

Quantitative Bayesian reconstruction of the NAO impact in the central Iberian Peninsula for the last two millennia

Sánchez-López et al

1 2.3 A Bayesian model to reconstruct NAO

To produce reconstructions of NAO, we follow the Bayesian modelling approach of Parnell et al (2015, 2016). Whilst Parnell et al (2015) based their framework on reconstructing multivariate temperature and moisture measurements from pollen, the method they developed can directly transfer over to other proxies and climate variables. Indeed Cahill et al (2016) used the same approach to reconstruct sea level from Foraminifera. In all cases the counts of the proxy are created for a set of sediment layers (depths) in a core. In our approach the proxy is the XRF chemical compositions taken from the core at Címera Lake, and we reconstruct univariate NAO.

We describe the mathematical details of the model in this section. However, we do not go through the full details of the model fitting process, as this involves some convoluted mathematics which is fully described in Parnell et al 2015. However we do provide all the code we use to create the reconstructions at XXX (URL LINK TO CODE?).

The notation we use is as follows:

- $NAO(t)$ is the North Atlantic Oscillation value at time t . The goal of our model is to estimate this value for a set of chosen times.
- XRF_{ij} (expressed as counts per second) represents the chemical element j measured at slice i of the CIM12-04A core. We have $i = 1, \dots, 647$ slices and $j = 1, \dots, 13$ elements.
- We superscript both the quantities above with m and f so that XRF^m refers to the modern XRF data set, with associated known NAO^m , and XRF^f to refer to the fossil data set, for which we wish to estimate NAO^f .

- t_i refers to the age of slice i . The ages in our core are all given in years AD
- θ refers to the set of parameters which govern the relationship between NAO and the XRF measurements, and the dynamics of how NAO changes over time.

Our model proceeds by creating a Bayesian joint posterior distribution:

$$p(\text{NAO}^f, \theta | \text{NAO}^m, \text{XRF}) \propto p(\text{XRF} | \text{NAO}) \times p(\text{NAO} | \theta) \times p(\theta)$$

The term on the left hand side of the equation is the posterior distribution and represents the probability distribution of fossil NAO given the data. The terms on the right hand side represent respectively the likelihood (the probability distribution of XRF given NAO), the prior distribution on NAO dynamics, and the prior distribution on the parameters governing the relationship between XRF and NAO.

For the likelihood, we standardise all the XRF values (by chemical element) and fit a multivariate Gaussian polynomial regression model. This means, for the 13 chemical elements we use:

$$[\text{XRF}_{i,1}, \dots, \text{XRF}_{i,13}] | \text{NAO}(t_i) \sim N(M_i, \Sigma)$$

where $M_i = [\mu_{i,1}, \dots, \mu_{i,13}]$ with $\mu_{ik} = \beta_{0k} + \beta_{1k}\text{NAO}(t_i) + \beta_{2k}\text{NAO}(t_i)^2$, and Σ is a covariance matrix which captures the extra dependence between elements not explained by differences in NAO. We also tried Gamma GLM models with the unstandardised data, but there was no discernable improvement in performance, and this added extra unnecessary parameters and transformations.

We set the prior distribution on NAO as a continuous-time random walk which should reasonably match climate behavior over the reconstructed time period (as in e.g., Haslett et al., 2006). Other choices are available, such as long-memory or long-tailed stochastic processes (e.g. Parnell et al 2015 who use a Normal inverse Gaussian process). Our prior distribution is:

$$\text{NAO}(t_i) \sim N(\text{NAO}(t_{i-1}), \sigma^2(t_i - t_{i-1}))$$

where σ^2 is a parameter representing the variance of the NAO increments for a unit time.

Finally we set vague prior distributions on the remaining parameters:

$$\beta_{0k}, \beta_{1k}, \beta_{2k} \sim N(0, 10), \sigma \sim U(0, 10), \Sigma^{-1} \sim Wishart(I_{13}, 14)$$

2 2.4 Model fitting and validation

The above model is hard to fit using the default tools for Bayesian model fitting due to the large number of parameters and the high dimension of the data. Instead, as stated above, we follow the approach of Parnell et al (2015) which involves a computational approximation to fit the model in three steps. The first step involves fitting the model to the modern data only. The second step involves estimating NAO for the fossil layers, and the third step involves constraining the estimated fossil NAO values according to the random walk model.

We fit the model in R (R Core Team 2016) and use the JAGS software (Just Another Gibbs Sampler; Plummer, 2003). The performance of the fitting algorithm can be determined by looking at the Brooks-Gelman Rubin \hat{R} statistic (Brooks and Gelman, 1998; Gelman and Rubin, 2001). We run the algorithm until all \hat{R} values are less than 1.05.

We evaluate the performance of the model itself by testing the modern relationship between NAO and XRF (step 1 as outlined above). Whilst the relationship between NAO and XRF for each chemical element is likely to be weak, we would hope that NAO would be estimated with some accuracy when these are taken together. To check this we created predicted NAO values from the modern model using the modern data. The results are shown in the Figure below. If the model is estimating NAO correctly there should be only 5% of observations outside the 95% interval, and 50% outside the 50% interval. In our validation we observe 2.9% of observation outside the 95% interval, and 44.9% outside the 50% interval. These indicate good performance of the model.

Finally, we show the complete reconstruction in the Figure below. The reconstruction is created using a 10-year time grid, and includes both 95% and 50% uncertainty intervals. The large uncertainty during the period 1200–1260 is due to the lack of proxy data during this period.

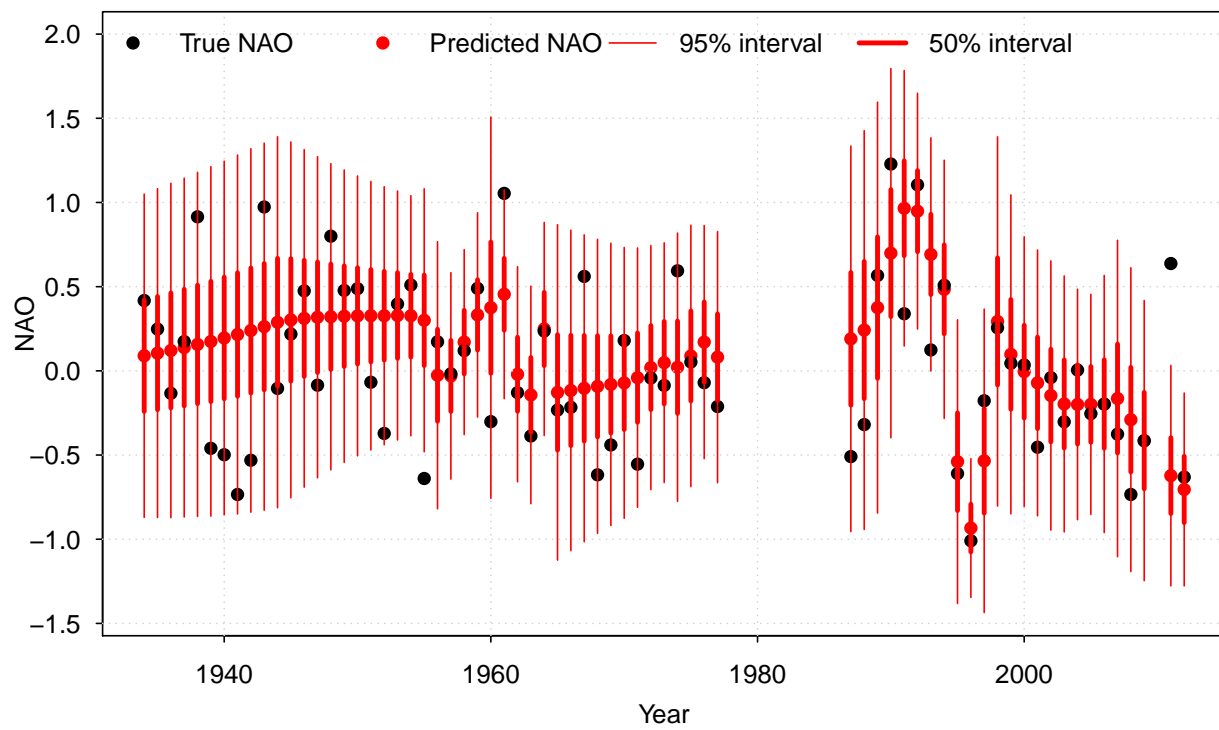


Figure 1:

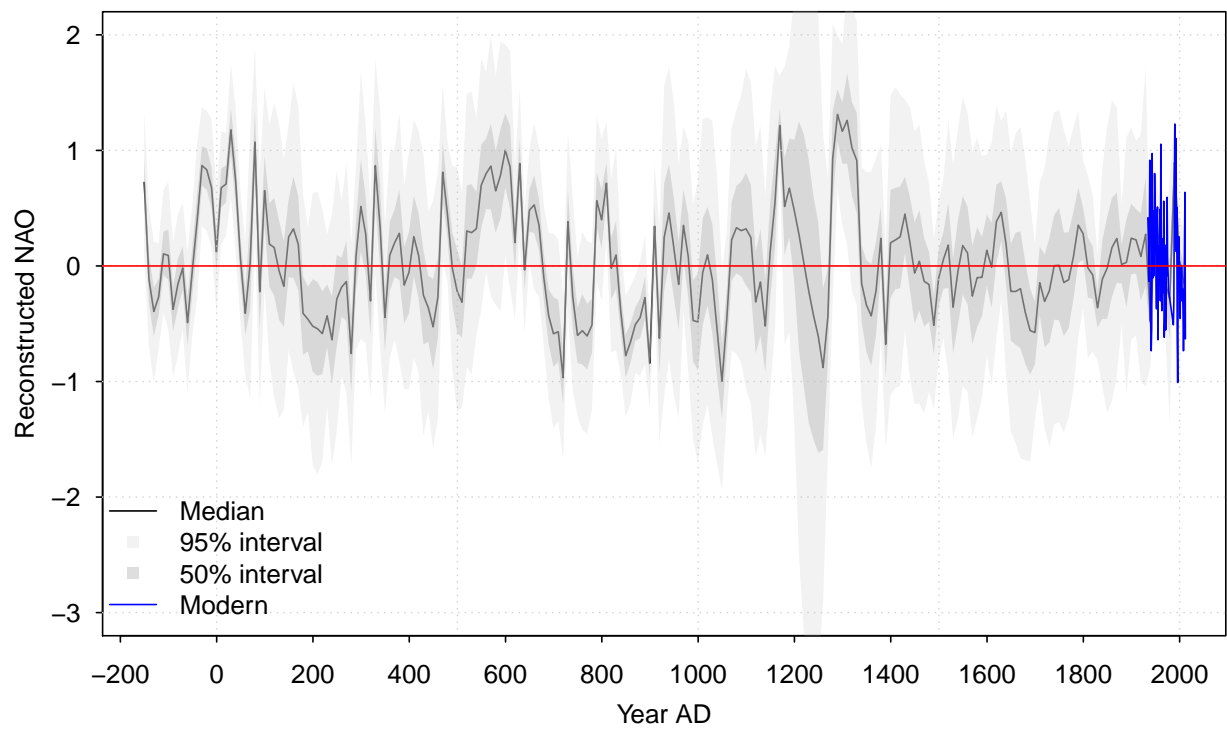


Figure 2: