Clim. Past Discuss., 11, C2904–C2915, 2016
www.clim-past-discuss.net/11/C2904/2016/

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.

N. Cahill et al.

niamh.cahill.2@ucdconnect.ie

Received and published: 14 January 2016

We would like to thank the reviewer for his detailed and well informed response. All of the comments are useful and valid. However, in some cases the comments are very specific to the example case study and adding detailed discussion to the manuscript will take away from the focus of the paper i.e., the new B-TF model. Therefore, we suggest that in some instances including references that will direct readers to the relevant literature is more appropriate.

The reviewer comments are in red, with our response immediately below in plain black font.

Specific Comments:

It is important to acknowledge that the authors use a modern analogue test and are working with the best material available to them.

We will add the following text to the beginning of section 4.1 to ensure that this is clear.

'The transfer function approach is under-pinned by foraminiferal surveys of modern taxa that seek to reliably quantify the relationships between individual species and sample elevation relative to the tidal frame. The precise and accurate quantification of these vertical relationships is fundamental to the reliability of the resulting transfer function and its reconstructions.'

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'.

This comment is indirectly referring to a long running debate in the sea-level literature about using local scale vs regional scale training sets since both have advantages and limitations. In this paper we apply the B-TF model to the regional dataset to be consistent with the original analysis presented in Kemp at al., 2013. To address this comment we will add a reference to the relevant Kemp et al, 2013 paper where the regional vs local debate is discussed for this specific dataset. We will also update Section 4.1 of the manuscript to include the following text:

'Ideally, surface foraminiferal samples should be collected from a range of elevations and depositional environments analogous to those represented in the fossil material under investigation (Horton and Edwards, 2005). Kemp et al. (2013) compiled a modern training set from twelve salt marshes in southern New Jersey (Figure 1). The twelve sites were selected to span a wide range of physiographic settings including brackish marshes located up to 25 km from the coast with a strong fluvial influence. The sites share a common climate and oceanographic regime and therefore constitute a regional-scale training set. The aim of a regional scale approach is to gather a wide range of modern analogues that will permit the transfer function to perform reliably even if past environmental conditions at the study site differed significantly from those operating at the present day (Horton and Edwards, 2005).'

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?

We agree with the reviewer and we note that the suggestion to modify Figure 2 was also made by reviewer 1. Therefore, we have produced an updated version of Figure 2 where samples are organized by elevation in SWLI units. We will add the necessary discussion to section 5.1.1 as suggested.

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.

A similar comment was also made by reviewer 1 and so we reiterate some of our thoughts here. The small uptake in probability for M.fusca occurs because of 3 samples found between 130 and 135 SWLI. The revised Figure 2 shows that these samples are unusual. This comment again indirectly refers to a debate in the literature concerning the degree to which samples should be screened. It is left to the judgement of the researcher to decide whether or not to screen unusual samples or retain them. We recognize that the uptake in the response is an artefact of using the B-TF model. The B-TF does not assign a pre-determined response curve and therefore, as suggested by the reviewer, is more susceptible to these 'outliers'. However, we did not remove the unusual samples because they were not removed from the original analysis presented in Kemp et al., 2013.

To help with addressing this comment we will update the relevant parts of the discussion to include the following text:

'While a more flexible response curve is more appropriate from a statistical point of view i.e., we are allowing the data to determine the shape of the curve, this may result in outliers being given undue importance in determining the trend. An advantage of assuming a pre-determined response curve (i.e., Gaussian for the WA-TF) is that it is less susceptible to outliers.'

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

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.

Yes, this statement is correct. The 2σ uncertainty is not referring to the spread of the residuals, rather the uncertainty intervals for each predicted elevation (shown as vertical lines on Figure 5). We will update the manuscript with a clarification as follows:

'The average of the 2σ prediction uncertainties (shown as vertical lines on Figure 5) is larger for the WA-TF than for the B-TF. This suggests that while the WA-TF is more successful than the B-TF at capturing the true elevation within the prediction uncertainty bounds, part of the reason for this may be because the intervals are larger. Prediction performance is often penalised for large intervals (e.g., Gneiting and Raftery, 2007).'

There are a couple of large outliers in the B-TF and the clustering around the key high marsh values (100 SWLI) of the WA-TF seems tighter?

We agree that the WA-TF appears 'tighter' around 80-100 SWLI than the B-TF. We suggest that the WA-TF performs better here because the average SWLI for the modern dataset is ~ 90, and, by the very nature of the WA-TF, estimates of elevation will be drawn towards the average SWLI value. We will make a note of this in the

results.

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.

The 2σ uncertainties that we refer to are the uncertainties surrounding the individual predicted elevations. The top panel includes these uncertainty bounds and we state in the text that for the B-TF 90% of the time the true value is within the prediction interval vs. 92% for the WA-TF.

S3: Variability does not mean accuracy.

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.

We agree with the reviewer. We will change our language to say that the B-TF is more sensitive to population changes (i.e., if foraminifera assemblages change then PME should change). We will remove accuracy statements from the discussion text.

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 δ^{13}C is actually to dampen the variability back down again.

We feel that by addressing the previous comment and removing accuracy statements we have also dealt with this comment i.e., the B-TF is more sensitive to changes but that does not imply greater accuracy.

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

We agree with the reviewer and we will make it clear in the text that there are many sources of uncertainty that have not been quantified.

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.

We agree with the reviewer regarding this point. We do not wish to overstate our conclusions and have removed some text from this page accordingly. However, based on the Mean Squared Error (MSE) results described in Section 5.2 we will retain a statement saying that the best reconstruction is obtained from the multi proxy B-TF model.

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?

Yes the mid-point has a higher probability density and the true value is not equally likely to occur everywhere in the 'error envelope'. The probability density deceases as you move towards the bounds of the 95% uncertainty region. However, upon reading this comment we have decided to update Figure 7 to display the uncertainties with boxes, which helps to illustrate that the true value can in fact occur with some probability throughout the entire uncertainty region. An example of displaying RSL reconstructions in this manor can be found in Cahill et al., 2015.

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 palaeo-marsh 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'.

We use the data as presented by Kemp et al., 2013. Sample layers are 1cm thick for all analyses (foraminifera and δ^{13}C, modern and fossil). We use the age-depth model without modification or reinterpretation from the relevant Kemp et al., 2013 publication. Further discussion can be found there. We will add a sentence to refer readers that are interested in the nuances and specifics of age-depth modelling salt marsh sediments using plant macrofossils back to the original papers.

T2: Clarification is needed regarding how uncertainty associated with floral distributions, tide levels and δ^{13}C is treated within the modeling framework.

We will update Section 3.3 to more explicitly state how $\delta^{13}C$ is treated within the modelling framework. We are happy to add references to the discussion that will refer readers to areas of debate and uncertainty surrounding the use of δ^{13}C.

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?

The use of count data is necessary for the model we have chosen to use. We assume that the data follow a multinomial distribution and this particular distribution requires count data rather than relative abundance data. We will update the manuscript to clarify this.

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?

All three modules of the model are applied to both cores separately and then the results are simply combined. We will add text to the the first paragraph of Section 5 to clarify this.

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?

Assuming that the reviewer is referring to the expanding uncertainty in the EIV-IGP rate results in Figure 7, this is an edge effect that occurs as a results of using a Gaussian process model. The Gaussian process model learns about the present based upon what it knows about past and future observations. Therefore, for the most recent observations we have less data (i.e., no future data) on which to base our rate estimates, resulting in an increase in uncertainty.

Technical Corrections

We will address the technical corrections and we would like to thank the reviewer for taking the time to read the paper in enough detail to bring these to our attention.

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

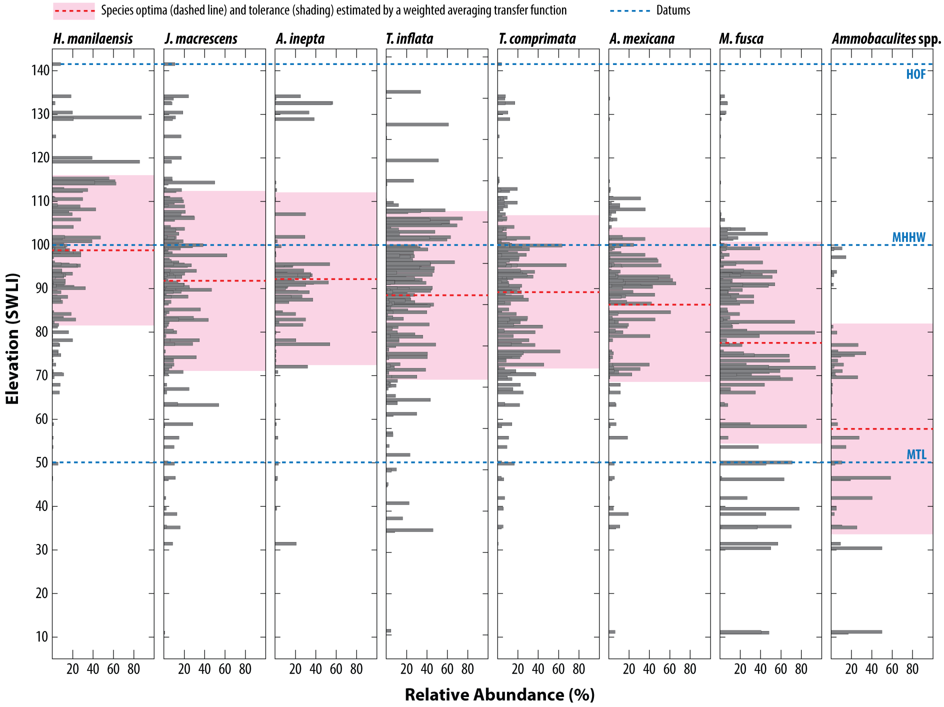

Fig. 1. Updated Version of Manuscript Figure 2

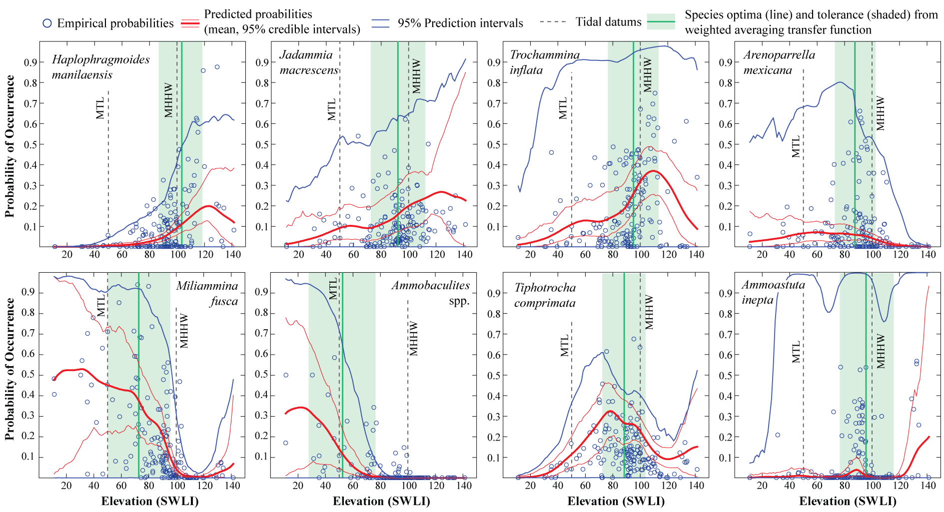

Fig. 2. Updated Version of Manuscript Figure 4