Joint inversion of proxy system models to reconstruct paleoenvironmental time series from heterogeneous data

Gabriel J. Bowen, Brenden Fisher-Femal, Gert-Jan Reinhart, Appy Sluijs, Caroline Lear

# Abstract

Paleoclimatic and paleoenvironmental reconstructions are fundamentally uncertain because no known proxy is a direct record of a single environmental variable of interest; all proxies are sensitive to multiple forcing factors. One productive approach to reducing proxy uncertainty is the integration of information from multiple proxy systems with complimentary, overlapping sensitivity to environmental one or more target variables. Most such analyses are conducted in an ad-hoc fashion, either through qualitative comparison to assess the similarity of single-proxy reconstructions or through step-wise quantitative interpretations where one proxy is used to constrain a variable relevant to the interpretation of a second proxy. Here we propose a formal framework for the integration of multiple proxies via the joint inversion of proxy system models. The “Joint Proxy Inversion” (JPI) method provides a statistically robust approach to producing self-consistent interpretations of multiproxy datasets, allowing full and simultaneous assessment of all proxy and model uncertainties to obtain quantitative estimates of past environmental conditions. Other benefits of the method include the ability to use independent information on climate and environmental systems to inform the interpretation of proxy data, to fully leverage information from unevenly- and differently-sampled proxy records, and to obtain refined estimates of proxy model parameters that are conditioned on paleo-archive data. Application of JPI to the marine Mg/Ca and δ18O proxy systems at two distinct timescales demonstrates many of the key properties, benefits, and sensitivities of the method, and produces new, statistically-grounded reconstructions of Neogene ocean temperature and chemistry from previously published proxy data. We suggest that JPI is a universally applicable method that can be implemented using proxy models of wide-ranging complexity to generate more robust, quantitative understanding of past climatic and environmental change.

# Introduction

Paleoenvironmental reconstructions, including reconstructions of past climate, provide a powerful tool to document the sensitivity of Earth systems to forcing, characterize the range of natural responses associated with different modes of global change, and identify key mechanisms governing these responses. Throughout the vast majority of the planet’s history, however, estimates of environmental conditions can only be obtained through proxy reconstructions. Proxy data are, as their name implies, indirect recorders of the environmental conditions we hope to estimate, and the estimates they provide are plagued by substantial, often poorly characterized, uncertainty.

The simplest proxy reconstructions typically focus on a single environmental variable of interest, inverting experimental or natural calibration datasets to obtain quantitative estimates of that variable and treating residual variance in the calibration as ‘noise’. In reality, however, no proxy system exists that is sensitive only to a single paleoenvironmentally-relevant variable, the majority of this noise reflects the uncharacterized influence of other environmental variables on the proxy system. Fossil leaf assemblages exhibit characteristic variability that can be associated with mean annual air temperature, but also correlate with and may be influenced many other environmental variables and evolutionary history (Greenwood et al. 2004, Royer et al. 2005). The saturation state of alkenones produced by marine phytoplankton is a sensitive recorder of water temperature, but characteristics of alkenones preserved in marine sediments appear to also be strongly affected by physiological factors, seasonality of production, and selective degradation (Conte et al. 1998, Conte et al. 2006). Even recently emerging clumped isotope techniques, which are in theory a direct recorder of the temperature of carbonate mineral formation, can be affected by factors such as growth-rate effects, carbonate system disequilibrium, and poorly constrained, potentially variable offsets between the environment of carbonate formation and more commonly targeted atmospheric temperature conditions (Passey et al. 2010, Saenger et al. 2012, Affek et al. 2014). Failure to recognize and consider the sensitivity of proxies to multiple environmental factors leads to two important problems in traditional proxy interpretations. First, considering only a single environmental variable in our interpretations maximizes the uncertainty in our reconstructions, which could be reduced if the influence of other variables is described and constrained. Second, unacknowledged sensitivity to multiple variables creates potential for biased proxy interpretations if variation in these variables is non-random over the period being reconstructed.

One productive approach to resolving the second issue is the development and use of proxy system models in the interpretation of proxy data (Evans et al. 2013). These models represent an attempt to mathematically describe the complex of environmental, physical, and biological factors that control how environmental signals are sampled, recorded, and preserved in proxy measurements. Several recent reviews and perspectives are available discussing the concepts underlying proxy system models and different ways that they have been applied to proxy interpretation, ranging from substitution for empirical calibrations in inverse estimation of environmental signals to formal integration within climate model data assimilation schemes (Evans et al. 2013). A growing number of proxy system models and modeling systems are being developed (e.g., Tolwinski-Ward et al. 2011, Stoll et al. 2012, Dee et al. 2015), and useful models span a range of complexity from empirically-constrained regressions to fully mechanistic, theory-based formulations. Key to any such model is accurate representation of uncertainty in each model component, which allows even relatively simple, potentially incomplete models to be used to obtain reconstructions with quantifiable uncertainty bounds.

Reducing the uncertainty of quantitative paleoenvironmental reconstructions requires adding constraints to proxy interpretations, as exemplified by the increasingly common development of multi-proxy reconstructions. In situations where two or more proxies share sensitivity to common or complimentary environmental variables, it stands to reason that the information provided by each can be used to refine interpretation of the multi-proxy suite. In practice, a variety of approaches have been used. Commonly, multi-proxy integration has been qualitative and focused on confirmation: trends reconstructed using one proxy system are cross-checked against a second, providing increased confidence in the reconstruction where the patterns match and further investigation where they don’t (e.g., Keating-Bitonti et al. 2011, Grauel et al. 2013). In other cases, proxies have been combined quantitatively, but usually in a stepwise fashion: one proxy system is used to estimate an environmental variable to which it is sensitive, and those reconstructed values are then used to constrain the interpretation of a second proxy formed at the same time (e.g., Fricke et al. 1998, Lear et al. 2000). Although it provides a simple strategy to combining complimentary proxy information, this approach does not fully leverage overlapping information that may be contained in multiple systems that respond to common forcing, is not conducive to robust quantification of uncertainty, and requires that both proxies sample coeval paleoenvironmental conditions.

Here we propose a general approach to proxy interpretation that leverages the benefits of proxy models and provides a formal statistical framework for proxy integration. The method, which we call Joint Proxy Inversion (JPI), cast in a Bayesian inferential framework, uses Markov Chain Monte Carlo methods to invert a hierarchy of proxy-system and environmental time series models. These models are simultaneously conditioned on proxy data and calibration datasets to obtain posterior estimates of the target paleoclimate time series with quantifiable uncertainty that accounts for all model components. We first describe the framework as applied to joint inversion of the marine Mg/Ca and δ18O proxy systems. We choose this system not because it is particularly challenging, but because it represents a well-established and widely-used multi-proxy system where diverse examples of proxy integration are available. We then present applications of JPI to two such datasets recording paleoceanographic conditions at two different timescales. These applications provide a platform allowing us to illustrate and discuss merits and limitations of JPI relevant to its broader application.

# Methods

Bayesian inversion of proxy system models has been demonstrated previously, for example to estimate site-specific model parameters (Tolwinski-Ward et al. 2013). JPI extends this approach to the simultaneous inversion of multiple proxy system and environmental time series models, conditioned on multiproxy data, to generate estimates of the posterior distribution of model parameters including the target paleoenvironmental time series.

## Data

Proxy and proxy model calibration datasets were compiled from published work (Fig. 1). Estimates from fluid inclusions, calcite veins, and echinoderm fossils (Lowenstein et al. 2001, Dickson 2002, Coggon et al. 2010) were combined with information on modern seawater Mg/Ca (de Villiers & Nelson 1999) to represent variation in seawater Mg/Ca since 110 Ma. For simplicity, and because of the relatively low sensitivity of the other paleoenvironmental variables to seawater Mg/Ca estimates, we use interpreted seawater Mg/Ca estimates given by these authors instead of developing formal models for each Mg/Ca proxy system. Because uncertainty exists in the form of the partitioning function between seawater and echinoderm carbonate, our dataset includes both the original estimates from Dickson (2002) and the reinterpreted estimates of Hasiuk and Lohmann (2010). The uncertainty associated with each estimate was approximated from the primary publication, and ranged from 0.03 mmol/mol for modern seawater to ~0.5 mmol/mol for some of the proxy estimates (1 σ, see data and code available at <https://github.com/SPATIAL-Lab/JPI_marine>).

Foram Mg/Ca and δ18O data were compiled from three Ocean Drilling Program (ODP) sites: site 806, Ontong Java Plateau (Bickert et al. 1993, Lear et al. 2003, Lear et al. 2015); site 1123, Chatham Rise (Elderfield et al. 2012), and site U1385, Iberian Margin (Birner et al. 2016). Data from site 806 constitute a low-resolution record from ~18 Ma to present, with an average sampling resolution of 1 sample per 240 and 180 kyr for Mg/Ca and δ18O, respectively, prior to 800 ka. Mg/Ca measurements were made on the benthic infaunal foraminifer *Oridorsalis umbonatus*. δ18O data represent a mix of benthic species, predominantly *Cibicidoides* spp, and we adopt the inter-species offset corrections applied by Lear et al. (2003) to correct data from other species to *Cibicidoides*-equivalent values. For the other two sites, data were extracted for the overlapping period of record (1.32 – 1.23 Ma), and comprise a set of higher-resolution records (sampling resolution between 1 per 110 and 1 per 1,700 years) spanning two glacial/interglacial cycles. Mg/Ca measurements were made on tests of the infaunal foraminifer *Uvigerina* spp at both sites, and δ18O data are from either *U.* spp (site 1123) or *Cibicidoides wuellerstorfi* (site U1385). Variance in the foram data, e.g., due to analytical effects and sample heterogeneity, was not estimated independently but rather treated as a model parameter and conditioned on the calibration and proxy data.

Calibration datasets were compiled to constrain the Mg/Ca and δ18O proxy system models. Mg/Ca calibration data for *O. umbonatus* are from the compilation of Lear et al. (2015), and include both modern coretop samples and samples from Paleocene and Eocene sediments of ODP site 690B. Data from site 690B include an adjustment for differences in cleaning procedures used for those samples (Lear et al. 2015). For *U.* spp we use the compilation of Elderfield et al. (2010) which consists of core-top samples exclusively. Each Mg/Ca datum is accompanied by a bottom water temperature (BWT) estimate based on syntheses of observational data (modern) or δ18O thermometry (paleo), the latter assuming ice-free conditions. We adopt both sets of estimates directly, applying a normally distributed uncertainty to the BWT values with a standard deviation of 0.2 and 1 °C for the modern and paleo data, respectively, to approximate the different quality of these estimates. For δ18O we used the compilation of Marchitto et al. (2014) including new and published coretop data for the genera *Cibicidoides* and *Uvigerina* (Shackleton 1974, Grossman & Ku 1986, Keigwin 1998). Estimates of BWT and δ18O of seawater from the original authors were adopted with an estimated uncertainty of 0.2 °C (1 σ) for BWT; as for Mg/Ca we do not attempt to constrain the uncertainty in the relationship between temperature and δ18O fractionation between seawater and calcite directly, but treat it as a model parameter.

The age of each pre-modern datum was taken from the primary source. Age uncertainties, where known, can be incorporated in the JPI analysis framework by treating ages as random variables rather than as fixed values and/or including proxy model components representing processes governing the time-integration of observations. To reduce the complexity of the current analysis we do not include such a treatment for any of the datasets. We note in the discussion several cases where including age uncertainty would produce a more robust analysis.

## Proxy models

The proxy system models used here are comprised of simple, empirically constrained transfer functions relating proxy data to contemporaneous environmental variables, and as such can be considered “observation models” in the terminology of Evans et al. (2013). The simplest model is that for seawater Mg/Ca proxy data, where, as noted above, we consider the interpreted data directly, giving:

. (1)

Here *MgCaswp*(*i*) is the *i*th proxy estimate, *N* represents the normal distribution, *MgCasw* is the paleo-seawater Mg/Ca value, and *tswp* and *σswp* are the age estimate and uncertainty associated with a proxy estimate.

We model foram Mg/Ca (*MgCaf*, including both calibration and proxy data) as a function of seawater chemistry and bottom water temperature:

, (2)

where *α1-3* and *τMgCaf* are the parameters and precision (1/σ2) associated with the transfer function, respectively, and other parameters are analogous to equation 1. In the absence of theoretical constraints, we assign normally distributed priors to the *α* parameters based on Bayesian regression of the expression for the mean in equation 2 against the calibration datasets. For *Oridorsalis* we assume paleo-seawater Mg/Ca of 1.5 mmol/mol in the Paleocene and Eocene for these initial estimates, and the prior estimates are *α1* ~ N[1.5, σ = 0.1], *α2* ~ N[0.1, σ = 0.01], and *α3* ~ N[-0.02, σ = 0.03]. For *Uvigerina* these distributions are *α1* ~ N[1.02, σ = 0.1] and *α2* ~ N[0.07, σ = 0.01], and the prior estimated for *α3* from the *Oridorsalis* data set was used because no calibration data were used that represented non-modern *MgCasw*. For both genera, the prior estimate on the precision of the foram Mg/Ca model, *τMgCaf*, is the gamma distribution Γ[shape = 2, rate = 1/30], which approximates the precision of the independent regressions.

Foram calibration and proxy δ18O values (*δ18Of*) are modeled similarly, with:

. (3)

Here *δ18Osw* is the modeled seawater isotope composition and *β1-3* are the transfer function coefficients. In this analysis we treat the scale conversion factor between the SMOW and PDB reference scales (Shackleton 1974) as implicit in the transfer function intercept term (*β1*), which is relevant only in comparing our posterior parameter estimates to other work. Prior estimates of the model parameters were obtained and specified as for Mg/Ca; these are *β1* ~ N[3.32, σ = 0.02], *β2* ~ N[-0.237, σ = 0.01], *β3* ~ N[0.001, σ = 0.0005] for *Cibicidoides* and *β1* ~ N[4.05, σ = 0.06], *β2* ~ N[-0.215, σ = 0.02], *β3* ~ N[-0.001, σ = 0.001] for *Uvigerina*. Because the amplitude of high-frequency (i.e. below the resolution of our model) *δ18Osw*variance in the long record from site 806 increased substantially with the onset of modern, 100 kyr glacial cycles, we modeled *τδ18Of*(*i*) separately for proxy data younger than 800 ka (prior on *τδ18Of* ~ Γ[6, 1]) and for all other proxy and calibration data (prior on *τδ18Of* ~ Γ[3, 1/30]). The former estimate is based on the observed proxy variance since 800 ka, whereas the former approximates the precision of the calibration relationships.

## Environmental models

The paleoenvironmental variables driving the observed proxy signals are represented as time series using a correlated random walk model. This parameterization is desirable in that it is minimally prescriptive (i.e. no preferred state or pattern of change is proscribed) but allows incorporation of constraints on (and extraction of inference about) two basic characteristics of the underlying environmental systems – namely their rates and directedness. The correlated random walk for variable *Y* is expressed as:

, (4)

where:

. (5)

In short, the variable follows a random walk in which the next value in the time series is a function only of the current value and a normally distributed error term *ϵY*, which has a temporal autocorrelation of *φY* and precision *τy*. This gives three independent parameters, *φY*, *τy*, and an initial value of *Y* at the beginning of the time series.

For seawater Mg/Ca, which is thought to evolve gradually in response to long-term tectonic and biogeochemical forcing (Wilkinson & Algeo 1989), we simulate the time series at 1 Myr steps from 110 Ma to present. Although the foram proxy data used here span only the interval from ~18 Ma to present, extending the seawater model over this longer temporal domain was necessary in order to generate a stable time series, conditioned on sparse seawater Mg/Ca proxy data, that spanned both the proxy records and the Paleogene-aged Mg/Ca proxy calibration data. Given that the modeled quantity is a ratio, we treat the error term in this time series model as a proportion, such that the change in *MgCasw* between two time steps is *MgCasw*(*t-1*) \* *ϵMgCasw*. We adopt priors that imply relatively slow change and strong temporal trends (*φMgCasw* is given by a uniform distribution, U[0.9, 1]; *τMgCasw* ~ Γ[100, 0.01]). We use a weak prior on the initial state of *MgCasw* at 110 Ma, U[0.75, 2], consistent with independent interpretations of Cretaceous proxy data (Coggon et al. 2010).

We select the bounds, resolution, and prior distributions for the bottom water temperature and δ18O time series models based on the properties of each record. For site 806 we use a time step of 50 kyr from 18 Ma to present, adequate to allow the time series model to adapt across the range of surpa-orbital timescales represented in the sample distribution. Prior estimates of the error term parameters were chosen to allow sampling across a range of weak to moderate autocorrelation states and error variances that were consistent with first-order interpretations of the proxy data (*φ* ~ U[0, 0.4] for both proxies; *τBWT* ~ Γ[20, 2]; *τδ18Osw* ~ Γ[10, 0.2]). We use weakly informative uniform priors for initial values at 18 Ma (*BWT*(-18) ~ U[3, 8], *δ18Osw*(-18) ~ U[-1, 1]). For the higher-resolution Pleistocene records, we bound the models between 1.32 and 1.235 Ma and adopt a time step of 1 kyr, accommodating orbital-scale changes in the parameter values. We adopt the same prior distributions for *τBWT* and *τδ18Osw* as in the long-term model, but use a broader prior on *φ* (U[0, 0.8] for both environmental variables) based on the expectation that temporal autocorrelation in temperature and seawater δ18O trends will be stronger at timescales of 1 kyr than at 50 kyr.

## Model inversion

The equations above were coded in the BUGS language (Lunn et al. 2012) and Markov Chain Mote Carlo was used to generate samples from the posterior distribution of all model parameters conditioned on the proxy and calibration datasets. The analysis was implemented in R version 3.5.1 (R Core Team 2018) using the rjags (Plummer 2018) and R2jags (Su & Yajima 2015) packages. Three chains were run in parallel. Convergence was assessed visually via trace plots and with reference to the Gelman and Rubin convergence factor (Rhat; Gelman & Rubin 1992) and effective sample sizes reported by rjags.

For the site 806 analysis, chains were run to a length of 1.5e6 samples with a burn-in period of 10e5 samples and thinning to retain a total of 5,000 posterior samples. All parameters showed strong convergence (Rhat << 1.05, effective sample size > 3,500) with the exception of some parts of the seawater Mg/Ca time series and the initialization period of the *BWT* and *δ18Osw* time series (i.e. prior to the first proxy observation). The long run and burn-in periods were dictated by the *MgCasw* time series values, which exhibited very strong autocorrelation as a result of their ‘stiff’ time series behavior and weak data constraints. Qualitative assessment showed no perceptible covariance between seawater MgCa and other parameters in the posterior samples, nor was the posterior distribution obtained from this inversion substantially different from one produced by inverting the MgCa proxy model alone (which was run to an effective sample size >4,000 beyond the initialization period); as a result, we do not believe the weaker sampling from the *MgCasw* posterior has a significant impact on our results or interpretations. The entire analysis took approximately 22 hours running on three cores of a Windows desktop computer.

For the Pleistocene data we conducted three different analyses, the first two inverting data from each site independently and the third inverting both records together. Because of the short time interval covered by these analyses we did not model the seawater Mg/Ca explicitly, but estimated paleo-seawater Mg/Ca values, where needed, from the posterior distributions of an independent inversion of the sweater Mg/Ca proxy data. Chains were run to 5e5 and 7.5e5 samples for the single- and multi-site analyses, respectively, using a burn in period of 1e4 samples and thinning to retain 5,000 posterior samples. All parameters showed strong convergence (Rhat << 1.05) and effective samples sizes were >4,000 for most parameters and >2,000 for all parameters excluding the initialization period of the time series (i.e. prior to the first observation). Total analysis time ranged from <1 hour (site 1123) to ~4 hours running three chains in parallel.

# Results and Discussion

## JPI paleoenvironmental reconstructions

The paleoenvironmental reconstructions obtained by applying JPI to the site 806 data are similar, to first order, to the reconstructions from Lear et al. (2015; hereafter L15) on which our analysis was modeled (Fig. 2). Our reconstruction shows strong support for ~2 °C of bottom-water warming at site 806 during the mid-Miocene Climatic Optimum (centered here on ~15.5 Ma), and although abrupt cooling followed this event, water temperatures warmed again by ~1 °C into the late Miocene. A strong and sustained multi-Myr cooling trend began at the site just prior to 5 Ma and persisted throughout the remainder of the record. Our median temperature estimates are most similar to those obtained by L15 using their “NBB” calibrations, which was based on the same compilation of calibration data used here. Our estimates of seawater Mg/Ca match those obtained by L15 using polynomial curve-fitting throughout most of the common period of analysis. Prior to 40 Ma our estimates diverge, reflecting the incorporating additional data from the Cretaceous in our analysis, but this difference does not impact other interpretations given that L15 did not use the curve-fit estimates from this part of the record in their analysis. 95% credible intervals estimated from JPI average 2.8 °C and 0.8 ‰, which is similar to the uncertainty bounds provided by L15 based on iterative estimation using different calibration functions. The width of the JPI CIs varies subtly across the time series, with somewhat narrower intervals during periods of dense sampling, e.g., in the late Pleistocene.

JPI paleoenvironmental time series for the single- and multi-site analysis of the Pleistocene data were nearly identical, with slightly broader credible intervals for both parameters (BWT and δ18Osw) and sites in the single-site analyses (not shown). The multi-site analysis showed coherent and slightly phase-shifted patterns of BWT variation across two glacial-interglacial cycles at the two sites, with the amplitude of variation being approximately twice as high and median BWT estimates 2 to 5 °C warmer at U1385 (Fig. 4A). In contrast, reconstructed δ18Osw values show greater glacial-scale variability at site 1123, with abrupt decreases of ~0.5‰ accompanying both glacial terminations, whereas the seawater δ18O time series reconstructed for site U1385 shows little response to the termination at ~1.295 Ma but also exhibits high-frequency variability not seen at 1123 (Fig. 4B). Both reconstructions are similar in nature to those provided by the original authors, though the absolute temperatures from the JPI analysis are ~1.5 °C warmer than those plotted by Birner et al. (2016). The origin of this difference is not immediately apparent, as our temperatures match those of Elderfield (2012) well and Birner et al. indicate that they have used the same proxy calibration as those authors. Neither original study presents quantitative uncertainty bounds on individual paleotemperature or δ18Osw estimates, but both provide bottom-line estimates of methodological uncertainty based on propagation of errors. The average width of our 95% CIs is actually somewhat narrower than the 2σ values of the original papers, and the JPI CIs are notably narrower for the U1385 record (2.3 °C, 0.6‰) than for 1123 (2.9 °C, 0.7‰; all estimates from the multi-site analysis).

## Time series properties

One visually striking difference between the JPI and L15 reconstructions is the higher BWT and δ18Osw variability implied by L15 (e.g., compare our Fig. 2 with their Fig. 7). As is common in traditional proxy interpretations, the L15 paleoenvironmental record treats each individual datum as an estimate of an independent environmental state, giving a reconstruction centered on ‘best estimates’ derived from each data point. In reality, however, the environmental states giving rise to the proxy data are not independent if autocorrelation exists at the resolution at which the time series is sampled. For BWT and δ18Osw this is true over a broad spectrum of resolutions including those considered here; for example values of these parameters are known to vary systematically over millions of years due to long-term fluctuations in Neogene climate and ice volume (REFS) and over tens to hundreds of thousands of years due to orbital forcing (REFS). This is commonly implicitly acknowledged in the presentation of traditional proxy reconstructions by including a smoothed representation of the record, obtained using a (usually arbitrary) smoothing filter (e.g., Elderfield et al. 2012).

JPI, in contrast, explicitly considers temporal autocorrelation of the underlying environmental variables, treating each proxy observation as a sample arising from one or more underlying, autocorrelated environmental time series. The properties of the time series themselves, rather than being assumed, are estimated using the proxy models and the data, meaning that the record produced is optimized to reflect the actual information content of the data. For very certain proxy models or densely distributed data that record high-frequency variability the reconstructed time series will express short-term changes in the environment, whereas reconstructions based on uncertain models or smooth or sparsely-sampled data will tend toward greater smoothing and reflect the actual information content of the proxies with respect to the longer-term evolution of the mean state of the system. This is nicely illustrated by comparison of JPI δ18Osw reconstructions for sites 1123 and U1385, where the sample density of the U1385 proxy record is approximately 15 times greater and the resultant time series reconstruction expresses much stronger variability at millennial timescales (Fig. 4B).

Another advantage of embedding time series models in JPI is that it offers an explicit framework for integration of differently-sampled proxy records. In most of the studies reviewed here foram δ18O values are more densely sampled than Mg/Ca. In a traditional, piece-wise interpretation of these proxy data, *δ18Osw* can only be estimated if paired oxygen and Mg/Ca data are available for a given core level. Thus, if Mg/Ca data are missing at a level either this value must be estimated, usually through linear interpolation, or the foram δ18O data excluded from the analysis. JPI eliminates the need to exclude or selectively interpolate data by linking all proxy measurements to a common set of continuous time series. The temporal interpolation required to integrate data sampled at different times is conducted for each environmental variable (which are in reality the quantities that are related in time), rather than for the proxy values themselves, as an explicit component of the analysis. One note of caution is warranted here: despite the advantages conferred by JPI, potential for artefacts to emerge from the integration of datasets with very different sampling densities remains. For example, the high-frequency variability in estimated seawater δ18O at site U1385 (Fig. 4B) stems from high-frequency variance in the over-sampled *δ18Of* record at this site, but without *MgCaf* at similar resolution it is impossible to determine whether the isotopic proxy record variance truly reflects millennial-scale changes in seawater δ18O or instead is driven by un-documented, high-frequency BWT variation.

A final outgrowth of the integration of proxy system and paleoenvironmental time series models via JPI is that the method provides quantitative uncertainty bounds that are linked to and reflect the stratigraphic distribution and density of proxy information. Because environmental parameters are modeled as continuous time series, estimates of central tendency and dispersion (e.g., credible intervals) are obtained throughout the reconstruction period. For time steps in which no observational data are available, the dispersion of posterior estimates increases consistent with the properties of the time series model (e.g., between ~55 and 80 Ma in the seawater Mg/Ca model; Fig. 3), providing quantitative estimates of the constraints provided by the data within these intervals. Moreover, because temporal autocorrelation of the environmental variables is considered, densely sampled data, even where samples are taken at different stratigraphic levels, places additive constraints on the true value of the environmental state. As a result, credible intervals in the posterior distribution adjust to reflect both the density and the strength of the proxy constraints. The result can be seen, for example, in the broader 95% CIs for the sparsely-sampled portion of the site 806 record between ~7 and 10 Ma (Fig. 2) or in the contrasting width of the CIs for the two Pleistocene sites (Fig. 4).

## Model properties

Bayesian inversion has previously been used to estimate proxy model parameter values in situations where these are poorly constrained (Tolwinski-Ward et al. 2013), and the joint inversion of proxy and environmental time series models performed in JPI can similarly be used to provide constraints on parameter values for all model components. Because the proxy system models used here were simple statistical formulations, and the calibration data themselves used to generate prior estimate on model parameters, the mean posterior estimates are generally quite similar to the priors (Fig. 5). The only notable exception is *β3* the second-order parameter in the *δ18Of* model, for which the posterior mean is shifted subtly toward zero (Fig. 5G). In general our prior estimates of parameter variance was slightly inflated to ensure that we did not over-constrain these values, and the posteriors show sharpening of the distributions for most parameters. This is particularly true for the proxy model precision (or variance) terms, where the posterior distributions are much more strongly constrained than the priors (Figs. 5D and H).

In all cases these refinements reflect a combination of the constraints offered by the calibration and down-core proxy data. Although at first consideration the relevance of the latter to calibrating proxy model parameters might not be apparent, keep in mind that the proxy model must not only be consistent with the calibration data but also explain the observed proxy data as a function of the underlying environmental variables. As a result, for a given set of proxy data and environmental time series model properties only a certain subset of proxy model parameter values will be plausible. This may be most apparent for the proxy model precision: this model parameter here explains not only the “noise” within the model calibration data, but also “noise” in individual proxy observations around the value expected for a given environmental state due to a range of factors (e.g., temporal variation in the environment at time scales below the time series model time step, biological or random variation in the environment-proxy relationship). Our analysis suggests that prior to the mid-Pleistocene transition, the proxy model variance implied by the full JPI inversion is similar to that estimated from the calibration data alone (solid curves in Figs. 5D and H), with slightly higher variance required by the *δ18Of* data. The site 806 *δ18Of* record, however, is much more densely sampled after 800 ka, and the combination of higher *δ18Osw* variability and dense sampling that more strongly records the variability following this transition requires a much higher proxy model variance (dashed lines in Fig. 5H). The proxy calibration data offer no constraints on this value, rather the JPI posterior estimates the parameter value to reconcile the environmental time series (representing the longer-term evolution of the mean system state) with the variance expressed in the proxy observations themselves.

Because the JPI analysis involves sampling of all model parameters simultaneously, it also can identify and account for correlation among parameters. The proxy system model parameter estimates for site 806 provide a clear example (Fig. 6). The posterior distributions show strong correlation between the seawater Mg/Ca sensitivity term (*α3*) and both the intercept and sensitivity terms (*α1* and *α2*) in the *MgCaf* model and between the first- and second-order terms (*β2* and *β3*) in the *δ18Of* model. This is not at all surprising: in all cases these terms are interactive and for a given estimate of the model calibration a change in one will generally be offset by a change in the other. Accounting for this covariance is important in assessing the uncertainty of proxy reconstructions, however, and may in part account for the more optimistic uncertainty estimates obtained here relative to those from published work based on propagation of errors assuming independence of parameters, in that the latter approach will ‘double-count’ uncertainty associated with correlated parameters.

JPI also provides posterior estimates on the environmental time series model parameters, and these distributions can provide information complimentary to the reconstructed time series themselves. Comparing prior and posterior estimates at all three study sites (Fig. 7), the analysis provides strong posterior constraints on the error autocorrelation (i.e. directedness of change) and the error variance (i.e. magnitude of change among time steps) for *δ18Osw*, but posterior estimates of *BWT* error variance are only subtly different from the priors (middle column). Interestingly, the error variance estimates are quite similar for both environmental variables at all sites despite the ~2 order of magnitude difference in the resolution of the time series models and data density, suggesting scale-independence of short-term rates of change in these systems.

In contrast, the error autocorrelation term, which reflects the directedness of environmental change across multiple model time steps, shows substantial variation among the data sets (Fig. 7, left column). The lowest posterior values were obtained for the long record at site 806, consistent with the assumption that inertia would be weaker for these variables at the longer time scales (i.e. 50 kyr time steps) reflected in this analysis. Across all scales, posterior distributions for autocorrelation were skewed lower for *δ18Osw* than for *BWT*. Although this may in part reflect the greater expression of short-term variance in more densely sampled *δ18O­f* records, the result holds at site 1123, where the sample distributions for *MgCaf* and *δ18O­f* are identical, implying that changes in *δ18Osw* are generally more chaotic than those of *BWT*. The strongest error autocorrelation is inferred for *BWT* at site U1385, where the data strongly support highly coherent, high-amplitude cyclic variation in *BWT* across the two glacial cycles sampled. In contrast, *δ18Osw* variation estimated at this site is only weakly directional and is features strong, chaotic, millennial-scale variability, reflected in a much lower posterior estimate for error autocorrelation (Fig. 7D).

## Derivative analyses

In this final section, we explore additional examples of how JPI results might be used to support inference or hypothesis testing in paleoenvironmental reconstruction. First, because JPI results provide integrated, self-consistent estimates of multiple environmental variables, they can be used to identify and characterize multivariate modes of environmental change in Earth’s past. Results from the site 806 analysis, for example, demonstrate non-linear coupling between changes in *BWT* and *δ18Osw* since the mid-Miocene (Fig. 8). These patterns, including limited coupling between *δ18Osw* and BWT change prior to ~5 Ma and strong bottom water cooling accompanied by a modest *δ18Osw* decrease into the Pleistocene, were previously noted by L15. What is apparent here, however, is that the proxy reconstructions appear to suggest that the system transitioned between at least three semi-stable states during this time. Jumps between a mid-Miocene warm, low-*δ18Osw* state, late Miocene warm, high- *δ18Osw* state, and Plio-Pleistocene cool state were in each case relatively abrupt, with the system spending the majority of the reconstruction period within, rather than between, states. Patterns of short-term correlation between *BWT* and *δ18Osw* appear to vary among these states, as well, with positive correlation between these variables dominating the first two states and the classical negative correlation characteristic of coupled temperature and ice volume changes only expressed during the final one (Fig. 8, dots).

JPI results also provide a sound basis for testing hypotheses of change within or between proxy records. As with the evaluation of reconstruction uncertainty, the important concept here is that multiple parameter values within individual samples of the posterior are not independent, but instead reflect the covariance of parameters as constrained by the data and models used. Consider the case where we want to assess the magnitude and amount of change in site 806 bottom water temperature relative to the modern (core top) value. Traditionally, we might take a central estimate of the modern value, e.g., the median shown by the left terminus of the red line in Fig. 9A, and compare it with the reconstructed distribution of values at one or more points in the past to ask whether it is or is not consistent with that distribution. This implicitly assumes that the true environmental parameter values at the two times (modern and some past time) are independent of each other, and gives the set of probabilities shown in dotted blue in Fig. 9A. In reality, however, the *BWT* values at the two times are not independent, as discussed above. We can account for this by framing the analysis in terms of change within individual posterior samples (Fig. 9A, solid blue line).

The resulting estimates show interesting, if subtle, contrasts with the traditional approach. At short time lags (less than ~400 kyr) the within-sample comparison actually implies somewhat higher probability of significant change. This reflects the influence of error autocorrelation in the time series model: within an individual posterior sample, directional change is likely to persist over multiple time steps, meaning that the ‘signal to noise ratio’ over short periods is higher if estimated based on within-sample vs. between-sample change. Beyond this time frame, however, the relationship between methods inverts, and the traditional method assuming independence gives exaggerated estimates of the significance of change. Beyond the scale of significant time series error autocorrelation, the variance of change estimated from the within-sample comparison is actually slightly greater than that estimated between samples, reflecting the fact that some possible BWT trajectories within the posterior ‘wander’ across the distribution of possible values over time, increasing the dispersion of the change estimates. The net result is that in this case, using a one-sided 95% credible interval threshold (equivalent to *p*=0.05), one would estimate that site 806 bottom water temperatures diverged from modern some 100 – 200 kyr earlier using the traditional approach than with the more appropriate within-sample analysis.

Our final example involves cross-site comparison. Here, we similarly ask whether seawater δ18O values were significantly different at sites 1123 and U1385 throughout the period of study based on comparisons of the posteriors from the multi-site analysis or the two single-site JPI analyses (Fig. 9B). The results show that the assessment which assumes independence of estimates at the two sites (the latter one) uniformly under-estimates the significance of the difference between the sites. This can be explained intuitively in terms of the impact of other model parameters on posterior estimates of *δ18Osw* values at both sites – in a given sample from the posterior of the multi-site analysis, if one of the *δ18Of* proxy system model parameters deviates from the central estimate, for example, it will similarly impact the seawater isotope reconstructions at both sites. As a result, the variance of the between-site differences is reduced in the comparison based on the multi-site analysis, producing stronger results in the post-hoc tests of difference. In this example the choice of approach would have little impact on inferences drawn based on the 95% credible interval, but at the 99% level several parts of the time series would be considered different using the multi-site comparison and not different with the traditional approach (Fig. 9B).

# Conclusion

Traditional approaches to proxy interpretation suffer from broad and poorly characterized uncertainty and potential biases related to the sensitivity of proxies to multiple environmental factors. Proxy system modeling and multi-proxy reconstruction provide partial solutions to these issues, but a robust, accessible framework for integrating these two approaches in the development of paleoenvironmental reconstructions has been lacking. Here we propose combining multiple proxy system models and proxy records, along with simple time series models for environmental variables, in a Bayesian inversion framework. This approach is broadly generalizable to any set of proxies for which appropriate forward models can be written, and confers many of the advantages of more complex data assimilation methods that leverage Earth system models (Evans et al. 2013) while remaining independent of the assumptions embedded in these models and flexible enough to be applied over a wide range of systems and time scales. Our simple illustration of the method based on the coupled Mg/Ca and δ18O systems in benthic foraminifera demonstrates this flexibility through applications to two contrasting time scales and both single- and multi-site proxy records. Although this example system, and the nature of the proxy models used here, is relatively simple, the example illustrates how JPI can be easily applied to widely used proxy systems to give improved characterization of uncertainty and interpretation of records. Implementations similar to those demonstrated here could easily and immediately become standard practice in the interpretation of most paleoenvironmental proxy data. As the underlying proxy system models mature, JPI-based interpretations can be revised and refined to incorporate new understanding and/or leverage additional proxy types, minimizing, but also accurately representing, bias and uncertainty in our paleoenvironmental reconstructions.

# References

Affek HP, Matthews A, Ayalon A, Bar-Matthews M, Burstyn Y, et al. 2014. Accounting for kinetic isotope effects in Soreq Cave (Israel) speleothems. Geochimica et Cosmochimica Acta 143: 303-18

Bickert T, Berger W, Burke S, Schmidt H, Wefer G. 1993. Late Quaternary stable isotope record of benthic foraminifers: Sites 805 and 806, Ontong Java Plateau 1. Proceedings of the Ocean Drilling Program, Scientific Results 130: 411-20

Birner B, Hodell DA, Tzedakis PC, Skinner LC. 2016. Similar millennial climate variability on the Iberian margin during two early Pleistocene glacials and MIS 3. Paleoceanography 31: 203-17

Coggon RM, Teagle DAH, Smith-Duque CE, Alt JC, Cooper MJ. 2010. Reconstructing past seawater Mg/Ca and Sr/Ca from mid-ocean ridge flank calcium carbonate veins. Science 327: 1114-7

Conte MH, Sicre M-A, Rühlemann C, Weber JC, Schulte S, et al. 2006. Global temperature calibration of the alkenone unsaturation index (UK′37) in surface waters and comparison with surface sediments. Geochemistry, Geophysics, Geosystems 7

Conte MH, Thompson A, Lesley D, Harris RP. 1998. Genetic and physiological influences on the alkenone/alkenoate versus growth temperature relationship in Emiliania huxleyi and Gephyrocapsa oceanica. Geochimica et Cosmochimica Acta 62: 51-68

de Villiers S, Nelson BK. 1999. Detection of Low-Temperature Hydrothermal Fluxes by Seawater Mg and Ca Anomalies. Science 285: 721

Dee S, Emile-Geay J, Evans MN, Allam A, Steig EJ, Thompson DM. 2015. PRYSM: An open-source framework for PRoxY System Modeling, with applications to oxygen-isotope systems. Journal of Advances in Modeling Earth Systems 7: 1220-47

Dickson JAD. 2002. Fossil Echinoderms As Monitor of the Mg/Ca Ratio of Phanerozoic Oceans. Science 298: 1222

Elderfield H, Ferretti P, Greaves M, Crowhurst S, McCave IN, et al. 2012. Evolution of Ocean Temperature and Ice Volume Through the Mid-Pleistocene Climate Transition. Science 337: 704

Elderfield H, Greaves M, Barker S, Hall IR, Tripati A, et al. 2010. A record of bottom water temperature and seawater δ18O for the Southern Ocean over the past 440kyr based on Mg/Ca of benthic foraminiferal Uvigerina spp. Quaternary Science Reviews 29: 160-9

Evans MN, Tolwinski-Ward SE, Thompson DM, Anchukaitis KJ. 2013. Applications of proxy system modeling in high resolution paleoclimatology. Quaternary Science Reviews 76: 16-28

Fricke HC, Clyde WC, O'Neil JR, Gingerich PD. 1998. Evidence for rapid climate change in North America during the latest Paleocene thermal maximum; oxygen isotope compositions of biogenic phosphate from the Bighorn Basin (Wyoming). Earth and Planetary Science Letters 160: 193-208

Gelman A, Rubin DB. 1992. Inference from iterative simulation using multiple sequences. Statistical Science 7: 457-72

Grauel A-L, Leider A, Goudeau M-LS, Müller IA, Bernasconi SM, et al. 2013. What do SST proxies really tell us? A high-resolution multiproxy (UK′37, TEXH86 and foraminifera δ18O) study in the Gulf of Taranto, central Mediterranean Sea. Quaternary Science Reviews 73: 115-31

Greenwood DR, Wilf P, Wing SL, Christophel DC. 2004. Paleotemperature estimation using leaf-margin analysis: Is Australia different? Palaios 19: 129-42

Grossman EL, Ku TL. 1986. Oxygen and carbon isotope fractionation in biogenic aragonite: temperature effects. Chemical Geology 59: 59-74

Hasiuk FJ, Lohmann KC. 2010. Application of calcite Mg partitioning functions to the reconstruction of paleocean Mg/Ca. Geochimica et Cosmochimica Acta 74: 6751-63

Keating-Bitonti CR, Ivany LC, Affek HP, Douglas P, Samson SD. 2011. Warm, not super-hot, temperatures in the early Eocene subtropics. Geology 39: 771-4

Keigwin LD. 1998. Glacial-age hydrography of the far northwest Pacific Ocean. Paleoceanography 13: 323-39

Lear CH, Coxall HK, Foster GL, Lunt DJ, Mawbey EM, et al. 2015. Neogene ice volume and ocean temperatures: Insights from infaunal foraminiferal Mg/Ca paleothermometry. Paleoceanography 30: 1437-54

Lear CH, Elderfield H, Wilson PA. 2000. Cenozoic deep-sea temperatures and global ice volumes from Mg/Ca in benthic foraminiferal calcite. Science 287: 269-87

Lear CH, Rosenthal Y, Wright JD. 2003. The closing of a seaway: ocean water masses and global climate change. Earth and Planetary Science Letters 210: 425-36

Lowenstein TK, Timofeeff MN, Brennan ST, Hardie LA, Demicco RV. 2001. Oscillations in Phanerozoic seawater chemistry: Evidence from fluid inclusions. Science 294: 1086-8

Lunn D, Jackson C, Best N, Thomas A, Spiegelhalter D. 2012. The BUGS Book: A Practical Introduction to Bayesian Analysis: CRC Press / Chapman and Hall

Marchitto TM, Curry WB, Lynch-Stieglitz J, Bryan SP, Cobb KM, Lund DC. 2014. Improved oxygen isotope temperature calibrations for cosmopolitan benthic foraminifera. Geochimica et Cosmochimica Acta 130: 1-11

Passey BH, Levin NE, Cerling TE, Brown FH, Eiler JM. 2010. High-temperature environments of human evolution in East Africa based on bond ordering in paleosol carbonates. PNAS 107: 11245-9

Plummer M. 2018. rjags: Bayesian graphical models using MCMC. R package version 4-8. <https://CRAN.R-project.org/package=rjags>.

R Core Team. 2018. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>

Royer DL, Wilf P, Janesko DA, Kowalski EA, Dilcher DL. 2005. Correlations of climate and plant ecology to leaf size and shape: Potential proxies for the fossil record. American Journal of Botany 92: 1141-51

Saenger C, Affek HP, Felis T, Thiagarajan N, Lough JM, Holcomb M. 2012. Carbonate clumped isotope variability in shallow water corals: Temperature dependence and growth-related vital effects. Geochimica et Cosmochimica Acta 99: 224-42

Shackleton NJ. 1974. Attainment of isotopic equilibrium between ocean water and the benthonic foraminifera genus Uvigerina: isotopic changes in the ocean during the last glacial. Colloques Internationaux du C.N.R.S 219: 203-9

Stoll HM, Müller W, Prieto M. 2012. I-STAL, a model for interpretation of Mg/Ca, Sr/Ca and Ba/Ca variations in speleothems and its forward and inverse application on seasonal to millennial scales. Geochemistry, Geophysics, Geosystems 13

Su Y-S, Yajima M. 2015. R2jags: Using R to Run 'JAGS'. R package version 0.5-7. <https://CRAN.R-project.org/package=R2jags>.

Tolwinski-Ward SE, Anchukaitis KJ, Evans MN. 2013. Bayesian parameter estimation and interpretation for an intermediate model of tree-ring width. Climate of the Past 9: 1481-93

Tolwinski-Ward SE, Evans MN, Hughes MK, Anchukaitis KJ. 2011. An efficient forward model of the climate controls on interannual variation in tree-ring width. Climate Dynamics 36: 2419-39

Wilkinson BH, Algeo TJ. 1989. Sedimentary carbonate record of calcium-magnesium cycling. American Journal of Science 289: 1158-94