https://doi.org/10.5194/egusphere-2024-2183
Preprint. Discussion started: 16 October 2024

PaleoSTeHM v1.0-rc: a modern, scalable spatio-temporal hierarchical modeling framework for paleo-environmental data

Yucheng Lin[1], Robert E. Kopp[1,2], Alexander Reedy[1,2], Matteo Turilli[3,4], Shantenu Jha[2,3,5], and Erica L. Ashe[1]

[1]Department of Earth & Planetary Sciences, Rutgers University, Piscataway, NJ, USA
[2]Rutgers Climate and Energy Institute, Rutgers University, New Brunswick, NJ, USA
[3]Department of Electrical and Computer Engineering, Rutgers University, Piscataway, NJ, USA
[4]Computational Science Initiative, Brookhaven National Laboratory, Upton, NY, USA
[5]Computational Science Department, Princeton Plasma Physics Laboratory, Princeton, NJ, USA

Correspondence: Yucheng Lin (yc.lin@rutgers.edu) and Robert E. Kopp (robert.kopp@rutgers.edu)

Abstract. Geological records of past environmental change provide crucial information for assessing long-term climate variability, non-stationarity, and nonlinearities. However, reconstructing spatio-temporal fields from these records is statistically challenging due to their sparse, indirect, and noisy nature. Here, we present PaleoSTeHM, a scalable and modern framework for spatio-temporal hierarchical modeling of paleo-environmental data. This framework enables the implementation of flexible statistical models that rigorously quantify spatial and temporal variability from geological data with clear distinguishing between measurement and inferential uncertainty from process variability. We illustrate its application by reconstructing temporal and spatio-temporal paleo sea-level changes across multiple locations. Using various modeling and analysis choices, PaleoSTeHM demonstrates the impact of different methods on inference results and computational efficiency. Our results highlight the critical role of model selection in addressing specific paleo-environmental questions, showcasing the PaleoSTeHM framework's potential to enhance the robustness and transparency of paleo-environmental reconstructions.

1 Introduction

As humans push the planet's climate and biosphere increasingly far outside the range of our species' experience, the geological record provides critical out-of-sample data against which to test the models used to project future environmental change. Yet, as an environmental record, the geological data is quite sparse and often noisy and indirect. Reconstructing paleo-environmental fields is thus a critical and challenging statistical task (Tingley et al., 2012).

From an analytical perspective, spatio-temporal hierarchical statistical models provide a natural, conceptually straightforward framework for reconstructing paleo-environment (Ashe et al., 2019; Cressie and Wikle, 2015; Tingley et al., 2012). Hierarchical statistical models, often employed within a Bayesian framework, decompose the various sources of random variation contributing to individual observations into distinct levels, thereby providing a clearer articulation of the assumptions underlying the statistical analysis. It has been increasingly used to model paleo-climate fields from geological proxies (e.g., temperature and precipitation from tree rings and corals). These applications have proven crucial in assessing the robustness

https://doi.org/10.5194/egusphere-2024-2183
Preprint. Discussion started: 16 October 2024

of scientific knowledge of past climate and placing changes in the modern, instrumentally observed period in the context of longer-term variability. For example, they have shown an increasing influence of ice melt and thermal expansion on GMSL since 1860 CE (Walker et al., 2021), that GMSL rise over the 20th century was faster than during any century in at least 3000 years (Kemp et al., 2018; Kopp et al., 2016), and that several early 21st century Arctic summers exhibited warmth unprecedented in at least 600 years (Tingley and Huybers, 2013). There is substantial community demand for the use of such techniques. For example, in the past few years, numerous papers have used temporal or spatiotemporal hierarchical models with Gaussian Process (GP) priors to interpret paleo sea-level proxies (Tan et al., 2023; Khan et al., 2022; Vacchi et al., 2021).

To meet the demand of the paleo-environment community, this paper describes PaleoSTeHM v1.0, which is designed to support the flexible and high-performance implementation of spatiotemporal hierarchical modeling for paleo-environmental data. PaleoSTeHM is a framework built using modern machine learning architecture and in the spirit of open science (e.g., Pollack et al., 2024). It is designed so users can select not only various modeling choices, such as change-point for temporal analysis or GP for spatiotemporal analysis, but also analysis choices, including fully Bayesian, empirical Bayesian, and variational Bayesian analysis (more details in section 2), to investigate different research questions asked, with different types of data and spatiotemporal scales (e.g., local to global, years to millennia) considered.

2 Hierarchical statistical modeling

This section briefly describes a basic theory about hierarchical modeling of the paleo-environment, using paleo sea level as an illustrative example. For more systematic introductions on hierarchical statistical modeling of paleo sea-level and paleoclimate, readers can refer to Ashe et al. (2019) and Tingley et al. (2012).

Based on Bayes' theorem, the conditional probability (see definition in Table 1) of the observed data (y) can be inverted from the conditional probability of unknown parameter(s) or process(es) (θ):

$$p(\theta|y) = \frac{p(y|\theta)p(\theta)}{p(y)} \tag{1}$$

where p denotes 'probability' and $|$ represents 'given'. The likelihood function, $p(y|\theta)$ represents the probability of observing the data y given the parameter(s) or process(es) θ of the model. The prior distribution, $p(\theta)$, captures *a priori* beliefs about the unknown parameter(s) or process(es) before any data is observed. The term $p(y)$, known as the marginal likelihood, is the probability of the observed data averaged over all possible parameters or processes.

Given the observations, the posterior distribution, $p(\theta|y)$, reflects the updated beliefs about the parameter(s) or process(es) after considering the data. Since the marginal likelihood $p(y)$ is often intractable and remains constant for static observations, we use the simplified form of Bayes' theorem: the posterior distribution is proportional to the product of the likelihood and the prior:

$$p(\theta|y) \propto p(y|\theta)p(\theta) \tag{2}$$

https://doi.org/10.5194/egusphere-2024-2183
Preprint. Discussion started: 16 October 2024

Table 1. Definitions of relevant terms in this study. This paper employs terminology based on Ashe et al. (2019).

Term	Meaning
analysis choices	decisions in how to implement a specific model structure
conditional probability	distribution of a random quantity, given another (unknown) random quantity
continuous core	near-continuous records from a single core of sediment or a single coral reef
covariance function	defines prior beliefs about the relationship or correlation between variables
data level	model representations of the relationship between the phenomenon and observed data
errors-in-variable (EIV)	a fully-Bayesian framework that accounts for measurement uncertainty in independent variables
error	difference between a measurement and the true value
Gaussian Process (GP)	a stochastic process that generalizes the multivariate Gaussian distribution to continuous time and space, defined by mean and covariance functions
hierarchical model	a statistical framework that partitions the multiple random effects that lead to individual observations into different levels
hyperparameter	parameter of a prior distribution
isotropy	a property of having identical statistical characteristics in all directions
likelihood	a way to measure how well a statistical model explains observed data
Markov Chain Monte Carlo (MCMC)	techniques used to generate random variables, perform complicated calculations, and simulate complicated distributions through randomsampling in Bayesian models
parameter level	model representations of prior beliefs about parameters used to control the behavior of a statistical and/or physical model at different levels of the hierarchy
physical model	a class of model based on physical principles to describe natural phenomena, typically using mathematical representation of a system or process that uses numbers and equations to describe physical conditions
process level	model representations of the underlying processes responsible for the data generation
posterior distribution	a type of conditional probability that results from updating the prior probability with observational information summarized by the likelihood
prior distribution	the assumed probability distribution before any observational evidence is taken into account, which can be subjective based on a priori knowledge, or uninformative
residuals	the difference between an observed and a modeled or predicted value
smoothness	the characteristic of a process that reflects how smoothly it changes over time or space, often controlled by the kernel's differentiability in Gaussian Process models
space-time separability	a property of processes where the spatial and temporal components of the covariance function are treated as independent, so the covariance is expressed as a product of purely spatial and purely temporal functions
stationarity	a property of processes or signals where their statistical properties, like mean and variance, remain consistent over time or space
uncertainty	model representation of how the model prediction might differ from the true value
white noise	serially uncorrelated random variation (zero mean and finite variance)

https://doi.org/10.5194/egusphere-2024-2183
Preprint. Discussion started: 16 October 2024

where $\propto$ indicates 'is proportional to.'

A basic hierarchical statistical model for paleo sea level distinguishes the fundamental RSL change from both its inherent variability and the observational noise. Hierarchical models achieve probabilistic uncertainty estimation for time series and/or spatial fields by inverting conditional probabilities. Each level of the model quantifies uncertainties independently, necessitating careful evaluation of their respective sources. Generally, three levels are defined: the data level, the process level, and the parameter level:

$$p(f,\theta_s,\theta_d \mid y) \propto \underbrace{p(y \mid f,\theta_d)}_{\text{data model}} \cdot \underbrace{p(f \mid \theta_s)}_{\text{process model}} \cdot \underbrace{p(\theta_d,\theta_s)}_{\text{parameter model}} \tag{3}$$

The data level defines the relationship between the latent (unobserved) RSL process (f) and the observed RSL data (instrumental and/or proxy), y, while accounting for measurement, inferential (e.g., uncertainties arising from converting a proxy's elevation to a distribution of RSL), and dating uncertainties (often inherited from geochronology techniques). This level represents the probability distribution of observing a particular sea-level height at a given age, conditioned on the underlying latent process and the associated measurement, inferential, and dating uncertainties, encapsulated by the data level parameters, θ_d.

The process level distinguishes the underlying phenomenon of interest and its inherent variability, from the noisy observation captured at the data level. This model integrates scientific understanding and associated uncertainties into the estimation of the true RSL process using conditional parameters, θ_s. These parameters may represent unobserved physical model parameters (e.g., Earth's rheology in a glacial isostatic adjustment model; Table 1), statistical model parameters (such as the linear rate in a linear sea-level model), or hyperparameters (parameters of a prior distribution, such as length scale and variance in a model with a Gaussian Process prior). At the foundational level, the parameter model specifies the prior distribution for all unknown parameters, effectively capturing the essential characteristics of both the data and process levels through the unobserved parameters.

In addition to constructing models at the data, process, and parameter levels, often referred to as modeling choices (Ashe et al., 2019), it is essential to choose an appropriate analysis choice for a specific model (Table 1). This involves decisions regarding the implementation of a model structure, such as deterministic methods including least-squares analysis and likelihood maximization (Wilks, 1938; Aitken, 1936), or probabilistic methods like Bayesian analysis (Hastings, 1970). Analysis choices are also integral to addressing how measurement uncertainties, particularly those arising from geochronological techniques, are incorporated and managed within the model. This ensures that the uncertainty is properly quantified and reflected in the final analysis outputs (Ashe et al., 2019). Several factors, including the complexity of the problem, the size and resolution of the data available, the computational resources at hand, and the extent of prior knowledge applicable to the modeling effort, should guide the selection of modeling and analytical choices.

https://doi.org/10.5194/egusphere-2024-2183
Preprint. Discussion started: 16 October 2024

3 Model description

This section provides a comprehensive overview of PaleoSTeHM, detailing its foundational model implementation (section 3.1), the basic architecture for a typical PaleoSTeHM experiment (section 3.2) and the development of PaleoSTeHM modules (sections 3.3, 3.4 and 3.5).

3.1 Model implementation

PaleoSTeHM is designed to be a functionally extensible and high-performing toolkit for modeling paleo data. It is fully open-source and developed under a four-layer structure to maintain a flexible and generic design that is agile to future development (Figure 1). The core toolkit and development reside in L3, comprising modules that integrate existing capabilities from L2, enabling PaleoSTeHM to function across various computing platforms defined in L1. L2 employs Python as the user interface language and utilizes a high-performance machine learning platform as the execution back-end. Atop L3 is the PaleoSTeHM User layer (L4), facilitating interaction with external packages and tools, thereby supporting practical applications and the resolution of scientific inquiries.

PaleoSTeHM modules were built upon Pyro (Bingham et al., 2019), a universal probabilistic programming language supported by PyTorch (Paszke et al., 2017), a popular machine learning library for artificial intelligence applications. Therefore, PaleoSTeHM not only supports probabilistic programming but also leverages an ecosystem of existing machine learning capabilities, including auto-differentiation, GPU acceleration, and modern optimization algorithms. By utilizing such advanced machine learning platforms, we have developed three core PaleoSTeHM modules in the L2 layer: (1) the Modeling Choices module, which incorporates multiple options for data, process, and parameter level modeling; (2) the Gaussian Process kernel module, a sub-module of the Modeling Choices module that supports kernel construction of a process model using GP priors; and (3) the Analysis Choices module, which incorporates multiple methods to consider temporal uncertainty and perform Bayesian inference (Figure 1). These modules provide flexible and efficient options for spatio-temporal hierarchical modeling of paleo-environmental applications.

We anticipate PaleoSTeHM interacting with external packages and/or tools for practical applications and addressing scientific questions on the PaleoSTeHM User layer (L3, Figure 1). Here, 'External Packages' refer to external Python libraries, which provide various pre-processing and post-processing data functions. For example, in PaleoSTeHM tutorials (see section 4), we use Scipy (Virtanen et al., 2020) for interpolation and Matplotlib (Hunter, 2007) for visualization. 'Tools' represent frameworks and services adapted by other developers to integrate PaleoSTeHM capabilities into their toolkits (e.g., Framework for Assessing Changes To Sea level (FACTS); Kopp et al., 2023). Such plug-in implementations will make it easy for users drawn from any of the PaleoSTeHM categories to use, extend, or contribute to core capabilities for various scientific applications.

https://doi.org/10.5194/egusphere-2024-2183
Preprint. Discussion started: 16 October 2024

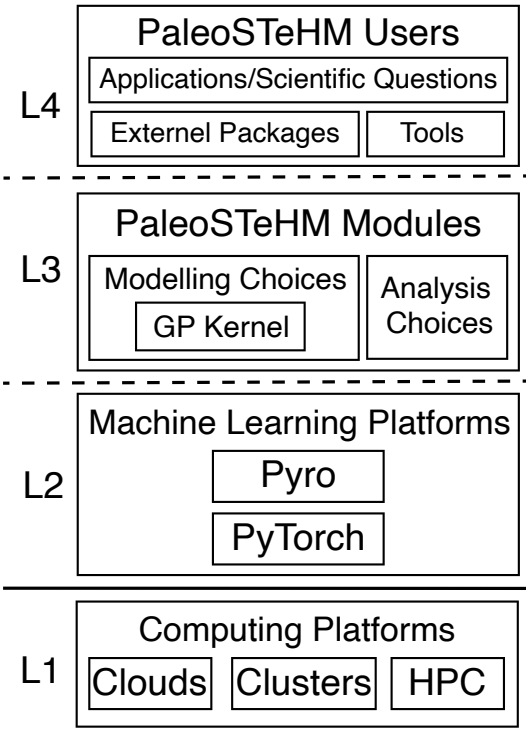

Figure 1. Schematic illustration of the four-layer structure of PaleoSTeHM. L1 specifies various computing platform (clouds, clusters and HPC), L2 comprises machine learning platforms (Pyro and PyTorch), L3 includes PaleoSTeHM modules (Modeling Choices, GP kernel, and Analysis Choices, see Figure 2), and L4 consists of the user layer, facilitating interaction with external packages and tools for practical applications and scientific inquiries.

3.2 PaleoSTeHM experiment architecture

Constructing and training a hierarchical model within PaleoSTeHM consists of five sequential selection steps. Each step corresponds to different modeling and analysis choices introduced in section 2. Typical PaleoSTeHM experiment steps include: (1) selecting data-level models for paleo-environmental data; (2) choosing an appropriate process-level model to describe the latent process; (3) defining prior distributions for each model parameter; (4) selecting a temporal uncertainty treatment method; and (5) choosing a Bayesian inference method (Figure 2). These five steps reflect core functionalities developed within three PaleoSTeHM modules, shown in Figure 1. To support the effective selection of modeling and analytical choices provided by PaleoSTeHM for various paleo-environmental applications. The fundamental theories and example applications for each modeling option will be introduced in section 3.3.

https://doi.org/10.5194/egusphere-2024-2183
Preprint. Discussion started: 16 October 2024

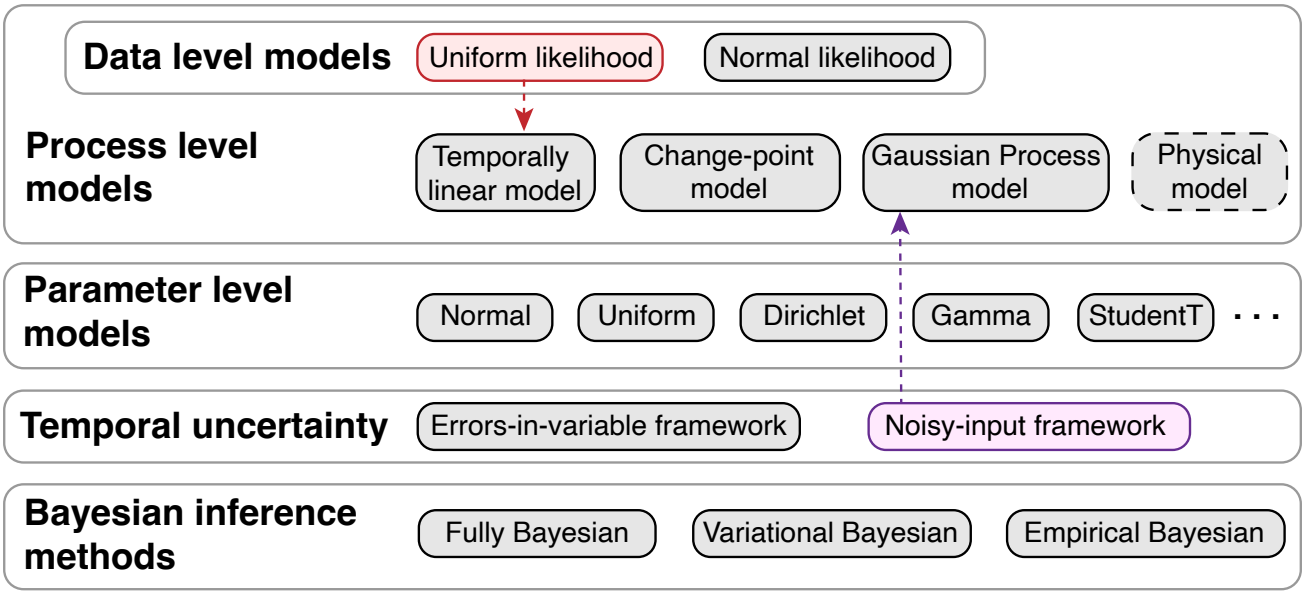

Figure 2. Schematic illustration of the PaleoSTeHM experiment architecture described in this paper. The large boxes represent five steps to build a hierarchical model, and it should be noted that the data-level model is specified within each process-level model in PaleoSTeHM v1.0. The smaller boxes indicate different modeling choices within each step. Grey boxes denote available choices that apply to other grey boxes in different steps. Red and purple boxes represent a specific data level model and temporal uncertainty treatment method corresponding to a specific process level model (temporally linear and Gaussian Process models), as indicated by colored arrows. The dashed grey box (physical model) highlights that no specific physical model is implemented in PaleoSTeHM. Instead, PaleoSTeHM utilizes outputs from other physical models (see section 3.3.2.)

3.3 Modeling Choices module

As mentioned above, spatio-temporal hierarchical modeling experiments begin with selecting an appropriate modeling choice for a specific problem. This module provides multiple commonly used temporal or spatio-temporal modeling choices used in paleo-environmental studies (Figure 2). We will briefly introduce the fundamental theories for each modeling choice and provide examples of paleo-environment studies that adopted such a model. While we do not include a specific section for parameter-level modeling, leveraging the ecosystem of Pyro and Pytorch enables users to easily define prior probabilities for data and process-level model parameters using most of the commonly used probability distributions (Figure 2).

3.3.1 Data level modeling

The data level of a hierarchical statistical model characterizes the relationship between true (unobserved) target signals and uncertain observations due to multiple error sources. For example, in reconstructing past sea-level changes, the data level addresses uncertainties arising from elevation measurements, indicative range, and leveling errors (Khan et al., 2017). Ad-

https://doi.org/10.5194/egusphere-2024-2183
Preprint. Discussion started: 16 October 2024

ditionally, proxy data are often subject to inherent temporal uncertainties stemming from various geochronological methods (e.g., radiocarbon Reimer et al., 2020; Heaton et al., 2020). This hierarchical structure can be formally expressed as:

$$y_i = f(x_i, t_i) + \epsilon_i^y \tag{4}$$

$$t_i = \hat{t}_i + \epsilon_i^t \tag{5}$$

where x_i is the noise-free spatial location of i-th observation, t_i is its true age, $\hat{t}_i$ is the mean observational age, and ϵ_i^t and ϵ_i^y are uncertainties in the age measurement and target signal reconstruction. For paleo-environmental studies, a commonly made assumption is that both ϵ_i^t and ϵ_i^y are multivariate normally distributed with zero mean and heteroscedastic covariance, so ϵ^y can be expressed as:

$$\epsilon^y \sim \mathcal{N}(0, \Sigma_y) \tag{6}$$

$$\Sigma_y = \begin{pmatrix} var(y_1) & cov(y_1, y_2) & \cdots & cov(y_1, y_n) \\ cov(y_2, y_1) & var(y_2) & \cdots & cov(y_2, y_n) \\ \vdots & \vdots & \ddots & \vdots \\ cov(y_n, y_1) & cov(y_n, y_2) & \cdots & var(y_n) \end{pmatrix} \tag{7}$$

where n indicates the number of observations available, $var(\cdot)$ represents the variance of specific data, and $cov(\cdot, \cdot)$ stands for covariance between two data points, which is often assumed to be 0 when all data are assumed to be independently distributed. Whereas strong covariance could potentially exist for some paleo-environmental data, such as sedimentary records collected from the same core or dated using age-depth modeling technique (Cahill et al., 2015; Blaauw, 2010).

In PaleoSTeHM v1.0, the data-level model is specified within each process-level model, which is assumed to be normally and independently distributed (Figure 2). For illustrative purposes, PaleoSTeHM v1.0 also includes an implementation of uniform likelihood together with a temporally linear model (see Figure 2 and section 4.1). For specific problems requiring different likelihood structures, users can replace the likelihood sampling code within each process-level model with most of the standard probability distributions supported by Pyro.

3.3.2 Process level modeling

The process level is a hierarchical layer where the variability of the paleo-environment is modeled and, in certain cases, decomposed. The process level reflects a scientific understanding of environmental change processes. PaleoSTeHM v1.0 offers multiple process-level models for temporal or spatio-temporal data analysis.

https://doi.org/10.5194/egusphere-2024-2183
Preprint. Discussion started: 16 October 2024

3.3.3 Temporally linear models

Starting with temporal data analysis, probably the most straightforward method for estimating linear trends and the average rate of paleo-environmental change is to fit a linear model to the observed data over time. For example, Engelhart et al. (2009) and Lin et al. (2021) applied linear regression to discrete paleo sea-level data to estimate the average rate of RSL change during the Common Era and Meltwater Pulse 1A. Over that period, the observations were qualitatively assessed to be well represented by a linear trend. A temporally linear model can be expressed as:

$$f(t) = \alpha + \beta t \tag{8}$$

where $f(t)$ is the modeled true RSL, β is the constant rate of change in RSL, and α is the intercept.

3.3.4 Change-point models

Change-point models describe a single time series by partitioning it into distinct, contiguous segments, each characterized by a linear trend over time. These models are widely used to identify the timing of abrupt changes in past climate conditions. For instance, Caesar et al. (2021) and Kemp et al. (2015) employed a change-point model to determine the onset of reduced strength in the Atlantic Meridional Overturning Circulation and the commencement of modern sea-level rise in Connecticut. With m change points, it can be written as:

$$f(t) = \begin{cases} \alpha_1 + \beta_1(t - \gamma_1), & \text{when } t < \gamma_1 \\ \alpha_{j-1} + \beta_j(t - \gamma_{j-1}), & \text{when } \gamma_{j-1} < t < \gamma_j \\ \alpha_m + \beta_{m-1}(t - \gamma_m), & \text{when } \gamma_m \leq t \end{cases} \tag{9}$$

where γ_k represents the change point, α_k denotes the expected value of RSL at that change point, and β_j indicates the rate of RSL change for each of the $m + 1$ segments. This model incorporates a continuity constraint ensuring that α_k equals α_{k-1} plus the product of β_{k-1} and the difference between γ_k and γ_{k-1}. In PaleoSTeHM, the change-point model is implemented to allow users to specify any number of change points (i.e., m in equation 9) in the model.

3.3.5 Gaussian Process models

Gaussian Process (GP) modeling is a nonparametric and Bayesian approach that has been frequently used to infer temporal (or spatio-temporal) variation of paleo-environmental change, including magnitude and rate (Ashe et al., 2019). In models with GP priors, the relationships among any set of points (e.g., over time or across both space and time) are described by a multivariate normal distribution, fully characterized by a mean function and a covariance function (or kernel). Unlike parametric models such as linear or change-point models used for spatio-temporal analysis, GP models offer greater flexibility because the shape of the curve is determined by the covariance matrix, which is inferred based on the data rather than being constrained by a predefined functional form.

https://doi.org/10.5194/egusphere-2024-2183
Preprint. Discussion started: 16 October 2024

The GP model has gained considerable traction in paleo-environmental science, largely owing to its proficiency in extracting meaningful insights from relatively small datasets. It utilizes a nonparametric framework to interpret intricate data patterns effectively. For example, Kay et al. (2021) utilized a GP model to assess herbivore richness for different latitudes in Argentina. Apart from that, Walker et al. (2021) estimated the trend and rate of RSL change across the US Atlantic coast with a GP model. A spatio-temporal GP model, which is defined by its mean function, $\mu(t)$ and covariance function (i.e., kernel) $K(X, X')$, can be expressed as:

$$f(X) \sim GP(\mu(X), K(X, X')) \tag{10}$$

here X indicates spatio-temporal information of a specific date, for which spatial information is frequently represented by latitude and longitude values with no uncertainty. A popular choice for many paleo-environmental studies is using the zero-mean function, indicating $\mu(X) = 0$ everywhere. In this case, the predictions are only determined by covariance function $K(X, X')$, which defines prior expectations about how information is shared between points in different time and space, which typically decays as the time and space differences increase (Rasmussen and Williams, 2006).

Constructing the covariance function is a pivotal and challenging step in a GP model, as it significantly influences the outcome of the inference results. Yet, its justification can sometimes be complex (Stein, 2012). PaleoSTeHM addresses this by incorporating a 'GP kernel' module under the Modeling Choices module, designed to offer more flexibility and customization extendability. This module provides a user-friendly platform for creating and managing GP kernels, streamlining the process of model construction and enhancing the adaptability of the analysis to diverse problems. For paleo-environmental applications, multiple choices of building kernels have been adopted in various studies (e.g., Walker et al., 2021; Hay et al., 2015; Kopp et al., 2016, 2014, 2009), and some examples will be in section 4.2.

3.3.6 Physical models

A physics-based model simulates real-world changes with predictive capabilities anchored in the causal mechanisms delineated by the laws of physics (see Table 1). Comparatively, statistical models mostly depend on data-driven correlations, often overlooking foundational physical principles (e.g., mass or energy conservation). Examples in paleo-environment research include using global circulation models to understand the response of the climate system to different climate forcings (Kageyama et al., 2018) and employing ice sheet dynamic models to quantify past ice sheet response to climate change (DeConto and Pollard, 2016; Tarasov et al., 2012). In the realm of paleo sea-level change modeling, the glacial isostatic adjustment (GIA) model is a widely adopted tool to characterize sea-level changes driven by the gravitational, rotational, and deformational (GRD) effects resulting from the redistribution of ice and water mass (e.g., Lin et al., 2023a; Whitehouse, 2018). The predictive power of such a model is contingent upon underlying formulation and core physical parameters (Kendall et al., 2005), such as the history of ice sheet fluctuations and the rheological properties of the Earth's interior for a GIA model. Validating the physical model against observational data will allow a more accurate representation of spatial teleconnections of sea-level change, including those linked to sea-level fingerprints (Lin et al., 2021), in stark contrast to statistical models that might merely presume correlation diminishes with distance (Walker et al., 2021).

https://doi.org/10.5194/egusphere-2024-2183
Preprint. Discussion started: 16 October 2024

Although PaleoSTeHM does not include a specific type of physics-based model (Figure 2), it offers multiple options to incorporate physical model outputs into final estimates (see examples in section 4.2). Users can use PaleoSTeHM to probabilistically calibrate physical model ensembles conditioned upon observational data. PaleoSTeHM also supports using a physical model as a mean function in a GP model. In this context, the GP covariance function essentially models the residuals—those processes not captured by the physical model—between observations and the predicted mean function. Additionally, PaleoSTeHM facilitates the construction of sampling covariance functions derived from a physical model ensemble, further enhancing its utility in model integration and assessment (Hay et al., 2015).

3.4 Gaussian Process Kernel module

The GP Kernel module in PaleoSTeHM is a cornerstone for modeling spatial and temporal variations in paleo-environmental data based on GP priors (Figure 1). It encompasses a variety of commonly used kernels in paleo-environmental studies, including the linear (or dot-product) kernel (Khan et al., 2017), radial basis function kernel (Cahill et al., 2015), rational quadratic kernel (Turner et al., 2023; Hay et al., 2015), Matérn kernel (Walker et al., 2021; Kopp et al., 2016), and periodic kernel (Meltzner et al., 2017). These kernels characterize features such as stationarity, isotropy, smoothness, and periodicity in Gaussian processes (see definitions in Table 1; Ashe et al., 2019). Detailed kernel information is given in Table 2.

Each kernel possesses unique characteristics and necessitates specific parameters (Table 2). For instance, the linear kernel produces linear trends identical to a temporally linear model, suitable for modeling signals with long temporal length scales (e.g., tectonic and GIA in Common Era and future sea level modeling; Kopp et al., 2016, 2014). The radial basis kernel and the Matérn family of kernels are highly generalizable and allow specification of the degree of differentiability (Table 2), making them suitable for representing physical processes with different levels of smoothness. For example, the GRD effects related to GIA are spatio-temporally smooth, while sediment compaction-induced sea-level rise can be much more localized and rough (i.e., less differentiable).

In the GP Kernel module of PaleoSTeHM v1.0, all kernels are designed for process-level modeling, except for the temporal and spatial white noise kernels, which add additional serially uncorrelated uncertainty at the data level. Apart from the linear and white noise kernels, all included kernels are stationary and isotropic (see Table 1). To enhance kernel construction flexibility, PaleoSTeHM supports combining different kernels, either additively, multiplicatively, or both. Designed for spatio-temporal data analysis, all GP kernels in PaleoSTeHM support temporal data (represented as a 1-dimensional vector), and most of kernels support spatial data (represented as 2-dimensional matrix including latitude and longitude; see Table 2). The spatial correlation is computed for spatial kernels based on the 1-dimensional geographical radial distance between data points. Users can choose to build a temporal or spatial kernel by switching a parameter in each kernel function.

3.5 Analysis Choices module

To accommodate diverse computational resources and varying requirements for the trade-off between modeling robustness and computational demands, the Analysis Choices module offers multiple methods for Bayesian inference of model parameters as defined in the Modeling Choices module (Figure 2). This flexibility ensures users can optimize their analyses based on avail-

https://doi.org/10.5194/egusphere-2024-2183
Preprint. Discussion started: 16 October 2024

Kernel Name	**Supports Spatial Data***	**Differentiability**	**Equation**
Radial Basis Function	Yes	Infinitely differentiable	$k(X,X') = \sigma^2 \exp\left(-\frac{1}{2}\frac{\lvert X-X'\rvert^2}{\ell^2}\right)$
Rational Quadratic	Yes	Infinitely differentiable	$k(X,X') = \sigma^2\left(1+\frac{\lvert X-X'\rvert^2}{2\alpha\ell^2}\right)^{-\alpha}$
Periodic	No	Infinitely differentiable	$k(X,X') = \sigma^2 \exp\left(-2\frac{\sin^2(\pi(X-X')/p)}{\ell^2}\right)$
2/1 Matérn	Yes	Non-differentiable	$k(X,X') = \sigma^2 \exp\left(-\frac{\lvert X-X'\rvert}{\ell}\right)$
3/2 Matérn	Yes	Once differentiable	$k(X,X') = \sigma^2\left(1+\sqrt{3}\frac{\lvert X-X'\rvert}{\ell}\right)\exp\left(-\sqrt{3}\frac{\lvert X-X'\rvert}{\ell}\right)$
5/2 Matérn	Yes	Twice differentiable	$k(X,X') = \sigma^2\left(1+\sqrt{5}\frac{\lvert X-X'\rvert}{\ell}+\frac{5}{3}\frac{\lvert X-X'\rvert^2}{\ell^2}\right)\exp\left(-\sqrt{5}\frac{\lvert X-X'\rvert}{\ell}\right)$
Linear (or dot product)	No	Once differentiable	$k(t,t') = \sigma^2(t-\gamma)\cdot(t'-\gamma)$
Sampling covariance kernel	Yes	Not applicable	$k(X,X') = Cov(m(X),m(X'))^{\#}$
Polynomial	No	d times differentiable	$k(t,t') = \sigma^2(\gamma + t\cdot t')^d$
Constant	No	Not applicable	$k(X,X') = \sigma^2$
Temporal white noise	No	Not applicable	$k(t,t') = \sigma^2\delta(t,t')$
Spatial white noise	Yes	Not applicable	$k(x,x') = \sigma^2\delta(x,x')$

Table 2. Summary of Gaussian Process kernels in PaleoSTeHM. Here, X represents spatio-temporal information, incorporating both the age and coordinates of the data; t denotes the age of the data; and x indicates the spatial coordinates. *All GP kernels can calculate temporal covariance, except spatial white noise kernel. #$Cov(m(X),m(X'))$ indicates sampling covariance between different physical models. For the polynomial kernel, d represents the degree of the polynomial, an integer determining the complexity of the model. Parameters: σ^2 = variance; ℓ = a positive characteristic length-scale parameter; α = a scale mixture parameter, when $\alpha \to \infty$, rational quadratic kernel is equivalent to radial basis function kernel; γ = offset or shift parameter, adjusting the baseline level of the kernel's output. p = periodicity parameter for the Periodic kernel, defining the cycle length of repeating patterns.

able technology and specific modeling needs. Unlike deterministic methods (e.g., least-squares), which have been extensively implemented in other packages, PaleoSTeHM focuses on developing Bayesian probabilistic approaches that more effectively manage the inherent uncertainties associated with paleo data.

https://doi.org/10.5194/egusphere-2024-2183
Preprint. Discussion started: 16 October 2024

3.5.1 Fully Bayesian analysis

A fully Bayesian analysis requires assigning prior probability distributions to all model parameters, allowing them to take on a range of probable values. These priors can either incorporate informative prior knowledge or remain uninformative and vague. Since the posterior distribution is shaped by both the priors and the likelihood of the observed data, it often becomes complex and analytically intractable. Markov Chain Monte Carlo (MCMC) methods are crucial in this case as they enable the efficient exploration and approximation of the posterior distribution. PaleoSTeHM supports two advanced MCMC samplers: Hamiltonian Monte Carlo (HMC; Neal et al., 2011) and the No-U-Turn sampler (NUTS; Hoffman et al., 2014), which provide more efficient sampling performance than traditional Metropolis-Hastings MCMC (Hastings, 1970).

HMC significantly enhances the sampling efficiency over traditional Metropolis-Hastings MCMC by utilizing gradients of the probability distribution to inform the sampling process. This method reduces autocorrelation between successive samples, increasing the effective sample size per iteration and enabling faster convergence. Based on HMC, NUTS further improves upon this by automatically adjusting the path length and effectively managing the step size. NUTS eliminates the need for manual tuning of these parameters, facilitating more efficient exploration of complex, high-dimensional distributions typical in Bayesian analysis.

Compared to other analysis choices such as empirical Bayesian models or variational Bayesian models (details provided below), a fully Bayesian model offers a more comprehensive estimation of the relative uncertainties associated with model parameters (Piecuch et al., 2017). It also offers a direct framework for sample age measurement uncertainty in an EIV manner (Table 1). However, the nature of MCMC-based samplers means they are computationally more demanding. Particularly within the EIV framework, where the number of sampling parameters increases linearly with data size, this leads to a polynomial increase in the computational power required (Belloni and Chernozhukov, 2009), which can be significant and unaffordable when dealing with large datasets or complex models.

3.5.2 Empirical Bayesian analysis

Unlike Fully Bayesian analysis, which requires full probability distributions for prior and posterior, Empirical Bayesian analysis offers a practical alternative. This approach approximates a fully Bayesian treatment where parameters at the highest level of the hierarchy are fixed at their most likely values rather than being integrated out. This optimization is typically achieved using the maximum likelihood estimate, leading to a posterior distribution that is conditional on the data and these optimized parameters:

$$p(f|y,\hat{\theta}_s,\hat{\theta}_d) \propto p(y|f,\hat{\theta}_d)p(f|\hat{\theta}_s) \tag{11}$$

here, the posterior probability of the latent processes f is inferred, assuming that the hyperparameters at the data and process levels ($\hat{\theta}_d$ and $\hat{\theta}_s$) are known and fixed. While the existing code base allows for explicit bounds to be set on hyperparameters for the maximum likelihood estimate (e.g., Ashe et al., 2019; Kopp et al., 2016), it does not provide for an explicit prior distribution for the parameters. By leveraging Pyro's variational inference capabilities (details below), PaleoSTeHM enables users not

https://doi.org/10.5194/egusphere-2024-2183
Preprint. Discussion started: 16 October 2024

only to optimize hyperparameters using the maximum likelihood estimate but also to define prior distributions for each model parameter explicitly. This allows optimization to be conducted in a maximum *a posteriori* probability estimation manner, assuming the variational distribution is a Dirac delta function. In PaleoSTeHM, by default, the optimization is achieved using Adam, a stochastic optimizer (Kingma and Ba, 2014). While empirical Bayesian analysis generally requires fewer computational resources than fully Bayesian methods, it is important to note that, assuming hyperparameters at the data and process levels are known and fixed may lead to substantial underestimation in the inference uncertainty (Piecuch et al., 2017).

3.5.3 Variational Bayesian analysis

Considering the computational expense required to perform MCMC in fully Bayesian analysis and the limitations of Empirical Bayesian methods that fail to account for the uncertainty of hyperparameters, PaleoSTeHM also supports variational Bayesian analysis, which emerges as an efficient intermediary. Rather than directly sampling from the posterior distribution through MCMC, variational Bayesian methods aim to approximate the true probability distribution ($p(f,\theta_s,\theta_d|y)$) with a simpler, parametric probability distribution ($q(f|\phi)$). Thus, Bayesian inference is transformed from a sampling challenge into an optimization problem—known as variational inference—requiring significantly fewer computational resources while facilitating uncertainty estimation.

In PaleoSTeHM, variational Bayesian analysis is achieved by optimizing the variational parameters ϕ to minimize the Kullback-Leibler (KL) divergence, a metric to effectively measure the difference between two distributions:

$$\phi = \arg\min_{\phi} \mathrm{KL}\left[q(f,\theta_s,\theta_d|\phi)||p(f,\theta_s,\theta_d|y)\right] \tag{12}$$

Adam facilitates this minimization, and the variational distribution for PaleoSTeHM is a normal distribution by default. In contrast to MCMC-based fully Bayesian analysis, which often requires computational power that increases polynomially with the number of data points, the optimization-driven approach of variational Bayesian analysis generally scales linearly. Consequently, variational methods can handle larger datasets more effectively, making them suitable for large-scale problems prohibitively for full Bayesian analysis.

3.5.4 Incorporation of temporal uncertainty

PaleoSTeHM provides two methods to incorporate temporal uncertainty into final estimations. The first method uses EIV framework (Cahill et al., 2015), which directly incorporates temporal uncertainty through MCMC sampling of the distribution. The second approach adopts the noisy-input framework (McHutchon and Rasmussen, 2011), which applies a first-order Taylor series approximation—a linear expansion around each input point—to account for errors in the independent variable, time, thereby translating these into equivalent errors in the dependent variable:

$$f(x_i,t_i) \approx f(x_i,\hat{t}_i) + \epsilon_i^t \frac{\partial f(x_i,\hat{t}_i)}{\partial t} \tag{13}$$

here $\hat{t}_i$ and ϵ_i^t are the same as in equation 5, standing for mean observational age and age uncertainty, respectively. The integration of temporal uncertainty within PaleoSTeHM is executed alongside each process level model (Figure 2). All process-

https://doi.org/10.5194/egusphere-2024-2183
Preprint. Discussion started: 16 October 2024

level models are implemented using an EIV framework, while for the GP models, both EIV and noisy-input frameworks are available (Figure 2).

4 Results

This section presents illustrative results using a tutorial format to enhance PaleoSTeHM's usability. All codes and data are accessible and actively managed on the PaleoSTeHM GitHub page (see code and data availability).

Firstly, we demonstrate various data levels, process levels, and analysis choice modeling techniques for time series analysis using coral reef data from the Great Barrier Reef and salt marsh data from New Jersey and North Carolina. Subsequently, we provide examples of reconstructing spatio-temporal sea-level changes in the US Atlantic coast using different process-level models. The prior and posterior distributions and analysis choice for each model are provided in Table A1. It should be noted this section only briefly describes the modeling results; for a more systematic analysis of paleo-environmental modeling results based on different statistical techniques, the user can refer to Ashe et al. (2019), PAGES2k Consortium (2019), and Tingley et al. (2012).

4.1 Time series analysis

4.1.1 Data level modeling

Although numerous paleo-environmental applications commonly assume that proxy reconstruction uncertainties are normally distributed (Ashe et al., 2019; Khan et al., 2019; Tingley et al., 2012), certain types of proxies may exhibit different forms of uncertainty. A typical example is the coral reef sea-level indicator, where reconstruction uncertainty can also be represented by a uniform distribution (Lin et al., 2021) according to species-specific living-habitat range (Hibbert et al., 2016).

To illustrate the impact of the data level model on inference results, we apply a temporally linear model within an EIV framework to coral reef data from the Great Barrier Reef (Yokoyama et al., 2018). We use two alternative data-level models. The first can be expressed as:

$$\epsilon_1^y \sim U(\tau_l - \omega_1, \tau_u + \omega_1) \tag{14}$$

where U indicates a uniform distribution between lower and upper ranges defined by specific coral species (τ_l and τ_u) and an additional white noise, defined by hyperparameter ω_1, the second data level model can be represented as:

$$\epsilon_2^y \sim N(\mu_2, \sqrt{\sigma_2^2 + \omega_2^2}) \tag{15}$$

where N indicates a normal distribution with mean μ_2 and a standard deviation σ_2, both of which are determined by specific coral species, and ω_2 is an additional white noise hyperparameter. The same prior distributions for each parameter are used for both data-level models, which are represented as non-informative uniform distributions.

For both models, the posterior distribution is determined by 11,000 posterior samples drawn from a NUTS sampler, with the first 1,000 samples discarded as burn-in steps. It can be seen in Figure 3 that, although the inference results from different

https://doi.org/10.5194/egusphere-2024-2183
Preprint. Discussion started: 16 October 2024

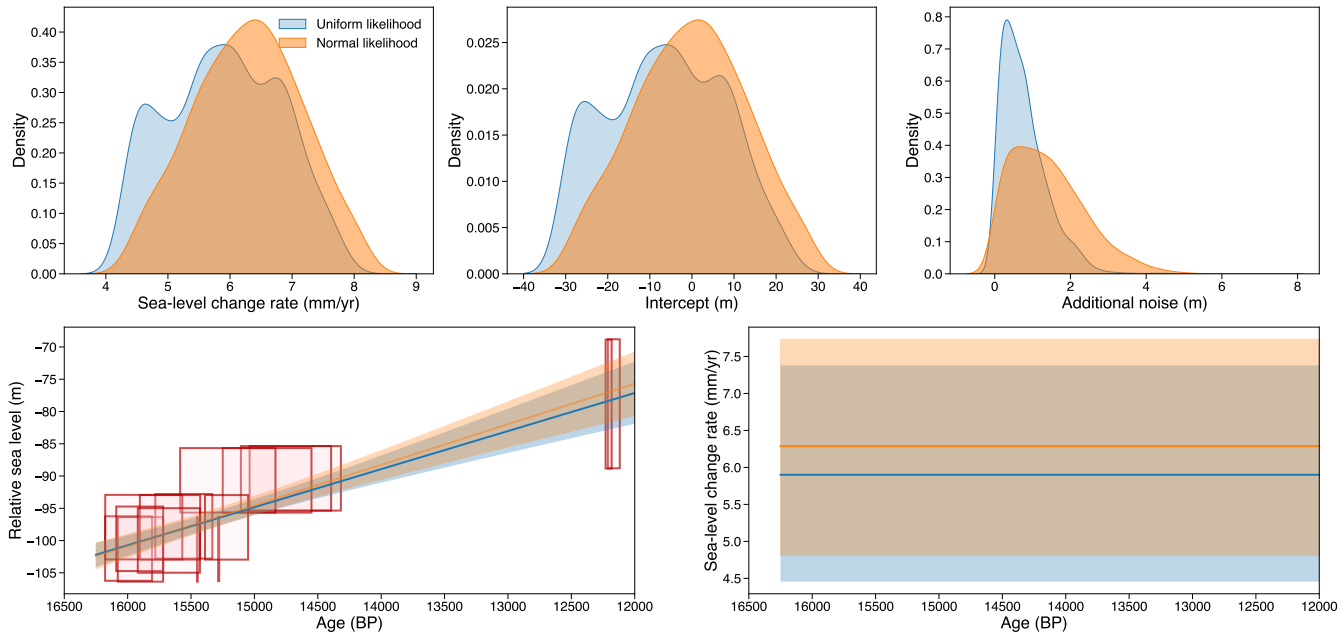

Figure 3. Data level models impact on temporal sea-level change inference at the Great Barrier Reef. The upper panel displays the posterior probability density functions of model parameters, assuming either a uniform likelihood (blue) or a normal likelihood (orange). The bottom panel presents the inferred mean sea-level trends and rates along with a 90% credible interval, where sea-level data are represented by red boxes with horizontal range representing $\pm 2\sigma$ age uncertainty and vertical range indicating the reconstructed maximum and minimum sea-level range determined by coral species. CI = credible interval, BP = before present.

data level models are overall similar, there are still some noticeable differences in the inferred sea-level change trend and rate. Uniform and normal likelihoods yield average sea-level rates of 5.91 mm/yr (4.45-7.38 mm/yr; 90% credible interval, CI) and 6.29 mm/yr (4.81-7.73 mm/yr), respectively. These likelihood assumptions also produce considerably different additional noise parameter distributions.

4.1.2 Process level modeling

To demonstrate the impact of different process level models on inferring paleo sea-level time series, we will use the same data level model together with process level models introduced in section 3.3.2, employing non-informative priors (Table A1). The sea-level data used here is a near-continuous record from single cores of salt-marsh sediment from Leeds Point (New Jersey) covering the Common Era (Kemp et al., 2013). For this database, a normal likelihood data level model is adopted with sea-level reconstruction uncertainties provided by the original study. Here we will test three process level models: (a) temporally linear model; (b) change-point model (assuming 3 change points); (c) Gaussian Process model with an RBF kernel. Posterior distributions for model a and b were sampled through a variational Bayesian manner, while model c was sampled using an empirical Bayesian approach.

https://doi.org/10.5194/egusphere-2024-2183
Preprint. Discussion started: 16 October 2024

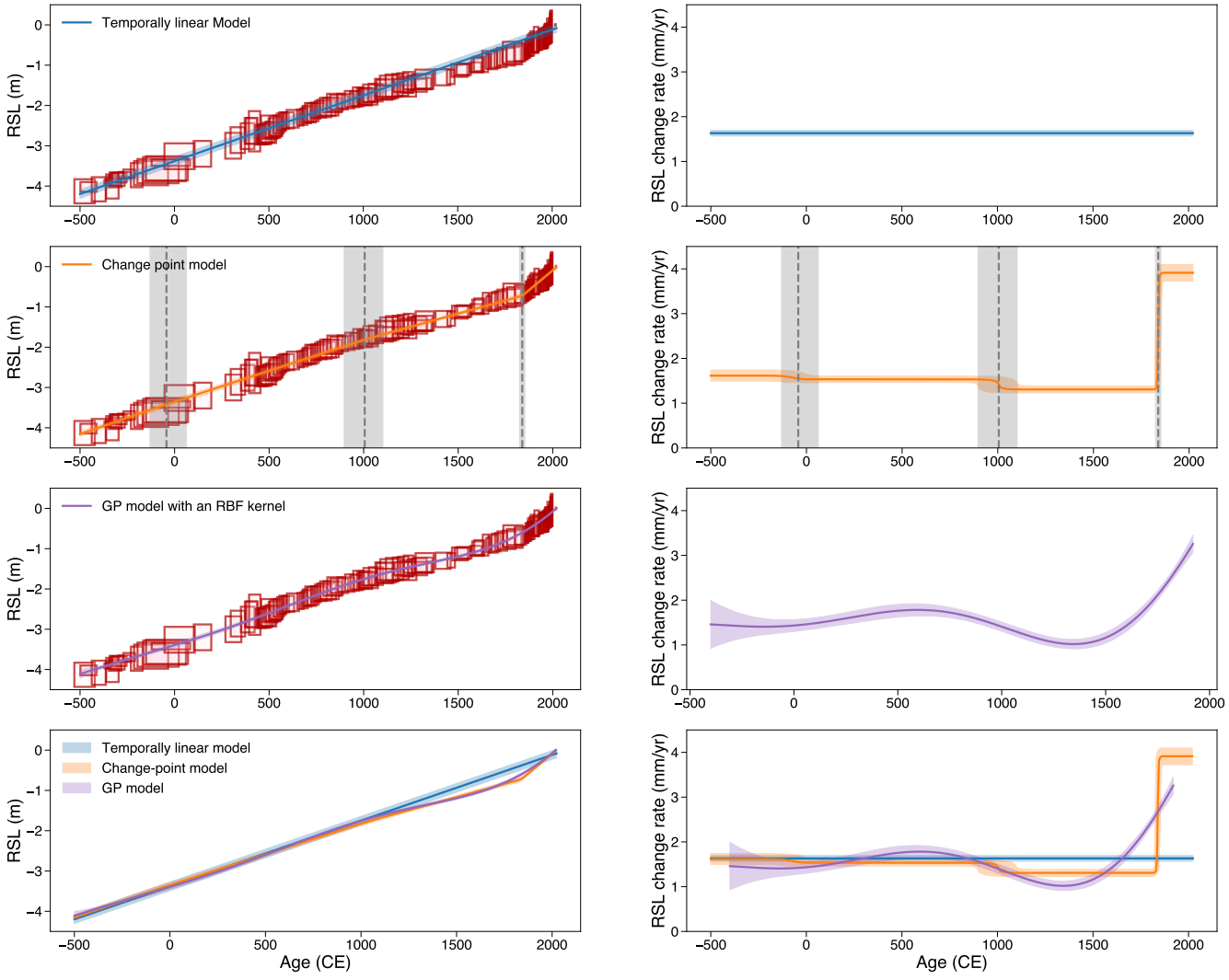

Figure 4. Process level models impact on temporal sea-level change inference at New Jersey. Common Era sea-level comparison of linear model (row 1), change-point model (assuming 3 change points; row 2), and Gaussian Process model with an RBF kernel, where input data are continuous cores. Output includes estimates of RSL (column 1) and rates of RSL change (column 2), which are each shown with mean and 90% credible intervals. CE = Common Era.

Figure 4 shows estimated RSL trends and rates of RSL change for each process model. The results differ significantly due to fundamentally different model formulations. The temporally linear model can only estimate an averaged trend and rate of sea-level change and will never predict an accelerated RSL change. Comparatively, the change-point model presents a noticeable change in RSL rate from 1.49 mm/yr (1.26-1.70 mm/yr) between -500 CE and 1839 CE (1824-1852 CE) to 3.91 mm/yr (3.72-4.10 mm/yr) after 1839 CE. Such an estimate is suitable for finding the time of emergence for various environmental

https://doi.org/10.5194/egusphere-2024-2183
Preprint. Discussion started: 16 October 2024

change problems (e.g., Walker et al., 2022; Caesar et al., 2021; Lyu et al., 2014). As a non-parametric approach, the GP model can produce continuous distributions of RSL change rates over time, allowing for the estimation of multiple inflection points (Walker et al., 2022). However, using the RBF kernel, which is infinitely differentiable, can lead to overly smooth changes in RSL rate—an assumption inappropriate in many environmental statistics (Stein, 2012).

4.1.3 Analysis choices

Using similar near-continuous sediment core data from Sand Point, North Carolina (Kemp et al., 2011), we illustrate the effects of analysis choices on RSL inference. Here, we only use a subset of the original data to better demonstrate the difference between various analysis choices. The adopted data and process level model employ a normal likelihood with a GP model using an RBF kernel (Table A1). The hyperparameters will be sampled using empirical, fully Bayesian, and variational Bayesian methods. For the fully Bayesian method, the posterior distribution is determined by 2,200 posterior samples drawn from a NUTS sampler, with the first 200 samples discarded as burn-in steps. For the empirical and variational Bayesian methods, the hyperparameters were optimized using the Adam optimizer over 500 iterations (Kingma and Ba, 2014). The run times of each implementation are reported on a 2023 MacBook Pro with an Apple M2 Pro chip.

Figure 5 compares posterior distributions of RSL trend and rate of change and the computational time for each analysis choice. The empirical Bayesian method requires the least computational power, only providing a point estimate of hyperparameters without accounting for underlying uncertainty. Although more computationally demanding, the fully Bayesian method captures the hyperparameter uncertainties effectively. As an intermediary, variational Bayesian method requires slightly more computational time compared to empirical method but can derive a variational posterior distribution that is largely similar to that obtained by the fully Bayesian method through MCMC sampling. In contrast, the point estimate by the empirical Bayesian method lies at the third percentile of the posterior hyperparameter distributions by the fully Bayesian method, indicating a strong bias.

Because of the near continuous sea-level data with smoothly rising sea-level trend in North Carolina, the inference results from these three methods are similar. However, given that geological sea-level data is often sparsely distributed across both spatial and temporal domains and may subject to abrupt change in rate, neglecting the underlying uncertainty of hyperparameters by empirical Bayesian method can result in a significant underestimation of the final inference uncertainty compared with fully Bayesian method.

4.2 Spatio-temporal analysis

The spatio-temporal analysis is a common challenge in paleo-environmental studies, for example, how to reconstruct continuous spatiotemporal signals from sparse and noisy data. For this perspective, PaleoSTeHM offers a range of options from purely statistical to purely physical approaches. Utilizing a sea-level database containing 1,043 proxy records spanning from 11 ka to the present, compiled by Ashe et al. (2019) from previous studies (Kemp et al., 2017a, b, 2015, 2014, 2013; Khan et al., 2017; Engelhart and Horton, 2012), here we attempt to recover the spatio-temporal RSL pattern along with its associated uncertainty.

https://doi.org/10.5194/egusphere-2024-2183
Preprint. Discussion started: 16 October 2024

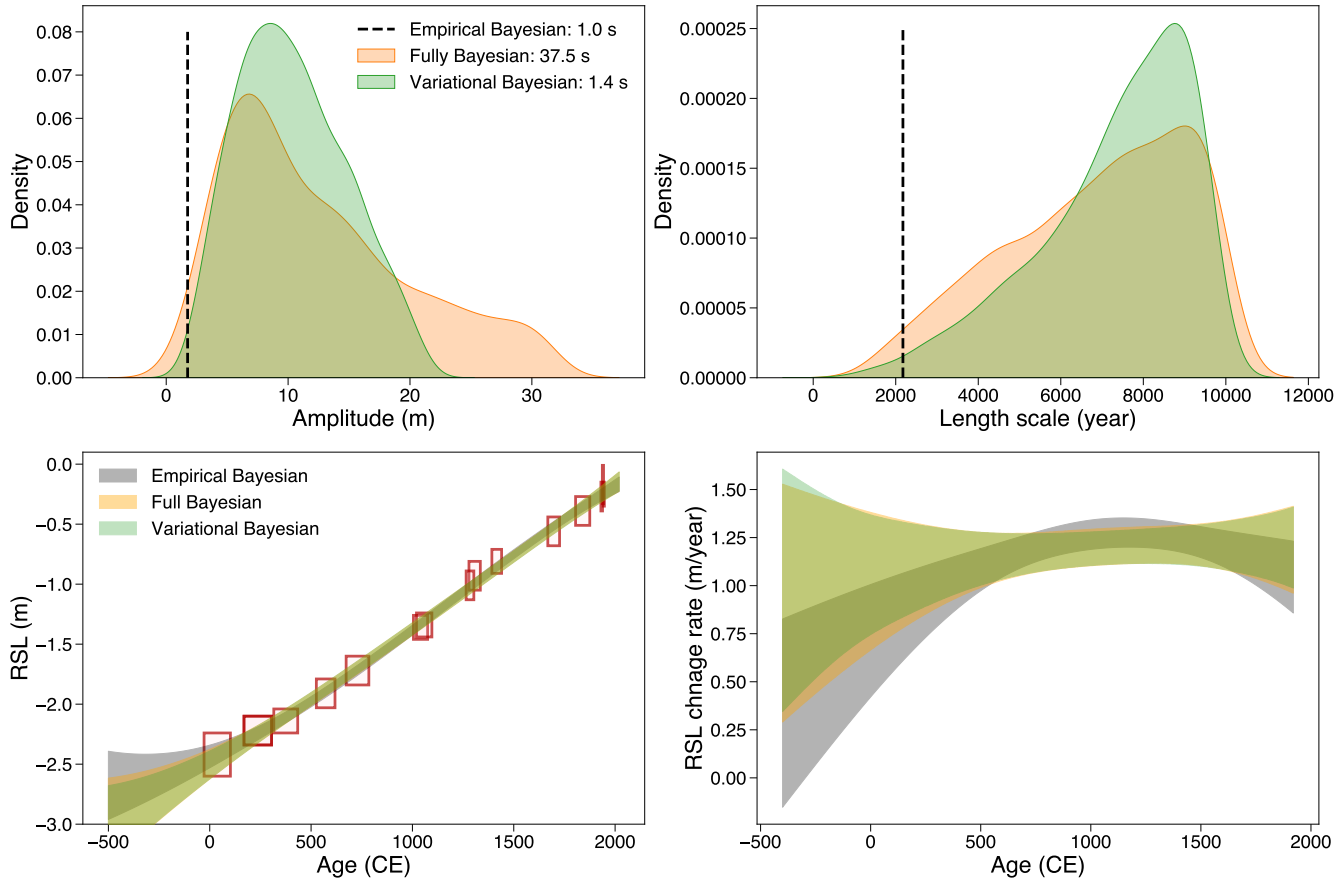

Figure 5. Analysis choices impact on temporal sea-level change inference at North Carolina. Row 1, GP model hyperparameters optimization results along with the required computational time in second (based on a 2023 MacBook Pro with an Apple M2 Pro chip). Row 2, common Era sea-level comparison between three analysis choices, the results indicate 90% credible interval of RSL change trend (left) and rate (right).

We demonstrate four process level models that have been used in previous studies: (i) a GP model with a zero mean function and multiple isotropic kernels (Ashe et al., 2019); (ii) a GP model with the mean function determined by a GIA model and multiple isotropic kernels (Walker et al., 2021; Kopp et al., 2016); (iii) a GP model with a zero mean function and a sampling covariance kernel determined by a GIA model ensemble (Kopp et al., 2009); and (iv) a purely GIA model ensembe (Lin et al., 2023a). All models assume a data-level model with a normal likelihood determined by RSL reconstruction uncertainty and an additional white noise term. For this analysis, we implement a noisy-input framework to address temporal uncertainty and use the empirical Bayesian method to optimize hyperparameters for models *i*, *ii*, and *iii*, while model *iv* is optimized through the variational Bayesian method (Table A1).

For model *i*, we follow the kernel structure as in Ashe et al. (2019), which can be expressed as:

$$f(X) \sim GP(0, K_1(X, X')) \tag{16}$$

https://doi.org/10.5194/egusphere-2024-2183
Preprint. Discussion started: 16 October 2024

$$K_1(X, X') = g(t) + r(x,t) + l(x,t) \tag{17}$$

where $g(t)$ represents a spatially-uniform covariance function, while $r(x,t)$ and $l(x,t)$ are regional and local varying isotropic covariance functions, respectively. These are characterized by a 3/2 Matérn temporal kernel for $g(t)$, and a product of a 3/2 Matérn temporal kernel and a 1/2 Matérn spatial kernel for $r(x,t)$ and $l(x,t)$, which are differentiated by prior distributions of hyperparameters.

Similarly, model *ii* can be written as:

$$f(X) \sim GP(\text{GIA}(X), K_2(X, X')) \tag{18}$$

$$K_2(X, X') = r(x,t) + l(x,t) \tag{19}$$

here, the mean expectation is determined by RSL prediction from ICE_7G ice model with VM5a Earth model (Roy and Peltier, 2018; Peltier et al., 2015), and $r(x,t)$ and $l(x,t)$ are the same as in equation 17. We do not include $g(t)$ as we assume the mean function already captures it.

Model *iii* can be denoted as:

$$f(X) \sim GP(0, K_3(X, X')) \tag{20}$$

$$K_3(X, X') = Cov(m(X), m(X')) \cdot exp(-|t - t'|/\tau^2) \tag{21}$$

here, $Cov(\cdot)$ here indicates a sampling covariance function through a physical model ensemble m, where m includes an ensemble of forward GIA models: (a) ICE_7G ice model with VM5a Earth model; (b) PaleoMIST ice model (Gowan et al., 2021) with 71 km lithosphere and 0.3 and 70 $\times$ 10^{21} Pa s upper and lower mantle viscosity; and (c) ANU ice model (Lambeck et al., 2014) with 71 km lithosphere and 1 and 10 $\times$ 10^{21} Pa s upper and lower mantle viscosity. To expand the variability of physical model predictions, we create six synthetic GIA model outputs by enlarging or shrinking these three GIA model outputs by 1.5. Therefore, this physical model ensemble consists of predictions from nine models. More details about the physics-based GIA model used here can be found in Lin et al. (2023b). To stabilize the estimate and reduce variability related to finite sample size, we applied a temporal Gaussian taper function to this kernel, controlled by a parameter τ. Following Hay et al. (2015); Kopp et al. (2009), we set τ to 3000 years.

And lastly, model *iv* can be written as a weighted mean of different physical models:

$$f(X) = \sum_{n=1}^{N} \nu_i \text{GIA}_i(X) \tag{22}$$

https://doi.org/10.5194/egusphere-2024-2183
Preprint. Discussion started: 16 October 2024

$$\nu \sim \text{Dirichlet}(\alpha_d) \tag{23}$$

In this model, ν represents the relative weights associated with each GIA model. These probabilities follow a Dirichlet distribution (or multivariate beta distribution) characterized by a concentration parameter α_d. A value greater than 1 for α_d indicates a preference for a more evenly distributed probability across all models. In contrast, a value less than 1 indicates a preference for more concentrated probabilities on fewer models. For this experiment, we set α_d according to each GIA model prediction fit to RSL observation (using weighted root mean square as a metric, see Table A1).

A comparison of RSL inference results between different spatio-temporal process level models is provided in Figure 6. At the purely statistical end of the process model spectrum, model *i* correlates RSL from various locations and times based solely on their spatio-temporal proximity, a property derived from the adopted isotropic kernels. According to model *i*, the RSL change along the US Atlantic coast during the Holocene was dominated by a spatially uniform signal (produced by the regional common kernel, $g(t)$, in equation 17; Table A1), which contributed to more than 25 m of RSL rise. In contrast, $r(x,t)$ and $l(x,t)$ only produce up to 5 m of spatially variable RSL signal, resulting in virtually no spatial pattern in the mean RSL prediction of this model. In the temporal domain, multiple studies have demonstrated that GP models like model *i* can accurately recover multi-millennial sea-level variation trends at locations with abundant sea-level observations, such as Florida, as shown in Figure 6a (Tang et al., 2023; Ashe et al., 2019; Cahill et al., 2015). However, spatial inferences based on isotropic and stationary kernels are often considered overly simplistic (Stein, 2005a), partly due to the sparse nature of geological data and the complexity of environmental change mechanisms. As see in Figure 6i, geological sea-level data are mostly collected across paleo-coastal areas. Therefore, RSL inferences from model *i* are only representative of coastal areas (as opposed to terrestrial or marine areas) and cannot adequately reflect the physical knowledge of paleo sea-level change (e.g., the RSL uncertainty caused by the existence of the Laurentide Ice Sheet).

Model *ii* uses a deterministic GIA model (ICE_7G ice model with VM5a Earth model) as the GP mean function. By harnessing a physics-based model, model *ii* captures intricate spatial sea-level variation patterns due to the GIA-induced GRD effects (Figure 4). In this setup, the covariance functions describe residuals between the GIA model and RSL observations (mostly captured by $r(x,t)$ in equation 19). At -5500 CE, the GIA model underestimates ~10 m RSL at New Jersey (Figure A1), which may reflect oversimplified physics (e.g., 3D rheology; Austermann et al., 2013), biased sampling of physical parameters (such as poorly-constrained ice history), or missing physical processes in the GIA model (e.g., sediment isostatic adjustment; Lin et al., 2023a). Because model *ii* assumes no uncertainty in GIA modeling, the uncertainty quantification here also relies solely on the radial distance from RSL data points (Figure 6j).

Model *iii* employs a kernel constructed by sampling covariance between different forward GIA models based upon alternative ice and Earth models. The incorporation of relevant physics in GP kernel construction enables model *iii* to capture anisotropic behaviors, non-stationarities, heterogeneities and teleconnections (Table 1) inherent in the physical processes of RSL change that cannot be easily described by normal classes of covariance function (Table 2). For example, the size of the Laurentide Ice Sheet is positively correlated with RSL around the northern Great Lakes while negatively correlated with RSL

https://doi.org/10.5194/egusphere-2024-2183
Preprint. Discussion started: 16 October 2024

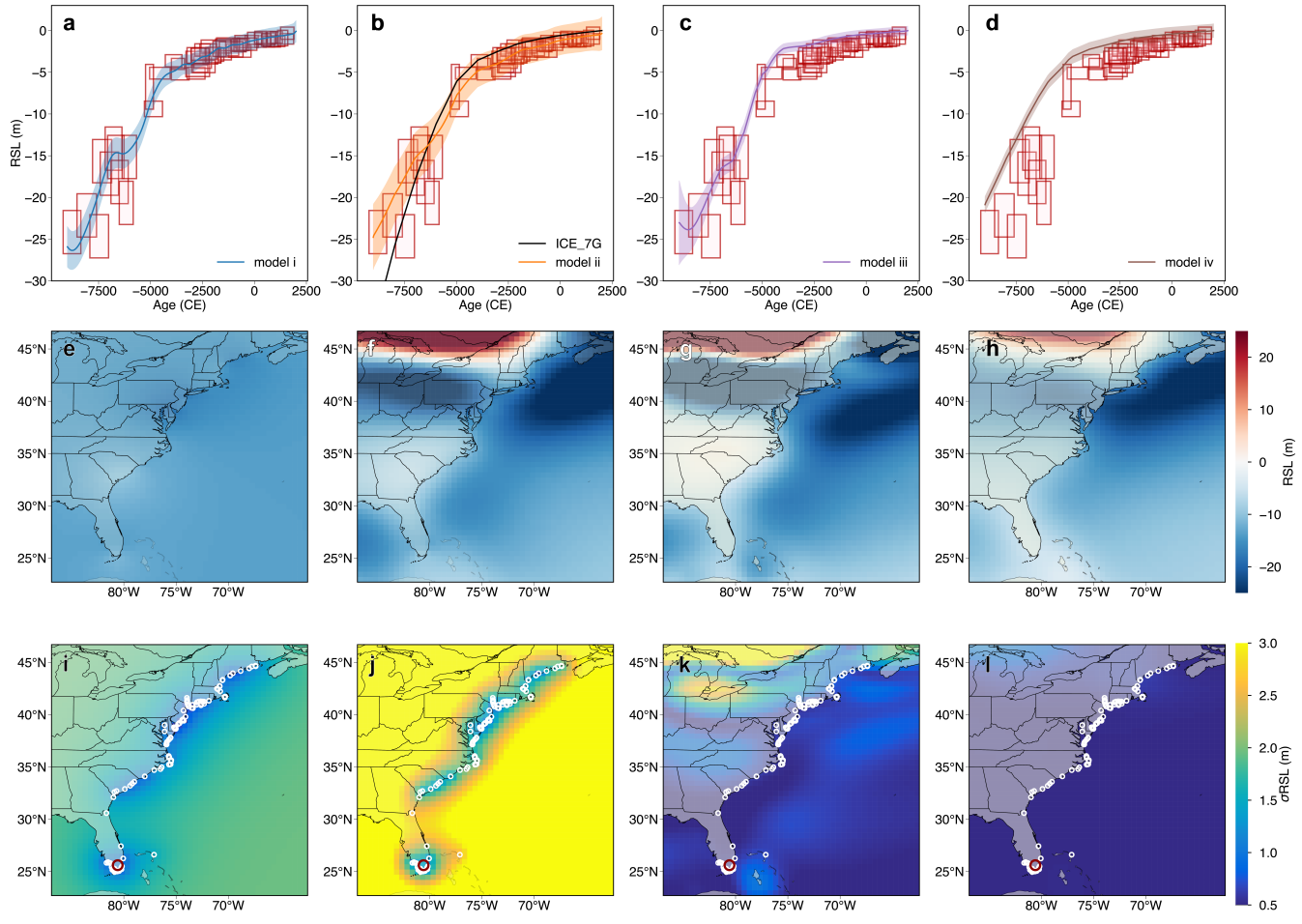

Figure 6. Process level models impact on spatio-temporal sea-level change inference along the US Atlantic coast. Each column represents a different process level model. (a-d) Row 1 displays the mean and 90% credible or confidence interval for RSL predictions at Florida (indicated by a red dot in Row 3). For Model ii, the GP mean function, determined by the ICE7G model, is depicted with a black line (b). (e-h) Row 2 shows the mean RSL prediction for the year -5,500 CE. (i-l) Row 3 illustrates the standard deviation of the RSL prediction for the year -5,500 CE, where white dots indicate locations where each sea-level data collected.

in peripheral bulge regions like New Jersey (Figure 6g). Whereas potential problems for this method include the computational burden required for thoroughly sampling physical model parameters, and structural errors in the physical model, such as over-simplifications or missing processes. Also, this method's posterior mean and standard deviation are not directly interpretable as model *iv*.

Lastly, model *iv* represents the purely physical end of the process level spectrum. It is equivalent to a linear combination of physical models according to data-model misfits (e.g., chi-square; Lin et al., 2021; Li et al., 2020; Lambeck et al., 2014). The mean and uncertainty estimated by this method reflect the parametric uncertainty of a certain physical model, allowing for

https://doi.org/10.5194/egusphere-2024-2183
Preprint. Discussion started: 16 October 2024

direct physical interpretation of prediction results, such as calculating posterior distributions of global ice history. However, this method is susceptible to model structural errors, similar to model *iii*. For instance, the uncertainty here is underestimated due to the limited sample size of physical parameters (only nine models were used here). Additionally, due to the limited direct constraints on ice history, the millennial-resolution ice models used in forward GIA modeling cannot capture centennial sea-level variation as effectively as models *i-iii* (Figure 6).

5 Discussion

5.1 Generalization for paleo-environmental problems

Although all functionality demonstrations in this paper focus on paleo-sea-level applications, the great flexibility of the hierarchical model any statistical model can be interpreted as a hierarchical model- enables the PaleoSTeHM modeling framework to be readily applicable to most paleo-environmental problems. Employing a hierarchical model to address various paleo-environmental problems enhances transparency by clearly distinguishing between modeling assumptions and analysis methods, as well as between process variability and observational noise.

Because of the commonality between paleo sea-level and many other paleo-environmental datasets that are characterized by sparse and discrete, such as paleo temperature (PAGES2k Consortium, 2017), past ice sheet thickness (Small et al., 2019) and sediment deposition depth (Wang et al., 2018), the data and process level models introduced this paper can be readily generalized to those paleo-environmental fields. For example, Tingley et al. (2012) and Stein (2005b) suggested it is reasonable to use a GP model to describe latent some space-time climate processes like annual mean surface temperature anomalies and daily wind speed; Lin et al. (2023a) applied a spatial GP model to recover the spatial pattern of Holocene coral reef depth based on a Holocene coral reef deposition depth database across the Great Barrier Reef (Hinestrosa et al., 2022); and Caesar et al. (2018) implemented a change-point model on multiple proxy datasets to detect significant reductions in the strength of the Atlantic Meridional Overturning Circulation.

Beyond the process level models featured in PaleoSTeHM v1.0, various approaches have been employed for paleo-environmental analyses. Common techniques for addressing problems in this field include principal component analysis, equivalent to the empirical orthogonal function method when temporal aspects are considered, autoregressive models, and generalized additive models. For instance, Shakun and Carlson (2010) used an empirical orthogonal function approach to detect modes of deglacial temperature variability, and Piecuch et al. (2017) adopted a degree-1 autoregressive model to reconstruct sea-level evolution using tide gauge data, and Simpson (2018) developed a series of generalized additive models to model paleo-ecological time series. While the reimplementation of these models in PaleoSTeHM is beyond the scope of this paper, doing so would benefit from the framework's multiple analysis options and its capacity for smooth integration with flexible data and parameter-level models.

https://doi.org/10.5194/egusphere-2024-2183
Preprint. Discussion started: 16 October 2024

5.2 Future developments

From a scientific perspective, numerous promising directions exist for further development of PaleoSTeHM.

Existing data-level models only support a common class of likelihood. However, in paleo-environmental studies, it is typical for proxy data to be subject to complex likelihoods. For instance, organic matter that has been radiocarbon-dated undergoes a calibration procedure to account for the time-evolving atmospheric carbon concentration, which can yield a data chronology characterized by multi-modal distributions that significantly differ from each other. Similarly, it is common for paleo-environmental studies to use multiple types of proxy data with different likelihoods to infer a common signal. Recently, new approaches have been developed to account for nonparametric proxy distributions within a hierarchical modeling framework (e.g., Ashe et al., 2022), which could better characterize the underlying uncertainty but can be computationally expansive.

The current GP Kernel module incorporates commonly used kernel options that are stationary, isotropic, and space-time separable (definitions in Table 1). While these assumptions simplify calculations significantly, they may not be suitable for some environmental applications. For example, temperature and dew point variations often exhibit strong non-stationary behavior influenced by diverse geographic and atmospheric conditions (Poppick and Stein, 2014). Furthermore, temperature anomalies over the last two millennia (Mann et al., 2008) demonstrate strong space-time interactions, which cannot be captured by a space-time separable kernel (Tingley et al., 2012). Developing a scientifically richer class of kernel structures could be an important future advancement for PaleoSTeHM. However, given the fundamental differences across various paleo-environmental problems, generalizing sophisticated kernel structures to multiple fields remains challenging.

Another outstanding issue for GP based process level models is scalability, the standard GP models included in PaleoSTeHM v1.0 cannot scale well to large data sets ($>$10 thousands data points). Thus, implementing alternative classes of GP models within PaleoSTeHM to model large data sets, especially when incorporating modern environmental observations, which often consist of millions of data points, is an important next step for PaleoSTeHM to develop in the future. Some potentially efficient methods include sparse GP (Quinonero-Candela and Rasmussen, 2005), stochastic variational GP (Hensman et al., 2013), and exact GP with black-box matrix-matrix inference (Wang et al., 2019).

Building upon machine learning infrastructure, another promising direction for the future development of PaleoSTeHM is integrating spatio-temporal hierarchical modeling with machine learning-based emulators as a process-level model. An emulator indicates a statistical model that mimics the behavior of the physics-based simulator but is computationally cheap to run (Reichstein et al., 2019), which is particularly useful for fast sensitivity analysis, model parameter calibration, and derivation of confidence intervals for the estimate. The use of statistical emulators trained by physical models will enable hierarchical models to capture the non-stationary physical systems better and enable better interpretation of the modeling results. For paleo-environment, Holden et al. (2019) presents a GP-based emulator for an atmosphere-ocean general circulation model with intermediate-complexity, and (Lin et al., 2023b) developed a neural network-based emulator for GIA-induced global sea-level change.

More broadly, PaleoSTeHM has been developed by a small team specialized in modeling paleo sea-level changes over multi-millennial time scales. Moving forward, a critical objective is to expand PaleoSTeHM into a larger-scale paleo-environmental

https://doi.org/10.5194/egusphere-2024-2183
Preprint. Discussion started: 16 October 2024

community project, where modules are developed autonomously by diverse research teams. The design of PaleoSTeHM, which allows modules to act as wrappers for independently developed code, is specifically intended to facilitate this collaborative effort.

6 Conclusion

Paleo-environmental records provide critical out-of-sample information essential for contextualizing current global changes and testing models used to simulate future environmental scenarios. However, our understanding of past environmental changes is often complicated by the sparse nature of geological records, geochronological uncertainties, and the indirect relationships between proxies and ecological variables. Hierarchical modeling offers a conceptually straightforward framework to address these challenges, though the limited availability of user-friendly software often hinders it. PaleoSTeHM offers a flexible and open-source platform that facilitates the rapid and easy implementation of hierarchical models for paleo-environmental applications. The inclusion of multiple process-level models in PaleoSTeHM allows it to be readily applicable across a broad spectrum of paleo-environmental studies. In contrast, its flexibility allows for customization to meet the specific needs of diverse paleo-environmental problems, such as using different Gaussian Process kernels or substituting alternative process-level models.

Code and data availability. The development version of PaleoSTeHM is available under an MIT license in a Git version-controlled repository at https://github.com/radical-collaboration/PaleoSTeHM (last access: 31 May 2024). The latest release is archived on Zenodo with the identifier https://doi.org/10.5281/zenodo.12730141 (Lin et al., 2024). Documentation of PaleoSTeHM is available at https://paleostehm.org/. All codes required to generate results and figures shown in section 4 are available in the repository.

Author contributions. YL developed the PaleoSTeHM architecture and modules. REK and SJ conceived the project, and REK supervised and administered it. AR guide software engineering of PaleoSTeHM, ELA provided YL guidance in statistical modeling. All authors contributed to code development and the writing and editing of the paper.

Competing interests. The authors declare no competing interests.

Acknowledgements. We thank Chris Piecuch, Roger Creel, Fiona Hibbert, Jeremy Ely, Sönke Dangendorf, and the participants of PaleoSTeHM workshops for their valuable feedback in the development of PaleoSTeHM. Yucheng Lin, Robert E. Kopp, Alexander Reedy, and Shantenu Jha are supported by the U.S. National Science Foundation under awards 2002437 and 2148265. The authors acknowledge

https://doi.org/10.5194/egusphere-2024-2183
Preprint. Discussion started: 16 October 2024

PALSEA, a working group of the International Union for Quaternary Sciences (INQUA) and Past Global Changes (PAGES), which received support from the Swiss Academy of Sciences and the Chinese Academy of Sciences.

https://doi.org/10.5194/egusphere-2024-2183
Preprint. Discussion started: 16 October 2024

References

Aitken, A. C.: IV.—On least squares and linear combination of observations, Proceedings of the Royal Society of Edinburgh, 55, 42–48, 1936.

Ashe, E. L., Cahill, N., Hay, C., Khan, N. S., Kemp, A., Engelhart, S. E., Horton, B. P., Parnell, A. C., and Kopp, R. E.: Statistical modeling of rates and trends in Holocene relative sea level, Quaternary Science Reviews, 204, 58–77, 2019.

Ashe, E. L., Khan, N. S., Toth, L. T., Dutton, A., and Kopp, R. E.: A statistical framework for integrating nonparametric proxy distributions into geological reconstructions of relative sea level, Advances in Statistical Climatology, Meteorology and Oceanography, 8, 1–29, 2022.

Austermann, J., Mitrovica, J. X., Latychev, K., and Milne, G. A.: Barbados-based estimate of ice volume at Last Glacial Maximum affected by subducted plate, Nature Geoscience, 6, 553–557, 2013.

Belloni, A. and Chernozhukov, V.: On the computational complexity of MCMC-based estimators in large samples, 2009.

Bingham, E., Chen, J. P., Jankowiak, M., Obermeyer, F., Pradhan, N., Karaletsos, T., Singh, R., Szerlip, P., Horsfall, P., and Goodman, N. D.: Pyro: Deep universal probabilistic programming, Journal of machine learning research, 20, 1–6, 2019.

Blaauw, M.: Methods and code for 'classical'age-modelling of radiocarbon sequences, Quaternary geochronology, 5, 512–518, 2010.

Caesar, L., Rahmstorf, S., Robinson, A., Feulner, G., and Saba, V.: Observed fingerprint of a weakening Atlantic Ocean overturning circulation, Nature, 556, 191–196, 2018.

Caesar, L., McCarthy, G., Thornalley, D., Cahill, N., and Rahmstorf, S.: Current Atlantic meridional overturning circulation weakest in last millennium, Nature Geoscience, 14, 118–120, 2021.

Cahill, N., Kemp, A. C., Horton, B. P., and Parnell, A. C.: Modeling sea-level change using errors-in-variables integrated Gaussian processes, 2015.

Cressie, N. and Wikle, C. K.: Statistics for spatio-temporal data, John Wiley & Sons, 2015.

DeConto, R. M. and Pollard, D.: Contribution of Antarctica to past and future sea-level rise, Nature, 531, 591–597, 2016.

Engelhart, S. E. and Horton, B. P.: Holocene sea level database for the Atlantic coast of the United States, Quaternary Science Reviews, 54, 12–25, 2012.

Engelhart, S. E., Horton, B. P., Douglas, B. C., Peltier, W. R., and Törnqvist, T. E.: Spatial variability of late Holocene and 20th century sea-level rise along the Atlantic coast of the United States, Geology, 37, 1115–1118, 2009.

Gowan, E. J., Zhang, X., Khosravi, S., Rovere, A., Stocchi, P., Hughes, A. L., Gyllencreutz, R., Mangerud, J., Svendsen, J.-I., and Lohmann, G.: A new global ice sheet reconstruction for the past 80 000 years, Nature communications, 12, 1–9, 2021.

Hastings, W. K.: Monte Carlo sampling methods using Markov chains and their applications, 1970.

Hay, C. C., Morrow, E., Kopp, R. E., and Mitrovica, J. X.: Probabilistic reanalysis of twentieth-century sea-level rise, Nature, 517, 481–484, 2015.

Heaton, T. J., Köhler, P., Butzin, M., Bard, E., Reimer, R. W., Austin, W. E., Ramsey, C. B., Grootes, P. M., Hughen, K. A., Kromer, B., et al.: Marine20—the marine radiocarbon age calibration curve (0–55,000 cal BP), Radiocarbon, 62, 779–820, 2020.

Hensman, J., Fusi, N., and Lawrence, N. D.: Gaussian processes for big data, arXiv preprint arXiv:1309.6835, 2013.

Hibbert, F. D., Rohling, E. J., Dutton, A., Williams, F. H., Chutcharavan, P. M., Zhao, C., and Tamisiea, M. E.: Coral indicators of past sea-level change: A global repository of U-series dated benchmarks, Quaternary Science Reviews, 145, 1–56, 2016.

Hinestrosa, G., Webster, J. M., and Beaman, R. J.: New constraints on the postglacial shallow-water carbonate accumulation in the Great Barrier Reef, Scientific Reports, 12, 1–18, 2022.

https://doi.org/10.5194/egusphere-2024-2183
Preprint. Discussion started: 16 October 2024

Hoffman, M. D., Gelman, A., et al.: The No-U-Turn sampler: adaptively setting path lengths in Hamiltonian Monte Carlo., J. Mach. Learn. Res., 15, 1593–1623, 2014.

Holden, P. B., Edwards, N. R., Rangel, T. F., Pereira, E. B., Tran, G. T., and Wilkinson, R. D.: PALEO-PGEM v1. 0: a statistical emulator of Pliocene–Pleistocene climate, Geoscientific Model Development, 12, 5137–5155, 2019.

Hunter, J. D.: Matplotlib: A 2D graphics environment, Computing in Science & Engineering, 9, 90–95, https://doi.org/10.1109/MCSE.2007.55, 2007.

Kageyama, M., Braconnot, P., Harrison, S. P., Haywood, A. M., Jungclaus, J. H., Otto-Bliesner, B. L., Peterschmitt, J.-Y., Abe-Ouchi, A., Albani, S., Bartlein, P. J., et al.: The PMIP4 contribution to CMIP6–Part 1: Overview and over-arching analysis plan, Geoscientific Model Development, 11, 1033–1057, 2018.

Kay, R. F., Vizcaíno, S. F., Bargo, M. S., Spradley, J. P., and Cuitiño, J. I.: Paleoenvironments and paleoecology of the Santa Cruz Formation (early-middle Miocene) along the Río Santa Cruz, Patagonia (Argentina), Journal of South American Earth Sciences, 109, 103 296, 2021.

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., Horton, B. P., Vane, C. H., Bernhardt, C. E., Corbett, D. R., 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, 2013.

Kemp, A. C., Bernhardt, C. E., Horton, B. P., Kopp, R. E., Vane, C. H., Peltier, W. R., Hawkes, A. D., Donnelly, J. P., Parnell, A. C., and Cahill, N.: Late Holocene sea-and land-level change on the US southeastern Atlantic coast, Marine Geology, 357, 90–100, 2014.

Kemp, A. C., Hawkes, A. D., Donnelly, J. P., Vane, C. H., Horton, B. P., Hill, T. D., Anisfeld, S. C., Parnell, A. C., and Cahill, N.: Relative sea-level change in Connecticut (USA) during the last 2200 yrs, Earth and Planetary Science Letters, 428, 217–229, 2015.

Kemp, A. C., Hill, T. D., Vane, C. H., Cahill, N., Orton, P. M., Talke, S. A., Parnell, A. C., Sanborn, K., and Hartig, E. K.: Relative sea-level trends in New York City during the past 1500 years, The Holocene, 27, 1169–1186, 2017a.

Kemp, A. C., Kegel, J. J., Culver, S. J., Barber, D. C., Mallinson, D. J., Leorri, E., Bernhardt, C. E., Cahill, N., Riggs, S. R., Woodson, A. L., et al.: Extended late Holocene relative sea-level histories for North Carolina, USA, Quaternary Science Reviews, 160, 13–30, 2017b.

Kemp, A. C., Wright, A. J., Edwards, R. J., Barnett, R. L., Brain, M. J., Kopp, R. E., Cahill, N., Horton, B. P., Charman, D. J., Hawkes, A. D., et al.: Relative sea-level change in Newfoundland, Canada during the past 3000 years, Quaternary Science Reviews, 201, 89–110, 2018.

Kendall, R. A., Mitrovica, J. X., and Milne, G. A.: On post-glacial sea level–II. Numerical formulation and comparative results on spherically symmetric models, Geophysical Journal International, 161, 679–706, 2005.

Khan, N. S., Ashe, E., Horton, B. P., Dutton, A., Kopp, R. E., Brocard, G., Engelhart, S. E., Hill, D. F., Peltier, W., Vane, C. H., et al.: Drivers of Holocene sea-level change in the Caribbean, Quaternary Science Reviews, 155, 13–36, 2017.

Khan, N. S., Horton, B. P., Engelhart, S., Rovere, A., Vacchi, M., Ashe, E. L., Törnqvist, T. E., Dutton, A., Hijma, M. P., and Shennan, I.: Inception of a global atlas of sea levels since the Last Glacial Maximum, Quaternary Science Reviews, 220, 359–371, 2019.

Khan, N. S., Ashe, E., Moyer, R. P., Kemp, A. C., Engelhart, S. E., Brain, M. J., Toth, L. T., Chappel, A., Christie, M., Kopp, R. E., et al.: Relative sea-level change in South Florida during the past˜ 5000 years, Global and Planetary Change, 216, 103 902, 2022.

Kingma, D. P. and Ba, J.: Adam: A method for stochastic optimization, arXiv preprint arXiv:1412.6980, 2014.

Kopp, R. E., Simons, F. J., Mitrovica, J. X., Maloof, A. C., and Oppenheimer, M.: Probabilistic assessment of sea level during the last interglacial stage, Nature, 462, 863–867, 2009.

Kopp, R. E., Horton, R. M., Little, C. M., Mitrovica, J. X., Oppenheimer, M., Rasmussen, D., Strauss, B. H., and Tebaldi, C.: Probabilistic 21st and 22nd century sea-level projections at a global network of tide-gauge sites, Earth's future, 2, 383–406, 2014.

https://doi.org/10.5194/egusphere-2024-2183
Preprint. Discussion started: 16 October 2024

Kopp, R. E., Kemp, A. C., Bittermann, K., Horton, B. P., Donnelly, J. P., Gehrels, W. R., Hay, C. C., Mitrovica, J. X., Morrow, E. D., and Rahmstorf, S.: Temperature-driven global sea-level variability in the Common Era, Proceedings of the National Academy of Sciences, 113, E1434–E1441, 2016.

Kopp, R. E., Garner, G. G., Hermans, T. H., Jha, S., Kumar, P., Reedy, A., Slangen, A. B., Turilli, M., Edwards, T. L., Gregory, J. M., et al.: The Framework for Assessing Changes To Sea-level (FACTS) v1. 0: a platform for characterizing parametric and structural uncertainty in future global, relative, and extreme sea-level change, Geoscientific Model Development, 16, 7461–7489, 2023.

Lambeck, K., Rouby, H., Purcell, A., Sun, Y., and Sambridge, M.: Sea level and global ice volumes from the Last Glacial Maximum to the Holocene, Proceedings of the National Academy of Sciences, 111, 15 296–15 303, 2014.

Li, T., Wu, P., Wang, H., Steffen, H., Khan, N. S., Engelhart, S. E., Vacchi, M., Shaw, T. A., Peltier, W. R., and Horton, B. P.: Uncertainties of glacial isostatic adjustment model predictions in North America associated with 3D structure, Geophysical Research Letters, 47, e2020GL087 944, 2020.

Lin, Y., Hibbert, F. D., Whitehouse, P. L., Woodroffe, S. A., Purcell, A., Shennan, I., and Bradley, S. L.: A reconciled solution of Meltwater Pulse 1A sources using sea-level fingerprinting, Nature communications, 12, 1–11, 2021.

Lin, Y., Whitehouse, P. L., Hibbert, F. D., Woodroffe, S. A., Hinestrosa, G., and Webster, J. M.: Relative sea level response to mixed carbonate-siliciclastic sediment loading along the Great Barrier Reef margin, Earth and Planetary Science Letters, 607, 2023a.

Lin, Y., Whitehouse, P. L., Valentine, A. P., and Woodroffe, S. A.: GEORGIA: A graph neural network based EmulatOR for glacial isostatic adjustment, Geophysical Research Letters, 50, e2023GL103 672, 2023b.

Lin, Y., Kopp, R., Reedy, A., Jha, S., and Turilli, M.: PaleoSTeHM v1.0: a modern, scalable spatio-temporal hierarchical modeling framework for paleo-environmental data (v1.0), https://doi.org/10.5281/zenodo.12730140, 2024.

Lyu, K., Zhang, X., Church, J. A., Slangen, A. B., and Hu, J.: Time of emergence for regional sea-level change, Nature Climate Change, 4, 1006–1010, 2014.

Mann, M. E., Zhang, Z., Hughes, M. K., Bradley, R. S., Miller, S. K., Rutherford, S., and Ni, F.: Proxy-based reconstructions of hemispheric and global surface temperature variations over the past two millennia, Proceedings of the National Academy of Sciences, 105, 13 252–13 257, 2008.

McHutchon, A. and Rasmussen, C.: Gaussian process training with input noise, Advances in neural information processing systems, 24, 2011.

Meltzner, A. J., Switzer, A. D., Horton, B. P., Ashe, E., Qiu, Q., Hill, D. F., Bradley, S. L., Kopp, R. E., Hill, E. M., Majewski, J. M., et al.: Half-metre sea-level fluctuations on centennial timescales from mid-Holocene corals of Southeast Asia, Nature Communications, 8, 14 387, 2017.

Neal, R. M. et al.: MCMC using Hamiltonian dynamics, Handbook of markov chain monte carlo, 2, 2, 2011.

PAGES2k Consortium: A global multiproxy database for temperature reconstructions of the Common Era, Scientific data, 4, 2017.

PAGES2k Consortium: Consistent multidecadal variability in global temperature reconstructions and simulations over the Common Era, Nature Geoscience, 12, 643–649, https://doi.org/10.1038/s41561-019-0400-0, 2019.

Paszke, A., Gross, S., Chintala, S., Chanan, G., Yang, E., DeVito, Z., Lin, Z., Desmaison, A., Antiga, L., and Lerer, A.: Automatic differentiation in PyTorch, 2017.

Peltier, W., Argus, D., and Drummond, R.: Space geodesy constrains ice age terminal deglaciation: The global ICE-6G_C (VM5a) model, Journal of Geophysical Research: Solid Earth, 120, 450–487, 2015.

https://doi.org/10.5194/egusphere-2024-2183
Preprint. Discussion started: 16 October 2024

Piecuch, C. G., Huybers, P., and Tingley, M. P.: Comparison of full and empirical B ayes approaches for inferring sea-level changes from tide-gauge data, Journal of Geophysical Research: Oceans, 122, 2243–2258, 2017.

Pollack, A., Campbell, J. E., Condon, M., Cooper, C., Coronese, M., Doss-Gollin, J., Hegde, P., Helgeson, C., Kwakkel, J., Lesk, C., et al.: Peer-reviewed climate change research has a transparency problem. The scientific community needs to do better, 2024.

Poppick, A. and Stein, M. L.: Using covariates to model dependence in nonstationary, high-frequency meteorological processes, Environmetrics, 25, 293–305, 2014.

Quinonero-Candela, J. and Rasmussen, C. E.: A unifying view of sparse approximate Gaussian process regression, The Journal of Machine Learning Research, 6, 1939–1959, 2005.

Rasmussen, C. E. and Williams, C. K.: Gaussian processes for machine learning, vol. 2, MIT press Cambridge, MA, 2006.

Reichstein, M., Camps-Valls, G., Stevens, B., Jung, M., Denzler, J., Carvalhais, N., et al.: Deep learning and process understanding for data-driven Earth system science, Nature, 566, 195–204, 2019.

Reimer, P. J., Austin, W. E., Bard, E., Bayliss, A., Blackwell, P. G., Ramsey, C. B., Butzin, M., Cheng, H., Edwards, R. L., Friedrich, M., et al.: The IntCal20 Northern Hemisphere radiocarbon age calibration curve (0–55 cal kBP), Radiocarbon, 62, 725–757, 2020.

Roy, K. and Peltier, W.: Relative sea level in the Western Mediterranean basin: A regional test of the ICE-7G_NA (VM7) model and a constraint on late Holocene Antarctic deglaciation, Quaternary Science Reviews, 183, 76–87, 2018.

Shakun, J. D. and Carlson, A. E.: A global perspective on Last Glacial Maximum to Holocene climate change, Quaternary Science Reviews, 29, 1801–1816, 2010.

Simpson, G. L.: Modelling palaeoecological time series using generalised additive models, Frontiers in Ecology and Evolution, 6, 149, 2018.

Small, D., Bentley, M. J., Jones, R. S., Pittard, M. L., and Whitehouse, P. L.: Antarctic ice sheet palaeo-thinning rates from vertical transects of cosmogenic exposure ages, Quaternary Science Reviews, 206, 65–80, 2019.

Stein, M. L.: Nonstationary spatial covariance functions, Unpublished technical report, 2005a.

Stein, M. L.: Space–time covariance functions, Journal of the American Statistical Association, 100, 310–321, 2005b.

Stein, M. L.: Interpolation of spatial data: some theory for kriging, Springer Science & Business Media, 2012.

Tan, F., Khan, N. S., Li, T., Meltzner, A. J., Majewski, J., Chan, N., Chutcharavan, P. M., Cahill, N., Vacchi, M., Peng, D., et al.: Holocene relative sea-level histories of far-field islands in the mid-Pacific, Quaternary Science Reviews, 310, 107 995, 2023.

Tang, Y., Zheng, Z., Huang, K., Chen, C., Chen, Z., Lu, H., Wu, W., Lin, X., Zhang, X., and Li, H.: Holocene Evolution of the Pearl River Delta: Mapping Integral Isobaths and Delta Progradation, Journal of Marine Science and Engineering, 11, 1986, 2023.

Tarasov, L., Dyke, A. S., Neal, R. M., and Peltier, W. R.: A data-calibrated distribution of deglacial chronologies for the North American ice complex from glaciological modeling, Earth and Planetary Science Letters, 315, 30–40, 2012.

Tingley, M. P. and Huybers, P.: Recent temperature extremes at high northern latitudes unprecedented in the past 600 years, Nature, 496, 201–205, 2013.

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.

Turner, F. E., Buck, C. E., Jones, J. M., Sime, L. C., Vallet, I. M., and Wilkinson, R. D.: Reconstructing the Antarctic ice-sheet shape at the Last Glacial Maximum using ice-core data, Journal of the Royal Statistical Society Series C: Applied Statistics, 72, 1493–1511, 2023.

Vacchi, M., Joyse, K. M., Kopp, R. E., Marriner, N., Kaniewski, D., and Rovere, A.: Climate pacing of millennial sea-level change variability in the central and western Mediterranean, Nature communications, 12, 4013, 2021.

https://doi.org/10.5194/egusphere-2024-2183
Preprint. Discussion started: 16 October 2024

Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., van der Walt, S. J., Brett, M., Wilson, J., Millman, K. J., Mayorov, N., Nelson, A. R. J., Jones, E., Kern, R., Larson, E., Carey, C. J., Polat, İ., Feng, Y., Moore, E. W., VanderPlas, J., Laxalde, D., Perktold, J., Cimrman, R., Henriksen, I., Quintero, E. A., Harris, C. R., Archibald, A. M., Ribeiro, A. H., Pedregosa, F., van Mulbregt, P., and SciPy 1.0 Contributors: SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python, Nature Methods, 17, 261–272, https://doi.org/10.1038/s41592-019-0686-2, 2020.

Walker, J. S., Kopp, R. E., Shaw, T. A., Cahill, N., Khan, N. S., Barber, D. C., Ashe, E. L., Brain, M. J., Clear, J. L., Corbett, D. R., et al.: Common Era sea-level budgets along the US Atlantic coast, Nature Communications, 12, 1841, 2021.

Walker, J. S., Kopp, R. E., Little, C. M., and Horton, B. P.: Timing of emergence of modern rates of sea-level rise by 1863, Nature Communications, 13, 966, 2022.

Wang, K., Pleiss, G., Gardner, J., Tyree, S., Weinberger, K. Q., and Wilson, A. G.: Exact Gaussian processes on a million data points, Advances in neural information processing systems, 32, 2019.

Wang, Z., Saito, Y., Zhan, Q., Nian, X., Pan, D., Wang, L., Chen, T., Xie, J., Li, X., and Jiang, X.: Three-dimensional evolution of the Yangtze River mouth, China during the Holocene: impacts of sea level, climate and human activity, Earth-Science Reviews, 185, 938–955, 2018.

Whitehouse, P. L.: Glacial isostatic adjustment modelling: historical perspectives, recent advances, and future directions, Earth surface dynamics, 6, 401–429, 2018.

Wilks, S. S.: The large-sample distribution of the likelihood ratio for testing composite hypotheses, The annals of mathematical statistics, 9, 60–62, 1938.

Yokoyama, Y., Esat, T. M., Thompson, W. G., Thomas, A. L., Webster, J. M., Miyairi, Y., Sawada, C., Aze, T., Matsuzaki, H., Okuno, J., et al.: Rapid glaciation and a two-step sea level plunge into the Last Glacial Maximum, Nature, 559, 603–607, 2018.

https://doi.org/10.5194/egusphere-2024-2183
Preprint. Discussion started: 16 October 2024

Appendix A: Additional model information

Table A1. Summary of Model Characteristics. The posterior is reported with a mean value with 90% credible interval. NJ = New Jersey, NC = North Carolina, EIV = errors in variable, NI = noisy input.

Task	Analysis Choice	Data Level	Process Level	Parameter Level	Posterior
GBR coral time series	Fully Bayesian; EIV	Uniform likelihood with additional white noise	Temporally linear	$\alpha \sim U(-30, 30)$ m $\beta \sim U(-10, 10)$ mm/yr $\omega \sim U(0.0001, 10)$ m	-6.3 (-28.3, 16.2) 5.9 (4.5, 7.4) 0.8 (0.1, 2.0)
GBR coral time series	Fully Bayesian; EIV	Normal likelihood with additional white noise	Temporally linear	$\alpha \sim U(-30, 30)$ m $\beta \sim U(-10, 10)$ mm/yr $\omega \sim U(0.0001, 10)$ m	-0.2 (-22.8, 21.9) 6.3 (4.8, 7.7) 1.4 (0.01, 1.97)
NJ salt marsh time series	Variational Bayesian; EIV	Normal likelihood	Temporally linear	$\alpha \sim U(-5, 5)$ m $\beta \sim U(-10, 10)$ mm/yr	-3.38 (-3.47, -3.30) 1.63 (1.57, 1.69)
NJ salt marsh time series	Variational Bayesian; EIV	Normal likelihood	Change-point model	$\alpha_1 \sim U(-15, 0)$ m $\beta_1 \sim U(-10, 10)$ mm/yr $\beta_2 \sim U(-10, 10)$ mm/yr $\beta_3 \sim U(-10, 10)$ mm/yr $\beta_4 \sim U(-10, 10)$ mm/yr $\gamma_2 \sim U(-476, 1020)$ CE $\gamma_3 \sim U(23, 1518)$ CE $\gamma_4 \sim U(521, 2017)$ CE	-4.92 (-5.01, -4.84) 1.6 (1.5, 1.8) 1.53 (1.46, 1.60) 1.31 (1.22, 1.39) 3.92 (3.73, 4.10) -42.0 (-139.6, 58.2) 1004.3 (893.1, 1106.0) 1838.0 (1823.6, 1853.2)
NJ salt marsh time series	Empirical Bayesian; NI	Normal likelihood	Gaussian Process with one RBF kernel	$\ell \sim U(1, 5000)$ yr $\sigma \sim U(1, 22.4)$ m	1038 20.79
NC salt marsh time series	Empirical Bayesian; NI	Normal likelihood	Gaussian Process with one RBF kernel	$\ell \sim U(1, 10000)$ yr $\sigma \sim U(1, 100)$ m	2175 1.75
NC salt marsh time series	Fully Bayesian EIV	Normal likelihood	Gaussian Process with one RBF kernel	$\ell \sim U(1, 10000)$ yr $\sigma \sim U(1, 100)$ m	6889 (2845, 9766) 12.5 (3.4, 28.1)
NC salt marsh time series	Variational Bayesian; NI	Normal likelihood	Gaussian Process with one RBF kernel	$\ell \sim U(1, 10000)$ yr $\sigma \sim U(1, 100)$ m	7305 (3559, 9509) 10.4 (3.8, 18.4)

https://doi.org/10.5194/egusphere-2024-2183
Preprint. Discussion started: 16 October 2024

Task	Analysis Choice	Data Level	Process Level	Parameter Level	Posterior
US Atlantic spatio-temporal analysis	Empirical Bayesian; NI	Normal likelihood with additional white noise	Gaussian Process with a zero mean function and multiple isotropic kernels	$\omega \sim U(0.01, 10)$ m	0.02
				$\ell_g \sim U(100, 20000)$ yr	11567
				$\sigma_g \sim U(0.01, 33.3)$ m	30.7
				$\sigma_r \sim U(0.2, 10)$ m	1.7
				$\ell_{r,x} \sim U(319, 1593)$ km	345
				$\ell_{r,t} \sim U(500, 5000)$ yr	3254
				$\sigma_l \sim U(0.1, 3.3)$ m	0.14
				$\ell_{l,x} \sim U(64, 319)$ km	317.4
				$\ell_{l,t} \sim U(100, 2000)$ yr	1978
US Atlantic spatio-temporal analysis	Empirical Bayesian; NI	Normal likelihood with additional white noise	Gaussian Process with ICE7G as mean function and multiple isotropic kernels	$\omega \sim U(0.01, 10)$ m	0.02
				$\sigma_r \sim U(0.2, 10)$ m	6.0
				$\ell_{r,x} \sim U(319, 1593)$ km	1586
				$\ell_{r,t} \sim U(500, 5000)$ yr	4684
				$\sigma_l \sim U(0.1, 3.3)$ m	0.13
				$\ell_{l,x} \sim U(64, 319)$ km	312
				$\ell_{l,t} \sim U(100, 2000)$ yr	1990
US Atlantic spatio-temporal analysis	Empirical Bayesian; NI	Normal likelihood with additional white noise	Gaussian Process with zero mean and a sampling kernel determined by a GIA model ensemble	$\omega \sim U(0.01, 10)$ m	0.3
US Atlantic spatio-temporal analysis	Variational Bayesian; NI	Normal likelihood with additional white noise	A GIA model ensemble (consist of 9 individual models)	$\omega \sim U(0.01, 0.5)$	0.498 (0.497, 0.498)
				$\nu_1 \sim Beta(0.22, 0.78)$	0.04 (0.03, 0.05)
				$\nu_2 \sim Beta(0.02, 0.98)$	0. (0., 0.)
				$\nu_3 \sim Beta(0.02, 0.98)$	0. (0., 0.)
				$\nu_4 \sim Beta(0.09, 0.91)$	0. (0., 0.)
				$\nu_5 \sim Beta(0.01, 0.99)$	0. (0., 0.)
				$\nu_6 \sim Beta(0.02, 0.98)$	0. (0., 0.)
				$\nu_7 \sim Beta(0.16, 0.84)$	0.38 (0.35, 0.42)
				$\nu_8 \sim Beta(0.29, 0.71)$	0.36 (0.33, 0.38)
				$\nu_9 \sim Beta(0.17, 0.83)$	0.22 (0.20, 0.24)

https://doi.org/10.5194/egusphere-2024-2183
Preprint. Discussion started: 16 October 2024

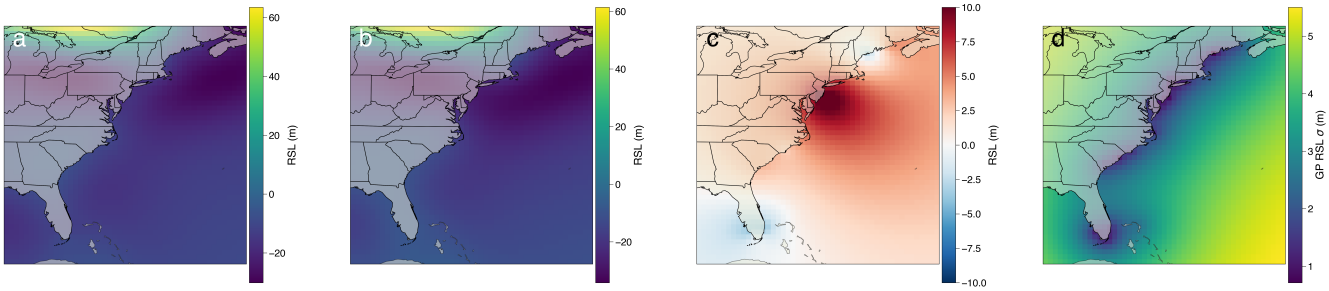

Figure A1. Model *ii* prediction on -5500 CE RSL at the US Atlantic coast. (a) Mean relative sea-level prediction, which is a combination of mean function (b) and covariance function (c). (b) Relative sea-level prediction by Gaussian Process mean function (described by a glacial isostatic adjustment model with ICE_7G ice model and VM5a Earth model). (c) Covariance function induced mean relative sea-level change. (d) Covariance function induced relative sea-level standard deviation.