

# Response to Reviews on A Bayesian Hierarchical Model for Reconstructing Relative Sea Level: From Raw Data to Rates of Change

January 24, 2016

We would like to thank the reviewers for their clear and constructive comments. Each comment helped to improve the quality of this paper. Please see below our response. **The reviewer comments are in red, with our response immediately below in plain black font.**

## In response to Reviewer 1

The values for the priors used in the model are not discussed in detail in this paper. I know there are many and at least some are discussed in previous work. I view this as important as some previous papers presenting Bayesian transfer functions have used very informative priors on the reconstructions, perhaps leading to artificially inflated cross-validation performance. This paper should make it clear that this is not case here.

The second last paragraph of Section 3.1 (Page 8) has been updated to include the following text:

*'A uniform prior is supplied for unknown PME suggesting all elevations in a given range are equally likely. For our New Jersey case study we chose a range of 0 to 150 SWLI. This range is specific to this case study and would need to be updated at different locations.'*

Section 3.2 (Page 8) has been added to the manuscript and includes a description of the priors for the  $\delta^{13}\text{C}$  proxy, noting that these are also specific to this case study. The priors in relation to the P-spline parameters are discussed in Section 3.1.

Figure 2 would be more informative if the SWLI were given, and the observations sorted by SWLI. I presume the observations are currently sorted by cluster rather than SWLI, which gives a misleading impression of how noisy the data are. The clusters add little in anything to the argument.

We have produced an updated version of Figure 2. In the new figure, samples are organized by SWLI as suggested by the reviewer.

I suspect the optima presented in Figure 4 are actually WAPLS-2 beta coefficients. It is misleading to present WAPLS-2 beta coefficients as if they were optima as they include a correction that accounts for secondary gradients and (mostly) edge effects. The WA optima could be shown.

We would like to thank the reviewer for pointing this out. The values presented are the coefficients for WA-PLS-1. We have re-run the analysis to obtain the species optima and tolerances and included them on Figure 2. Figure 4 and Section 5.1.1 have been updated accordingly. This update does not change the discussion or conclusions of the manuscript.

I don't understand how empirical probability of occurrence is being used when both the calibration and fossil data are relative abundance data.

The B-TF does not use relative abundance data it uses the raw species counts. We apologise that this was not made clear and have made edits throughout the manuscript to address this. The following text has also been added to the end of Section 4.1.1 (Para 2, Page 11).

*'However, an important difference is that the B-TF uses raw counts of foraminifera as the input, while the WA-TF model uses relative abundances expressed as percentages.'*

Two of the species show an uptick in probability of occurrence in at lowest SWLI which are ecologically questionable. Could these be an artefact?

We have made a note of the 'uptick' in Section 6 (Para 1, Page 23) by adding the following text:

*'However, our results suggest that the flexibility in the B-TF can, in some instances, make it more susceptible to unusual observations. For example, the *Miliammina fusca* response curve in the B-TF shows a slight increase in probability of occurrence above 120 SWLI (Figure 4) because of three unusual samples located above 130 SWLI in which this species occurs (Figure 2). This distribution may represent the advantage of a regional-scale training set in capturing natural variability caused by ecological complexity and/or the influence of secondary environmental variables. Equally these samples could be an anomalous occurrence (e.g., tests washed in by a storm) that ought to be screened prior to analysis. Since Kemp et al. (2013a) retained these samples, we also elected to use them in developing the new B-TF.'*

Would it be possible to include information on salinity to further constrain the transfer function? Or are there insufficient data to do this well?

This is a good suggestion, unfortunately, salinity data wasn't collected and isn't available for this case study. However, in Section 4.1.1 (Page 12) we have referred readers to the Kemp et al., 2013 paper that includes a qualitative discussion about salinity and the foraminifera species specific to this case study.

## In response to Reviewer 2

### 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 have added the following text to the beginning of Section 4.1(Para 1, Page 10) to ensure that this is clear.

*‘The transfer function approach is under-pinned by surveys that quantify the modern, observable relationship between foraminifera and tidal elevation. Modern (surface) samples should be collected from depositional environments analogous to those represented in the fossil material under investigation (Horton and Edwards, 2005).’*

#### **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 et al., 2013. We have updated Section 4.1 (Para 2, Page 10) of the manuscript to include the text below and referenced the relevant Kemp et al, 2013 paper where the regional vs local debate is discussed for this specific dataset.

*‘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 spatial scope of modern training sets has frequently been discussed and it is widely accepted that regional-scale training sets capture natural variability in the distribution of foraminifera and provide a large suite of samples from which to draw modern analogs (e.g., Horton and Edwards, 2005). The principal advantage of local-scale training sets is that they produce more precise reconstructions, but at the expense of offering only a narrow range of modern analogs (e.g., Horton and Edwards, 2005). Kemp et al. (2013b) described and discussed the effect of using regional- and local-scale modern training sets to reconstruct RSL in New Jersey and concluded that it was necessary to use the regional data to provide sufficient modern analogs for interpreting assemblages*

*of foraminifera preserved in sediment cores. This led Kemp et al. (2013a) to use the regional-scale training set when they produced Common Era RSL reconstructions from two locations in southern New Jersey and their analysis showed that the closest modern analogs for core samples were drawn from 11 of the 12 sites in the modern training set. Since it is not our aim to investigate the relative utility of local and regional-scale modern training sets we use the same data as Kemp et al. (2013a) to ensure that the new B-TF can be compared directly with existing results.'*

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.

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 uptick in *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 judgment 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.

Please see the response to Reviewer 1 for how and where we addressed this comment in the manuscript.

## **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  $2\sigma$  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 have updated Section 5.1.2 (Para 1, Page 18) with the following clarification:

*'The average of the  $2\sigma$  prediction uncertainties is slightly larger for the WA-TF (28 SWLI) than for the B-TF (21 SWLI). 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  $\sim 90$ , and, by the very nature of the WA-TF estimates of elevation will be drawn towards the average SWLI value. We have made a note of this in Section 5.1.2 (Para 2, Page 19).

*'In contrast, the WA-TF performs better at elevations between 80 to 100 SWLI because the distribution of samples in the modern training set is focused at  $\sim 90$  SWLI, and, by the very nature of the WA-TF, estimates of elevation will be drawn towards this average value.'*

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  $2\sigma$  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 have changed our language to suggest that the B-TF is more sensitive to population changes (i.e. if foraminifera assemblages change then PME should change.). We have removed 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  $\delta^{13}\text{C}$  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 have added the following text to Section 5 (Page 15).

*'It should be noted however that other sources of uncertainty that are inherent in RSL reconstructions (e.g., the influence of sediment compaction or tidal range changes) were not considered or accounted for.'*

**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 MSE results described in Section 5.2 (Para 2, Page 20) 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 decreases as you move towards the bounds of the 95% uncertainty region. However, upon reading this comment we have updated Figure 7 to display the uncertainties as boxes to help illustrate that the true value can in fact occur with some probability throughout the entire uncertainty region.

## 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 1-cm thick for all analyses (foraminifera and  $\delta^{13}\text{C}$ , 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 have updated the relevant parts of the manuscript to make this clear.

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

Section 3.2 (Page 8) has been added to the manuscript and includes a description of the priors for the  $\delta^{13}\text{C}$  proxy, noting that these are also specific to this case study.

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 and it offers the advantage over relative abundance since the resulting multinomial distribution can take account of the fact that large counts give reduced uncertainty. We have updated Section 3 (Para 1, Page 4).

*'A B-TF model using a penalized spline (P-spline) as a non-parametric model of the multinomial response of foraminifera counts to tidal elevation. The multinomial distribution can take account of the fact that large counts give reduced uncertainty...'*

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 have added the following text to Section 5 (Page 15) to make this clear:

*'The Cape May Courthouse and Leeds Point RSL reconstructions were developed separately to one another and were then merged into a single, regional-scale dataset for analysis...'*

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 result of using a Gaussian process model. The Gaussian process model learns about the present rates based upon what it knows about past and future sea level 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. To address this comment we have added the following text to Section 5.2 (Para 1, Page 20).

*'All reconstructions show an expanding uncertainty for the most recent rates. This is an edge effect that occurs as a result of using a Gaussian process model on the rates. These rates are estimated based upon their relationship (through an integration step) to the observations. As we approach the edge of the data, there are fewer constraints from the observations. This means that there are multiple rate trajectories that are consistent with recent data and so the uncertainty in the rate process increases.'*

## Technical Corrections

We did not analyse samples with no modern analogues to insure consistency with the original Kemp et al. 2013. analysis. We have made this clear throughout the manuscript. We have defined  $T_c$ =Tiphotrocha comprimata in the Figure 3 description. We have addressed all other 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.

## List of Changes

Locations of Relevant changes as stated in the above response are listed below and highlighted in red in the marked-up manuscript version.

- Section 3.1 (Page 8)
- Section 3.2 (Page 8)
- Section 4.1.1 (Para 2, Page 11)
- Section 6 (Para 1, Page 23)
- Section 4.1.1 (Page 12)
- Section 4.1 (Para 1, Page 10)
- Section 4.1 (Para 2, Page 10)



- Section 5.1.2 (Para 1, Page 18)
- Section 5.1.2 (Para 2, Page 19)
- Section 5 (Page 15)
- Section 5.2 (Para 2, Page 20)
- Section 3.2 (Page 8)
- Section 3 (Para 1, Page 4)
- Section 5 (Page 15)
- Section 5.2 (Para 1, Page 20)

# A Bayesian Hierarchical Model for Reconstructing Relative Sea Level: From Raw Data to Rates of Change

Niamh Cahill<sup>1</sup>, Andrew C. Kemp<sup>2</sup>, Benjamin P. Horton<sup>3</sup>, and Andrew C. Parnell<sup>1</sup>

<sup>1</sup>School of Mathematical Sciences, CASL, Earth Institute, University College Dublin

<sup>2</sup>Dept. of Earth and Ocean Sciences, Tufts University

<sup>3</sup>Department of Marine & Coastal Sciences and Institute of Earth, Ocean, & Atmospheric Sciences, Rutgers University

<sup>4</sup>The Earth Observatory of Singapore and the Asian School of the Environment, Nanyang Technological University

*Correspondence to:* Niamh Cahill (niamh.cahill.2@ucdconnect.ie)

**Abstract.** We present a Bayesian hierarchical model for reconstructing the continuous and dynamic evolution of relative sea-level (RSL) change with quantified uncertainty. The reconstruction is produced from biological (foraminifera) and geochemical ( $\delta^{13}\text{C}$ ) sea-level indicators preserved in dated cores of salt-marsh sediment. Our model is comprised of three modules: (1) A new Bayesian transfer function for the calibration of foraminifera into tidal elevation, which is flexible enough to formally accommodate additional proxies (bulk-sediment ( $\delta^{13}\text{C}$ )); (2) An existing chronology developed using the Bchron age-depth model, and (3) An existing errors-in-variables integrated Gaussian process (EIV-IGP) model for estimating rates of sea-level change. Our approach is illustrated using a case study of Common Era sea-level variability from New Jersey, U.S.A. We develop a new Bayesian transfer function (B-TF) using foraminifera, with and without the additional ( $\delta^{13}\text{C}$ ) proxy and compare our results to those from a widely-used weighted-averaging transfer function (WA-TF). The formal incorporation of a second proxy into the B-TF model results in smaller vertical uncertainties and improved accuracy for reconstructed RSL. The vertical uncertainty from the multi-proxy B-TF is  $\sim 28\%$  smaller on average compared to the WA-TF. When evaluated against historic tide-gauge measurements, the multi-proxy B-TF most accurately reconstructs the RSL changes observed in the instrumental record (mean square error =  $0.003 \text{ m}^2$ ). The Bayesian hierarchical model provides a single, unifying framework for reconstructing and analysing RSL change through time. This approach is suitable for reconstructing other paleoenvironmental variables (e.g. temperature) using biological proxies.

## 1 Introduction

Paleoenvironmental reconstructions describe Earth's response to past climate changes and offer a context for current trends and analogs for anticipated future changes (e.g., Mann et al., 2009). Reasoning by analogy underpins the use of biological proxies for paleoenvironmental reconstructions (e.g., Rymer, 1978; Jackson and Williams, 2004; Bradley, 2015). The ecological preferences of biological proxies observed in modern environments are used to derive a paleoenvironmental reconstruction from their counterparts preserved in dated sediment cores under the assumption that their ecological preferences were unchanged through time (e.g., Juggins and Birks, 2012). This approach commonly utilizes data consisting of one environmental variable and counts from multiple proxy species (e.g., Imbrie and Kipp, 1971; Fritz et al., 1991; Birks, 1995). Numerical techniques known as transfer functions formalize the relationship between biological assemblages and the environmental variable. This step is termed calibration. To quantify environmental change through time it is necessary to combine the paleoenvironmental reconstruction with a chronology of sediment deposition and an appropriate methodology to describe temporal trends. These three components can be developed and applied independently of one another or assimilated in a single framework.

Relative sea-level (RSL) reconstructions can constrain the relationship between temperature and sea level and reveal the long-term, equilibrium response of ice sheets to climate forcing (e.g., Dutton et al., 2015). Some of the most accurate and precise RSL reconstructions are produced from salt-marsh foraminifera. Foraminifera are used as sea-level proxies because species have different ecological preferences for the frequency and duration of tidal submergence, which is primarily a function of tidal elevation (e.g., Scott and Medioli, 1978; Horton and Edwards, 2006; Edwards and Wright, 2015). Under conditions of RSL rise, salt marshes accumulate sediment to maintain an elevation in the tidal frame. The resulting sedimentary sequence is an archive of past RSL changes that may be accessed by collecting sediment cores. After extraction, these sediment cores are sliced into layers (samples), from which foraminifera are counted. The transfer functions commonly used to reconstruct RSL impose a single ecological response to tidal elevation on all species of foraminifera (or other biological groups such as diatoms). Other analyses performed on the same layers can provide a multi-proxy approach to reconstructing RSL, although this often relies on informal approaches to combine results from independent proxies (e.g., Kemp et al., 2013a; Gehrels, 2000; Nelson et al., 2008). For example, on organogenic salt marshes on the U.S. Atlantic coast the primary source of organic carbon is *in-situ* plant material and measurements of bulk sediment  $\delta^{13}\text{C}$  reflect the dominant plant community (e.g. Kemp et al., 2012). Since salt-marsh plants are also sea-level proxies the  $\delta^{13}\text{C}$  values can be used to reconstruct RSL. Some sediment layers are dated using radiocarbon or recognition of pollution and vegetation clearance markers of known age. Since there are typically fewer dated layers than total layers, a statistical age-depth model is used to estimate the age of undated layers with uncertainty (e.g., Bronk Ramsey, 2008; Haslett and Parnell, 2008; Blaauw and Christen,

2011). Although Bayesian age-depth models and methods for estimating rates of sea-level change already exist, Bayesian methods are yet to be applied in the calibration phase of reconstructing RSL, preventing the appropriate propagation of uncertainties.

We develop a Bayesian transfer function (B-TF) to reconstruct RSL using foraminifera (**expressed as raw counts**) and measurements of bulk sediment  $\delta^{13}\text{C}$  from salt-marsh sediment. This model allows each species of foraminifera to have a different ecological response to tidal elevation and provides a formalized approach to combine multiple proxies and consequently reduce reconstruction uncertainty. Following the framework of Parnell et al. (2015) we combine this new B-TF with an existing chronology module (Bchron), and an existing process module (the Errors-In Variables Integrated Gaussian Process (EIV-IGP) model of Cahill et al., 2015) to create a holistic Bayesian hierarchical model. Through application of the model to a case study of Common Era and instrumental RSL change in New Jersey (U.S.A), we compare the utility of the B-TF with an existing weighted averaging transfer function (WA-TF) approach.

## **2 Previous calibration methods**

Transfer functions are empirically-derived equations for reconstructing past environmental conditions from the abundance of multiple species. The term refers not to a single numerical method, but to a range of regression-based techniques that are classified into two categories depending on whether the underlying model maps environmental variables to species abundances (classical calibration) or vice versa (inverse calibration). Classical approaches are underpinned by the ecologically-intuitive assumption that the distribution of species is driven by environmental variables (e.g., Birks, 2012). Inverse approaches initially gained popularity because of their reduced computational complexity (e.g., Birks, 2010) resulting in quicker processing compared to classical methods. Furthermore, inverse methods often demonstrate equal or superior performance when compared to classical approaches (e.g., Toivonen et al., 2000; ter Braak and Juggins, 1993; Korsman and Birks, 1996)). The parameters in transfer functions are estimated using empirical data (a modern training set) from environments likely to be analogous to those encountered in core material (e.g., Juggins and Birks, 2012) and are treated as fixed and known. Studies seeking to reconstruct RSL from salt-marsh sediment employ transfer functions developed using a modern training set of paired observations of tidal elevation and microfossil assemblages (most commonly foraminifera or diatoms) to reconstruct RSL from their counterparts preserved in sediment cores (e.g., Horton et al., 1999; Gehrels, 2000; Edwards and Horton, 2006; Kemp et al., 2013b; Barlow et al., 2014). Although the different types of transfer function have advantages and weaknesses compared to one another, these regression-based techniques share the limitations of applying a single response form to all species and treating model parameters as fixed and known. Applying a single response form can result in misleading or inac-

curate paleoenvironmental reconstructions if the response curve is not appropriate for all species (Greig-Smith, 1983) and does not account for the inherent uncertainty in model parameters that results from ecological noise and the influence of secondary environmental variables, which in RSL reconstructions can include salinity, pH and sediment texture and composition (e.g., Shennan et al., 1996; Zong and Horton, 1999; Horton et al., 1999).

Bayesian calibration methods are inherently classical and have recently been given growing attention to produce paleoenvironmental reconstructions using biological proxies (e.g., Toivonen et al., 2000; Vasko et al., 2000; Haslett et al., 2006; Li et al., 2010; Tingley et al., 2012; Tolwinski-Ward et al., 2013, 2015; Parnell et al., 2015). Toivonen et al. (2000) and Vasko et al. (2000) developed a Bayesian model to reconstruct temperature from chironomid counts. Haslett et al. (2006) adopted elements of the model proposed by Toivonen et al. (2000) in a more complex Bayesian hierarchical model for reconstructing multivariate climate histories from pollen counts. Li et al. (2010) proposed a Bayesian hierarchical model to reconstruct temperature from pollen using a multi-proxy approach. Similarly, Tingley et al. (2012) considered a Bayesian hierarchical space-time model for inferring climate processes. More recently, Tolwinski-Ward et al. (2013, 2015) and Parnell et al. (2015) expanded on the aforementioned approaches of Haslett et al. (2006) and Tingley et al. (2012) for reconstructing climate variables. To date, Bayesian methods have not been used for reconstructing RSL using biological proxies.

### **3 A Bayesian hierarchical model for reconstructing and analysing former sea levels**

We now describe our statistical model, which produces estimates of RSL and associated rates from raw inputs including foraminifera counts and radiocarbon dates (or dates produced from recognition of pollution and vegetation clearance markers of known age) from a sediment core. There are two advances from existing approaches:

1. A B-TF model using a penalized spline (P-spline) as a non-parametric model of the multinomial response of foraminifera counts to tidal elevation. **The multinomial distribution can take account of the fact that large counts give reduced uncertainty** and the P-spline model allows for multi-modal and non-Gaussian species response to environmental variables;
2. A full hierarchical model which incorporates the B-TF, a chronology model accounting for time uncertainty, and a rich stochastic process for quantifying sea level rate changes.

We use the JAGS package (Just Another Gibbs Sampler; Plummer, 2003) to fit the model via Gibbs sampling.

We start by outlining our notation:

- $y$  are the observed foraminifera (expressed as raw counts) from the sediment core.  $y_{il}$  is the count of species  $l$  in layer  $i$ . We denote  $y_i$  the L-vector of foraminifera counts for each layer  $i$  in the sediment core, where  $i = 1, \dots, N$  layers and  $l = 1, \dots, L$  species;
- $r$  are the observed radiocarbon dates in the sediment core.  $r_k$  is the  $k^{th}$  radiocarbon date,  $k = 1, \dots, K$ . Usually  $K \ll N$ . Due to the nature of radiocarbon, these are given in radiocarbon years rather than calendar years. A known calibration curve is used to transform the radiocarbon ages into calendar ages as part of the chronology model;
- $d$  are the observed depths in the sediment core.  $d_i$  is the depth associated with layer  $i$ ;
- $e$  is paleo marsh elevation (PME), which is the tidal elevation at which a layer originally accumulated.  $e_i$  is the PME for sediment core layer  $i$ ;
- $s$  is RSL.  $s$  has a deterministic relationship with  $e$  and  $d$  given some fixed parameters  $\omega$  so that  $s = g_\omega(e, d)$ . Producing  $s$  will require correcting PME for sample tidal elevation (a function of sediment core depth).  $s_i$  is the RSL for sediment core layer  $i$ ;
- $t$  represents the calendar ages (in years before present (1950); BP) of all layers in the sediment core. It is unknown and estimated with uncertainty as part of the chronology module from the radiocarbon dates  $r$  (and non radiocarbon dates) and observed depths  $d$ .  $t_i$  represents the age of sediment core layer  $i$ ;
- $y^m$  are the observed modern foraminifera counts.  $y_{jl}^m$  is the abundance of species  $l$  in surface sample  $j$ .  $y_j^m$  is an L-vector of modern foraminifera counts for modern sample  $j$  with  $j = 1, \dots, J$  modern samples.  $T_j = \sum_{l=1}^L y_{jl}^m$  are the row totals of species counts for calibration sample  $j$  in the matrix of species abundances;
- $e^m$  are the observed modern tidal elevations.  $e_j^m$  is the tidal elevation for surface sample  $j$ . Together  $y^m$  and  $e^m$  are used to calibrate the relationship between foraminifera abundance and tidal elevation;
- $z$  is the sediment core  $\delta^{13}\text{C}$  where  $z_i$  is the  $\delta^{13}\text{C}$  for layer  $i$ . We include this as a secondary proxy though it is an optional part of the model and can be removed if unavailable in other sediment cores;
- $\theta$  and  $\alpha$  are a set of parameters governing the relationship between foraminifera (expressed as raw counts) and tidal elevation and  $\delta^{13}\text{C}$  and tidal elevation respectively;
- $\psi$  are a set of parameters governing the sedimentation process (i.e. linking age and depth);
- $\phi$  are a set of parameters governing the RSL process, including its smoothness and variability;

Using the notation above we create a Bayesian hierarchical model to produce a posterior distribution of our parameters given data:

$$\begin{aligned}
& p(s, e, t, \theta, \psi, \phi, \alpha | y^m, e^m, y, r, d, \omega, z) \propto \\
& \underbrace{p(y|e, \theta)p(z|e, \alpha)}_{\text{Fossil Data Model}} \times \underbrace{p(y^m|e^m, \theta)}_{\text{Modern Calibration Model}} \times \underbrace{p(r|t, \psi, d)}_{^{14}\text{C Calibration}} \times \underbrace{p(t|d, \psi)}_{\text{Chronology Model}} \times \underbrace{p(s|e, d, t, \phi, \omega)}_{\text{Sea level Model}} \\
& \times \underbrace{p(\theta)}_{\text{Modern Calibration Prior}} \times \underbrace{p(\psi)}_{\text{Chronology Prior}} \times \underbrace{p(\phi)}_{\text{Sea Level Prior}} \times \underbrace{p(\alpha)}_{\delta^{13}\text{C Prior}}
\end{aligned}$$

Before describing the components of the model that we use, we note that this is an extremely complex and computationally demanding model to fit, being of very high dimension with rich stochastic processes being required for many of the sub-models. We follow Parnell et al. (2015) in making some simplifying assumptions. First, we assume that the calibration parameters  $\theta$  can be learnt solely from the modern calibration data  $y^m$  and  $e^m$ . Thus, the sediment core data contains no further information about this relationship. This is a common assumption in many palaeoclimate studies (see e.g. Haslett et al., 2006; Tolwinski-Ward, 2015). Second, we assume that the model can be modularised into three parts: the aforementioned calibration, chronology and process modules. This is a conservative assumption and follows from the restriction on the calibration parameters.

Following these assumptions, we obtain the three modules:

$$p(\theta | y^m, e^m) \propto p(y^m | e^m, \theta) p(\theta) \quad (\text{calibration module})$$

$$p(t, \psi | r, d) \propto p(r | d, t, \psi) p(t | d, \psi) p(\psi) \quad (\text{chronology module})$$

$$p(s, e, t, \theta, \psi, \phi, \alpha | y^m, e^m, y, r, d, \omega, z) \propto p(y|e, \theta) p(\theta | y^m, e^m) p(t, \psi | r, d) p(s|e, \omega, d, t, \phi) p(\phi) p(z|e, \alpha) p(\alpha) \quad (\text{process module})$$

We note that if there is no additional  $\delta^{13}\text{C}$  proxy information then  $z$  and  $\alpha$  and hence the last two terms on the RHS of the process module are removed from the equation.

### 3.1 The calibration module: multinomial P-splines (B-TF)

In this module we aim to estimate the parameters  $\theta$  that govern the relationship between foraminifera and tidal elevation by using the model as specified in the Section 3. The probability density function  $p(y_j^m | e_j^m, \theta)$  used as the likelihood here provides the data-generating mechanism from which foraminifera abundances can be simulated given tidal elevation. The likelihood we use for the modern model is:

$$y_{j1}^m, y_{j2}^m, \dots, y_{jL}^m | T_j, p_{j1}, p_{j2}, \dots, p_{jL} \sim \text{Multinomial}(p_{j1}, p_{j2}, \dots, p_{jL}, T_j), \quad (1)$$

where  $p_j = \{p_{j1}, p_{j2}, \dots, p_{jL}\}$  is the vector containing the probability of finding species  $l$  at the tidal elevation associated with sample  $j$ .

The probability vectors  $p_j$  are estimated from a latent response  $\lambda_{jl}$  (i.e. the response of species  $l$  for sample  $j$ ) which is a function of tidal elevation  $e_j^m$ .  $\lambda_l$  is a  $J$ -vector including the latent response of species  $l$  for all samples  $j$ . The relationship between probability of foraminifera species occurrence and tidal elevation is expected to be non-linear so we model these using P-splines (De Boor, 1978; Dierckx, 1993) via a softmax transformation. The softmax transformation is given as:

$$p_{jl} = \frac{\exp(\lambda_{jl})}{\sum_{l=1}^L \exp(\lambda_{jl})} \quad (2)$$

$\lambda_l$  are given  $P$ -spline prior distributions. P-splines are created from B-spline basis functions penalised to produce a smooth curve. The B-spline basis functions are constructed from piecewise polynomial functions that are differentiable to a given degree  $q$ , here cubic. The component cubic B-spline basis functions look like individual Gaussian curves, however, they will be non-zero only over the range of  $q + 2$  knots; this has numerous computational advantages. We refer to the B-spline matrix as  $B$ . The columns of  $B$  are the tidal elevations  $e^m$ , transformed by the appropriate basis function. The resulting relationship is:

$$\lambda_l = B\beta_l + \epsilon_l \quad (3)$$

$B$  is a  $J \times M$  matrix of basis functions where  $M$  is the number of knots, and  $J$  is the number of modern samples. To obtain the penalised smooth behaviour for  $\lambda_l$  we apply a prior such that the first differences of  $\beta_l$  are normally distributed with mean 0 and precision  $\tau_\beta$ . The parameter  $\tau_\beta$  controls how close the weights are related to each other and will therefore control smoothness.

An error term,  $\epsilon_l \sim N(0, \nu_l)$ , is added to the mean for  $\lambda_l$  here to ensure that we do not encounter problems with over-dispersion by under or over-estimating the variance in the observed data. We do not assume a constant variance; to account for the changing variation in the data the precision parameters  $\nu_l$  are also estimated using P-splines, i.e.,  $\nu_l = \exp(B\gamma_l)$ . This allows the variance to adapt given the data and will allow it to increase/decrease where necessary. Similarly to Equation 3, the basis functions are penalised by parameters  $\gamma_l$  to produce  $\nu_l$  and we apply a prior such that the first differences of  $\gamma_l$  are normally distributed with mean 0 and precision  $\tau_\gamma$ . Therefore, the calibration model has parameters  $\theta = \{\beta_l, \gamma_l, \tau_\gamma, \tau_\beta; l = 1, \dots, L\}$ , which can be fitted in a single Bayesian model for all species simultaneously.

The B-TF produces posterior estimates for the multinomial probability vector  $p$  for each modern sample. For each species of foraminifera, we compare the probability of species occurrence (at each modern observed tidal elevation) estimated from the B-TF, with the empirical probability of foraminifera species occurrence estimated from the observed data. The model vs. empirical probability comparison provides evidence to support the validity of the model. The comparison indicates if the model is capable of capturing the within-species variability of occurrence probabilities across



changing tidal-elevations. Once run, the B-TF can produce estimates of PME for each layer in the sediment core from this relationship. A uniform prior is supplied for unknown PME suggesting all elevations in a given range are equally likely. For our New Jersey case study we chose a range of 0 to 150 SWLI. This range is specific to this case study and would need to be updated at different locations.

We evaluate the performance of the B-TF via 10-fold cross validation on the modern data, where the data are divided up into 10 randomly drawn equal size sections (known as folds) which are removed in turn. We create predictions for the removed sections repeatedly until every observation has an out-of-sample prediction value. To allow direct and meaningful comparison between models, we also cross validated the WA-TF using the same approach on the same randomly drawn folds.

### 3.2 The inclusion of $\delta^{13}\text{C}$

In cases where the secondary  $\delta^{13}\text{C}$  proxy is available, the posterior estimated for  $e_i$  will include the likelihood  $p(z|e, \alpha)$ . This is a normal likelihood  $z_i \sim N(\mu_i, \tau_z)$  where the precision  $\tau_z$  is constant and  $\mu_i$  will correspond to the dominant  $\delta^{13}\text{C}$  value at  $e_i$ .  $\delta^{13}\text{C}$  reflect dominant plant communities on a marsh and the observed modern boundaries between communities can correspond to a tidal datum (TD). As a result  $\delta^{13}\text{C}$  measured in bulk sediment can be related to tidal elevation as follows;

$$\mu_i = \begin{cases} \mu_1, & \text{if } e_i \leq TD \\ \mu_2, & \text{if } e_i \geq TD \\ \mu_3, & \text{otherwise} \end{cases}$$

where,  $\mu_i$ 's are given informative uniform priors with upper and lower limits corresponding to the maximum and minimum  $\delta^{13}\text{C}$  values represented in a given elevational range. The prior information required here is location specific. For example, for the New Jersey case study sites the TD is MHHW. At these sites  $\delta^{13}\text{C}$  can range anywhere from -28‰ to -22‰ at elevations above MHHW and anywhere from -18.9‰ to -12‰ at elevations below MHHW. Therefore, the prior for  $\mu_1 \sim U(-28, -22)$  and  $\mu_2 \sim U(-18.9, -12)$ . In cases where  $\delta^{13}\text{C}$  values fall in the range -21.9‰ to -19‰ they can't provide information about elevation in relation to MHHW, thus  $\mu_3 \sim U(-21.9, -19)$ .

### 3.3 The chronology module: Bchron

The chronology module is concerned with estimating the ages  $t$  of the foraminifera in the sediment core. These ages will necessarily be uncertain, since the radiocarbon dates  $r$  (and non-radiocarbon dates) are observed with uncertainty. An interpolation step is then required to obtain estimated ages

at all depths, which adds further uncertainty. A useful constraint is that age must increase with depth (the law of superposition; Nicolaus, 1669) so a monotonic stochastic process is used. Bchron (Haslett and Parnell, 2008) assumes that the integrated sedimentation rate (i.e., the accumulation of sediment over a fixed period of time) arises as the realisation of a Compound Poisson–Gamma (CPG) process. Bchron calibrates the radiocarbon (and non-radiocarbon) dates, estimates the parameters of the CPG (here  $\psi$ ) and identifies outliers. Other age-depth models are available (see Parnell et al. 2011 for a review), but Bchron was designed specifically for use in palaeoenvironmental reconstructions.

Once Bchron has been run, we obtain a joint posterior distribution of ages for every layer in the sediment core, which we denote as  $p(t|r, d, \psi)$ . Each individual chronology sample from Bchron satisfies the law of superposition. However, we approximate the age of each layer in the posterior, i.e.,  $p(t_i|r, d, \psi)$  as a normal distribution, so that  $t_i|r, d, \psi \sim N(\mu_{t_i}, \sigma_{t_i}^2)$ . This may seem like a severe relaxation, since the ages of layers may now overlap, but we find this has minimal effect on the resulting RSL reconstructions since the ages are further updated during the process module. Further simulations justifying this assumption have been carried out using chronological models in late Holocene sea level reconstructions from saltmarsh sediments (Parnell and Gehrels, 2015).

### 3.4 The process module: errors-in-variables integrated Gaussian process (EIV-IGP)

Our final step is to transform PME,  $e_i$  into RSL,  $s_i$  via the relationship  $s_i = g_\omega(e_i, d_i)$ . We then have a set of bivariate probability distributions for each layer consisting of pairs  $(t_i, s_i)$  which represent the raw layer-by-layer estimates of RSL and age. To use these in the EIV-IGP framework of Cahill et al. (2015), we approximate each bivariate probability distribution as bivariate Gaussian. The model makes use of two well known statistical approaches. Firstly, the EIV approach (Dey et al., 2000) accounts for measurement error in the explanatory variable, here time. The EIV approach is necessary when dealing with proxy reconstructions that include temporal uncertainty from dating the sediment core. Secondly, the Gaussian process approach (Rasmussen, 2006) is useful for non-linear regression problems and is a practical approach to modelling time series data. A Gaussian process is fully specified by a mean function (set to zero) and a covariance function that relates the observations to one another.

We use an integrated Gaussian process approach (Holsclaw et al., 2013; Cahill et al., 2015). A Gaussian process prior is placed on the rates of sea-level change and the mean of the distribution assumed for the observed data is derived from the integral of the rate process. This integrated approach is useful when there is interest in the rate process as the analysis allows for estimates of instantaneous rates of RSL change. Furthermore, the current RSL estimate is derived as the integral of all the previous RSL rates that have occurred, matching the physical behaviour of sea-level evolution over time. By embedding the integrated Gaussian process (IGP) model in an errors-in-variables

(EIV) framework (which takes account of time uncertainty), we can estimate rates with quantified uncertainty. We use the same priors for the parameters  $\phi$  as described in Cahill et al. (2015).

#### 4 Case study: New Jersey RSL

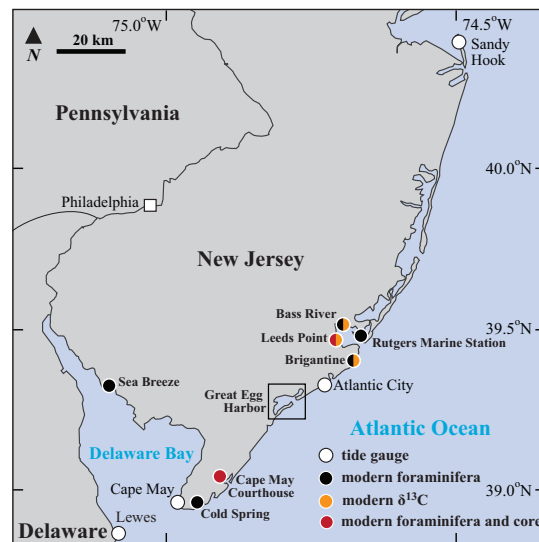
On the Atlantic coast of southern New Jersey (Figure 1), salt marshes form in quiet-water, depositional settings and display a zonation of plants into distinct vertical zones corresponding to ecologically important tide levels. Elevations below mean tide level (MTL) are not vegetated and the inorganic sediment is comprised of silt and fine sand, often with shell material. Low salt-marsh environments between MTL and mean high water (MHW) are vegetated by *Spartina alterniflora* (tall form), which is a  $C_4$  plant. Sediment in this zone is organic grey silt and clay. High salt-marsh environments exist between MHW and highest astronomical tide (HAT). This zone is typically a wide, flat meadow vegetated by *Spartina patens* and *Distichlis spicata* ( $C_4$  species). The sediment deposited in this zone is brown peat with abundant plant remains. The transition between high salt marsh and the freshwater upland is vegetated by  $C_3$  plants such as *Phragmites australis*, *Iva frutescens*, *Schoenoplectus americanus*, and *Typha augustifolia*. This community exists at tidal elevations above mean higher high water (MHHW), including freshwater environments above the reach of tidal influence, and occurs with black, amorphous, organic sediment.

##### 4.1 Modern training set

The transfer function approach is under-pinned by surveys that quantify the modern, observable relationship between foraminifera and tidal elevation. Modern (surface) samples should be collected from depositional environments analogous to those represented in the fossil material under investigation (Horton and Edwards, 2005). At individual sites sampling should span the full range of tidal elevations from shallow sub-tidal settings through to supra-tidal zones. Kemp et al. (2013a) compiled a modern training set from twelve salt marshes in southern New Jersey (Figure 1). At stations along each transect, surface (0-1 cm) sediment samples were collected to count foraminifera. The tidal elevation of each sample was measured in the field. Since the great diurnal tidal range (MLLW to MHHW) varies among sites in the study region tidal elevation is expressed as a standardized water level index (SWLI; e.g. Horton et al., 1999), where a value of 0 corresponds to MLLW and 100 is MHHW. This modern training set was used here without modification from Kemp et al., 2013a.

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 spatial scope of modern training sets has frequently been discussed and it is widely accepted that regional-scale training sets capture natural variability in the distribution of foraminifera and provide

a large suite of samples from which to draw modern analogs (e.g., Horton and Edwards, 2005). The principal advantage of local-scale training sets is that they produce more precise reconstructions, but at the expense of offering only a narrow range of modern analogs (e.g., Horton and Edwards, 2005). Kemp et al. (2013b) described and discussed the effect of using regional- and local-scale modern training sets to reconstruct RSL in New Jersey and concluded that it was necessary to use the regional data to provide sufficient modern analogs for interpreting assemblages of foraminifera preserved in sediment cores. This led Kemp et al. (2013a) to use the regional-scale training set when they produced Common Era RSL reconstructions from two locations in southern New Jersey and their analysis showed that the closest modern analogs for core samples were drawn from 11 of the 12 sites in the modern training set. Since it is not our aim to investigate the relative utility of local- and regional-scale modern training sets we use the same data as Kemp et al. (2013a) to ensure that the new B-TF can be compared directly with existing results. However, an important difference is that the B-TF uses raw counts of foraminifera as the input, while the WA-TF model uses relative abundances expressed as percentages.



**Figure 1.** Location of study sites in southern New Jersey, USA. The distribution of modern foraminifera was described at 12 different salt marshes including five in Great Egg Harbor (not located with symbols in the figure). Bulk surface sediment  $\delta^{13}\text{C}$  values were measured at three sites and cores for sea-level reconstruction (filled circles) were collected at Leeds Point and at Cape May Courthouse.

#### 4.1.1 Modern counts of foraminifera

The modern dataset comprised 172 paired observations of 18 foraminiferal species (including many zeros) and tidal elevation. The foraminifera count sizes ranged from 8 to 307 dead individuals.

The highest occurrence of foraminifera in the modern dataset is at 141.5 SWLI. Higher samples were devoid of foraminifera and interpreted as being located above marine influence. This modern training set demonstrates that foraminifera (like plants) form distinct assemblages that correspond to elevation in the tidal frame (e.g., Scott and Medioli, 1978), but with a secondary influence of salinity (e.g., de Rijk, 1995). **Kemp et al., 2013b provide a qualitative discussion on the influences of salinity on the foraminifera species specific to our case study.** Throughout southern New Jersey, low-marsh environments are occupied by *Miliammina fusca* and *Ammobaculites* spp. (Figure 2). High salt-marsh environments are characterized by a number of foraminiferal assemblages including groups dominated by *Trochammina inflata*, *Arenoparella mexicana*, and *Tiphotrecha comprimata*. High salt marshes at sites with strong fluvial influence and correspondingly low (brackish) salinity are occupied by *Ammoastuta inepta* (Figure 2). At some sites elevations above MHHW are characterized by a group of foraminifera in which *Haplophragmoides manilaensis* is the dominant species. The pattern (uniform low marsh and diverse high marsh) and composition of these assemblages is similar to those identified elsewhere on the U.S. Atlantic coast (e.g., Murray, 1991; Gehrels, 1994; Kemp et al., 2009a; Wright et al., 2011; Edwards et al., 2004).

#### 4.1.2 Modern bulk-sediment $\delta^{13}\text{C}$ measurements

In the mid-Atlantic and northeastern U.S.A the low salt-marsh and high salt-marsh zones are dominated by  $\text{C}_4$  species such as *Spartina alterniflora*, *Spartina patens*, and *Distichlis spicata*, while the transitional marsh and surrounding upland zones are dominated by  $\text{C}_3$  species. In New Jersey the boundary between  $\text{C}_3$  and  $\text{C}_4$  plant communities corresponds to MHHW and  $\delta^{13}\text{C}$  measured in bulk sediment can be used to reconstruct RSL by determining if a sample formed above or below the MHHW tidal datum. **Based on a dataset of surface (0-1 cm) bulk sediment  $\delta^{13}\text{C}$  from three sites in southern New Jersey (Figure 1) and presence or absence of foraminifera, Kemp et al. (2013a) recognized three types of sediment that were likely to be encountered in cores of organic coastal sediment and to ensure comparability we use their interpretation here without modification.**

1. Samples with  $\delta^{13}\text{C}$  values more depleted than  $-22.0\text{‰}$  and with foraminifera present formed at tidal elevations from 100-150 SWLI. The lower limit of this range corresponds to MHHW and the upper limit is conservatively set to extend slightly beyond the observed highest occurrence of foraminifera (141.5 SWLI) in the modern dataset.
2. Samples with  $\delta^{13}\text{C}$  values less depleted than  $-18.9\text{‰}$  formed at tidal elevations from 0-100 SWLI since  $\text{C}_4$  plants are dominant below MHHW. This interpretation is the same if foraminifera are present or absent.
3. Samples with intermediate  $\delta^{13}\text{C}$  values between  $-18.9\text{‰}$  and  $-22.0\text{‰}$  provide no additional information and if foraminifera are present these samples are interpreted as having formed at 0-150 SWLI (MLLW to slightly above the highest observed occurrence of foraminifera).



**Figure 2.** Dataset of modern foraminifera described from a total of 172 surface sediment samples from 12 different sites. The data are presented as relative abundances. Only the abundance of the eight most common species are shown. Modified from Kemp et al. (2013a). Pink dashed lines and shading indicate the species optima and tolerances estimated by the WA-TF. Tidal datums are indicated by blue dashed lines.

## 4.2 Proxy data

Cores of salt-marsh sediment were recovered from two sites in southern New Jersey (Cape May Courthouse and Leeds Point; Figure 1) and sliced into 1-cm thick samples. Three types of data were generated for each sediment core that we use as originally presented by Kemp et al. (2013a).

### 4.2.1 Fossil counts of foraminifera

In the Cape May Courthouse core *Jadammina macrescens* and *Trochammina inflata* were the dominant species from 1.72 m to 1.29 m (Figure 3A, upper panel). Foraminifera were absent at 1.25 m to 1.12 m. Between 1.10 m and 0.33 m *Jadammina macrescens* was the dominant species, while samples in the interval from 0.31 m to 0.05 m included *Trochammina inflata*, *Tiphrotrocha comprimata*, and *Jadammina macrescens*. Two samples near the top of the core with total counts of 194 (at 0.03 m) and 174 (at 0.05 m) included 34 and 37 *Miliammina fusca* respectively. Counts of foraminifera throughout the core ranged from 16 to 194 per sample with an average of 98. In the lower part of

the Leeds Point core (2.85 m to 3.95 m), *Jadammina macrescens* was the most common species and occurred with *Tiphotrecha comprimata* and *Trochammina inflata* (Figure 3A, lower panel). Within this section, at  $\sim 3.00$  m there was an unusual occurrence of 24 *Miliammina petila* (out of a total of 101) and from 2.82 to 2.95 m 30% of the counted individuals were *Miliammina fusca*. From 2.82 m to 1.85 m *Trochammina inflata* was the dominant species. The uppermost section of the Leeds Point core (1.73 m to 1.20 m) was comprised of a near mono-specific assemblage of *Jadammina macrescens*. Counts of foraminifera in this core ranged from 4 to 127 per sample with an average of 70.

The assemblages of foraminifera preserved in each core were compared to those in the modern training set. If the dissimilarity between a core sample and its closest modern analog (measured using the Bray-Curtis metric) exceeded the 20<sup>th</sup> percentile of dissimilarity among all possible pairings of modern samples then the core sample was deemed to lack a suitable modern analogue and was excluded from further analysis by Kemp et al. (2013a). We did not reconstruct PME for these samples because they may lack ecological plausibility (e.g., Jackson and Williams, 2004). On Figure 3, these samples therefore lack a PME reconstruction (panels B, C, and E).

#### 4.2.2 Fossil bulk-sediment $\delta^{13}\text{C}$ measurements

In the Cape May Courthouse core all samples were less depleted than  $-18.9\text{‰}$  and were interpreted as having formed below MHHW in a salt marsh dominated by  $\text{C}_4$  plants (Figure 3D, upper panel). In the lowermost part of the Leeds Point core (below 3.35 m) the presence of foraminifera and bulk sediment  $\delta^{13}\text{C}$  values more depleted than  $-22.0\text{‰}$  indicate that the sediment accumulated above MHHW in an environment dominated by  $\text{C}_3$  plants, but below the highest occurrence of foraminifera (Figure 3D, lower panel). Between 3.31 m and 2.86 m bulk-sediment  $\delta^{13}\text{C}$  values were variable and interpreted to record the transition from highest salt marsh to high salt-marsh environments. Above 2.86 m, all samples were less depleted than  $-18.9\text{‰}$  and were interpreted as having formed below MHHW in a salt marsh dominated by  $\text{C}_4$  plants.

#### 4.2.3 Age-depth profile estimated by Bchron

Age-depth models for the Cape May Courthouse and Leeds Point cores were previously developed by Kemp et al., 2013a using Bchron and are used here in the chronology module without modification (Figure 3F). The Cape May Courthouse core was dated by recognition of pollution markers, the appearance of *Ambrosia* pollen as a land clearance marker, and radiocarbon dating of in situ and identifiable plant macrofossils. These data were combined into a single age-depth model that estimated the age of every 1-cm thick sediment sample in the core with an average uncertainty of  $\sim 30$  years for the period since  $\sim 700$  CE (Figure 3F, upper panel). Anthropogenic modification of the Leeds Point site limited dating and RSL reconstruction to the interval from  $\sim 500$  BC to  $\sim 1750$

CE. The core was dated using only radiocarbon measurements performed on *in situ* and identifiable plant macrofossils. These data were combined into a single age-depth model that estimated the age of every 1-cm thick sediment sample in the core with an average uncertainty of ~50 years (Figure 3F, lower panel).

### 4.3 Instrumental data

A tide gauge is an instrument that automatically measures the sea surface height with reference to a control point on the land many times during a day. These measurements are averaged to obtain annual values to minimize the effects of weather and tidal variability. In New Jersey (Figure 1), tide-gauge data are available since 1911 CE when the Atlantic City tide gauge was installed. The Sandy Hook, Cape May, and Lewes (Delaware) tide gauges began measurements in 1932 CE, 1966 CE, and 1919 CE respectively. A single regional record was compiled by averaging annual data from these four local tide gauges. RSL is zero between 2000-2010 CE to be roughly equivalent to the age of a surface sample in the core (by definition RSL=0 m when the core was collected). The resulting record minimizes spatial variability and the influence of decadal-scale RSL variability (Douglas, 1991).

## 5 Results

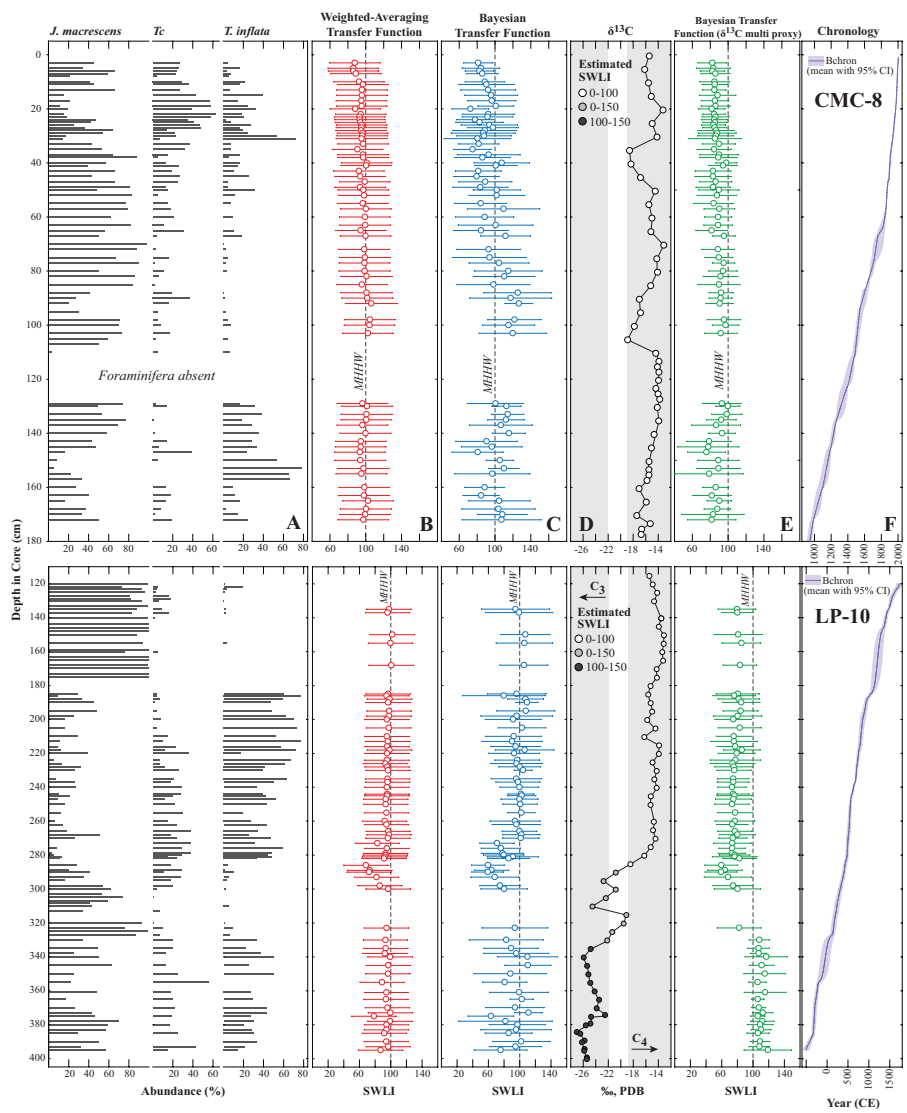
The Cape May Courthouse and Leeds Point RSL reconstructions were developed separately to one another and were then merged into a single, regional-scale dataset for analysis using the EIV-IGP model to capture the continuous and dynamic evolution of RSL change while taking account of the quantified uncertainty in both sea level and age reconstructions. It should be noted however that other sources of uncertainty that are inherent in RSL reconstructions (e.g., the influence of sediment compaction or tidal range changes) were not considered or accounted for.

### 5.1 The Bayesian Transfer Function

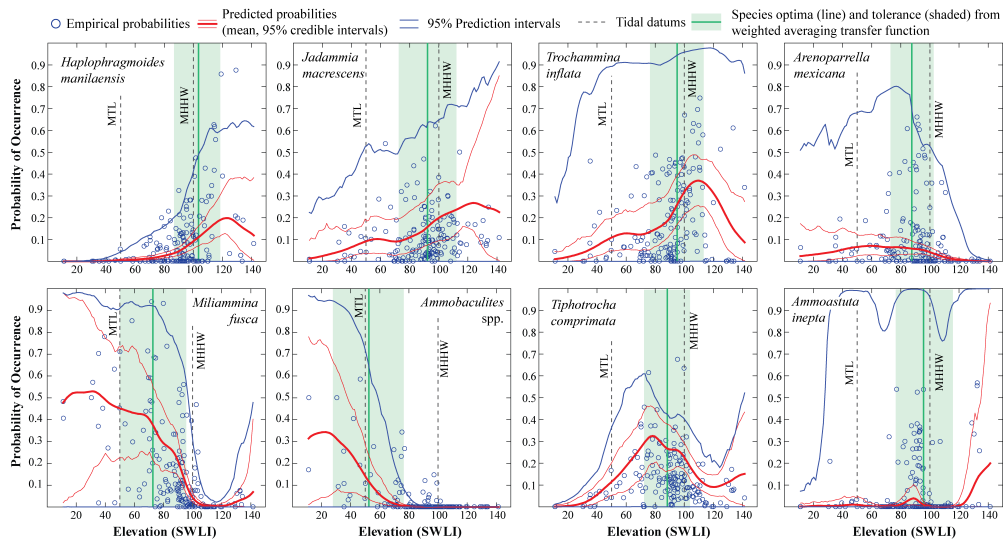
#### 5.1.1 Species-response curves

The B-TF estimated a response curve (mean with a 95% credible interval) for each species of foraminifera (expressed as raw counts) to tidal elevation (expressed as SWLI) from the modern training set of 172 samples (Figure 4). The response curves are estimated from a multinomial distribution (in which species compositions are considered as a whole) parameterized by a probability vector  $p$ , which is the probability of a species being present at a given tidal elevation. The multinomial model (described in detail in Section 3.1) utilizes the combined species information from these observed responses to estimate PME. The species prediction intervals (dashed red lines in Figure 4) will aid in providing uncertainty for the PME estimates.





**Figure 3.** The abundance of the the three most common species (*Tc*= *Tiphotrocha comprimata*) of foraminifera found in cores from Cape May Courthouse (CMC-8) and Leeds Point (LP-10) are represented by the horizontal gray bars (A). Paleo marsh elevation, in standardised water level index (SWLI) units, was reconstructed using the weighted average transfer function (B), the Bayesian transfer function (C) and the multi-proxy Bayesian transfer function (E). Mid points of the reconstruction are shown as white circles with the bars representing  $\pm 2\sigma$  uncertainty. Vertical dashed lines show the elevation of the mean higher high water (MHHW) tidal datum. Stable carbon isotope concentrations ( $\delta^{13}\text{C}$ ) for bulk sediment are parts per thousand (‰) relative to the Vienna Pee Dee Belemnite (VPDB) standard (D). Bchron provided the chronology for both cores (F). Note that some samples with counts of foraminifera lack a corresponding PME reconstruction because they lack a suitable modern analog using the criteria applied by Kemp et al. (2013).



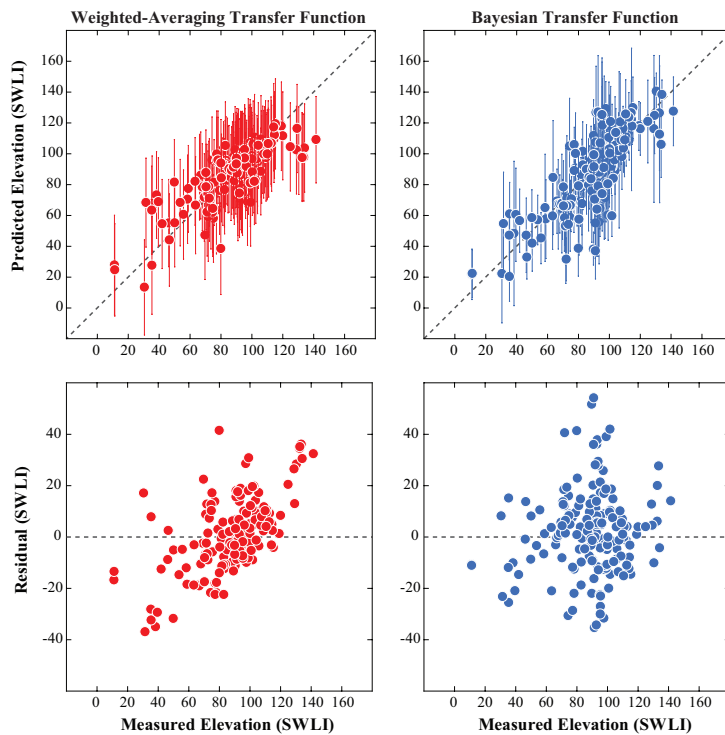
**Figure 4.** The response of foraminifera species to elevation estimated from the modern training set of raw counts using the Bayesian transfer function. The blue circles represent the probabilities of species occurrence as determined from the count data (empirical probabilities). The response probabilities of occurrence estimated by the Bayesian transfer function model are shown in with a mean (heavy red line), a credible interval for the mean (light red line), and a prediction interval (blue line). The green vertical lines and shading represent the species optima and tolerances determined from the weighted average transfer function. Tidal datums are indicated by vertical dashed black lines.

Broadly, we identify two forms of species-response curve in southern New Jersey. First, a skewed, unimodal form describes the distribution of *Haplophragmoides manilaensis* with a maximum probability of occurrence of  $\sim 0.2$  at 123 SWLI compared to a WA-TF species optimum of 99 SWLI. *Jadammina macrescens* and *Trochammina inflata* also have a skewed, unimodal form with the highest probability of occurrences found in high salt-marsh environments at 124 SWLI ( $p \sim 0.3$ ) and 110 SWLI ( $p \sim 0.4$ ) respectively. The optima for these species were situated at lower elevations by the WA-TF (92 SWLI for both species). Both *Ammobaculites spp.* (highest probability of occurrence of  $\sim 0.3$  at 22 SWLI) and *Miliammina fusca* (highest probability of occurrence  $\sim 0.5$  at 31 SWLI) have skewed unimodal distributions with maximum probability of occurrence in low salt-marsh environments. The WA-TF estimated species optima of 58 and 78 SWLI respectively for these two species. *Ammoastuta inepta* also has a skewed, unimodal form (maximum probability of occurrence of  $\sim 0.2$  at 140 SWLI). This species generally has a low probability of being present because its distribution in southern New Jersey is restricted to sites with brackish salinity such as those located up river in Great Egg Harbor (Figure 1). Relatively few samples from these environments are included in the modern training set and therefore in the dataset as a whole it is a rare, but ecologically-important, species. Second, a unimodal Gaussian-like form describes the distributions of *Arenoparrella mex-*

*icana* (maximum probability of occurrence of  $\sim 0.07$  between 60 and 70 SWLI) and *Tiphotrocha comprimata* (maximum probability of occurrence of  $\sim 0.3$  at 78 SWLI). For these species the WA-TF indicated optima of 86 and 89 SWLI respectively.

### 5.1.2 Cross validation of the modern data

Performance of the new B-TF and existing WA-TF was judged using 10-fold cross validation (Figure 5). The uncertainty bounds ( $\pm 2\sigma$ ; shown as vertical lines on Figure 5) for predicted elevations contained the true elevation for 90% of samples using the B-TF, compared to 92% for the WA-TF. The average of the  $2\sigma$  prediction uncertainties is slightly larger for the WA-TF (28 SWLI) than for the B-TF (21 SWLI). 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).



**Figure 5.** Cross validation of the modern training set for the weighted average transfer function (red) and the Bayesian transfer function (blue). Upper panels are elevations in standardised water level index (SWLI) units, with lines representing  $\pm 2\sigma$  uncertainty for prediction. Lower panels show the (observed-predicted) residuals.

The pattern of residuals in the WA-TF displayed a structure in which the elevation of low salt-marsh samples is over predicted (negative residuals) and the elevation of high salt-marsh samples is

under predicted (positive residuals; Figure 5). For example, the WA-TF showed an average residual of -16.6 between  $\sim 10$  and  $\sim 70$  SWLI and an average residual of 22.5 between  $\sim 120$  and  $\sim 140$  SWLI. This structure is absent in the B-TF, suggesting that it is better suited to reconstructing values close to the extremes of the sampled elevational gradient than the WA-TF. **In contrast, the WA-TF performs better at elevations between 80 to 100 SWLI because the distribution of samples in the modern training set is focused at  $\sim 90$  SWLI, and, by the very nature of the WA-TF estimates of elevation will be drawn towards this average value.**

### 5.1.3 PME reconstructions

We reconstructed PME in the Cape May Courthouse and Leeds Point cores using the WA-TF and B-TF models. At Cape May Courthouse, the B-TF estimated an average PME close to MHHW (SWLI=100) of 96.2 SWLI, with a standard deviation 14.1 SWLI (Figure 3C, upper panel). The WA-TF also estimated an average PME close to MHHW of 96.7 SWLI, with a standard deviation of 4.1 SWLI (Figure 3B, upper panel). The  $2\sigma$  uncertainties for the B-TF reconstructions ranged from 15.1 to 45.7 and are more variable than those from the WA-TF (28.1 to 29.0 SWLI). At Leeds Point, the B-TF estimated an average PME close to MHHW of 92.8 SWLI, with a standard deviation 12.8 SWLI (Figure 3C, lower panel). The WA-TF estimated PME close to MHHW for all samples except for the 3.00-2.80 m interval where *Miliammina fusca* was present and PME reconstructions were correspondingly lower (Figure 3B, lower panel). The  $2\sigma$  uncertainties for the B-TF reconstructions ranged from 11.5 to 59.8 SWLI and were more variable than those from the WA-TF (28.0 to 28.5 SWLI). Comparison of the two models shows that the B-TF typically reconstructed PME with greater variability among samples than the WA-TF model. Within their uncertainties the PME reconstructions for the B-TF and WA-TF overlap for all samples in both sediment cores.

### 5.1.4 Multi-proxy reconstruction of PME

On salt-marshes in the mid-Atlantic and northeast coast of the U.S.A. measurements of  $\delta^{13}\text{C}$  in bulk-organic sediment are useful sea-level proxies because they readily distinguish between sediment that accumulated above MHHW in an environment dominated by  $\text{C}_3$  plants and sediment that accumulated below MHHW in an environment dominated by  $\text{C}_4$  plants. This additional paleoenvironmental information was combined with the B-TF to generate a multiproxy reconstruction of PME. The inclusion of the  $\delta^{13}\text{C}$  proxy did not treat MHHW as a hard bound for PME, rather, if a sample is dominated by  $\text{C}_3$  plants then the probability of PME being above MHHW increases. Likewise, if a sample is dominated by  $\text{C}_4$  plants then the probability of PME being below MHHW increases.

For Cape May Courthouse using the downcore  $\delta^{13}\text{C}$  values as a secondary proxy in the B-TF estimated an average PME of 87.5 SWLI (a reduction of 9 SWLI compared to the original B-TF). The  $2\sigma$  PME uncertainties were reduced by 32% on average and by up to  $\sim 60\%$  for some samples

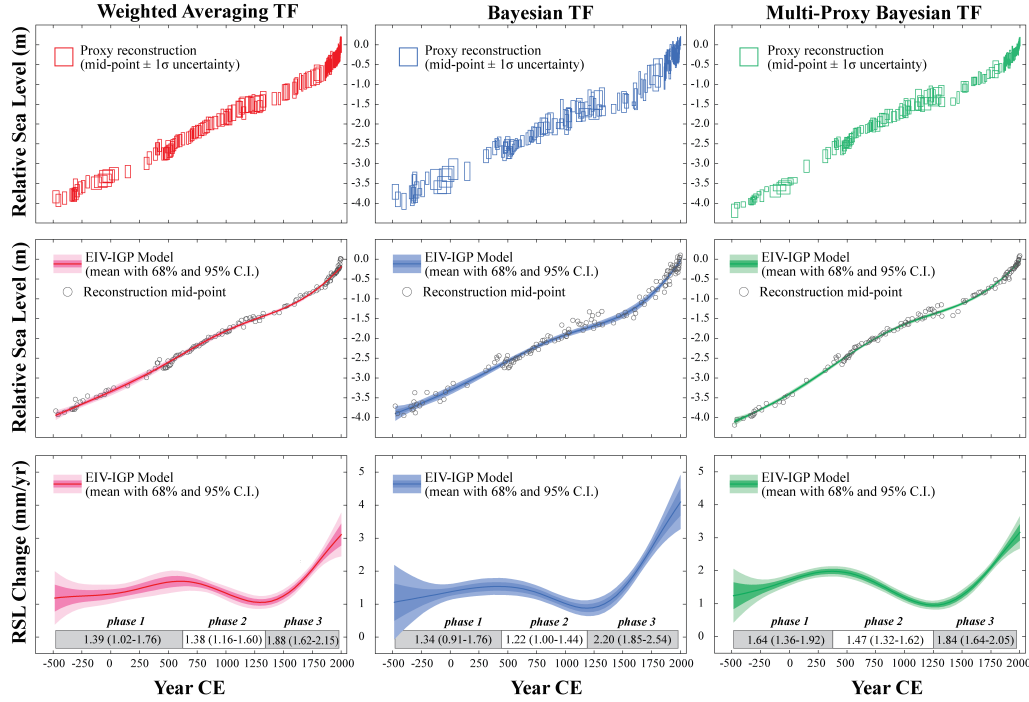
(Figure 3E, upper panel). For the Leeds Point core, the inclusion of the secondary  $\delta^{13}\text{C}$  proxy resulted in an average PME reconstruction of 86.1 SWLI (a reduction of 7 SWLI compared to the original B-TF). The  $2\sigma$  PME uncertainties decreased by  $\sim 25\%$  on average (Figure 3E, lower panel). However, for samples where  $\delta^{13}\text{C}$  values and the presence of foraminifera indicate deposition in the transitional marsh (above MHHW, but below the highest occurrence of foraminifera) the uncertainty was reduced by an average of  $\sim 50\%$  and by up to  $\sim 70\%$  for some samples because the constraint on PME changed from 0-150 SWLI to 100-150 SWLI. These results demonstrate that incorporating a second line of proxy evidence in the B-TF framework is an efficient and formalized way to reduce uncertainty in RSL reconstructions in some (but not all) sedimentary environments.

## 5.2 Comparison among reconstructions

We applied the EIV-IGP model to the RSL reconstructions produced from the WA-TF, B-TF and multi-proxy B-TF to describe RSL trends along the coast of southern New Jersey since  $\sim 500$  BCE (Figure 6). The WA-TF and B-TF models show  $\sim 3.9$  m of RSL rise compared to  $\sim 4.1$  m of RSL rise for the multi-proxy B-TF model. The multi-proxy B-TF reconstructed lower RSL at the beginning of the record compared to the B-TF and the WA-TF, because of the additional constraint placed on the lowermost section of the Leeds Point core by  $\delta^{13}\text{C}$  values that indicate deposition at 100-150 SWLI. All of the reconstructions show three phases of RSL behavior (Figure 6). The period from  $\sim 500$  BCE to  $\sim 500$  CE is characterized by a continuous increase in the rate of RSL rise. The second stage ( $\sim 500$  CE to  $\sim 1400$  CE) shows a decline in rates of RSL rise. After 1400 CE there is a continuous acceleration in the rate of RSL rise until reaching historic rates, which are unprecedented for at least 2500 years. **All reconstructions show an expanding uncertainty for the most recent rates. This is an edge effect that occurs as a result of using a Gaussian process model on the rates. These rates are estimated based upon their relationship (through an integration step) to the observations. As we approach the edge of the data, there are fewer constraints from the observations. This means that there are multiple rate trajectories that are consistent with recent data and so the uncertainty in the rate process increases.**

There are some differences among the three reconstructions. For example, the B-TF shows the highest modern rate of rise at 4.1 mm/yr (95% C.I. 3.27-4.92 mm/yr) in 2000 CE compared to 3.16 mm/yr (95% C.I. 2.68-3.65 mm/yr) and 3.11 mm/yr (95% C.I. 2.45-3.77 mm/yr) for the multi-proxy B-TF and the WA-TF respectively. The B-TF consistently estimated RSL lower than the multi-proxy B-TF and the WA-TF between  $\sim 1400$  to  $\sim 1700$ , resulting in the observed difference in the rates into the 21<sup>st</sup> century. When compared to the observed tide-gauge data for the last  $\sim 100$  years from New Jersey (Figure 7), the quality of the estimated RSL mid-point reconstructions can be assessed using mean squared error (MSE). For the multi-proxy B-TF the MSE was 0.003 m<sup>2</sup>, compared to 0.014

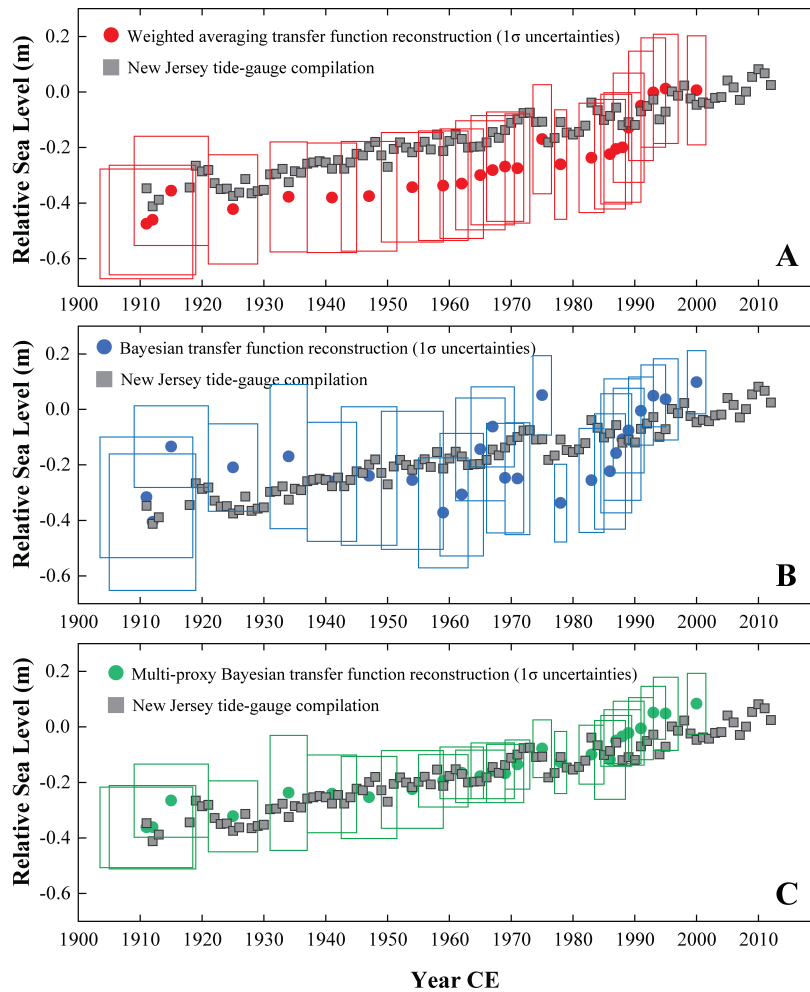
$m^2$  for the B-TF and the WA-TF, indicating that the multi-proxy B-TF mid-points provide better estimates for RSL in comparison to the B-TF and the WA-TF.



**Figure 6.** The EIV-IGP model results for reconstructions produced using the B-TF, the WA-TF and the multi-proxy Bayesian transfer function. The upper panel shows individual data points (represented by rectangular boxes that illustrate the 95% confidence region) and include age and relative sea-level uncertainties. The middle panels show the posterior fit of the errors-in-variables integrated Gaussian process model to the relative sea-level reconstructions. Solid line represents the mean fit with the 68% and 95% confidence intervals (C.I.) denoted by shading. The lower panels are the rates of relative sea-level (RSL) change. Shading denotes 68% and 95% confidence intervals (C.I.) for the posterior mean of the rate process. The average rate for each phase of the reconstruction is given (in mm/yr) with a 95% confidence interval.

## 6 Discussion

The B-TF provides an alternative to the (non-Bayesian) regression-based transfer functions commonly used for reconstructing RSL (e.g., Horton et al., 1999; Gehrels, 2000; Barlow et al., 2014) and in conjunction with the previously developed chronology and process modules enables RSL to be reconstructed in an entirely Bayesian framework. A key difference between the B-TF and existing transfer functions (e.g. the WA-TF) is the modeled relationship between species of foraminifera and tidal elevation. The number and type of species-response curves estimated by the B-TF model stands



**Figure 7.** Comparison of the weighted average transfer function (A), the Bayesian transfer function (B) and the multi-proxy Bayesian transfer function (C) relative sea-level reconstruction with tide gauge data observed in the New Jersey region. The boxes represent  $\pm 1\sigma$  uncertainty region for the reconstruction.

in contrast to the WA-TF, which assumes a unimodal Gaussian form for all species. The optima and tolerance estimated for each species by the WA-TF show overlap with the B-TF species-response curves, particularly those that have a Gaussian form such as *Tiphotrecha comprimata*. However, this form is only appropriate for two of the eight dominant species in the southern New Jersey training set. The flexible, species-specific response curves provided by the B-TF are more appropriate given that models based on the assumption of a single response do not adequately explain the ecological behavior of the dominant species in New Jersey, other assemblages of salt-marsh foraminifera (e.g., Edwards and Horton, 2006), or species from other biological groups used in RSL reconstructions such as diatoms (e.g., Zong and Horton, 1999; Gehrels, 2000). **However, our results suggest that the**

flexibility in the B-TF can, in some instances, make it more susceptible to unusual observations. For example, the *Miliammina fusca* response curve in the B-TF shows a slight increase in probability of occurrence above  $\sim 120$  SWLI (Figure 4), because of three unusual samples located above 130 SWLI in which this species occurs (Figure 2). This distribution may represent the advantage of a regional-scale training set in capturing natural variability caused by ecological complexity and/or the influence of secondary environmental variables. Equally these samples could be an anomalous occurrence (e.g., tests washed in by a storm) that ought to be screened prior to analysis. Since Kemp et al. (2013a) retained these samples, we also elected to use them in developing the new B-TF.

The implication of the flexibility of the B-TF is illustrated in the cross validation results. The WA-TF displayed edge effects (a tendency to bias PME predictions towards the mean of the training data), which is a common artifact of using weighted average based methods (e.g., ter Braak and Juggins, 1993; Birks, 1995). Our B-TF does not suffer from this prediction bias and outperformed the WA-TF in the upper and lower extremes of tidal elevation. The consequences of such an improvement are significant where true PMEs lie close to the ends of the sampled environmental gradient. For example, on subduction zone coastlines such as the Pacific Northwest coast of North America, cyclical tectonic activity contributes to reconstructed RSL trends (e.g., Nelson et al., 1996). During a slow (100s to 1000s of years) inter-seismic phase, accumulation of strain results in uplift of the coast (RSL fall). Conversely, the strain is released during an instantaneous co-seismic phase in which the coastline subsides (RSL rise). These processes cause significant and very rapid shifts in depositional environment that can span the full elevational range of coastal environments from sub-tidal settings to supra-tidal, freshwater uplands. In these settings, samples that are analogous to those at the ends of the sampled environmental gradient are frequently encountered in core material. In contrast, the sediment sequences targeted for reconstructing Common Era RSL on passive margins (e.g., New Jersey) are commonly comprised of unbroken sequences of high salt-marsh peat that are less susceptible to edge effects.

Further motivation for the development of the B-TF lies in the **quantification of PME uncertainty**. Non-Bayesian transfer function methods (e.g. the WA-TF model) assume that model parameters are fixed and known. Therefore, they do not incorporate uncertainty into the estimation of the PME reconstruction itself, rather, the uncertainty is produced separately either before or after PME was estimated. This uncertainty is the root mean square error from two sources (S1 and S2; Birks et al., 1990; Juggins and Birks, 2012). The S1 contribution is sample-specific and is the standard deviation of bootstrapped PME reconstructions. The S2 contribution is the difference between observed and predicted tidal elevations established by cross validation of the modern training set (Figure 5). The PME uncertainties reconstructed by the WA-TF model show very little variability among samples ( $2\sigma$  ranges from 28.0 to 29.0 SWLI). This pattern arises because the contribution from the sample-



specific (S1) uncertainty is very small compared to the model uncertainty (S2) which is common to all samples. As a result, the PME reconstructions for all core samples have very similar uncertainties despite biological variability in species composition.

Alternatively, Bayesian methods explicitly model the uncertainty associated with individual reconstructions. Uncertainty for PME (and other unknown parameters) is included in the probability model through prior distributions. Assuming distributions for unknown parameters (in contrast to non-Bayesian approaches that use point estimates) allows the parameter uncertainty from the calibration step to be formally propagated into the reconstruction step. Therefore, estimates of PME produced by the B-TF take fuller account of the uncertainties related to the model and its parameters than non-Bayesian approaches. Consequently, the uncertainties estimated by the B-TF (excluding a secondary proxy) show more pronounced variability among core samples ( $2\sigma$  uncertainties range between 15.1 to 45.7 SWLI) than the WA-TF model. This variability arises from the observed response distribution of each species to tidal elevation (estimated from the modern data; Figure 4). For each individual species there is variability in both the uncertainty of the mean response curves and in the prediction intervals (i.e. uncertainty is greater in some parts of the elevational gradient than at others).

The variability of reconstructed PME from the B-TF may reflect a more ecologically plausible reconstruction than the WA-TF model. For example, the key, high salt-marsh species *Jadammina macrescens* and *Trochammina inflata* vary in relative abundance from 0% (absent) to 100% and approximately 80%, respectively (Figure 3). Despite this variability there was a pronounced lack of variability reconstructed PME using the WA-TF model (average of  $\sim 95$  SWLI with a standard deviation of 5.5). In contrast, PME reconstructions from the B-TF are also estimated at an average of  $\sim 95$  SWLI, but with a larger standard deviation of 13.1. Intuitively the higher degree of PME variability reconstructed by the B-TF model is a better reflection of the species composition changes observed in the two sediment cores than the near stationary PME values reconstructed by the WA-TF.

The majority of quantitative RSL reconstructions employ a single proxy (e.g., Kemp et al., 2011). A number of other proxies are available to support RSL reconstructions primarily produced from salt-marsh foraminifera. Additional biological proxies could include different groups of organisms with a relationship to tidal elevation such as diatoms (e.g., Zong and Horton, 1999; Shennan et al., 1994) or testate amoebae (e.g., Charman et al., 2010; Roe et al., 2002). These organisms can be incorporated as presence/absence data or as species counts from a modern training set of paired observations of species abundance and tidal elevation. A number of lithological proxies (e.g., Nelson, 2015) are also available which can be qualitative (such as field and lab-based descriptions of sediment as high marsh or low marsh) or quantitative (such as measurements of organic content;

e.g., Plater et al., 2015) and may provide thresholds in a similar fashion to sediment geochemistry in New Jersey. Although secondary proxies are often available to provide additional and independent constraints, a barrier to their use is the lack of an accessible and formal framework for combining multiple proxies with appropriate consideration of uncertainty. A strength of our B-TF is its ability to accommodate these secondary proxy sources. In the example from New Jersey we primarily used a biological proxy (assemblages of foraminifera), but amended the model to include information from a geochemical proxy (bulk sediment  $\delta^{13}\text{C}$ ). On average this approach reduced the uncertainty for PME reconstructions by  $\sim 28\%$ . The reduction in uncertainty consequently provides more precise estimates of RSL and the rate of RSL change through time. This is highlighted in the reconstruction of sea level between  $\sim 500$  BCE and 500 CE where the uncertainty for rate estimates was reduced by 25% for the multi-proxy B-TF compared to the WA-TF and the B-TF. These results highlight the specific utility of bulk sediment  $\delta^{13}\text{C}$  measurements as a sea-level indicator along the mid and northeast Atlantic coast the U.S.A. and the general utility of employing a multi-proxy approach to reconstructing RSL where the goal is to produce reconstructions with the best possible precision. A practical and intuitive means to illustrate the improved performance of the multi-proxy B-TF over the WA-TF model is to compare RSL reconstructions with long-term measurements made by nearby tide gauges (e.g., Kemp et al., 2009b, 2013b; Barlow et al., 2014; Long et al., 2014; Leorri et al., 2008). We compare the reconstruction provided by the WA-TF, B-TF and multi-proxy BTF with regional tide-gauge measurements from New Jersey (Figure 7). The tide-gauges measured about 30 cm of RSL change over the period 1911 to 2012 CE. For each transfer function method, the tide-gauge observations fall within the 95% reconstruction uncertainty bounds, with the multi-proxy B-TF providing notably improved mid-point estimates as indicated by the calculated MSEs.

## 7 Conclusions

To accurately reconstruct the continuous and dynamic evolution of relative sea-level change, we developed a Bayesian hierarchical model comprised of three formally interconnected modules. (1) A B-TF for the calibration of foraminifera into tidal elevation, which is flexible enough to formally accommodate additional proxies such as bulk sediment  $\delta^{13}\text{C}$ . (2) An existing chronology developed from a Bchron age-depth model. (3) An existing EIV-IGP model for estimating rates of sea-level change. Previous reconstructions treated these three components as independent and employed existing approaches that were developed in a variety of numerical frameworks.

Our new B-TF provides an alternative to existing transfer functions. The relationship between species of salt marsh foraminifera and tidal elevation was described using a regional-scale modern training set ( $n = 172$ ) comprised of paired observations of species abundance and elevation, from 12 salt-marshes in southern New Jersey, U.S.A. Results from the B-TF show that six of the eight most

dominant foraminifera do not conform to the unimodal, Gaussian response curve prescribed by the WA-TF and other existing transfer functions.

We applied the transfer functions to cores of salt-marsh sediment that were recovered from two sites in southern New Jersey. The flexible approach utilized in the B-TF results in more variability in reconstructed PME and associated uncertainty among samples than the WA-TF model. This variability is consistent with observed changes in foraminiferal population in core samples and we propose that the B-TF produces a more complete evaluation of uncertainty than the WA-TF model.

The B-TF allows results from additional, independent sea-level proxies to be formally incorporated alongside the primary biological proxy to produce a multi-proxy reconstruction. In New Jersey, we used bulk sediment  $\delta^{13}\text{C}$  values to determine if a core sample formed above or below the MHHW tidal datum. The addition of a second proxy reduced reconstruction uncertainty by an average of 28% and up to  $\sim 70\%$  for samples that formed in some salt-marsh sub environments (the transition from high salt marsh to upland specifically).

We assessed the ability of the multi-proxy B-TF, B-TF and the WA-TF to reconstruct RSL through comparison with observed tide-gauge data from New Jersey. Results showed that the  $2\sigma$  uncertainty bounds for all reconstructions capture the observations from the tide gauge. However, the multi-proxy B-TF provides improved estimates ( $\text{MSE} = 0.003 \text{ m}^2$ ) for the reconstructed RSL mid points compared to the B-TF and the WA-TF ( $\text{MSE} = 0.014 \text{ m}^2$ ). This practical test suggests that the multi-proxy B-TF is the best approach for generating accurate and precise RSL reconstructions using salt-marsh sediment on the mid- and northeast Atlantic coast of the U.S.A.

*Acknowledgements.* We are grateful to the editor Eduardo Zorita, the anonymous reviewer and Robin Edwards (the second reviewer) for their comments that greatly improved the early version of the paper. This research was supported by the Structured PhD in Simulation Science which is funded by the Programme for Research in Third Level Institutions (PRTL) Cycle 5 and co-funded by the European Regional Development Fund, and the Science Foundation Ireland Research Frontiers Programme (2007/RFP/MATF281) and also supported by the National Science Foundation awards EAR and OCE 1458904.

## References

- Barlow, N. L. M., Long, A. J., Saher, M. H., Gehrels, W. R., Garnett, M. H., and Scaife, R. G.: Salt-marsh reconstructions of relative sea-level change in the North Atlantic during the last 2000 years, *Quaternary Science Reviews*, 99, 1–16, 2014.
- Birks, H. J.: Overview of numerical methods in palaeolimnology, in: *Tracking Environmental Change Using Lake Sediments: Data Handling and Numerical Techniques*, edited by Birks, H., Lotter, A., Juggins, S., and Smol, J., vol. 5 of *Tracking environmental change using lake sediments*, book section 2, Springer, 2012.
- Birks, H. J. B.: Quantitative Palaeoenvironmental Reconstructions, in: *Statistical Modelling of Quaternary Science Data*, edited by Maddy, D. and Brew, J. S., vol. Technical Guide 5 of *Technical Guide*, pp. 161–254, Quaternary Research Association, Cambridge, 1995.
- Birks, H. J. B., Line, J. M., Juggins, S., Stevenson, A. C., and ter Braak, C. J. F.: Diatoms and pH Reconstruction, *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 327, 263–278, 1990.
- Birks, J.: Calibration, Transfer Functions and Environmental Reconstructions, in: *PAGES*, 2010.
- Blaauw, M. and Christen, J. A.: Flexible paleoclimate age-depth models using an autoregressive gamma process, *Bayesian Anal.*, pp. 457–474, 2011.
- Bradley, R. S.: Chapter 1 - Paleoclimatic Reconstruction, in: *Paleoclimatology (Third Edition)*, edited by Bradley, R. S., pp. 1–11, Academic Press, San Diego, third edition edn., 2015.
- Bronk Ramsey, C.: Radiocarbon Dating: Revolutions in Understanding\*, *Archaeometry*, 50, 249–275, 2008.
- Cahill, N., Kemp, A. C., Horton, B. P., and Parnell, A. C.: Modeling Sea-Level Change using Errors-in-Variables Integrated Gaussian Processes, *Annals of Applied Statistics*, 9, 547–571, 2015.
- Charman, D. J., Gehrels, W. R., Manning, C., and Sharma, C.: Reconstruction of recent sea-level change using testate amoebae, *Quaternary Research*, 73, 208–219, 2010.
- De Boor, C.: *A Practical Guide to Splines*, Springer, 1978.
- de Rijk, S.: Salinity control on the distribution of salt marsh foraminifera (Great Marshes, Massachusetts), *The Journal of Foraminiferal Research*, 25, 156–166, 1995.
- Dey, D. K., Ghosh, S. K., and Mallick, B. K.: *Generalized Linear Models A Bayesian Perspective*, Marcel Dekker, Inc, 2000.
- Dierckx, P.: *Curve and Surface Fitting with Splines*, Clarendon Press, Oxford, 1993.
- Douglas, B. C.: Global sea level rise, *Journal of Geophysical Research: Oceans*, 96, 6981–6992, 1991.
- Dutton, A., Carlson, A. E., Long, A. J., Milne, G. A., Clark, P. U., DeConto, R., Horton, B. P., Rahmstorf, S., and Raymo, M. E.: Sea-level rise due to polar ice-sheet mass loss during past warm periods, *Science*, 349, 2015.
- Edwards, R. and Wright, A. J.: Foraminifera, in: *Handbook of Sea-Level Research*, edited by Shennan, I., Long, A. J., and Horton, B. P., book section 13, pp. 191–217, John Wiley & Sons, 2015.
- Edwards, R. J. and Horton, B. P.: Developing detailed records of relative sea-level change using a foraminiferal transfer function: an example from North Norfolk, UK, *Philosophical Transactions of the Royal Society A - Mathematical Physical and Engineering Sciences*, 364, 973–991, 2006.
- Edwards, R. J., Wright, A. J., and van de Plassche, O.: Surface distributions of salt-marsh foraminifera from Connecticut, USA: modern analogues for high-resolution sea level studies, *Marine Micropaleontology*, 51, 1–21, 2004.

- Fritz, S. C., Juggins, S., Battarbee, R. W., and Engstrom, D.: Reconstruction of past changes in salinity and climate using a diatom-based transfer function, *Nature*, 352, 706–708, 1991.
- Gehrels, W. R.: Determining Relative Sea-Level Change from Salt-Marsh Foraminifera and Plant Zones on the Coast of Maine, U.S.A., *Journal of Coastal Research*, 10, 990–1009, 1994.
- Gehrels, W. R.: Using foraminiferal transfer functions to produce high-resolution sea-level records from salt-marsh deposits, Maine, USA, *The Holocene*, 10, 367–376, 2000.
- Gneiting, T. and Raftery, A. E.: Strictly Proper Scoring Rules, Prediction, and Estimation, *Journal of the American Statistical Association*, 102, 359–378, 2007.
- Greig-Smith, P.: Correlation with Habitat Factors, in: *Quantitative Plant Ecology*, edited by Anderson, D. J., Greig-Smith, P., and Pitelka, F. A., book section 5, pp. 129–145, University of California Press, Berkeley and Los Angeles, 1983.
- Haslett, J. and Parnell, A. C.: A simple monotone process with application to radiocarbon-dated depth chronologies, *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 57, 399–418, 2008.
- Haslett, J., Whitley, M., Bhattacharya, S., Salter-Townshend, M., Wilson, S. P., Allen, J. R. M., Huntley, B., and Mitchell, F. J. G.: Bayesian palaeoclimate reconstruction, *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 169, 395–438, 2006.
- Holsclaw, T., Sanso, B., Lee, H. K. H., Heitmann, K., Habib, S., Higdon, D., and Alam, U.: Gaussian Process Modeling of Derivative Curves, *Technometrics*, 55, 57–67, 2013.
- Horton, B. P. and Edwards, R. J.: The application of local and regional transfer functions to the reconstruction of Holocene sea levels, north Norfolk, England, *The Holocene*, 15, 216–228, 2005.
- Horton, B. P. and Edwards, R. J.: Quantifying Holocene sea-level change using intertidal foraminifera: lessons from the British Isles, *Cushman Foundation for Foraminiferal Research, Special Publication*, 40, 97, 2006.
- Horton, B. P., Edwards, R. J., and Lloyd, J. M.: UK intertidal foraminiferal distributions: implications for sea-level studies, *Marine Micropaleontology*, 36, 205–223, 1999.
- Imbrie, J. and Kipp, N. G.: A new micropalaeontological method for quantitative paleoclimatology: application to a late Pleistocene Caribbean core, in: *The Late Cenozoic Glacial Ages*, edited by Turekian, K. K., pp. 71–181, Yale University Press, New Haven and London, 1971.
- Jackson, S. T. and Williams, J. W.: Modern analogs in Quaternary paleoecology: Here today, gone yesterday, gone tomorrow?, *Annual Review of Earth and Planetary Sciences*, 32, 495–537, 2004.
- Juggins, S. and Birks, H. J.: Quantitative Environmental Reconstructions From Biological Data, in: *Tracking Environmental Change Using Lake Sediments: Data Handling and Numerical Techniques*, edited by Birks, H., Lotter, A., Juggins, S., and Smol, J., vol. 5 of *Tracking environmental change using lake sediments*, pp. 431–494, Springer, 2012.
- Kemp, A. C., Horton, B. P., Corbett, D. R., Culver, S. J., Edwards, R. J., and van de Plassche, O.: The relative utility of foraminifera and diatoms for reconstructing late Holocene sea-level change in North Carolina, USA, *Quaternary Research*, 71, 9–21, 2009a.
- Kemp, A. C., Horton, B. P., Culver, S. J., Corbett, D. R., van de Plassche, O., Gehrels, W. R., Douglas, B. C., and Parnell, A. C.: Timing and magnitude of recent accelerated sea-level rise (North Carolina, United States), *Geology*, 37, 1035–1038, 2009b.

- Kemp, A. C., Horton, B. P., Donnelly, J. P., Mann, M. E., Vermeer, M., and Rahmstorf, S.: Climate related sea-level variations over the past two millennia, *Proceedings of the National Academy of Sciences*, 108, 11 017–11 022, 2011.
- Kemp, A. C., Vane, C. H., Horton, B. P., Engelhart, S. E., and Nikitina, D.: Application of stable carbon isotopes for reconstructing salt-marsh floral zones and relative sea level, New Jersey, USA, *Journal of Quaternary Science*, 27, 2012.
- Kemp, A. C., Horton, B. P., Vane, C. H., Corbett, D. R., Bernhardt, C. E., Engelhart, S. E., Anisfeld, S. C., Parnell, A. C., and Cahill, N.: Sea-level change during the last 2500 years in New Jersey, USA, *Quaternary Science Reviews*, 81, 90–104, 2013a.
- Kemp, A. C., Telford, R. J., Horton, B. P., Anisfeld, S. C., and Sommerfield, C. K.: Reconstructing Holocene sea-level using salt-marsh foraminifera and transfer functions: lessons from New Jersey, USA, *Journal of Quaternary Science*, 28, 617–629, 2013b.
- Korsman, T. and Birks, H. J.: Diatom-based water chemistry reconstructions from northern Sweden: a comparison of reconstruction techniques, *Journal of Paleolimnology*, 15, 65–77, 1996.
- Leorri, E., Horton, B. P., and Cearreta, A.: Development of a foraminifera-based transfer function in the Basque marshes, N. Spain: Implications for sea-level studies in the Bay of Biscay, *Marine Geology*, 251, 60–74, 2008.
- Li, B., Nychka, D. W., and Ammann, C. M.: The value of multiproxy reconstruction of past climate, *Journal of the American Statistical Association*, 105, 883–895, 2010.
- Long, A. J., Barlow, N. L. M., Gehrels, W. R., Saher, M. H., Woodworth, P. L., Scaife, R. G., Brain, M. J., and Cahill, N.: Contrasting records of sea-level change in the eastern and western North Atlantic during the last 300 years, *Earth and Planetary Science Letters*, 388, 110–122, 2014.
- Mann, M. E., Zhang, Z., Rutherford, S., Bradley, R. S., Hughes, M. K., Shindell, D., Ammann, C., Faluvegi, G., and Ni, F.: Global Signatures and Dynamical Origins of the Little Ice Age and Medieval Climate Anomaly, *Science*, 326, 1256–1260, 2009.
- Murray, J. W.: *Ecology and Palaeoecology of Benthic Foraminifera*, Elsevier, Amsterdam, 1991.
- Nelson, A. R.: Coastal sediments, in: *Handbook of Sea-Level Research*, edited by Shennan, I., Long, A. J., and Horton, B. P., pp. 47–65, John Wiley & Sons, 2015.
- Nelson, A. R., Jennings, A. E., and Kashima, K.: An earthquake history derived from stratigraphic and microfossil evidence of relative sea-level change at Coos Bay, southern coastal Oregon, *Geological Society of America Bulletin*, 108, 141–154, 1996.
- Nelson, A. R., Sawai, Y., Jennings, A. E., Bradley, L., Gerson, L., Sherrod, B. L., Sabeau, J., and Horton, B. P.: Great-earthquake paleogeodesy and tsunamis of the past 2000 years at Alsea Bay, central Oregon coast, USA, *Quaternary Science Reviews*, 27, 747–768, 2008.
- Nicolaus, N.: *Nicolai Stenonis De solido intra solidum naturaliter contento dissertationis prodromus., Ex typographia sub signo stellae, Florentiae, 1669.*
- Parnell, A. C. and Gehrels, W. R.: Using chronological models in late Holocene sea-level reconstructions from salt marsh sediments, in: *Handbook of Sea-Level Research*, edited by Shennan, I., Long, A. J., and Horton, B. P., book section 32, pp. 500–513, John Wiley & Sons, 2015.

- Parnell, A. C., Buck, C. E., and Doan, T. K.: A review of statistical chronology models for high-resolution, proxy-based Holocene palaeoenvironmental reconstruction, *Quaternary Science Reviews*, 30, 2948–2960, 2011.
- Parnell, A. C., Sweeney, J., Doan, T. K., Salter-Townshend, M., Allen, J. R. M., Huntley, B., and Haslett, J.: Bayesian inference for palaeoclimate with time uncertainty and stochastic volatility, *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 64, 115–138, 2015.
- Plater, A. J., Kirby, J. R., Boyle, J. F., Shaw, T., and Mills, H.: Loss on ignition and organic content, in: *Handbook of Sea-Level Research*, edited by Shennan, I., Long, A. J., and Horton, B. P., pp. 312–330, John Wiley & Sons, 2015.
- Plummer, M.: JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling, in: *Proceedings of the 3rd International Workshop on Distributed Statistical Computing*, 2003.
- Rasmussen, C. E.: *Gaussian Processes for Machine Learning*, MIT Press, 2006.
- Roe, H. M., Charman, D. J., and Gehrels, R. W.: Fossil testate amoebae in coastal deposits in the UK: implications for studies of sea-level change, *Journal of Quaternary Science*, 17, 411–429, 2002.
- Rymer, L.: The use of uniformitarianism and analogy in palaeoecology, particularly pollen analysis, in: *Biology and Quaternary Environments*, edited by Walker, D. and Guppy, J. C., pp. 245–257, Australian Academy of Sciences, Canberra, 1978.
- Scott, D. B. and Medioli, F. S.: Vertical zonations of marsh foraminifera as accurate indicators of former sea levels, *Nature*, 272, 528–531, 1978.
- Shennan, I., Innes, J. B., Long, A. J., and Zong, Y.: Late Devensian and Holocene relative sealevel changes at Loch nan Eala, near Arisaig, northwest Scotland, *Journal of Quaternary Science*, 9, 261–283, 1994.
- Shennan, I., Rutherford, M. M., Innes, J. B., and Walker, K. J.: Late glacial sea level and ocean margin environmental changes interpreted from biostratigraphic and lithostratigraphic studies of isolation basins in northwest Scotland, *Geological Society, London, Special Publications*, 111, 229–244, 1996.
- ter Braak, C. J. and Juggins, S.: Weighted averaging partial least squares regression (WA-PLS): an improved method for reconstructing environmental variables from species assemblages, *Hydrobiologia*, 269–270, 485–502, 1993.
- Tingley, M. P., Craigmile, P. F., Haran, M., Li, B., Mannshardt, E., and Rajaratnam, B.: Piecing together the past: statistical insights into paleoclimatic reconstructions, *Quaternary Science Reviews*, 35, 1 – 22, 2012.
- Toivonen, H. T. T., Mannila, H., Korhola, A., and Olander, H.: APPLYING BAYESIAN STATISTICS TO ORGANISM-BASED ENVIRONMENTAL RECONSTRUCTION, *Ecological Applications*, 11, 618–630, 2000.
- Tolwinski-Ward, S., Tingley, M., Evans, M., Hughes, M., and Nychka, D.: Probabilistic reconstructions of local temperature and soil moisture from tree-ring data with potentially time-varying climatic response, *Climate Dynamics*, 44, 791–806, 2015.
- Tolwinski-Ward, S. E.: Uncertainty quantification for a climatology of the frequency and spatial distribution of North Atlantic tropical cyclone landfalls, *Journal of Advances in Modeling Earth Systems*, 7, 305–319, 2015.
- Tolwinski-Ward, S. E., Anchukaitis, K. J., and Evans, M. N.: Bayesian parameter estimation and interpretation for an intermediate model of tree-ring width, *Climate of the Past*, 9, 1481–1493, 2013.

- Vasko, K., Toivonen, H. T., and Korhola, A.: A Bayesian multinomial Gaussian response model for organism-based environmental reconstruction, *Journal of Paleolimnology*, 24, 243–250, 2000.
- Wright, A. J., Edwards, R. J., and van de Plassche, O.: Reassessing transfer-function performance in sea-level reconstruction based on benthic salt-marsh foraminifera from the Atlantic coast of NE North America, *Marine Micropaleontology*, 81, 43 – 62, 2011.
- Zong, Y. and Horton, B. P.: Diatom-based tidal-level transfer functions as an aid in reconstructing Quaternary history of sea-level movements in the UK, *Journal of Quaternary Science*, 14, 153–167, 1999.