TECHNOMETRICS SUPPLEMENTARY MATERIALS

Posterior Confidence Intervals in Linear Calibration Problems: Calibrating the Thompson Ice Core Index (Extended Version)

J. Huston McCulloch[1]

Economics Department

Ohio State University

Oct. 31, 2010

Keywords: Bayesian interval, credible interval, diffuse prior, paleoclimatology, oxygen-18, Medieval Warm Period, Hockey Stick

Abstract

[1] The author is indebted to Stephen McIntyre, Jussi Collin, participants in the OSU Econometrics Seminar and Statistics Colloquium, and several JASA and NSF referees for helpful comments and suggestions.
Author's e-mail: mcculloch.2@osu.edu.

In calibration problems, an exogenous state variable and an endogenous response variable or proxy are both observed in a set of calibration observations. We wish to make inferences about the unobserved state variable from an additional observation on the response variable under a diffuse prior. Hoadley (1970) argued that an informative prior is required in order to obtain a proper posterior distribution. Hunter and Lamboy (1981) proposed a solution, but were sharply criticized at the time. This paper presents a new derivation of the Hunter-Lamboy posterior distribution under a diffuse prior that meets these objections. At the same time, it is shown that Hoadley's approach was based on a subtle inconsistency in the application of Bayes' Rule.

The reinstated Hunter-Lamboy posterior is applied to the problem of calibrating the ice core index of Thompson et al. (2003) to instrumental temperatures. It is found, contrary to the famous claim of Gore (2006), that this index is in fact uninformative about the question of whether Medieval Warm Period was warmer or cooler than the present.

It is shown that the "classical" confidence intervals proposed by Fieller (1954) are a good approximation to the posterior confidence intervals when the calibration slope coefficient is highly significant relative to the desired confidence interval tail probability. However, when the slope is only marginally significant, the Fieller intervals become increasingly distorted and then meaningless. The proposed posterior confidence intervals simply become wider as the slope loses significance, but remain bounded.

Extensions to multiple proxies, sequentially organized data, and prior restrictions on the calibration slope coefficient are outlined but not implemented.

I. Introduction

In the classical calibration problem, as reviewed by Osborne (1991) and Brown (1993), an exogenous *state variable* x_i is measured indirectly by an endogenous *response variable* y_i that has an affine, but noisy, relation to it:

$$y_i = \alpha + \beta x_i + \varepsilon_i . \tag{1}$$

We have a set of *calibration* observations on both for $i = 1, \ldots n$. In the benchmark case considered here, it is assumed that the measurement errors ε_i are i.i.d. Gaussian with $E(\varepsilon_i \mid x_i) = 0$ and $\mathrm{var}(\varepsilon_i) = \sigma^2$. We wish to make inferences about the unobserved state variable x' for an additional observation or observations outside the calibration set, for which we have only the observed response y'.[2] This response obeys the same rule,

$$y' = \alpha + \beta x' + \varepsilon' . \tag{2}$$

Such calibration problems arise in such diverse fields as thermodynamics, pharmacology, urology, and chemical analysis (*cf.* Brown 1993), zoology (du Plessis and van der Merwe 1996), and paleoclimatology (*e.g.* Mann *et al*. 1999, Kaufman *et al.* 2009). In econometrics, applications could include the reconstruction of historical Gross National Product from fragmentary records (*e.g.* Romer 1989), or the historical standard of living from height data (*e.g.* Steckel 1995).

Solving (2) for x' yields

$$x' = (y' - \alpha - \varepsilon')/\beta , \tag{3}$$

so that the natural or "classical" calibration estimator of x', proposed by Eisenhart (1939), is

$$\hat{x}' = (y' - \hat{\alpha})/\hat{\beta} , \tag{4}$$

[2] I here use x' and y' for the reconstruction values, following the nomenclature of Brown (1982).

where $\hat{\alpha}$ and $\hat{\beta}$ are the Ordinary Least Squares (OLS) estimates of α and β from (1). The expression in (4) is also the Maximum Likelihood estimator of x' (Hoadley 1970, Brown 1993).

The present paper presents a new derivation of the Hunter and Lamboy (HL, 1981) posterior distribution for x' under an uninformative, diffuse prior. The resulting confidence intervals (CIs), or credible intervals as they are often called in the Bayesian context, are always bounded and contiguous, and may readily be constructed with equal tail probabilities.

The traditional or "classical" method of computing confidence regions advocated by most authors is based on a theorem of Fieller (1954). However, this approach leads to unsatisfactory sets that may be unbounded at one or even both ends, and may even be discontiguous. Even when they are bounded intervals, the classical CIs are in most cases excessively skewed away from the mean of the calibration values of the state variable, so that the two tail probabilities are unequal. Hoadley (1970) attempts to construct a Bayesian alternative, but is unable to do this without an unnecessarily informative prior.

The reinstated Hunter-Lamboy method is applied to the problem of calibrating the Thompson et al. (2003) ice core isotope ratio index to instrumental temperature. This index was made famous in Al Gore's *An Inconvenient Truth* (2006), where it was said to provide definitive confirmation that temperatures during the Medieval Warm Period (MWP) were lower than those at present. It is found that the index does imply a point estimate of temperature that is uniformly below the 1961-1990 instrumental average from the beginning of the series in 1001-1010 AD until the 1930's. However, the 95% posterior CI for the temperature anomaly typically extends from at least -1.8°C to

+1.2°C, so that the reconstruction is in fact uninformative. Although this index does not point to a MWP, it by no means rules one out, contrary to the claim of Gore. The multiproxy temperature reconstruction of Loehle and McCulloch (2008) instead shows a statistically significant MWP, relative to the bimillenial average, during most of 817-1036 AD plus 1249-1265 AD, as well as a significantly cool Little Ice Age (LIA) during most of 1442-1746 AD.

Section II of this paper presents the new derivation of the HL posterior distribution for the unobserved state variable, contrasts it with the approach of Hoadley (1970), and answers the criticisms that were raised against the HL approach when it was first published. Section III constructs a close approximation to the Thompson et al. (2003) ice core isotope index from source data. Section IV calibrates this index to global temperature, and computes posterior 50% and 95% CIs. The concluding section V compares these intervals to the classical Fieller regions, and discusses extensions of the method of Section II.

II. Posterior Confidence Intervals for the Classical Calibration Estimator

The joint prior distribution for x' and the regression parameters $\mathrm{p}(x', \alpha, \beta, \sigma^2)$ that governs the entire exercise may be factored as follows:

$$\mathrm{p}(x', \alpha, \beta, \sigma^2) = \mathrm{p}(x' \mid \alpha, \beta, \sigma^2)\,\mathrm{p}(\alpha, \beta \mid \sigma^2)\,\mathrm{p}(\sigma^2). \tag{5}$$

Bayes' Rule implies that the distribution of x', conditional on y' and the true parameters, is

$$\mathrm{p}(x' \mid y', \alpha, \beta, \sigma^2) \propto \mathrm{p}(y' \mid x', \alpha, \beta, \sigma^2)\,\mathrm{p}(x' \mid \alpha, \beta, \sigma^2), \tag{6}$$

so that in the benchmark case in which the prior density component $\mathrm{p}\left(x' \mid \alpha, \beta, \sigma^2\right)$ is diffuse and therefore uninformative (*cf*. Zellner 1971: 41-53),

$$\begin{aligned}
\mathrm{p}\left(x' \mid y', \alpha, \beta, \sigma^2\right) &\propto \mathrm{p}\left(y' \mid x', \alpha, \beta, \sigma^2\right) \\
&\propto \exp\left(-\frac{\left(y' - \alpha - \beta x'\right)^2}{2\sigma^2}\right) \\
&= \exp\left(-\frac{\left(x' - \left(y' - \alpha\right)/\beta\right)^2}{2\sigma^2/\beta^2}\right) \\
&\sim \mathrm{N}\left(\left(y' - \alpha\right)/\beta,\ \sigma^2/\beta^2\right),
\end{aligned} \tag{7}$$

whence

$$\mathrm{p}\left(\begin{pmatrix} \beta x' \\ \beta \end{pmatrix} \mid y', \alpha, \beta, \sigma^2\right) \sim \mathrm{N}\left(\begin{pmatrix} y' - \alpha \\ \beta \end{pmatrix}, \sigma^2 \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}\right). \tag{8}$$

The distribution of the OLS estimators $\hat{\alpha}$ and $\hat{\beta}$, conditional on the true parameters, is

$$\mathrm{p}\left(\begin{pmatrix} \hat{\alpha} \\ \hat{\beta} \end{pmatrix} \mid \alpha, \beta, \mathbf{S}, \sigma^2\right) \sim \mathrm{N}\left(\begin{pmatrix} \alpha \\ \beta \end{pmatrix}, \sigma^2 \mathbf{S}\right),$$

where

$$\mathbf{S} = \begin{pmatrix} s_{\alpha\alpha} & s_{\alpha\beta} \\ s_{\alpha\beta} & s_{\beta\beta} \end{pmatrix} = \begin{pmatrix} n & \sum x_i \\ \sum x_i & \sum x_i^2 \end{pmatrix}^{-1}.$$

Then under the standard diffuse prior for $\mathrm{p}\left(\alpha, \beta \mid \sigma^2\right)$ (Zellner 1971, ch. 3),

$$\mathrm{p}\left(\begin{pmatrix} \alpha \\ \beta \end{pmatrix} \mid \hat{\alpha}, \hat{\beta}, \mathbf{S}, \sigma^2\right) \sim \mathrm{N}\left(\begin{pmatrix} \hat{\alpha} \\ \hat{\beta} \end{pmatrix}, \sigma^2 \mathbf{S}\right), \tag{9}$$

whence

$$\mathrm{p}\left(\begin{pmatrix} y' - \alpha \\ \beta \end{pmatrix} \mid y', \hat{\alpha}, \hat{\beta}, \mathbf{S}, \sigma^2\right) \sim \mathrm{N}\left(\begin{pmatrix} y' - \hat{\alpha} \\ \hat{\beta} \end{pmatrix}, \sigma^2 \begin{pmatrix} s_{\alpha\alpha} & -s_{\alpha\beta} \\ -s_{\alpha\beta} & s_{\beta\beta} \end{pmatrix}\right). \tag{10}$$

Since $\beta x' - (y' - \alpha) = -\varepsilon'$ is independent of the regression data, combining (8) with (10) yields

$$\mathrm{p}\left(\begin{pmatrix}\beta x' \\ \beta\end{pmatrix} \mid y', \hat{\alpha}, \hat{\beta}, \mathbf{S}, \sigma^2\right) \sim \mathrm{N}\left(\begin{pmatrix}y' - \hat{\alpha} \\ \hat{\beta}\end{pmatrix}, \sigma^2 \begin{pmatrix}s_{\alpha\alpha} + 1 & -s_{\alpha\beta} \\ -s_{\alpha\beta} & s_{\beta\beta}\end{pmatrix}\right). \qquad (11)$$

It follows from $x' \equiv (\beta x')/\beta$ that

$$\mathrm{p}\left(x' \mid y', \hat{\alpha}, \hat{\beta}, \mathbf{S}, \sigma^2\right) \sim \mathrm{R2N}\left(\begin{pmatrix}y' - \hat{\alpha} \\ \hat{\beta}\end{pmatrix}, \sigma^2 \begin{pmatrix}s_{\alpha\alpha} + 1 & -s_{\alpha\beta} \\ -s_{\alpha\beta} & s_{\beta\beta}\end{pmatrix}\right), \qquad (12)$$

where $\mathrm{R2N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ indicates the distribution of the ratio of two normally distributed random variables with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$ (Fieller 1932, Hinckley 1969). This distribution has as special cases the Cauchy distribution when the numerator and denominator both have mean 0, and the Gaussian distribution when the denominator has zero variance. In general it is heavy-tailed, with tails intermediate between these two cases, and is skewed unless the numerator or denominator has mean 0. It may even be bimodal if the numerator is large relative to its standard deviation and the denominator small relative to its standard deviation. See Marsaglia (1965) for graphs of the density in the special case of zero correlation. The Cumulative Distribution Function (CDF) $\mathrm{P}\left(x' \mid y', \hat{\alpha}, \hat{\beta}, \mathbf{S}, \sigma^2\right)$, as required for confidence intervals on x', may easily be computed in terms of the bivariate normal CDF, as shown in Appendix I of the present paper, equation (26).

It may easily be seen that conditional on the true parameters, the classical point estimate $\hat{x}'$ in (4) has the same R2N distribution as (12), only with the OLS estimates of β and $x'\beta$ replaced by their true values:

$$\mathrm{p}\left(\hat{x}' \mid y', \alpha, \beta, \mathbf{S}, \sigma^2\right) \sim \mathrm{R2N}\left(\begin{pmatrix} x'\beta \\ \beta \end{pmatrix}, \sigma^2 \begin{pmatrix} s_{\alpha\alpha} + 1 & -s_{\alpha\beta} \\ -s_{\alpha\beta} & s_{\beta\beta} \end{pmatrix}\right). \qquad (13)$$

The R2N posterior distribution for x' in (12) is therefore particularly natural and intuitive. Williams (1969) notes, in the context of (13), that the R2N distribution has undefined mean and variance. In fact, it lies the domain of attraction of the Cauchy distribution, so that even the Generalized (stable domain of attraction) Law of Large Numbers cannot be relied upon to make the distribution of the average more compact than that of the contributions. Care should therefore be exercised in merging multiple classical estimates by simple averaging as in Loehle and McCulloch (2008), or even by GLS as in Brown (1982).

In practice, σ^2 is unknown and must be estimated from the variance of the residuals

$$s^2 = \sum_{i=1}^{n} e_i^2 / (n-2),$$

where $e_i = y_i - \hat{\alpha} - \hat{\beta}x_i$. Under a diffuse and therefore uninformative prior for $\log(\sigma^2)$ so that $\mathrm{p}\left(\sigma^2\right) \propto \sigma^{-2}$ (Zellner 1971: 41-53),

$$\mathrm{p}\left(\sigma^2 \mid s^2\right) \sim (n-2)s^2 / \chi^2_{(n-2)}.$$

Then

$$\mathrm{p}\left(\begin{pmatrix} \alpha \\ \beta \end{pmatrix} \mid \hat{\alpha}, \hat{\beta}, \mathbf{S}, s^2\right) \sim \mathrm{BVT}\left(\begin{pmatrix} \hat{\alpha} \\ \hat{\beta} \end{pmatrix}, s^2\mathbf{S}, n-2\right), \qquad (14)$$

where BVT($\boldsymbol{\mu}$, $\mathbf{V}$, ν) represents the elliptical bivariate Student t distribution with mean $\boldsymbol{\mu}$, covariance matrix estimate $\mathbf{V}$, and degrees of freedom ν (*cf.* Zellner 1971, ch. 3, appendix B.2; Genz 2004). It follows from (11) and (14) that

$$\mathrm{p}\left(x' \mid y', \hat{\alpha}, \hat{\beta}, \mathbf{S}, s^2\right) \sim \mathrm{R2T}\left(\begin{pmatrix} y'-\hat{\alpha} \\ \hat{\beta} \end{pmatrix}, s^2 \begin{pmatrix} s_{\alpha\alpha}+1 & -s_{\alpha\beta} \\ -s_{\alpha\beta} & s_{\beta\beta} \end{pmatrix}, n-2\right), \qquad (15)$$

where $\mathrm{R2T}(\boldsymbol{\mu}, \mathbf{V}, \nu)$ represents the distribution of the ratio of two elliptical bivariate Student t random variables with mean vector $\boldsymbol{\mu}$, covariance matrix estimate $\mathbf{V}$, and ν degrees of freedom. Its CDF $\mathrm{P}\left(x' \mid y', \hat{\alpha}, \hat{\beta}, \mathbf{S}, s^2\right)$ may easily be computed in terms of the elliptical bivariate Student t CDF, as also shown in Appendix I of the present paper.[3]

The present paper first uses Bayes' Rule to derive $\mathrm{p}\left(x' \mid y', \alpha, \beta, \sigma^2\right)$ from $\mathrm{p}\left(y' \mid x', \alpha, \beta, \sigma^2\right)$ under a diffuse prior for $\mathrm{p}\left(x' \mid \alpha, \beta, \sigma^2\right)$. Then the regression parameters are estimated using the calibration data and the true values are integrated out to obtain $\mathrm{p}\left(x' \mid y', \hat{\alpha}, \hat{\beta}, \mathbf{S}, s^2\right)$. Hoadley (1970: 361-364) likewise considers the use of a diffuse prior for x'. However, he instead first estimates the regression parameters to obtain the classical Student t distribution with n-2 degrees of freedom for y' as a function of x':

$$\mathrm{p}\left(y' \mid x', \hat{\alpha}, \hat{\beta}, \mathbf{S}, s^2\right) \propto \mathrm{v}(x')^{-1/2}\left(1 + \frac{\left(y' - \hat{\alpha} - \hat{\beta}x'\right)^2}{(n-2)\mathrm{v}(x')}\right)^{-(n+1)/2}, \qquad (16)$$

where

$$\mathrm{v}(x') = s^2\left(s_{\alpha\alpha} + 1 + 2x's_{\alpha\beta} + x'^2 s_{\beta\beta}\right) \qquad (17)$$

is the estimated variance of y' as a function of x'. Hoadley then notes that holding y' constant and considering this expression as a function of x', it has tails that are asymptotically proportional to $1/|x'|$ and therefore integrate to infinity when multiplied

[3] Hunter and Lamboy (1981) report the exact formula for the density of the ratio of two Student t random variables, but then one must still integrate this density numerically in order to compute CIs, so that it is easier just to use the BVT CDF to compute the R2T CDF directly.

by a uniform prior. He concludes that "an improper uniform prior on [x'] leads to a nonsensical Bayes estimator," and therefore rejects the uniform prior. Brown (1982; 1993, p. 98) and Osborne (1991) concur with Hoadley that a proper prior is required in the case, considered here, of a single response variable.

Hoadley consequently introduces an informative prior that is based on the assumption that x' is drawn from the marginal distribution of the calibration values, and shows that this leads to the "inverse" estimator favored by Krutchkoff (1967), based on a calibration regression in which the independent variable x is regressed on the dependent variable y. Although there are applications in which such a prior would be justified, the present paper is primarily concerned with the important case of an uninformative diffuse prior.[4]

However, there is a subtle fallacy in Hoadley's reasoning: Because the Student t density (16) is conditioned on the regression data, Hoadley's prior for x' must be as well – Bayes' Rule requires

$$\mathrm{p}\left(x' \mid y', \hat{\alpha}, \hat{\beta}, \mathbf{S}, s^2\right) \propto \mathrm{p}\left(y' \mid x', \hat{\alpha}, \hat{\beta}, \mathbf{S}, s^2\right) \mathrm{p}\left(x' \mid \hat{\alpha}, \hat{\beta}, \mathbf{S}, s^2\right),$$

not

$$\mathrm{p}\left(x' \mid y', \hat{\alpha}, \hat{\beta}, \mathbf{S}, s^2\right) \propto \mathrm{p}\left(y' \mid x', \hat{\alpha}, \hat{\beta}, \mathbf{S}, s^2\right) \mathrm{p}(x').$$

It is not immediately obvious why conditioning the prior for x' on the regression data would make any difference. However, consider generalizing the joint prior (5) that governs the entire exercise to be conditioned on an information set Ω that may or may not be empty:

[4] In the case of a sequential state variable that follows a stochastic process with stationary increments, an informative prior that depends on the nearest of the calibration state variables is appropriate, as discussed below. Even then, however, this prior should be applied before the regression coefficients are integrated out, and not afterwards.

$$\mathrm{p}\left(x', \alpha, \beta, \sigma^2 \mid \Omega\right) = \mathrm{p}\left(x' \mid \alpha, \beta, \sigma^2, \Omega\right) \mathrm{p}\left(\alpha, \beta \mid \sigma^2, \Omega\right) \mathrm{p}\left(\sigma^2 \mid \Omega\right). \qquad (18)$$

Since Hoadley's prior for x' implicitly depends on the regression data, his Ω must also contain the regression data in each of the other terms in (18), including $\mathrm{p}\left(\alpha, \beta \mid \sigma^2, \Omega\right)$. However, conditioning the prior for the regression parameters on the regression results is inconsistent with using the same regression results to inform the posterior for the parameters. Hoadley's sequence of operations is therefore invalid.

It has been shown here that there are no real problems with the Bayesian approach under a diffuse prior, provided that a diffuse prior for the conditional density $\mathrm{p}\left(x' \mid \alpha, \beta, \sigma^2\right)$ that arises from the factorization (5) is applied *before* the regression parameters are integrated out, rather than a diffuse prior for the conditional density $\mathrm{p}\left(x' \mid \hat{\alpha}, \hat{\beta}, \mathbf{S}, s^2\right)$ *after* the true regression parameters have been eliminated.

Hunter and Lamboy (HL, 1981) derived, in density form, exactly the same R2T posterior distribution for x' as we do above, but using a diffuse prior for $\eta = \alpha + \beta x'$ instead of for x' itself as in the above derivation. In the same issue of *Technometrics*, Hill (1981), Lawless (1981), and Orban (1981) sharply criticized HL under the assumption that they were treating the prior for their η as being independent of that for the regression parameters. Hill noted that if the prior for η is uniform, the implicit prior for x' must be proportional to $|\beta|$, so that the priors for η and x' cannot both be independent of the regression parameters, and argues that the latter is the more natural assumption.

However, independence of the priors, as insisted upon by the HL critics, is in fact nowhere required. The full joint prior for the exercise may either be expressed in terms of x' as in (5), or in terms of the HL η as

$$\mathrm{p}(\eta,\alpha,\beta,\sigma^2) = \mathrm{p}(\eta \mid \alpha,\beta,\sigma^2)\mathrm{p}(\alpha,\beta \mid \sigma^2)\mathrm{p}(\sigma^2). \quad (19)$$

The assumption that our $\mathrm{p}(x' \mid \alpha,\beta,\sigma^2)$ is constant with respect to x' is equivalent to the HL assumption that their implicit $\mathrm{p}(\eta \mid \alpha,\beta,\sigma^2)$ is constant with respect to η. It is immaterial whether either of these is proportional to $|\beta|$ or its inverse, since any such constant will simply drop out in the normalizing integration when Bayes' Rule is applied as in (7).[5]

III. The Thompson Ice Core Index

The reinstated Hunter-Lamboy CI method is illustrated by calibrating the millenial ice core isotope ratio index of Lonnie Thompson *et al.* (2003) to recent global instrumental temperatures. This index was famously described by Al Gore in his Nobel Peace Prize-winning *An Inconvenient Truth* (2006: 60-65), as follows:

> Scientist Lonnie Thompson takes his team to the tops of glaciers all over the world. They dig core drills down into the ice, extracting long cylinders filled with ice that was formed year by year over many centuries.
>
> Lonnie and his team of experts .... can .... measure the exact temperature of the atmosphere each year by calculating the ratio of different isotopes of oxygen (oxygen-16 and oxygen-18), which provides an ingenious and highly accurate thermometer. ....
>
> The thermometer to the right measures temperatures in the Northern Hemisphere over the past 1,000 years. .... [T]he so-called global-warming skeptics often say that global warming is really an illusion reflecting nature's cyclical fluctuations. To support their view, they frequently refer to the Medieval Warm Period.
>
> But as Dr. Thompson's thermometer shows, the vaunted Medieval Warm Period (the third little red blip from the left, below) was tiny compared to

---

[5] HL (1981: 327) erroneously accept that the difference between their approach and that of Hoadley (1970) is that they assume a uniform prior on their η, whereas Hoadley considers a uniform prior for x'. In fact, the difference is that HL condition their prior for η or equivalently x' on the true parameters, whereas Hoadley inappropriately conditions his prior for x' on the parameter estimates, as noted above.

> the enormous increases in temperature of the last half century (the red peaks at the far right of the chart).
>
> These global warming skeptics ... launched a fierce attack against another measurement of the 1,000-year correlation between CO_2 and temperature known as "the hockey stick," a graphic image representing the research of Michael Mann and his colleagues. But in fact, scientists have confirmed the same basic conclusions in multiple ways – with Thompson's Ice core record as one of the most definitive. (Gore 2006: 60-64).

McIntyre and McKitrick (2003, 2005) had argued that the "hockey stick" reconstruction of Mann *et al.* (1999) is based on invalid stripbark bristlecone treering data, as well as improper use of Principal Components Analysis. As it happens, the graph that Gore (2006) presents as "Dr. Thompson's Thermometer" actually *was* the disputed "hockey stick" temperature reconstruction for which it was supposed to provide definitive independent confirmation, spliced together with an instrumental temperature record as if they were a single series, and had nothing to do with Thompson's ice core data.[6]

Thompson, who served as a member of the Science Advisory Board for *An Inconvenient Truth*, had indeed published a similar-looking graph based on decadal averages of ice core oxygen isotope ratios in Figure 7 of Thompson *et al*. (2003), which is reproduced in Figure 1 below.

---

[6] In fact, two of the 12 series employed by Mann *et al.* (1999) for the crucial 1000-1400AD period were Quelccaya $\delta^{18}O$ and precipitation series, so that the hockey stick is not completely independent of even the true Thompson ice core index.

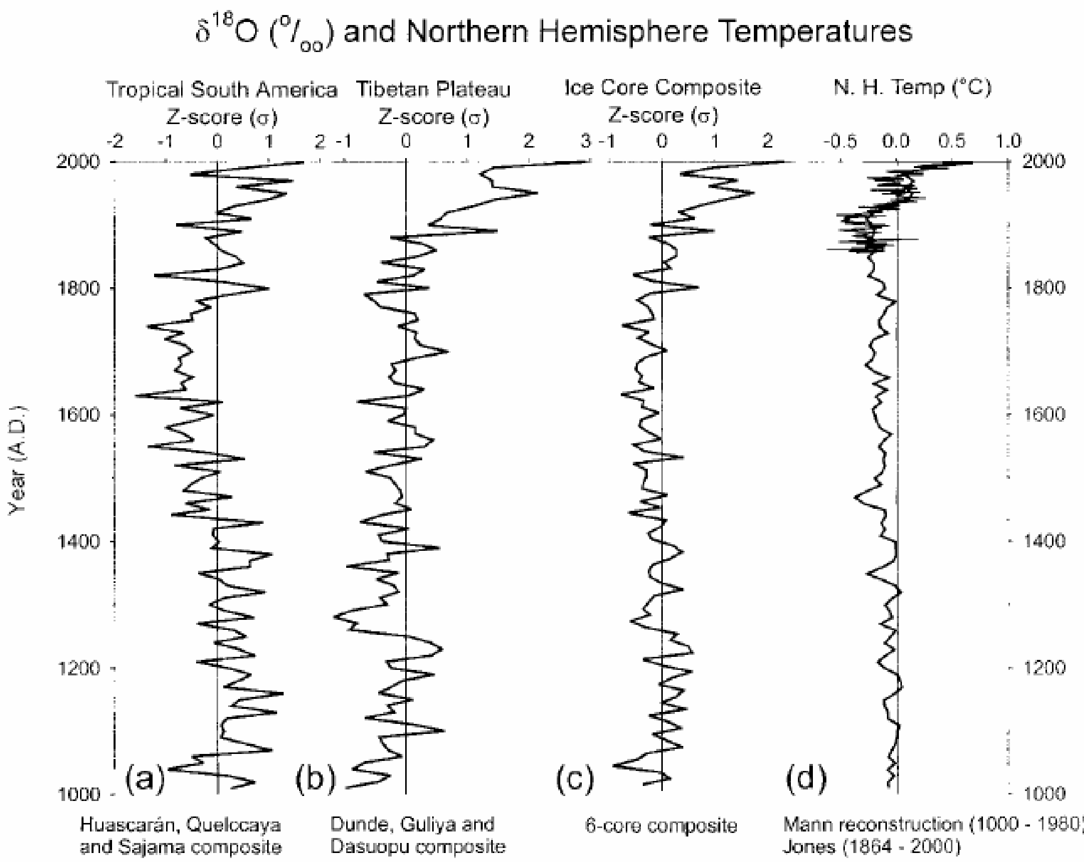

Figure 7. Regional composites, shown as z-scores, for the last millennium were constructed from the decadal averages of $\delta^{18}O_{ice}$ from three Andean ice cores (a) and three Tibetan ice cores (b). The composite of all six low latitude cores is shown in (c). The measured (Jones et al., 1999) and reconstructed (Mann et al., 1999) Northern Hemisphere temperatures are shown in (d) and are plotted as deviations (°C) from their respective 1961–1990 means. Note that the decadal average of $\delta^{18}O_{ice}$ for 1991 to 2000 is based on the 1991 to 1997 annual values for the Dasuopu core drilled in 1997 and on the 1991–1997 annual values for the Sajama drilled in 1997. The Quelccaya $\delta^{18}O_{ice}$ history has been updated to 2000 by drilling new shallow cores.

Figure 1.
Thompson *et al.* (2003), Fig. 7, with original caption.

Panel d) of Fig. 1 is the disputed Mann *et al.* "hockey stick" reconstruction itself, overlain with a Northern Hemisphere temperature index. Gore was supposed to have used panel c), but mistakenly used panel d) of the same figure instead, merging its two lines into a single graph.[7] The slightly positive values around 1320 AD in panel d)

[7] This substitution was confirmed by Lonnie Thompson in response to a question by the author at a seminar at Ohio State University, Jan. 11, 2008.

became the "third little red blip from the left" representing the "vaunted Medieval Warm Period" according to Gore.

The actual Low Latitude composite ice core index in panel c) does turn up sharply in the 20th century, much like the instrumentally augmented hockey stick. However, the axes at the top of panels a)-c) do not indicate temperature, but merely composite Z-scores computed from $\delta^{18}O_{ice}$ measurements. Thompson *et al*. in fact did not calibrate these Z-scores to temperature, let alone provide CI's for such a calibration, so that "Dr. Thompson's Thermometer" was not yet a thermometer at all. The present paper attempts to fill this deficiency, by calibrating this index to instrumental temperature, and computing confidence intervals.

Thompson *et al.* provide decadally averaged data for 5 of the 6 cores used in their article in a spreadsheet online at <bprc.osu.edu/Icecore/Climatic-change-2003-Fig5-table.XLS>. Unfortunately, however, the Himalayan composite Z-score (henceforth HCZ) series is missing, along with the data for Quelccaya, so that even though the last column is identified as "6 core composite", it is in fact based only on the two included Andean cores, and is definitely not the series plotted in panel c) of their Fig. 7.

On March 3, 2008, I e-mailed Lonnie Thompson and several of his co-authors requesting the missing series in this spreadsheet, but received no reply. Nevertheless, it is possible to at least approximately reconstruct the 6-core composite from the five decadal $\delta^{18}O_{ice}$ series in the spreadsheet, plus archived annual values for the Quelccaya Summit core, with the help of formulas embedded in the spreadsheet.

Figure 2 shows the decadally averaged $\delta^{18}O_{ice}$ values from the spreadsheet for the three 3 Himalayan cores, Guliya, Dunde and Dasuopu.[8] Each decadal value is identified by the date at the top of its interval. These dates are even multiples of 10, so that the decades used for these three series are 1001-1010, 1011-1020, etc. Note that only Dasuopu has data for the final decade of the study, 1991-2000 (through 1997). Guliya ends with 1981-1990, and Dunde with 1981-1987. All three of these Himalayan series visually match those in Figure 5 of Thompson *et al.* (2003).

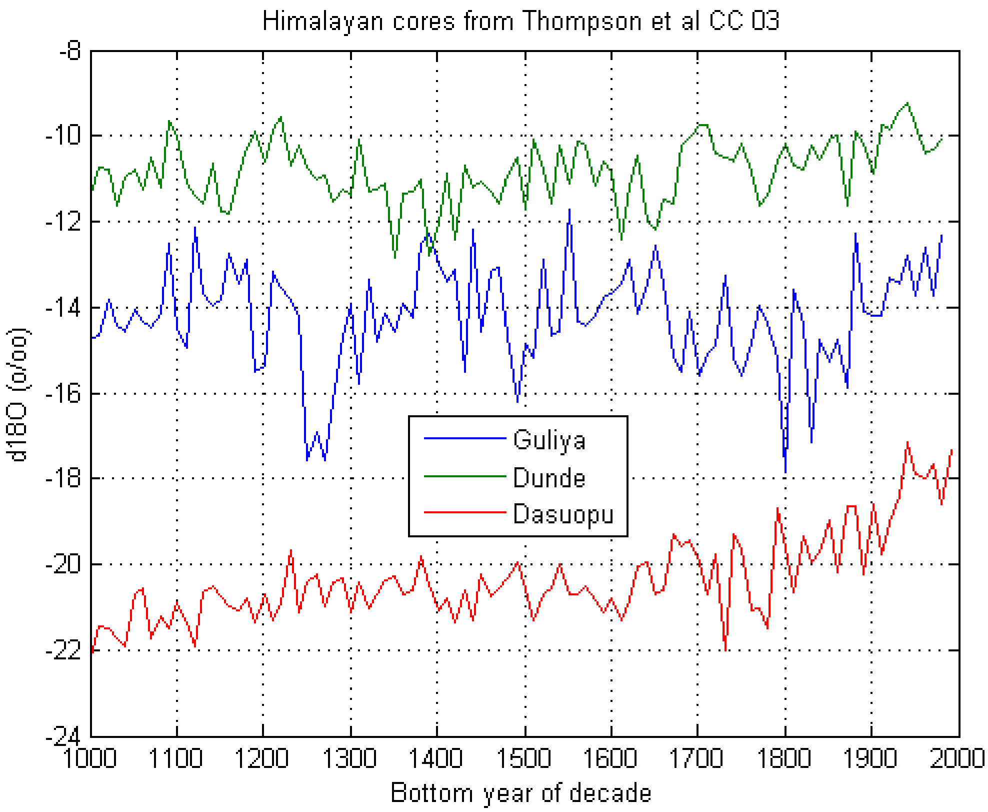

Figure 2.
Decadal averages of $\delta^{18}O_{ice}$ for the three Himalayan cores: Guliya, Dunde, and Dasuopu.

---

[8] The decadally averaged data from this spreadsheet for the 3 Himalayan cores has been archived at ftp://ftp.ncdc.noaa.gov/pub/data/paleo/icecore/trop/dasuopu/dasuopu-d18o.txt, <... /dunde/dunde-d18o.txt>, and <... /guliya/guliya-d18o.txt>. However, the decadal Andean data from the spread sheet is not archived there except inadvertently at the lower right side of the Guliya file.

IPCC AR4 WGI expert reviewer Stephen McIntyre (2006, 2007) notes that several inconsistent versions of Guliya and Dunde have been used in the climate literature. The present paper simply uses the version in the spreadsheet constructed for Thompson *et al.* (2003).

Only two of the three Andean cores used, Sajama and Huascarán Core 2, are tabulated in the online spreadsheet. Each of these decadal values is associated with two dates identified as "bottom", and "to decade", but both are even multiples of 10, starting with (1000, 1010), (1010, 1020), etc. Evidently these Andean cores were averaged over decades *beginning* with an even multiple of 10 such as 1000-1009, 1010-1019, etc., so that the header "to decade" refers to the first year of the *following* decade. This interpretation is reinforced by an annotation that the most recent value for Huascarán Core 2 covers the period 1990 to 1993.

The third Andean core is identified by Thompson *et al.* (2003) only as "Quelccaya," even though annual $\delta^{18}O_{ice}$ data is archived at ftp.ncdc.noaa.gov for two different Quelccaya cores, Core 1 and the Summit Core. It was found that averages of the Summit Core for decades *beginning* with even multiples of 10 gave a perfect visual fit to the graph of the "Quelccaya" series in Fig. 6 of Thompson *et al.* (2003), while decades *ending* with even multiples of 10 had obvious differences. Core 1 had obvious differences under either interpretation, and clearly was not the source of their Fig. 6. The average of the two cores also has obvious differences.

The archived annual data for the Quelccaya Summit Core (Thompson *et al.* 2005) only extend to 1984, yet Thompson *et al.* (2003) use data for Quelccaya through 2000 based on newer shallow cores (see caption to their Fig. 7 in Fig. 1 above) that are not available numerically. Accordingly, these two values were read visually off Fig. 6 of Thompson *et al.* (2003) as -17.67 per mil for 1980-1989 and -16.64 per mil for 1990-1999.

Decadal averages of $\delta^{18}O_{ice}$ for the three Andean cores, Quelccaya Summit, Sajama, and Huascarán 2, are shown in Figure 3 below. The Sajama series is a good visual match to that in Figure 6 of Thompson *et al.* (2003). The relative sizes of the local peaks in the Huascarán 2 series since 1790 are not quite the same as in Figure 6 of Thompson *et al.* (2003), although the general shape of the two series are very similar. As noted above, Quelccaya is a perfect match, but only provided the Summit Core is used with decades beginning with 0.

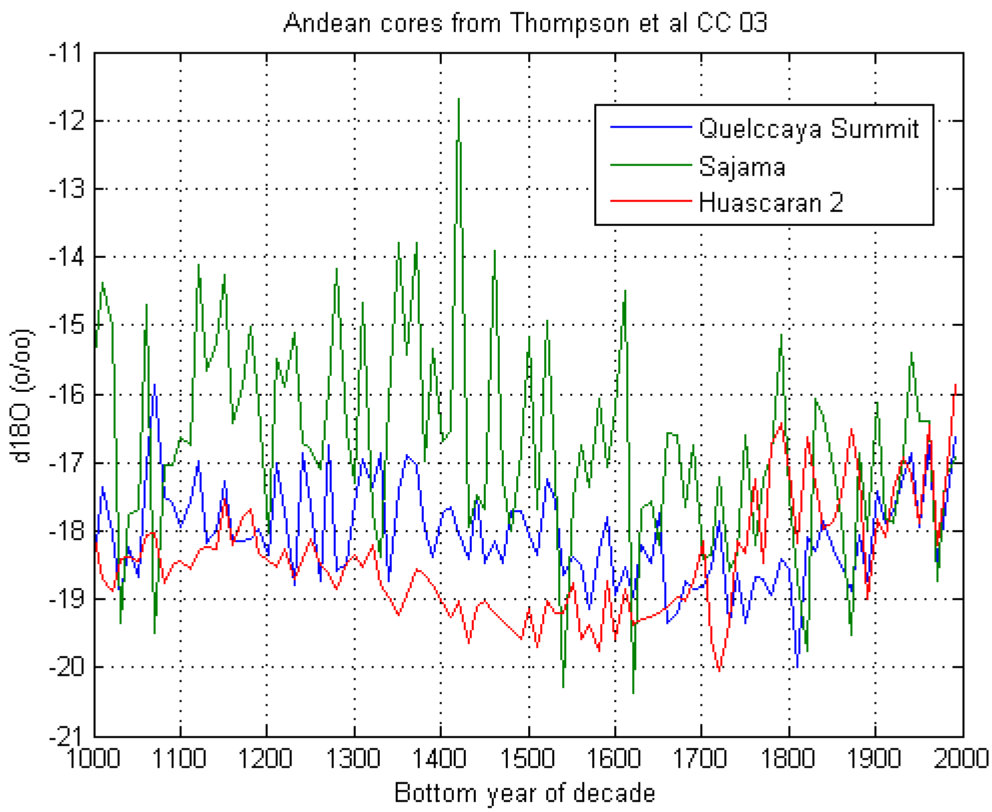

Figure 3
Decadal averages of $\delta^{18}O_{ice}$ for the three Andean cores Quelccaya Summit, Sajama, and Huascarán 2.

The ice-core indices shown in Fig. 1 are composite Z-scores are supposed to be derived somehow from the decadal $\delta^{18}O_{ice}$ data shown in Figs. 2 and 3 above, but

Thompson *et al.* (2003) do not provide details of how they were computed. Were the $\delta^{18}O_{ice}$ values averaged and then Z-scores computed, or were the averages computed from Z-scores? Were the means and variances for the Z-scores computed before or after decadal averaging? Were they computed using each entire series, or just the portion after 1000 AD used for the study? And were the means and variances computed to the end of each series, or only using the common portions?

Although the provided data spreadsheet is incomplete, it does contain embedded formulas that reveal how the Z-scores would have been computed if the data were present: First, each decadally averaged series was converted to Z-scores using its mean and standard deviation for all its available decades, but restricted to the period of the study, i.e. beginning with 1001-1010 for the Himalayan cores and with 1000-1009 for the Andean cores. These Z-scores are shown in Figures 4 and 5 below:

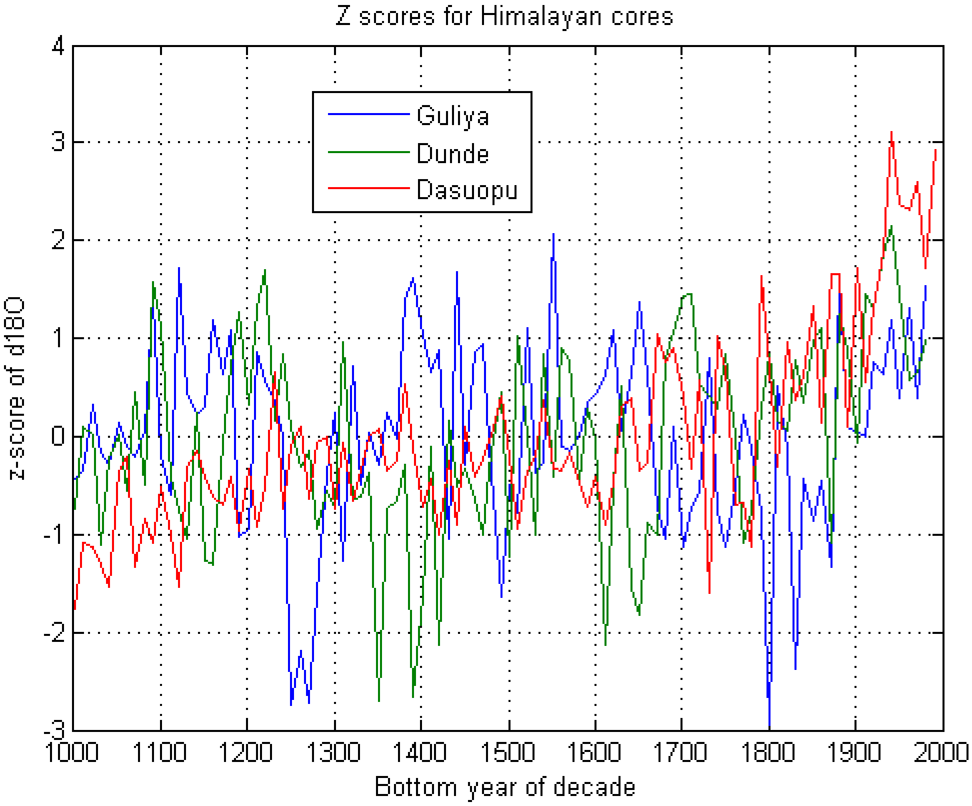

Fig. 4
Z-scores for the three Himalayan cores.

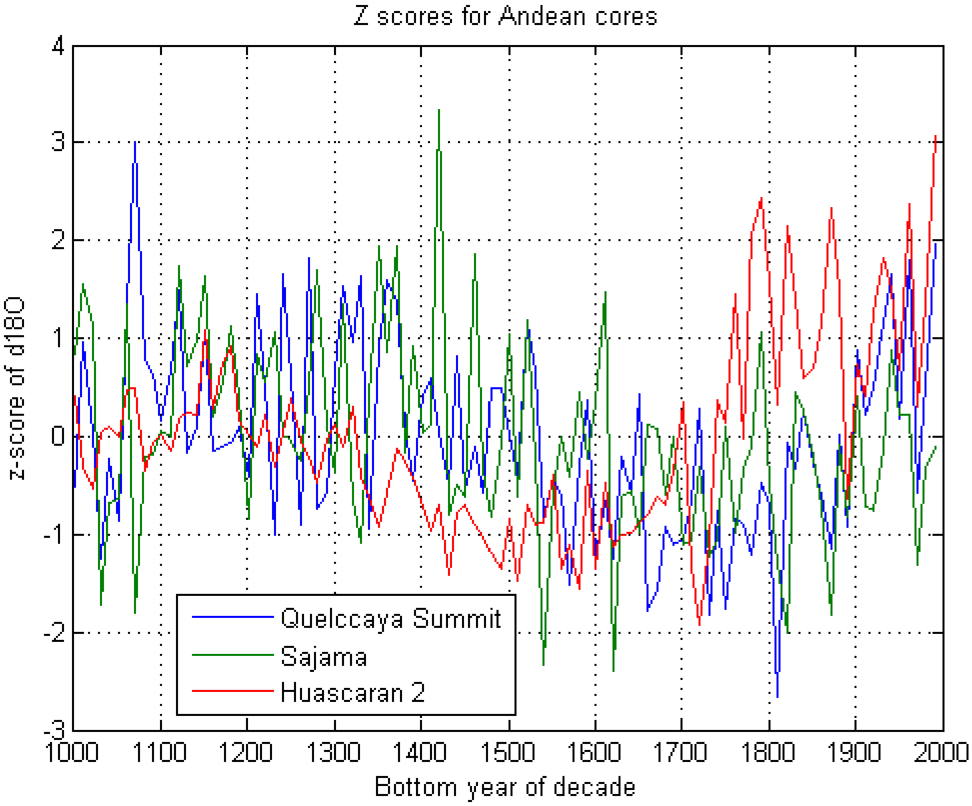

Fig. 5
Z-scores for the three Andean cores.

Then, regional composite Z-scores were computed by averaging each region's Z-scores, *as available*. The resulting Andean Composite Z-score series (ACZ), shown in Fig 6 below, is thus simply the average of the three component Z-scores for all 100 decades. Unfortunately, however, this series is not a perfect visual match to Fig. 1a of Thompson et al. (2003) (Fig. 1a above). Although most of the high frequency wiggles match, the local high in the 1720s and the record low in the 1810s are greatly attenuated in the Thompson et al. version. Furthermore, although the 20th century is generally the highest in the Thompson version, it is not quite as high as the 12th-14th centuries in our emulation. Perhaps Thompson et al. in fact used Quelccaya Core 1 or the average of the two Quelccaya cores in the construction of the index, even though the Summit core was clearly what was used in their Fig. 6.

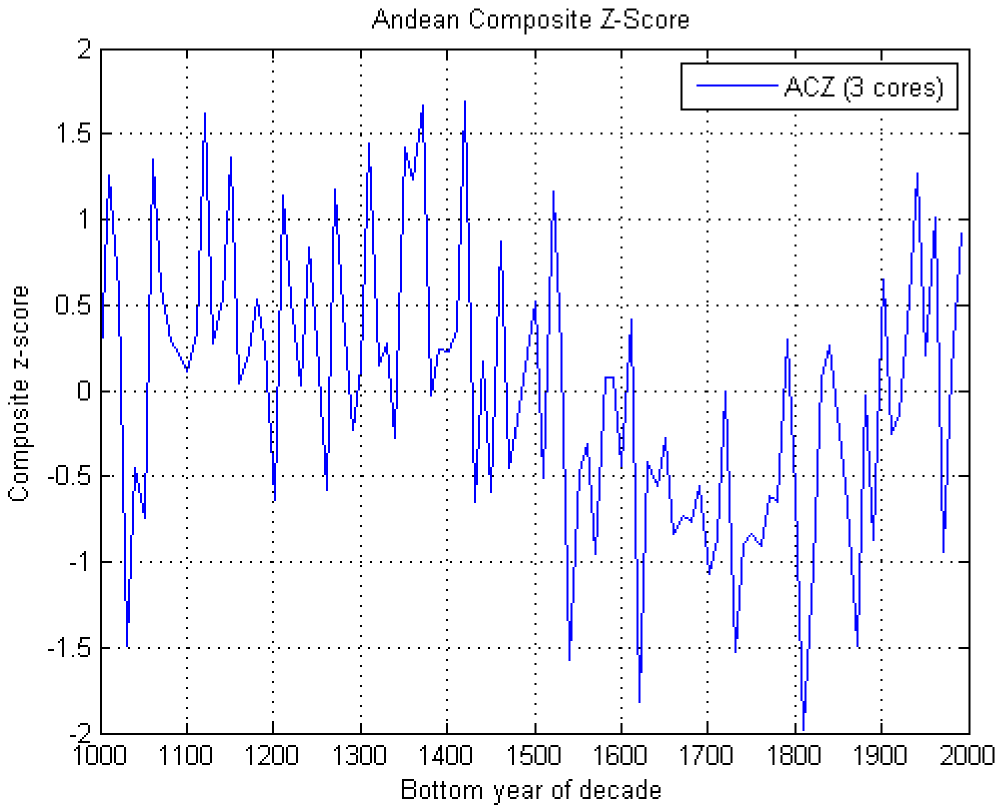

Fig. 6
Andean Composite Z-score series ACZ.

The emulated Himalayan Composite Z-score (HCZ) series, shown in Fig. 7 below, is based on all 3 Himalayan cores for the first 99 decades (shown in blue), and then abruptly is based only on Dasuopu for the final decade (shown in red), after the other two drop out. The line segment connecting the point representing the 1980s (3 cores) and that representing the 1990s (Dasuopu only) is also in red. The resulting series is a perfect visual match to Panel (b) of Fig. 1.

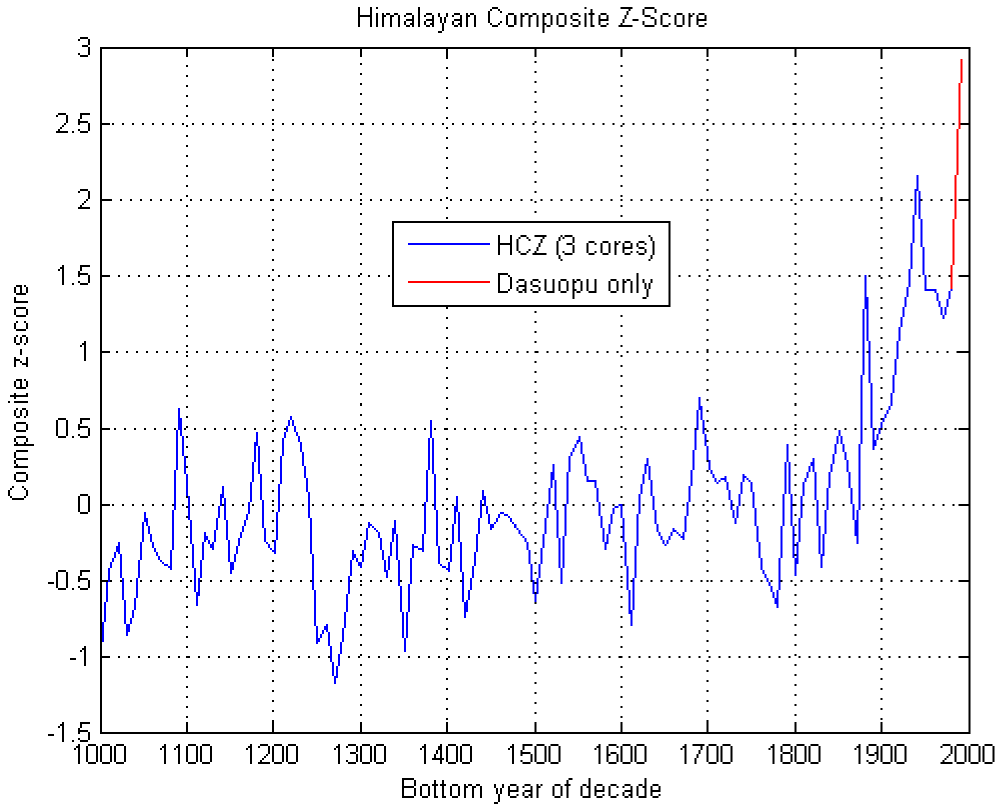

Fig. 7.
Himalayan Composite Z-score series (HCZ), using all three Himalayan cores 1001-1990 (in blue), and Dasuopu only, 1991-2000 (in red). The line segment connecting 1981-1990 (3 cores) and 1991-2000 (Dasuopu only) is also in red.

Finally, the spreadsheet formulas show that two regional composites were averaged with equal weights to obtain a single Low Latitude Composite Z-Score (LCZ), as shown in Figure 8 below. The final decade, as well as the line segment connecting the 1980s (6 cores) to the final decade (4 cores), are plotted in red. This series is a very good visual match to the "6-core composite" series in Panel (c) of Figure 1 above, including in particular, the unprecedentedly high value for the final decade of the 1990s. Our replication of the series in panel c) in Figure 7 of Thompson *et al.* (2003) is therefore reasonably successful, despite its absence from the data spreadsheet, and despite some inconsistencies in the Andean sub-index.

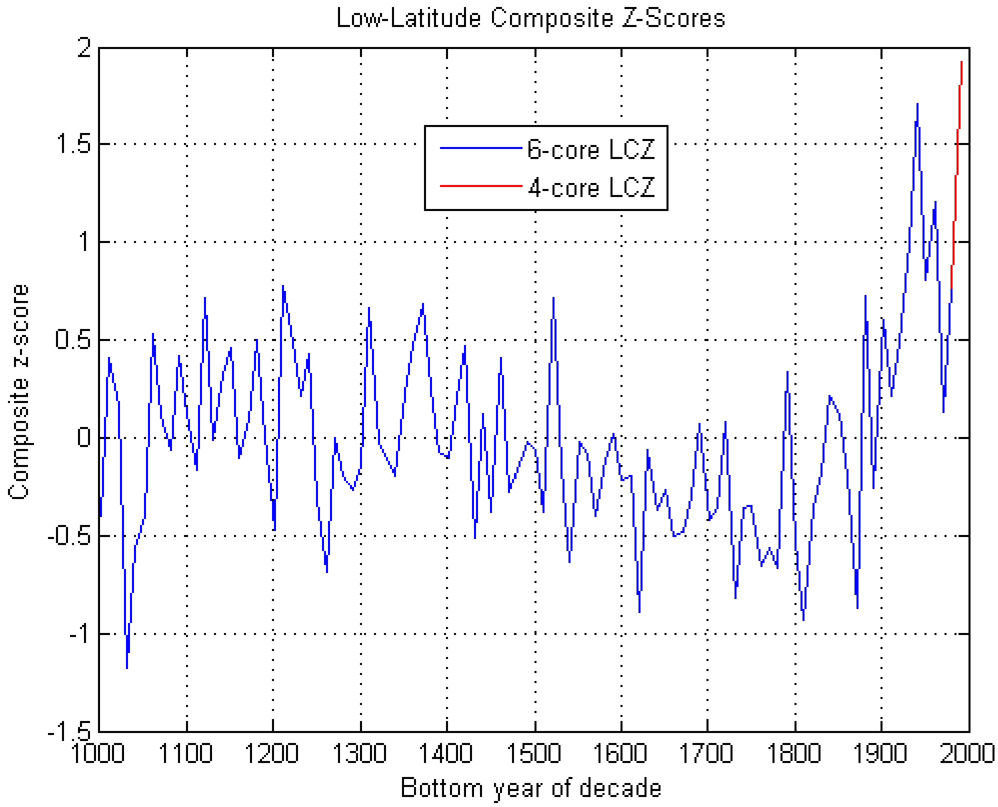

Fig. 8.
Low-Latitude Composite Z-Scores, based on 6 cores for 1001-1990 (in blue), and on 4 cores for 1991-2000 (in red). The line segment connecting 1981-1990 (6 cores) and 1991-2000 (4 cores) is also in red.

It is not clear why one would want to average Z-scores in this manner, since it makes inefficient use of the data. However, this appears to be a common practice in paleoclimatology (see *e.g.* Kaufman *et al.* 2009). The primary goal of the present paper is merely to replicate and calibrate the Thompson *et al.* (2003) composite ice core series, and not to improve upon it. Section V below suggests how this data might be used more efficiently.

It is singular that although the LCZ series in panel (c) of Figure 1 ends on an unprecedentedly high value for the 1990s, and panels (a) and (b) of Figure 1 above shows that this is being driven by its HCZ component rather than the ACZ component, none of the individual Himalayan Z-scores shown in Fig. 4 ends on a record high value. This only comes about as a statistical artifact, however, because two of the Himalayan series drop out in the last decade, leaving only Dasuopu, which was running higher or "warmer" than the other series to represent the region by itself. The effect is accentuated because when HCZ and ACZ are averaged together to obtain LCZ, the net weight on Dasuopu suddenly changes from 1/6 in the first 99 decades to 1/2 in the final decade.

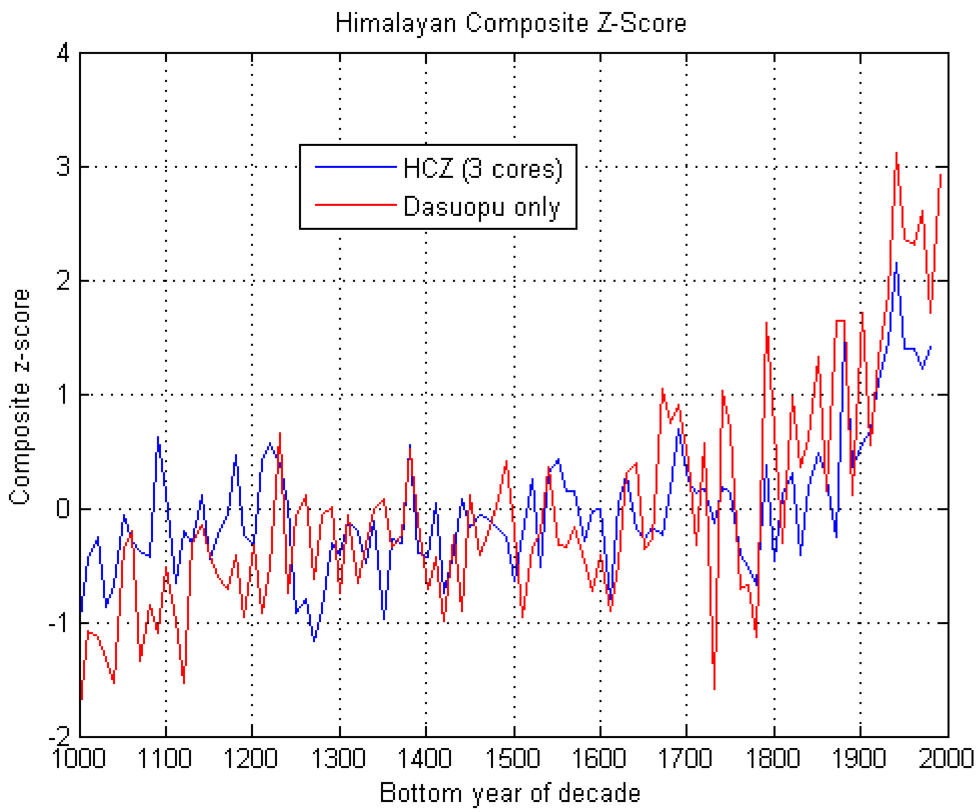

Fig. 9
Himalayan Composite Z-score (HCZ), computed from all 3 Himalayan cores (in blue), and from Dasuopu only (in red).

Figs. 9 and 10 show HCZ and LCZ for the full period 1001-2000 AD, computed both from all 6 cores (LCZ6) as well as from only the 4 cores that make it to the final decade of the study period (LCZ4). Clearly the 1990s are not any "warmer" than the 1940s when either HCZ or LCZ is computed on a consistent basis, despite the impression given in panels b) and c) of Fig. 1.

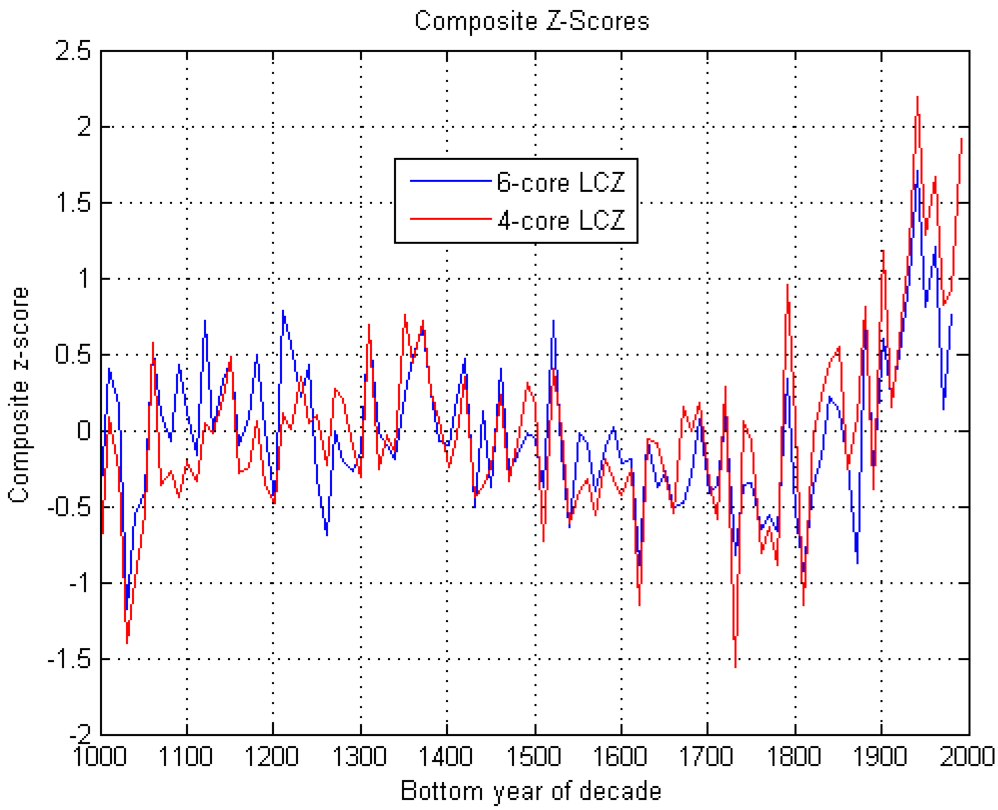

Fig. 10
Low-Latitude Composite Z-score (LCZ), computed from all 6 cores (in blue), and from only the 4 cores available in the last decade (in red).

The changing composition of LCZ would not alter its expected response to temperature if $\delta^{18}O_{ice}$ were a universally valid and linear indicator of annual average temperature, aside from a location shift to compensate for differences in latitude and/or altitude. However, some of the cores, such as Quelccaya Summit and Dasuopu, are strongly correlated with annual average global temperatures, while others, such as Sajama and Dunde, are not, so that the response is not universal. Some of the sites may be reflecting historical variations in the seasonality or altitude of snow formation, or in precipitation patterns as the air is transported to the site, rather than average annual temperatures, so that the correlation, if any, with local or global annual temperature can only be determined empirically on a site-by-site basis. The makeup of the composite series therefore can make a substantial difference for its correlation with temperature, so that in fact the final decade of LCZ (based on 4 cores with 1/2 weight on Dasuopu) is a quite different series than the preceding 99 decades (based on 6 cores with equal weights on all), and cannot be used in its calibration.

IV. Calibration to Global Temperature

Thompson *et al*. (2003) state that "Comparison of this [LCZ] ice core composite with the Northern Hemisphere proxy record (1000-2000 A.D.) reconstructed by Mann et al. (1999) and measured temperatures (1856-2000) reported by Jones et al. (1999) suggests the ice cores have captured the decadal scale variability in the global temperature trends." Of course, it would be circular to calibrate LCZ to the hockey stick and then to put it forward as independent confirmation of the hockey stick, so only the comparison to instrumental temperatures is relevant for our purposes.

Although Thompson et al. (2003) compare their LCZ favorably to a Northern Hemisphere instrumental temperature index, a global temperature index would more appropriate for univariate calibration, given that half of the six cores are from the Southern Hemisphere. Figure 11 below shows annual averages for 1850 – 2009 of the CRUTEM3vGL global land air index (variance adjusted) of Brohan *et al.* (2006), a global sequel to the Jones et al. (1999) series to which Thompson *et al.* compare their series.[9]

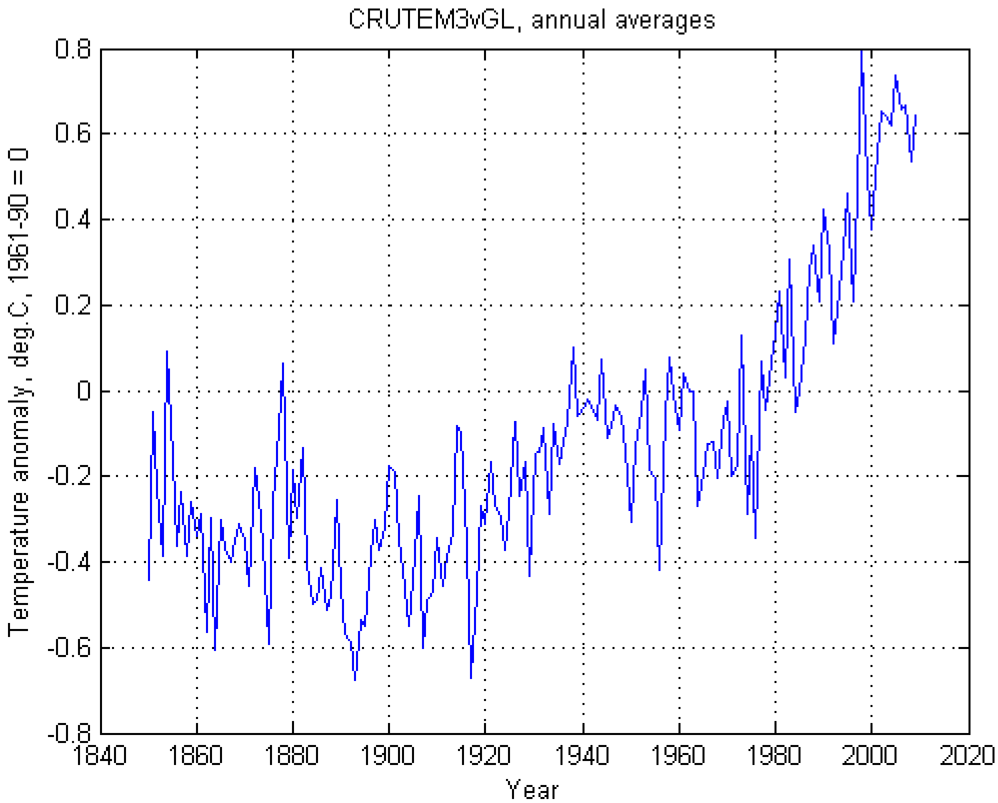

Fig. 11.
CRUTEM3vGL global land air temperature index, annual averages, 1850-Oct. 2009.

The widely used CRU series are partially based on confidential weather data that has not been made publicly available (see Met Office 2009). Since they are not scientifically replicable, they should only be used with caution. However, the alternative GISStemp series, produced by NASA/GISS, is only computed back to 1880, and hence has three fewer decades than the CRU series. Preliminary calculations show that the

9 Source <http://www.cru.uea.ac.uk/cru/data/temperature/crutem3vgl.txt>. The annual average plotted for 2009 is incomplete, running only through October. However, the calibration only uses the portion through 2000.

correlations with GISStemp are much weaker than with CRUTEM3vGL, but this is more likely due to the smaller sample size than to any difference in behavior since 1880.

Since half of the Thompson *et al.* decadal ice core series are based on decades *ending* in a year divisible by 10, while the other half use decades *beginning* in a year divisible by 10, it is a matter of indifference which convention is used for the temperature series. This paper arbitrarily adopts the former convention, i.e. 1851-1860 etc., and treats LCZ as if it were computed consistently on this basis.

Since the 4-core LCZ series LCZ4 does not necessarily have the same relationship to temperature as the 6-core series LCZ6, the two must be calibrated separately to temperature. Table 1 below shows the results of regressing the 14 decadal values of LCZ6 for 1851-1990 on the corresponding decadal averages of the temperature series. For purposes of comparison, the 4-core index LCZ4, using only Dasuopu for the Himalayan region with a 1/2 weight as in 1991-2000, was computed for all 15 post-1850 decades 1851-2000, and also regressed on the temperature series, as shown in the last two columns of the same table.

Table 1
Decadal Regression of 6- and 4- core LCZ on CRUTEM3vGL,
1851-1990 and 1851-2000, resp.

	LCZ6		LCZ4	
	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\alpha}$	$\hat{\beta}$
coef.	0.842	1.794	1.219	2.146
s.e.	0.256	0.934	0.195	0.692
t-stat	3.290	1.921	6.252	3.101
p(t-stat)	0.006	0.079	0.000	0.008
n	14		15	
R^2	0.235		0.425	
s^2	0.363		0.354	
$s^2 s_{\alpha\beta}$	0.186		0.083	

r_1	-0.005	-0.007
DW	1.948	2.005

It may be seen that there is a positive and weakly significant correlation between LCZ6 and instrumental temperature (2-tailed p = 0.079), and a positive and highly significant correlation between LCZ4 and temperature (2-tailed p = 0.008). The first order serial correlation coefficients of the residuals r_1 are essentially 0, and the Durbin-Watson statistics DW are very close to 2, so that positive serial correlation is not an issue with these regressions, at least not at this decadal frequency.

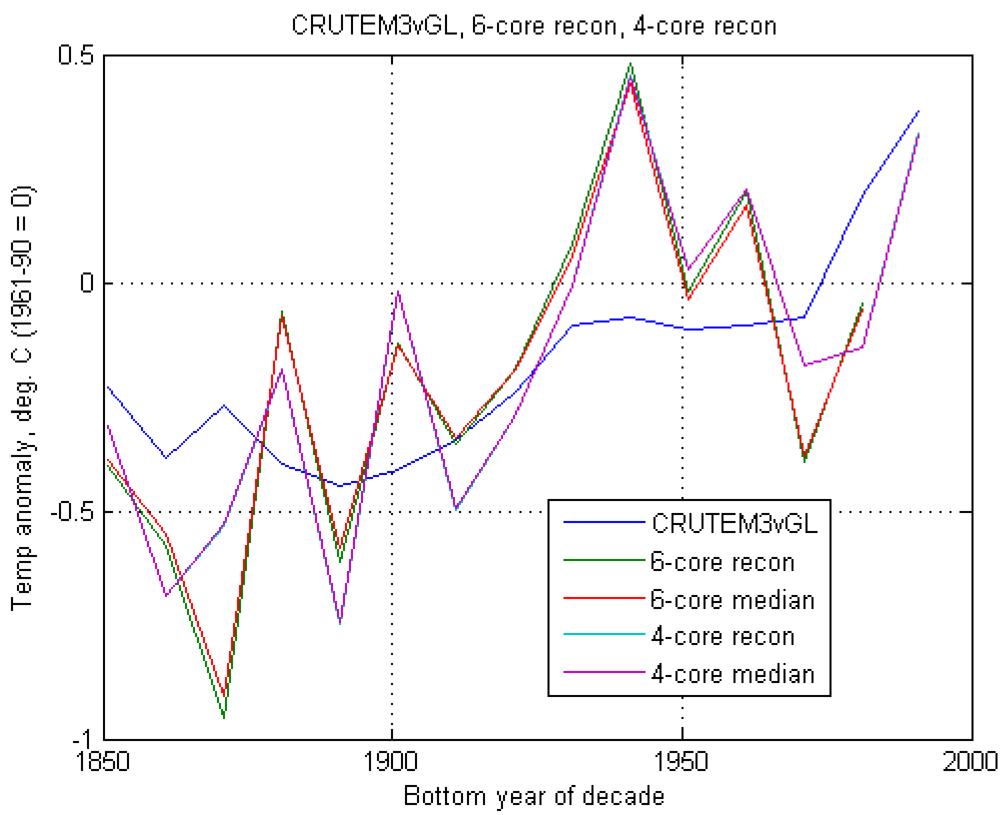

Fig. 12
CRUTEM3vGL global temperature index, with reconstructions using the 6-core LCZ and the 4-core LCZ, along with posterior median estimates.

Fig. 12 shows the decadal averages of CRUTEM3vGL used for the calibration, along with the 6- and 4-core reconstructed temperatures using the classical estimator (4). Also shown are the medians of the posterior distribution for the 6- and 4- core reconstruction. In most cases the classical reconstruction and posterior median are almost indistinguishable. Since the classical estimator (4) is much easier to compute and explain, it is natural to focus on it as the point estimate of reconstructed temperature. It is conjectured that the classical estimator coincides with the posterior mode.

Even though LCZ4 exceeded all previous values of LCZ6 during the decade of the 1990s, the temperatures reconstructed from it are somewhat lower than those of the 1940s using LCZ6. This comes about because the estimates of both the intercept and slope are higher for LCZ4 than for LCZ6 in Table 1. Both these factors work to reduce the reconstructed temperature as a function of the index.

Figure 13 shows the temperature reconstruction for 1001-2000, using the classical point estimate (4) rather than the posterior median. The reconstruction is based on LCZ6 for 1001-1990 (in blue), and on LCZ4 for 1991-2000 (in red). The line segment connecting 1981-1990 (6 cores) and 1991-2000 (4 cores) is also in red. Although the final decade has a decidedly warm point estimate, it is not as warm as the 1940s, contrary to the impression given in Figs. 1 and 8.

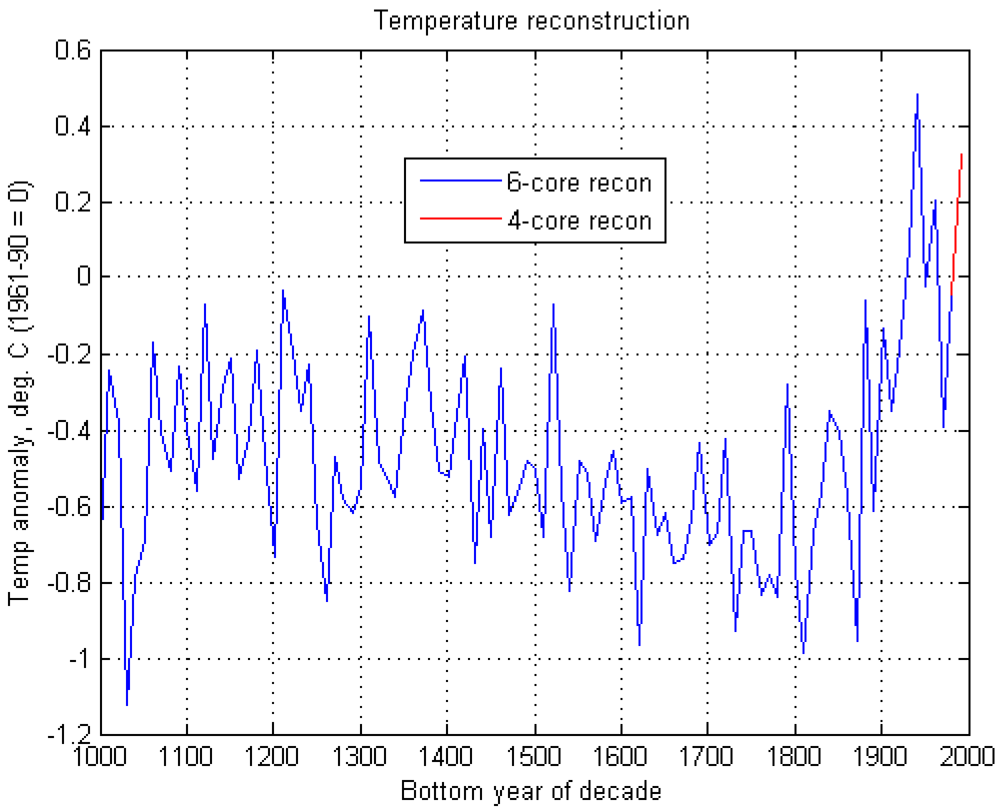

Fig. 13.

Reconstructed temperature point estimates, using (4) with LCZ6 for 1001-1990 (in blue) and LCZ4 for 1991-2000 (in red). The line segment connecting 1981-1990 (6 cores) and 1991-2000 (4 cores) is also in red.

Of course, since we have relatively precise instrumental temperatures for 1850 to the present, the reconstruction in Fig. 13 is not our best estimate of global temperatures for these years. The full reconstruction is provided here, and with its confidence intervals below, merely to show how the full "Dr. Thompson's thermometer" calibrates to temperature.

The posterior cumulative distribution functions (CDFs) for the LCZ6 reconstruction for the 99 decades 1001-1010 to 1981-1990 were computed in steps of .01

from -5.00 to +5.00 °C, as shown in Fig. 14 below.[10] Because the slope coefficient is only weakly significant for this reconstruction (p = .079), the tails of the posterior distributions are extremely heavy and slow to converge to 0 and 1 for low and high temperatures.

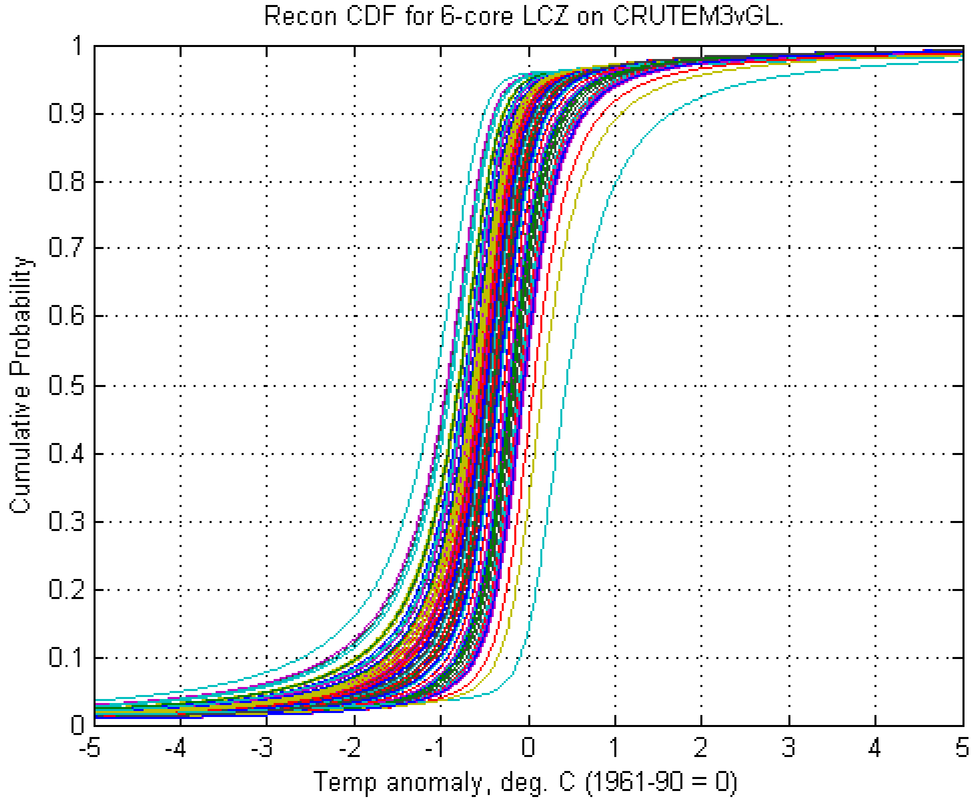

Fig. 14.
Posterior CDFs for 99 LCZ6 temperature reconstructions, 1001-1010 to 1981-1990.

Fig. 15 below shows the posterior CDF for the LCZ4 reconstruction, computed for the sake of comparison for the 15 calibration decades 1851-1860 to 1991-2000. Since the slope coefficient in this regression is highly significant (p = 0.008), the distribution is

[10] CRU (<http://www.cru.uea.ac.uk/cru/data/temperature/>) gives the precision of CRUTEM3vGL (2 standard errors) as approximately 0.05°C since 1951 and about 0.2°C in 1851, with gradual improvement from 1860 to 1950 except for wartimes, so that steps of 0.01°C are overkill. Nevertheless, using Matlab's *mvtcdf* as described in Appendix I, the program for all 100 reconstruction dates runs in just a few minutes on a desktop PC.

nearly Student t with 13 degrees of freedom, and the tails converge much more quickly to 0 and 1 for low and high temperatures than with LCZ6.

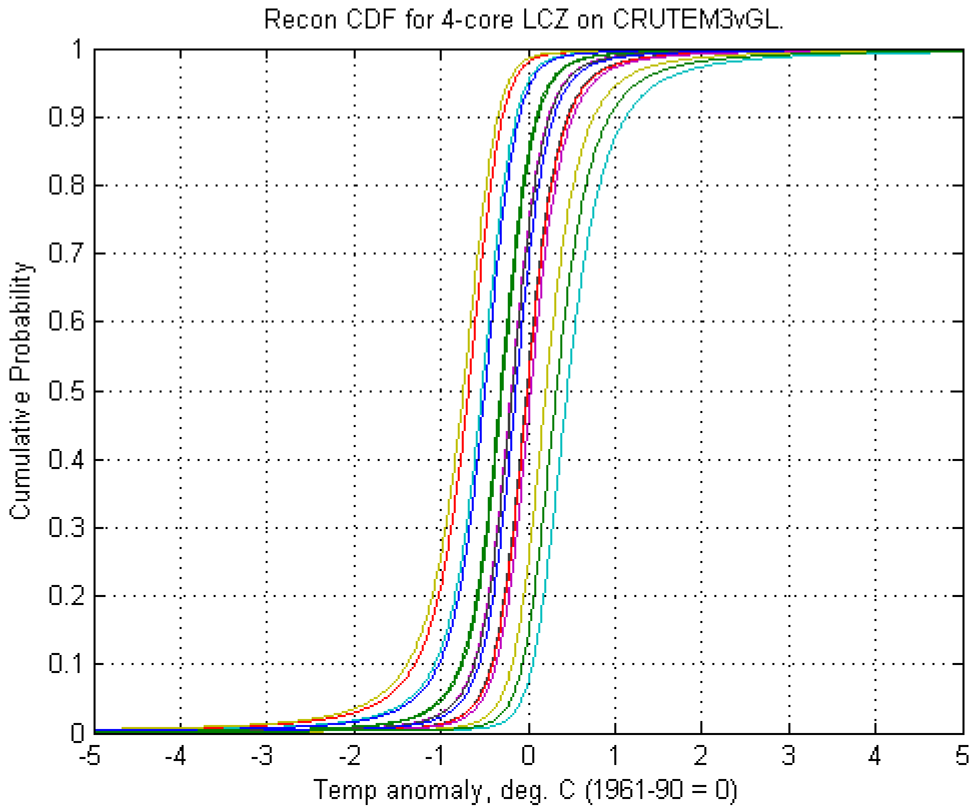

Fig. 15.
Posterior CDF for 15 LCZ4 reconstructions, 1851-1860 to 1991-2000.

The posterior CDFs in Figs. 14 and 15 were inverted by linear interpolation at probabilities 0.025, 0.25, 0.50, 0.75, and 0.0975. The resulting 50% and 95% LCZ6 Confidence Intervals (CIs) are plotted in blue for 1001-1990 in Fig. 16 below, along with the "classical" point estimates from Fig. 13. Corresponding values for LCZ4 are plotted in red for 1991-2000. The line segments connecting 1981-1990 (6 cores) and 1991-2000 (4 cores) are also shown in red.

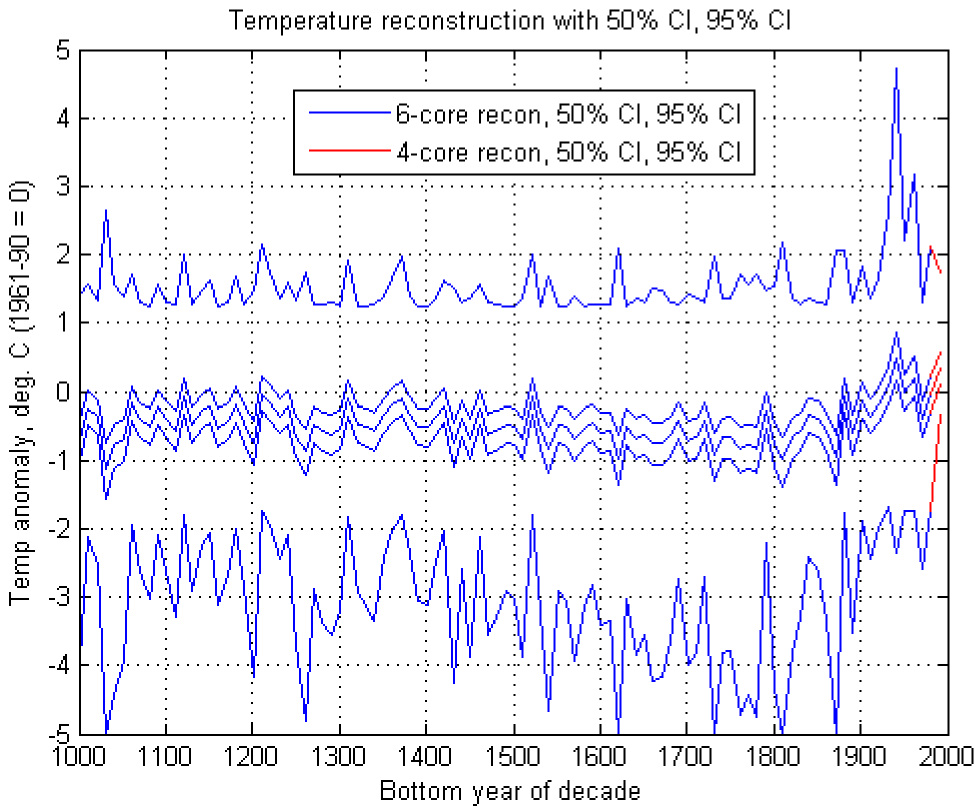

Fig. 16.
LCZ6 temperature reconstruction for 1001-1990, with 50% and 95% Confidence Intervals, in blue. LCZ4 temperature reconstruction for 1991-2000, with 50% and 95% Confidence Intervals, in red. Line segments connecting 1981-1990 (6 cores) and 1991-2000 (4 cores) are also in red.

The partial decade 2001-2009 was the warmest in the instrumental record used, at +0.64°C relative to 1961-90 = 0. It may be seen from Fig. 16 that temperatures throughout the period 1001-1850 could have been at least 1.2°C warmer than 1961-90, or at least 1.7°C colder. The Medieval Warm Period (MWP), or even the Little Ice Age (LIA) for that matter, therefore could well have been even warmer than the most recent decade, so far as this reconstruction goes. (Since 2001-2009 was not used in the calibration, it is not shown in Fig. 12.)

An even better Thompson $\delta^{18}O_{ice}$ series to calibrate should have been the 7-core index of Thompson *et al.* (2006). This index adds a seventh, Himalayan core, Puruogangri, and extends back 2000 years, where three of the cores still have data. By all rights, it should supersede the 6-core, 1000 year index of Thompson *et al.* (2003). Unfortunately, however, the supporting data sets that accompany that paper only tabulate the seven cores back 400 years, as five-year averages, so that the different subsets of the seven cores that are active in different periods cannot be separately calibrated, as they must be. Furthermore, the decadally averaged Z-score indices in Data Set 3 do not match the averages of the illustrative five-year average Z-score indices for either region that are tabulated in Data Set 2, so that there is no linear relationship between the archived 2000-year composite Z-score index and the illustrative 400-year $\delta^{18}O_{ice}$ data on which it is supposed to be based, as documented by McCulloch (2009a, 2009b). Three e-mails to Lonnie Thompson and most of his co-authors asking for the complete data and clarification of this inconsistency received no reply. The present study therefore focuses instead on the 1000-year 6-core $\delta^{18}O_{ice}$ index of Thompson *et al*. (2003), which is at least approximately replicable, in terms of its fully tabulated decadal component series.

V. Further Calibration Issues

This concluding section discusses a number of further calibration issues: i) the classical joint sampling confidence intervals discussed by Hoadley (1970) and Brown (1993), ii) efficient multiproxy calibration, iii) a more powerful sequential prior approach; and iv) efficient use of prior information about the sign of the calibration slope coefficient. No attempt is made at present to implement the proposed solutions to issues ii) – iv).

i) The traditional or "classical" alternative approach to calibration CI construction is based on Fieller's (1954) confidence region for the ratio of two Student t random variables. This approach first considers a 100(1-γ) CI for y' as a function of x', using the standard Student t distribution (16) for $\mathrm{p}\left(y' | x', \hat{\alpha}, \hat{\beta}, \mathbf{S}, s^2\right)$. These CIs bound y' by a pair of hyperbolic functions of x':

$$\hat{\alpha} + \hat{\beta}x' - t^c \,\mathrm{v}(x')^{1/2} \le y' \le \hat{\alpha} + \hat{\beta}x' + t^c \,\mathrm{v}(x')^{1/2},$$

where $t^c = \mathrm{T}_{n-2}^{-1}(1-\gamma/2)$ is the Student t 2-tailed critical value for test size γ, $\mathrm{T}_\nu(t)$ is the standard Student t CDF with ν degrees of freedom, and $\mathrm{v}(x')$ is as defined in (17). Let $p = 2\left(1 - T_{n-2}\left(\hat{\beta} / \left(s\sqrt{s_{\beta\beta}}\right)\right)\right)$ be the two-tailed "*p*-value" of the test statistic for the hypothesis that $\hat{\beta} = 0$. If $p < \gamma$ so that the slope is significantly different from 0 at level γ, solving the boundaries for x' as a function of the observed proxy value y' yields a quadratic equation with two real roots $\xi_1 < \xi_2$ where y' intersects the two hyperbolas, and the confidence region is the bounded interval (ξ_1, ξ_2). But if $p = \gamma$, this interval becomes either (ξ_1, ∞), ($-\infty$, ξ_2) or ($-\infty$, ∞), depending on the signs of $\hat{\beta}$ and $y' - \bar{y}$, where $\bar{y}$ is the mean of the calibration y_i-values. And if $p > \gamma$, so that the slope is insignificantly different from 0 at level γ, either y' twice intersects one of the hyperbolas and the confidence region is the discontiguous set (∞, ξ_2) ∪ (ξ_1, ∞), or else there are no real roots, in which case the confidence region is the entire real line. Hoadley (1970) dryly remarks that this set "possesses inherent difficulties."

Brown (1982) proposes that (ξ_1, ξ_2) is a "respectable interval provided the t-test of the hypothesis $\beta = 0$ is rejected." Indeed, if say $\hat{\beta} > 0$ and it were known *with perfect certainty* that β has the same sign, then for any $x' < \xi_1$, the probability that y' could have been as high as its observed value would be less than or equal to $\gamma/2$, while for any $x' > \xi_2$, the probability that y' could have been as low as its observed value would also be less than or equal to $\gamma/2$. The hitch, however, is that there in fact is always probability $p/2$ that β and $\hat{\beta}$ have *opposite* signs.

The Bayesian approach, on the other hand, takes full account of the possibility that β and $\hat{\beta}$ might have opposite signs, and never gives unbounded CIs, let alone discontiguous confidence regions. When $p > \gamma$, as in the LCZ6 example above, the Bayesian CI is simply somewhat wider than it otherwise would be, as is only natural.

Fig. 17 compares the posterior and Fieller 50% CIs for our LCZ calibration. Since for LCZ6, $p = .079 << .50$, the Fieller 50% region is an interval and the two are very similar. For the final decade, based on LCZ4, $p = .0008$, so that the similarity is even closer. In general, the classical CI is a good approximation to the Bayesian CI when $p << \gamma$, so that there is only negligible chance, relative to γ, that β and $\hat{\beta}$ can have opposite signs. Indeed, Hunter and Lamboy (1981) themselves actually advocate it as an approximation.

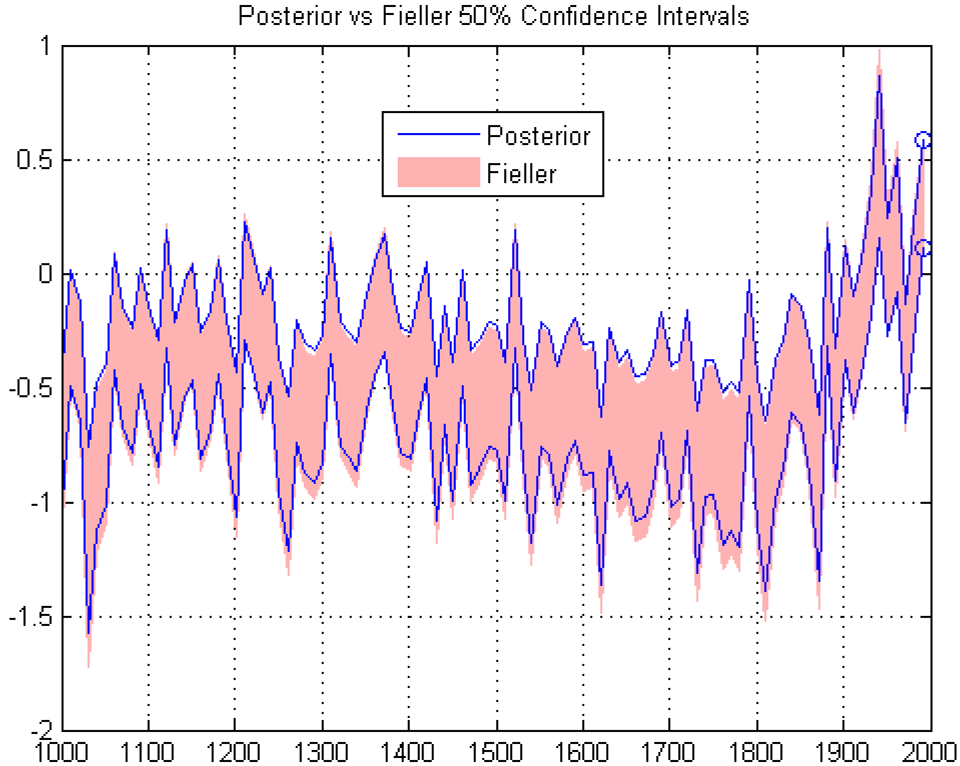

Fig. 17.
Comparison of posterior (lines) and Fieller (shaded) 50% confidence intervals.

However, when the slope is only marginally significant, or even insignificant, relative to the desired γ, as is often the case in paleoclimate contexts, the exact Bayesian interval is to be preferred. Fig. 18 compares the posterior and Fieller 95% confidence regions for our LCZ calibration. Since for LCZ6, we now have $p = .079 > \gamma = .05$, the Fieller region is unreasonably either the entire real line or else a pair of semiinfinite intervals. For the final decade, however, $p = .0008 << .05$, so that the Fieller 95% region becomes an interval, and is a good approximation to the posterior interval.

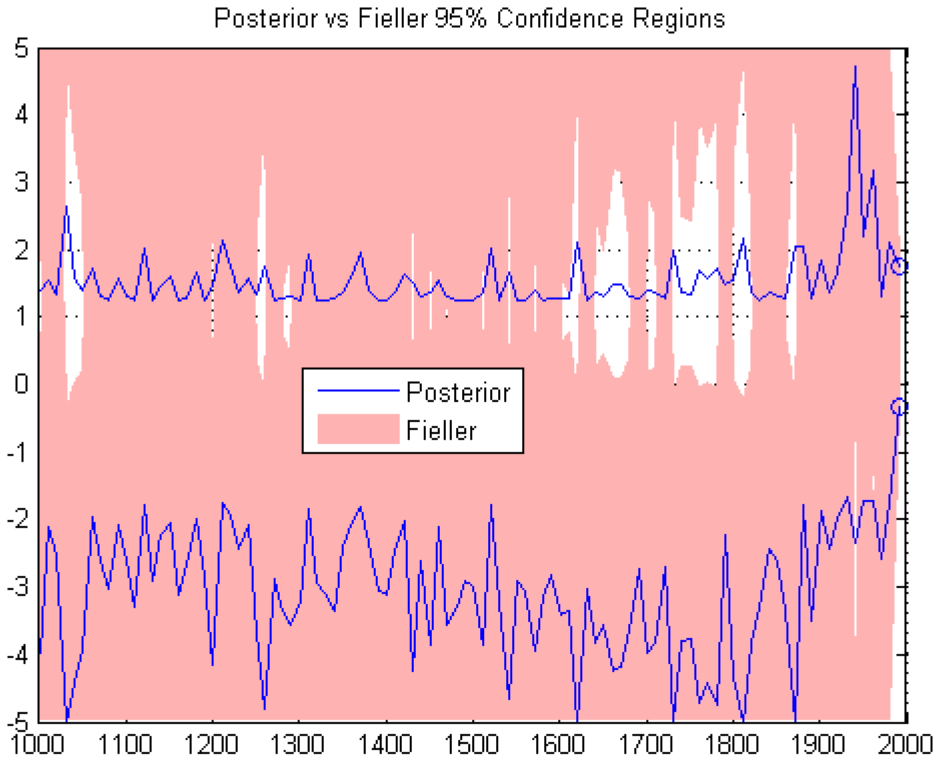

Fig. 18.

Comparison of posterior (lines) and Fieller (shaded) 95% confidence regions.

ii) When several proxies are available, and there is an *a priori* expectation that they will have the same quantitative relationship to the state variable net of a location shift, their average may simply be univariately calibrated to the state variable. However, when, as in the case of the Thompson *et al.* (2003) ice core data, the individual proxies respond unequally to the state variable, and sometimes not at all, a multiproxy approach should be more efficient than arbitrarily averaging either the raw proxies themselves, or their Z-scores as in Thompson *et al.* (2003, 2006) and Kaufman *et al.* (2009). Of course, one may not simply disregard the insignificant proxies, or those that give the "wrong" sign slope, for then there will be selection bias that will invalidate conventional significance tests.

Suppose there are q proxies, with y_{ij} being the i-th calibration-period observation on the j-th proxy, and y'_j being the reconstruction-period observation on the j-th proxy. Assume first that

$$y_{ij} = \alpha_j + \beta_j x_i + \varepsilon_{ij}\,, \tag{20}$$

where $\varepsilon_{ij} \sim N(0,\sigma_i^2)$ and are independent across proxies. Let $\mathbf{y}_j = (y_{1j},\ldots,y_{nj})^{\mathrm{T}}$ and $\mathbf{x} = (x_1,\ldots,x_n)^{\mathrm{T}}$, so that $\hat{\alpha}_j$, $\hat{\beta}_j$, and $s_j^2 = \hat{\sigma}_j^2$ are implicit in $\mathbf{y}_j$ and $\mathbf{x}$, and $\mathrm{p}(x' \mid y'_j, \mathbf{x}, \mathbf{y}_j)$ may be computed under a diffuse prior as in Section 2 of the text . Then under a diffuse prior for x',

$$\mathrm{p}(x' \mid y'_1,\ldots y'_q, \mathbf{x}, \mathbf{y}_1,\ldots \mathbf{y}_q) \propto \prod_{j=1}^{q} \mathrm{p}(x' \mid y'_j, \mathbf{x}, \mathbf{y}_j). \tag{21}$$

If the regression errors are *not* independent across proxies, as preliminary calculations show to be the case with the Thompson *et al.* ice core data, then (21) is invalid. However, if we have $q \leq n$-2 proxies, we may obtain errors that are uncorrelated across proxies, and therefore independent under our Gaussian assumption, by instead regressing each proxy $j \geq 2$ on the state variable plus proxies 1 ... j-1 during the calibration period i = 1, ... n, as follows:

$$y_{ij} = \alpha_j + \beta_j x_i + \sum_{h=1}^{j-1} \gamma_{jh} y_{ih} + \varepsilon_{ij}\,.$$

The calibration OLS estimates of these parameters are implicit in the partial calibration data $\mathbf{y}_1$ ... $\mathbf{y}_j$ and $\mathbf{x}$. Using the reconstruction values of the proxies, define the orthogonal innovation of proxy j and its estimate by

$$y_j^* = y'_j - \sum_{h=1}^{j-1} \gamma_{jh} y'_h,$$

$$\hat{y}_j^* = y'_j - \sum_{h=1}^{j-1} \hat{\gamma}_{jh} y'_h$$

for $j \geq 2$ and $y_1^* = \hat{y}_1^* = y_1$. Then $\mathrm{p}\left(x' \mid \hat{y}_j^*, \mathbf{x}, \mathbf{y}_1, \ldots \mathbf{y}_j\right)$ may be computed as in (11), simply by adding the appropriate terms to (6) – (10). Equation (21) above then generalizes to

$$\mathrm{p}\left(x' \mid y'_1, \ldots y'_q, \mathbf{x}, \mathbf{y}_1, \ldots \mathbf{y}_q\right) \propto \prod_{j=1}^{q} \mathrm{p}\left(x' \mid \hat{y}_j^*, \mathbf{x}, \mathbf{y}_1, \ldots \mathbf{y}_j\right). \tag{22}$$

Equation (21) or (22) may be evaluated numerically either by calculating the R2T densities directly as in Hunter and Lamboy (1981) or by finite differences from the CDF, and then numerically integrating the product of posterior densities to normalize and obtain CIs. A transformation such as Student t with 2*q*-1 degrees of freedom may be useful in order to capture the entire tails.

When the proxies are significantly correlated with the state variable even after conditioning on the preceding proxies, and if they tell a consistent story about the reconstruction state variable, the multiproxy posterior distribution (22) will tend to be much tighter about its mode than the single-proxy reconstructions would be. But when the proxies are significant yet tell an *in*consistent story about x', the multiproxy posterior density will tend to be more spread out, or even multimodal. CIs computed from this distribution will then tend to be wide enough to include all, or at least most, of the individual modes.

As long as all the proxies have the same coverage across observations, the order in which they are arranged should make no difference for the multiproxy posterior distribution. However, if they have different coverage, as is the case for the calibration period with two of the Thompson *et al.* (2003) proxies, and as is the case during the 2000-year reconstruction period for several of the proxies in Thompson *et al.* (2006), it is important to arrange them in decreasing order of coverage, so that all the required conditioning proxies are available at each step. If a proxy has more coverage in the reconstruction period than the others but less in the calibration period, or vice versa, it may unfortunately be necessary to give up some data points.

In the multiproxy case, the point estimate of x' may be taken as the posterior median, since if the proxies are inconsistent there may be multiple posterior modes. The GLS estimator suggested by Brown (1982: eq. 2.16), which generalizes the single-proxy classical estimator in equation (4), is far simpler to compute. However, it may be problematic because it is essentially a weighted average of single proxy estimates, each of which has Cauchy-like heavy tails with undefined mean. The Law of Large Numbers therefore does not ensure that the distribution of the average will be more compact than the distributions of the contributing values.

When the proxies do not have independent simple regression errors and $q \geq n$-1, as is for example the case with the Kaufman *et al.* (2009) data, (22) becomes invalid, since eventually zero and even negative degrees of freedom will be encountered. In this case, it may be appropriate to reduce the dimensionality of the proxy data set by careful use of Principal Components Analysis (Preisendorfer 1988, Mann *et al*. 1999, McIntyre and McKitrick 2005). Such an approach may also be preferred even when $q < n$-2, if n/q is not large.

iii). Section II above treats the problem of inferring a single reconstruction value for an observation that has no sequential relationship to the calibration data. In a time series or other sequential context, however, there may be valuable information from what is known or has been inferred about the adjacent observations.

In particular, let us suppose that the calibration state observations x_t, now with a subscript t = 1, ... n to indicate time, follow a random walk so that

$$x_t = x_{t-1} + \eta_t , \qquad (23)$$

where η_t ~ iid N(0, τ^2) and are independent of the regression errors. The random walk "signal" variance may be estimated from the calibration data by

$$\begin{aligned}\hat{\tau}^2 &= \sum_{t=2}^{n} \left(x_t - x_{t-1}\right)^2 / (n-1) \\ &\sim \tau^2 \chi^2_{n-1} / (n-1).\end{aligned}$$

Suppose we have a time series of calibration proxy values y_t, t = 1, ... n, as well as an adjoining time series of reconstruction proxy values y'_t, t = 1, ... T, with $y'_{T+1} = y_1$ etc., as is ordinarily the case in paleoclimate temperature reconstructions. We wish to infer the T state values x'_t, t = 1, ... T, which also follow the random walk (23), with $x'_{T+1} = x_1$ etc. Let $\mathbf{x}'$ and $\mathbf{y}'$ represent the corresponding vectors of reconstruction values.

The random walk implies that the distribution of the state variable k periods before x_1 or k periods after x_n, conditional on the calibration data, has a normal distribution with mean x_1 or x_n and variance $k\tau^2$, and that the unconditional distribution has infinite variance. It therefore motivates the improper diffuse prior assumption of Section II for an isolated reconstruction date that is so far from the calibration data that the calibration values become uninformative. However, when the reconstruction date is either close to the calibration data or surrounded by other reconstruction dates, conditioning on this information can greatly improve the precision of the reconstruction.

Conditional on the true parameter values, the reconstruction proxy values, and the calibration state vector, the posterior distributions of the x_t' are Gaussian, with means and variances determined by a simple Kalman smoother (e.g. Harvey 1989, McCulloch 2005). This meshes the information in a forward Kalman filter that starts at $t = 1$ with a diffuse prior as in Section II, with the information in a reverse Kalman filter that starts at $t = T$ with a prior that is governed by x_1, and works backwards. Then the true parameters may be integrated out to find the distributions of the x_t' conditional on the parameter estimates.

To simplify the notation, the following development of the Kalman filter and smoother is implicitly conditional on the true parameter values α, β, σ^2, and τ^2, until otherwise indicated.

The forward Kalman filter is initialized with a diffuse prior as in section II, so that as in (7),

$$\begin{aligned} \mathrm{p}(x_1' \mid y'_1) &\sim N\left((y_1' - \alpha)/\beta, \sigma^2/\beta^2\right) \\ &= N(c_1 + d_1\alpha, v_1), \end{aligned}$$

with

$$c_1 = y_1'/\beta, \quad d_1 = -1/\beta, \quad v_1 = \sigma^2/\beta^2 .$$

Now suppose that for any $t = 2, \ldots T$, the mean of filter density for t-1 is likewise affine in α:

$$\mathrm{p}(x_{t-1}' \mid y_1', \ldots y_{t-1}') \sim N(c_{t-1} + d_{t-1}\alpha,\, v_{t-1}) .$$

Then (23) implies that the predictive density for time t is

$$\mathrm{p}\left(x_t' \mid y_1',\ldots y_{t-1}'\right) \sim N\left(c_{t-1} + d_{t-1}\alpha,\, v_{t-1} + \tau^2\right).$$

Bayes' Rule then implies that the filter density for *t* itself is also affine in α:

$$\begin{aligned}
\mathrm{p}\left(x_t' \mid y_1',\ldots y_t'\right) &\propto \mathrm{p}\left(y_t' \mid x_t', y_1',\ldots y_{t-1}'\right)\mathrm{p}\left(x_t' \mid y_1',\ldots y_{t-1}'\right) \\
&= \mathrm{p}\left(y_t' \mid x_t'\right)\mathrm{p}\left(x_t' \mid y_1',\ldots y_{t-1}'\right) \\
&\sim N\left(\left(y_t' - \alpha\right)/\beta,\, \sigma^2/\beta^2\right)N\left(c_{t-1} + d_{t-1}\alpha,\, v_{t-1} + \tau^2\right) \\
&\sim N\left(c_t + d_t\alpha,\, v_t\right),
\end{aligned}$$

with

$$\begin{aligned}
v_t &= \left(\beta^2/\sigma^2 + \left(v_{t-1} + \tau^2\right)^{-1}\right)^{-1} \\
g_t &= \left(\beta^2/\sigma^2\right)v_t, \\
c_t &= g_t\left(y_t'/\beta\right) + \left(1 - g_t\right)c_{t-1}, \\
d_t &= g_t\left(-1/\beta\right) + \left(1 - g_t\right)d_{t-1}.
\end{aligned}$$

The reverse Kalman filter works by backward induction, beginning with time T, using a prior based on the observed value of $x_1 = x_{T+1}'$ and (23):

$$\mathrm{p}\left(x_T' \mid \mathbf{x}\right) = \mathrm{p}\left(x_T' \mid x_1\right) \sim N\left(x_1, \tau^2\right)$$

Bayes' Rule then implies that the reverse filter for t = T is again affine in α:

$$\begin{aligned}
\mathrm{p}\left(x_T' \mid y_T', \mathbf{x}\right) &\propto \mathrm{p}\left(y_T' \mid x_T', \mathbf{x}\right)\mathrm{p}\left(x_T' \mid \mathbf{x}\right) \\
&= \mathrm{p}\left(y_T' \mid x_T'\right)\mathrm{p}\left(x_T' \mid \mathbf{x}\right) \\
&\sim N\left(\left(y_T' - \alpha\right)/\beta,\, \sigma^2/\beta^2\right)N\left(x_1,\, \tau^2\right) \\
&\sim N\left(c_T^* + d_T^*\alpha,\, v_T^*\right); \\
v_T^* &= \left(\beta^2/\sigma^2 + 1/\tau^2\right)^{-1}, \\
g_T^* &= \left(\beta^2/\sigma^2\right)v_T^* \\
c_T^* &= g_T^*\left(y_T'/\beta\right) + \left(1 - g_T^*\right)x_1, \\
d_T^* &= g_T^*\left(-1/\beta\right).
\end{aligned}$$

Reasoning analogous to that for the forward filter then implies that the reverse filter is likewise affine in α for all t:

$$
\begin{aligned}
\mathrm{p}\left(x_t' \mid y_t',\ldots y_T',\mathbf{x}\right) &\sim N\left(c_t^* + d_t^*\alpha,\, v_t^*\right); \\
v_t^* &= \left(\beta^2/\sigma^2 + \left(v_{t+1}^* + \tau^2\right)^{-1}\right)^{-1} \\
g_t^* &= \left(\beta^2/\sigma^2\right)v_t^*, \\
c_t^* &= g_t^*\left(y_t'/\beta\right) + \left(1-g_t^*\right)c_{t+1}^*, \\
d_t^* &= g_t^*\left(-1/\beta\right) + \left(1-g_t^*\right)d_{T+1}^*.
\end{aligned}
$$

Finally, merging together the independent information in $\mathrm{p}\left(x_t' \mid y_1',\ldots y_t'\right)$ and $\mathrm{p}\left(x_t' \mid y_{t+1}',\ldots y_T',\mathbf{x}\right) \sim N\left(c_{t+1}^* + d_{t+1}^*\alpha,\, v_{t+1}^* + \tau^2\right)$ yields the Kalman smoother

$$
\begin{aligned}
\mathrm{p}\left(x_t' \mid \mathbf{y}',\mathbf{x},\alpha,\beta,\sigma^2,\tau^2\right) &\sim N\left(C_t + D_t\alpha,\, V_t\right); \\
V_t &= \left(1/v_t + 1/\left(v_{t+1}^* + \tau^2\right)\right)^{-1}, \\
G_t &= V_t/v_t, \\
C_t &= G_t c_t + \left(1-G_t\right)c_{t+1}^*, \\
D_t &= G_t d_t + \left(1-G_t\right)d_{t+1}^*.
\end{aligned}
\tag{24}
$$

Smoother point estimates $\hat{x}_t'$ analogous to the classical pointwise estimator (4) may be found by evaluating the mean $C_t + D_t\alpha$ using the parameter estimates in place of the parameters themselves.

In order obtain valid confidence intervals, it remains to integrate out the unknown parameters. Unfortunately, this is no longer determined by the R2N distribution, even conditional on the two variances, since β now enters in a complicated way into C_t, D_t, and V_t. However, the dependence on the intercept α is still simple, since it only appears in the affine expression for the mean. Equation (9) implies

$$
\mathrm{p}\left(\alpha \mid \beta,\hat{\alpha},\hat{\beta},\mathbf{S},\sigma^2,\tau^2\right) \sim \mathrm{N}\left(\hat{\alpha} + \left(s_{\alpha\beta}/s_{\beta\beta}\right)\left(\beta-\hat{\beta}\right),\, \left(1-\rho_{\alpha\beta}\right)\sigma^2 s_{\alpha\alpha}\right),
$$

where $\rho_{\alpha\beta} = s_{\alpha\beta}/\left(s_{\alpha\alpha}s_{\beta\beta}\right)^{1/2}$. Therefore[11]

---

[11] Note that the information set $\{\alpha,\beta\}$ is equivalent to $\{\alpha,\beta,\hat{\alpha},\hat{\beta}\}$, since once we know both α and β, $\hat{\alpha}$ and $\hat{\beta}$ add no information. Therefore the information set $\{\alpha,\beta\}$ may be taken to implicitly include $\{\hat{\alpha},\hat{\beta}\}$.

$$\mathrm{p}\left(x_t' \mid \mathbf{y}', \mathbf{x}, \hat{\alpha}, \beta, \hat{\beta}, \mathbf{S}, \sigma^2, \tau^2\right) \sim \mathrm{N}\left(C_t + D_t\left(\hat{\alpha} + \left(s_{\alpha\beta} / s_{\beta\beta}\right)\left(\beta - \hat{\beta}\right)\right),\ V_t + D_t^2\left(1 - \rho_{\alpha\beta}\right)\sigma^2 s_{\alpha\alpha}\right) \qquad (25)$$

The remaining three parameters, β, σ^2, and τ^2, must be integrated out numerically to obtain $\mathrm{p}\left(x_t' \mid \mathbf{y}', \mathbf{x}, \hat{\alpha}, \hat{\beta}, \mathbf{S}, s^2, \hat{\tau}^2\right)$, using

$$\mathrm{p}\left(\beta, \sigma^2, \tau^2 \mid \hat{\beta}, s_{\beta\beta}, s^2, \hat{\tau}^2\right) = \mathrm{p}\left(\beta \mid \hat{\beta}, s_{\beta\beta}, \sigma^2\right)\mathrm{p}\left(\sigma^2 \mid s^2\right)\mathrm{p}\left(\tau^2 \mid \hat{\tau}^2\right)$$

and a uniform prior on log(τ^2). These integrals may be performed using *ternary integration*, as described in Appendix II. Although this is a triple integral, it is probably adequate to use ternary integration with only a modest number of points over each parameter, *e.g.* m = 10 for β and m = 4 for the two variances, making a total of 160 combinations to evaluate. (In fact, the sequential structure of the state variable means that the proxy data contain some indirect information about σ^2 and τ^2 that is not contained directly in s^2 or $\hat{\tau}^2$. However, the gain from exploiting this indirect information is likely to be small.)

The resulting sequential reconstruction will, by construction, equal the observed state data during the calibration period. In order to make this trivial relationship clear, the calibration portion should be plotted with a different color or symbol than the nontrivial portion. At the end of the reconstruction period, the reconstruction will be centered nearly on x_1, with a very tight distribution. Moving back into the reconstruction period, however, the point estimates will begin to look like a smoothed version of the pointwise reconstruction, with smooth CIs that gradually widen. However, because the information in adjacent proxy values is being taken into account, the CIs will never become as wide as those for the pointwise reconstruction. Simply smoothing the pointwise CIs is therefore not equivalent to computing a CI for the smoothed reconstruction. Furthermore, since the Kalman smoother point estimate is already optimally smoothed according to the empirical signal/noise ratio, there is no need for any *ad hoc* further smoothing of the point estimate.

In order to illustrate how informative the proxy data is by itself (conditional on the calibration regression but otherwise not using the calibration state values), the reconstruction may also be constructed using a diffuse prior for x'_{T+n} in the reverse filter. This will resemble a smoothed version of the pointwise reconstruction clear up to the end of the calibration period as in Figure 19 above, since the pointwise reconstruction takes no account of the time series relationship between the calibration values and the reconstruction values.

Tingley and Huybers (TH 2009) propose a similar sequential Bayesian estimator, with the addition features of being multiproxy and spatially disaggregated, as well as explicitly treating the instrumental data as containing measurement error. However, TH assume that the state variable follows a potentially stationary AR(1) process with unknown AR parameter. Although this model nests the random walk (23), its likelihood is 0 when the AR process has a unit root, so that it will only give posterior density to stationary values, with the result that the reconstruction will be unnecessarily tied to the mean of the calibration state variables.

Also, rather than exploiting the Kalman filter/smoother to integrate out the state variables and one of the hyperparameters as in (25) above, thereby leaving only three hyperparameters to integrate out deterministically, TH perform the entire normalizing integration randomly using a Gibbs sampler. There are some problems for which Monte

Carlo integration is the only practical solution, but it would only add unnecessary sampling error to the calculation the simple cases discussed here.

iv. If the slope coefficient is known *a priori* to be nonnegative, (7) and (11) in the pointwise single-proxy case of Section II may simply be evaluated with the restricted R2N formula (27) or restricted R2T formula (29) in Appendix I in place of the unrestricted formula (26) or (28). The pointwise multiproxy case of Section V.ii above is similar.

In the single proxy sequential case of section V.iii, β must already be integrated out numerically in (25), so that all that is necessary, when the slope is assumed to be nonnegative, is to replace its Gaussian distribution (conditional on σ^2) with the conditional counterpart.

Data and Matlab programs for this project are online via http://www.econ.ohio-state.edu/jhm/AGW/Thompson6/

APPENDIX I

The Distribution of the Ratio of Two Normal or Student t Random Variables

The distribution of the ratio of two correlated normal random variables was developed by Fieller (1932) but is re-developed here just for fun (and somewhat more simply). The formula is easily extended to the ratio of two Student t random variables.

Let $X = Y/Z$, where Y and Z have the bivariate normal distribution

$$(Y, Z) \sim \mathrm{N}((\mu_Y, \mu_Z), \mathbf{\Sigma}_{YZ}), \qquad \mathbf{\Sigma}_{YZ} = \begin{pmatrix} \sigma_{YY} & \sigma_{YZ} \\ \sigma_{YZ} & \sigma_{ZZ} \end{pmatrix}.$$

For any real x, define

$$W = xZ - Y$$

so that

$$\begin{aligned} \mu_W &= x\mu_Z - \mu_Y \\ \sigma_{WZ} &= x\sigma_{ZZ} - \sigma_{YZ} \\ \sigma_{WW} &= x^2\sigma_{ZZ} + \sigma_{YY} - 2x\sigma_{YZ} \end{aligned}$$

Then the CDF of X is

$$\begin{aligned} \mathrm{F}(x) &= \mathrm{P}(X < x) \\ &= \mathrm{P}(Y < xZ,\ Z > 0) + \mathrm{P}(Y > xZ,\ Z < 0) \\ &= \mathrm{P}(W > 0,\ Z > 0) + \mathrm{P}(W < 0,\ Z < 0) \\ &= \Phi((0,0), -(\mu_W, \mu_Z), \mathbf{\Sigma}_{WZ}) + \Phi((0,0), (\mu_W, \mu_Z), \mathbf{\Sigma}_{WZ}), \end{aligned} \qquad (26)$$

where $\Phi((x, y), (\mu_X, \mu_Y), \mathbf{\Sigma}_{XY})$ is the bivariate normal CDF with general mean vector and covariance matrix. Matlab function *mvncdf*, based on Genz(2004), accommodates non-zero means and non-unit variances, so that it may be applied directly to (26), without a location-scale transformation.

Hinckley (1969) provides the formula for the density, but this is not required for the present problem of computing confidence intervals.

If the denominator Z (in the present context β) is known to be nonnegative, we instead have

$$\begin{aligned}
G(x) &= P(X < x \mid Z > 0) \\
&= P(X < x, Z > 0) / P(Z > 0) \\
&= P(Y < xZ, Z > 0) / P(Z > 0) \\
&= P(W > 0, Z > 0) / P(-Z < 0) \\
&= \Phi((0,0), -(\mu_W, \mu_Z), \boldsymbol{\Sigma}_{WZ})) / \Phi\left(\mu_Z / \sqrt{\sigma_{ZZ}}\right),
\end{aligned} \tag{27}$$

where $\Phi(z)$ (with one scalar argument) is the standard univariate normal CDF.

If instead (Y, Z) have the general bivariate Student t distribution with means (μ_Y, μ_Z), estimated covariance matrix $\boldsymbol{\Sigma}_{YZ}$, ν degrees of freedom, and CDF $T_\nu((y,z), (\mu_Y, \mu_Z), \boldsymbol{\Sigma}_{YZ})$ (see Zellner 1971, app. B.2; Genz 2004), then by the same reasoning,

$$\begin{aligned}
F(x) &= P(X < x) \\
&= T_\nu((0,0), -(\mu_W, \mu_Z), \boldsymbol{\Sigma}_{WZ})) + T_\nu((0,0), (\mu_W, \mu_Z), \boldsymbol{\Sigma}_{WZ})) \\
&= T_\nu((w,z), \mathbf{C}_{WZ}) + T_\nu(-(w,z), \mathbf{C}_{WZ}),
\end{aligned} \tag{28}$$

where $w = \mu_W / \sigma_{WW}^{1/2}$, $z = \mu_Z / \sigma_{ZZ}^{1/2}$, $\mathbf{C}_{WZ} = \begin{pmatrix} 1 & \rho_{WZ} \\ \rho_{WZ} & 1 \end{pmatrix}$, $\rho_{WZ} = \sigma_{WZ} / \sqrt{\sigma_{WW}\sigma_{ZZ}}$, and the standard bivariate Student t CDF $T_\nu((w,z), \mathbf{C}_{WZ})$ (with two arguments) is equivalent to $T_\nu((w,z), (0,0), \mathbf{C}_{WZ})$ in terms of the generalized Student t CDF (with three arguments). Although Matlab function *mvtcdf*, also based on Genz (2004), will rescale a covariation matrix to a correlation matrix, it does not at the same time rescale the arguments or accommodate a non-zero location vector, so that the user must perform the location-scale transformation as in (28).

If the denominator Z is known to be nonnegative, we have

$$\begin{aligned} \mathrm{G}(x) &= \mathrm{P}(X < x \mid Z > 0) \\ &= \mathrm{T}_\nu\left((w, z), \mathbf{C}_{WZ}\right) / \mathrm{T}_\nu(z). \end{aligned} \tag{29}$$

in place of (27).

APPENDIX II

Ternary Integration

The integrals required to obtain the final smoother CDF $\mathrm{P}\left(x_t' \mid \mathbf{y}', \mathbf{x}, \hat{\alpha}, \hat{\beta}, \mathbf{S}, s^2, \hat{\tau}^2\right)$ from the conditionally normal smoother CDF $\mathrm{P}\left(x_t' \mid \mathbf{y}', \mathbf{x}, \hat{\alpha}, \beta, \hat{\beta}, \mathbf{S}, \sigma^2, \tau^2\right)$ implied by (25) have the form

$$V = \int_{-\infty}^{\infty} \mathrm{G}(x)\,\mathrm{p}(x)\,dx\,,$$

where p(x) is a probability density function with CDF P(x). Using the transformation s = P(x), so that ds = p(x)dx,

$$V = \int_0^1 \mathrm{G}(\mathrm{P}^{-1}(z))\,ds = \int_0^1 \mathrm{f}(z)\,dz \qquad (30)$$

for $\mathrm{f}(z) = \mathrm{G}(\mathrm{P}^{-1}(z))$.

Integration of (30) by trapezoids or by Simpson’s rule requires evaluating the integrand at the n+1 endpoints of n equally spaced intervals, and then giving positive weights, {.5, 1, ... 1, .5}/n or {1, 4, 2, ...4, 1}/(3n), respectively, to each of these values. However, if the integrand is singular at either limit of integration, neither of these methods will work at all. This occurs in the present context when $x_t' = \hat{x}_t'$, $\sigma^2 = 0$, and $\tau^2 = \infty$, since there the density integrand is infinite, and the CDF integrand is undefined. Even in the vicinity of $x' = \hat{x}'$, either of these methods will be ill-conditioned at $\sigma^2 = 0$ and $\tau^2 = \infty$ (z = 1 and z = 0, resp.).

Integration by rectangles instead evaluates the function at the n center points of these intervals, and then gives each value equal weight:

$$V(n) = \frac{1}{n}\sum_{h=1}^{n} \mathrm{f}(z_h),$$
$$z_h = (h - .5)/n.$$

This is more primitive than Simpson's Rule since it does not take the convexity or concavity of the integrand into account, but at least it is computable when the integrand is singular at one of the endpoints. $V(n)$ will converge to $V = V(\infty)$, but only very slowly, so that this method is not very satisfactory by itself. Nevertheless, it can give very good results if it is extrapolated appropriately to $n = \infty$.

If n is chosen to be divisible by 3, $V(n/3)$ will require only a subset of the points used by V(n), namely z_2 and every third point thereafter, so that there is essentially no extra cost to computing it. If n is sufficiently large, $V(n)$ will be closer to V than is V(n/3), so that for some a >0, V may be obtained by extrapolation from these two values:

$$\begin{aligned} V &= V(n) + a\big(V(n) - V(n/3)\big) \\ &= (1+a)V(n) - aV(n/3) \\ &= \sum_{h=1}^{n} w_h f(s_h) \end{aligned}$$

where

$$w_h = \begin{cases} (1+a)/n, & h = 1(\mathrm{mod}\,3) \text{ or } 3(\mathrm{mod}\,3) \\ (1-2a)/n, & h = 2(\mathrm{mod}\,3). \end{cases}$$

The exact value of a is of course unknown, and in the present application varies with x_t'. Nevertheless, $a = 1/2$ was found to work well for recovering the Student t density from a mixture of normal densities, and therefore presumably also for the present problem, which has Student t as a special case in the limit $s_{\beta\beta} \to 0$ when $\hat{\beta} \neq 0$ and

$\tau^2 = \infty$. This precise value has the considerable additional virtue that $w_h = 0$ for $h = 2 \pmod 3$, so that in fact the integrand only needs to be evaluated at the $m = (2/3)n$ points s_h for which the base 3 or *ternary* representation of h ends in either 1 or 0. In this *ternary integration*, each of the $m = (2/3)n$ included points receives equal weight $w_h = (3/2)/n = 1/m$.

In terms of the m included points, ternary integration may be written more succinctly as,

$$V \approx \frac{1}{m}\sum_{i=1}^{m} f(s_i),$$
$$s_i = \begin{cases} (i - 2/3)/m, & i \text{ odd}, \\ (i - 1/3)/m, & i \text{ even}, \end{cases}$$

for even values of m.

REFERENCES

Brohan, P., J.J. Kennedy, I. Harris, S.F.B. Tett, and P.D. Jones (2006). Uncertainty estimates in regional and global observed temperature changes: a new dataset from 1850. *J. Geophysical Research* **111** D12106, DOI: 10.1029/2005JD006548.

Brown, Philip J. (1982). Multivariate Calibration. *Journal of the Royal Statistical Society* **B 44** (3): 287-321.

__________ (1993). *Measurement, Regression, and Calibration*. Oxford: Clarendon Press.

Du Plessis, J.L., and A. J. van der Merwe (1996). Bayesian calibration in the estimation of the age of rhinoceros. *Annals of the Institute of Statistical Mathematics* **48**: 17-28.

Eisenhart, C. (1939). The interpretation of certain regression methods and their use in biological and industrial research. *Annals of Mathematical Statistics* **10**: 162-186.

Fieller, E.C. (1932). The distribution of the index in a normal bivariate population. *Biometrika* **24**: 428-40.

Fieller, E.C. (1954). Some problems in interval estimation. *Journal of the Royal Statistical Society*, Ser. B, **16**: 175-185.

Genz, Alan (2004). Numerical computation of rectangular bivariate and trivariate normal and t probabilities. *Statistics and Computing* **14**: 251-260.

Gore, Al (2006). *An Inconvenient Truth*. New York: Rodale.

Harvey, Andrew C. (1989). *Forecasting, Structural Time Series Models and the Kalman Filter*. Cambridge University Press.

Hill, Bruce M. (1981). Discussion [of Hunter and Lamboy (1981)]. *Technometrics* **23**: 335-338.

Hoadley, Bruce. (1970). A Bayesian look at inverse linear regression. *Journal of the American Statistical Association* **65**: 356-369.

Hinckley, D. V., (1969). On the ratio of two correlated normal random variables. *Biometrika,* **56**: 635–639.

Hunter, William G., and Warren F. Lamboy (1981). A Bayesian analysis of the linear calibration problem. *Technometrics* **23**: 323-350.

Kaufman, D.S., D.P. Schneider, N.P. McKay, C.M. Ammann, R.S. Bradley, K.R. Briffa, G.H. Miller, B. L. Otto-Bliesner, J. T. Overpeck, B.M. Vinther, and Arctic Lakes 2k Project Members (2009). Recent warming reverses long-term Arctic cooling. *Science* **325** (no. 5945, 4 Sept.): 1236-1239. DOI: 10.1126/science.1173893.

Krutchkoff, R.G. (1967). Classical and inverse regression methods of calibration. *Technometrics* **9**: 425-439.

Lawless, J.F. (1981). Discussion [of Hunter and Lamboy (1981)]. *Technometrics* **23**: 334-338.

Loehle, Craig, and J. Huston McCulloch (2008). Correction to: A 2000-year global temperature reconstruction based on non-tree ring proxies. *Energy & Environment* **19** (1): 93-100.

Mann, Michael E., Raymond S. Bradley, and Malcolm K. Hughes (1999). Northern hemisphere temperatures during the past millennium: Inferences, uncertainties, and limitations, *Geophysical Research Letters*, **26**(6), 759–762.

Marsaglia, George (1965). Ratios of normal variables and sums of uniform variables. *Journal of the American Statistical Association* **60**: 193-204.

McCulloch, J. Huston (2005). The Kalman foundations of Adaptive Least Squares, with application to U.S. Inflation. Ohio State University working paper. <www.econ.ohio-state.edu/jhm/papers/KalmanAL.pdf>.

__________ (2009a). Irreproducible Results in Thompson *et al.* (*PNAS* 2006), "Abrupt Tropical climatic Change: Past and Present." *Energy and Environment* **20** (3): 369-375.

__________ (2009b). Further Results, posted May 11, 2009, on <www.econ.ohio-state.edu/jhm/AGW/Thompson/>.

McIntyre, Stephen (2006). Juckes, Yang, Thompson and PNAS: Guliya. *Climate Audit*, 12/3/2006. http://climateaudit.org/2006/12/03j/juckes-yang-thompson-and-pnas-guliya.

McIntyre, Stephen (2007). Dunde: Will the real Slim Shady please stand up? *Climate Audit*, 4/12/2007. http://climateaudit.org/2007/04/12/the-dunde-fiasco.

McIntyre, Stephen and Ross McKitrick (2003). Corrections to the Mann *et al.* (1998). Proxy Data Base and Northern Hemisphere Average Temperature Series. *Energy and Environment* **14** (6): 751-71.

__________ (2005). Hockey sticks, principal components, and spurious significance. *Geophysical Research Letters* **32** (3). DOI: 10.1029/2004GL021750.

Met Office [of UK] (2009). Release of global-average temperature data. Press release dated 5 Dec. 2009, online at http://www.metoffice.gov.uk/corporate/pressoffice/2009/pr20091205.html.

Orban, John E. (1981) Discussion [of Hunter and Lamboy (1981)]. *Technometrics* **23**: 342-343.

Osborne, Christine (1991). Statistical Calibration: A Review. *International Statistical Review* **59**: 309-336.

Preisendorfer, Rudolph W. (1988). *Principal Component Analysis in Meteorology and Oceanography*. Amsterdam: Elsevier.

Romer, Christina (1989). The prewar business cycle reconsidered: new estimates of Gross National Product, 1869-1908. *Journal of Political Economy* **97**: 1-37.

Steckel, Richard H. (1995). Stature and the standard of living. *Journal of Economic Literature* **33**: 1903-1940.

Thompson, Lonnie G., Ellen Mosley-Thompson, M.E. Davis, P.-N. Lin, K. Henderson, and T.A. Mashiotta (2003). Tropical Glacier and Ice Core Evidence of Climate Change on Annual to Millennial Time Scales. *Climatic Change* **59:** 137-155. Partial data online at <http://bprc.osu.edu/Icecore/Climatic-change-2003-Fig5-table.XLS>.

Thompson, L.G., *et al.* (2005). Quelccaya Summit Core Particle Sizes, Conductivity, Delta 18-O and Accumulation. Data file archived 2/23/2005 at <ftp://ftp.ncdc.noaa.gov/pub/data/paleo/icecore/trop/quelccaya/q83summ.txt>.

Thompson, Lonnie G., Ellen Mosley-Thompson, Henry Brecher, Mary Davis, Blanca León, Don Les, Ping-Nan Lin, Tracy Mashiotta, and Keith Mountain (2006). Abrupt tropical climate change: Past and present. *Proceedings of the National Academy of Sciences, USA* (*PNAS*) 103 (July 11): 10536-10543. Data Sets online via <http://www.pnas.org/cgi/content/full/0603900103/DC1> (accessed Nov. 25, 2007)

Tingley, Martin P., and Peter Huybers (2009). A Bayesian algorithm for reconstructing climate anomalies in space and time. Part 1: Development and applications to paleoclimate reconstruction problems. Forthcoming, *Journal of Climate*. Preprint created 10/30/09 online at <www.people.fas.harvard.edu/~tingley/BARCAST_Part1.pdf>.

Williams, E.J. (1969). A note on regression methods in calibration. *Technometrics* **11**: 189-192.

Zellner, Arnold (1971). *An Introduction to Bayesian Inference in Econometrics.* New York: Wiley.