Clim. Past Discuss., 9, C1837–C1841, 2013
www.clim-past-discuss.net/9/C1837/2013/

Climate of the Past Discussions

Interactive comment on "Qualitative and quantitative reconstruction of surface water characteristics and recent hydrographic changes in the Trondheimsfjord, central Norway" *by* G. Milzer et al.

R.J. Telford

richard.telford@bio.uib.no

Received and published: 26 August 2013

Milzer et al present dinocyst assemblages from three short cores from Trondheimsfjord and use these records to reconstruct summer and winter sea surface temperature (SST), sea surface salinity (SSS), and annual productivity with the modern analogue technique (MAT). Unfortunately, as is usual in the dinocyst literature, this paper does not consider the problems that MAT has and which makes the reported performance over-optimistic, and thus the reconstructions are even more uncertain that they appear to be. There are also other issues with the numerical methods.

The first problem with MAT is that it gives biased estimates of the root mean squared error of prediction (RMSEP) if the observations are not evenly distributed along the environmental gradient (Telford and Birks, 2011a). In under-sampled portions of the gradient, the uncertainty can be much larger than the RMSEP as there are few available analogues. Conversely, in over-sampled portions of the gradient, the uncertainty can be smaller than the RMSEP. How much of a problem this is will depend on which part of the gradient is being reconstructed. Telford and Birks (2011a) develop some tools that help explore this problem.

The second problem is that estimates of transfer-function performance statistics assume that the calibration-set observations are independent, whereas the dinocyst calibration-set exhibits strong spatial autocorrelation (Telford, 2006). This autocorrelation will make the transfer function appear to perform better than is justified by the data (Telford and Birks, 2005; 2009), especially if MAT is used (Telford and Birks, 2005). In extremis, environmental variables with no ecological relevance will appear to be possible to reconstruct. Telford and Birks (2009) introduce tools, available in R, for estimating how severely autocorrelation is affecting the performance statistics.

Guiot and de Vernal (2011) tried to show that MAT is unique in not being affected by spatial autocorrelation. However, they tested which transfer function performs best once spatial autocorrelation has been removed by excluding spatially close analogues. Their analysis has no relevance to the question of how severe a problem spatial autocorrelation is for Milzer et al and other studies that have not controlled for the effects of spatial autocorrelation (Telford and Birks, 2011b).

A third problem is that variability in environmental variables other than the one of interest can contaminate the reconstructions (Juggins 2013). As Milzer et al are reconstructing five variables, some of which are unlikely to vary much relative to the ecological tolerance of the taxa, they need to consider this problem. It is doubtful that there is sufficient information in the downcore data to make five independent reconstructions.

The uncertainty on the SST and SSS reconstructions is very wide. This is perhaps not surprising given that, as Milzer et al write, "Dinoflagellates in coastal environments are generally adapted to large sea-surface temperature (SST) and salinity (SSS) ranges due to the seasonally varying relative inïňĆuence of marine/coastal and continental waters." For most of the cores, the uncertainty on the reconstructions (e.g. 20-33 psu, 8-18°C) is much wider than could plausibly be predicted for variability in Trondheimsfjord during the twentieth century: the reconstructions have limited or no utility. Indeed such broad uncertainties are unlikely to be a useful constraint for sea surface conditions in at least the late Holocene.

Milzer et al use a Wilcoxon test to compare the reconstructed SST and SSS with instrumental data. This is essentially testing if the reconstructions and measured variables have different medians. The more interesting question of whether the variability in the reconstructions correlates with the instrumental data is ignored. The median could be approximately correct if one of the nearby analogues in the calibration-set, for example the observation from Hemnefjord, is heavily weighted.

Including the dinocyst assemblages from Milzer et al (2013) in the calibration-set, as suggested on page 4575, would almost certainly improve the fit between the reconstructions and the instrumental data, but most of this improvement would be because of the lack of independence between the calibration-set and the fossil data. Any improvement for the twentieth century could not be expected to hold back through the Holocene.

It is not clear why or how Milzer et al use NMDS to find the clusters of downcore observations highlighted on the CA. A cluster algorithm would be a more obvious and (somewhat) objective choice, or simply showing the NMDS. If the CA is to be used, the scale should be the same on both axes, and the 16 "factors" found by CA simply reflects that there are 17 taxa – not a result worth reporting.

The paper reports a relationship between NAO index and dinocyst assemblages, but

no test is given of this relationship. A simple test would be to correlate the first axis of the ordination with the NAO index.

This paper shares many of its weaknesses with similar recent dinocyst literature. The simplest strategy for improving the paper is to drop the environmental reconstructions as these still need considerable work. As it stands, the paper does not persuade me that it has established "a solid basis for the future investigation of Holocene paleoclimate and paleoceanographic variability."

Minor points

I cannot find any dinocyst data at the url given.

Several of the statistics given in tables two and three are of little interest and not discussed, others can be read from figure 9. These tables can either be deleted or condensed.

The data handling is probably mis-described in figure 2. The red lines cannot be "the 15 pt. running average of annual SSTs" as they show intra-annual variability. Are they the running average of weekly data? The black polynomials on the SST and SSS plots look to be overfitted – a LOESS would probably be preferable. The version of R used should be cited.

References

Guiot, J, de Vernal, A. 2011. Is spatial autocorrelation introducing biases in the apparent accuracy of paleoclimatic reconstructions? Quaternary Science Reviews 30, 1965–1972.

Juggins, S. 2013 Quantitative reconstructions in palaeolimnology: new paradigm or sick science? Quaternary Science Reviews 64: 20–32.

Telford RJ. 2006. Limitations of dinoflagellate cyst transfer functions. Quaternary Science Reviews 25: 1375-1382.

Telford RJ, Birks HJB. 2005. The secret assumption of transfer functions: problems of spatial autocorrelation in evaluating model performance. Quaternary Science Reviews 24: 2173-2179.

Telford RJ, Birks HJB. 2009. Evaluation of transfer functions in spatially structured environments. Quaternary Science Reviews 28: 1309-1316.

Telford RJ, Birks HJB. 2011a. Effect of uneven sampling along an environmental gradient on transfer-function performance. Journal of Paleolimnology 46, 99-106.

Telford RJ, Birks HJB. 2011b. QSR Correspondence "Is spatial autocorrelation introducing biases in the apparent accuracy of palaeoclimatic reconstructions?" Quaternary Science Reviews, 30: 3210-3213.

Interactive comment on Clim. Past Discuss., 9, 4553, 2013.