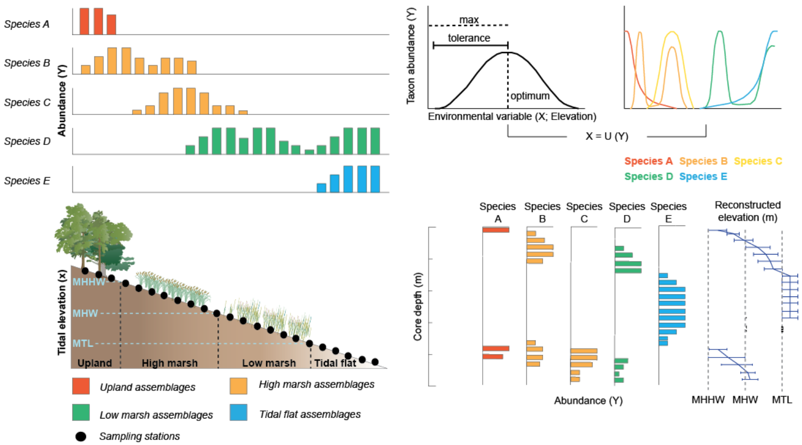
A BAYESIAN STATISTICAL MODEL FOR RECONSTRUCTING AND ANALYSING FORMER SEA LEVELS

33rd annual meeting of the Irish Mathematical Society

Niamh Cahill, Andrew Parnell, Andrew Kemp, Benjamin Horton

January 14th 2021

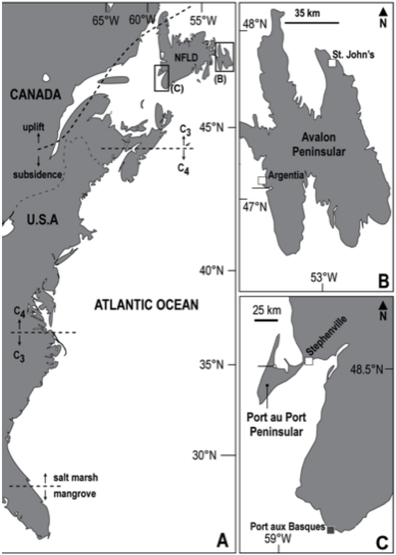
Reconstructing Relative Sea Level

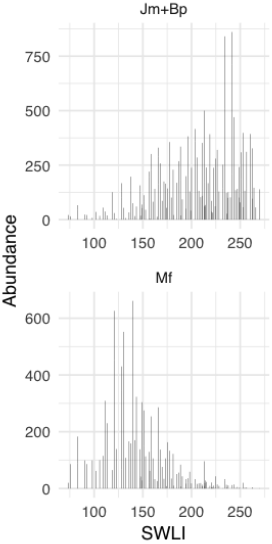


[1]

[1] Figure credit: Dr Isabel Hong

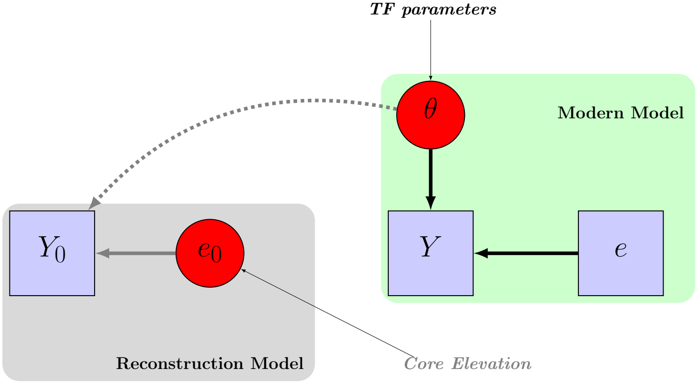
Case Study: Newfoundland



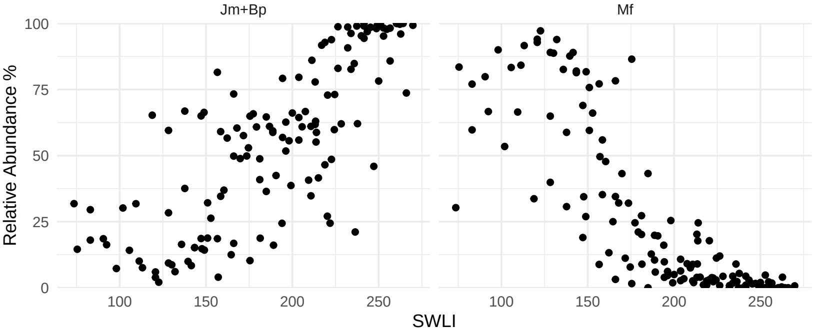


A Bayesian Transfer Function (BTF)

$$\underbrace{P(parameters|Data)}_{\text{Posterior}} \propto \underbrace{P(Data|parameters)}_{\text{Likelihood}} \underbrace{P(parameters)}_{\text{Prior}}$$



Species Response Curves



The Process Model

- For modelling puposes we consider a latent species response variable
- Each latent species variable has a functional relationship with elevation:

$$\mathbf{s}_j = f_{\theta_j}(\text{elevation}) + \epsilon_j$$

- f is a penalised-spline function and θ_j is a vector of parameters controlling the shape of the spline for species j.
- $\epsilon_{ij} \sim N(0, \sigma_j^2)$ and is added to account for overdispersion (i.e, the data here are likely to exhibit more variation than may be capturped with the data model).

The Data Model

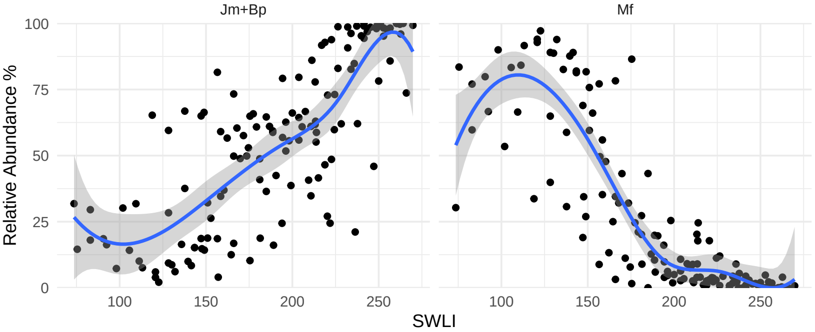
- The observed species abundances are multinomial

$$(y_{i1}, ..., y_{iM}) \sim Multi(p_{i1}, ..., p_{iM}, N_i)$$

- The probabilites of the multinomial distribution are estimated as a function of the latent species variable via a softmax transformation where

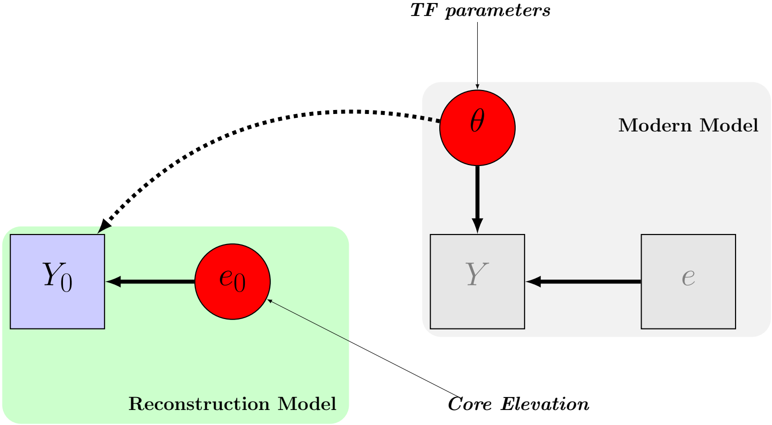
$$p_{ij} = \frac{e^{s_{ij}}}{\sum_{j=1}^{M} e^{s_{ij}}}$$

Species Response Curves

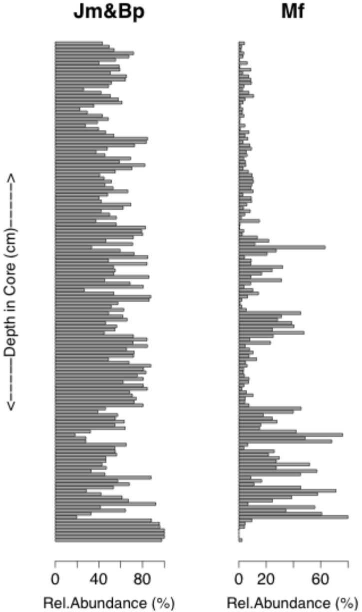


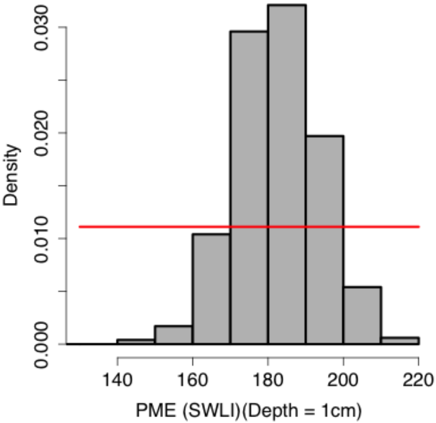
A Bayesian Transfer Function (BTF)

$$\underbrace{P(parameters|Data)}_{\text{Posterior}} \propto \underbrace{P(Data|parameters)}_{\text{Likelihood}} \underbrace{P(parameters)}_{\text{Prior}}$$

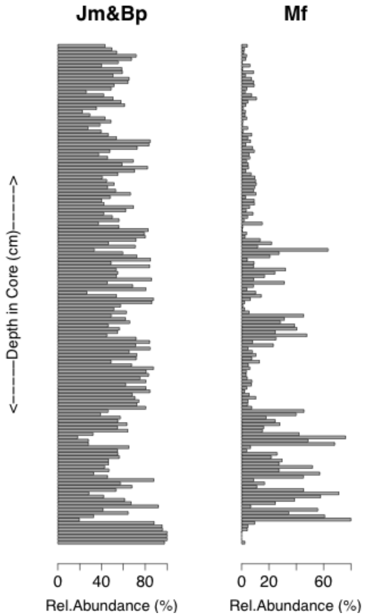


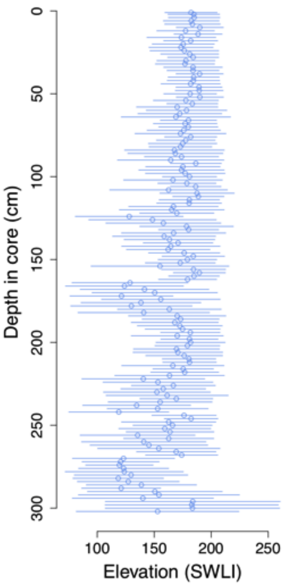
Core Results for Newfoundland





Core Results for Newfoundland

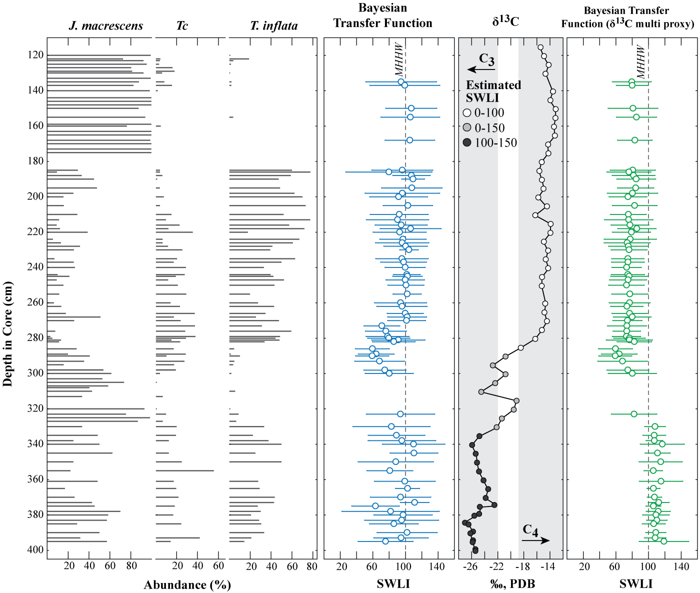




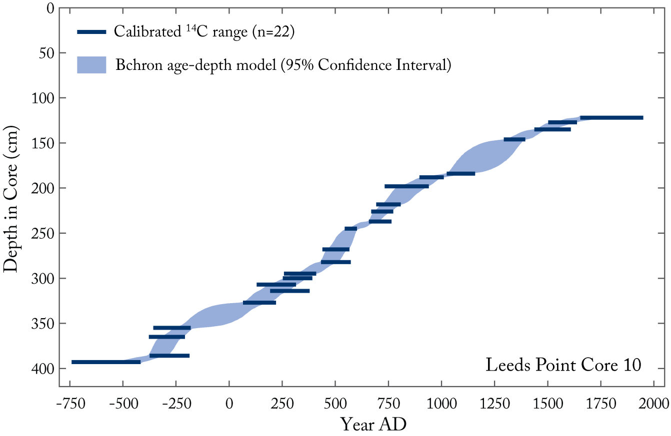
Using Secondary Proxies

A Case Study from New Jersey, USA

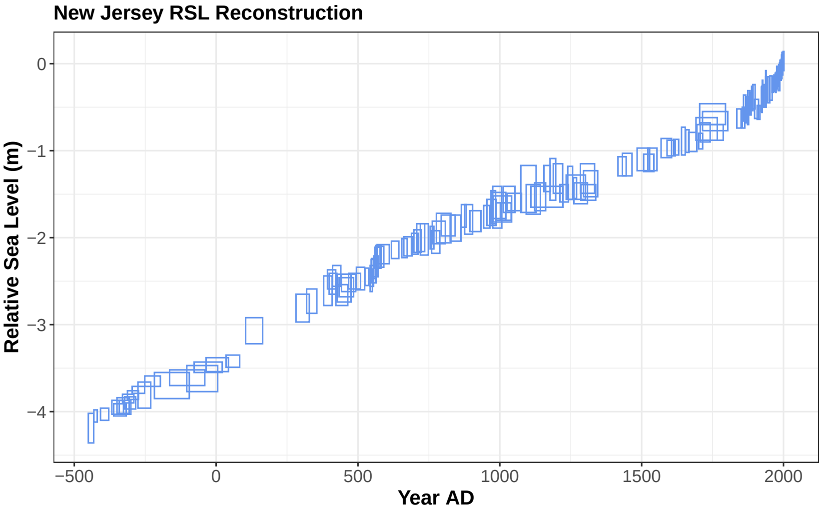
Core Results for New Jersey



Developing a Chronology for New Jersey



A RSL Reconstruction for New Jersey



Gaussian processes

Consider the multivariate normal distribution for some k-dimension random vector y

$$y \sim MVN(0, \Sigma)$$

The matrix Sigma currently looks like this:

$$\Sigma = \begin{bmatrix} \sigma^2 & 0 & \dots & 0 \\ 0 & \sigma^2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma^2 \end{bmatrix}$$

- The *key idea* in Gaussian Processes is to change the off-diagonals of Σ so that the you get higher correlations between y_i and y_j when t_i and t_j are close.

Gaussian Processes cont.

- A zero-mean Gaussian process is defined as follows:

$$y \sim MVN(0, K)$$

- one commonly used option is the squared exponential covariance function:

$$K_{i,j} = \sigma_g^2 exp\Big(- \rho^2 (t_i - t_j)^2 \Big)$$

- σ_g^2 should be high for functions that cover a broad range on the y-axis
- if t_i is distant from t_j then $K_{i,j} \approx 0$
- the effect of $t_i - t_j$ on $K_{i,j}$ will depend on ρ

The Process model

- We assume a GP model for the rate of sea-level change

$$\omega(t) \sim MVN(0, K)$$

Where $K_{i,j} = \sigma_g^2 exp\Big(- \rho^2 (t_i - t_j)^2 \Big)$

- Then the sea-level process is obtained by integrating the rate process

$$s(t) = \int_0^t \omega(u) du$$

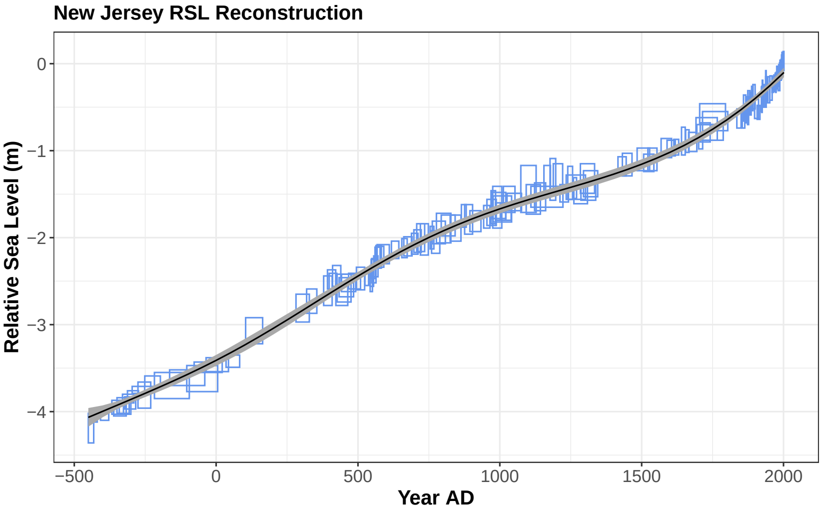
The Data Model

- The observed sea-level values are assumed to be normally distributed

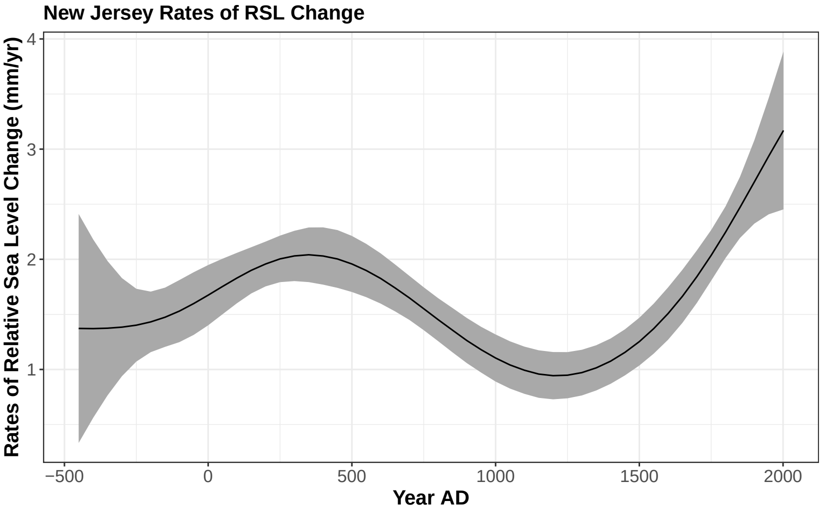
$$y_i \sim N(s(t_i), \sigma_i^2)$$

where σ_i^2 will capture the variation of the observed sea-level values around the mean.

Model Results for New Jersey



Model Results for New Jersey



References

N Cahill, A C. Kemp, B P. Horton and A C. Parnell. A Bayesian Hierarchical Model for Reconstructing Relative Sea Level: From Raw Data to Rates of Change.Climate of the Past} , 12(2):525-542, 2016.

N Cahill, A C. Kemp, B P. Horton and A C. Parnell. Modeling sea-level change using errors-in-variables integrated Gaussian processes. Annals of Applied Statistics, 9(2): 547-571, 2015.