Clim. Past Discuss., 10, C1031–C1034, 2014
www.clim-past-discuss.net/10/C1031/2014/

Climate of the Past Discussions
Open Access

Interactive comment on "Implication of methodological uncertainties for Mid-Holocene sea surface temperature reconstructions" *by* I. Hessler et al.

R.J. Telford (Referee)

richard.telford@bio.uib.no

Received and published: 11 July 2014

The mid-Holocene (MH) is a key target for climate model simulations. If the models are able to simulate the MH climate, it gives some encouragement that they are able to make good projections of future climate. Palaeoclimate data are required to validate the simulations. Because discrepancies between the climate model simulations and the palaeoclimate data could be due to problems in either the models or the data, it is important to understand the uncertainties in the palaeoclimate data. Hessler et al. attempt to do this with a new compilation of MH sea-surface temperature (SST) proxies, exploring the impact of reference state and the time span covered by the MH window,

and find that the uncertainties are larger than the signal and that the different proxies are inconsistent. The paper will be a valuable contribution to the literature, but there are other sources of uncertainty that ought to be addressed.

Hessler et al. compile a variety of SST proxies, which for unknown reasons they decide to call sensors, a word I would associate with instrumental data, but exclude diatoms on the grounds that there is not a harmonised data set. This is unfortunate as diatoms can reasonably be expected to be sensitive to SST rather than temperature at a greater depth, unlike several of the other proxies included.

Foraminifera and dinocyst assemblages are used to reconstruct summer, winter and mean annual SST. This may be possible in parts of the ocean, but it is doubtful at high-latitudes where the vast majority of biological production occurs during the warm season. This problem is acknowledged deep in the discussion "The derived seasonal SST reconstructions are not independent but necessarily reïňĆect the covariance among seasonal SSTs in the modern ocean (Kucera et al., 2005a). This is patently unlikely in the case of the MH"

Such concerns are likely to be overlooked by users of the compilation. It would be better to evaluate whether it is possible to make meaningful reconstructions of seasonal SST, and if not, omit them from the analysis.

SST reconstructions from planktonic foraminifera are calculated using both the modern analogue technique (MAT) and artificial neural networks (ANN). ANN is described as permitting extrapolation beyond the range of parameters in the calibration set. I have not seen this issue explored.

A few cores in the compilation have results from multiple proxies, permitting a direct comparison between the proxies. Unfortunately the results are not consistent, even allowing for the uncertainty in the reconstructions. There are several possible reasons for this that are not fully explored in the paper.

CPD

10, C1031–C1034, 2014

1) Chronological control. The minimum number of radiocarbon dates or other stratigraphic markers is two. Presumably stricter criteria resulted in the exclusion of too many records. With such weak chronological constraints, errors of 1000 years or more can be expected in some cores. This would be a particular problem for the shorter MH window used. I would like the authors to give some consideration to this problem, but suspect the impact is relatively minor.

2) Palaeoclimatologists have adopted methods that allow them to report the lowest possible uncertainty on their reconstructions. These low uncertainties make the reconstruction appear good until different proxies are compared, as in Hessler et al., and the reconstructions are not consistent with each other. It is trivial to show that the root mean squared error of prediction of the transfer functions for the dinocysts and planktonic foraminifera is biased low, perhaps by a factor of two, because of spatial autocorrelation violating the assumption of independent between the test and calibration set during cross-validation (Telford and Birks 2005, QSR) and other problems. If the uncertainties were correctly shown, the results might appear less inconsistent.

3) Hessler et al. treat all the proxies as being sensitive to (seasonal) SST. However, some of the proxies are probably more sensitive to sub-surface temperatures. There is, for example, good evidence that the dominant foraminifera in the Nordic Seas live sub-surface and that the Holocene temperature trends at the surface and in the sub-surface are different (Andersson et al. 2010 COP). Even if the proxies were perfect, they would appear inconsistent if they are sensitive to different aspects of the water column. Note, if the biotic assemblages have been calibrated against a suboptimal depth/season, the uncertainty will have been inflated.

4) The great mismatch between alkenones and other proxies suggests there may be undiagnosed biases in one of more of the proxies.

It may be difficult to account for these issues, but if it could be done, I think the results of this paper would be less pessimistic.

CPD
10, C1031–C1034, 2014

Hessler et al. use a significance test to determine how many records have SST anomalies significantly different from zero. My understanding is that records where the absolute mean anomaly (from cores where there are three or more analyses in the MH window) is greater than twice the standard error of the analyses plus the uncertainty on the reconstruction are deemed significant. This is a non-standard test and its behaviour is not explored. It is equivalent to a t-test for large samples (if the proxy uncertainty is ignored) as the t-distribution resembles a Gaussian distribution in this case, but not when the sample is small as the t-distribution has heavy tails when the sample is small. One critical aspect is whether the reconstruction errors are independent. If the errors are not independent, such that if one analysis is too warm all will be too warm, this test will be too liberal. Conversely, if the errors are independent the test will be too strict. My feeling is that the degree to which the error is independent will be method specific. Methods where analytical error or aliasing of the annual signal are high will have more independent errors. If the errors are independent, a one-sample t-test can be used – the reconstruction uncertainty is already accounted for in the estimate of the variance. If the errors are not independent, I think a modified t-test can be used, adding the reconstruction uncertainty in quadrature to the standard error. The reconstruction uncertainty could be scaled according to the degree it is thought to be independent. I do not know what effect this would have on the results.

Figures S1-S5 are missing from the supplementary material.

If the proxies cannot be reconciled at the MH, questions need to be asked about the utility of palaeoceanographic proxies for Holocene reconstructions. I am hoping that either the issues I raise above help reconcile the proxies, or that one or more of the proxies can be discounted and the remainder are coherent.

Interactive comment on Clim. Past Discuss., 10, 1747, 2014.