Clim. Past Discuss., 11, C2424–C2431, 2015
www.clim-past-discuss.net/11/C2424/2015/

Climate of the Past
Discussions

Interactive comment on "A Bayesian hierarchical model for reconstructing relative sea level: from raw data to rates of change" *by* N. Cahill et al.

R. Edwards (Referee)

robin.edwards@tcd.ie

Received and published: 20 November 2015

GENERAL COMMENTS

This paper addresses the use of microfossils from salt-marsh environments as tools for relative sea-level (RSL) reconstruction. This is a subject of topical interest and an area of active research associated with several publications in high-impact journals. The subject matter is therefore of relevance and within the scope of CP. The paper uses previously published data to illustrate the application of a new Bayesian hierarchical model for microfossil-based RSL reconstruction. The new model builds upon two previously published models produced by the authors, but marks a significant step forward in two areas: 1) the development of a holistic numerical framework within which to com-

bine microfossil-based palaeomarsh-surface elevation (PMSE) reconstructions with a statistical age-depth model to produce RSL reconstructions with appropriately propagated uncertainties; 2) a formal framework within which secondary/additional proxies can be used to refine PMSE reconstructions. In this example, organic carbon isotope data from bulk sediment samples are used to refine PMSE reconstructions developed from salt-marsh foraminifera. The paper therefore presents a novel approach with great potential application in the field of RSL research (and other sediment-based palaeoecological studies). I anticipate its rapid adoption by the evolving field of quantitative palaeoenvironmental reconstruction.

The paper is well-written and material is clearly presented / explained. I note a couple of areas that I feel would benefit from minor clarification (see Specific Comments). I am unable to comment on the more technical aspects of the Bayesian statistics, but as far as I can follow the description, the approach appears logical and the conclusions are generally sound. I have some minor technical comments relating to the model itself and an additional couple of comments/questions relating to the interpretations and conclusions presented at the end of the paper (see Specific Comments). The bulk of these comments are arguably linked to attributes of the particular dataset used to illustrate the approach and therefore do not detract from the underlying quality of the hierarchical model which is the main focus of the paper. I recommend that the paper is published subject to minor correction / response to the specific and technical comments outlined below.

SPECIFIC COMMENTS

Note: An important caveat to the following points is that there is currently no consensus in the literature regarding how best to compile an appropriate modern training set for RSL reconstruction. It is important to acknowledge that the authors use a modern analogue test and are working with the best material available to them. That said, I think that some aspects of the modern training set have an influence on some of the results/discussion in the paper, and I have tried to restrict my comments to those aspects. Ultimately, some of the points made may only be demonstrable by comparison using synthetic/'ideal' data. I am not suggesting this is required for this paper (although it may be a useful avenue for future work), but the conclusions should more explicitly acknowledge the potential influence that the training set may have had on the results.

S1: Composition of the training set & species response curves.

The modern dataset contains marshes of different 'types' reflecting their physiographic setting and environmental conditions (e.g. salinity). This is expressed as different high marsh assemblages of foraminifera (Section 4.1.1). Certain taxa are restricted to particular marshes whilst others are more cosmopolitan. For the cosmopolitan species their response curves are developed in the presence of different taxa at each site with the result that they are likely to express different ecological optima and tolerances (ie a modified realised niche). In other words, the 'elevation' signal produced by an individual taxon may vary between marsh 'types'.

Whilst it is technically possible to combine distributions of cosmopolitan species from multiple sites within a TF (Figure 4 and associated discussion about 'ecological plausibility') the net result may be rather misleading / uninformative. This is particularly the case when discussing the extent to which one TF variant or the other accurately captures the form of the response curve.

I suspect what we are seeing in several cases is a composite response curve. I think that Section 5.1.1 needs some revision in light of this, especially when referring to 'ecological plausibility'. To assist with this, it would be useful to have a plot that showed the species distributions against SWLI – perhaps a modification of Figure 2? In many instances there will be multiple samples for a given SWLI, so this need not be a strict plot along a SWLI axis, but rather a general ordering by height (I suspect Fig. 2 is already part of the way there). In this way, consistent vertically zoned foraminifera will have much less 'noisy' distributions than those taxa that occur across a range of heights (poor elevation indicators). Given that the SWLIs at the top end of the

gradient are really high (140+!) there could be additional issues with correct alignment/standardisation due to the use of MHHW as the upper datum (cf. Highest Occurrence of Forams – see Wright et al., 2011), or the reworking of material. This may show up in a replot (although coding by site would really be needed to pull this out, and that may not be graphically feasible).

As a final note, some of the samples look decidedly odd and, whilst they may not screen out on statistical grounds (ie count size) could have a distorting effect on the B-TF plots. For example, the secondary increase in probability of occurrence at the top of the elevation gradient for M. fusca is not ecologically plausible given what we know about its distribution: this has all the hallmarks of an 'in-wash' signal? One strength of fitting an underlying response curve (ie WA-TF) is that it is less susceptible to being pulled by this kind of outlier.

S2: Comparison of the TF performance (Section 5.1.2, Fig. 5, bits of the discussion.)

From Fig. 5 there appears a large spread in the residuals for the B-TF, but the text says the WA-TF has a larger average 2sigma uncertainty. Is this correct? Eyeballing the performance graphs would lead me to prefer the WA-TF. There are a couple of large outliers in the B-TF and the clustering around the key high marsh values ($\sim$100 SWLI) of the WA-TF seems tighter? Is the higher av. uncertainty for WA-TF driven by the consistent under/over prediction at the ends of the environmental gradient? These are well-known edge effects that can be addressed in some instances through the use of WA-PLS. The scatter at the top end is also potentially the fingerprint of in-washing etc and links into the final comment above. Sometimes outliers are important as they are indicators of issues with the training set. It is perhaps a little over-simplistic to say that the B-TF performs better close to the extremes of the gradient – it may just be easily fooled by dodgy data. It would perhaps also be useful to include as dashed lines the size of the 2sigma uncertainty band on the residual plots: ultimately if estimates are within error, they are fine.

S3: Variability does not mean accuracy The paper makes an excellent point regarding the general insensitivity of the WA-TF error envelope and the fact that the B-TF has greater capacity to recognise intervals in which reconstructions are possible with greater (or lesser) degrees of precision. However, I did not follow the logic of the section in the discussion which equates the presence of variability in the PME reconstruction as being more accurate, simply because the foram assemblages exhibit variability. This implies that even subtle changes in foram assemblage must translate into a reliable elevation signal. This is clearly not the case, especially given the nature of the modern training set as discussed above. Both cores are dominated by T. inflata and J. macrescens which lack strong elevation signals in the modern training set (Fig. 4).

The variability expressed by the B-TF may simply be an expression of its sensitivity to 'noise' in the data which it extracts as an (erroneous) elevation signal. The statement of Engelhart & Horton (2012) that 'high marsh' forms between MHW and HAT is a useful back-of-the-envelope generalisation, but not an appropriate measure against which to assess TF performance. There is no basis for concluding that the core material analysed had to accumulate across the full height range within which high marsh sediments may form. Similarly, as shown in Fig 7, the B-TF scatters around the instrumental data but shows no correlation with real rate changes and appears to overestimate variability. The impact of adding the δ13C is actually to dampen the variability back down again.

S4: The abstract / text refers to the model as reconstructing RSL with "fully quantified uncertainty." Whilst I understand what the authors are saying here, it is perhaps worth explicitly noting at some point in the text that there are many sources of uncertainty in the resulting RSL that are not quantified (e.g. the influence of sediment compaction, tidal range changes, GIA, altered species-environment response, taphonomic effects etc etc). These uncertainties are inherent to all microfossil-based approaches and so are not particular problems unique to the material presented here. However, given that the outputs from this kind of model are likely to be referred to / adopted by scientists outside the 'palaeo' world, it may be useful making this point somewhere in the

manuscript.

S5: Evidence for decadal- multidecadal reconstructions (pg 4879) This may be overstating things a little. There is actually limited evidence that the TF picks out changes evident in the instrumental record. I think if this is to be demonstrated there needs to be a bit more analysis. What is the temporal resolution of the TF record (c. 1 sample / 5-10 yrs?) What happens if you smooth the instrumental record to emulate this resolution – at the moment its annual right?

Is there any added significance about the mid-point vs any other point within the error envelope (ie is a reconstruction more likely to be correct at the mid point or is the 'true' value equally likely to reside anywhere within the error envelope)? Is this the same for WA-TF, B-TF and multi-proxy B-TF?

TECHNICAL COMMENTS

T1: Clarification is required regarding 'layer thickness' and the manner in which the various data are combined within the modelling framework. Are foraminifera assumed to be surface indicators (1 cm thick)? Is the bulk organic isotope data assumed to represent the signature of a ?1 cm thick surface layer? Is the organic material used for radiocarbon dating also assumed to represent the past marsh surface, or is a palaeomarsh surface correction applied to deal with the fact that sub-surface plant fragments have been used? If so, has an uncertainty term(s) been included in this (and if so what)? This latter correction does influence the uncertainties attached to Bchron and the output from the 'Chronology Module'.

T2: Clarification is needed regarding how uncertainty associated with floral distributions, tide levels and δ13C is treated within the modelling framework. The paper appears to use 'hard' cut-off values of a certain ‰ depletion as thresholds discriminating three elevation zones. Is this correct? Floral zones are known to exhibit (sometimes significant) variability with respect to tidal elevation: indeed, this was one of the drivers for the use of microfossils which were suggested to have better constrained elevation relationships. The manuscript states “The inclusion of the δ13C did not treat MHHW as a hard bound for PME..” so I am guessing that something is going on behind the scenes? It would be useful to have a couple of words on this, especially given that the improvement in performance is solely due to the inclusion of δ13C as a second proxy.

T3: What is the reason for using abundance (ie number of foraminifera) rather than relative abundances (ie proportion of the total count) in the modelling framework? With the exception of the decrease in foraminiferal abundance at the upper limit of marine influence, there is no environmental/ecological significance (that I am aware of) of absolute numbers of forams in a sediment, nor any particular linkage with elevation. Does the model use count size to assess the statistical significance of the foram data such that very low counts are treated with greater uncertainty than counts of several hundred individuals? Given that it is standard practice to use relative abundance in TF approaches, it would be useful to clarify the rationale behind using absolute counts.

T4: Clarification is required regarding how the two cores are combined to produce the RSL reconstructions (ie how does Fig 3 translate to Figs 6&7?). How is the overlap dealt with? Are all the data included? Given that age-depth relationships for both cores will be different, I’m not sure I follow how the ‘chronology module’ handles this. Similar question for how the dynamic evolution of RSL works when you splice records? As a more general question, why does the uncertainty envelope expand toward the present? Surely we know age – elevation precisely for core top and this should squeeze errors down?

Technical Corrections

Abstract Ln 12: New Jersey, USA.

Pg4854; Ln5-6: “..estimate the age of undated layers with uncertainty.” You mean with a quantified uncertainty term right?

Fig 2: See comments in S1

Fig 3: Can samples without analogues be plotted on the figure but shaded differently? It would be interesting to see how the WA-TF vs B-TF handle no analogue situations. Tc needs to be spelt out in full. Is it possible to indicate dated points for reference just to get a feel for the chronological control?

Fig 5: See comment in S2

Interactive comment on Clim. Past Discuss., 11, 4851, 2015.