



Inferring climate variability from nonlinear proxies: application to palaeo-ENSO studies

J. Emile-Geay¹ and M. Tingley²

¹Department of Earth Sciences & Center for Applied Mathematical Sciences, University of Southern California, Los Angeles, CA, USA

²Departments of Statistics & Meteorology, Pennsylvania State University, State College, PA, USA

Correspondence to: J. Emile-Geay (julienege@usc.edu)

Received: 30 May 2015 – Published in *Clim. Past Discuss.*: 8 July 2015

Revised: 30 November 2015 – Accepted: 30 November 2015 – Published: 15 January 2016

Abstract. Inferring climate from palaeodata frequently assumes a direct, linear relationship between the two, which is seldom met in practice. Here we simulate an idealized proxy characterized by a nonlinear, thresholded relationship with surface temperature, and we demonstrate the pitfalls of ignoring nonlinearities in the proxy–climate relationship. We explore three approaches to using this idealized proxy to infer past climate: (i) methods commonly used in the palaeoclimate literature, without consideration of nonlinearities; (ii) the same methods, after empirically transforming the data to normality to account for nonlinearities; and (iii) using a Bayesian model to invert the mechanistic relationship between the climate and the proxy. We find that neglecting nonlinearity often exaggerates changes in climate variability between different time intervals and leads to reconstructions with poorly quantified uncertainties. In contrast, explicit recognition of the nonlinear relationship, using either a mechanistic model or an empirical transform, yields significantly better estimates of past climate variations, with more accurate uncertainty quantification. We apply these insights to two palaeoclimate settings. Accounting for nonlinearities in the classical sedimentary record from Laguna Pallcacocha leads to quantitative departures from the results of the original study, and it markedly affects the detection of variance changes over time. A comparison with the Lake Challa record, also a nonlinear proxy for El Niño–Southern Oscillation, illustrates how inter-proxy comparisons may be altered when accounting for nonlinearity. The results hold implications for how univariate, nonlinear recorders of normally distributed climate variables are interpreted, compared

to other proxy records, and incorporated into multiproxy reconstructions.

1 Introduction

A principal goal of palaeoclimatology is to infer climate information from geochemical, physical, or lithological signals embedded in various proxy archives. Implicit in most analyses of palaeoclimate records is the assumption that the observations are linearly related to the target climate quantity, so that traditional calibration approaches (e.g. Brown, 1994) allow climate to be inferred given the proxy values.

The climate system, however, is rife with nonlinearities, as are the processes translating climate signals into proxy records (see e.g. Evans et al., 2013, for a review). Examples of nonlinear processes known to markedly distort climate signals include biological threshold effects on tree growth (Vaganov et al., 2006; Anchukaitis et al., 2006; Tolwinski-Ward et al., 2011), karst effects on speleothem $\delta^{18}\text{O}$ records (Baker et al., 2012; Jex et al., 2013), and hydrodynamic effects on flood proxies. Nonlinearities are especially pronounced in terrestrial proxy records from the tropics, where temperature experiences its lowest dynamic range and precipitation its highest dynamic range, resulting in distributions that are non-normal, with strong positive skew. These records are frequently interpreted as reflecting some aspect of the El Niño–Southern Oscillation (ENSO) phenomenon, involving hydrological balance (Conroy et al., 2008, 2009), river runoff (Rodbell et al., 1999; Moy et al., 2002; Rein et al., 2004, 2005), or wind speed (Wolff et al., 2011) – all nonlinear functions of ENSO state. These proxies can feature statis-

tical distributions that are different from those of the climate phenomenon they purport to record, and different from one another. What are the consequences of inferring changes in ENSO state from such land-based, nonlinear records? More generally, do nonlinear proxies allow for reliable statements about changes in climate variability, and if so, under which conditions? Can nonlinearity be overcome to correctly infer changes in the underlying climate?

Even when the target climate quantity is well approximated by a normal distribution, nonlinearities often manifest themselves as non-normality in the proxy distribution. This is because a normally distributed proxy is only expected if the proxy is a linear recorder of a normally distributed climate variable, as the linear transform of a Gaussian random vector is also Gaussian. Temperature fluctuations, especially if averaged over a month or more, typically obey normal statistics, so linear temperature proxies tend to obey normal statistics as well. In contrast, tropical run-off proxies, for instance, are nonlinearly related to land precipitation, which is heavily skewed (thus non-normal) and only indirectly linked to sea-surface temperature (SST). Such records are not directly amenable to analysis using common techniques assuming linearity and normality (e.g. spectral analysis, principal component analysis, least-squares regression methods, correlation analysis, and parametric hypothesis testing).

One strategy to overcome this difficulty is to use generalized linear models (GLMs; e.g. McCullagh and Nelder, 1989; Lindsey, 1997) that enable linear inference methods despite nonlinear relationships between predictors and predictands. Such methods require the specification of the functional relationship between proxy and climate, which may not always be known. Even if these *link functions* are known, GLMs require a case-by-case treatment, with each proxy potentially displaying a different relation to climate. Such a case-specific treatment can become prohibitive in multiproxy studies, which range from graphical comparisons of a half-dozen records in a stack to assess the synchronous nature of climate fluctuations (e.g. Conroy et al., 2008, Fig. 5, or Kirby et al., 2014, Fig. 5) to using several hundreds of proxies to reconstruct temperature over the late Holocene using statistical models (e.g. Mann et al., 1998; Luterbacher et al., 2004; Moberg et al., 2005; Mann et al., 2008; Kaufman et al., 2009; Mann et al., 2009; Barriopedro et al., 2011; Emile-Geay et al., 2013a, b; Tingley and Huybers, 2013). In such cases, it would be desirable to account for nonlinearity so that nonlinear proxies may be used alongside linear proxies without recourse to individual treatment.

This article draws attention to the pitfalls of ignoring nonlinearity in the proxy–climate relationship and explores a number of approaches to using nonlinear proxies to infer climate variability when the underlying climate obeys normal statistics. Our objective is not to consider all of the vagaries of fitting a straight line through noisy data, including calibration vs. regression, measurement errors in predictor variables, and various regularization techniques. For in-

depth discussions of these issues, relevant and complementary to the problem at hand, see Brown (1994) and Christiansen (2014). Here we focus on the related challenge of inferring climate variability from proxy archives that are nonlinear, univariate recorders of the target climate variable. We consider both simple, generic strategies that are amenable to multiproxy studies, and more case-specific treatments.

Section 2 presents a simple toy model for a nonlinear proxy, used to illustrate the effects of nonlinearity on inferences about climate variability. Section 3 outlines various practical solutions to the problem, which we then apply to the Laguna Pallcacocha and Lake Challa records in Sect. 4. Section 5 provides discussion and concluding remarks.

2 Nonlinearity in climate proxies

2.1 A simple proxy model

To illustrate the challenges posed by nonlinear recorders of climate, we consider an idealized model for a run-off proxy that displays a univariate and stationary but nonlinear response to ENSO-induced rainfall. Sediment-based run-off proxies are expected to feature nonlinearities for at least four reasons: sediment mobilization is a nonlinear function of flow speed (Rose, 1993); flow speed is a nonlinear function of rainfall (e.g. Maidment, 1993); rainfall is a nonlinear function of sea-surface temperature (e.g. Lengaigne and Vecchi, 2010); and finally, we assume that rainfall at the location of the proxy occurs only during the warm phase of ENSO, inducing an asymmetric relationship with climate.

To characterize ENSO state, we use December–January–February averages of the NINO3.4 index (average sea-surface temperature anomaly in the region (5° N–5° S; 170° W–120° E)) from the Extended Reconstructed Sea Surface Temperature (ERSSTv3) data set (Smith et al., 2008). Although NINO3.4 statistics are well known to be skewed at monthly scales, the seasonal averaging results in a nearly Gaussian distribution, so that for this example non-normality in the proxy arises on account of the proxy–climate relationship rather than the underlying climate. To add stochasticity, we assume that the idealized land-based proxy is sensitive to the local expression of NINO3.4 via $c'(t) = c(t) + \epsilon_1(t)$, where $c(t)$ is the NINO3.4 index, $c'(t)$ is the local expression of ENSO variability, and for convenience ϵ_1 is a zero-mean Gaussian white-noise process with variance δ^2 . An additive noise ϵ_2 , in the proxy units, represents measurement error. Our idealized model for the run-off proxy then takes the form

$$p(t) = \begin{cases} \alpha [c(t) + \epsilon_1(t)]^\beta + \epsilon_2(t) & c'(t) \geq 0 \\ \epsilon_2(t) & \text{otherwise} \end{cases}, \quad (1)$$

where $\alpha > 0$ and $\beta > 1$. The exact value of β is relatively unimportant in the context of our example; in the following, we choose $\alpha = 1$ and $\beta = 3$, though our results are qualitatively insensitive to this choice.

As we expect the error introduced by the local vs. global ENSO signal to be much larger, we neglect measurement error in the remainder for the sake of simplicity of exposition. We also ignore chronological errors, though they would be important in nature. This example may be viewed as a best-case scenario: it is the most parsimonious representation of the fact that the ENSO signal is global but experienced locally by the proxy, with deviations controlled by δ .

A proxy generated according to Eq. (1) selectively records El Niño events, and it does so with great sensitivity – in this respect it is an ideal proxy. Fig. 1 shows the NINO3.4 index and a pseudo-proxy series derived from it using Eq. (1). It is qualitatively similar to run-off proxy records commonly interpreted as reflecting ENSO variability (e.g. Rodbell et al., 1999; Moy et al., 2002; Rein et al., 2004). The pseudo-proxy correctly identifies maxima in the target NINO3.4 index but exaggerates their relative magnitudes, while the minima in the target series are clipped at zero. Analysis of this skewed pseudo-proxy will thus lead to correct conclusions regarding the *timing* of positive excursions in the NINO3.4 index, but the nonlinearity and thresholding can potentially result in overestimates of changes in ENSO variability. Quantitatively, the ratio of the NINO3.4 sample variances within the 1950–1999 period and the 1900–1949 period is 1.25, while the distribution of variance ratios for 10 000 realizations of the pseudo-proxy, differing in their noise alone, is peaked at higher values (Fig. 3), with a median of 5.25 and interquartile range (Wilks, 2011, p. 26) of 6.20. Indeed, the variance ratio for 86 % of the realizations of $p(t)$ is more than 50 % higher than that calculated from the original NINO3.4 series, indicating the large impact of the nonlinear transform (Eq. 1) on the second-moment properties of the proxy as compared with the climate. Note that this assessment considered positive values of the proxies only, but results are nearly identical without this restriction.

2.2 Coping with nonlinearity

Our aim is to explore the quantitative information about climate variability that can reliably be inferred from a proxy record generated according to Eq. (1), and to explore the pitfalls of failing to account for the nonlinear relationship with climate. A full review of all existing inference strategies in the presence of nonlinearity and non-normality is beyond the scope of this paper; instead, we focus on three workable approaches that may be useful in palaeoclimatology.

- *Functional transformation*: when the transfer function between proxy and climate is known a priori, it may be leveraged to infer climate. Indeed, in the absence of noise, Eq. (1) expresses a one-to-one mapping between proxy values and local climate when the latter displays positive excursions. This relationship over positive values is therefore invertible and may be used for inference, for instance by specifying this relationship as a

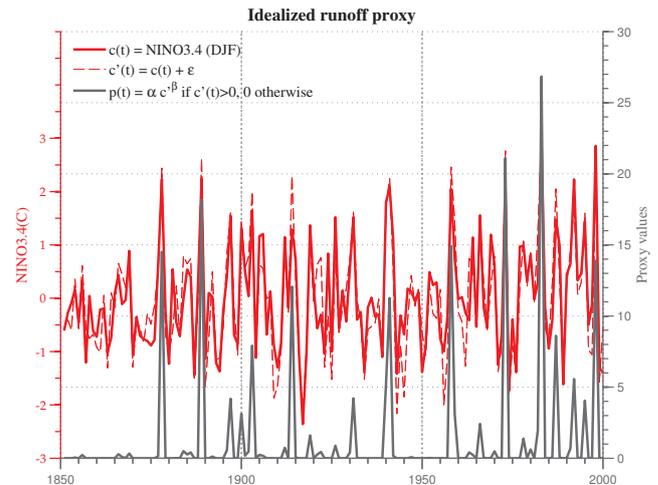


Figure 1. A nonlinear recorder of ENSO activity. Solid red: December–January–February average of the NINO3.4 index from the ERSSTv3 data set (Smith et al., 2008), standardized to Z scores; dashed red: standardized NINO3.4 with additive $\mathcal{N}(0, 0.25)$ noise; grey: the noisy NINO3.4 signal transformed according to Eq. (1), with $\alpha = 1$ and $\beta = 3$.

link function within a GLM (Nelder and Wedderburn, 1972). Linear regression can then be used to perform inference on climate given the proxy in a manner that accounts for nonlinearities. However, as all negative excursions are mapped to zero, the GLM framework cannot be used over such points. A related approach is to embed the mechanistic model describing how climate signals are recorded in the proxy archives within a hierarchical model. Under some circumstances, such models may be inverted using Bayes’ rule to yield quantitative information about the underlying climate (e.g. Tolwinski-Ward et al., 2011, 2013). This is similar in spirit to the GLM framework, as both exploit the functional relationship between proxy and climate. A limitation common to both frameworks is that the parametric form of the proxy–climate relationship must be known, which may not be the case in practice.

One important difference between the two approaches concerns the choice of dependent vs. independent variable. Under the GLM framework, there is ambiguity in the set-up of the model, as the goal is to infer climate from proxies but the proxies are best understood as the dependent variable. Bayes’ rule, in contrast, naturally inverts the etiologically correct specification of climate as the dependent variable to infer climate conditional on the proxies (Brown, 1994; Tingley et al., 2012; Christiansen, 2014).

- *Empirical transformation*: when a parametric relationship is not known a priori, or when parameter estimation is unreliable, empirical transforms offer a useful alter-

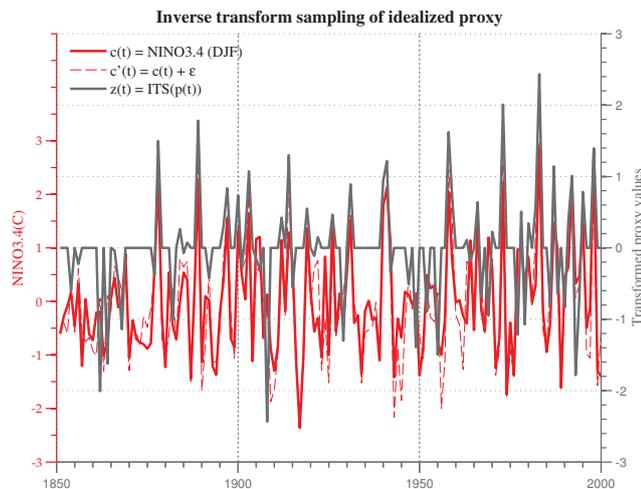


Figure 2. Correcting for non-normality using inverse transform sampling. Solid red: the normalized NINO3.4 index; dashed red: standardized NINO3.4 with additive $\mathcal{N}(0, 0.25)$ noise; grey: the transformed version of the pseudo-proxy depicted in Fig. (1). Because the distribution of $p(t)$ has a mass point at 0, we compute the transform only on those time points where $p(t) > 0$, leaving it 0 otherwise. Code for implementing the transform is available at <https://github.com/CommonClimate/common-climate/blob/master/gaussianize.m>.

native. For example, the well-established power transforms (Box and Cox, 1964) render non-normal time series approximately normal, but the estimation of their exponents can be unstable in the presence of noise (Bickel and Doksum, 1981). An alternative is afforded by inverse transform sampling (ITS; e.g. van Albada and Robinson, 2007), a robust approach for converting any distribution to a standard normal. The transform in this case proceeds by evaluating the inverse normal cumulative distribution at the percentiles of the observations according to the fitted cumulative distribution function; the resulting values are then approximately normally distributed. Because it is quantile-based, this method is less sensitive to outliers than power transforms. It is also non-parametric and hence more widely applicable. We note that the widely used standardized precipitation index (SPI; McKee et al., 1993, 1995), often used to characterize drought conditions, amounts to a parametric ITS assuming a gamma distribution for precipitation measurements.

We now apply those approaches to better understand past ENSO variability as depicted by nonlinear proxies.

2.3 Application to variance changes

In the absence of information on the proxy–climate link, ITS is computationally simple and non-parametric, and allows the transformed time series to be analysed using classical

tools. Fig. 2 shows the result of applying this transform to a pseudo-proxy record generated according to Eq. (1). The amplitude of positive excursions in the transformed series is now comparable to the original NINO3.4 series, and in repeated experiments the distribution of variance ratios between the early and late twentieth century (Fig. 3) is centred at 1.5, much closer to the true value of 1.25, and the interquartile range is now reduced to 0.75 (compare with 5.25 and 6.20, respectively, using the raw proxy values). Similarly, the probability of overstating the true change in variance by more than 50 % is only 22 % (vs. 86 % using the raw proxy values). Similar results are obtained by inverting Eq. (1), showing that application of the correct transform can effectively correct for nonlinearity.

However, while the transform adequately corrects for the cubic nonlinearity, it cannot overcome the fundamental limitation that this idealized proxy only records El Niño events, while information about La Niña events is irretrievably lost. Proxies generated according to Eq. (1) thus represent a lossy transformation of the underlying climate signal, one that no amount of statistical ingenuity can ever rectify. A good probabilistic analysis should nonetheless yield reasonable estimates of positive climate excursions, and accurate error bounds for all climate excursions.

3 Inference from nonlinear proxies: three approaches

We compare the relative merits of three approaches to reconstructing past climate from a proxy generated according to Eq. (1), given a calibration period when both climate and proxy observations are available. The problem is more general than the estimation of variance changes considered above, since it seeks to estimate climate values and quantify the associated uncertainties. We note that the problem of statistical calibration of noisy predictors has received extensive attention in the statistical (e.g. Brown, 1994) and climate (e.g. Christiansen, 2014, and references therein) literature. Our goal is neither to present an exhaustive list of possibilities nor to re-invent the wheel, but rather to investigate some choice practical methods that may be of use to palaeoclimatologists.

3.1 Three approaches to inference

- Method 1 (RAW): nonlinearity is ignored, and the proxy series $p(t)$ is used as a predictor in a standard linear regression (an assumption implicit to many palaeoclimate studies).
- Method 2 (ITS): nonlinearity is recognized, empirically corrected via inverse transform sampling, and used as a predictor in a standard linear regression.
- Method 3 (BPM): nonlinearity is recognized, and a probability model that allows for Bayesian inversion of

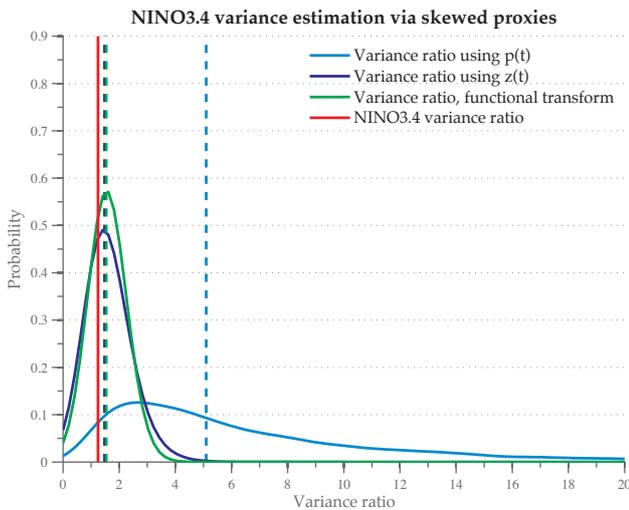


Figure 3. Distributions of estimated NINO3.4 variance ratios within the periods 1900–1949 and 1950–1999, using positive proxy excursions. Light blue: distribution of ratios from 10 000 samples of $p(t)$ derived according to Eq. (1) with noise variance set to 0.25 (compare with the dark grey curve in Fig. 1); dark blue: distribution of ratios estimated from $z(t)$, the empirical transform of each sample of $p(t)$ (compare with the dark grey curve in Fig. 2); green: distribution of ratios estimated from the functional transform $((p/\alpha)^\beta)$. In each case, the median is indicated with a dashed vertical line of the same colour, while the ratio estimated from the original NINO3.4 series is indicated in red.

the structural dependence of the proxy on the climate (Eq. 1) is constructed. The model, described extensively in Appendix A, is used to infer the posterior distribution of climate values given proxy observations, $f(c|p)$.

These three cases involve increasing levels of sophistication but do not attempt to cover all possible choices. Instead, our purpose here is to illustrate the danger of ignoring nonlinearity (RAW), outline an optimal (but difficult) solution (BPM), and find a practical compromise between the two (ITS).

The experimental sample is composed of 10 000 surrogate climate time series $c(t)$ of length $n = 300$ years generated according to an AR(1) process, with mean zero and a lag-1 autocorrelation of $\gamma = 0.7$ to mimic the serial correlation typical of observed sea-surface temperature time series; conclusions are not sensitive to this choice. The innovation standard deviation, σ , is chosen so that the standard deviation of $c(t)$ is unity. We add $\epsilon_1 \sim N(0, \delta)$, with $\delta = 0.5 \times \sigma$, to produce $c'(t)$, and the idealized proxy series is then generated according to Eq. (1), with $\alpha = 1$ and $\beta = 3$.

In the RAW and ITS cases, inference is made via linear regression, where the regression model is trained on a calibration interval formed from the most recent 150 time points. To minimize erroneous inference when $p = 0$, the regression model is fit using only proxy observations that are greater than zero. We adopt a heuristic approach to the $p(t) = 0$ ob-

servations in the prediction interval, assigning to them the value \bar{c}_- , the calibration sample mean of $c(t)$ calculated over time points for which $p(t) = 0$. While somewhat unorthodox in the regression context, the divergent treatments in the $p = 0$ and $p > 0$ cases is a reasonable ad hoc procedure warranted by the thresholding behaviour of the proxy.

For the two regression cases, uncertainties are quantified via 95 % prediction intervals (PIs) following standard regression theory (e.g. Wilks, 2011, Eq. 7.22), where the variance of residuals is again computed separately for $p = 0$ and $p > 0$. In the Bayesian model, the posterior mean serves as the optimal prediction of $c(t)$, and 95 % credible intervals (Jaynes, 1976; Gelman et al., 2004) are readily obtained from the posterior samples.

For simplicity, we assume in the Bayesian treatment that all model parameters ($\alpha, \beta, \mu, \sigma, \delta$) are known. Doing so permits for closed-form expressions for the required posterior distributions that aid the interpretation of results. In practical applications these parameters would be inferred from the data as well, via a hierarchical model and Markov chain Monte Carlo-based inference (e.g. Gelman et al., 2004). By constructing the Bayesian model to match the proxy-generating process, in both structure and parameter values, we give it an unfair advantage over the other methods. We use this best-case scenario as a benchmark, expected to provide the best estimates of climate given the skewed proxy measurements, as well as reliable estimates of uncertainty. Appendix B explores the role played by the prior specification of the variance of the climate process, σ^2 .

3.2 Results

Results of performing inference on $c(t)$ using each of the three approaches are plotted in Fig. 4 for one realization of the climate and noise processes. The raw regression overestimates positive excursions (cf. time points 110 to 130), while in contrast the amplitudes of positive excursions as inferred from the regression on the transformed proxies and using the Bayesian inversion of the mechanistic model are in reasonable agreement with $c(t)$. The raw regression could thus exaggerate any changes in climate variability prior to the calibration period, a direct consequence of linearizing the non-linear climate-proxy relationship.

We investigate the properties of the three inference methods, and the sensitivity of the results to the ratio δ/σ , using residual plots (Fig. 5). For small noise levels ($\delta = 0.1\sigma$), and for $p > 0$, there is most structure in the residual plot for the RAW method and least for the BPM, indicating that the RAW and BPM approaches provide, respectively, the poorest and best fits to the data. The curvature of the residual plot for $p > 0$ is positive for the RAW regression, as the non-linearity magnifies large excursions. In contrast, the residual plot for the ITS regression displays negative curvature, indicating that the transform overly compresses large excursions and overly magnifies small ones. In the least noisy case

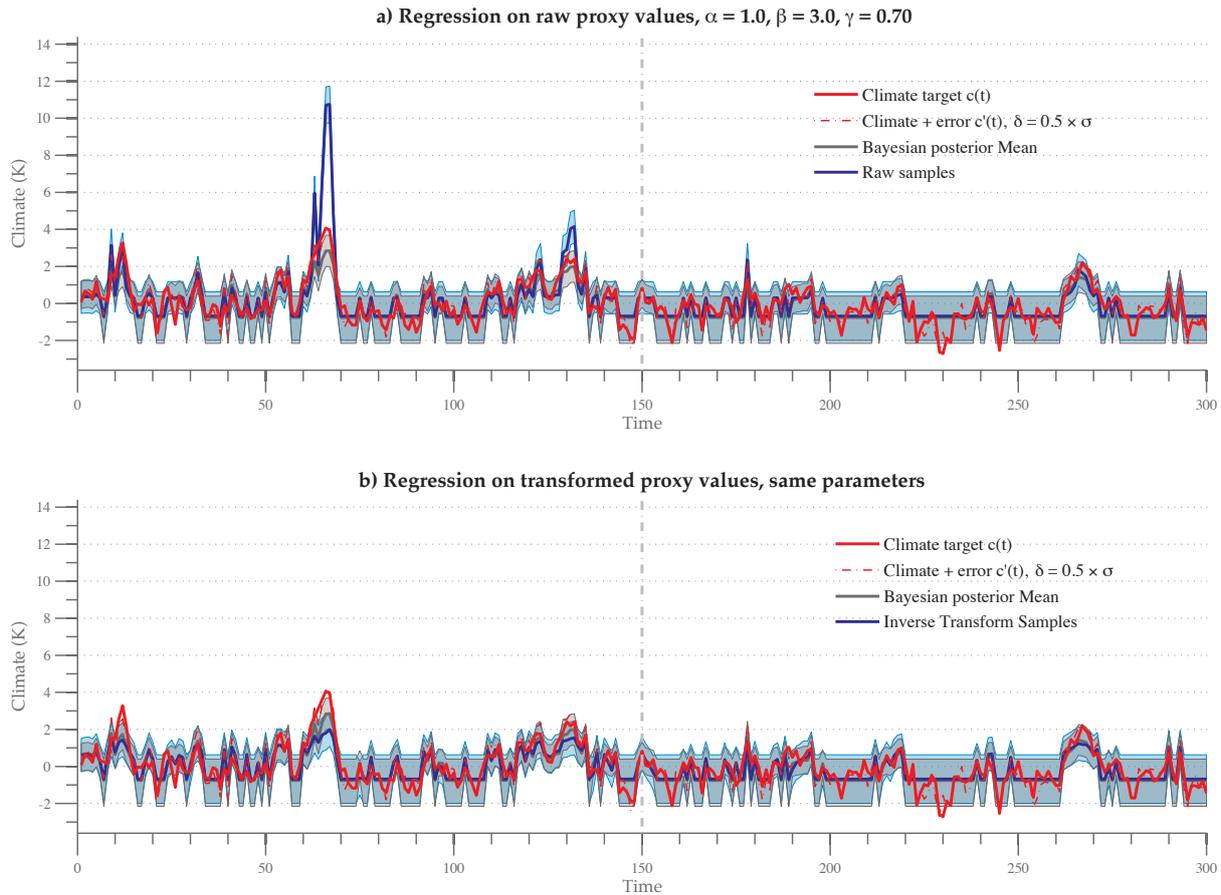


Figure 4. Inferences on simulated $c(t)$ using three different methods. **(a)** Inference using regression on raw proxy values (blue) and the Bayesian model (grey). **(b)** Inference using regression on transformed proxy values (blue) as well as the Bayesian model (grey). In both panels, $c(t)$ and $c'(t)$ are shown in solid and dashed red, respectively; blue shading represents 95 % frequentist prediction intervals; and grey intervals represent pointwise 95 % credible intervals from the Bayesian posterior distribution (see, for example, Tingley and Huybers, 2013).

($\delta = 0.1\sigma$), the Bayesian model yields near-perfect estimates for $p > 0$, while for $p = 0$ the distribution of residuals is similar for all three methods.

As the noise level increases (middle and bottom rows), the residual plots become less structured and the vertical spread in each case becomes comparable to that of the $p = 0$ case. For higher noise levels, there are large positive outliers in the residuals from the RAW proxy regression, as expected since the method makes no account of the nonlinearity. In contrast, the ITS regression and Bayesian inversion yield generally smaller residuals even for high noise levels.

We further investigate the diagnostic features of the inference procedures using the 10 000-member ensemble of realizations of the climate and noise processes for the intermediate noise level $\delta = 0.5\sigma$ (Table 1). To quantify the properties of the uncertainty intervals, we first evaluate their width, considering both the $p = 0$ and $p > 0$ cases for each inference method (top row). All three methods show larger uncertainties in the $p = 0$ case (around 2.7°C) with small differences between methods, consistent with the lack of information of-

ferred by the proxy in this regime. When $p > 0$, uncertainties are reduced for all methods, as expected from the information provided by the proxy. The raw proxy regression results in the widest uncertainty intervals, while those from the Bayesian inversion are narrowest.

To see how well such intervals encompass the true climate fluctuations, we calculate their actual coverage rates (Guttman, 1970), to be compared with a nominal coverage rate of 95 % (Table 1, rows 3–4). For the two regression approaches, the uncertainties are slightly too permissive in both cases, with actual coverage rates of about 94 %. The coverage rate for the mechanistic model is correct for the $p = 0$ case but only about 90 % when $p > 0$. In contrast to RAW and ITS, the mechanistic model does not involve two separate calibrations and solutions for the $p = 0$ and $p > 0$ cases. As such, the coverage rates are more appropriately assessed without division into the two cases, yielding PIs with a coverage rate of 93 with a standard error of $\pm 2\%$, compatible with the nominal rate of 95 %.

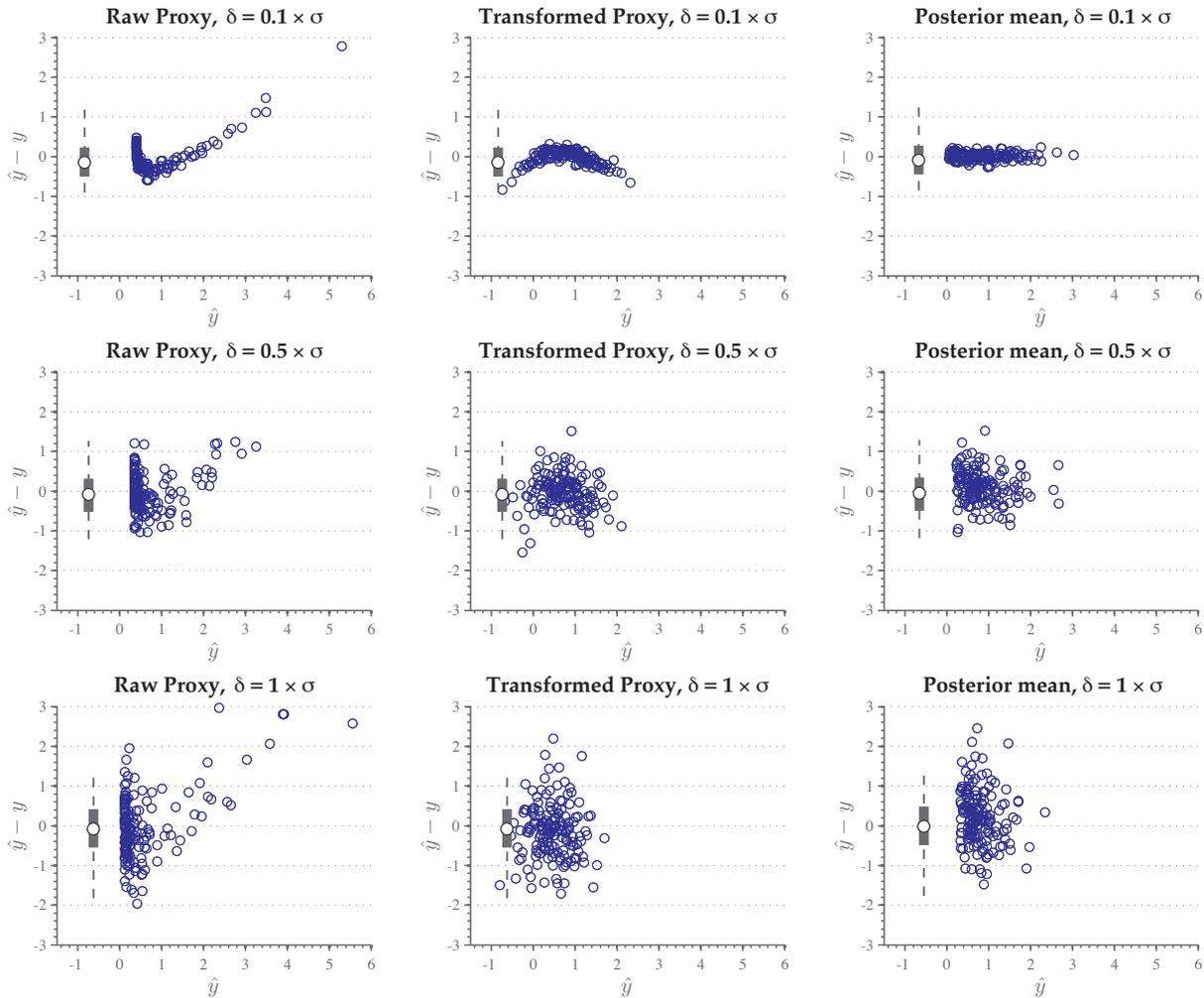


Figure 5. Residuals (differences between estimated and true values of the climate) as a function of the estimated climate. The three inference methods (columns) are compared for three values of δ (rows). Blue circles: residuals when $p > 0$; box plots: the distribution of residuals for $p = 0$. Whiskers extend to the 2.5 % and 97.5 % quantiles; boxes denote the 25 % and 75 % quantiles; circles mark the median. The middle row corresponds to the simulated proxy plotted in Fig. 4.

While coverage rates can always be increased by widening PIs, a useful probabilistic model should yield predictions that are both sharp and on point. A complementary view comes from the interval score (Gneiting and Raftery, 2007, Eq. 43), which balances coverage with sharpness. That is, the score rewards the narrowest intervals that encompass the true values at the nominal rate (here, 95 %) and should be as small as possible. Computed scores (Table 1, rows 5–6) are lowest for the BPM, whose slightly permissive coverage is more than compensated for by smaller widths; the scores are also the least variable for this method. At the opposite end, the raw regression displays the largest and most variable scores, while the ITS regression displays scores much closer to the BPM, and with comparable variability, particularly when data speak ($p > 0$).

The bias in the mean is quite small for all methods (rows 7–8) but is slightly more variable for the two regression methods (raw and transformed) than for the Bayesian model. The bias in the Bayesian inference for $p > 0$ results from the balance between the erfc term in Eq. (A5) giving greater weight to more positive values and the informative prior shrinking the posterior mean towards the prior mean (Eq. A6); given the particular parameters, the erfc terms has a stronger effect and there is a slight positive bias. More appropriately, assessing the Bayesian inference jointly for the $p = 0$ and $p > 0$ cases results in a bias that is not inconsistent with zero (0.05 with standard error 0.08). Changing the prior sharpness for the climate process has competing effects on the bias (and other diagnostics) in the $p = 0$ and $p > 0$ cases; these issues are investigated further in Appendix B.

Table 1. Inference diagnostics derived from 10 000 realizations of the climate process $c(t)$, for each of the three inference approaches, with $\delta = 0.5$. CI width is the width of the 95 % uncertainty intervals produced by each method (see Fig. 5) averaged over the 300 time points. The coverage rate should be compared to its nominal value of 95 %. The interval score should be as small as possible. The bias is defined as the difference between the estimated and true climate, averaged over the entire interval. All statistics have been stratified according the sign of p to account for the varying amounts of uncertainty in each case. For the top three rows, we report the mean and standard error (s.e.) of each distribution, obtained from the 10 000 realizations. The variance ratio statistics pertain to the distributions plotted in Fig. 6; variance ratios for $p = 0$ are not reported because all methods produce constant estimates in that case.

Quantity	Statistic	RAW			ITS			BPM		
		$p = 0$	$p > 0$	total	$p = 0$	$p > 0$	total	$p = 0$	$p > 0$	total
CI width (°C)	mean	2.71	2.07	2.39	2.71	1.84	2.28	2.74	1.60	2.17
	s.e.	0.33	0.21	0.21	0.33	0.17	0.20	0.00	0.01	0.08
Coverage rate	mean	0.94	0.94	0.94	0.94	0.94	0.94	0.95	0.91	0.93
	s.e.	0.05	0.04	0.03	0.05	0.04	0.03	0.03	0.04	0.02
Interval score	mean	03.51	2.99	3.27	3.51	2.34	2.94	3.27	2.32	2.79
	s.e.	0.82	1.33	0.81	0.82	0.40	0.49	0.50	0.38	0.34
Bias (°C)	mean	-0.00	+0.02	+0.01	-0.00	-0.01	-0.00	-0.01	+0.11	+0.05
	s.e.	0.19	0.12	0.12	0.19	0.11	0.12	0.14	0.07	0.08
Variance ratio	mode	–	0.40	–	–	0.96	–	–	0.96	–
	$\mathbb{P}(R > 1.5)$	–	0.29	–	–	0.07	–	–	0.07	–

Finally we summarize how closely each method estimates the variance ratio between the calibration ($t > 151$) and validation ($t \leq 150$) intervals (Table 1, rows 9–10). We calculate the ratio of sample variances between the two periods for each of 10 000 inferences of $c(t)$ using each method. We restrict our attention to the case where $p > 0$, as when $p = 0$ each method estimates $c(t)$ by a constant value and the variance of the corresponding climate cannot be estimated. To remove the effect of differences in sample variances calculated over the two intervals for realizations of the actual climate $c(t)$, we divide the ratio inferred from each pseudo-proxy by the ratio calculated from the corresponding realization of $c(t)$.

An ideal estimator of past climate should therefore result in a distribution of normalized variance ratios that is tightly distributed about unity, and indeed we find this to be the case for regression on the transformed proxies and for the Bayesian inversion (Fig. 6, blue and green curves; Table 1). In contrast, regression on the raw proxies results in a much broader distribution, peaked at about 0.4, with a much heavier upper tail (Fig. 6, cyan curve; Table 1). More tellingly, regression on the raw proxies results in a normalized variance ratio in excess of 1.5 in 29 % of experiments, more than 4 times as frequently as for the other two methods (Table 1, last row): in other words, regression on the raw proxies ups the risk of overestimating changes in variability by a factor of 4 as compared to the other methods.

Taken together, the results of our numerical experiments lead to a number of general observations about the potential pitfalls of ignoring nonlinearity in proxies, as well as the strengths and weaknesses of potential solutions. A failure to account for strong nonlinearity in a proxy often presents itself in structured residuals (Fig. 5) and incurs a high risk

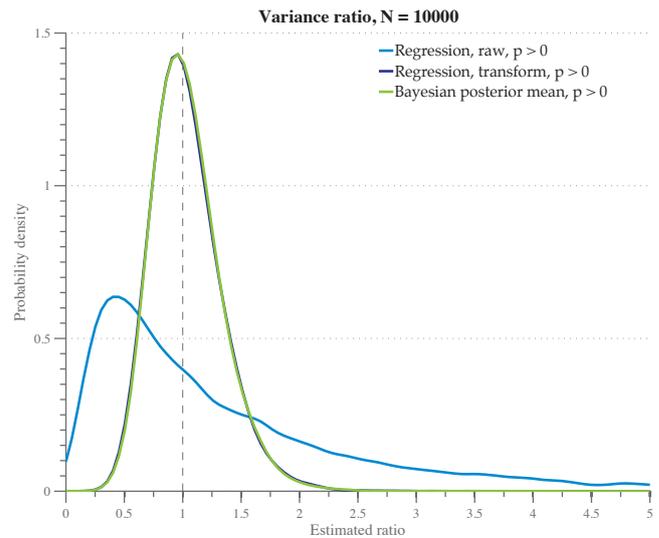


Figure 6. Distribution of variances ratios between the calibration and validation intervals, as inferred from each method and normalized by the same ratio calculated from the actual realization of $c(t)$. Cyan: raw regression; blue: regression on transformed pseudo-proxies; green: Bayesian inversion. The black dashed line denotes unity, which, due to the normalization, is the result under perfect inference for each realization of $c(t)$. Distributions are based on 10 000 realizations of $c(t)$ with $\delta = 0.5$.

of overstating changes in climate variability (Table 1, last row). Since palaeoclimate proxies are almost never considered in isolation, but (rightfully so) often stacked against comparable records, it is necessary that all proxies be approximately linearly related before interpreting discrepancies between records. This caveat is especially applicable to mul-

tiproxy analyses, as a climate reconstruction based on linear regression that includes such nonlinear proxy predictors would inherit the same potential pitfalls, though they may be attenuated by the presence of other sources of information.

Applying ITS to the nonlinear proxy record prior to regression analysis results in a much better fit according to the residual plots, as well as more faithful estimates of changes in variance. Such a transformed proxy series will thus be less likely to lead to spurious conclusions about past climate variability if included in a multiproxy reconstruction or a proxy intercomparison.

An important shortcoming of the regression methods used here is that inference is generally limited to a best-estimate of past climates and an uncertainty interval, whereas the Bayesian framework naturally provides the full distribution of climate conditional on the proxy observations (for further discussion see Tingley et al., 2012). However, we note that the excellent performance of the Bayesian inversion in all numerical experiments performed here is to be expected given that the inference model is constructed using the correct mechanistic model and parameter values. In practical applications, the parametric form of the mechanistic model would likely not be an exact representation of the data-generating mechanism, while the parameters themselves would need to be estimated from the data. This can be computationally demanding when no closed-form solution exists for the posterior distribution of model parameters (as in this case) and would naturally broaden the posterior distributions, i.e. lead to more uncertainties about the reconstructed climate.

4 Applications to published records

4.1 Laguna Pallcacocha

We now investigate how inferences made from the iconic Laguna Pallcacocha record of river run-off from the southern Ecuadorian Andes (Moy et al., 2002) are affected by different considerations of nonlinearity. The primary variable of interest here is the red colour intensity of the sediment, which is controlled by sediment delivery to the lake. Episodes of alluvial deposition are hypothesized to occur predominantly during El Niño events (Rodbell et al., 1999), similar to our idealized proxy (Sect. 2). Our intent here is not to dispute *qualitative* conclusions made from the Laguna Pallcacocha record – they turn out to be robust across the analyses we consider – but to show how nonlinearity may affect the *quantitative* inference made from such a proxy, and we view the analysis primarily as an illustrative example.

The time series of Pallcacocha red colour intensity displays pronounced positive skew (Fig. 7a), closely following a generalized extreme value distribution (GEV, Fig. 7b). In contrast, the distribution of December–January–February averages of NINO3 values is fairly close to a Gaussian, so using the Pallcacocha red colour record to infer changes in ENSO variability requires inferring a normally distributed climate

variable from a strongly non-normal proxy time series – as per the experiments in Sects. 2 and 3. Fitting a mechanistic model would necessitate determining both its structural form and parameters, so for simplicity we explore the consequences of an empirical transform (ITS). By design, applying the transform to the empirical distribution produces a nearly Gaussian series (Fig. 7d).

We investigate how the transform affects the time series of sample variances computed on disjoint, 100-year segments (Fig. 8). While the 100-year variance time series calculated from the raw series suggests an abrupt increase in ENSO-related variance at about 5000 years BP (grey dashed line), the series calculated from transformed values displays a more gradual increase. To establish if either or both series record significant increases in variance around 5000 years BP, we compute the variance ratio between two segments delineated by the grey line (centred at 3150 BCE), with lengths that vary between 200 and 3600 years. A window length of 200 years, for instance, means that we compare variances over the period 3350–3151 and 3150–2951 BCE. As the statistic of interest is a variance *ratio*, results are unchanged under *linear* transformation of the series: while accounting for the nonlinearity is necessary for meaningful inference, calibrating to climate units is not.

On account of the non-normal and serially correlated nature of both series, a standard F test for the significance of the observed variance ratio is not appropriate. Instead, we assess the significance of the variance ratio for each series via a block bootstrap resampling plan (Efron and Tibshirani, 1993; Kunsch, 1989), using a block length of 25 years (Table 2). That is, we randomly sample with replacement the original time series in 25-year blocks, compare variances before and after the year 3150, and do so $N = 2000$ times to obtain a null distribution, smoothed by kernel averaging (e.g. Wilks, 2011, Chap 4). The P value for each observed ratio (the probability of observing a ratio at least as large as that observed from chance alone) is then estimated from the bootstrap-derived null distributions. The results (Table 2) show two salient features: firstly, the variance ratio is generally larger with the raw than with the transformed series, as would be expected from Sects. 2 and 3. Secondly, the variance ratio estimated from the raw series is deemed significant at the 5 % level for all segment lengths, while the transformed series requires segment lengths of at least 1000 years to make this conclusion. Hence, with the raw series one would detect a significant shift at the boundary in a matter of 200 years, while when using the power-transformed series, this transition would be seen to occur much more gradually, taking at least 1000 years to become manifest. We note that the change in variance is largely driven by the record's accumulation rate (Rodó and Rodríguez-Arias, 2004), so we refrain from interpreting this aspect of the record.

Following Moy et al. (2002), we next apply Morlet wavelet analysis (Torrence and Compo, 1998), modified to account for energy conservation (Liu et al., 2007), to both the raw

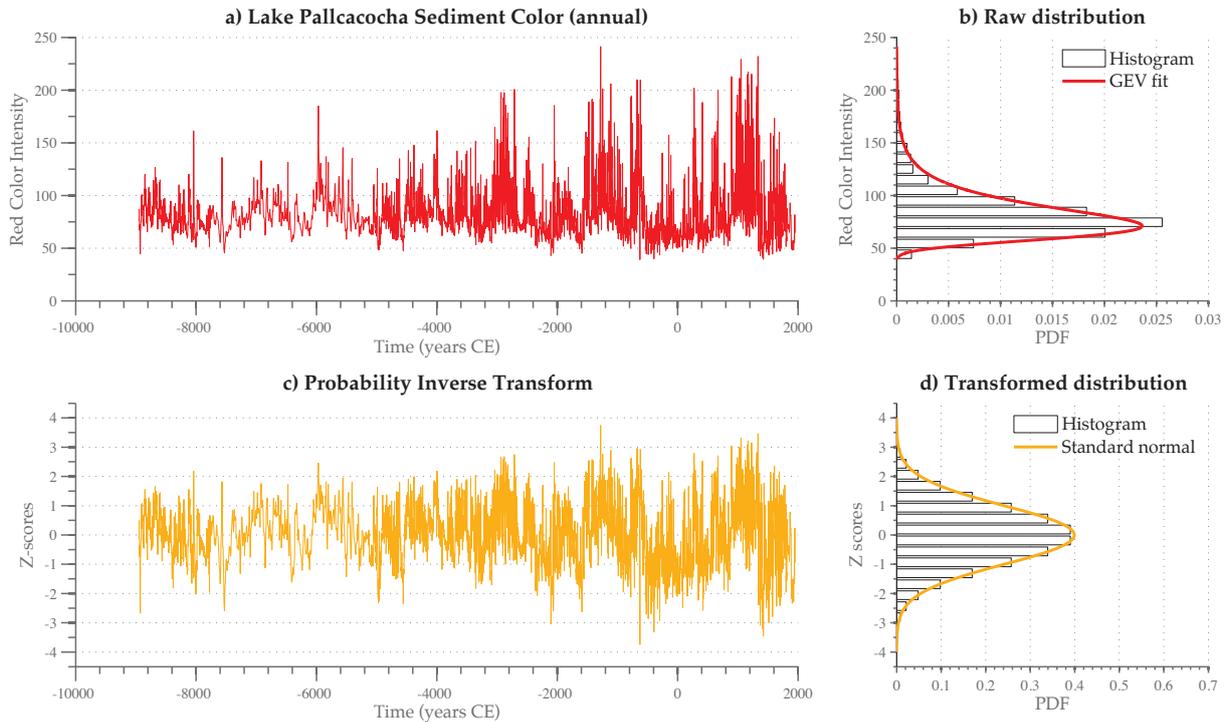


Figure 7. Raw and transformed time series of the Laguna Pallacocha record (Moy et al., 2002). **(a)** Raw time series, interpolated at 2-year intervals; **(b)** empirical histogram (bars) and parametric fit to a GEV distribution with parameters $(k_P, \sigma_P, \mu_P) = (0.07, 15.59, 72.13)$ (red line); **(c)** transformed time series with $\lambda = 0.31$; **(d)** empirical histogram of the transformed time series (bars) and parametric fit to a standard normal (red line).

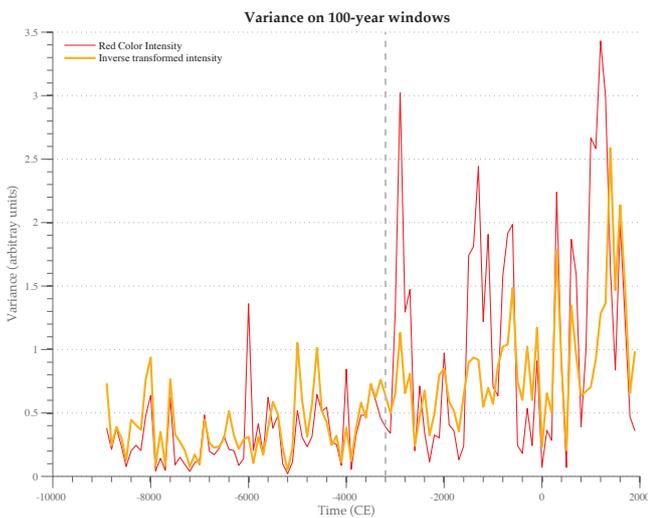


Figure 8. Variances calculated over non-overlapping 100-year segments of the Pallacocha red colour intensity record (thick cyan line) and after applying the empirical transform (thin blue line). Both series were standardized prior to analysis.

and transformed series. Results are qualitatively similar in both cases (Fig. 9): interannual variability is muted prior to the mid-Holocene, while millennial and centennial variabil-

ity is relatively persistent. However, the two wavelet transforms differ in detail, in particular regarding which “islands” of multidecadal power are significant with respect to a red noise benchmark (Torrence and Compo, 1998). In the latter 5000 years of the record, high-frequency variability appears more consistently significant with the empirical transform than without, indicating that the mid-Holocene shift for high-frequency variance is more consistently significant under the transform, even though the shift in total variance appears less substantial.

Our analysis does not contradict the qualitative conclusions of the original study by Moy et al. (2002). Indeed, the transformed series also suggests that ENSO events “become more frequent over the Holocene until about 1200 years ago, and then decline towards the present” (Moy et al., 2002). Wavelet analysis accounting for energy conservation (Liu et al., 2007) also suggests a statistically significant modulation in the 2000-year range, though it is nearly equal in amplitude to a ~ 500 yr broadband peak and lies mostly within the cone of influence. Nonetheless, results of the statistical analyses are markedly changed from a quantitative standpoint when nonlinearity is recognized and the proxy series transformed to approximate normality. Applying classical estimators to the raw series may lead to erroneous conclusions about the magnitude of changes in the variability of the un-

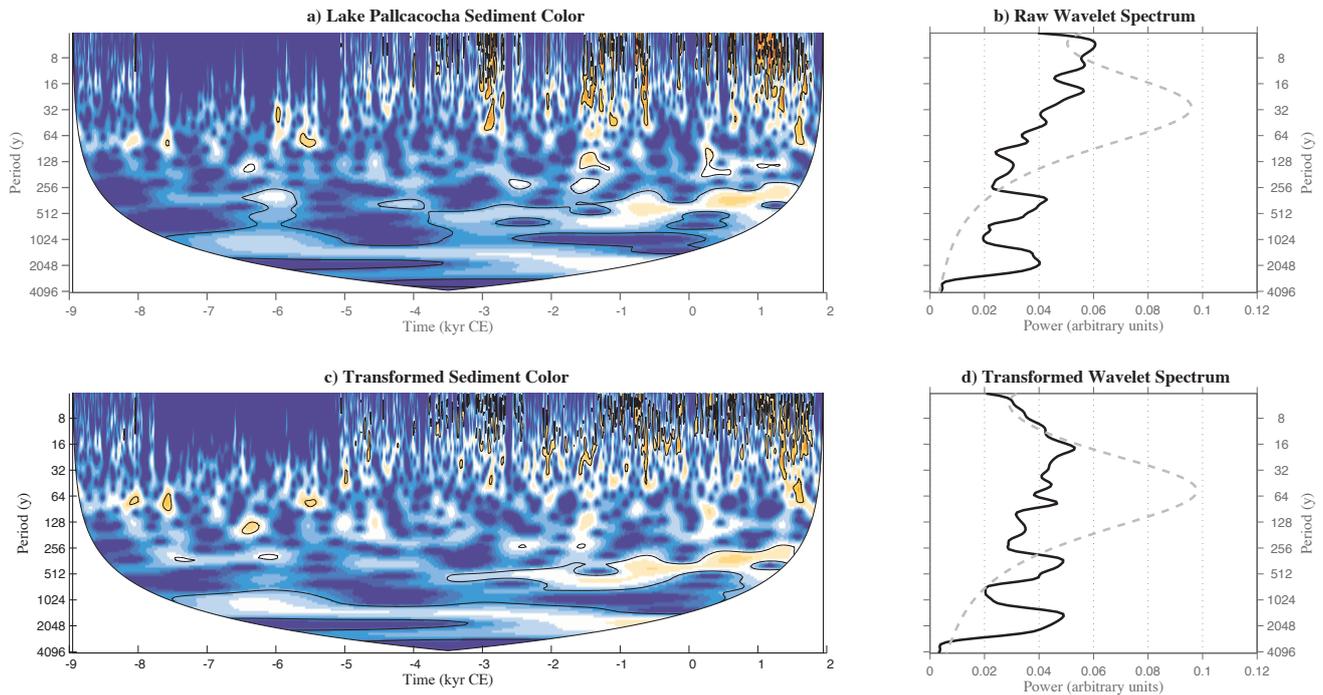


Figure 9. Wavelet analysis of the Laguna Pallacocha record (Moy et al., 2002). **(a)** Morlet wavelet coefficients for raw values of the red colour intensity series, interpolated at 2-year intervals; **(b)** global wavelet spectrum (black), with the AR(1) benchmark plotted in grey. **(c, d)** As in **(a, b)** but for the transformed time series. Colour bars are omitted since units are arbitrary.

derlying climate, as well as their localization in frequency space. The empirical transform counterbalances the nonlinearities inherent to run-off proxies such as the Pallacocha red colour intensity record, and the resulting transformed series is nearly normal and more likely to have an approximately linear relationship with climate (cf. the numerical examples of Sects. 2 and 3). One is therefore able to interpret changes in variability estimated from the transformed proxy record as indicating changes in climate variability (cf. Fig. 6).

4.2 Lake Challa

Recently, Wolff et al. (2011) used varved thickness at Lake Challa (equatorial east Africa) to retrace changes in ENSO activity since the Last Glacial Maximum. The proxy is interpreted as follows: El Niño (La Niña) events are associated with wetter (drier) conditions in east Africa and decreased (increased) surface wind speeds, which drive turbulent mixing in the lake, thereby bringing nutrients to the surface and boosting primary productivity. Thicker varves are therefore indicative of La Niña conditions, and thinner varves of El Niño conditions, resulting in a large anti-correlation ($r = -0.48$) between varve thickness and NINO3.4 SST (Wolff et al., 2011). The raw varve thickness data are plotted in Fig. 10a for the past 3000 years. The histogram is highly skewed and closely fits a log-normal distribution with parameters $(\mu, \sigma) = (0.33, 0.31)$ (Fig. 10b). The transformed

series (Fig. 10c) closely follows a standard normal distribution (Fig. 10d).

As the raw Lake Challa series has been used as a proxy for ENSO activity (Emile-Geay et al., 2013a, b), it is worth asking how much it resembles the Lake Pallacocha red colour intensity. Fig. 11 plots the two time series, as well as their standard deviations on sliding 40 yr windows, before and after transforming each record to normality. While Fig. 11c shows a fair amount of structure in each time series (peaks near 500 BCE and 2000 CE in Lake Challa varve thickness variability, quasi-periodic modulations of red colour intensity variability with an approximate recurrence time of 300 years), the transformed series (Fig. 11d) show different features: variability in red colour intensity now displays more pronounced late Holocene excursions than its untransformed counterpart, while variations in varve thickness now appear to recur at centennial, rather than millennial, scales.

This illustrative comparison is necessarily limited, as the proxies are located at either end of the Indo-Pacific system and are sensitive to different aspects of tropical Pacific SST (Challa responds more to central Pacific anomalies; Pallacocha responds more to variability along the Peruvian coast, though see Rodbell et al. (2008) for a non-climatic interpretation), which are not simply related to each other. In addition, the Pallacocha sensitivity to SST is highly nonlinear, as noted before, as is the Challa relationship: Wolff et al. (2011) note that “the varve thickness record is particularly sensitive

Table 2. Significance of variance changes across the 3150 BCE boundary (grey line in Fig. 8) in Laguna Pallcacocha red colour intensity. The table lists P values for a block bootstrap test (25-year blocks, $N = 2000$ realizations) of the null hypothesis that the variance jump across the boundary could have arisen from chance alone. The test is deemed significant at the α level whenever $P \leq \alpha$ and is conducted on time segments of variable length, to identify how long an interval is needed to detect a shift in variance. Results are virtually identical to 10-year blocks.

Statistic	Segment length (yr)	200	400	800	1000	1400	1800	3600
RAW	variance ratio	5.082	3.277	2.411	2.048	1.822	2.288	2.501
	P value	$< 10^{-3}$	$< 10^{-3}$	$< 10^{-3}$	$< 10^{-3}$	$< 10^{-3}$	$< 10^{-3}$	$< 10^{-3}$
ITS	variance ratio	1.275	1.182	1.155	1.304	1.359	1.360	1.797
	P value	0.199	0.218	0.118	0.012	0.003	$< 10^{-3}$	$< 10^{-3}$

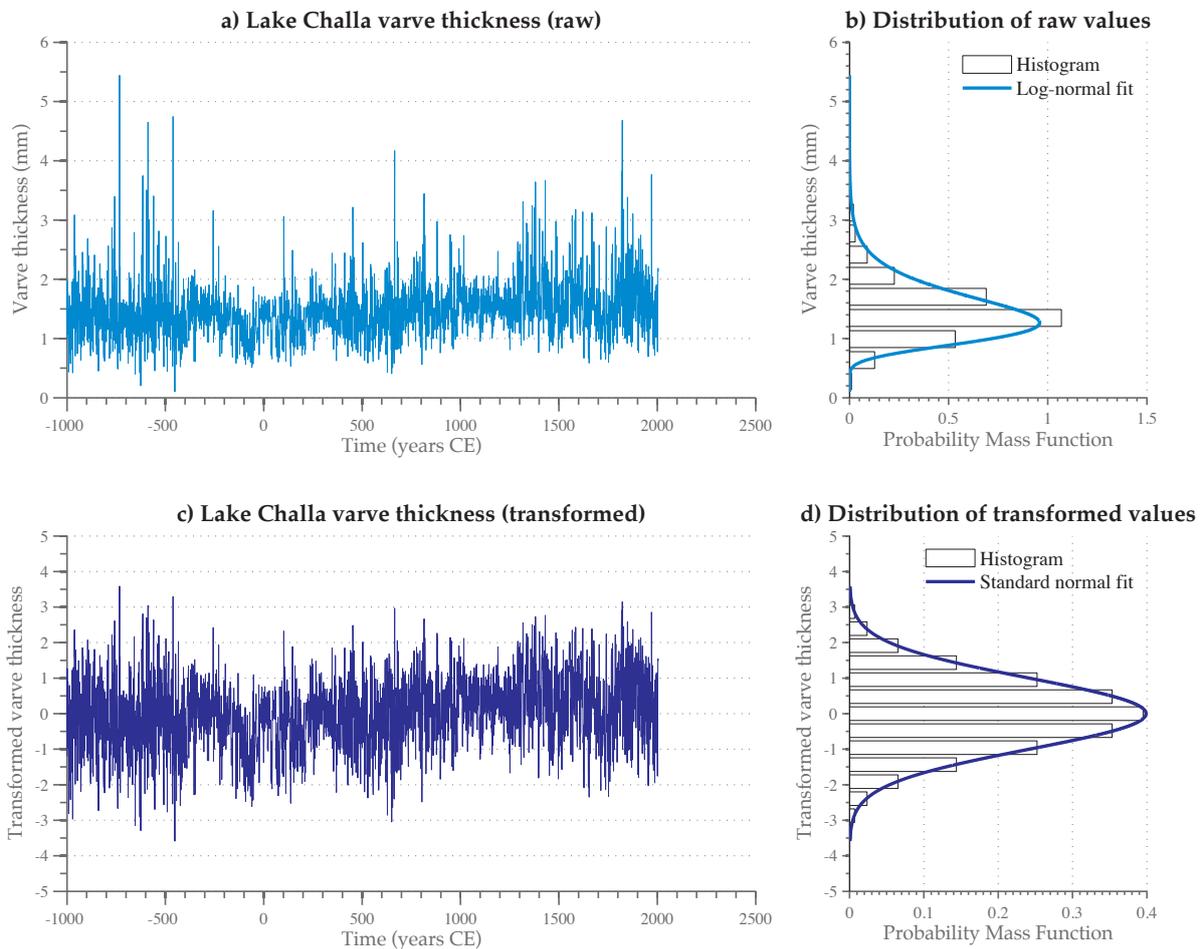


Figure 10. Raw and transformed time series of Lake Challa varve thickness (Wolff et al., 2011). **(a)** Raw time series, interpolated at yearly intervals; **(b)** empirical histogram (bars) and parametric fit to a log-normal distribution (cyan line); **(c)** transformed time series; **(d)** empirical histogram of the transformed time series (bars) and parametric fit to a standard normal (blue line).

to La Niña conditions and shows some evidence for saturation during El Niño years”. The main message is that the visual impression changes with the application of the transforms. Given that each proxy is non-normal due to a non-linear relationships with normally distributed SSTs, there is

a strong rationale for comparing the transformed rather than the untransformed series.

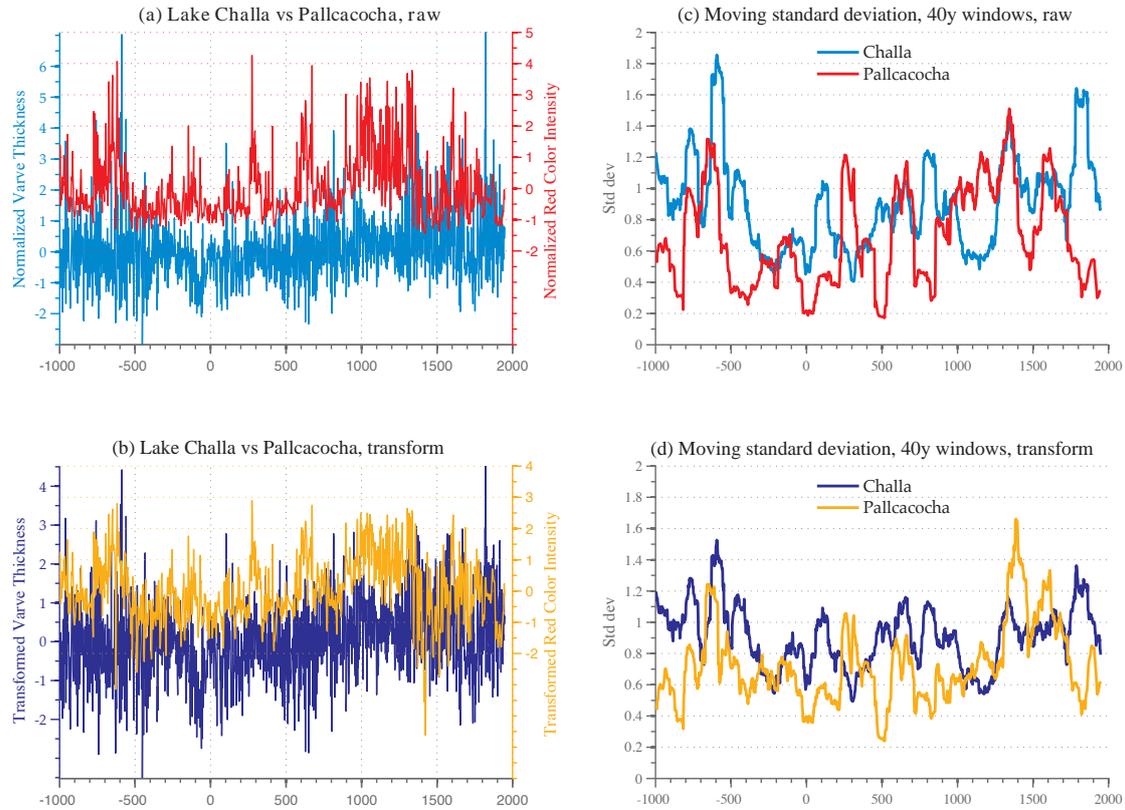


Figure 11. Comparison of Lake Challa and Laguna Pallcacocha records. Top row: raw (untransformed) records. Bottom row: transformed records. Left: time series interpolated at biannual resolution. Right: standard deviation on sliding 40 yr windows.

5 Discussion

Much of palaeoclimatology is concerned with the characterization of climate variability in the pre-instrumental era. Many proxy archives exhibit nonlinear relationships to the underlying climate, whereas many commonly used tools assume, either explicitly or implicitly, that the proxies are linearly related to a normally distributed climate variable. As illustrated for an idealized proxy (Sects. 2 and 3), and for ENSO-sensitive proxies (Sect. 4), direct variance estimates on different segments of nonlinear proxy records generally give a misleading picture of climate variability.

There are a number of techniques that allow for meaningful conclusions about changes in the underlying climate from such nonlinear proxy archives. Applying an empirical transform to such a series can render it approximately normal, so that standard data-analytic tools may be applied. In the context of the pseudo-proxy study (Sect. 2), the inverse transform sampling approach corrected for nonlinearity and reduced the risk of overstating variance changes. In the context of inferring climate values from proxy values (Sect. 3), the transform provided a regression fit comparable to that of an optimal benchmark and led to normally distributed resid-

uals, in accordance with the assumptions of the regression framework. These examples also highlighted the dangers of ignoring nonlinearity in the proxy–climate relationship.

Applying the transform to the Lake Pallcacocha record (Sect. 4.1) allows conclusions from a wavelet analysis and estimates of changes in variance to be interpreted in terms of climate, even without explicit knowledge of the mechanism responsible for the nonlinearity. Transformation allows for classical tools to be correctly applied to the analysis of palaeoclimate records, and it is a simpler approach than specifying and fitting a mechanistically informed statistical model. Transformation to normality may thus be seen as a pragmatic compromise between scientific and statistical rigour on the one hand and the need for practical solutions to analyse nonlinear proxy time series on the other. A limitation of the transform approach is that it is blind to the data-generating mechanism and may conflate various sources of nonlinearity.

We thus stress that generic recipes are no substitute for a mechanistic understanding, and ultimately the scientific interpretation of a proxy in terms of its climatic, ecological, or geological controls should guide the choice of statistical methodology. The example of Sect. 3 illustrates that explic-

itly modelling the mechanism giving rise to non-normality yields much better estimates of the underlying climate. In the example considered here, both the functional form and parameters of the mechanistic model were assumed known, while in real applications the form would likely be an approximation and the parameters would be estimated as part of a fully Bayesian analysis (Gelman et al., 2004; Tingley and Huybers, 2010a, b; Tolwinski-Ward et al., 2013), together with their associated uncertainties.

In instances where many proxies are used as predictors of past climates, such as climate reconstructions of the Common Era (e.g. Mann et al., 1998; Luterbacher et al., 2004; Moberg et al., 2005; Mann et al., 2008; Kaufman et al., 2009; Mann et al., 2009; Barriopedro et al., 2011; Emile-Geay et al., 2013b; Tingley and Huybers, 2013), individually modelling each proxy may prove challenging. Ideally, one would apply mechanistic forward models to each proxy (e.g. Dee et al., 2015) and combine them within a fully Bayesian inference procedure (Tingley et al., 2012). Until a full suite of validated models and computationally accessible Bayesian tools are available, however, empirical transforms provide a simple expedient to ensure that all input series fit the assumption of linear regression from a multivariate normal model that underpins most climate field reconstructions. We stress that such transforms are not a panacea, however, as they do not address other modelling challenges with proxy data, including time uncertainty (Anchukaitis and Tierney, 2012; Combour et al., 2014), the heterogeneous distribution of observations in space and time, spatial covariance modelling (Tingley and Huybers, 2010a; Guillot et al., 2015), and the differing spatial and temporal averaging inherent to the proxies (Tingley et al., 2012; Hanhijärvi et al., 2013).

Finally, while this article has focused on nonlinear proxy records with power-law type relationships, it is worth pointing out that a number of other valuable climate proxies may deviate from linearity in other ways. In particular, records based on proportions (e.g. pollen counts, lithological fractions, fractions of certain faunal assemblages), being in the range $[0, 1]$, also involve a nonlinear transform of Gaussian inputs like temperature and therefore require specific inference tools. Another key factor complicating inference from climate proxies is the existence of multiple influences on the measured variable, e.g. temperature and soil moisture controls on tree-ring width (e.g. Anchukaitis et al., 2006; Vaganov et al., 2006; Tolwinski-Ward et al., 2011; Evans et al., 2014) or temperature and seawater composition controls on the oxygen isotopic composition of biocarbonates (e.g. Thompson et al., 2011; Russon et al., 2013; Dee et al., 2015). We will explore solutions to these problems in future work.

6 Conclusions

In this article we showed that nonlinearity fundamentally alters the information content of climate proxies and that it must be dealt with, lest some erroneous inferences be made. Though we advocate for mechanistic modelling whenever possible, we showed that a simple empirical transform (ITS) can often be sufficient to remedy many of the ills of nonlinearity, and we recommend that palaeoclimatologists working with nonlinear proxies adopt such simple transforms in their work. Matlab code for implementing the transform is available at <https://github.com/CommonClimate/common-climate/blob/master/gaussianize.m>.

Appendix A: Bayesian model derivation

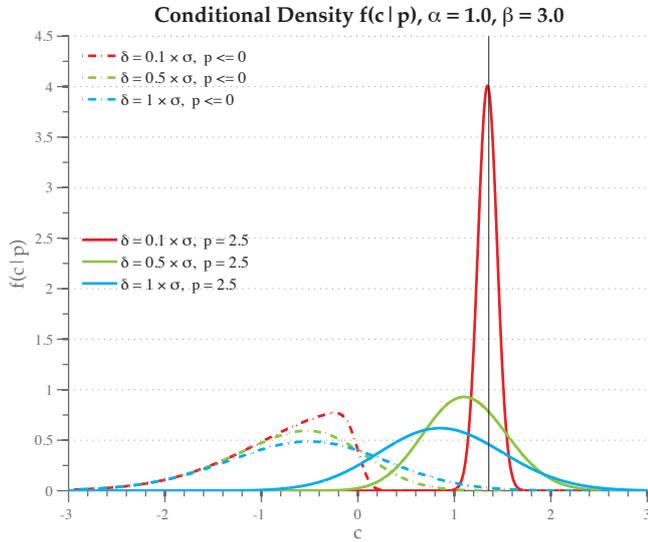


Figure A1. Posterior probability density functions of climate, according to Eqs. (6, 7), with $\mu = 0$, $\sigma = 1$, $\alpha = 1$, and $\beta = 3$ and the noise level δ taking on the values $[0.1, 0.5, 1] \times \sigma$. Solid and dashed lines correspond, respectively, to proxy values of $p = 2.5$ and $p = 0$. The black vertical line marks the point $(p/\alpha)^{1/\beta}$, the true value of c in the absence of noise.

Consider the proxy model of Eq. (1) and assume for simplicity that both c and ϵ are normally distributed, with $c \sim \mathcal{N}(\mu, \sigma^2)$ and $\epsilon \sim \mathcal{N}(0, \delta^2)$.

We are interested in inferring $c(t)$ from $p(t)$ and, more specifically, in finding the posterior probability distribution of climate given the proxy, $f(c|p)$. Using Bayes' theorem, the posterior distribution $f(c|p)$ can be re-expressed as a function of the distribution of the proxy conditional on climate, $f(p|c)$, and the prior distribution of climate states, $\pi(c)$:

$$f(c|p) \propto f(p|c)\pi(c). \tag{A1}$$

In what follows, we set the prior on the climate states, $\pi(c)$, to follow the same distribution as the process used to generate the climate itself. The prior is therefore informative and unbiased, and the posterior will be a mix of the information supplied by the prior and information from the data. As the prior distribution and the distribution of climate itself are the same in this illustrative example, in the absence of any information from the data, the posterior distribution of the climate will converge to the distribution used to generate climate, $\mathcal{N}(\mu, \sigma^2)$.

On account of the thresholding behaviour of the proxy, it is necessary to distinguish two cases in calculating the posterior distribution of climate.

– Case 1:

$$\begin{aligned} p(t) &= 0 \\ f(c|p = 0) &\propto \mathbb{P}(p = 0|c)\pi(c) \propto \mathbb{P}(c' < 0|c)\pi(c) \\ f(c|p = 0) &\propto \mathbb{P}(c + \epsilon < 0|c)\pi(c) \propto \mathbb{P}(\epsilon < -c|c)\pi(c) \\ f(c|p = 0) &\propto \int_{-\infty}^{-c} e^{-\frac{\xi^2}{2\delta^2}} d\xi \cdot e^{-\frac{(c-\mu)^2}{2\sigma^2}} \end{aligned} \tag{A2}$$

With some rearrangements, the integral may be rewritten with the help of erfc, the complementary error function¹:

$$f(c|p = 0) \propto \text{erfc}\left(\frac{c}{\delta\sqrt{2}}\right) e^{-\frac{(c-\mu)^2}{2\sigma^2}}, \tag{A3}$$

and the normalization constant can then be calculated numerically. The posterior is the product of two terms. The first gives the likelihood that $p = 0$ given any value of c , while the second is the prior on c . Note that the likelihood increases as c decreases, and in the absence of the prior the uncertainty interval for c would be unbounded from below. The prior distribution thus provides important regularization and allows for uncertainty estimates with finite width when $p = 0$.

– Case 2:

$$\begin{aligned} p(t) &> 0 \\ f(c|p > 0) &\propto \mathbb{P}(p|p > 0, c)\mathbb{P}(p > 0|c)\pi(c) \\ f(c|p > 0) &\propto \mathbb{P}(p = \alpha c'^{1/\beta}|c)\mathbb{P}(p > 0|c)\pi(c) \\ f(c|p > 0) &\propto \mathbb{P}\left(\epsilon = \left(\frac{p}{\alpha}\right)^{1/\beta} - c\right) \mathbb{P}(p > 0|c)\pi(c) \\ f(c|p > 0) &\propto e^{-\frac{1}{2}\left(\frac{(p/\alpha)^{1/\beta} - c}{\delta}\right)^2} \\ &\cdot \mathbb{P}(c + \epsilon > 0|c) \cdot e^{-\frac{1}{2}\left(\frac{c-\mu}{\sigma}\right)^2} \\ f(c|p > 0) &\propto e^{-\frac{1}{2}\left(\frac{(p/\alpha)^{1/\beta} - c}{\delta}\right)^2} e^{-\frac{1}{2}\left(\frac{c-\mu}{\sigma}\right)^2} \cdot \mathbb{P}(\epsilon > -c|c) \\ f(c|p > 0) &\propto e^{-\frac{1}{2}\left(\frac{c-\mu'}{\sigma'}\right)^2} \cdot \int_{-c}^{\infty} e^{-\frac{\xi^2}{2\delta^2}} d\xi \\ f(c|p > 0) &\propto e^{-\frac{1}{2}\left(\frac{c-\mu'}{\sigma'}\right)^2} \cdot \text{erfc}\left(\frac{-c}{\delta\sqrt{2}}\right) \end{aligned} \tag{A4}$$

where

$$\mu' = \frac{\mu\delta^2 + (p/\alpha)^{1/\beta}\sigma^2}{\delta^2 + \sigma^2} \text{ and } \sigma'^2 = \frac{\delta^2\sigma^2}{\delta^2 + \sigma^2} \tag{A6}$$

¹erfc(x) = $\frac{2}{\sqrt{\pi}} \int_x^{\infty} e^{-u^2} du$

The posterior is once more the product of two terms. The first is a normal density that combines the information from the prior and the likelihood of the observed value of p . As is common in such settings (e.g. Gelman et al., 2004, chap 2), the mean of this normal distribution is an inverse variance-weighted mean of the prior mean and the transformed proxy value, while the variance is the harmonic mean of the prior and noise variances. The second term in the posterior gives the likelihood of observing a non-zero value of p , conditional on the value of the climate.

Examples of the posterior distribution are shown in Fig. A1, for different relative values of δ and σ . In the limit of no noise ($\delta \rightarrow 0$), the erfc terms converge to step functions at zero and ensure that, in the $p = 0$ case, the distribution is $\pi(c)$ for $c \leq 0$ and 0 otherwise – that is, the negative half of a scaled normal distribution. Similarly, in the $p > 0$ case, the posterior distribution converges to a delta function centred at $(p/\alpha)^{\frac{1}{\beta}}$, as the variance in Eq. (A6) is zero. In the case $\delta = \sigma$, the distributions for both $p > 0$ and $p = 0$ are much flatter and display substantial overlap, indicating the influence of the informative prior. Finally, in the limit as $\delta \rightarrow \infty$, the posterior converges to the prior $\pi(c)$ and the posterior distribution is the same in both the $p > 0$ and $p = 0$ cases.

Appendix B: Alternate prior choices

The choice of prior affects both point estimates and uncertainty intervals and is the dominant control on the inference in data-poor situations – such as the $p = 0$ case for the example discussed in the text. Here we explore, both qualitatively and quantitatively, the influence of the prior standard deviation, σ_p , for the climate process on the inference of climate conditional on proxy values.

To build intuition into the role of the prior, consider the case where $\sigma_p \rightarrow 0$, corresponding to perfect prior information. Provided that $\delta > 0$, the posteriors for both $p = 0$ and $p > 0$ then converge to delta functions at μ , and the data have no influence on the posterior. In contrast, as $\sigma_p \rightarrow \infty$, corresponding to an uninformative prior, the posterior for $p > 0$ has the same form as for finite σ_p , but now the mean and variance (Eq. A6) are, respectively, $(p/\alpha)^{(1/\beta)}$ and δ^2 . For the $p = 0$ case, Eq. (A3) then converges to something proportional to erfc, which is not a proper probability distribution as its integral on the real line is infinite. An informative prior therefore plays two roles: providing regularization when $p = 0$ so that uncertainty widths are finite, and shrinking the posterior mean towards the prior mean when $p > 0$, while likewise reducing the posterior variance.

Table B1 presents the same statistics as Table 1, but for the three following situations: (i) Bayesian posterior distribution with $\sigma_p = \sigma$ (true standard deviation); (ii) DBL: $\sigma_p = 2 \times \sigma$; and (iii) HLF: $\sigma_p = \frac{1}{2} \times \sigma$. The wider (DBL) prior results in substantial negative bias for $p = 0$, as the lower tail of the posterior expands towards more negative values. For the $p > 0$ case, the wider prior results in a positive bias, as there is more posterior mass farther from zero and insufficient shrinkage towards the true climate mean. With more weight given to the observed value of p , rather than the prior, the posterior approaches a normal with mean $(p/\alpha)^{(1/\beta)}$ and variance δ^2 – despite the fact that a particularly large observed p is more likely due to the error term than the climate. The situation is reversed for the HLF prior, with positive bias when $p = 0$ and negative bias when $p > 0$.

Although the uncertainty intervals are wider under the DBL as compared with the CTL prior, the coverage rates are largely unchanged between the two. There is a trade-off in this case, as the wider prior results in an increased probability of small magnitude values of c not being covered by the intervals, and a decreased probability that large magnitude values of c fall outside the intervals. For the HLF prior, the uncertainty intervals are narrower and coverage rates lower as compared with the CTL prior. For both the DBL and HLF, the interval scores are larger than for CTL, indicating that, of the three priors, the latter results in the best compromise between coverage rate and interval width.

Table B1. Inference diagnostics derived from 1000 realizations of the climate process $c(t)$, for three values of the prior standard deviation σ_p , with $\delta = 0.5$. CTL: $\sigma_p = \sigma$ (true standard deviation); DBL: $\sigma_p = 2 \times \sigma$; HLF: $\sigma_p = \frac{1}{2} \times \sigma$.

Quantity	Statistic	CTL			DBL			HLF		
		$p = 0$	$p > 0$	total	$p = 0$	$p > 0$	total	$p = 0$	$p > 0$	total
CI width	mean	2.74	1.60	2.17	4.86	1.73	3.29	1.62	1.28	1.45
	s.e.	0.00	0.01	0.08	0.00	0.02	0.23	0.00	0.01	0.02
Coverage rate	mean	0.95	0.91	0.93	0.96	0.90	0.93	0.72	0.77	0.74
	s.e.	0.03	0.04	0.02	0.03	0.04	0.02	0.07	0.06	0.05
Interval score	mean	3.29	2.32	2.80	5.24	2.52	3.86	6.99	3.65	5.39
	s.e.	0.52	0.39	0.35	0.30	0.41	0.31	2.36	0.98	1.39
Bias	mean	-0.01	+0.11	+0.05	-0.84	+0.24	-0.30	+0.42	-0.17	+0.13
	s.e.	0.14	0.07	0.08	0.14	0.07	0.07	0.14	0.09	0.12
Variance ratio	mode	-	0.96	-	-	0.96	-	-	0.96	-
	$\mathbb{P}(R > 1.5)$	-	0.07	-	-	0.07	-	-	0.06	-

Acknowledgements. J. Emile-Geay and M. Tingley are grateful to Peter Craigmile and Carl Wunsch for comments on an earlier version of this manuscript. J. Emile-Geay acknowledges funding from NSF grant DMS-1025464. The code and data necessary to reproduce the results of this study, including figures and tables, will be made available at <http://climdyn.usc.edu/Publications.html> upon publication.

Edited by: V. Rath

References

- Anchukaitis, K. J. and Tierney, J. E.: Identifying coherent spatiotemporal modes in time-uncertain proxy paleoclimate records, *Clim. Dyn.*, 41, 1291–1306, doi:10.1007/s00382-012-1483-0, 2012.
- Anchukaitis, K. J., Evans, M. N., Kaplan, A., Vaganov, E. A., Hughes, M. K., Grissino-Mayer, H. D., and Cane, M. A.: Forward modeling of regional scale tree-ring patterns in the southeastern United States and the recent influence of summer drought, *Geophys. Res. Lett.*, 33, L04705, doi:10.1029/2005GL025050, 2006.
- Baker, A., Bradley, C., Phipps, S. J., Fischer, M., Fairchild, I. J., Fuller, L., Spötl, C., and Azcurra, C.: Millennial-length forward models and pseudoproxies of stalagmite $\delta^{18}\text{O}$: an example from NW Scotland, *Clim. Past*, 8, 1153–1167, doi:10.5194/cp-8-1153-2012, 2012.
- Barriopedro, D., Fischer, E. M., Luterbacher, J., Trigo, R. M., and García-Herrera, R.: The Hot Summer of 2010: Redrawing the Temperature Record Map of Europe, *Science*, 332, 220–224, doi:10.1126/science.1201224, 2011.
- Bickel, P. J. and Doksum, K. A.: An Analysis of Transformations Revisited, *J. Am. Stat. Assoc.*, 76, 296–311, doi:10.1080/01621459.1981.10477649, 1981.
- Box, G. and Cox, D.: An analysis of transformations, *J. Roy. Stat. Soc., Ser. B*, 26, 211–252, 1964.
- Brown, P. J.: Measurement, Regression, and Calibration, vol. 12 of Oxford Statistical Science Series, Oxford University Press, USA, 216 pp., 1994.
- Christiansen, B.: Straight Line Fitting and Predictions: On a Marginal Likelihood Approach to Linear Regression and Errors-In-Variables Models, *J. Climate*, 27, 2014–2031, doi:10.1175/JCLI-D-13-00299.1, 2014.
- Comboul, M., Emile-Geay, J., Evans, M. N., Mirnategui, N., Cobb, K. M., and Thompson, D. M.: A probabilistic model of chronological errors in layer-counted climate proxies: applications to annually banded coral archives, *Clim. Past*, 10, 825–841, doi:10.5194/cp-10-825-2014, 2014.
- Conroy, J. L., Overpeck, J. T., Cole, J. E., Shanahan, T. M., and Steinitz-Kannan, M.: Holocene changes in eastern tropical Pacific climate inferred from a Galápagos lake sediment record, *Quaternary Sci. Rev.*, 27, 1166–1180, doi:10.1016/j.quascirev.2008.02.015, 2008.
- Conroy, J. L., Restrepo, A., Overpeck, J. T., Steinitz-Kannan, M., Cole, J. E., Bush, M. B., and Colinvaux, P. A.: Unprecedented recent warming of surface temperatures in the eastern tropical Pacific Ocean, *Nat. Geosci.*, 2, 46–50, doi:10.1038/ngeo390, 2009.
- Dee, S., Emile-Geay, J., Evans, M. N., Allam, A., Steig, E. J., and Thompson, D. M.: PRYSM: An open-source framework for PROXY System Modeling, with applications to oxygen-isotope systems, *Journal of Advances in Modeling Earth Systems*, 7, 1220–1247, doi:10.1002/2015MS000447, 2015.
- Efron, B. and Tibshirani, R. J.: *An Introduction to the Bootstrap*, Chapman & Hall, New York, 456 pp., 1993.
- Emile-Geay, J., Cobb, K., Mann, M., and Wittenberg, A. T.: Estimating Central Equatorial Pacific SST variability over the Past Millennium. Part 1: Methodology and Validation, *J. Clim.*, 26, 2302–2328, doi:10.1175/JCLI-D-11-00510.1, 2013a.
- Emile-Geay, J., Cobb, K., Mann, M., and Wittenberg, A. T.: Estimating Central Equatorial Pacific SST variability over the Past Millennium. Part 2: Reconstructions and Implications, *J. Clim.*, 26, 2329–2352, doi:10.1175/JCLI-D-11-00511.1, 2013b.
- Evans, M. N., Tolwinski-Ward, S. E., Thompson, D. M., and Anchukaitis, K. J.: Applications of proxy system modeling in high resolution paleoclimatology, *Quaternary Sci. Rev.*, 76, 16–28, doi:10.1016/j.quascirev.2013.05.024, 2013.
- Evans, M. N., Smerdon, J. E., Kaplan, A., Tolwinski-Ward, S. E., and Gonzalez-Rouco, J. F.: Climate field reconstruction uncertainty arising from multivariate and nonlinear properties of predictors, *Geophys. Res. Lett.*, 41, 9127–9134, doi:10.1002/2014GL062063, 2014.
- Gelman, A., Carlin, J. B., Stern, H. S., and Rubin, D. B.: *Bayesian Data Analysis*, Chapman and Hall, New York, NY, 2nd edn., 675 pp., 2004.
- Gneiting, T. and Raftery, A. E.: Strictly Proper Scoring Rules, Prediction, and Estimation, *J. Am. Stat. Assoc.*, 102, 359–378, doi:10.1198/016214506000001437, 2007.
- Guillot, D., Rajaratnam, B., and Emile-Geay, J.: Statistical paleoclimate reconstructions via Markov random fields, *Ann. Applied. Statist.*, 324–352, doi:10.1214/14-AOAS794, 2015.
- Guttman, I.: *Statistical tolerance regions: classical and Bayesian*, vol. 26, Griffin London, 160 pp., 1970.
- Hanhijärvi, S., Tingley, M., and Korhola, A.: Pairwise comparisons to reconstruct mean temperature in the Arctic Atlantic Region over the last 2,000 years, *Clim. Dynam.*, 1–22, doi:10.1007/s00382-013-1701-4, 2013.
- Jaynes, E. T.: Confidence Intervals vs. Bayesian Intervals, in: *Foundations of Probability Theory, Statistical Inference, and Statistical Theories of Science*, edited by: Harper, W. L. and Hooker, C. A., 175–257, D. Reidel, Dordrecht, 1976.
- Jex, C. N., Phipps, S. J., Baker, A., and Bradley, C.: Reducing uncertainty in the climatic interpretations of speleothem $\delta^{18}\text{O}$, *Geophys. Res. Lett.*, 40, 2259–2264, doi:10.1002/grl.50467, 2013.
- Kaufman, D. S., Schneider, D. P., McKay, N. P., Ammann, C. M., Bradley, R. S., Briffa, K. R., Miller, G. H., Otto-Bliesner, B. L., Overpeck, J. T., Vinther, B. M., and Arctic Lakes 2k Project Members: Recent Warming Reverses Long-Term Arctic Cooling, *Science*, 325, 1236–1239, doi:10.1126/science.1173983, 2009.
- Kirby, M. E., Feakins, S. J., Hiner, C. A., Fantozzi, J., Zimmerman, S. R. H., Dingemans, T., and Mensing, S. A.: Tropical Pacific forcing of Late-Holocene hydrologic variability in the coastal southwest United States, *Quaternary Sci. Rev.*, 102, 27–38, doi:10.1016/j.quascirev.2014.08.005, 2014.
- Kunsch, H. R.: The jackknife and the bootstrap for general stationary observations, *The Annals of Statistics*, 17, 1217–1241, 1989.

- Lengaigne, M. and Vecchi, G. A.: Contrasting the termination of moderate and extreme El Niño events in coupled general circulation models, *Clim. Dyn.*, 35, 299–313, doi:10.1007/s00382-009-0562-3, 2010.
- Lindsey, J. K.: Applying generalized linear models, Springer texts in statistics, Springer, New York, 256 pp., 1997.
- Liu, Y., San Liang, X., and Weisberg, R. H.: Rectification of the Bias in the Wavelet Power Spectrum, *J. Atmos. Ocean. Tech.*, 24, 2093–2102, doi:10.1175/2007JTECHO511.1, 2007.
- Luterbacher, J., Dietrich, D., Xoplaki, E., Grosjean, M., and Wanner, H.: European Seasonal and Annual Temperature Variability, Trends, and Extremes Since 1500, *Science*, 303, 1499–1503, doi:10.1126/science.1093877, 2004.
- Maidment, D. R. (Ed.): Handbook of Hydrology, McGraw-Hill, New York, NY, USA, 1424 pp., 1993.
- Mann, M. E., Bradley, R. S., and Hughes, M. K.: Global-scale temperature patterns and climate forcing over the past six centuries, *Nature*, 392, 779–787, doi:10.1038/33859, 1998.
- Mann, M. E., Zhang, Z., Hughes, M. K., Bradley, R. S., Miller, S. K., Rutherford, S., and Ni, F.: Proxy-based reconstructions of hemispheric and global surface temperature variations over the past two millennia, *Proceedings of the National Academy of Sciences*, 105, 13252–13257, doi:10.1073/pnas.0805721105, 2008.
- Mann, M. E., Zhang, Z., Rutherford, S., Bradley, R. S., Hughes, M. K., Shindell, D., Ammann, C., Faluvegi, G., and Ni, F.: Global Signatures and Dynamical Origins of the Little Ice Age and Medieval Climate Anomaly, *Science*, 326, 1256–1260, doi:10.1126/science.1177303, 2009.
- McCullagh, P. and Nelder, J. A.: Generalized linear models (Second edition), London: Chapman & Hall, 532 pp., 1989.
- McKee, T. B., Doeskin, N. J., and Kleist, J.: The Relationship of Drought Frequency and Duration to Time Scales, in: Proc. 8th Conf. on Applied Climatology, 179–184, American Meteorological Society, Boston, Massachusetts, 1993.
- McKee, T. B., Doeskin, N. J., and Kleist, J.: Drought Monitoring with Multiple Time Scales, in: Proc. 9th Conf. on Applied Climatology, 15–20 January 1995, 233–236, American Meteorological Society, Boston, Massachusetts, 1995.
- Moberg, A., Sonechkin, D. M., Holmgren, K., Datsenko, N. M., and Karlen, W.: Highly variable Northern Hemisphere temperatures reconstructed from low and high-resolution proxy data, *Nature*, 433, 613–617, 2005.
- Moy, C., Seltzer, G., Rodbell, D., and Anderson, D.: Variability of El Niño/Southern Oscillation activity at millennial timescales during the Holocene epoch, *Nature*, 420, 162–165, 2002.
- Nelder, J. A. and Wedderburn, R. W. M.: Generalized Linear Models, *J. R. Stat. Soc. Ser. A-G.*, 135, 370–384, doi:10.2307/2344614, 1972.
- Rein, B., Lückge, A., and Sirocko, F.: A major Holocene ENSO anomaly during the Medieval period, *Geophys. Res. Lett.*, 31, L17211, doi:10.1029/2004GL020161, 2004.
- Rein, B., Lückge, A., Reinhardt, L., Sirocko, F., Wolf, A., and Dullo, W.-C.: El Niño variability off Peru during the last 20,000 years, *Paleoceanography*, 20, PA4003, doi:10.1029/2004PA001099, 2005.
- Rodbell, D. T., Seltzer, G. O., Anderson, D. M., Abbott, M. B., Enfield, D. B., and Newman, J. H.: An 15,000-Year Record of El Niño-Driven Alluviation in Southwestern Ecuador, *Science*, 283, 516–520, doi:10.1126/science.283.5401.516, 1999.
- Rodbell, D. T., Seltzer, G. O., Mark, B. G., Smith, J. A., and Abbott, M. B.: Clastic sediment flux to tropical Andean lakes: records of glaciation and soil erosion, *Quaternary Sci. Rev.*, 27, 1612–1626, doi:10.1016/j.quascirev.2008.06.004, 2008.
- Rodó, X. and Rodríguez-Arias, M.-A.: El Niño Southern Oscillation: Absent in the Early Holocene?, *J. Climate*, 17, 423–426, doi:10.1175/1520-0442(2004)017<0423:ENOAIT>2.0.CO;2, 2004.
- Rose, C.: Erosion and Sedimentation, in: Hydrology and Water Management in the Humid Tropics, edited by Bonell, M., Hufschmidt, M. M., and Gladwell, J. S., International Hydrology Series, chap. 14, pp. 301–343, Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, http://dx.doi.org/10.1017/CBO9780511564468.018, 1993.
- Russon, T., Tudhope, A. W., Hegerl, G. C., Collins, M., and Tindall, J.: Inter-annual tropical Pacific climate variability in an isotope-enabled CGCM: implications for interpreting coral stable oxygen isotope records of ENSO, *Clim. Past*, 9, 1543–1557, doi:10.5194/cp-9-1543-2013, 2013.
- Smith, T., Reynolds, R., Peterson, T. C., and Lawrimore, J.: Improvements to NOAA's Historical Merged Land-Ocean Surface Temperature Analysis (1880–2006), *J. Clim.*, 21, 2283–2296, 2008.
- Thompson, D. M., Ault, T. R., Evans, M. N., Cole, J. E., and Emile-Geay, J.: Comparison of observed and simulated tropical climate trends using a forward model of coral $\delta^{18}\text{O}$, *Geophys. Res. Lett.*, 38, L14706, doi:10.1029/2011GL048224, 2011.
- Tingley, M. P. and Huybers, P.: A Bayesian Algorithm for Reconstructing Climate Anomalies in Space and Time. Part 1: Development and applications to paleoclimate reconstruction problems, *J. Clim.*, 23, 2759–2781, doi:10.1175/2009JCLI3016.1, 2010a.
- Tingley, M. P. and Huybers, P.: A Bayesian Algorithm for Reconstructing Climate Anomalies in Space and Time. Part 2: Comparison with the Regularized Expectation-Maximization Algorithm, *J. Clim.*, 23, 2782–2800, doi:2009JCLI3016.1, 2010b.
- Tingley, M. P. and Huybers, P.: Recent temperature extremes at high northern latitudes unprecedented in the past 600 years, *Nature*, 496, 201–205, doi:10.1038/nature11969, 2013.
- Tingley, M. P., Craigmile, P. F., Haran, M., Li, B., Mannshardt, E., and Rajaratnam, B.: Piecing together the past: statistical insights into paleoclimatic reconstructions, *Quaternary Sci. Rev.*, 35, 1–22, doi:10.1016/j.quascirev.2012.01.012, 2012.
- Tolwinski-Ward, S. E., Evans, M. N., Hughes, M. K., and Anchukaitis, K. J.: An efficient forward model of the climate controls on interannual variation in tree-ring width, *Clim. Dynam.*, 36, 2419–2439, doi:10.1007/s00382-010-0945-5, 2011.
- Tolwinski-Ward, S. E., Anchukaitis, K. J., and Evans, M. N.: Bayesian parameter estimation and interpretation for an intermediate model of tree-ring width, *Clim. Past*, 9, 1481–1493, doi:10.5194/cp-9-1481-2013, 2013.
- Torrence, C. and Compo, G. P.: A Practical Guide to Wavelet Analysis, *B. Am. Meteor. Soc.*, 79, 61–78, 1998.
- Vaganov, E. A., Hughes, M. K., and Shashkin, A. V.: Growth dynamics of conifer tree rings, vol. 183, Springer-Verlag, New York, NY, 358 pp., 2006.
- van Albada, S. and Robinson, P.: Transformation of arbitrary distributions to the normal distribution with application to EEG test-retest reliability, *J. Neurosci. Meth.*, 161, 205–211, doi:10.1016/j.jneumeth.2006.11.004, 2007.

- Wilks, D. S.: Statistical Methods in the Atmospheric Sciences: an Introduction, Academic Press, San Diego, 676 pp., 2011.
- Wolff, C., Haug, G. H., Timmermann, A., Damsté, J. S. S., Brauer, A., Sigman, D. M., Cane, M. A., and Verschuren, D.: Reduced Interannual Rainfall Variability in East Africa During the Last Ice Age, *Science*, 333, 743–747, doi:10.1126/science.1203724, 2011.