
Spatiotemporal modeling of European paleoclimate using doubly sparse Gaussian processes

Seth D. Axen^{1*} Alexandra Gessner¹ Christian Sommer²
Nils Weitzel³ Álvaro Tejero-Cantero¹

¹Cluster of Excellence Machine Learning ²Heidelberg Academy of Sciences and Humanities

³Department of Geosciences
University of Tübingen, Germany

Abstract

Paleoclimatology—the study of past climate—is relevant beyond climate science itself, such as in archaeology and anthropology for understanding past human dispersal. Information about the Earth’s paleoclimate comes from simulations of physical and biogeochemical processes and from proxy records found in naturally occurring archives. Climate-field reconstructions (CFRs) combine these data into a statistical spatial or spatiotemporal model. To date, there exists no consensus spatiotemporal paleoclimate model that is continuous in space and time, produces predictions with uncertainty, and can include data from various sources. A Gaussian process (GP) model would have these desired properties; however, GPs scale unfavorably with data of the magnitude typical for building CFRs. We propose to build on recent advances in sparse spatiotemporal GPs that reduce the computational burden by combining variational methods based on inducing variables with the state-space formulation of GPs. We successfully employ such a doubly sparse GP to construct a probabilistic model of European paleoclimate from the Last Glacial Maximum (LGM) to the mid-Holocene (MH) that synthesizes paleoclimate simulations and fossilized pollen proxy data.

1 Introduction

Paleoclimate reconstructions are important for understanding climate processes and variability but are also valuable to other scientific domains such as archaeology or paleoecology. For example, past climate variability has been a driver for human migration, adaption, and cultural innovation [1, 2, 3]. For reconstructing the climate of the past, there are two main sources of information. The first source is the known dynamics of the climate system encoded as a climate model. Paleoclimate states are generated from these models through numerical simulations under reconstructed or theorized boundary conditions called forcings. The simulation outputs are typically represented as a spatial grid of values of climate variables like mean annual temperature or total annual precipitation. Some simulations estimate climate variables at one or a few relevant time slices, while others do so at regular intervals (Fig. B.1). The second source of information is proxy data from geolocated natural archives that contain information about the time of their deposition, where the proxy data informs the corresponding climate state at that time (Fig. B.1). Such archives include ice cores, mineral deposits in caves, and lake sediment cores. Fossilized pollen grains, found within different types of archives, are the most abundant terrestrial proxy due in part to their comparatively high durability. These pollen grains are classified into taxonomic groups and compared to modern species distributions. Relating the modern taxa to modern climate allows one to generate paleoclimate reconstructions with existing techniques [4, 5, 6]. Because pollen are found at archaeological sites worldwide, they are particularly compelling for studying the environmental context of human evolution.

*E-mail: seth@sethaxen.com

Climate field reconstructions (CFRs) infer spatial or spatiotemporal fields of climate variables with statistical models. Most Bayesian CFRs consider only the more data-rich recent past, i.e. the last two thousand years (2 ka), and most of these use information only from proxies and not from simulations [7, 8]. Purely spatial probabilistic reconstructions of paleoclimate from simulations and pollen have been built by Weitzel et al. [9, 10]. Osman et al. [11] used ensemble Kalman filtering to model surface temperature based on multiple simulation time slices from one climate model and sea surface temperature proxy data from the LGM to the present; this study did not explicitly model temporal autocorrelations. To our knowledge, the only fully Bayesian spatiotemporal model for CFR of paleoclimate has been presented by Weitzel [12] but applied only to synthetic data. Their parametric hierarchical model has $O(10^5)$ parameters, which makes inference challenging.

Currently there is no consensus CFR available that is global, spatiotemporal, probabilistic, uses the majority of available proxy data and simulations, and can be queried at an arbitrary location and time. Our goal is ultimately to construct such a model and make it available to the research community. As a first step, we present here a spatiotemporal Gaussian process model of mean annual temperature across Europe from the *Last Glacial Maximum* (LGM; 21 ka) to the *mid-Holocene* (MH; 6 ka) based on simulations from various paleoclimate models and reconstructions from fossilized pollen (cf. Section 3). Inference on such a large dataset ($N > 6 \cdot 10^5$, cf. Section 2) is made tractable using a combination of inducing variables and a Markovian structure in the temporal domain [13, 14], using *doubly sparse* GPs (cf. Section 4). With only 100 spatial and six temporal inducing points, our model interpolates well in space and time between data points and comes with calibrated uncertainty estimates.

2 Dataset

In this work we focus on mean annual temperature as the climate variable of interest using fossilized pollen grains as proxy data. We constrain the spatial domain to the European continent during the period from LGM to MH. All data is from publicly available databases.

Pollen data We use 41,986 site-specific reconstructions of mean annual temperature from fossilized pollen collected at 826 sites and 1,607 time slices from the LegacyClimate 1.0 database [15].

Simulation data We use a total of ten models with different output coverage in space and time. Three models have been simulated only at MH [16, 17, 18], while five are available at both LGM and MH but no other time slices [19, 20, 21, 22, 23]. All of the above have a spatial grid spacing of 10 arcminutes (< 15 km) and were downloaded from WorldClim 1.4 [24]. Two models have been simulated at time slices spaced by 1,000 years ranging from LGM to MH and have a grid spacing of 30 arcminutes (< 46 km) [25, 26].

Simulations and pollen-derived reconstructions provide a total of $N = 661,028$ data points. We preprocessed the data by using radial basis functions to spatially interpolate within each lower-resolution MH simulation; these interpolations were then averaged to form a function $m(x)$. This function $m(x)$ was then subtracted from all data; this smoothed out the steep gradients in climate variables around mountain ranges and the coastline. Thus the data we use during modeling is the deviation from this empirical model.

3 Description of probabilistic model

Prior Let $C(x; t)$ be a climate variable of interest at coordinates $x = (\text{longitude}; \text{latitude})$ and time t . We adopt the following zero-mean, factorized GP prior for $\bar{C}(x; t) := C(x; t) - m(x)$:

$$\bar{C}(x; t) \sim GP(0; k_x(x; x')k_t(t; t'));$$

where k_x is a Matérn- $3/2$ spatial kernel with different latitudinal and longitudinal length scales, and k_t is a Matérn- $1/2$ temporal kernel corresponding to an Ornstein-Uhlenbeck process, whose state-space representation has dimension $d = 1$.

Likelihood Let $Y_{s;p}$ be either a site-specific reconstruction or a simulated value at site s with coordinates x_s and time t_p . Furthermore, let $\bar{Y}_{s;p} = Y_{s;p} - m(x_s)$. We approximate for now that all data are independently and identically (i.i.d.) normally distributed

$$\bar{Y}_{s;p} \mid \bar{C}(x_s; t_p) \sim N(\bar{C}(x_s; t_p); \sigma^2);$$

(a)

(b)

Figure 1: Predictive error of model $\hat{C}(x; t)$. (a) Results of validation when fitting the model to a single multi-time step simulation [26]. For each model t , a single time slice was held out from the simulation during training and compared with the corresponding posterior predictive (PP) means at the same time slice and coordinates. Error is computed as $(PP \text{ mean} - \text{data})$. Shown are the central intervals of the error distributions containing 80%, 60%, 40%, and 20% of values, as well as the median. (b) Kernel density estimate (KDE) of normalized posterior predictive (PP) error distribution. Error is computed as $(PP \text{ mean} - \text{data}) / (PP \text{ std})$. The KDE is computed using ArviZ [29].

The same is used for all simulations as well as the pollen-based reconstructions.

This model yields an exact GP posterior \bar{C} , which allows the resulting posterior distribution of C to be queried at any spatiotemporal point (x, t) . However, the large number of data points makes the $O(N^3)$ inversion of the kernel Gram matrix intractable. Therefore, we turn to a sparse GP approximation for inference.

4 Approximate inference

To keep inference computationally feasible, we use a sparse variational GP approximation that is compatible with the time-series structure of our model. Such algorithms are known as doubly sparse GP approximations because they combine two approaches to obtain sparsity: 1) a set of M inducing variables and 2) the state-space representation of Markovian GPs, which allows for linear-time inference using Kalman updates [13]. With a variational distribution on the inducing states that retains the chain structure, the marginal posterior predictions can be computed efficiently using filtering methods. For spatiotemporal models, these methods require a kernel that is separable into the product of a spatial kernel and a Markovian temporal kernel. Among multiple recent algorithmic approaches to sparse Markovian GPs [14] we choose $\mathcal{S}^2\text{CVI}$ (doubly sparse conjugate-computation variational inference) as implemented in markovflow [27]. $\mathcal{S}^2\text{CVI}$ defines a variational posterior $f_{\bar{C}}$ that is optimized using natural gradient updates and has been reported to be more efficient and numerically stable than alternative approaches [14]. The time complexity of evaluating the variational objective and its gradient using $\mathcal{S}^2\text{CVI}$ is $O((M_t + N_b)(M_s d)^3)$, for M_s spatial inducing points, M_t temporal inducing points, dimension d of state-space representation, and batch size N_b [14]. This method has not been used previously for problems of the scale of ours, and no public code exists of applications such as geospatiotemporal modeling where there is more than one spatial dimension.

Experimental details We trained $\mathcal{S}^2\text{CVI}$ with $M_s = 100$ spatial and $M_t = 6$ temporal inducing points over 30 epochs with a minibatch size of $N_b = 1;000$. At each iteration, the parameters of the variational posterior are first updated using natural gradients, and then all hyperparameters are updated using Adam [28]. The coordinates of the spatial inducing points, which are constrained to stay within the spatial domain of the data, are treated as hyperparameters. All predictions from the variational posterior $f_{\bar{C}}$ were shifted by $m(x)$ to obtain the approximate posterior $f_{\bar{C}}$, and these are used in all subsequent analyses.

5 Results

Validation We validated the general approach by performing “leave-one-time-slice-out” inference with a single multi-time slice simulation [26] as data (Fig. 1a). The average posterior predictive error

Figure 2: Comparison of mean annual temperature from one of the regularly temporally spaced simulations [26] \bar{Y} ; top) with the mean of the variational posterior approximation $\bar{Q}(x; t)$; bottom) at several time slices. At 0.5 ka, there is no simulated data available, so only the model's prediction is shown. The grey area in the simulation indicates grid points that are masked due to presence of ocean or an ice sheet. Because no observations are available in these regions, the model's predictions there are not expected to be reliable.

comparing each posterior predictive mean to the corresponding measurement in the held-out time slice across all time slices is 0.06 with a 95% central posterior predictive interval of (-1.86, 2.24 C). The mean absolute posterior predictive error is 0.69

Inference Training on the entire dataset took 36 hours on an NVIDIA V100 GPU. 98.9% of all data fell within three standard deviations of the mean of the posterior predictive distribution (99.1% for simulations and 95.9% for pollen; Fig. 1b). The mean of the variational posterior both spatially and temporally interpolates between the simulations (Fig. 2).

6 Outlook

We are working on several immediate improvements to the model. Instead of using a weighted interpolation of simulations as an empirical model to be subtracted from data, we will use a spatial interpolation of measurements from the early 1900s. To allow capturing remaining large-scale spatial and temporal trends, we will further introduce a non-zero parametric prior mean function to the prior on \bar{C} . We will also incorporate prior beliefs about the values of the fitted parameters by setting weakly informative priors. As a step toward extending the model to accommodate different data sources, we will use separate likelihood functions for proxy data and simulations. In the medium term, we would like to drop the i.i.d. assumption for simulations by including a sparse covariance matrix to account for the spatial covariance of the simulated data. Moreover, climate variables often co-vary, so we plan to jointly model mean annual temperature and total annual precipitation using a multi-output GP. Dating of proxies has its own uncertainty, which is especially relevant for pollen and is currently unaccounted for in our model; another extension would be to include this uncertainty in the variational objective [30, 31]. In the future we will expand the geographical extent of the model to provide predictions of climate with global land coverage between LGM and MH.

7 Conclusion

Paleoclimate modeling lacks consensus spatiotemporal models that combine simulations and proxies and adequately report uncertainty estimates. Progress in this area is hindered by the computational expense of training such models. We demonstrate that state-of-the-art sparse GP algorithms enable spatiotemporal modeling in realistic applications in the climate sciences where large datasets of different modalities are common. While the result already possesses the desired properties for a consensus model, we describe model improvements that will enhance its usefulness for geosciences communities. Our paleoclimate synthesis will ultimately be made available as a free, queryable web app to facilitate discovery in applied fields such as archaeology and paleoecology.

Acknowledgments and Disclosure of Funding

Funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy – EXC number 2064/1 – project number 390727645. NW acknowledges funding by the DFG, project number 395588486. CS acknowledges funding by the ROCEEH project (The Role of Culture in Early Expansions of Humans) of the Heidelberg Academy of Sciences. We thank Vincent Adam and Elizaveta Seminova for useful discussions.

References

- [1] Axel Timmermann et al. “Climate effects on archaic human habitats and species successions”. In: *Nature* 604.7906 (2022), pp. 495–501. ISSN: 1476-4687 URL: <https://www.nature.com/articles/s41586-022-04600-9>.
- [2] Axel Timmermann and Tobias Friedrich. “Late Pleistocene climate drivers of early human migration”. In: *Nature* 538.7623 (2016), pp. 92–95.
- [3] Peter B deMenocal. “Climate and human evolution”. *Science* 331.6017 (2011), pp. 540–542.
- [4] Patrick J Bartlein et al. “Pollen-based continental climate reconstructions at 6 and 21 ka: a global synthesis”. In: *Climate Dynamics* 37.3 (2011), pp. 775–802.
- [5] Christian Ohlwein and Eugene R Wahl. “Review of probabilistic pollen-climate transfer methods”. In: *Quaternary Science Reviews* 34 (2012), pp. 17–29.
- [6] Manuel Chevalier et al. “Pollen-based climate reconstruction techniques for late Quaternary studies”. In: *Earth-Science Reviews* 210 (2020), p. 103384.
- [7] Martin P Tingley et al. “Piecing together the past: statistical insights into paleoclimatic reconstructions”. In: *Quaternary Science Reviews* 35 (2012), pp. 1–22.
- [8] Gregory J Hakim et al. “The last millennium climate reanalysis project: Framework and first results”. In: *Journal of Geophysical Research: Atmospheres* 121.12 (2016), pp. 6745–6764.
- [9] Nils Weitzel, Andreas Hense, and Christian Ohlwein. “Combining a pollen and macrofossil synthesis with climate simulations for spatial reconstructions of European climate using Bayesian filtering”. In: *Climate of the Past* 15.4 (2019), pp. 1275–1301.
- [10] Nils Weitzel et al. A state estimate of Siberian summer temperature and moisture availability during the Last Glacial Maximum combining pollen records and climate simulation. 2022. URL: <https://eartharxiv.org/repository/view/3197>.
- [11] Matthew B. Osman et al. “Globally resolved surface temperatures since the Last Glacial Maximum”. In: *Nature* 599.7884 (Nov. 2021), pp. 239–244. ISSN: 1476-4687 URL: <https://doi.org/10.1038/s41586-021-03984-4>.
- [12] Nils Weitzel. “Climate field reconstructions from pollen and macrofossil syntheses using Bayesian hierarchical models”. PhD thesis. Rheinische Friedrich-Wilhelms-Universität Bonn, Jan. 2020 URL: <https://hdl.handle.net/20.500.11811/8249>.
- [13] Vincent Adam et al. “Doubly sparse variational Gaussian processes”. *International Conference on Artificial Intelligence and Statistics*. PMLR, 2020, pp. 2874–2884.
- [14] William Wilkinson, Arno Solin, and Vincent Adam. “Sparse Algorithms for Markovian Gaussian Processes”. In: *Proceedings of The 24th International Conference on Artificial Intelligence and Statistics*. Ed. by Arindam Banerjee and Kenji Fukumizu. Vol. 130. *Proceedings of Machine Learning Research*. PMLR, Apr. 2021, pp. 1747–1755. <https://proceedings.mlr.press/v130/wilkinson21a.html>.
- [15] Ulrike Herzschuh et al. “LegacyClimate 1.0: A dataset of pollen-based climate reconstructions from 2594 Northern Hemisphere sites covering the late Quaternary”. *Earth System Science Data Discussions* 2022 (2022), pp. 1–29. URL: <https://essd.copernicus.org/preprints/essd-2022-38/>.
- [16] Aurore Voldoire et al. “The CNRM-CM5.1 global climate model: description and basic evaluation”. In: *Climate Dynamics* 40.9 (May 2013), pp. 2091–2121. ISSN: 1432-0894 URL: <https://doi.org/10.1007/s00382-011-1259-y>.

- [17] Jean-Louis Dufresne et al. “Climate change projections using the IPSL-CM5 Earth System Model: from CMIP3 to CMIP5”. In: *Climate Dynamics* 40.9 (May 2013), pp. 2123–2165. ISSN: 1432-0894 URL: <https://doi.org/10.1007/s00382-012-1636-1>
- [18] Tongwen Wu et al. “The Beijing Climate Center Climate System Model (BCC-CSM): the main progress from CMIP5 to CMIP6”. In: *Geoscientific Model Development* 12.4 (2019), pp. 1573–1600 URL: <https://gmd.copernicus.org/articles/12/1573/2019/>
- [19] Peter R. Gent et al. “The Community Climate System Model Version 4”. *Journal of Climate* 24.19 (2011), pp. 4973–4990 URL: <https://journals.ametsoc.org/view/journals/clim/24/19/2011jcli4083.1.xml>
- [20] The HadGEM2 Development Team: G. M. Martin et al. “The HadGEM2 family of Met Office Unified Model climate configurations”. In: *Geoscientific Model Development* 4.3 (2011), pp. 723–757 URL: <https://gmd.copernicus.org/articles/4/723/2011/>
- [21] Seiji Yukimoto et al. “A New Global Climate Model of the Meteorological Research Institute: MRI-CGCM3 – Model Description and Basic Performance”. *Journal of the Meteorological Society of Japan. Ser. 10A* (2012), pp. 23–64.
- [22] Tetsuo Sueyoshi et al. “Set-up of the PMIP3 paleoclimate experiments conducted using an Earth system model, MIROC-ESM”. In: *Geoscientific Model Development* 6.3 (2013), pp. 819–836. URL: <https://gmd.copernicus.org/articles/6/819/2013/>
- [23] Johann H. Jungclauss, Katja Lohmann, and Davide Zanchettin. “Enhanced 20th-century heat transfer to the Arctic simulated in the context of climate variations over the last millennium”. In: *Climate of the Past* 10.6 (2014), pp. 2201–2218 URL: <https://cp.copernicus.org/articles/10/2201/2014/>
- [24] Robert J. Hijmans et al. “Very high resolution interpolated climate surfaces for global land areas”. In: *International Journal of Climatology* 25.15 (2005), pp. 1965–1978. eprint: <https://rmets.onlinelibrary.wiley.com/doi/pdf/10.1002/joc.1276> URL: <https://rmets.onlinelibrary.wiley.com/doi/abs/10.1002/joc.1276>
- [25] Robert M. Beyer, Mario Krapp, and Andrea Manica. “High-resolution terrestrial climate, bioclimate and vegetation for the last 120,000 years”. *Scientific Data* 7.1 (July 2020), p. 236. ISSN: 2052-4463 URL: <https://doi.org/10.1038/s41597-020-0552-1>
- [26] Mario Krapp et al. “A statistics-based reconstruction of high-resolution global terrestrial climate for the last 800,000 years”. *Scientific Data* 8.1 (Aug. 2021), p. 228. ISSN: 2052-4463. URL: <https://doi.org/10.1038/s41597-021-01009-3>
- [27] Vincent Adam et al. Markov flow. commit:fc82c0a. Sept. 5, 2022. URL: <https://github.com/secondmind-labs/markovflow>
- [28] Diederik P. Kingma and Jimmy Ba. Adam: A Method for Stochastic Optimization. 2014. URL: <https://arxiv.org/abs/1412.6980>
- [29] Ravin Kumar et al. “ArviZ: a unified library for exploratory analysis of Bayesian models in Python”. In: *Journal of Open Source Software* 4.33 (2019), p. 1143. URL: <https://doi.org/10.21105/joss.01143>
- [30] Andreas C. Damianou, Michalis K. Titsias, and Neil D. Lawrence. “Variational Inference for Latent Variables and Uncertain Inputs in Gaussian Processes”. *Journal of Machine Learning Research* 17.42 (2016), pp. 1–62. URL: <http://jmlr.org/papers/v17/damianou16a.html>
- [31] Mark van der Wilk et al. A framework for interdomain and multioutput Gaussian processes. 2020. URL: <https://arxiv.org/abs/2003.01115v1>

A Results of hyperparameter optimization

The spatial kernel of the GP prior k_x has two length scales in units of degrees: a longitudinal one λ_{lon} and a latitudinal one λ_{lat} . After optimization, these took the values $\lambda_{\text{lon}} = 19.6^\circ$ and $\lambda_{\text{lat}} = 13.2^\circ$. The temporal kernel k_t has a single length scale λ_t , which was optimized to $\lambda_t = 9,900$ years. The global standard deviation of the product kernel k was optimized to 2.9°C . The standard deviation of the likelihood was optimized to a value of 1.6°C . The optimized spatial inducing point locations are shown in Fig. A.1.

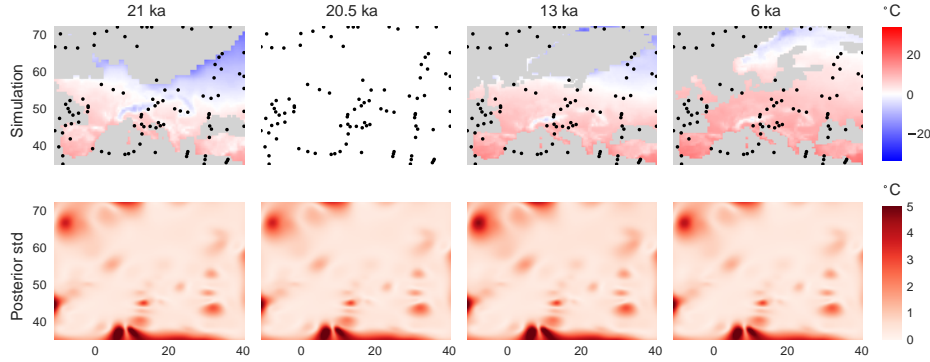


Figure A.1: Comparison of mean annual temperature from one of the regularly temporally spaced simulations [26] (*top*) with the standard deviation of the variational posterior approximation (*bottom*). The black dots in the top row are the optimized locations of the spatial inducing points, which are shared across all time slices.

B Overview of data and consensus model

Multiple spatiotemporally gridded simulations are combined with reconstructions from fossilized pollen proxies to construct a consensus model (Fig. B.1). While the data are discrete in space and time, the consensus model is continuous in both space and time, and at any spatiotemporal point, the marginal posterior distribution of temperature can be queried.

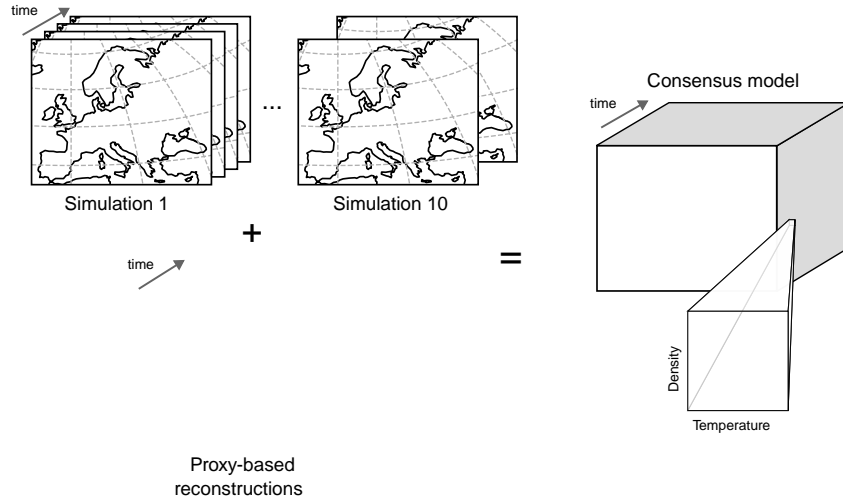


Figure B.1: Overview of data and consensus model. All colors represent temperature. For proxy-based reconstructions (*bottom*), each marker represents a reconstruction at a single spatiotemporal point. Multiple reconstructions at different points in time within a single archive appear stacked.