Reply to Comments on "Reconstruction of the extra-tropical NH mean temperature over the last millennium with a method that preserves low-frequency variability" by A. Moberg.

BO CHRISTIANSEN, *

Danish Meteorological Institute, Copenhagen, Denmark

FREDRIK CHARPENTIER LJUNGQVIST

Department of History, Stockholm University, Stockholm, Sweden

* *Corresponding author address:* Bo Christiansen, Danish Meteorological Institute, Danish Climate Centre, Lyngbyvej 100, DK-2100 Copenhagen Ø, Denmark.

E-mail: boc@dmi.dk

1. Introduction

Christiansen (2011a) introduced a multi-proxy temperature reconstruction method, LOC, that is designed to avoid the serious underestimation of low-frequency variability that has been demonstrated for previously applied methods, see, e.g., von Storch et al. (2004); Zorita et al. (2007); Christiansen et al. (2009). The method is based on first reconstructing local temperatures at the geographical positions of the proxies and then averaging the local temperatures to obtain a large-scale (e.g., the Northern Hemispheric) mean. It is important that the linear regression applied to obtain the local reconstructions takes temperature as the independent variable, a choice that is known as indirect regression or classical calibration. Furthermore, LOC combines the local reconstructions into a large-scale construction by a simple mean avoiding the complications of more complex spatial covariance models. Many details, e.g., about the screening of the proxies and the relation to calibration, can be found in Christiansen (2011a,b) and Christiansen and Ljungqvist (2011b).

The skills of the LOC method depend on averaging as pointed out already in Christiansen (2011a) where it is stated that temporal smoothing will reduce the influence of the noise if the decorrelation time of the noise is shorter than the decorrelation time of the temperature and that spatial averaging furthermore will dampen the influence of the noise. It was also repeatedly pointed out that the good skill of the reconstructed low-frequency variability comes with the disadvantage of an exaggerated high-frequency variability. Furthermore Christiansen (2011a) found only small differences in the skills regarding low-frequency diagnostics when comparing LOC reconstructions with 57 and 23 pseudo-proxies.

In his comment Moberg (2011) correctly notice that the indirect regression on which the LOC method is based will overestimate the variance on all frequencies and he discusses how the temporal structure of the noise is important for the strength of this overestimation. These are valid considerations that we will like to discuss in more details in this Reply. As mentioned above an important factor is the decorrelation time of the noise, or more generally, its spectral structure. Although, this effect was already to a large extent included in the reconstruction method comparison (Christiansen et al. 2009), in the estimations of the skills of LOC (Christiansen 2011a,b), and in the calculations of the confidence intervals on the real-world LOC reconstructions (Christiansen and Ljungqvist 2011b,a) there are still unanswered questions about how the LOC method behaves under different types of noise.

We will focus on the confidence intervals on the smoothed reconstructions. Such confidence intervals were considered in previous work (Christiansen and Ljungqvist 2011b; Christiansen 2011b; Christiansen and Ljungqvist 2011a) with LOC. We also believe that the confidence intervals are more easily interpreted than the variance spectra used by Moberg (2011) as they are directly related to the significance of the temperature anomalies under, e.g., the Little Ice Age or the Medieval Warm Period.

We begin in section 2 with some theoretical considerations about the bias/variance trade-off leading to a highly idealized expression of the confidence intervals as function of the strength of the temporal and spatial smoothing, the amount of noise in the proxies, and the decorrelation time of the noise. In section 3 we present ensemble pseudo-proxy experiments giving more reliable confidence intervals for different values of these important variables. In section 4 we review previous real-world LOC reconstructions with different number of proxies. We close in section 5 with the conclusions and some additional considerations.

2. Some theoretical considerations

The relevant expression for our purpose is Eq. A6 in Christiansen (2011a). This equation relates the local LOC reconstruction, T^I, to the target local temperature, T,

$$T^I = T + \xi_{eq}/\lambda, \qquad (1)$$

where ξ_{eq} is the noise which is assumed independent of T. This result holds in the asymptotic limit of a large calibration interval and is derived under the assumption that the proxies are noisy versions of the local temperature $P = \lambda T + \xi_{eq}$. Here λ is related to the correlation coefficient c between P and T by $\sigma^2_{\xi_{eq}}/\lambda^2 = \sigma^2_T(1-c^2)/c^2$ (σ^2_x denotes the variance of x).

The equation above is the key to understand the behavior of LOC and is the mathematical background for the discussion in Moberg (2011).

As discussed by Moberg (2011) the variance of T^I is larger than the variance of T, as T and ξ_{eq} are independent. However, it is important to note that the noise term in Eq. 1 is additive so the noise has the form of modulations on top of the correct signal T. This means that the expectation of T^I is T even in the absence of any smoothing. Thus there is no bias in Eq. 1 or, in other words, T^I is a consistent estimator. If direct regression had been used instead of indirect regression the expression for the local reconstruction is given by Eq. A5 in Christiansen (2011a). Here, there is in addition to additive noise also an multiplicative term which is between 0 and 1 and which dampens both the signal T and the noise. The expectation of T^D is lower than T, so T^D is a biased estimator. The difference becomes totally transparent for small λ where the noise variance diverges and dominates T^I while the bias dominates T^D which approaches zero. The essential point for LOC is that the variance can be dampened by averaging while the bias cannot. This point was elaborated on in details in Christiansen and Ljungqvist (2011b) and Christiansen (2011b) and the averaging can include both spatial and temporal components.

Let us assume that all correlations between proxies and local temperatures are c and that the decorrelation time of the noise is τ everywhere. Averaging over N independent proxies and a time-interval of length Ω we get from Eq. 1 an estimate for the width of the 95 % confidence interval

$$\Delta_{95} = 2\sigma_T \sqrt{1/c^2 - 1}/\sqrt{\Omega/\tau}/\sqrt{N} \tag{2}$$

where σ_T^2 is the typical variance of the local temperatures. As expected there are three effects here. If the proxies are not well related to local temperature, c will be low and Δ_{95} correspondingly large. Both an increased length of the time interval Ω and an increased number of proxies will reduce Δ_{95}. However, the decorrelation length reduces the effect of the temporal smoothing. For AR1 noise we have the estimate $\tau = (1+\rho)/(1-\rho)$, where ρ is the lag-1 autocorrelation coefficient (see, e.g., von Storch and Zwiers 1999).

Typical correlations between local temperatures and proxies are centered around 0.4 (Mann et al. (2007); Christiansen et al. (2009). See also Figs. 6 and 7 of Christiansen and Ljungqvist (2011b)). Of the 65 annually resolved proxies in Christiansen and Ljungqvist (2011a) the lag-1 autocorrelation coefficients of the noise have a mean of 0.30 and the mean of their absolute values is 0.34 (lag-1 autocorrelations of magnitude larger than 0.22 are significantly different from zero to the 5% level). Here we have estimated the noise as the residual between proxy and local temperature in the calibration period 1880-1960. The same numbers for the 21 annually resolved proxies in Christiansen and Ljungqvist (2011b) are 0.35 and 0.37. These numbers agree well with the estimates of Mann et al. (2007) and McShane and Wyner (2011). See also the discussion in Smerdon (2011).

With $\rho = 0.32$, $c = 0.4$, $N = 23$, $\Omega = 50$ year, and $\sigma_T = 1°$C we find from the above equation that $\Delta_{95} = 0.19°$C. This number is obviously optimistic as it does not consider, e.g., the effects of the finite calibration interval and the spatial sampling which will be included in the pseudo-proxy experiments in the next section.

3. Ensemble pseudo-proxy experiments

Ensemble pseudo-proxy experiments were used in Christiansen (2011a,b) to test the skills of the LOC method and in Christiansen and Ljungqvist (2011b,a) to provide confidence intervals on the 50-year smoothed real-world reconstructions. These pseudo-proxy studies included the effect of both temporal and spatial averaging and showed that low-frequency variability and trends were well reconstructed without bias and with relatively small variance. They also demonstrated that the confidence intervals shrink with increasing number of proxies. In particular, the 95% confidence intervals on the 50-year smoothed reconstructions had widths of 0.57°C (23 proxies, Christiansen and Ljungqvist 2011b), 0.6 and 0.4°C (17 and 47 proxies, Christiansen and Ljungqvist 2011a) making features like the 17th century anomaly significantly colder than the calibration period.

In the ensemble pseudo-proxy experiments of Christiansen et al. (2009); Christiansen (2011a); Christiansen and Ljungqvist (2011b); Christiansen (2011b); Christiansen and Ljungqvist (2011a) the "realistic" noise were calculated individually for each pseudo-proxy from the residual between a corresponding real-world proxy and the observed local temperature. This residual was phase-scrambled to give noise with the same auto-correlation spectra as the residual. The temporal spectrum of the noise therefore vary from proxy to proxy and preserves the local spectral structure without the potential limitation of prescribing a AR1 structure. The caveat here is that the residual has to be calculated from the observed period and therefore cannot confidently include timescales longer than 80-100 years.

Using the proxy compilations of Christiansen and Ljungqvist (2011b,a) we have performed ensemble pseudo-proxy experiments to investigate more closely the effect of temporal and spatial smoothing as well as the influence of the type of the noise. We base the ensemble pseudo-proxy experiments on a forced climate model (ECHAM4-OPYC3) experiment covering the last 500 years (Stendel et al. 2006); for more details see Christiansen et al. (2009); Christiansen (2011a). We focus on the 95 % confidence intervals on the extra-tropical Northern Hemisphere (NH) mean as these can be easily interpreted and compared to the estimates of Christiansen and Ljungqvist (2011b,a). The 95 % confidence intervals are plotted in Fig. 1 as function of the smoothing interval for "realistic" noise as described above and AR1 noise with $\rho = 0.32$ and $\rho = 0.64$. This is done for the two different proxy compilations of sizes 23 and 47 (before screening 40 and 91). As expected the confidence intervals shrink both when the number of proxies grow and when the smoothing interval, Ω, increase. Also as expected, we find that the temporal smoothing is most efficient for white noise and less efficient for $\rho = 0.64$ compared to $\rho = 0.32$. It is in particular interesting that the results for "realistic" noise and $\rho = 0.32$ are very similar. This suggests that the noise is adequately modelled by an AR1 process with $\rho = 0.32$ and that the low-frequency noise induced by this process is of minor importance for the 95 % confidence intervals.

Note, that while the widths of these confidence intervals are larger than the theoretical estimates in the previous section they do follow the functional form very closely. The square-root dependence on Ω is clearly visible in Fig. 1 and the square-root dependence on N is demonstrated by the fact that confidence intervals for 47 proxies (full curves) are approximately a factor $\sqrt{47/23}$ wider than the confidence intervals for 23 proxies (dashed curves). From Eq. 2 is also follows that the confidence intervals for $\rho = 0.32$ and $N = 23$ are identical to those for $\rho = 0.64$ and $N = 54$. This is in agreement with the closeness of the red dashed and the green solid curves in Fig. 1.

4. Experience from real-world reconstructions

While the pseudo-proxy experiments give much more realistic estimates of the skills and confidence intervals than the theoretical arguments in section 2 they do include important assumptions. Among those assumptions are the spectral structure of the noise (as described above) and the linear dependence between proxy and local temperature. Another assumption is that the surface temperature field from the climate model experiment is a good representation of the real temperature field.

In this section we will therefore try to estimate the effect of the potential low-frequency noise on LOC by studying how stable the LOC based real-world reconstructions are to the degree of spatial averaging, i.e., to the number of included proxies.

Christiansen and Ljungqvist (2011a) (Figs. 6 and 8) show LOC based reconstructions reaching back to AD 1500 based on different number of proxies (from 17 to 55 after screening). We note that they all agree on a cold minimum around AD 1600 close to 1.1 °C lower than the temperature in the reference period, AD 1880–1960 (50-y smoothed). Christiansen and Ljungqvist (2011b) presented a LOC reconstruction of the extra-tropical NH mean temperature in the last millennium based on 23 proxies (selected by screening 40 proxies). Also in this study a cold anomaly approximately 1.1 °C below the AD 1880–1960 level was found around AD 1600. The different reconstructions disagree more about the temperature minimum in the 19th century. However, there does not seem to be a systematic difference with the number of proxies; in fact the coldest reconstruction is the one based on the most proxies (red curve in Fig. 8 of Christiansen and Ljungqvist (2011a)). We also note that the LOC reconstructions based on different numbers of proxies agree within the pseudo-proxy based confidence levels (see last section). These results suggest that the effect of low-frequency noise in the real-world reconstructions is small and that LOC reconstructs 50-year smoothed NH mean values well when based of 23 proxies or more.

5. Conclusions

As Moberg (2011) correctly notice the indirect regression on which LOC is based will overestimate the variance on all timescales. LOC dampens this noise by both spatial and temporal averaging. While spatial averaging works on all timescales, temporal averaging is effective only on timescales larger than the decorrelation time of the noise. Previous work on skills and confidence levels for LOC reconstructions has shown that the low-frequency variability is confidently reconstructed.

In this Reply we have focused on questions about how the LOC method behaves under different types of noise and on how the temporal and spatial averaging suppress the influence of the noise. We have

- From theoretical arguments derived an expression that shows the competing effects of the correlations between proxies and local temperatures, the number of proxies, the temporal smoothing and the decorrelation timescale of the noise on the width of the confidence intervals. While this expression underestimates the width, as it is based on strong assumptions, it catches the functional form well.

- Performed ensemble pseudo-proxy experiments with different kinds of noise. These pseudo-proxy experiments mimic the real-world reconstructions in most details and do not depend on the assumptions behind the theoretical expression. These pseudo-proxy experiments confirmed that our earlier results about the confidence intervals, based on "realistic" noise estimated from the residuals between proxies and local temperatures, also hold when the noise is AR1 with observed values of the lag-1 coefficient. As shown in Christiansen and Ljungqvist (2011b,a) real-world reconstructions based on LOC is generally more variable than other reconstructions even when the widths of the confidence intervals are taken into account.

- Compared previous (Christiansen and Ljungqvist 2011b,a) real-world reconstructions based on LOC including different numbers of proxies. We do not find any systematic reduction of the 50-year smoothed variability with increasing number of proxies. Furthermore, the reconstructions based on different numbers of proxies fall within the confidence intervals estimated by the ensemble pseudo-proxy approach. This indicates that around 25 independent proxies with correlations to the local temperatures centered around 0.4 is enough to produce a reliable 50-year smoothed NH mean LOC reconstruction.

For our proxy compilations we observe values of the AR1 coefficient close to what is found earlier (Mann et al. 2007; McShane and Wyner 2011). Pseudo-proxy experiments based on AR1 noise with $\rho = 0.32$ gives the same results as our previous experiments with realistic noise confirming that low-frequency noise is probably not a large problem. Estimates of the noise in the proxies are limited by the brevity of the period where we have observed local temperatures and estimates of the low-frequency parts of the noise is necessarily based on assumptions. However, there is little evidence that typical AR1 coefficients can reach the size of 0.71 as suggested by Moberg (2011). It should also be noted that if proxies are polluted with considerable amounts of low-frequency noise all reconstruction efforts will be jeopardized as the calibration process will be subject to large uncertainties due to the limited number of degrees of freedom.

Moberg (2011) briefly mentions the discussion of error-in-variable models in Tingley and Li (2011); Christiansen (2011b). The reasons why LOC ignores the noise on the temperature is discussed in Appendix B of Christiansen (2011a) and again in Christiansen (2011b) We repeat the main arguments here (but see, e.g., Cheng and Ness (1999) for a thorough review of error-in-variable models). Presumably the dominating noise originates from the equation error and not the measurement errors. In the indirect regression used in LOC the equation error and the measurement error on the proxies enter as a sum. We therefore expect this sum to dominate over the measurement error on the temperatures. It is also important that the temperature needs to be reconstructed from the noisy proxies and not from the noise-free proxies and that the reconstructed temperature will be compared to the noisy temperature in the calibration period.

Finally, Moberg (2011) mentions that Tingley and Li (2011) shows that LOC can be embedded in a Bayesian framework. This is not surprising as, in the language of Tingley et al. (2011), LOC is a model for which we have used ordinary least-square regression for inference. The Bayesian framework is another inference method which has the advantage that it allows a consistent estimation also of the confidence intervals. However, in contrast to the ensemble pseudo-proxy approach the Bayesian framework will not assess the bias related to the model choice (Christiansen 2011b).

Acknowledgments.

This work was supported by the Danish Climate Centre at the Danish Meteorological Institute.

REFERENCES

Cheng, C.-L. and J. W. V. Ness, 1999: *Statistical regression with measurement error*, Kendall's Library of Statistics, Vol. 6. Arnold, 262 pp.

Christiansen, B., 2011a: Reconstructing the NH mean temperature: Can underestimation of trends and variability be avoided? *J. Climate*, **24**, 674–692.

Christiansen, B., 2011b: Reply to "Comments on ..." by Tingley and Li. *J. Climate*, **submitted**, available as preprint at http://web.dmi.dk/fsweb/solar-terrestrial/staff/boc/loc_reply.pdf.

Christiansen, B. and F. C. Ljungqvist, 2011a: The extra-tropical NH temperature in the last two millennia: Reconstructions of low-frequency variability. *Clim. Past*, **submitted**, available as preprint at http://web.dmi.dk/fsweb/solar-terrestrial/staff/boc/loc_2000.pdf.

Christiansen, B. and F. C. Ljungqvist, 2011b: Reconstruction of the extra-tropical NH mean temperature over the last millennium with a method that preserves low-frequency variability. *J. Climate*, **accepted**, available as preprint at http://web.dmi.dk/fsweb/solar-terrestrial/staff/boc/millennium_reconstr_rev2.pdf.

Christiansen, B., T. Schmith, and P. Thejll, 2009: A surrogate ensemble study of climate reconstruction methods: Stochasticity and robustness. *J. Climate*, **22**, 951–976.

Mann, M. E., S. Rutherford, E. Wahl, and C. Ammann, 2007: Robustness of proxy-based climate field reconstruction methods. *J. Geophys. Res.*, **112**, D12 109, doi:10.1029/2006JD008272.

McShane, B. B. and A. J. Wyner, 2011: A statistical analysis of multiple temperature proxies: Are reconstructions of surface temperatures over the last 1000 years reliable? *Ann. Appl. Stat.*, **5**, 5–44.

Moberg, A., 2011: Comments on ”reconstruction of the extra-tropical NH mean temperature over the last millennium with a method that preserves low-frequency variability”. *J. Climate*, **submitted**.

Smerdon, J. E., 2011: Climate models as a test bed for climate reconstruction methods: pseudoproxies. *Interdisciplinary Reviews: Climate Change*, **submitted**, available as preprint at `http://www.ldeo.columbia.edu/~jsmerdon/papers/2011_wires_smerdon.pdf`.

Stendel, M., I. A. Mogensen, and J. H. Christensen, 2006: Influence of various forcings on global climate in historical times using a coupled atmosphere-ocean general circulation model. *Clim. Dyn.*, **26**, 1–15.

Tingley, M. P., P. F. Craigmile, M. Haran, B. Li, E. Mannshardt-Shamseldin, and B. Rajaratnam, 2011: Piecing together the past: Statistical insights into paleoclimatic reconstructions. *Quat. Sci. Rev.*, **submitted**, available as preprint at `http://www.martintingley.com/wp-content/uploads/2011/08/PiecingTogether_TechReport-1.pdf`.

Tingley, M. P. and B. Li, 2011: Comments on ”Reconstructing the NH mean temperature: Can underestimation of trends and variability be avoided?”. *J. Climate*, **submitted**, available as preprint at `http://www.martintingley.com/wp-content/uploads/2011/08/Comment_on_Christiansen.pdf`.

von Storch, H., E. Zorita, J. M. Jones, Y. Dimitriev, F. González-Rouco, and S. F. B. Tett, 2004: Reconstructing past climate from noisy data. *Science*, **306**, 679–882.

von Storch, H. and F. W. Zwiers, 1999: *Statistical analysis in climate research*. Cambridge University Press, 484 pp.

Zorita, E., F. González-Rouco, and H. von Storch, 2007: Comments on "Testing the fidelity of methods used in proxy-base reconstructions of past climate". *J. Climate*, **20**, 3693–3698.

List of Figures

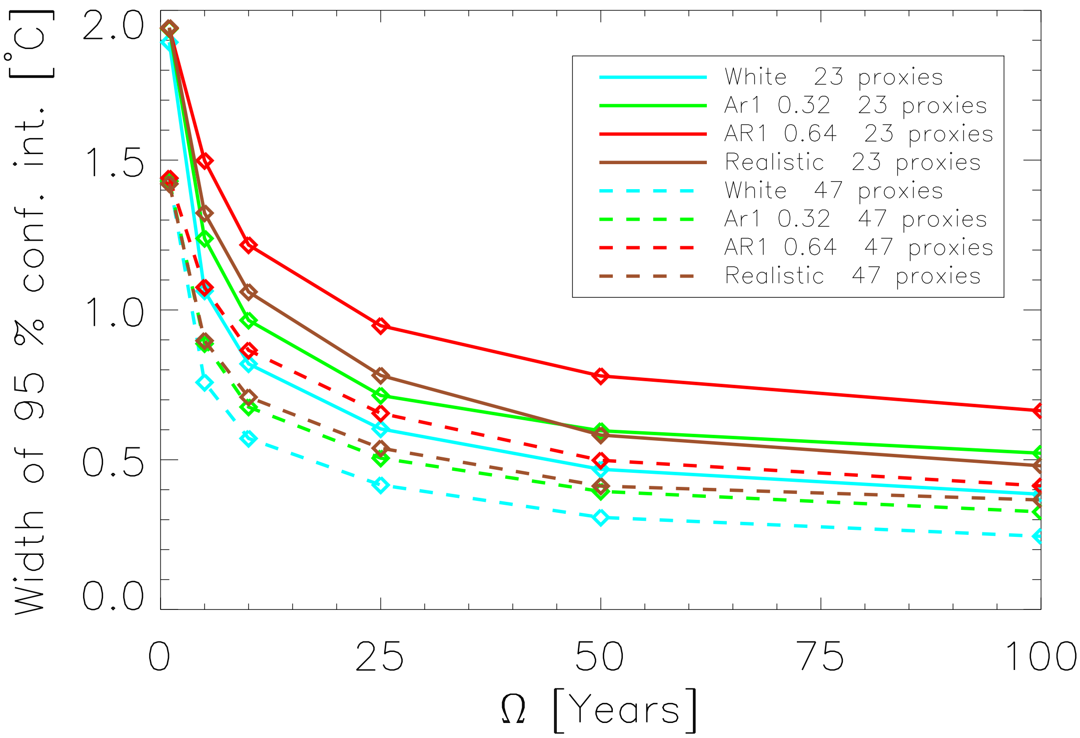

FIG. 1. The width of the 95 % confidence intervals [° C] as function of temporal smoothing, Ω [Years]. Full curves are based on the compilation of 40 proxies from Christiansen and Ljungqvist (2011b) (of which 23 passed the t-test) and dashed curves are based on a compilation of 91 proxies from Christiansen and Ljungqvist (2011a) (of which 47 passed the t-test). The confidence intervals are calculated with the ensemble pseudo-proxy method described in Christiansen et al. (2009); Christiansen (2011a); Christiansen and Ljungqvist (2011b).