxygen isotopic composition) reveal fluctuations due to the counting errors in the mass spectrometer and other sources (e.g., imperfect standard material used by the device, biological effects in the sample, or chronological uncertainties of the archive).

Observational error is usually described by a zero-mean stochastic process, and in climate time series analysis (Mudelsee 2014) it has become practice to decompose the process into a variability component (possibly dependent on time, T), such as $S_X(T)$, which is multiplied by a zero-mean and unit-standard deviation noise process, for example, $X_{\text{noise}}(T)$; the combined error process is then given by $S_X(T) \cdot X_{\text{noise}}(T)$, and the time-dependent standard deviation of that process equals $S_X(T)$. If one recognizes that in practice, the observations are available not at continuous time, T , but rather at a finite number, n , of time points, $T(i)$, $i = 1, \dots, n$, and if one further considers that the response variable, Y , in the calibration period is also made with observational errors, then one arrives at the real-world (i.e., with noise) linear calibration model,

$$Y(i) = \beta_0 + \beta_1 \cdot [X(i) - S_X(i) \cdot X_{\text{noise}}(i)] + S_Y(i) \cdot Y_{\text{noise}}(i). \quad (2)$$

Here, i is a counter which runs from 1 to n , the size of the bivariate calibration time series sample $\{t(i), x(i), y(i)\}_{i=1}^n$, and the discrete notation, such as $X(i)$ for

$X(T(i))$, is a handy abbreviation. We note that the temporal spacing, which is given by $[t(i) - t(i - 1)]$, $i = 2, \dots, n$, need not be constant. Note also that this paper follows statistical convention and denotes a theoretical process with a capital letter, such as $X(i)$, and an observed value with a small letter, such as $x(i)$.

Our statistical task is estimation, namely, to determine β_0 and β_1 on the basis of $\{t(i), x(i), y(i)\}_{i=1}^n$. Since the calibration observations are influenced by errors, the parameter estimates ($\hat{\beta}_0$ and $\hat{\beta}_1$) are not exactly the same as the true values, and the “hat” reminds the researcher of that fact. We note that the noise term, $S_Y(i) \cdot Y_{\text{noise}}(i)$, also accommodates proxy errors. For example, in coral paleoclimatology, the predictor variable oxygen isotopic composition may be influenced not only by the desired response variable sea-surface temperature (SST) but also by other climate variables such as the oxygen isotopic composition of seawater, a parameter that is closely related to salinity (Brocas et al. 2019). The problem of predictor noise also plagues climate modeling. For field reconstructions, two versions of a climate variable (e.g., air temperature) are related to each other: one version is the instrumental measurement, the other the climate model output. (See, for example, Mann et al. (2007); Ammann et al. (2010) and Tingley et al. 2012, Sect. 6 therein.)

The calibration model (Eq. 2) is called an errors-in-variables regression, and because there is predictor noise (i.e., $S_X(i) \neq 0$), the estimation of the calibration parameters β_0 and β_1 is not straightforward (Mudelsee 2014). In particular, ignored predictor noise may lead to a negative bias of the absolute value of the slope estimate, β_1 . This would further mean that in the case of a paleoclimatic coral proxy record, the inferred temperature amplitudes would be too small. Section 2 presents two powerful (i.e., unbiased) procedures for estimation and prediction; one applies to the situation of homoscedastic errors (both $S_X(i)$ and $S_Y(i)$ constant), the other to the more difficult situation of heteroscedastic errors ($S_X(i)$ or $S_Y(i)$ not constant). Section 3 introduces a bootstrap resampling technique that is able to obtain uncertainty measures that are also reliable in the presence of non-Gaussian distributional shapes and autocorrelations in the noise components $X_{\text{noise}}(i)$ and $Y_{\text{noise}}(i)$ —features that are rather the norm than the exception in climate sciences (Mudelsee 2014). Section 4 illustrates the mathematical methods for estimation, prediction, and uncertainty determination on paleoclimatic data from the coral archive. The conclusions (Sect. 5) are directed to the practitioner who has to find the optimal calibration line in the presence of predictor noise. The Appendix gives a brief description of the linear calibration software, LINCAL, which is used to implement the concepts described in the present paper.

2 Estimation and Prediction

An estimation is a mathematical procedure, a recipe, to determine the parameters, β_0 and β_1 , of the errors-in-variables calibration model (Eq. 2) on the basis of a time series sample, $\{t(i), x(i), y(i)\}_{i=1}^n$. A prediction is the estimation for a new predictor value, $x(n + 1)$, of the response value, $\hat{y}(n + 1)$, which is achieved by means of the parameter estimates, $\hat{\beta}_0$ and $\hat{\beta}_1$, via the formula

$$\hat{y}(n + 1) = \hat{\beta}_0 + \hat{\beta}_1 \cdot x(n + 1). \quad (3)$$

Typically, one uses many $x(n+1)$ values over a large range to construct a calibration curve.

For the homoscedastic error type, this paper investigates the ordinary least-squares (OLS) estimation procedure, which is enhanced by a bias correction (Sect. 2.1). For the heteroscedastic error type, it employs a weighted estimation procedure, which allows more accurate data points (i.e., with smaller standard deviation) to contribute more strongly to the estimation than less accurate data (Sect. 2.2). A crucial point is that—irrespective of whether there is homo- or heteroscedasticity—some prior knowledge about the size of the standard deviation(s), $S_X(i)$ and $S_Y(i)$, has to be available. The temporal aspect is relevant only for uncertainty determination (Sect. 3), not for estimation. This means that the estimation procedures presented herein are also applicable to experimental situations where no time values, $t(i)$, are available. These four cases—homoscedastic/heteroscedastic errors and availability/unavailability of time values—are also treated separately in LINCAL software (Appendix). We briefly mention two other procedures (Sect. 2.3) which are occasionally found in practical applications. However, both have exhibited poor performance as regards the major target set in the present paper, namely, the delivery of an unbiased calibration slope.

2.1 Ordinary Least-Squares Estimation with Bias Correction

Assume homoscedasticity, that is, $S_X(i) = S_X$ is a constant. The simple OLS estimation minimizes the unweighted sum of squares,

$$SSQ(\beta_0, \beta_1) = \sum_{i=1}^n [y(i) - \beta_0 - \beta_1 \cdot x(i)]^2. \quad (4)$$

This yields the estimators

$$\begin{aligned} \hat{\beta}_{1, \text{OLS}} = & \left\{ \left[\sum_{i=1}^n x(i) \right] \cdot \left[\sum_{i=1}^n y(i) \right] / n - \sum_{i=1}^n x(i) \cdot y(i) \right\} \\ & \times \left\{ \left[\sum_{i=1}^n x(i) \right]^2 / n - \sum_{i=1}^n x(i)^2 \right\}^{-1} \end{aligned} \quad (5)$$

and

$$\hat{\beta}_{0, \text{OLS}} = \left[\sum_{i=1}^n y(i) - \hat{\beta}_{1, \text{OLS}} \sum_{i=1}^n x(i) \right] / n. \quad (6)$$

However, if the noise components in the calibration model, $X_{\text{noise}}(i)$ and $Y_{\text{noise}}(i)$, are statistically independent—which is typically the case since the observations usually stem from two independent sources—and also of zero mean and Gaussian shape, then standard textbooks (Draper and Smith 1981, Sect. 2.14 therein) explain that the slope

estimator (Eq. 5) is biased downwards as $E(\hat{\beta}_1) = \kappa \cdot \beta_1$, where E is the expectation operator and $\kappa \leq 1$ is the attenuation factor. The latter is given by

$$\kappa = \left(1 + S_X^2 / \text{VAR}[X_{\text{true}}(i)]\right)^{-1}, \quad (7)$$

where VAR is the variance operator, and $\{X_{\text{true}}(i)\}_{i=1}^n$ are the true (but observed only with error) predictor points. From error propagation (Mudelsee 2014) it follows that $\text{VAR}[X(i)] = \text{VAR}[X_{\text{true}}(i)] + S_X^2$. This, finally, leads to the unbiased slope estimator and intercept estimator,

$$\hat{\beta}_{1,\text{OLSBC}} = \hat{\beta}_{1,\text{OLS}} / \left\{1 - S_X^2 / \text{VAR}[X(i)]\right\}, \quad (8)$$

and

$$\hat{\beta}_{0,\text{OLSBC}} = \left[\sum_{i=1}^n y(i) - \hat{\beta}_{1,\text{OLSBC}} \sum_{i=1}^n x(i) \right] / n, \quad (9)$$

respectively. The acronym OLSBC stands for OLS with bias correction. It is evident that bias correction for the slope (Eq. 8) requires prior knowledge about S_X .

2.2 Weighted Least-Squares for Both Variables Estimation

The linear calibration model (Eq. 2) shows that the two noise components can be combined as $S_Y(i) \cdot Y_{\text{noise}}(i) - \beta_1 \cdot S_X(i) \cdot X_{\text{noise}}(i)$. This combination hints at the estimation approach by attaching weights to the observations of both variables (Deming 1943; Lindley 1947). Over the years, the variant suggested by York (1966) and others became the standard, namely the minimization of

$$SSQWXY(\beta_0, \beta_1) = \sum_{i=1}^n \frac{[y(i) - \beta_0 - \beta_1 \cdot x(i)]^2}{S_Y(i)^2 + \beta_1^2 \cdot S_X(i)^2}. \quad (10)$$

We abbreviate this estimation procedure as WLSXY. Also, this estimation type requires prior knowledge, namely about $S_Y(i)$ and $S_X(i)$.

However, WLSXY minimization of $SSQWXY(\beta_0, \beta_1)$ is numerically difficult because the slope, β_1 , appears in the denominator of the least-squares sum. The routine Fitexy (Press et al. 1992) parameterizes the slope as $\beta'_1 = \tan^{-1}(\beta_1)$, scales the $y(i)$ values, and uses Brent's search with a starting value for the slope from an initial OLS estimation. Other relevant papers were published by Reed (1989, 1992) and Squire (1990). We follow those authors in the use of WLSXY for estimation, but not for parameter error determination—for that purpose, bootstrap resampling (Sect. 3) is employed instead.

2.3 Other Estimation Procedures

OLS regression (Sect. 2.1) can be geometrically described as a procedure for finding the best-fit regression line via the minimization of the sum of the squares of the vertical distances (i.e., in the Y -direction) between data points and the regression line. OLS estimation can also be applied in the other direction (i.e., of X on Y), which would utilize the horizontal distances (i.e., in the X -direction) and yield a different estimation, as standard textbooks (Davis 1986) explain. Reduced major axis (RMA) regression can be seen as a compromise between the OLS variants, since it involves the minimization of the sum of the products of the vertical and horizontal distances (Davis 1986). The RMA estimators, finally, are

$$\hat{\beta}_{1,\text{RMA}} = (SS_Y / SS_X)^{1/2}, \quad (11)$$

$$\hat{\beta}_{0,\text{RMA}} = \bar{Y} - \hat{\beta}_{1,\text{RMA}} \cdot \bar{X}, \quad (12)$$

where

$$SS_Y = \sum_{i=1}^n y(i)^2 - \left[\sum_{i=1}^n y(i) \right]^2 / n, \quad (13)$$

$$SS_X = \sum_{i=1}^n x(i)^2 - \left[\sum_{i=1}^n x(i) \right]^2 / n, \quad (14)$$

$$\bar{Y} = \sum_{i=1}^n y(i) / n, \quad (15)$$

and

$$\bar{X} = \sum_{i=1}^n x(i) / n. \quad (16)$$

Another estimation approach, also geometrically straightforward to understand, is the Wald–Bartlett (WB) procedure (Mudelsee 2014). The idea is to divide the set of data points into three groups of the same size according to the size of the $x(i)$ values. The center for the first group (smallest $x(i)$ values) is given via the respective $x(i)$ and $y(i)$ means for the group, and analogously for the third group (largest $x(i)$ values). The line that connects both centers then defines the WB slope estimate. The problem with the RMA and WB procedures is that they deliver biased estimations for the calibration slope, β_1 . This was previously shown in the case of WB (Mudelsee 2014) and will be seen in the case of RMA (Sect. 3.2).

3 Uncertainty Determination

In statistical language, the estimated values for the intercept and slope are called point estimates. In the real world, with limited size of noisy data, the point estimates deviate from the true values. However, one can measure the typical sizes of such deviations. Statistical methodology has therefore developed procedures to determine interval estimates, or confidence intervals (CIs) (Robinson 1982), as a type of uncertainty measure that accompanies an estimate. In fact, a serious interpretation of the significance of an estimate is difficult to make without an uncertainty measure.

Classical CIs are those that can be derived with pencil and paper. The construction of classical CIs for the linear errors-in-variables model (Eq. 2) in earlier work (York 1966; Fuller 1987) made a number of assumptions from the following:

1. Gaussian distributional shapes of the noise components, $X_{\text{noise}}(i)$ and $Y_{\text{noise}}(i)$;
2. absence of autocorrelation in the noise components;
3. absence of correlation between $X(i)$ and $X_{\text{noise}}(i)$ and between $Y(i)$ and $Y_{\text{noise}}(i)$; and
4. absence of correlation between $X_{\text{noise}}(i)$ and $Y_{\text{noise}}(i)$.

Some authors (York 1969; Freedman 1984; Freedman and Peters 1984; Carroll et al. 2006) treat non-Gaussian errors (point 1) and the correlation effects (points 3 and 4). However, allowance for autocorrelations (point 2) seems to have been made by none, and this is a crucial issue since climate fluctuations typically exhibit autocorrelation (memory) besides non-Gaussian shapes.

In the present paper, the interest is in linearly relating two processes, $X(i)$ and $Y(i)$, and the sample, $\{t(i), x(i), y(i)\}_{i=1}^n$, may contain the time. Analyses of climate data by numerous authors document that non-Gaussian distributions and autocorrelation phenomena are typical of climate processes (see, e.g., Trenberth (1984); Perron and Sura (2013) and Mudelsee (2014)). We cannot expect the classical method to yield accurate results (CIs) for climate data. Therefore, this paper presents a computing-intensive CI construction method that is based on bootstrap resampling (Efron 1979; Mudelsee 2014). The computational steps are detailed at the algorithmic level (Sect. 3.1).

3.1 Bootstrap Algorithm

The algorithm for CI construction for the estimated errors-in-variables regression parameters is shown below. Explanations and illustrations for various steps are also given (Sect. 3.1.1).

3.1.1 Remarks

In step 1, if time values, $t(i)$, are unavailable, then in step 5, set $l = 1$. In step 3, see Fig. 1 for the definition of the residuals.

The overall goal of step 5 is to find a block length that is suitable to preserve the autocorrelation properties of the data-generating process (Künsch 1989). The ordinary bootstrap (Efron 1979), which generates resamples with preserved distributional shape

1.	Bivariate time series	$\{t(i), x(i), y(i)\}_{i=1}^n$
2.	Parameter estimates from OLSBC or WLSXY	$\hat{\beta}_0, \hat{\beta}_1$
3.	Residuals	$e_X(i), e_Y(i)$
4.	Fit values	$x_{\text{fit}}(i) = x(i) - e_X(i),$ $y_{\text{fit}}(i) = y(i) - e_Y(i)$
5.	Bias-corrected AR(1) parameters estimated on residuals, block length selection	\hat{a}'_X, \hat{a}'_Y l
6.	Resampled residuals, pairwise-MBB with l	$\{e_X^{*b}(i), e_Y^{*b}(i)\}_{i=1}^n$ (b , counter)
7.	Resample	$x^{*b}(i) = x_{\text{fit}}(i) + e_X^{*b}(i),$ $y^{*b}(i) = y_{\text{fit}}(i) + e_Y^{*b}(i), i = 1, \dots, n$
8.	Replication, parameters	$\hat{\beta}_0^{*b}, \hat{\beta}_1^{*b}$
9.	Replication, prediction	$\hat{y}^{*b}(n+1) = \hat{\beta}_0^{*b} + \hat{\beta}_1^{*b} \cdot [x(n+1) + S_X \cdot \mathcal{E}_{N(0, 1)}(n+1)]$
10.	Go to step 6 until $b = B = 2,000$ replications exist each	$\{\hat{\beta}_0^{*b}\}_{b=1}^B, \{\hat{\beta}_1^{*b}\}_{b=1}^B, \{\hat{y}^{*b}(n+1)\}_{b=1}^B$
11.	Calculate CI on the basis of replications	

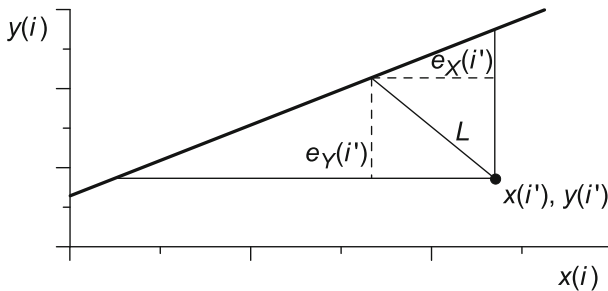


Fig. 1 CI construction, definition of residuals. The illustration is for a certain data point, $(x(i'), y(i'))$ (filled circle), from which the distance to the linear fit (thick, tilted line) is measured via line L ; the slope of L is equal to $-\lambda/\hat{\beta}_1$, where $\lambda = [S_Y(i')/S_X(i')]^2$ (York 1967). The residuals (dashed lines) are given by $e_X(i') = [\hat{\beta}_0 + \hat{\beta}_1 \cdot x(i') - y(i')]/[\lambda/\hat{\beta}_1 + \hat{\beta}_1]$ and $e_Y(i') = -\lambda \cdot e_X(i')/\hat{\beta}_1$; they can have positive or negative values. Modified after (Mudelsee 2014, Fig. 8.5 therein)

(because one randomly draws from the data with replacement), can then be augmented to preserve autocorrelation (over the length of the randomly drawn time blocks) as well. This renders the moving-block bootstrap (MBB) resampling as suited for climate data (Mudelsee 2014). The AR(1) model is also applicable to unevenly spaced time series if it is formulated by means of a parameter called persistence time (here called τ_X

or τ_Y); see Mudelsee (2002) for a description of the model and a numerical persistence time estimation procedure that includes bias correction and CI construction. With the help of the average temporal spacing, $\bar{d} = [t(n) - t(1)]/(n - 1)$, the bias-corrected persistence-time estimates, $\hat{\tau}'_X$ and $\hat{\tau}'_Y$, can then be converted into another pair of parameters called equivalent autocorrelation coefficients, $\hat{a}'_X = \exp(-\bar{d}/\hat{\tau}'_X)$ and $\hat{a}'_Y = \exp(-\bar{d}/\hat{\tau}'_Y)$. The idea of the equivalent autocorrelation coefficient is to have for the case of uneven spacing a persistence parameter that corresponds to the usual autocorrelation coefficient in the case of even spacing. Owing to the construction of the residuals (Fig. 1), it can be shown (Mudelsee 2014) that for the algorithm, both X and Y yield the same value, $\hat{a}'_X = \hat{a}'_Y$. This value is plugged into the block length selector by Carlstein (1986) and Sherman et al. (1998),

$$l = NINT \left\{ \left[6^{1/2} \cdot \hat{a}'_X / \left(1 - \hat{a}'_X{}^2 \right) \right]^{2/3} \cdot n^{1/3} \right\}, \quad (17)$$

where $NINT(\cdot)$ is the nearest integer function. We finally note that there exist other block length selectors that are also based on estimated autocorrelation properties of residuals (Mudelsee 2014), but for CI accuracy (i.e., how close to the nominal confidence level a CI performs in a Monte Carlo simulation experiment), this question is of limited relevance. In other words, the blocks just have to be long enough to preserve autocorrelation.

Step 6 is bootstrap resampling: draw with replacement random blocks of length l from the residuals (step 3). The blocks are allowed to overlap. The resample is filled from the left with the block elements until the sample size, n , has been reached. If necessary, some elements from the right of the last block are discarded. For example, let $n = 8$ and $l = 3$; a possible sequence would be $\{e_X^{*b}(i)\}_{i=1}^8 = \{e_X(3), e_X(4), e_X(5), e_X(1), e_X(2), e_X(3), e_X(6), e_X(7)\}$. This procedure is performed in a pairwise manner (Mudelsee 2014), which means that the random indices are taken for $e_Y^{*b}(i)$; in the example, $\{e_Y^{*b}(i)\}_{i=1}^8 = \{e_Y(3), e_Y(4), e_Y(5), e_Y(1), e_Y(2), e_Y(3), e_Y(6), e_Y(7)\}$. Since the adaptation of the MBB resampling is done for the residuals in a pairwise manner, this adapted bootstrap procedure is called pairwise-MBBres.

Step 9 shows that the usual step of calculating the replications (i.e., copies of the estimates) can also be done for prediction. The important new calculation step, which means one that goes beyond the book by Mudelsee (2014), is the addition of the random term $S_X \cdot \mathcal{E}_{N(0, 1)}(n + 1)$, where $\mathcal{E}_{N(0, 1)}(\cdot)$ denotes an independent standard Gaussian variable (mean zero, standard deviation unity). This addition serves to include the predictor uncertainty since $X(n + 1)$ is not exactly known but measured with error (standard deviation S_X). The Gaussian assumption should be well applicable to most calibration problems where data are obtained from measurement devices. This predictor uncertainty is combined with the parameter estimation uncertainties in order to yield a realistic quantification of the prediction uncertainty. Note further that the formulation of step 9 corresponds to the homoscedastic error type. In the case of heteroscedastic errors, one would use $S_X(n + 1)$ instead of S_X and make some assumptions. In the provided LINCAL software, for example, the user can select the

value of $S_X(n + 1)$. Evidently, if prior knowledge indicates other values, then the LINCAL source code should be adapted accordingly and recompiled.

The number of replications in step 10 is set as $B = 2,000$. Monte Carlo simulation experiments (Mudelsee 2014) show that this value is clearly sufficient for an accurate determination of the CI bounds. For CI construction (step 11), there exist two basic approaches (Efron and Tibshirani 1993). One is via the percentiles of the B replication values. The other, used in the present paper, is via the standard deviation over the replications, which is called bootstrap standard error. For example, in the case of the slope, it is given by

$$\widehat{\text{se}}_{\hat{\beta}_1^*} = \left\{ \sum_{b=1}^B \left[\hat{\beta}_1^{*b} - \langle \hat{\beta}_1^{*b} \rangle \right]^2 / (B - 1) \right\}^{1/2}, \quad (18)$$

where $\langle \hat{\beta}_1^{*b} \rangle = \sum_{b=1}^B \hat{\beta}_1^{*b} / B$. The bootstrap Student's t confidence interval then is (shown for the case of the slope)

$$\text{CI}_{\hat{\beta}_1, 1-2\alpha} = \left[\hat{\beta}_1 + t_\nu(\alpha) \cdot \widehat{\text{se}}_{\hat{\beta}_1^*}; \hat{\theta} - t_\nu(\alpha) \cdot \widehat{\text{se}}_{\hat{\beta}_1^*} \right], \quad (19)$$

where $t_\nu(\alpha)$ is the percentage point of the Student's t distribution function with $\nu = n - 2$ degrees of freedom, and $1 - 2\alpha = 95\%$ is the confidence level. The ν value is justified by the fact that the calibration problem comprises two parameters to be estimated. The confidence level employed is (besides values of 90% and 99%) a common choice in climate sciences (Mudelsee 2014). The LINCAL source code details the numerical calculation of $t_\nu(\alpha)$. An interesting alternative to uncertainty determination via the pairwise-MBBres procedure, pointed out by one of the reviewers of the present paper, could be via the effective sample size (Mudelsee 2014; Hu et al. 2017).

3.2 Monte Carlo Simulation Experiment

A Monte Carlo experiment is a computer simulation of a random process (Fishman 1996). The properties of the data-generating process, such as mean and standard deviation, can be prescribed. Hence, unlike in the real world, the truth is known. The random component is brought in by means of a random number generator (Mudelsee 2020). Since the prescribed parameter values are known, Monte Carlo simulation experiments can be used as objective tests of the estimation algorithms of the parameters and of the CI construction methods. The present paper studies the linear calibration model (Eq. 2) with parameters β_0 and β_1 . Emphasis is put on unbiasedness of the parameter estimators (Sect. 2) and accuracy of the bootstrap CIs (Sect. 3.1) under realistic conditions. Six different simulation experiments are carried out, starting with an easy setting and progressing to more challenging conditions.

The easy simulation setting with un-autocorrelated and homoscedastic Gaussian noise processes (Table 1) allows unbiased (i.e., within simulation noise) estimations, with both OLSBC and WLSXY, and for both intercept β_0 and slope β_1 . Additionally,

the bootstrap CIs show excellent coverage performance, which means that the nominal level of 95% is nearly achieved for sample sizes (n) as small as 50.

The so-called naive OLS method fails already for the easy simulation setting (Table 1) in the case of calibration slope estimation, which means that the empirical coverage is far off the nominal value (results not shown). This is due to the negative bias of OLS slope estimation in the presence of predictor noise (Sect. 2.1).

Prescription of a slightly more challenging setting, where $Y_{\text{noise}}(i)$ instead has a skewed, lognormal shape (Table 2), has within simulation noise no detectable performance reduction in terms of bias and coverage. This holds for OLSBC and WLSXY estimation, and for both intercept and slope.

However, the introduction of a heteroscedastic noise component (Table 3) demonstrates that a misapplication of the OLSBC estimation method brings a dramatic reduction in coverage accuracy for the slope. Also, the bias for $\hat{\beta}_1$ by means of OLSBC increases (compared to the previous two easy settings) by a factor of between roughly 2 and 100 (dependent on n). On the other hand, use of the indicated (because of the heteroscedasticity) WLSXY estimation yields acceptably accurate results. In that regard, empirical accuracy of 92–94% or 98–96% instead of nominally 95% should be acceptable for applied sciences.

Even further increased challenges for the method, the introduction of a lognormal shape (Table 4) and, additionally, of autocorrelation (Table 5) does not change the overall assessment: WLSXY performs acceptably well, while OLSBC fails because of the heteroscedasticity.

The summary assessment of the performance of the estimation methods for the errors-in-variables regression model (Eq. 2) is as follows. For homoscedastic noise processes, both OLSBC and WLSXY perform well, that is, they deliver (within simulation noise) unbiased calibration slopes and bootstrap CIs with an accuracy that is acceptable for applied sciences. The bootstrap resampling (pairwise-MBBres) allows one to take into account both non-Gaussian shapes and autocorrelation of the random components. On the other hand, RMA estimation yields, even in the easy situation with un-autocorrelated and homoscedastic Gaussian noise processes (Table 6), unacceptably inaccurate calibration slopes: positive bias (i.e., overestimation) and undercoverage of CIs.

Although the present paper attaches priority to an unbiased calibration slope estimation, the standard error, $\text{se}_{\hat{\beta}_1}$, should also be considered since this measures the spread (standard deviation) of the estimate. Both uncertainty measures can be readily combined to form the root mean squared error, $\text{RMSE}_{\hat{\beta}_1}$. Following Mudelsee (2014),

$$\text{RMSE}_{\hat{\beta}_1} = \left(\text{se}_{\hat{\beta}_1}^2 + \text{Bias}_{\hat{\beta}_1}^2 \right)^{1/2}. \quad (20)$$

For the easy Monte Carlo setting (un-autocorrelated and homoscedastic Gaussian noise processes) and large sample sizes (say, $n \geq 500$), the WLSXY estimation method also outperforms OLSBC and RMA in terms of $\text{RMSE}_{\hat{\beta}_1}$ (Table 7). For medium-sized samples ($50 \leq n < 500$), however, RMA performs no worse (or even better) than WLSXY or OLSBC. And for small sample sizes ($n \leq 50$), there are no positive effects of WLSXY or OLSBC in terms of $\text{RMSE}_{\hat{\beta}_1}$, not even when compared with OLS (Table

Table 1 Monte Carlo simulation experiment, linear errors-in-variables regression (Eq. 2) with un-autocorrelated and homoscedastic Gaussian noise processes

n	OLSBC		WLSXY			
	$\gamma_{\hat{\beta}_0}$	Bias $\hat{\beta}_0$	$\gamma_{\hat{\beta}_1}$	Bias $\hat{\beta}_1$	$\gamma_{\hat{\beta}_0}$	Bias $\hat{\beta}_1$
10	0.925	-0.001271	0.930	+0.036640	0.917	-0.001098
20	0.947	+0.000960	0.944	+0.013832	0.944	+0.000975
50	0.951	+0.000085	0.948	+0.005227	0.950	+0.000103
100	0.951	+0.000062	0.949	+0.002707	0.951	+0.000066
200	0.955	-0.000042	0.948	+0.001485	0.954	-0.000040
500	0.951	+0.000151	0.949	+0.000447	0.950	+0.000151
1,000	0.950	+0.000128	0.950	+0.000176	0.950	+0.000128

$n_{\text{sim}} = 47,500$ random samples were generated from $X_{\text{true}}(i) = \mathcal{E}_{N(0, 1)}(i)$ and $Y(i) = \beta_0 + \beta_1 \cdot X_{\text{true}}(i) + S_Y \cdot Y_{\text{noise}}(i)$, $i = 1, \dots, n$, with $\beta_0 = 1.0$ and $\beta_1 = 2.0$. Predictor noise was subsequently added: $X(i) = X_{\text{true}}(i) + S_X \cdot X_{\text{noise}}(i)$. The $X_{\text{noise}}(i)$ and $Y_{\text{noise}}(i)$ are mutually independent Gaussian processes without autocorrelations. Construction of bootstrap CIs used pairwise-MBBres resampling (Sect. 3.1), $B = 2,000$, and a confidence level of 95%. Prior knowledge of $S_X = 0.2$, $S_Y = 0.4$ was utilized in the estimations. The entries (rounded) show empirical coverage, γ (accurate to three digits), and empirical estimation bias dependent on sample size. The empirical coverage is given by the number of simulations, where the CI contains the true value, divided by n_{sim} . The coverage error is given by γ minus the nominal value (0.95). The empirical bias is given by the average of an estimate (e.g., $\hat{\beta}_1$) over the simulations minus the true value (e.g., β_1)

Table 2 Monte Carlo simulation experiment, linear errors-in-variables regression (Eq. 2) with un-autocorrelated and homoscedastic Gaussian/lognormal noise processes

n	OLSBC				WLSXY			
	$\gamma_{\hat{\beta}_0}$		$\gamma_{\hat{\beta}_1}$		$\gamma_{\hat{\beta}_0}$		$\gamma_{\hat{\beta}_1}$	
	Bias $\hat{\beta}_0$	Bias $\hat{\beta}_1$	Bias $\hat{\beta}_0$	Bias $\hat{\beta}_1$	Bias $\hat{\beta}_0$	Bias $\hat{\beta}_1$	Bias $\hat{\beta}_0$	Bias $\hat{\beta}_1$
10	0.923	-0.000265	0.931	+0.035848	0.914	-0.000368	0.908	+0.014392
20	0.941	-0.000914	0.944	+0.016476	0.937	-0.000829	0.928	+0.008725
50	0.946	-0.000204	0.951	+0.005006	0.946	-0.000375	0.931	+0.002987
100	0.949	+0.000022	0.956	+0.002329	0.949	+0.000021	0.927	+0.001883
200	0.948	-0.000035	0.958	+0.001319	0.948	-0.000041	0.923	+0.001369
500	0.951	+0.000233	0.963	+0.000402	0.951	+0.000240	0.918	+0.001072
1,000	0.951	+0.000046	0.965	+0.000231	0.951	+0.000047	0.916	+0.000989

Implementation details are as in Table 1, with the following exception: $Y_{\text{noise}}(i)$ has a lognormal distributional shape

Table 3 Monte Carlo simulation experiment, linear errors-in-variables regression (Eq. 2) with un-autocorrelated and heteroscedastic Gaussian noise processes

n	OLSBC			WLSXY				
	$\gamma_{\hat{\beta}_0}$	Bias $\hat{\beta}_0$	$\gamma_{\hat{\beta}_1}$	Bias $\hat{\beta}_1$	$\gamma_{\hat{\beta}_0}$	Bias $\hat{\beta}_0$	$\gamma_{\hat{\beta}_1}$	Bias $\hat{\beta}_1$
10	0.910	-0.000434	0.892	-0.022025	0.920	-0.000590	0.908	+0.004985
20	0.934	+0.000126	0.901	-0.026653	0.947	+0.000410	0.930	+0.001736
50	0.939	-0.000221	0.890	-0.028759	0.952	-0.000104	0.936	+0.000640
100	0.941	+0.000111	0.864	-0.029190	0.954	+0.000110	0.936	+0.000412
200	0.939	-0.000329	0.817	-0.029458	0.952	-0.000304	0.940	+0.000162
500	0.940	+0.000016	0.676	-0.029393	0.954	+0.000007	0.942	+0.000384
1,000	0.938	-0.000096	0.465	-0.029506	0.953	-0.000099	0.940	+0.000271

Implementation details are as in Table 1, with the following exceptions: $S_X(i) = 0.2$ for i even and $S_X(i) = 0.4$ for i uneven, $i = 1, \dots, n$ (with a prior knowledge for OLSBC that assumes constant $S_X = 0.4$)

Table 4 Monte Carlo simulation experiment, linear errors-in-variables regression (Eq. 2) with un-autocorrelated and heteroscedastic Gaussian/lognormal noise processes

n	OLSBC			WLSXY				
	$\gamma_{\hat{\beta}_0}$	Bias $\hat{\beta}_0$	$\gamma_{\hat{\beta}_1}$	Bias $\hat{\beta}_1$	$\gamma_{\hat{\beta}_0}$	Bias $\hat{\beta}_0$	$\gamma_{\hat{\beta}_1}$	Bias $\hat{\beta}_1$
10	0.902	+0.000279	0.896	-0.022434	0.915	-0.002715	0.913	+0.010469
20	0.922	+0.000590	0.910	-0.026991	0.938	-0.000054	0.933	+0.002225
50	0.927	+0.000116	0.911	-0.028298	0.941	-0.000077	0.940	+0.001175
100	0.935	+0.000139	0.903	-0.029038	0.949	+0.000117	0.948	+0.000841
200	0.934	-0.000011	0.878	-0.029187	0.948	-0.000045	0.953	+0.000535
500	0.938	-0.000178	0.796	-0.029374	0.952	-0.000168	0.957	+0.000440
1,000	0.936	-0.000020	0.667	-0.029465	0.952	-0.000027	0.962	+0.000318

Implementation details are as in Table 1, with the following exceptions: $S_X(i) = 0.2$ for i even and $S_X(i) = 0.4$ for i uneven, $i = 1, \dots, n$ (with a prior knowledge for OLSBC that assumes constant $S_X = 0.4$); and $Y_{\text{noise}}(i)$ has a lognormal distributional shape

Table 5 Monte Carlo simulation experiment, linear errors-in-variables regression (Eq. 2) with autocorrelated and heteroscedastic Gaussian/lognormal noise processes

n	OLSBC		WLSXY					
	$\gamma_{\hat{\beta}_0}$	Bias $\hat{\beta}_0$	$\gamma_{\hat{\beta}_1}$	Bias $\hat{\beta}_1$	$\gamma_{\hat{\beta}_0}$	Bias $\hat{\beta}_0$	$\gamma_{\hat{\beta}_1}$	Bias $\hat{\beta}_1$
10	0.842	-0.000435	0.887	-0.017342	0.860	-0.004460	0.903	-0.000050
20	0.882	-0.000556	0.908	-0.025074	0.901	-0.000817	0.929	+0.002641
50	0.897	-0.000139	0.911	-0.027622	0.915	-0.000023	0.941	+0.002113
100	0.906	+0.000203	0.903	-0.028559	0.922	+0.000199	0.947	+0.000966
200	0.911	+0.000070	0.878	-0.028925	0.927	+0.000062	0.953	+0.000751
500	0.922	+0.000147	0.799	-0.029501	0.938	+0.000158	0.959	+0.000297
1,000	0.925	-0.000054	0.673	-0.029490	0.940	-0.000054	0.963	+0.000278

Implementation details are as in Table 1, with the following exceptions: $X_{\text{noise}}(i)$ and $Y_{\text{noise}}(i)$ are AR(1) processes with even temporal spacing and autocorrelation parameters $\alpha_X = 0.25$ and $\alpha_Y = 0.15$; $S_X(i) = 0.2$ for i even and $S_X(i) = 0.4$ for i uneven, $i = 1, \dots, n$ (with a prior knowledge for OLSBC that assumes constant $S_X = 0.4$); and $Y_{\text{noise}}(i)$ has a lognormal distributional shape

Table 6 Monte Carlo simulation experiment, linear errors-in-variables regression (Eq. 2) with un-autocorrelated and homoscedastic Gaussian noise processes

n	RMA			
	$\gamma_{\hat{\beta}_0}$	$\text{Bias}_{\hat{\beta}_0}$	$\gamma_{\hat{\beta}_1}$	$\text{Bias}_{\hat{\beta}_1}$
10	0.914	−0.000383	0.909	+0.023350
20	0.940	+0.000237	0.930	+0.017742
50	0.944	−0.000445	0.928	+0.015891
100	0.944	+0.000332	0.925	+0.014998
200	0.945	+0.000165	0.913	+0.015126
500	0.944	−0.000061	0.881	+0.014530
1,000	0.943	+0.000050	0.822	+0.014672

Implementation details are as in Table 1, with the following exception: RMA estimation (Sect. 2.3) is employed instead of OLSBC or WLSXY

Table 7 Monte Carlo simulation experiment, linear errors-in-variables regression (Eq. 2) with un-autocorrelated and homoscedastic Gaussian noise processes

n	OLS		OLSBC	WLSXY	RMA
	$\text{Bias}_{\hat{\beta}_1}$	$\text{RMSE}_{\hat{\beta}_1}$	$\text{RMSE}_{\hat{\beta}_1}$	$\text{RMSE}_{\hat{\beta}_1}$	$\text{RMSE}_{\hat{\beta}_1}$
10	−0.048502	0.196151	0.261060	0.222625	0.193319
20	−0.049304	0.131203	0.145591	0.140027	0.122674
50	−0.049263	0.088191	0.085059	0.083019	0.075074
100	−0.049070	0.070697	0.059449	0.058395	0.053108
200	−0.048728	0.060534	0.041697	0.041004	0.038825
500	−0.048805	0.053804	0.026228	0.025860	0.026809
1,000	−0.048860	0.051367	0.018411	0.018088	0.021640

Implementation details are as in Table 1. The focus is on RMSE for slope estimation; results from the OLS method are also shown

7). It is emphasized that these assessments apply only to the easy setting; for more complex settings, the WLSXY method should be optimal. Finally, the so-called naive OLS calibration estimation method yields completely unacceptable $\text{RMSE}_{\hat{\beta}_1}$ values (Table 7). Evidently, this poor performance is due to the negative bias of the slope via the OLS estimation (Sect. 2.1).

4 Application

Among the various natural archives where information about past climate variations is stored, the coral archive is distinguished by (i) a comparably quick growth (i.e., a high accumulation rate of the carbonate material) and (ii) the presence of detectable yearly growth bands. This has allowed paleoclimatologists to infer climate variability on subseasonal timescales with a high relative temporal accuracy (i.e., with respect to an absolute fixpoint), thereby extending the instrumental record back in time by thousands and millions of years (Felis 2020). A mild deficit of the method stems from

often-reduced absolute temporal accuracy in fossil corals (i.e., the fixpoint date is uncertain) and the shortness of the records. Typical is a monthly or bimonthly spacing over a length of a few decades to centuries—corals provide accurate snapshots of subseasonal climate, possibly deeply back in time.

A frequently employed proxy variable is oxygen isotopic composition, which is measured by a mass spectrometer on the coral carbonate material. Conventionally, the isotopic values are reported in the delta notation,

$$\delta^{18}\text{O} = \left[(^{18}\text{O}/^{16}\text{O})_{\text{sample}} / (^{18}\text{O}/^{16}\text{O})_{\text{VPDB}} - 1 \right] \cdot 1,000\text{‰}, \quad (21)$$

where $(^{18}\text{O}/^{16}\text{O})$ is the number ratio of oxygen isotopes ^{18}O and ^{16}O , and VPDB stands for the Vienna Pee Dee Belemnite standard material, against which the sample is compared. Corals that grew in the surface water in tropical regions can therefore provide proxy records of SST (Felix 2020). As mentioned in Sect. 1, the real world in paleoclimatology is not perfect, and a number of uncertainties are present in coral proxy thermometry: imperfect devices and material (e.g., mass spectrometer, thermometer, or standard) and confounding climate variables (e.g., sea-surface salinity). Paired measurements, where the $\delta^{18}\text{O}$ data are augmented with information from another proxy (e.g., Sr/Ca elemental ratios) obtained on the same sample, can therefore improve proxy-derived inferences (Pfeiffer et al. 2019).

The paper by Brenner et al. (2017) presented an SST- $\delta^{18}\text{O}$ calibration obtained from modern *Isopora* corals on the Great Barrier Reef, off the northeastern coast of Australia. In total, the $\delta^{18}\text{O}$ data were measured on five corals from near Heron Island; between one and three corals supplied monthly resolved series between October 1971 and July 1976, while between one and two corals grew between September 2009 and August 2012. For these two time intervals, there are no missing data, and the sample size of the combined monthly series is $n = 94$. Brenner et al. (2017) utilized the Extended Reconstructed Sea Surface Temperature V3b (ERSST V3b) dataset from the National Oceanic and Atmospheric Administration (NOAA, Boulder, CO, USA), which is available at monthly resolution on a 2.0 degree latitude by 2.0 degree longitude global grid; Heron Island is presented by the grid cell centered at 22°S, 152°E. ERSST V3b is described in detail by Smith et al. (2008).

Since the original five $\delta^{18}\text{O}$ records from the various corals could possibly be shifted against each other due to local biological or microclimatological factors, Brenner et al. (2017) determined five centered $\delta^{18}\text{O}$ series by subtracting from each series the average $\delta^{18}\text{O}$ calculated over the shared intervals. These centered series were then averaged to construct a composite coral record. This procedure means that the target of the inference is just the calibration slope, not the intercept.

For an individual coral $\delta^{18}\text{O}$ measurement data point, Brenner et al. (2017) report accuracy of 0.11‰. We take this value for the calculation of the standard error, $s_X(i)$, of the composite coral $\delta^{18}\text{O}$ by means of error propagation. This means that if three corals contribute to an $x(i)$ value, then $s_X(i) = 0.11\text{‰}/\sqrt{3} \approx 0.064\text{‰}$; if two corals contribute, then $s_X(i) = 0.11\text{‰}/\sqrt{2} \approx 0.078\text{‰}$; and if one coral contributes, then $s_X(i) = 0.11\text{‰}$. As regards the SST standard error, Smith et al. (2008) give the SST error variances dependent on time and grid cell (available for download at <https://>

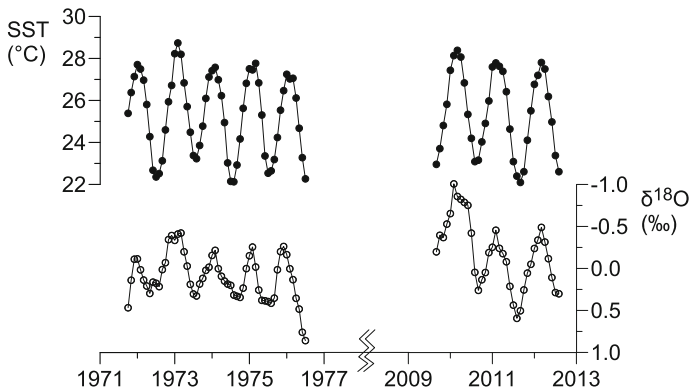


Fig. 2 Application, data. Filled symbols, SST; open symbols, coral $\delta^{18}\text{O}$. Note inverted $\delta^{18}\text{O}$ axis. See the main text (Sect. 4) for further explanations and Table 8 for numerical values

psl.noaa.gov/data/gridded/data.noaa.ersst.v3.html, last accessed November 21, 2022); and the standard error, $s_Y(i)$, is simply the square root of the variance. The time series are shown in Fig. 2, and the full dataset $\{t(i), x(i), y(i), s_X(i), s_Y(i)\}_{i=1}^{94}$ is given in Table 8. The heteroscedastic error components indicate that the estimation method to be employed for the linear calibration is WLSXY (Sect. 2.2).

The numerical values for the resulting calibration fit (Fig. 3) are as follows. The bias-corrected persistence-time estimate for SST and coral $\delta^{18}\text{O}$ is $\hat{\tau}'_Y = 1.2$ years and $\hat{\tau}'_X = 0.9$ years, respectively. This leads to a block length of $l = 9$ for pairwise-MBBres resampling (Eq. 17). The typical number of $B = 2,000$ resamplings is used for the bootstrap algorithm (Sect. 3.1). The sample size is $n = 94$. This setting should provide highly accurate Student's t CIs for the WLSXY-determined slope and also for the SST predictions, as the corresponding entries (for $n = 100$) in Table 5 indicate. Finally, the slope estimate with 95% CI is

$$\hat{\beta}_1 = -6.5 \text{ } ^\circ\text{C}/\text{‰} \left[-8.0 \text{ } ^\circ\text{C}/\text{‰}; -5.0 \text{ } ^\circ\text{C}/\text{‰} \right]. \quad (22)$$

The fitted calibration curve (Fig. 3) also serves for SST prediction from new coral $\delta^{18}\text{O}$ values, denoted as $x(n+1)$ in the statistical algorithm (Sect. 3.1). The assumed predictor uncertainty is $s_X(n+1) = 0.11\text{‰}$, which corresponds to a situation where just one coral is available for prediction of new SST values in the area of the Great Barrier Reef. For the application, a wide range of densely spaced $x(n+1)$ values are shown, between -1.5‰ and $+1.5\text{‰}$ (Fig. 3). The resulting 95% CIs for the predictions are plotted as a shaded band. The borders of the confidence band show a minimal degree of jaggedness (visible in the electronic version of the present paper by means of zooming), which stems from the fact that only $B = 2,000$ resamplings are performed (and not an infinite amount). As with any extrapolation, a sound degree of caution in the interpretation for data regions extending beyond the sampled range should be exercised.

The persistence time estimates of 1.2 years for SST and 0.9 years for coral $\delta^{18}\text{O}$ are rather long when compared with other instrumental series of similar length and

Table 8 Application, data

$t(i)$	$x(i)$	$y(i)$	$s_X(i)$	$s_Y(i)$	$t(i)$	$x(i)$	$y(i)$	$s_X(i)$	$s_Y(i)$
1971.750	+0.467	25.39	0.064	0.412	1,972.583	+0.216	22.52	0.064	0.346
1971.833	+0.140	26.38	0.064	0.387	1,972.667	+0.013	23.13	0.064	0.346
1971.917	−0.111	27.14	0.064	0.374	1,972.750	−0.069	24.60	0.064	0.346
1972.000	−0.114	27.71	0.064	0.374	1,972.833	−0.344	25.94	0.064	0.361
1972.083	+0.015	27.50	0.064	0.374	1,972.917	−0.393	26.72	0.064	0.361
1972.167	+0.137	26.97	0.064	0.374	1,973.000	−0.335	28.23	0.064	0.361
1972.250	+0.208	25.81	0.064	0.374	1,973.083	−0.410	28.74	0.064	0.346
1972.333	+0.297	24.28	0.064	0.374	1,973.167	−0.422	28.20	0.064	0.332
1972.417	+0.162	22.68	0.064	0.374	1,973.250	−0.199	26.84	0.064	0.316
1972.500	+0.176	22.36	0.064	0.346	1,973.333	−0.029	25.71	0.064	0.332
1973.417	+0.189	24.49	0.064	0.316	1,974.250	+0.092	26.23	0.064	0.387
1973.500	+0.304	23.38	0.064	0.316	1,974.333	+0.149	24.95	0.064	0.387
1973.583	+0.327	23.23	0.064	0.332	1,974.417	+0.185	23.03	0.064	0.387
1973.667	+0.184	23.86	0.064	0.332	1,974.500	+0.200	22.15	0.064	0.387
1973.750	+0.117	24.78	0.064	0.332	1,974.583	+0.316	22.13	0.064	0.374
1973.833	+0.022	26.10	0.064	0.361	1,974.667	+0.331	22.93	0.064	0.374
1973.917	−0.015	27.12	0.064	0.361	1,974.750	+0.344	24.17	0.064	0.374
1974.000	−0.156	27.42	0.064	0.374	1,974.833	+0.233	25.63	0.064	0.374
1974.083	−0.216	27.58	0.064	0.387	1,974.917	+0.001	26.82	0.064	0.374
1974.167	+0.002	26.99	0.064	0.387	1,975.000	−0.152	27.51	0.064	0.374
1975.083	−0.255	27.46	0.078	0.374	1,975.917	−0.262	26.47	0.078	0.400
1975.167	−0.018	27.77	0.078	0.374	1,976.000	−0.165	27.25	0.078	0.400
1975.250	+0.256	26.84	0.078	0.374	1,976.083	+0.003	27.03	0.078	0.400
1975.333	+0.379	25.31	0.078	0.374	1,976.167	+0.133	27.06	0.078	0.387
1975.417	+0.384	23.36	0.078	0.387	1,976.250	+0.355	26.12	0.078	0.387
1975.500	+0.393	22.54	0.078	0.387	1,976.333	+0.483	24.68	0.078	0.400
1975.583	+0.413	22.65	0.078	0.387	1,976.417	+0.759	23.28	0.078	0.400
1975.667	+0.352	23.19	0.078	0.387	1,976.500	+0.858	22.27	0.078	0.412
1975.750	+0.015	24.24	0.078	0.387	2,009.667	−0.198	22.96	0.110	0.510
1975.833	−0.201	25.54	0.078	0.387	2,009.750	−0.396	23.71	0.110	0.510
2009.833	−0.369	24.81	0.110	0.500	2,010.667	+0.260	23.16	0.078	0.520
2009.917	−0.529	25.82	0.110	0.500	2,010.750	+0.134	24.03	0.078	0.500
2010.000	−0.654	27.44	0.110	0.510	2,010.833	+0.048	24.91	0.078	0.500
2010.083	−1.007	28.14	0.110	0.520	2,010.917	−0.189	25.98	0.078	0.500
2010.167	−0.858	28.39	0.110	0.520	2,011.000	−0.253	27.60	0.078	0.500
2010.250	−0.823	28.07	0.110	0.510	2,011.083	−0.456	27.79	0.078	0.510
2010.333	−0.788	26.84	0.110	0.510	2,011.167	−0.240	27.63	0.078	0.510
2010.417	−0.752	25.34	0.110	0.520	2,011.250	−0.177	27.39	0.078	0.510
2010.500	−0.421	24.20	0.110	0.510	2,011.333	−0.080	26.42	0.078	0.520
2010.583	+0.046	23.09	0.110	0.520	2,011.417	+0.212	24.64	0.078	0.510

Table 8 continued

$t(i)$	$x(i)$	$y(i)$	$s_X(i)$	$s_Y(i)$	$t(i)$	$x(i)$	$y(i)$	$s_X(i)$	$s_Y(i)$
2011.500	+0.436	23.09	0.078	0.500	2,012.333	−0.117	26.19	0.078	0.480
2011.583	+0.593	22.40	0.078	0.500	2,012.417	+0.108	24.98	0.078	0.480
2011.667	+0.503	22.10	0.078	0.500	2,012.500	+0.286	23.38	0.110	0.480
2011.750	+0.255	22.61	0.078	0.490	2,012.583	+0.301	22.61	0.110	0.480
2011.833	+0.055	24.11	0.078	0.490					
2011.917	−0.050	25.51	0.078	0.490					
2012.000	−0.239	26.77	0.078	0.490					
2012.083	−0.338	27.20	0.078	0.480					
2012.167	−0.490	27.81	0.078	0.480					
2012.250	−0.315	27.50	0.078	0.480					

Time, $t(i)$, is given as year AD; predictor with standard error, $x(i) \pm s_X(i)$, is coral $\delta^{18}\text{O}$ in ‰ units; response with standard error, $y(i) \pm s_Y(i)$, is SST in °C units

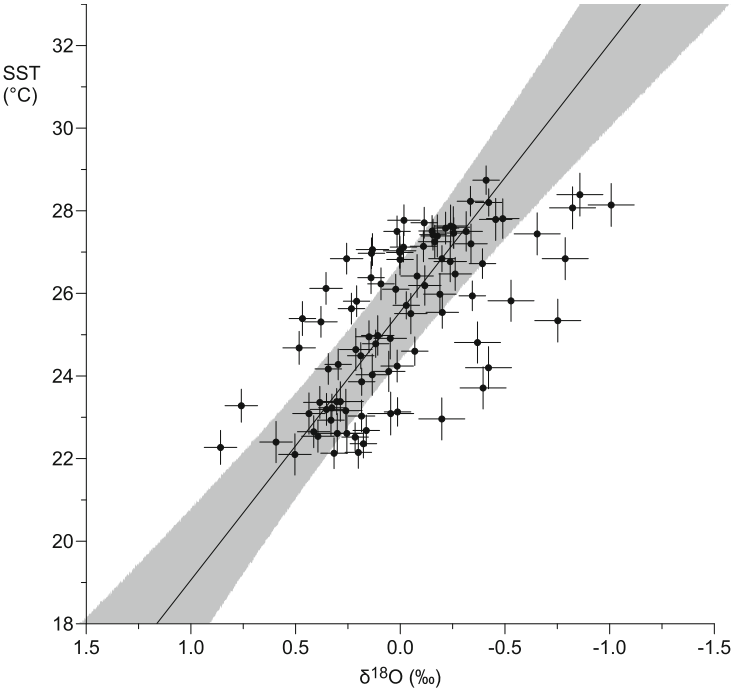


Fig. 3 Application, result. The data points (SST, coral $\delta^{18}\text{O}$) are shown as filled symbols, the standard errors as vertical and horizontal bars, the WLSXY-determined calibration curve as a solid line with a 95% Student's t CI band (shaded) for the prediction. Note the inverted $\delta^{18}\text{O}$ axis

resolution (Mudelsee 2014). Evidently, these positive autocorrelations reflect the seasonal cycles, which are clearly expressed in the time series data (Fig. 2). Although in principle a harmonic seasonal cycle model for the data could be formulated instead of the AR(1) (Sect. 3.1.1), such a model mis-specification should have negligible effects on the CI bounds owing to the fact that, in practice, the block length (determined as $l = 9$ in this application) must only be large enough to preserve the autocorrelations. However, the autocorrelations have to be taken into account for uncertainty determination. This is illustrated by means of a numerical experiment, where $l = 1$ (i.e., no preserved autocorrelation) is prescribed, which would lead to a clearly too narrow CI for the slope (as compared with the true CI given in Eq. 22), namely

$$\hat{\beta}_{1, l=1} = -6.5 \text{ } ^\circ\text{C}/\text{‰} \left[-7.3 \text{ } ^\circ\text{C}/\text{‰}; -5.7 \text{ } ^\circ\text{C}/\text{‰} \right]. \quad (23)$$

Still, however, it is wise to keep the seasonal data for the calibration (and to not perform an annual downsampling), since this means that (i) the sample size is kept and not reduced, and (ii) the calibration range of the coral $\delta^{18}\text{O}$ values is kept wide.

The calibration slope—SST change per coral $\delta^{18}\text{O}$ change—is estimated as $\hat{\beta}_1 = -6.5 \text{ } ^\circ\text{C}/\text{‰}$. The negative sign means that an increase in coral $\delta^{18}\text{O}$ indicates a cooling of SST. This estimate is obtained for the correct regression model (Eq. 2), namely, of the response variable Y (SST) on the predictor variable X (coral $\delta^{18}\text{O}$). The regression direction is correct because the task of the proxy, X , is to predict the response, SST. It may be instructive to quantify the effects of the selection of the incorrect regression direction (of X on Y). This would lead to a slope estimate of $-0.15\text{‰}/^\circ\text{C}$, which, when inverted, corresponds to $\hat{\beta}_{1, \text{incorrect}} = -6.7 \text{ } ^\circ\text{C}/\text{‰}$. This incorrect estimate is clearly within the CI for the correct estimate (Eq. 22). However, one cannot conclude that the selection of the incorrect regression direction has always negligible effects—this depends on the accuracy of the model, that is, the sizes of standard deviations and autocorrelations of the noise components in Eq. (2). Bearing this caveat about the incorrect regression direction in mind, one can compare the slope estimate of $\hat{\beta}_1 = -6.5 \text{ } ^\circ\text{C}/\text{‰}$ with 95% CI $[-8.0 \text{ } ^\circ\text{C}/\text{‰}; -5.0 \text{ } ^\circ\text{C}/\text{‰}]$ with values from previous studies. A remark on proxy system or forward modeling may be appropriate. The criticism expressed here on the incorrect regression direction is from a pure statistical viewpoint. It has nothing to do with the emerging field of proxy system modeling, which has led over the past few years to high-resolution insights into past climates. Proxy system modelers have the vision of full mathematical/physical descriptions of how the environment (e.g., climate) influences a sensor (e.g., $\delta^{18}\text{O}$) in an archive (e.g., a coral), which is then sampled and used for measurements; see, for example, the review by Evans et al. (2013).

The direct comparison partner with our result is the estimate from the paper by Brenner et al. (2017), who used the same data (kindly sent to the present author), monthly $\delta^{18}\text{O}$ measured on *Isopora* corals from Heron Island and SST from ERSST V3b. Brenner et al. (2017) gave preferences to the estimate $\hat{\beta}_{1, \text{incorrect}} = -5.4 \text{ } ^\circ\text{C}/\text{‰}$, which was determined via RMA regression. The other value obtained by these authors on the same data is $\hat{\beta}_{1, \text{incorrect}} = -8.1 \text{ } ^\circ\text{C}/\text{‰}$, which stems from generalized least-squares regression, which is a method similar to OLS (Mudelsee 2014). The present paper is left to note that for statistical/methodological reasons, the preference should

be on neither value but rather on Eq. (22) and $\hat{\beta}_1 = -6.5 \text{ } ^\circ\text{C}/\text{‰}$; however, the two values presented by Brenner et al. (2017) cover roughly the range of the 95% CI given in Eq. (22).

The same data as used here or by Brenner et al. (2017) were studied in a thesis by Lemley (2012); however, he used only a strongly reduced subset, as Fig. 12 of the thesis reveals. The statistical estimation technology was not clearly communicated, and the resulting calibration slope of $-6.7 \text{ } ^\circ\text{C}/\text{‰}$, which is in excellent agreement with the estimate from the present paper (Eq. 22), could well be a mere coincidence.

Also, Felis et al. (2014) presented a calibration of $\delta^{18}\text{O}$ measured on *Isopora*, including corals from Heron Island, but deviated as follows: they (i) used bulk coral values for reef settings with different average temperatures (so-called core-top calibration) along a latitudinal transect (from Papua New Guinea to Heron Island) instead of monthly values from records; (ii) employed another SST data product; and (iii) applied OLS and RMA regression techniques. The estimates of $\hat{\beta}_{1, \text{incorrect}} = -5.1 \text{ } ^\circ\text{C}/\text{‰}$ (for OLS) and $-4.6 \text{ } ^\circ\text{C}/\text{‰}$ (for RMA) obtained by Felis et al. (2014) therefore deviate from the value of $\hat{\beta}_1 = -6.5 \text{ } ^\circ\text{C}/\text{‰}$ from Eq. (22).

Finally, Nishida et al. (2014) established a calibration on data from cultivated *Isopora* corals in laboratory experiments where the water temperature could be advantageously controlled; the range employed in that study (between 21.1 and 29.5 $^\circ\text{C}$) is comparable to what is shown in Fig. 3. Their result ($\hat{\beta}_{1, \text{incorrect}} = -6.7 \text{ } ^\circ\text{C}/\text{‰}$) is apparently in excellent agreement with the result from the present paper (Eq. 22). To assess this, we consulted the JMP statistical software used by Nishida et al. (2014), which seems to perform OLS in an analysis-of-variance framework. Unfortunately, Nishida et al. (2014) presented only significance tests on the slope (but no CIs), and therefore we conclude that the excellent agreement may to some extent be spurious.

It is tempting to speculate about the general applicability of the obtained proxy calibration slope (Eq. 22) of $\hat{\beta}_1 = -6.5 \text{ } ^\circ\text{C}/\text{‰}$. Regions other than Heron Island? Unfortunately, published results on *Isopora* corals from elsewhere are sparse. Other coral species? (Brenner et al. 2017, Table 3 therein) give many estimates from *Porites* corals and other regions, and these values seem in rough agreement with the value from Eq. (22). This value is also, within uncertainty bounds, in agreement with the slope estimate ($-6.1 \text{ } ^\circ\text{C}/\text{‰}$) of a *Porites* calibration work from the Red Sea using the correct regression direction and OLS (Felis et al. 2000). Additionally, an often-cited calibration study on *Diploria* corals from the Atlantic Ocean reported similar values (Hetzinger et al. 2006). If the calibration slopes obtained from various regions and various coral species agree (within confidence limits), then one would be in a favorable position and would be allowed to calculate an overall calibration slope (from all data points), an estimate that would then enjoy wider applicability. In that regard, it is perhaps apt to remind readers of the work done by the fathers of isotope paleoclimatology: Epstein et al. (1953) went so far as to assume a universal calibration slope of $\hat{\beta}_1 = -4.3 \text{ } ^\circ\text{C}/\text{‰}$ for all marine carbonates. (To be more precise, these authors also assumed a quadratic term.) Notably, Epstein et al. (1953) applied the correct regression direction, namely of the response variable temperature on the predictor $\delta^{18}\text{O}$.

5 Conclusions

The linear calibration model is a powerful statistical tool that can be utilized to predict an unknown variable, Y , by means of observations of a proxy variable, X . Due to the nature of most prediction tasks, it is the relative change that is relevant for practical applications (i.e., the slope, which determines the change in Y per change in X), while the absolute value (i.e., the intercept) is less important. Note that the intercept is often arbitrarily constrained via use of anomaly values or employment of certain standards or units.

Since the calibration involves the estimation of regression model parameters on the basis of observed noisy datasets of a limited size, there are uncertainties involved. It is therefore of key importance to achieve a calibration slope estimation without bias. However, due to the fact that the proxy observations exhibit errors (i.e., predictor noise), advanced statistical routines have to be employed for correctly fitting the linear errors-in-variables calibration model (Eq. 2) to data. A similar quest for a state-of-the-art, data-driven uncertainty determination algorithm exists, which is rooted in the typically rather complex properties of the associated noise processes (i.e., for both Y and X), namely, non-Gaussian distributional shapes and autocorrelations. Such so-called problematic noise properties are ubiquitous in climate sciences, but they also appear in other branches of natural sciences, social sciences, medicine, or engineering.

The present paper shows that WLSXY estimation is able to deliver accurate and unbiased slope estimations under heteroscedasticity. In the case of homoscedasticity, besides WLSXY, the OLSBC estimation approach also performs well. On the other hand, estimation techniques to be avoided include OLS and RMA regression estimation, each of which may lead to severely biased calibration slope estimates. It is further necessary to take the correct regression direction into account, which is of the response variable, Y (e.g., SST), on the predictor or proxy variable, X (e.g., coral $\delta^{18}\text{O}$).

Both methods of choice, however, require prior knowledge about the size of the predictor noise; this means that for homoscedasticity, OLSBC requires information about S_X , and for heteroscedasticity, WLSXY about $S_X(i)$. If such knowledge is unavailable, then there is a problem with calibration since a biased slope estimate may result. However, the experimenter, who generates the prediction values, $X(i)$, has some control to reduce the bias. As Eq. (8) informs, such control can be gained by making $\text{VAR}[X(i)]$ large against a guessed value of S_X^2 ; then the results obtained using OLS lead to only a small bias. (Mudelsee 2014, Sect. 8.2.1, Sect. 8.3.2, Sect. 8.3.3 and Sect. 8.3.4 therein) presents ideas on modeling incomplete prior knowledge and several Monte Carlo simulation experiments on that theme. The Monte Carlo simulation results further demonstrate that a certain number of points (say, $n \geq 50$) may be necessary for the methodological advantages of OLSBC or WLSXY to take significant effect.

The present paper also introduces a powerful bootstrap resampling scheme (Sect. 3.1) which helps to obtain accurate CIs and prediction intervals under the above-mentioned problematic noise properties. A Monte Carlo simulation experiment with artificial time series reveals good coverage performance of 95% Student's t CIs under real-world conditions.

The errors-in-variables regression problem has long been known in climate sciences as well; see DeLong et al. 2007; Mann et al. 2007; Ammann et al. 2010; Mudelsee 2014, Ch. 8 therein; Thompson 2022, Sect. 3.6 therein and references cited in these works. The two methodological novelties presented in this paper are as follows. First, the bootstrap approach for CI construction (Sect. 3.1) is introduced to the peer-reviewed literature; the core of it has already been described in the book by Mudelsee (2014), but the present paper gives the full formula for the determination of the predictor uncertainty (Sect. 3.1). Second, the Monte Carlo simulation experiment (Sect. 3.2) presents new computer runs which show quantitatively that the WLSXY and OLSBC methods perform well for the proxy calibration problem in the presence of predictor noise, while the OLS and RMA methods fail. To the best of the author's knowledge, there exist no other Monte Carlo simulation experiments published in the peer-reviewed literature on such tests of uncertainty measures for proxy calibration. The present paper therefore takes a step beyond the state of the art as regards method validation for that estimation problem. In addition to those novelties, the present paper has a new paleoclimatic application (Sect. 4) and presents a novel calibration software tool (Appendix).

The data-analytical machinery developed in the present paper was applied to a published dataset of SST predictions by means of $\delta^{18}\text{O}$ measurements made on the coral climate archive. Evidently, there is room for future application of the proxy technique shown herein on the basis of new datasets employed. However, it should also be enlightening to reanalyze existing data series using the presented state-of-the-art methodological approach in a consistent manner. The approach itself is put into Fortran software (LINCAL), which is made publicly available as source code (see Appendix). As a reminder, paleoclimatic science has over the past few decades developed powerful laboratory-based approaches to improve the prediction quality of proxy variables (Bradley 1999; Cronin 2010; Brenner et al. 2017; Felis 2020). This means that the calibration results presented here (Fig. 3) are not the last word. Improved calibrations for coral $\delta^{18}\text{O}$ as an SST proxy are expected to come from paired measurements, that is, from $\delta^{18}\text{O}$ and Sr/Ca measured on the same sample.

Acknowledgements Logan Brenner kindly sent the data for the application. She, as well as Andrew Dolman, Thomas Felis, and Miriam Pfeiffer, provided helpful comments on the manuscript draft. Two anonymous reviewers made constructive remarks on the initial submission, which helped to create a revised manuscript version.

Funding Open Access funding enabled and organized by Projekt DEAL.

Declarations

Conflict of interest The author declares that he has no conflict of interest.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted

by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

Appendix: Calibration Software LINCAL

The software LINCAL can be used to estimate the parameters of a linear errors-in-variables calibration model (Eq. 2) and to predict the response, Y , over a certain range of values of the predictor, X . The software, as Windows executable and as Fortran 90 source code, is available at Zenodo (doi: 10.5281/zenodo.7440651) and the author's academic website (<https://www.manfredmudelsee.com/software/LINCAL>). These sites also contain a file (Brenner.txt) with the application data (Sect. 4) to serve testing of the software and reproducing the findings of the present paper. This appendix gives a short manual on how to run LINCAL; the source code contains a detailed documentation of the method and its numerical implementation.

It is convenient to (1) copy the executable on the computer into a folder of choice (say, C:/LINCAL), (2) open the Windows command prompt (by running cmd.exe from the Windows start menu), (3) navigate into the folder (by typing "cd C:/LINCAL" followed by "Enter"), and (4) run the software on the command line ("LINCAL.exe" followed by "Enter"). After a welcome screen ("Enter"), the four data cases—homoscedastic/heteroscedastic errors and availability/unavailability of time values—are explained to the user. Case 1 (homoscedasticity, unavailability of time values) requires an input data file (possibly including the path on the computer) containing $\{x(i), y(i)\}_{i=1}^n$ and a manual input of S_X and S_Y ; case 2 (homoscedasticity, availability of time values) requires an input data file containing $\{t(i), x(i), y(i)\}_{i=1}^n$ and a manual input of S_X and S_Y ; case 3 (heteroscedasticity, unavailability of time values) requires an input data file containing $\{x(i), y(i), S_X(i), S_Y(i)\}_{i=1}^n$; and case 4 (heteroscedasticity, availability of time values) requires an input data file containing $\{t(i), x(i), y(i), S_X(i), S_Y(i)\}_{i=1}^n$.

The estimation is performed by the software via OLSBC (homoscedasticity) or WLSXY (heteroscedasticity), while the prediction is carried out by means of the estimated parameters. Under homoscedasticity, the value of S_X is taken for the calculation of the prediction error (Sect. 3.1, step 9), while under heteroscedasticity, a new value, $S_X(n+1)$, can be selected. The prediction values, $x(n+1)$, are between $x_{\min} - f \cdot (x_{\max} - x_{\min})$ and $x_{\max} + f \cdot (x_{\max} - x_{\min})$, where x_{\min} is the minimum and x_{\max} the maximum of $\{x(i)\}_{i=1}^n$. The value of $f = 0.5$ makes the prediction interval sufficiently wide. The number of prediction points is 2,000, which achieves a sufficiently densely spaced prediction grid.

The output file (LINCAL.dat) contains references, information about the data case, input data file name, sample size, standard deviations, estimation result with 95% bootstrap Student's t CIs, persistence properties ($\hat{\tau}'_X, \hat{\tau}'_Y$), bootstrap block length, and the prediction values including upper and lower CI bounds.

References

- Ammann CM, Genton MG, Li B (2010) Technical note: correcting for signal attenuation from noisy proxy data in climate reconstructions. *Clim Past* 6(2):273–279
- Bradley RS (1999) *Paleoclimatology: reconstructing climates of the Quaternary*, 2nd edn. Academic Press, San Diego
- Brenner LD, Linsley BK, Potts DC (2017) A modern Sr/Ca- $\delta^{18}\text{O}$ -sea surface temperature calibration for *Isopora* corals on the Great Barrier Reef. *Paleoceanogr* 32(2):182–194
- Brocas WM, Felis T, Mudelsee M (2019) Tropical Atlantic cooling and freshening in the middle of the last interglacial from coral proxy records. *Geophys Res Lett* 46(14):8289–8299
- Carlstein E (1986) The use of subseries values for estimating the variance of a general statistic from a stationary sequence. *Ann Stat* 14(3):1171–1179
- Carroll RJ, Ruppert D, Stefanski LA, Crainiceanu CM (2006) *Measurement error in nonlinear models: a modern perspective*, 2nd edn. Chapman and Hall, Boca Raton
- Cronin TM (2010) *Paleoclimates: understanding climate change past and present*. Columbia University Press, New York
- Davis JC (1986) *Statistics and data analysis in geology*, 2nd edn. Wiley, New York
- DeLong KL, Quinn TM, Taylor FW (2007) Reconstructing twentieth-century sea surface temperature variability in the southwest Pacific: a replication study using multiple coral Sr/Ca records from New Caledonia. *Paleoceanogr* 22(4):4212. <https://doi.org/10.1029/2007PA001444>
- Deming WE (1943) *Statistical adjustment of data*. Wiley, New York
- Draper NR, Smith H (1981) *Applied regression analysis*, 2nd edn. Wiley, New York
- Efron B (1979) Bootstrap methods: another look at the jackknife. *Ann Stat* 7(1):1–26
- Efron B, Tibshirani RJ (1993) *An introduction to the bootstrap*. Chapman and Hall, London
- Epstein S, Buchsbaum R, Lowenstam HA, Urey HC (1953) Revised carbonate-water isotopic temperature scale. *Geol Soc Am Bull* 64(11):1315–1325
- Evans MN, Tolwinski-Ward SE, Thompson DM, Anchukaitis KJ (2013) Applications of proxy system modeling in high resolution paleoclimatology. *Quat Sci Rev* 76:16–28
- Falcy MR, McCormick JL, Miller SA (2016) Proxies in practice: calibration and validation of multiple indices of animal abundance. *J Fish Wildl Manag* 7(1):117–128
- Felis T (2020) Extending the instrumental record of ocean-atmosphere variability into the last interglacial using tropical corals. *Oceanogr* 33(2):68–79
- Felis T, Pätzold J, Loya Y, Fine M, Nawar AH, Wefer G (2000) A coral oxygen isotope record from the northern Red Sea documenting NAO, ENSO, and North Pacific teleconnections on Middle East climate variability since the year 1750. *Paleoceanogr* 15(6):679–694
- Felis T, McGregor HV, Linsley BK, Tudhope AW, Gagan MK, Suzuki A, Inoue M, Thomas AL, Esat TM, Thompson WG, Tiwari M, Potts DC, Mudelsee M, Yokoyama Y, Webster JM (2014) Intensification of the meridional temperature gradient in the Great Barrier Reef following the Last Glacial Maximum. *Nat Commun* 5(1):4102. <https://doi.org/10.1038/ncomms5102>
- Fishman GS (1996) *Monte Carlo: concepts, algorithms, and applications*. Springer, New York
- Freedman D (1984) On bootstrapping two-stage least-squares estimates in stationary linear models. *Ann Stat* 12(3):827–842
- Freedman DA, Peters SC (1984) Bootstrapping an econometric model: some empirical results. *J Bus Econ Stat* 2(2):150–158
- Fuller WA (1987) *Measurement error models*. Wiley, New York
- Hetzinger S, Pfeiffer M, Dullo W-C, Ruprecht E, Garbe-Schönberg D (2006) Sr/Ca and $\delta^{18}\text{O}$ in a fast-growing *Diploria strigosa* coral: evaluation of a new climate archive for the tropical Atlantic. *Geochem, Geophys, Geosyst* 7(10):Q10002. <https://doi.org/10.1029/2006GC003147>
- Hu J, Emile-Geay J, Partin J (2017) Correlation-based interpretations of paleoclimate data—where statistics meet past climates. *Earth Planet Sci Lett* 459:362–371
- Kotz S, Johnson NL, Read CB (eds) (1986) *Encyclopedia of statistical sciences*, vol 7. Wiley, New York
- Künsch RH (1989) The jackknife and the bootstrap for general stationary observations. *Ann Stat* 17(3):1217–1241
- Lemley GM (2012) *Assessing $\delta^{18}\text{O}$ in the coral genus Isopora for reconstructing Indo-Pacific regional and seasonal climate variability*. M.Sc. thesis, State University of New York, Albany, USA
- Lindley DV (1947) Regression lines and the linear functional relationship. *J R Stat Soc Suppl* 9(2):218–244

- Mahnken JD, Chen X, Brown AR, Vidoni ED, Billinger SA, Gajewski BJ (2014) Evaluating variables as unbiased proxies for other measures: assessing the step test exercise prescription as a proxy for the maximal, high-intensity peak oxygen consumption in older adults. *Int J Stat Probab* 3(4):25–34
- Mann ME, Rutherford S, Wahl E, Ammann C (2007) Robustness of proxy-based climate field reconstruction methods. *J Geophys Res* 112(D12):D12109. <https://doi.org/10.1029/2006JD008272>
- McClelland HLO, Halevy I, Wolf-Gladrow DA, Evans D, Bradley AS (2021) Statistical uncertainty in paleoclimate proxy reconstructions. *Geophys Res Lett* 48(15):e2021GL092773
- Mudelsee M (2002) TAUEST: a computer program for estimating persistence in unevenly spaced weather/climate time series. *Comput Geosci* 28(1):69–72
- Mudelsee M (2014) Climate time series analysis: classical statistical and bootstrap methods, 2nd edn. Springer, Cham
- Mudelsee M (2020) Statistical analysis of climate extremes. Cambridge University Press, Cambridge
- Nishida K, Iguchi A, Ishimura T, Sakai K, Suzuki A (2014) Skeletal isotopic responses of the scleractinian coral *Isopora palifera* to experimentally controlled water temperatures. *Geochem J* 48(4):e9–e14
- Perron M, Sura P (2013) Climatology of non-Gaussian atmospheric statistics. *J Clim* 26(3):1063–1083
- Pfeiffer M, Reuning L, Zinke J, Garbe-Schönberg D, Leupold M, Dullo W-C (2019) 20th century $\delta^{18}\text{O}$ seawater and salinity variations reconstructed from paired $\delta^{18}\text{O}$ and Sr/Ca measurements from a La Reunion coral. *Paleoceanogr Paleoclimatol* 34(12):2183–2200
- Press WH, Teukolsky SA, Vetterling WT, Flannery BP (1992) Numerical recipes in Fortran 77: the art of scientific computing, 2nd edn. Cambridge University Press, Cambridge
- Reed BC (1989) Linear least-squares fits with errors in both coordinates. *Am J Phys* 57(7):642–646, Erratum: Reed BC (1990) *Am J Phys* 58(2):189
- Reed BC (1992) Linear least-squares fits with errors in both coordinates. II: comments on parameter variances. *Am J Phys* 60(1):59–62
- Robinson GK (1982) Confidence intervals and regions. In: Kotz S, Johnson N, Read C (eds) *Encyclopedia of statistical sciences*, vol 2. Wiley, New York, pp 120–127
- Sherman M, Speed FM Jr, Speed FM (1998) Analysis of tidal data via the blockwise bootstrap. *J Appl Stat* 25(3):333–340
- Smith TM, Reynolds RW, Peterson TC, Lawrimore J (2008) Improvements to NOAA's historical merged land-ocean surface temperature analysis (1880–2006). *J Clim* 21(10):2283–2296
- Squire PT (1990) Comment on “Linear least-squares fits with errors in both coordinates,” by B. C. Reed [*Am. J. Phys.* 57, 642–646 (1989)]. *Am J Phys* 58(12):1209
- Thompson DM (2022) Environmental records from coral skeletons: a decade of novel insights and innovation. *WIREs Clim Change* 13(1):e745. <https://doi.org/10.1002/wcc.745>
- Tingley MP, Craigmile PF, Haran M, Li B, Mannshardt E, Rajaratnam B (2012) Piecing together the past: statistical insights into paleoclimatic reconstructions. *Quat Sci Rev* 35:1–22
- Trenberth KE (1984) Some effects of finite sample size and persistence on meteorological statistics. Part I: autocorrelations. *Mon Weather Rev* 112(12):2359–2368
- York D (1966) Least-squares fitting of a straight line. *Can J Phys* 44(5):1079–1086
- York D (1967) The best isochron. *Earth Planet Sci Lett* 2(5):479–482
- York D (1969) Least squares fitting of a straight line with correlated errors. *Earth Planet Sci Lett* 5(5):320–324