Environmental Modelling and Software 139 (2021) 105002

Contents lists available at ScienceDirect

Environmental Modelling and Software

journal homepage: http://www.elsevier.com/locate/envsoft

Precipitation reconstruction from climate-sensitive lithologies using Bayesian machine learning

Rohitash Chandra [d,b,*], Sally Cripps [b,d], Nathaniel Butterworth [c], R. Dietmar Muller [a]

[a] EarthByte Group, School of Geosciences, University of Sydney, NSW, 2006, Sydney, Australia
[b] Data Analytics for Resources and Environments, Australian Research Council - Industrial Transformation Training Centre, Australia
[c] Sydney Informatics Hub, University of Sydney, NSW, 2006, Sydney, Australia
[d] School of Mathematics and Statistics, University of Sydney, NSW, 2006, Sydney, Australia

ARTICLE INFO

Keywords:
Paleo-climate
Gaussian process
Bayesian methods
Forecasting
Precipitation

ABSTRACT

Although global circulation models (GCMs) have been used for the reconstruction of precipitation for selected geological time slices, there is a lack of a coherent set of precipitation models for the Mesozoic-Cenozoic period (the last 250 million years). There has been dramatic climate change during this time period capturing a supercontinent hothouse climate, and continental breakup and dispersal associated with successive greenhouse and ice-house climate periods. We present an approach that links climate-sensitive sedimentary deposits such as coal, evaporites and glacial deposits to a global plate model, reconstructed paleo-elevation maps and high-resolution GCMs via Bayesian machine learning. We model the joint distribution of climate-sensitive sediments and annual precipitation through geological time, and use the dependency between sediments and precipitation to improve the model's predictive accuracy. Our approach provides a set of 13 data-driven global paleo-precipitation maps between 14 and 249 Ma, capturing major changes in long-term annual rainfall patterns as a function of plate tectonics, paleo-elevation and climate change at a low computational cost.

1. Introduction

Palaeoclimatology refers to the study or reconstruction of ancient climates (Crowley and North, 1991; Bradley, 1999), often linked to the goal of understanding the current climate and its potential future trajectories (Hansen and Sato, 2012). The two primary variables used to define climate are temperature and precipitation. We focus on reconstructing the long-term history of precipitation, which is reflected in the geological record of climate-sensitive sedimentary deposits (Boucot et al., 2013a). Such a reconstruction involves several challenges. First, observational data constraining precipitation over geological time spans covering millions of years are sparse, both temporally and spatially (Boucot et al., 2013a). Second, the information from observational data must be fused together with knowledge of the geophysical processes in a logically consistent statistical framework or model (Birchfield et al., 1981; Crowley, 1988; Glancy et al., 1993; Patzkowsky et al., 1991; McGehee and Lehman, 2012; Stocker et al., 1992; Phipps et al., 2013; Ritz et al., 2011; Wang and Mysak, 2000; Contreras et al., 2019; Arıkan, 2015; Sellwood and Valdes, 2006). Third, the data is often noisy and becomes increasingly uncertain, the further we go back in time (Mann and Rutherford, 2002; Steiger et al., 2014; McIntyre and McKitrick, 2009). These characteristics increase levels of uncertainty about ancient climates, which must be accurately quantified for meaningful inference using the data and the model parameters.

The evolution of precipitation through geological time can be modelled using fully-coupled global circulation models (GCMs) (e.g. (Herold et al., 2011; Lunt et al., 2017; Baatsen et al., 2020)). However, a single model of this type for an individual geological time slice, typically takes several months to run on a high-performance computer. This limits the usefulness of this approach to develop models over geologic time. In addition, the preparation of initial and boundary conditions for such models is time-consuming. Only a limited number of geological time slices has been explored given the enormous computational resources for construction of a single model using GCMs. Some models focused on past hothouse climates, such as those in parts of the Miocene (Herold et al., 2011) and Eocene (Baatsen et al., 2020) periods. A major challenge in this area of research is developing improved methods to quantify climate model uncertainty. Combining climate proxies with Bayesian inference is seen as having great potential for assessing uncertainties and directly linking climate proxies with climate simulations

* Corresponding author. School of Mathematics and Statistics, University of Sydney, NSW, 2006, Sydney, Australia.
E-mail address: rohitash.chandra@unsw.edu.au (R. Chandra).

https://doi.org/10.1016/j.envsoft.2021.105002
Accepted 16 February 2021
Available online 25 February 2021

Table 1
Data description showing reconstruction timeslices (Ma) showing given precipitation (precip.) simulated data (number of grid samples) for Miocene and Eocene (Cao et al., 2019). The unavailable precipitation data is shown as not applicable (n/a) which our model will estimate. The number of locations for available deposits for coal, evaporites (eva.) and glacial (gla.) is shown which adds up to the present (N_p) number of deposit grids.

Era	Period	Epoch/Age	Timespan (Ma)	Timeslice (Ma)	Total (N_t)	Present (N_{D_t})	Precip.	Coal	Eva.	Gla.
Cenozoic	Neogene	Miocene	23.0–5.3	14	1763	335	1763	241	86	8
	Palaeogene	Oligocene	33.9–23.0	28	1761	282	n/a	229	52	1
		Eocene	47.8–33.9	38	1766	200	1766	146	50	4
		Early Eocene (Ypresian)	56.0–47.8	51	1748	278	n/a	219	58	1
		Palaeocene	66.0–56.0	61	1653	163	n/a	120	42	1
Mesozoic	Cretaceous	Late Cretaceous (Coniacian–Maastrichtian)	89.8–66.0	77	1490	236	n/a	151	85	0
		Late Cretaceous (Albian–Turonian)	113.0–89.8	101	1628	252	n/a	170	82	0
		Early Cretaceous (Berriasian–Aptian)	145.0–113.0	129	1650	292	n/a	185	102	5
	Jurassic	Late Jurassic	164.0–145.0	154	1630	180	n/a	85	95	0
		Early and Middle Jurassic	201.0–164.0	182	1675	330	n/a	249	81	0
	Triassic	Late Triassic	237.0–201.0	219	1731	217	n/a	142	75	0
		Middle Triassic	247.0–237.0	242	1594	84	n/a	21	63	0
		Early Triassic	252.0–247.0	249	1548	73	n/a	24	49	0

(Lunt et al., 2017). Although machine learning methods have been used in data-driven Earth system science (Reichstein et al., 2019) by coupling physical process models with observational data, the development of this field is still in its infancy. To our knowledge, a machine learning approach has not been used for the reconstruction of the distribution of climate belts and precipitation over geological time. This knowledge gap motivates our approach to fuse GCM precipitation simulations with proxies from the geological record to estimate paleo-precipitation back to the assembly of the supercontinent Pangea over 200 million years ago.

The prediction or reconstruction of precipitation through geological time is essential for understanding the evolution of Earth's climate, continental erosion and sedimentation, as well as the associated carbon cycle. Our work builds on previous studies investigating the connection between climate-sensitive lithologies and paleogeography, paleoclimate and paleoenvironments (Ziegler et al., 2003). Recently, Cao et al. (2019) reconstructed the most extensive available global-scale compilation of lithologic indicators of climate belts, including coals, evaporites and glacial deposits (Boucot et al., 2013a) and tested the sensitivity of their latitudinal distributions to the uneven distribution of continental areas through time. Using a purely data-driven approach, they were able to evaluate how the paleo-latitudinal distributions of climate-sensitive lithologies have changed through geological time in response to plate motions, mountain building, biological evolution and paleoclimate conditions. Given lithological indicators and additional global precipitation GCM simulations, we can extend this approach to use the same climate-sensitive indicators to develop estimates of paleo-precipitation.

Bayesian inference has been used extensively for the estimation and quantification of uncertainty in paleoclimate reconstructions (Haslett et al., 2006; Li et al., 2010; Tingley and Huybers, 2010; Carson et al., 2018; Ilvonen et al., 2016). The models for these reconstructions are often expressed in a hierarchical form; the top of the hierarchy specifies the model for observational data known as the likelihood, and the next layers in the hierarchy describe the latent processes which give rise to these observations. The latent processes have unknown parameters associated with them and the bottom layers of the hierarchy specify the prior distributions of these parameters (Tingley et al., 2012). Having specified a likelihood function and priors, inference regarding the quantities of interest proceeds via the posterior distribution. Obtaining this posterior distribution often involves performing a high dimensional integration, which has no closed-form analytic solution. As a result this distribution is usually approximated in two ways; 1.) sampling methods such as Markov chain Monte Carlo (MCMC) (Hastings, 1970; Geman and Geman, 1984; Gelfand and Smith, 1990; Chandra et al., 2019; Scalzo et al., 2019; Pall et al., 2020), and 2.) variational methods which provides more computationally efficient approximations (Jordan et al., 1999; Wainwright and Jordan, 2008).

Our goal is to reconstruct precipitation across the globe for selected time frames taken from the Miocene (14 million years ago) to 249 million years back in time. We present a Bayesian framework for inference to jointly model the predictive distributions of both precipitation and climate-sensitive deposits and place a Gaussian process (GP) prior over the spatial models for paleoclimate reconstruction. Our framework employs MCMC via Gibbs sampling for estimating the relevant posterior distributions such as the GP model hyperparameters as well as the joint predictive posterior distributions of the precipitation and deposits. We note that we have a spatiotemporal missing data problem where we have missing deposits (lithologies). Although it is common practice to impute missing values, such an approach ignores the uncertainty associated with these imputed values. To address this, we model the joint distribution of lithologies and precipitation and treat the missing values as unknown parameters. We account for the uncertainty surrounding the missing values by integrating (or marginalising) over all possible values using the MCMC framework.

2. Background

The Bayesian paradigm is a logically consistent framework to combine different pieces of information to make inference about unknown quantities via the posterior distribution. Information from expert opinion or prior knowledge can be incorporated via the prior distribution while information from data is incorporated via the likelihood function (Sen and Stoffa, 1996). The choice of likelihood function depends upon the type of observational data and assumptions regarding how the data is generated from the latent processes. The latent processes may be simple and deterministic; such as how the density of particular geological structures below the surface may impact gravity measurements made above the surface (Reid et al., 2013; Scalzo et al., 2019). The latent processes may be deterministic yet complex, such as the impact of environmental conditions on the development coralgal assemblages in a vertical reef drilled-core (Pall et al., 2020). The latent processes may also be stochastic such as earthquake fault rupture (Monterrubio-Velasco et al., 2019). The prior distribution of the parameters which govern these processes should be informed by the physical characteristics regarding the problem, such as precipitation values that must be positive and have a realistic upper limit (Chandra et al., 2019).

Gaussian process priors (GPP) play an important role in the Bayesian analysis of spatiotemporal data and have been used extensively for over 30 years. They are an established method for geostatistics that focuses

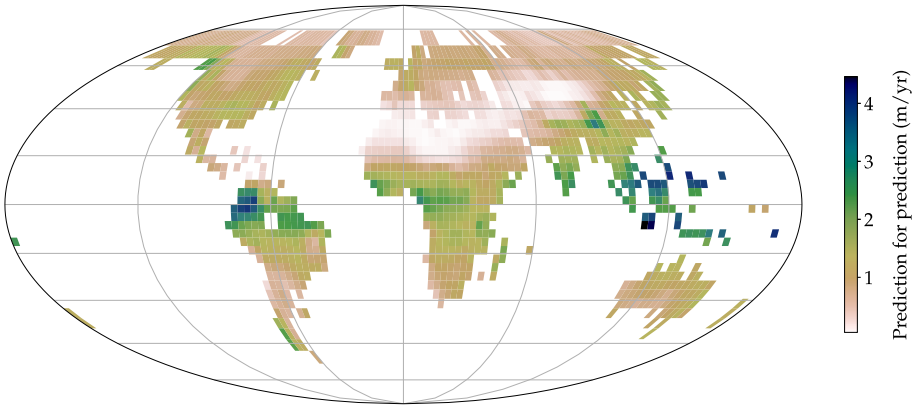


Fig. 1. Mid-Miocene precipitation model estimation (Herold et al., 2011).

on spatial or spatiotemporal datasets (Diggle et al., 1998; Chiles and Delfiner, 2009). A Gaussian process is used as a prior over an unknown function. Such a prior has many advantages. First, and most importantly, it avoids specifying any parametric form for the function, such as assuming the function is linear/polynomial or a power function. Second, it assumes that the posterior mean of the function changes smoothly, where the degree of smoothness is estimated from the data, so that over fitting is rarely problem. Third, uncertainty intervals for the posterior distribution of the unknown function are easily obtained from the iterates of the MCMC scheme. Lastly, it provides a means of obtaining estimates of the function, and the corresponding uncertainties, between observed data points.

Kriging is an example of a GP regression which is more familiar to geologists (Bernardo et al., 1998). GP regression has been used in many applications. In a recent application (Marchant et al., 2018), a GP regression was incorporated into a larger model to determine spatial-demographic insights for criminology. The dependency between crime rate and demographic indicators was assumed to be linear, while the dependency between spatial location and crime rate, after controlling for demographic information, was assumed to follow a GP. The advantage of this approach is that it provides a flexible fit to the data via the GP, while allowing for inference to be made around the impact of demographic indicators on crime. In this paper, we use this approach in modelling paleolithic precipitation; the relationship between precipitation and geographic variables, including lithologies can be modelled via linear regression, while the relationship between precipitation and spatial location can be modelled via a GP.

3. Data

This paper provides reconstruction of paleo-precipitation for selected timeslices between 14 and 249 million years ago (Ma) as shown in Table 1. The data used to do this can be grouped into three categories. The first category is lithological indicators, the second is precipitation simulations, and the third is paleo-coastlines and continental paleo-elevation.

3.1. Lithological indicators

We group available lithological indicators into three groups following Boucot et al. (2013a). Humid environments are represented by coal deposits and sediments with palm, mangrove, and crocodilian fossils. Arid environments are represented by evaporites, calcrete, bauxite, laterite, kaolinite, and oolitic ironstones. Glacial environments are represented by tillite, dropstones, and glendonites. We refer to these three types of lithologies as coal, evaporite and glacial, respectively. We note that we do not have all of these lithological indicators across the reconstruction timeslices. The lithologies depend on the climate conditions of the time slices and we have a significant amount of missing information as shown in Table 1.

In our analysis, the climate-sensitive geological indicators are reconstructed using a recent plate model (Matthews et al., 2016) paired with a paleomagnetic reference frame as employed by Cao et al. (2019), in order to produce data on the presence of lithology types for reconstruction timeslices from 14 Ma to 249 Ma. We note that we do not have all of these lithological indicators across the geological time period of interest. The lithologies depend on the climate conditions of the time frame and we have a lot of missing information which is evident from our sparse dataset as shown in Table 1.

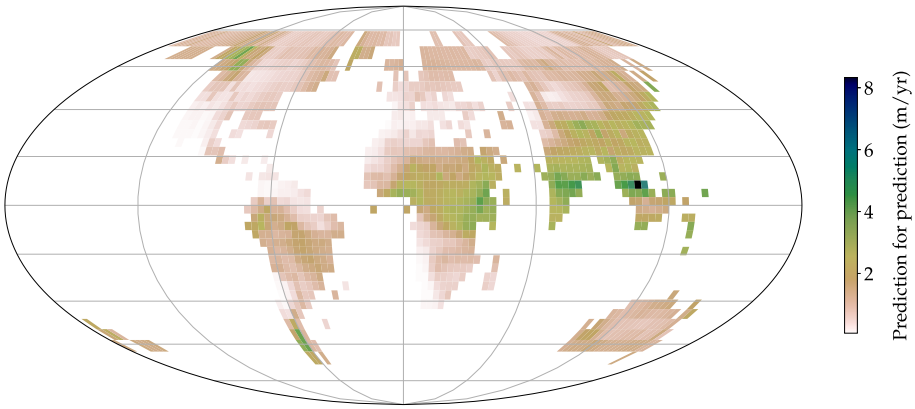


Fig. 2. Late Eocene precipitation model estimation (Hutchinson et al., 2018).

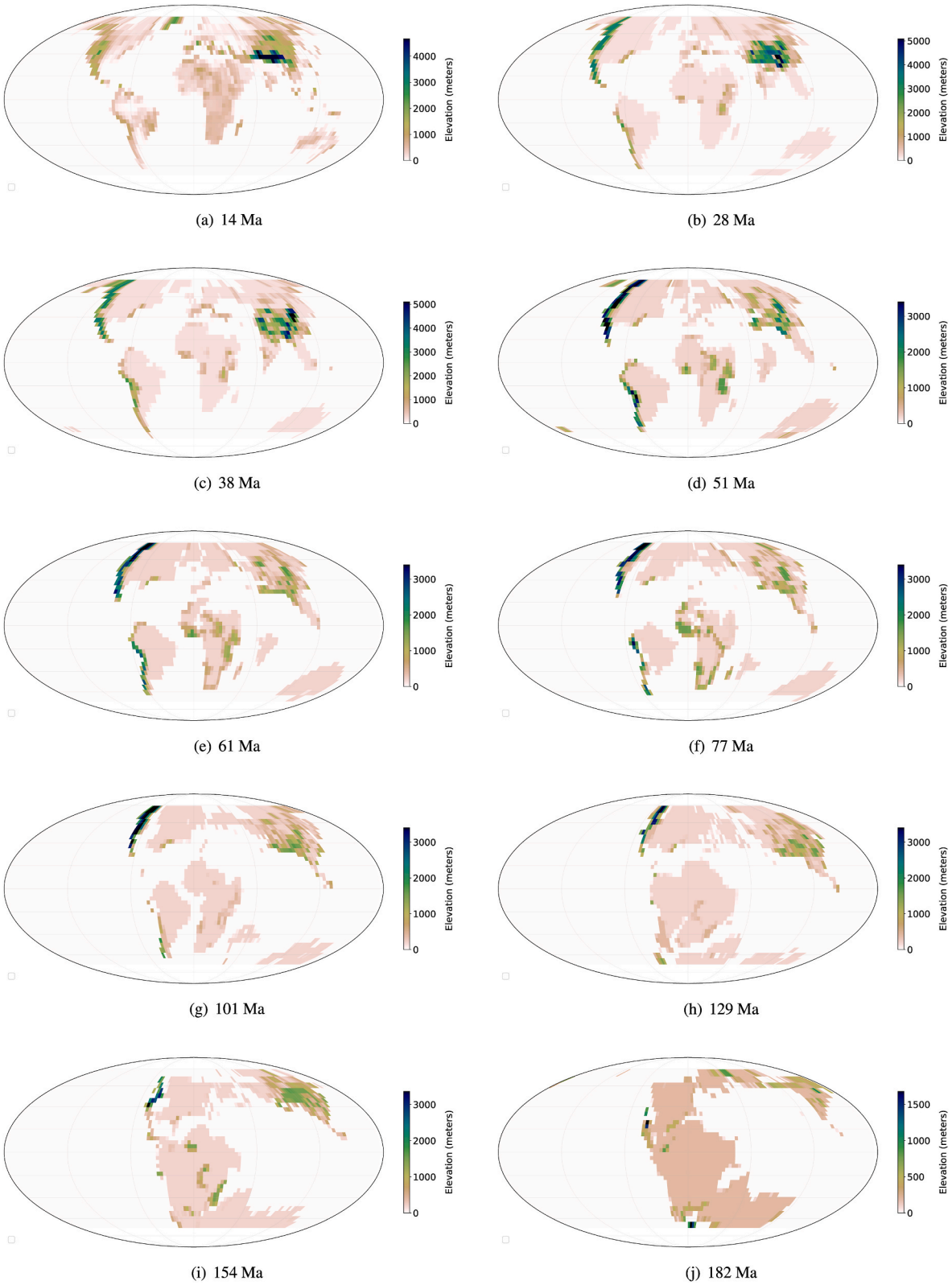
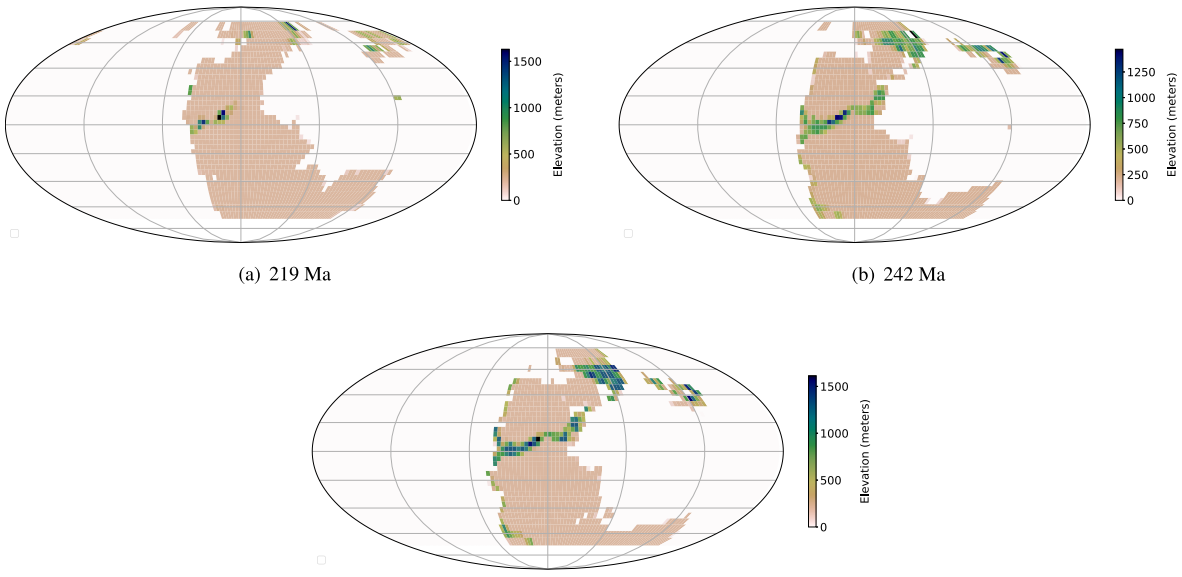


Fig. 3. Maps of paleo-elevation from the mid-Miocene (14 Ma) to the early Jurassic (182 Ma).

3.2. Precipitation simulation

The paleoclimate proxy data are complemented by global precipitation simulations from GCMs for the mid-Miocene (Herold et al., 2011) and late Eocene (Baatsen et al., 2020) epochs. Hence, we use simulated precipitation datasets and proxy data from the mid Miocene (14 Ma) and the late Eocene (38 Ma) for model training and validation, respectively. The precipitation estimates at these two times are extracted for the sites at which time-overlapping paleo-climate proxies are available from Boucot et al. (2013b).

Figs. 1 and 2 presents the precipitation estimates from the mid-Miocene and late Eocene epochs (Herold et al., 2011). Using the data



(a) 219 Ma (b) 242 Ma

(c) 249 Ma

Fig. 4. Maps of paleo-elevation in the Triassic (219–249 Ma).

from the mid Miocene and Eocene as training data, the goal is to provide precipitation estimates for the test data, namely those epochs without GCM-simulated precipitation.

These precipitation data of the Miocene and Eocene are available on the spatial characteristics of sites across the globe at a grid of 2.5° resolution in longitude. The latitude-spacing is 2.5° at the equator and increases toward the poles to preserve the grid-area for 13 reconstruction times, following Cao et al. (2019).

3.3. Paleo-coastlines and paleo-elevation

The paleogeographic maps of Cao et al. (2019) subdivide various timeslices in the geological past into different qualitative classes. The majority of continental areas above sea-level are classified as 'lowlands', with the remainder of the continents above sea-level subareas classified as 'mountains' (a further class, that of 'ice sheets' also exists for some Cenozoic timeslices but is not relevant for our analysis since no data points are available for Antarctica during these times). We use a simple approach to convert these maps of qualitative paleoenvironment into a quantitative representation of paleotopography through time (Ziegler et al. (1985)). First, we assign uniform elevation values to different classes: we assigned 200 meters (m) for lowland areas, to which we added an additional 1500 m of elevation for mountain ranges, 3000 m for non-collisional orogenic belts along the western margins of the Americas and 4500 m for the collisional orogen between India and Eurasia. To smooth the transition between mountains, lowlands and regions of the continents below sea-level, we applied a 1° buffer which is approximately 110 km (km), around the mountain regions to remove the lowland elevations. We replaced them with elevation values by linear interpolation between the surrounding mountain and lowlands elevations. Finally, we smoothed the grids with a 400 km wavelength low-pass Gaussian filter.

Our analysis is further underpinned by paleo-coastlines from the globalpaleo-environments of Cao et al. (2019) back to 249 Ma. Paleo-coastlines provide information such as the distance to shoreline relative to a location of interest, and orientation given by direction to the nearest paleo-shoreline (clockwise from North in degrees).

Figs. 3 and 4 present maps of paleo-elevation which are used in our proposed model for precipitation estimation. In developing our model, we only use the spatial characteristics of the sites above sea level in a given age, and thus the number of observations changes with each reconstruction time. The time-dependent site-specific model parameters include distance to the nearest paleo-shoreline, orientation given by direction to the nearest paleo-shoreline (clockwise from North in degrees given as cosine), and the paleo-elevation given by meters. We denote the number of sites with geological indicators for each time t to be N_t for $t = 1,\ldots,T$; where, the subscript t refers to the reconstruction time and $T = 13$ is the total number of times.

Table 1 provides details for the number of observations regarding the spatial characteristics for each reconstruction time (Ma), the number of observations on the deposit type (coal, evaporites, and glaciers and other associated deposits for humid, arid and glacial climates) denoted by N_{D_t}, and the number of precipitation estimates available for that era.

4. Methodology

In order to avoid underestimating the uncertainty surrounding imputing the missing lithologies, we model the joint posterior distribution of lithologies and precipitation. Hence, we estimate the marginal posterior distribution of lithologies given the data, and then model the conditional posterior distribution of precipitation. The following sections provide further details of the approach.

4.1. Model and priors

We wish to model the joint distribution of sedimentary deposits (referred to as deposits from hereon), denoted by $\mathscr{Z}$, and precipitation, denoted by $\mathscr{Y}$ across the globe for a number of reconstruction times in the past, conditional on some data. We denote this distribution as $\mathscr{P}(\mathscr{Z}, \mathscr{Y}|\text{data})$.

The joint distribution of $\mathscr{Z}$, and $\mathscr{Y}$ conditional on the data can be decomposed into the posterior marginal distribution of deposits, multiplied by the conditional distribution of precipitation. Here, the conditioning is on the deposits as well as the data, and we have

$$\mathscr{P}(\mathscr{Z}, \mathscr{Y}|\text{data}) = \mathscr{P}(\mathscr{Z}|\text{data}) \times \mathscr{P}(\mathscr{Y}|\mathscr{Z}, \text{data}).$$

We define the observed precipitation and deposits to be Y^{obs} and $\mathbf{Z}^{\text{obs}}$, respectively; where $Y^{\text{obs}} = (\mathbf{y}_{.1}^{\text{obs}}, \ldots, \mathbf{y}_{.T}^{\text{obs}})$, with $\mathbf{y}_{.t}^{obs} = (y_{1,t}^{\text{obs}}, \ldots, y_{N_t,t}^{\text{obs}})'$, and y_{st}^{obs} is the observed median annual precipitation at site s in time t, for $t = 1,\ldots,T$. The reconstruction times corresponding to the time indicator t are given in Table 1. Let $\mathbf{Z}^{\text{obs}} = (Z_{.1}^{\text{obs}}, \ldots, Z_{.T}^{\text{obs}})$ with

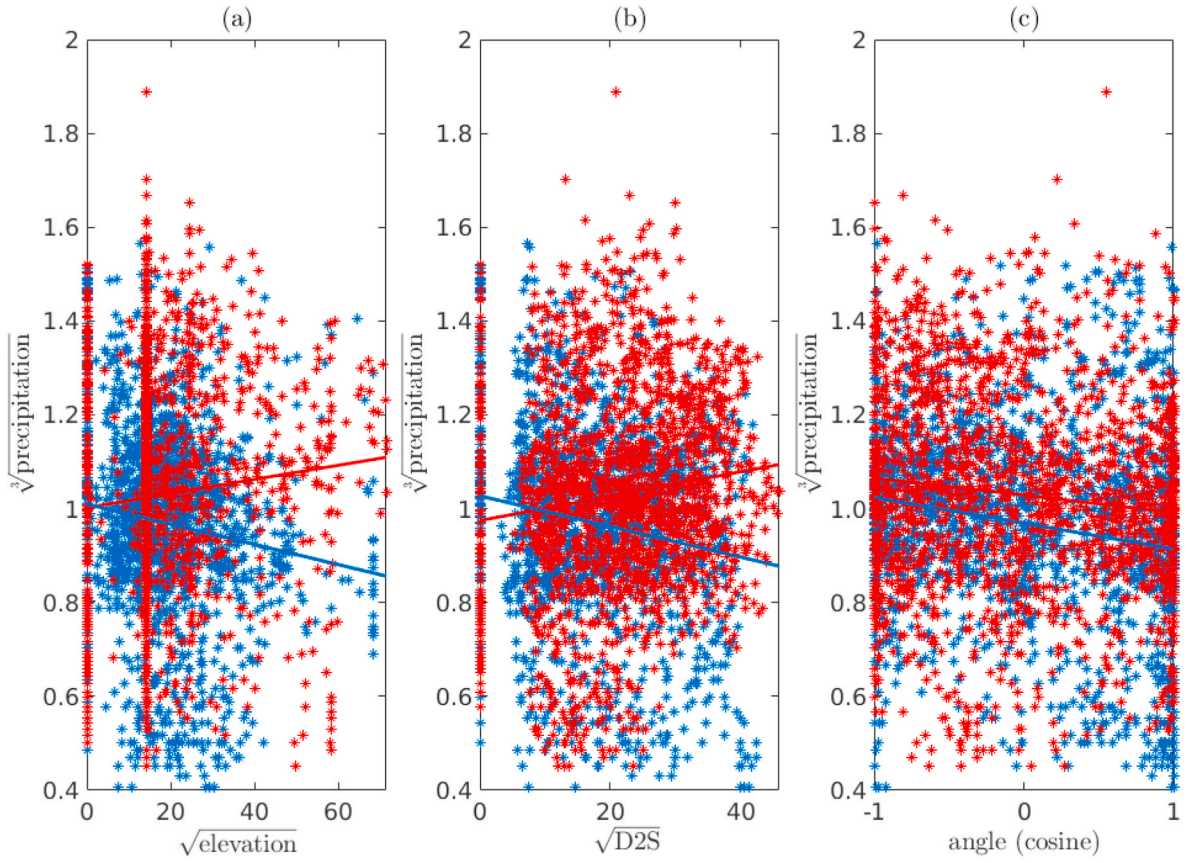


Fig. 5. Scatter plots for establishing the marginal relationship of the spatial covariates with respect to the precipitation in Miocene (blue) and Eocene (red) combined. The assumption of a linear relationship between $\sqrt[3]{\text{precipitation}}$ and the variables (a) ($\sqrt{\text{elevation}}$, (b) $\sqrt{\text{D2S}}$, and (c) *angle* (cosine) is reasonable.

$Z_{.t}^{\text{obs}} = (\mathbf{z}_{.1,t}^{\text{obs}}, \ldots, \mathbf{z}_{.N_t,t}^{\text{obs}})'$ with $\mathbf{z}_{.s,t}^{\text{obs}} = (z_{1,s,t}^{\text{obs}}, \ldots, z_{D,s,t}^{\text{obs}})'$; where $z_{d,s,t}^{\text{obs}} = 1$, if a deposit of type d was observed at location s in time t, and $z_{d,s,t}^{\text{obs}} = 0$ otherwise, for $d = 1, \ldots, D$, and D is the number of deposit types.

We also have data available on the spatial characteristics of the site, denoted collectively by $\mathbf{X}$, such as the latitude (*lat*), longitude (*lon*), elevation given in meters, the shortest distance to paleo-shoreline (*D2S*) given in kilometers, and orientation to nearest paleo-shoreline corresponding to *D2S* (given by *angle* (cosine)).

We transformed the variables so that the assumption of a linear relationship between the predictors and response, with normally distributed errors. It is convenient, from a computational and estimation point of view to assume that the noise has a Gaussian distribution. In order for this assumption to be valid, it is common practice to transform the data. We take the cube root transformation of the median annual precipitation data ($\sqrt[3]{\text{precipitation}}$), based on previous research (Stidd, 1953; Sanso and Guenni, 2000). We also take the square root to transform ($\sqrt{\text{elevation}}$) and $\sqrt{\text{D2S}}$, and hence use them as variables hereafter.

We note that we do not use longitude; however, it is included indirectly by using the distance to coastlines (this controls aridity in continental interiors) and by using the *angle* (cosine) of this distance to distinguish between the effect of eastern versus western coasts on precipitation. The idea for using the coastal orientation originates from the need to distinguish between continental locations close to a western or eastern shoreline, as these react differently to adjacent oceans. Climate and precipitation is moderated by the proximity to a western or eastern ocean because of the effect of the rotation of the Earth and the Coriolis force on ocean circulation, and where upwelling of deep, cold water occurs; i.e. preferentially along west-facing coastlines), making these coastal regions cooler and reducing precipitation relative to their eastern counterparts.

The array $\mathbf{X}$ is defined as $\mathbf{X} = (X_{.1}, \ldots, X_{.T})$ where $X_{.t} = (\mathbf{x}_{1t}, \ldots, \mathbf{x}_{N_t t})'$, with $\mathbf{x}_{s,t} = (x_{1st}, \ldots, x_{Kst})'$ and x_{kst}, is the value of covariate k at location s in time t for $t = 1, \ldots, T$, $s = 1, \ldots, N_t$, and $k = 1, \ldots, K$; where, K is the number of covariates.

In developing our proposed framework for modelling $\mathscr{P}(\mathscr{Z}, \mathscr{Y}|\text{data})$, it is useful to look at the relationship between the spatial characteristics of a site and the precipitation and deposit type, if recorded, at that site. Fig. 5 shows scatter plots of the predictor variables (a) ($\sqrt{\text{elevation}}$, (b) $\sqrt{\text{D2S}}$, and (c) *angle* (cosine), versus $\sqrt[3]{\text{precipitation}}$. We show the data for the Miocene in blue and Eocene in red. The figure shows that we can fit a linear regression line for each predictor variable, for each period.

We now outline our model for $\mathscr{P}(\mathscr{Z}|\text{data})$ and $\mathscr{P}(\mathscr{Y}|\mathscr{Z}, \text{data})$. The data consists of the observed precipitation (Y^{obs}), the observed deposits (Z^{obs}), and the spatial characteristics given by $\mathbf{X}$. In defining our model for $\mathscr{P}(\mathscr{Z}, \mathscr{Y}|\text{data})$, it will be useful to define subsets of the data; $\mathbf{X}^z \in \mathbf{X}$, given $\mathbf{X}^z = (lat, lon)$ and $\mathbf{X}^y \in \mathbf{X}$, given $\mathbf{X}^y = (lat, \sqrt{\text{elevation}}, \sqrt{\text{D2S}}, angle)$ used to predict $\mathscr{Z}$ and $\mathscr{Y}$, respectively.

4.1.1. Model for deposits

In this section, we describe our model for $\mathscr{P}(\mathscr{Z}|\text{data})$. There are many other locations for which there are topographical observations, than there are locations for there are deposit observations, see Table 1. Thus, our goal is to develop a flexible model for each time t to predict the probability of the presence of a deposit type d at a given location, conditional on the location of the observed deposits given by Z^{obs} and spatial characteristics, $\mathbf{X}^Z$.

We assume that the spatial dependence among deposits of type d in time t is induced by a Gaussian process ($\mathscr{GP}$) prior (Wahba, 1990) with mean function $U\boldsymbol{\alpha}_{d.t}$, with $U = (\mathbf{u}_1, \ldots, \mathbf{u}_{N_t})'$ and $\mathbf{u}_s = (1, lat_s, lon_s)'$, for

$s = 1,\ldots,N_t$, and $\boldsymbol{\alpha}_{d.t} = (\alpha_{0dt}, \alpha_{1dt}, \alpha_{2dt})'$ and covariance matrix $\tau_{dt}^2\Omega_t$. We use the reproducing kernel Hilbert space defined by a two-dimensional thinplate Gaussian process prior to construct Ω_t, expressed as a linear combination of basis functions, as described in Wood et al. (2013); Nychka (1988). To achieve computational feasibility, we truncate the basis expansion to the first $P_{dep} = 60$ basis vectors, where the subscript *dep* denotes the number of basis functions used in our model for deposits. We obtain the eigenvalue decomposition of $\Omega_t \approx Q_t R_t Q_t'$;where Q_t is a $N_t \times P_{dep}$ matrix of the first P_{dep} eigen-vectors, R_t is a $P_{dep} \times P_{dep}$ diagonal matrix with entries corresponding to the first 60 eigenvalues, and define $W_t = Q_t R_t^{1/2}$, $\varphi_{dt} \sim N(0, \tau_{dt}^2 I_{P_{dep}})$ so that $W_t\varphi_{dt}$ is approximately distributed $\mathcal{GP}(0, \tau_{dt}^2\Omega_t)$.

We model the probability of deposit d occurring at a location s and time t by

$$\Pr(z_{dst} = 1 | \mathbf{x}_{st}^z) = \Psi(\mathbf{u}_s\boldsymbol{\alpha}_{.dt} + \mathbf{w}_{st}\varphi_{dt}) \tag{1}$$

where $\Psi(\cdot)$ is the standard normal cumulative distribution function, known as a probit link function, and $\mathbf{w}_{st}$ is the s^{th} row of W_t. The probabilities in Equation (1) are formulated as probit models so that the sampling scheme of Albert and Chib (1993) can be used to obtain draws of $\boldsymbol{\alpha}_{dt}$ and φ_{dt} from the appropriate conditional posterior distribution as described in Section 4.2.

4.1.2. Model for precipitation conditional on deposits

Modelling the dependency of precipitation on latitude and longitude across such a large time horizon is problematic due to plate tectonics and changing boundaries of the continents. Accordingly, we do not include longitude since latitude plays the major role in climate conditions (Deutsch et al., 2008). Therefore, our set of spatial characteristics $\mathbf{X}^Y$ used to predict precipitation and use only *lat*, $\sqrt{\text{elevation}}$, $\sqrt{\text{D2S}}$, and *angle* as predictors. Additionally, we condition on the type of deposits which are present at the site, as these contain information regarding precipitation. For example, the locations where evaporites are present are likely to have lower than average precipitation, while those where coal deposits are found are likely to have higher than average precipitation.

In our model $\mathcal{P}(\mathcal{Y}|\mathcal{Z} = \mathbf{Z}, \mathbf{X}^Y)$, we assume that the observed precipitation in time t at location s denoted by $y_{s,t}$, conditional on observations in $\mathbf{x}_{st}^y$ and $\mathbf{z}_{st}$ arise from a true signal, plus some noise.

We obtain a non-parametric, yet parsimonious estimate of the dependence between precipitation and latitude by assuming this dependence is induced by a $\mathcal{GP}$ prior with mean $\boldsymbol{\mu}_{lat}$ and covariance matrix $\lambda^2\Sigma$ constructed using a low rank approximation to the reproducing kernel Hilbert space defined by an integrated Wiener process (Wood et al., 2013), expressed as a linear combination of basis functions. To achieve parsimony, we truncate the basis expansion to the first $P_{precip} = 10$ basis vectors only, where the subscript *precip* refers to precipitation. We obtain the eigenvalue decomposition of $\Sigma \approx QRQ'$, and define $L = QR^{1/2}$, $\boldsymbol{\delta} \sim N(0, \lambda I_{10})$, as in Section 4.1.1. Then, the function which maps latitude to precipitation $\mathbf{f}_{lat} = \boldsymbol{\mu}_{lat} + L\boldsymbol{\delta}$ is approximately distributed as $\mathcal{GP}(\boldsymbol{\mu}_{lat}, \lambda\Sigma)$. Note that we assume this function varies only with latitude and is constant across all reconstruction times. This assumption is justified for the time range investigated here. Cao et al. (2019) found that the latitudinal dependence of climate-sensitive lithologies differs for timeslices before 300 Ma when compared to timeslices after 300 Ma, where the palaeo-latitudinal distribution of climate proxies, particularly coal, has remained similar.

To formally define the model, let y_{st} be the cube root of precipitation at location s, at time t. Our model is

$$Y_{.t} = 1_{N_t}\beta_0 + Z_{.t}\boldsymbol{\beta}_Z + X_{.t}^y\boldsymbol{\beta}_X + L\boldsymbol{\delta} + e_{st} \text{ with } e_{st} \sim N(0, \sigma^2) \tag{2}$$

where $X_{.t}^y = (\mathbf{x}_{1t}^y, \ldots, \mathbf{x}_{N_t t}^y)$, $\mathbf{x}_{s,t}^y = (lat_s, \sqrt{elevation_{st}}, \sqrt{D2S_{st}}, angle_{st})$, for $s = 1, \ldots, N_t$ $\boldsymbol{\beta}_X = (\beta_{lat}, \beta_{\sqrt{D2S}}, \beta_{\sqrt{elevation}}, \beta_{angle})'$, and $\boldsymbol{\beta}_Z = (\beta_{coal}, \beta_{evap}, \beta_{glac})'$.

Let $\boldsymbol{\beta}^* = (\beta_0, \boldsymbol{\beta}_Z, \boldsymbol{\beta}_X)$ be the regression coefficients of variables which comprise the linear component of the mean of $Y_{.t}$ and let H_t be the matrix formed by formed by concatenating the vector 1_{N_t} with the matrices $Z_{.t}$, $X_{.t}^Y$ and so that $H_t = (1_{N_t}, Z_{.t}, X_{.t}^Y)$. We can decompose the mean of $Y_{.t}$ into its linear and non-linear components and write Equation (2) as

$$Y_{.t} = H_t\boldsymbol{\beta}^* + L\boldsymbol{\delta} + e_{st} \tag{3}$$

Our full model is now

$$Y_{.t} | \boldsymbol{\beta}^*, H_t, L_t, \boldsymbol{\delta}, \sigma^2 \sim N(H_t\boldsymbol{\beta}^* + L_t\boldsymbol{\delta}, \sigma^2 I_{N_t})$$

$$\boldsymbol{\beta}^* \sim N(0, c_\beta I)$$

$$\boldsymbol{\delta} \sim N(0, \lambda^2 I_{P_{prec}}),$$

$$Z_{d.t} \Big| U, W, \boldsymbol{\alpha}, \varphi \sim \prod_{s=1}^{Nt} N_t Be(\Phi(\mathbf{u}_s\boldsymbol{\alpha}_{dt} + \mathbf{w}_s\varphi_{dt})),$$

$$\boldsymbol{\alpha}_{dt} \sim N(0, c_\alpha I_3),$$

$$\varphi_{dt} \sim N(0, \tau_{dt}^2 I_{P_{dep}}),$$

$$\sigma \sim \text{Cauchy}_+$$
$$\lambda \sim \text{Cauchy}_+$$

$$\tau_{dt} \sim \text{Cauchy}_+ \tag{4}$$

for $d = 1, \ldots, 3$, $t = 1, \ldots, T$, and where Cauchy_+ is the folded Cauchy distribution which are commonly used prior for variance parameters with justifications given in (Gelmanet al., 2006). A graphical representation of the model appears in Fig. 6.

4.2. Estimation via MCMC sampling scheme

Our goal is to estimate the joint predictive distribution of precipitation and deposits across time and space. We take a Bayesian approach and use the joint posterior distribution $p(\mathcal{Y}, \mathcal{Z}|\text{data})$ and estimate the

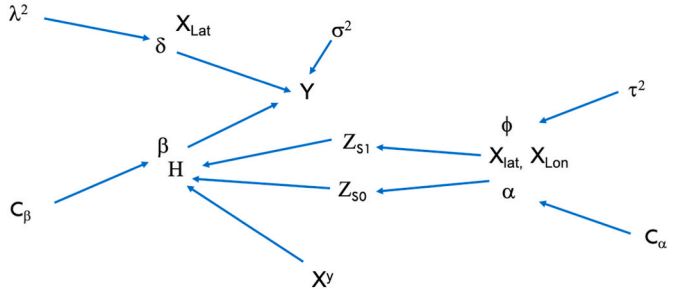


Fig. 6. A directed acyclic graph representation of the dependence between the model inputs, parameters, and outputs.

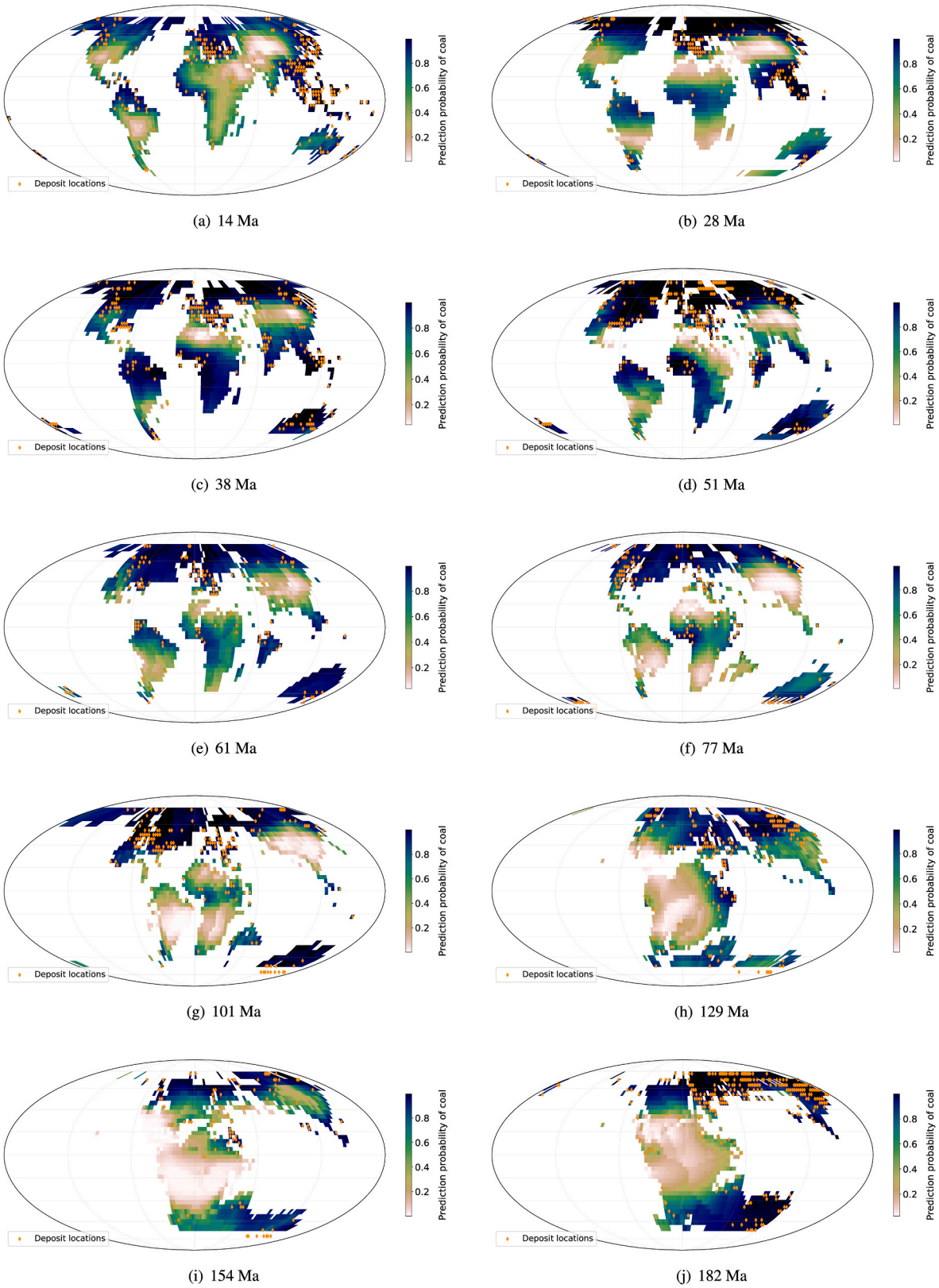


Fig. 7. Maps of predicted probability of coal deposits. The presence (observation) of a deposit used as part of the training data is given by the orange points on the grid.

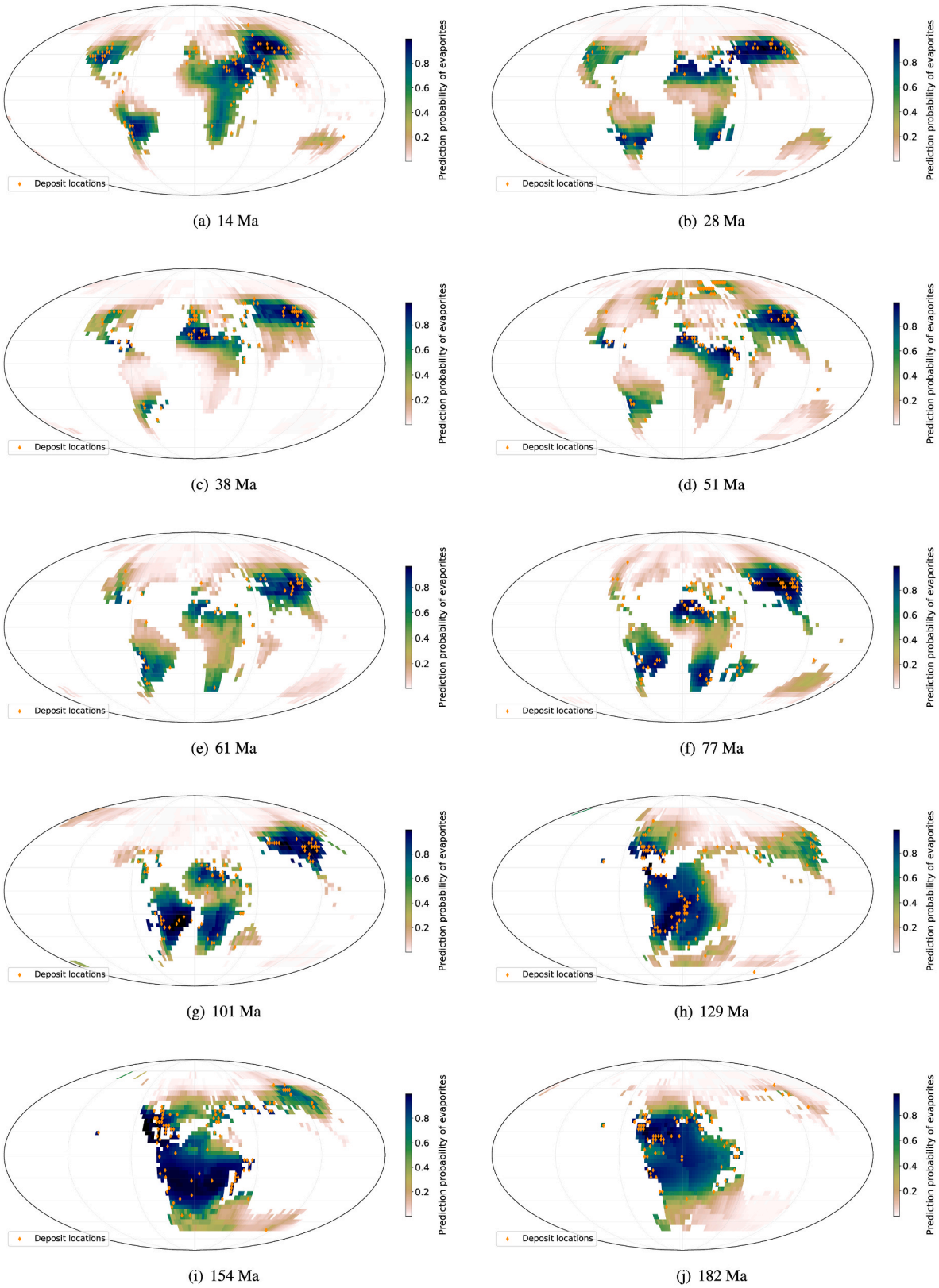


Fig. 8. Maps of predicted evaporite deposits given by probability. The presence (observation) of a deposit used as part of the training data is given by the orange points on the grid.

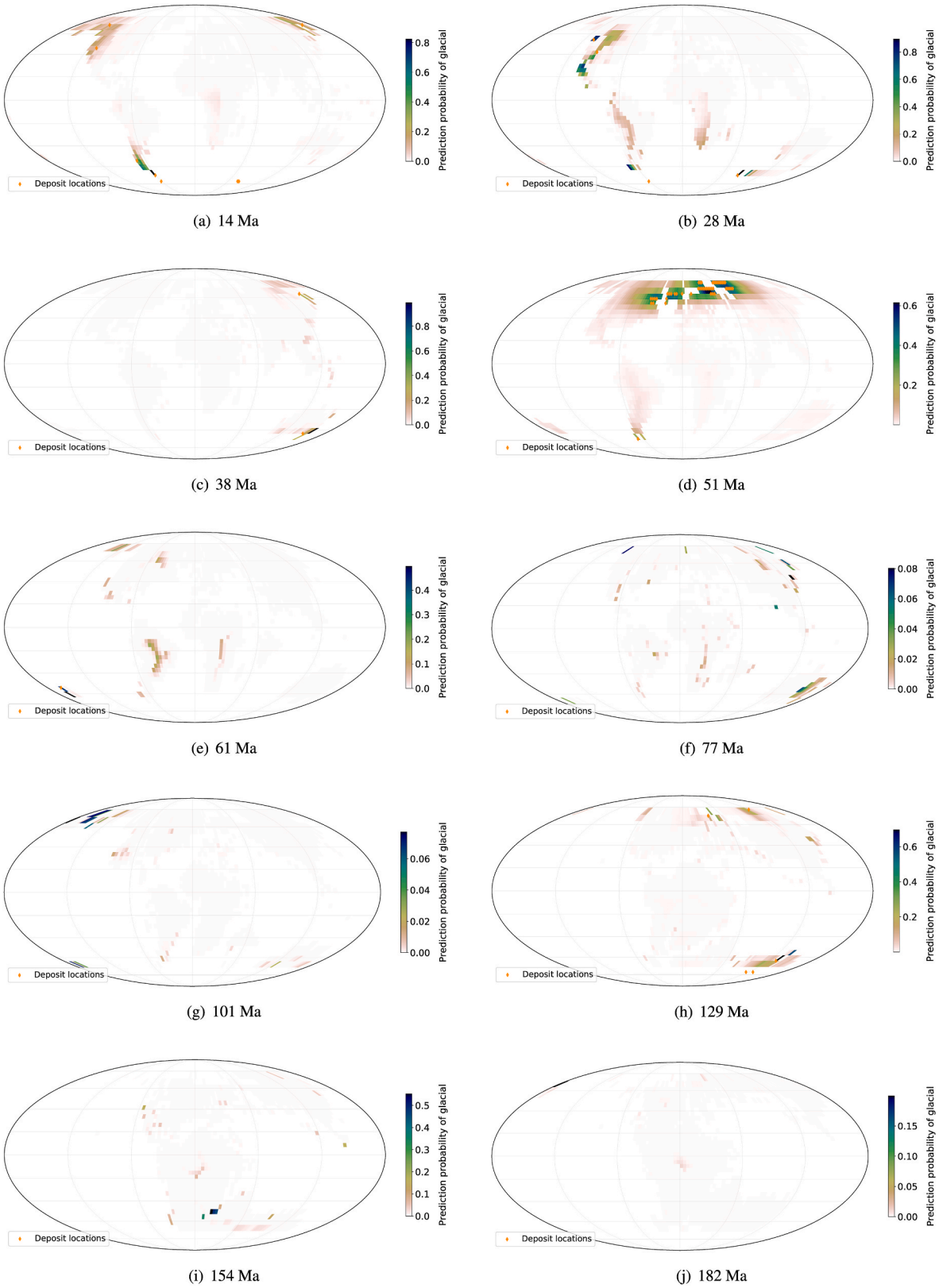


Fig. 9. Maps of predicted glacial deposits given by probability. The presence (observation) of a deposit used as part of the training data is given by the orange points on the grid.

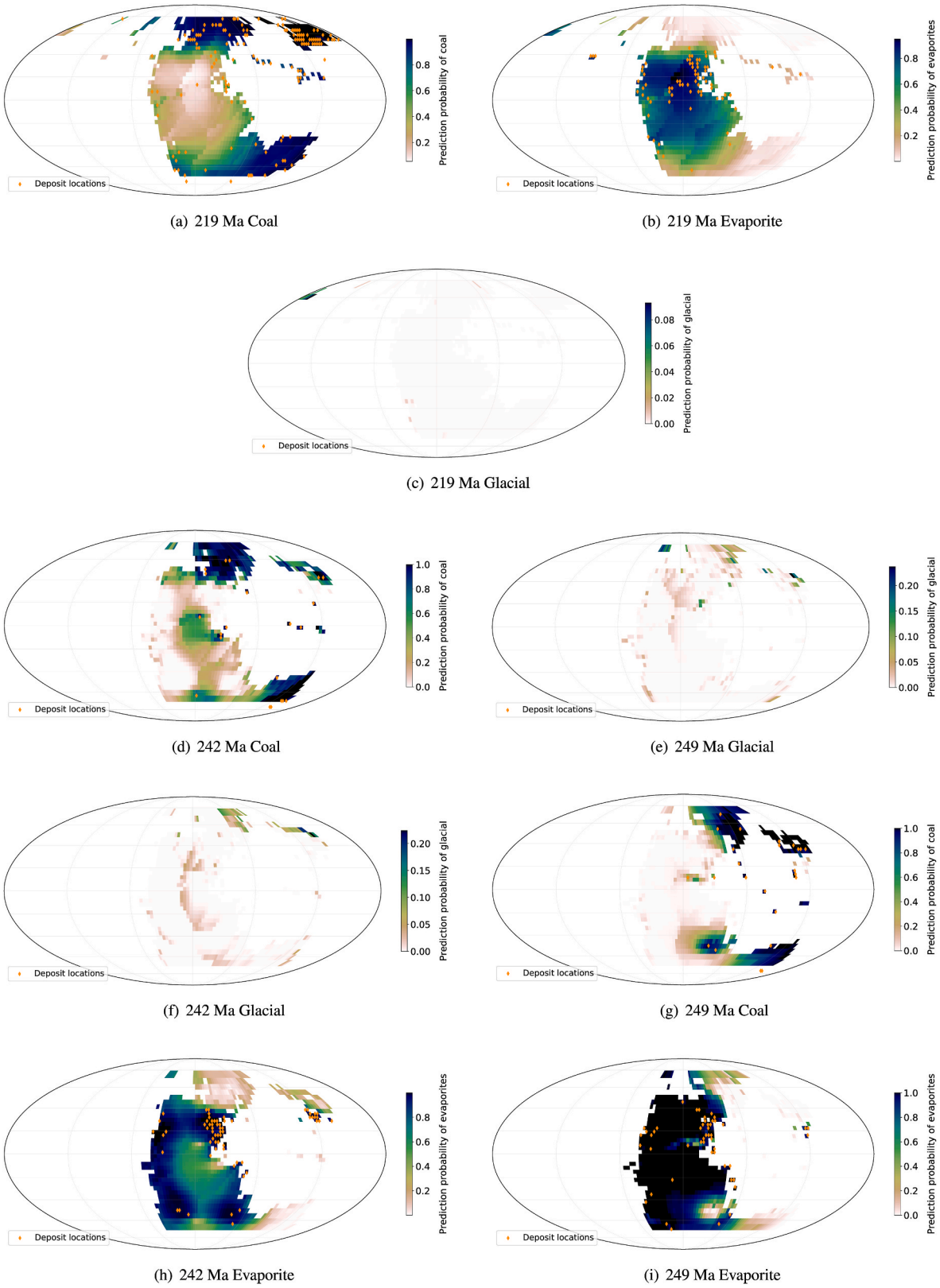


Fig. 10. Map of predicted deposits along with deposit location for 219 Ma to 249 Ma timeslices for respective deposits. The presence (observation) of a deposit which makes part of the training data is given by the orange points on the grid.

respective parameters via MCMC sampling scheme.

As mentioned earlier, the data consists of the observed values of precipitation and deposits, Y^{obs}, $\mathbf{Z}^{\text{obs}}$ as well as the observed factors, $\mathbf{X}$. We only have observations on precipitation for the Miocene $(t = 1)$ and Eocene $(t = 3)$. There is no precipitation data available for other timeslices and the predictive distribution $p(\mathscr{Y}, \mathscr{Z}|\text{data})$, is given by

$$p(\mathscr{Y}, \mathscr{Z}|\text{data}) = \int p(\mathscr{Y}, \mathscr{Z}|\boldsymbol{\Theta}, \text{data}) p(\boldsymbol{\Theta}|\text{data}) d\boldsymbol{\Theta} \quad (5)$$

where $\boldsymbol{\Theta}$ is a vector of unknown parameters which are needed to prescribe the joint posterior $p(\mathscr{Y}, \mathscr{Z}|\text{data})$. In this context, the unknown parameters are $\boldsymbol{\Theta} = (\boldsymbol{\beta}, \boldsymbol{\delta}, \mathbf{A}, \boldsymbol{\Phi}, \sigma^2, \mathbf{T}, \lambda)$; where $\mathbf{T} = (\boldsymbol{\tau}_{\cdot 1}^2, \ldots, \boldsymbol{\tau}_{\cdot T}^2)$ with $\boldsymbol{\tau}_{\cdot t}^2 = (\tau_{1t}^2, \tau_{2t}^2, \tau_{3t}^2)$, $\mathbf{A} = (\boldsymbol{\alpha}_{\cdot d1}, \ldots, \boldsymbol{\alpha}_{\cdot dN_t})$ with $\boldsymbol{\alpha}_{\cdot dt} = (\alpha_{0dt}, \alpha_{1dt}, \alpha_{2dt})$, and $\boldsymbol{\Phi} = (\varphi_{\cdot d1}, \ldots, \varphi_{\cdot dN_t})$ with $\varphi_{\cdot dt} = (\varphi_{1dt}, \ldots, \varphi_{P_{dep}dt})$, for $t = 1, \ldots, T$ and $d = 1, 2, 3$.

As stated in Section 4.1.1, the deposit data is sparse in space and time, and the missing values are part of the parameter space contained in $\boldsymbol{\Theta}$. We define the indicator variable $\gamma_{st} = 1$ to denote if the type of deposit is recorded at site s at time t and, $\gamma_{st} = 0$ otherwise. We further define the set $S_{1t} = \{s; \gamma_{st} = 1\}$ to be the collection of sites for which deposit information is recorded and define S_{0t} similarly. The matrix $Z_{\cdot t}$ is rearranged so that $Z_{\cdot t} = (Z_{S_1 t}, Z_{S_0 t})$ and $n_t = \sum_{i=1}^{N_t} \gamma_{st}$, is the number of sites for which deposit information is available at time t.

We approximate Equation (5) by

$$p\left(\mathscr{Y}, \mathscr{Z}|\text{data}\right) \approx \frac{1}{J} \sum_{j=1}^{J} p\left(y_{st}|\boldsymbol{\Theta}^{[j]}, \text{data}\right),$$

where $\boldsymbol{\Theta}^{[j]}$ are draws from the posterior distribution, $p(\boldsymbol{\theta}|\text{data})$, for $j = 1, \ldots, J$. We use MCMC sampling to estimate the parameters in the model and to facilitate the sampling scheme, we introduce the latent variables $\mathbf{v}_{d.t} = (v_{d1t}, \ldots, v_{dN_t t})$ for $t = 1, \ldots, T$ and $d = 1, 2, 3$ as shown below

$$v_{dst}|z_{dst} = 1 \sim N_{\mathbb{R}^+}(\mathbf{u}_s \boldsymbol{\alpha}_{\cdot dt} + \mathbf{w}_{st} \varphi_{dt}, 1)$$

$$v_{dst}|z_{dst} = 0 \sim N_{\mathbb{R}^-}(\mathbf{u}_s \boldsymbol{\alpha}_{\cdot dt} + \mathbf{w}_{st} \varphi_{dt}, 1)$$

where, the subscripts $\mathbb{R}^+$ and $\mathbb{R}^-$ denote that the distribution has support only on the positive and negative real number line, respectively. Then, conditional on v_{dst}'s, the regression coefficients $\boldsymbol{\alpha}_{dt}$ and φ_{dt} have distributions which are normally distributed.

The MCMC sampling scheme for a given number of iterations (n-iterates) proceeds as follows.

Initialize $(\sigma^2, \mathbf{T}, \lambda, A, \boldsymbol{\Phi}, \mathbf{Z}_{S_0 \cdot}) = (\sigma^{2[0]}, \mathbf{T}^{[0]}, \lambda^{[0]}, A^{[0]}, \boldsymbol{\Phi}^{[0]}, \mathbf{Z}_{S_0}^{[0]})$ then for. $j = 1, \ldots, \text{n} -$ iterates

1. For $t = 1, \ldots, T$ and $d = 1, \ldots, D$.
   (a) Draw $\mathbf{v}_{d.t}^{[j]}$ from. $p(\mathbf{v}_{d.t} \Big| Z_{S_1 t}, Z_{S_0 t}^{[j-1]}, U_t, W_t, \boldsymbol{\alpha}_{dt}^{[j-1]} \varphi_{dt}^{[j-1]})$
   (b) Draw $\boldsymbol{\alpha}_{dt}^{[j]}$ and $\varphi_{dt}^{[j]}$ jointly from.
   $p(\boldsymbol{\alpha}_{dt}, \varphi_{dt}, \Big| U_t, W_t, \mathbf{v}_{d.t}^{[j]}, \tau_{dt}^{2[j-1]})$. Note that conditional on $\mathbf{v}_{d.t}$, $\boldsymbol{\alpha}_{dt}$ and φ_{dt} are independent of $Z_{S_1 t}$, and $Z_{S_0 t}$.
   (c) Draw $\tau_{dt}^{2[j]}$ from. $p(\tau_{dt}^2 \Big| \varphi_{dt}^{[j]})$
   (d) Draw $Z_{S_0 t}^{[j]}$ from. $p(Z_{S_0 t} \Big| U_t, W_t, \mathbf{v}_{d.t}^{[j]}, \boldsymbol{\alpha}_{dt}^{[j]}, \varphi_{dt}^{[j]})$
   (e) Form. $H_t^{[j]} = (1_{N_t}, Z_{S_1 t}, Z_{S_0 t}^{[j]}, X_{\cdot t}^Y)$
2. Draw $\boldsymbol{\beta}^{*[j]}$ and $\boldsymbol{\delta}^{[j]}$ from
   $p(\boldsymbol{\beta}^*, \boldsymbol{\delta} \Big| Y, H_t^{[j]}, \sigma^{2[j-1]}, \lambda^{[j-1]})$.
3. Draw $\sigma^{2[j]}$ from $p(\sigma^2 \Big| Y, H_t^{[j]}, \boldsymbol{\beta}^{*[j]}, \boldsymbol{\delta}^{[j]})$
4. Draw $\boldsymbol{\lambda}^{2[j]}$ from $p(\boldsymbol{\lambda}^2 \Big| \boldsymbol{\delta}^{[j]})$
5. Repeat

We note that the proposed framework assumes that the dependency between precipitation and covariates is constant over time. The temporal structure is induced only because the values of these covariates change across time. This simplification is made because there is only precipitation data from two epochs, i.e. the mid-Miocene and the late Eocene, which is insufficient to directly estimate any temporal structure in precipitation.

Table 2
Misclassification performance (percentage) for given deposits (coal, evaporite and glacial) for the respective reconstruction timeslices. Note that the 0 percent misclassification indicate that there was no data present for the respective time of the glacial deposits. The results are not applicable (n/a) for cases where there was no data as shown in Table 1.

t	Timeslice (Ma)	Coal	Evaporite	Glacial
1	14	11.94	9.55	1.49
2	28	2.65	23.00	1.50
3	38	9.57	10.64	0.71
4	51	5.04	22.66	16.90
5	61	11.04	9.82	1.23
6	77	11.86	11.86	n/a
7	101	8.33	8.33	n/a
8	129	15.41	13.69	1.37
9	154	10.00	9.44	n/a
10	182	8.78	8.78	n/a
11	219	14.29	14.75	n/a
12	242	0	0	n/a
13	249	0	0	n/a

5. Results

5.1. Reconstruction of deposits

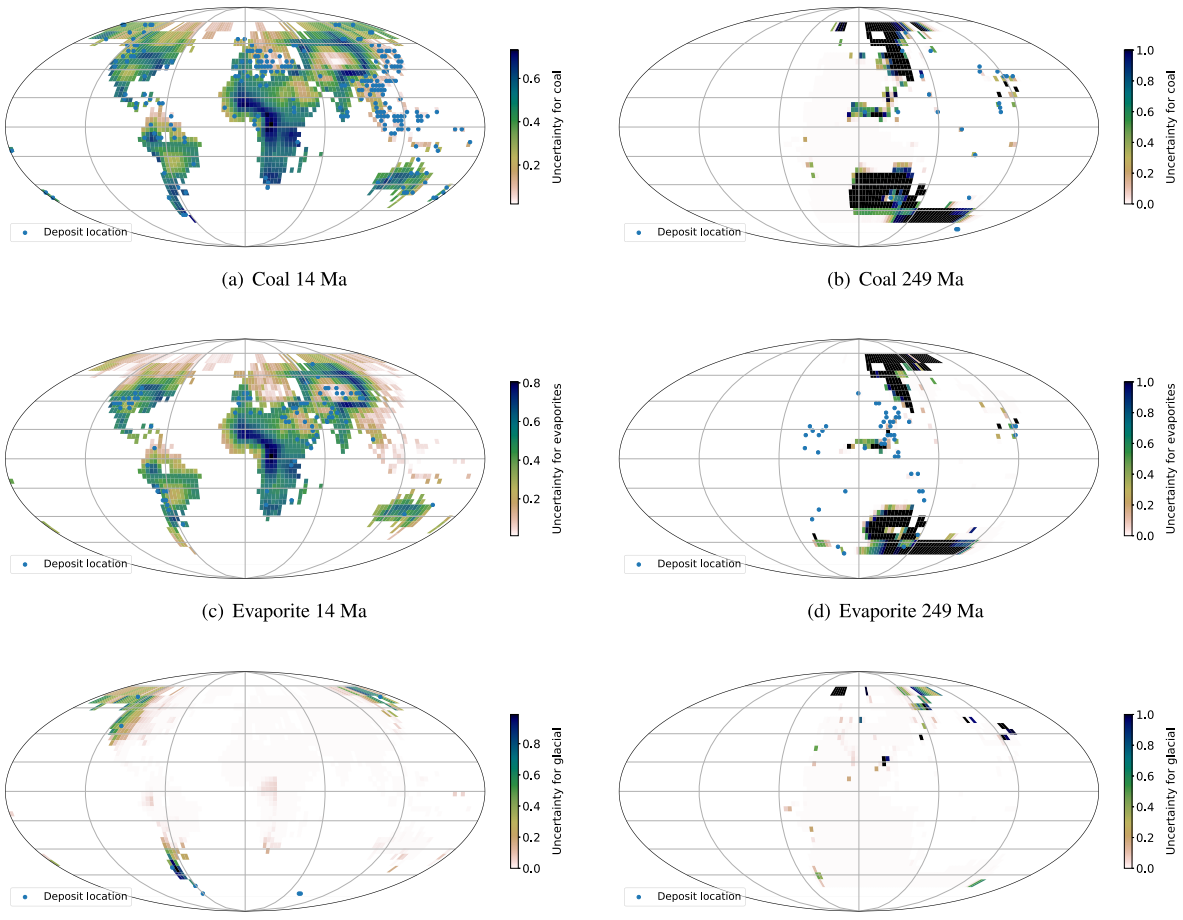
Recall that we are modelling the joint distribution $\mathscr{P}(\mathscr{Z}, \mathscr{Y}|\text{data}) = \mathscr{P}(\mathscr{Z}|\text{data}) \times \mathscr{P}(\mathscr{Y}|\mathscr{Z}, \text{data})$. Fig. 7 panels (a)-(j) presents the marginal posterior probability of a coal deposit across the globe for various reconstruction timeslices, namely $\mathscr{P}(\mathbf{z}_{1.t}|\text{data})$, for $t = 1, \ldots, 10$, for 14 Ma ($t = 1$), panel (a), to 182 Ma ($t = 10$), panel (j). Figs. 8 and 9 are analogous plots for evaporites and glacial deposits, namely $\mathscr{P}(\mathbf{z}_{2.t}|\text{data})$, and $\mathscr{P}(\mathbf{z}_{3.t}|\text{data})$, respectively. Fig. 10 presents the results for all three types of deposits for 242 Ma ($t = 11$) to 249 Ma ($t = 13$) timeslices, respectively. Note that the actual observations in the training data is given by the orange spots on the grid and the rest are estimations.

Table 2 gives the misclassification rates for the deposits (coal, evaporite and glacial) for the respective reconstruction times. Each time was treated independently and 90 percent of data was used for training and the remaining 10% for testing. We note that in some cases, the misclassification rate is not applicable (n/a), for glacial deposits where we have no observations as shown in Table 1.

Fig. 11 shows uncertainty estimates surrounding the posterior probability of deposits, for the first and the last timeslices only, namely 14Ma ($t = 1$) and 249 Ma ($t = 13$). Panels (a) and (b) show results for coal deposits, panels (c) and (d) for evaporites, and panels (e) and (f) for glacial deposits. As expected, the uncertainty is highest in locations for which no deposit information was recorded, and lowest in those locations where deposit information was available. In addition, the uncertainty is higher for earlier timeslices than later ones.

5.2. Paleo-precipitation estimation

One of the most important drivers of precipitation is latitude (Deutsch et al., 2008) since it has major effect on climate. Figs. 12 and 13 plot latitude versus precipitation for the Miocene ($t = 1$) and the Eocene ($t = 3$), respectively. The observations are colour-coded according to deposit information; black for coal, red for evaporites, yellow for glacial and blue for not available (NA). It is clear from these figures that the relationship between latitude and precipitation is not only non-linear, it is also not well approximated by parametric nonlinear forms, such as polynomials. To investigate the relationship between latitude and precipitation, we assume the dependence is induced by a $\mathscr{GP}$ prior, as discussed in Section 4.1.2. We provide plots of the 95% credible



(a) Coal 14 Ma (b) Coal 249 Ma

(c) Evaporite 14 Ma (d) Evaporite 249 Ma

(e) Glacial 14 Ma (f) Glacial 249 Ma

Fig. 11. We present two instances of uncertainties, first and last prediction for the time slice (14 and 249 Ma) given the respective deposits. The presence (observation) of a deposit which makes the training data is given by the orange spots on the grid.

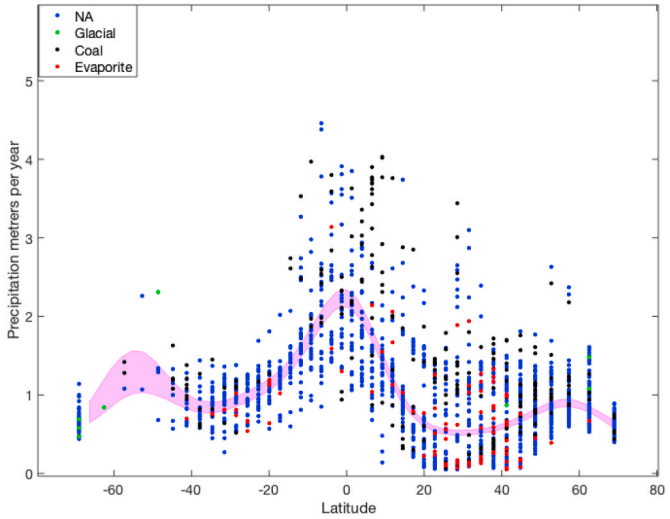


Fig. 12. We plot 95% credible intervals (shown in pink) for the annual median precipitation by the model as a function of latitude for the Miocene. The observations are colour coded according to deposit information; black for coal, red for evaporites, green for glacial and blue for not available (NA).

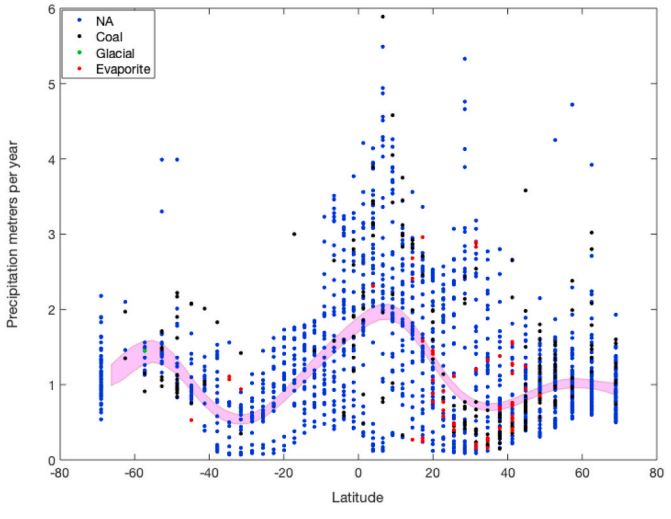


Fig. 13. We plot 95% credible intervals (shown in pink) for the annual median precipitation the model as a function of latitude for the Eocene. The observations are colour coded according to deposit information; black for coal, red for evaporites, green for glacial, and blue for not available (NA).

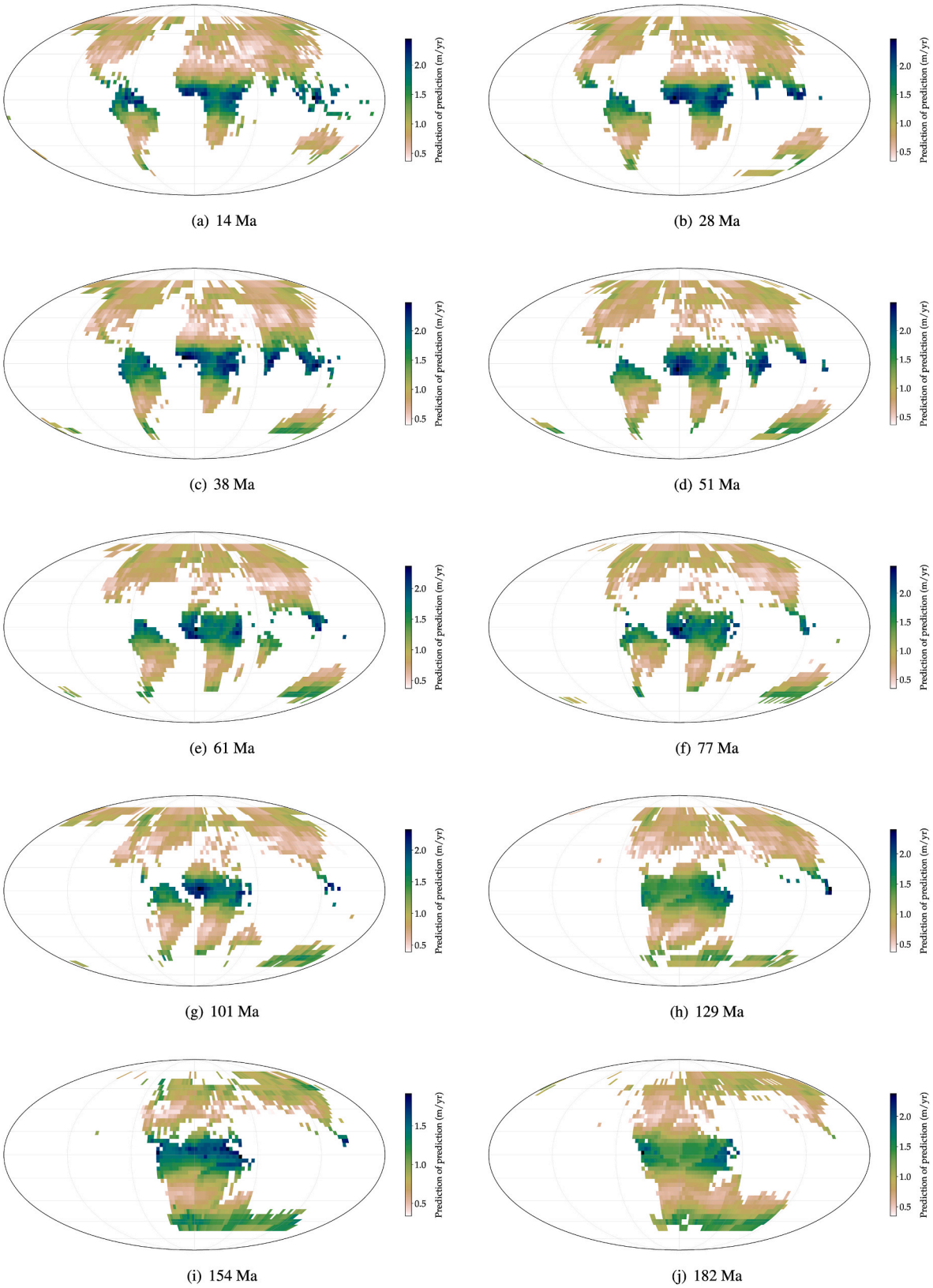
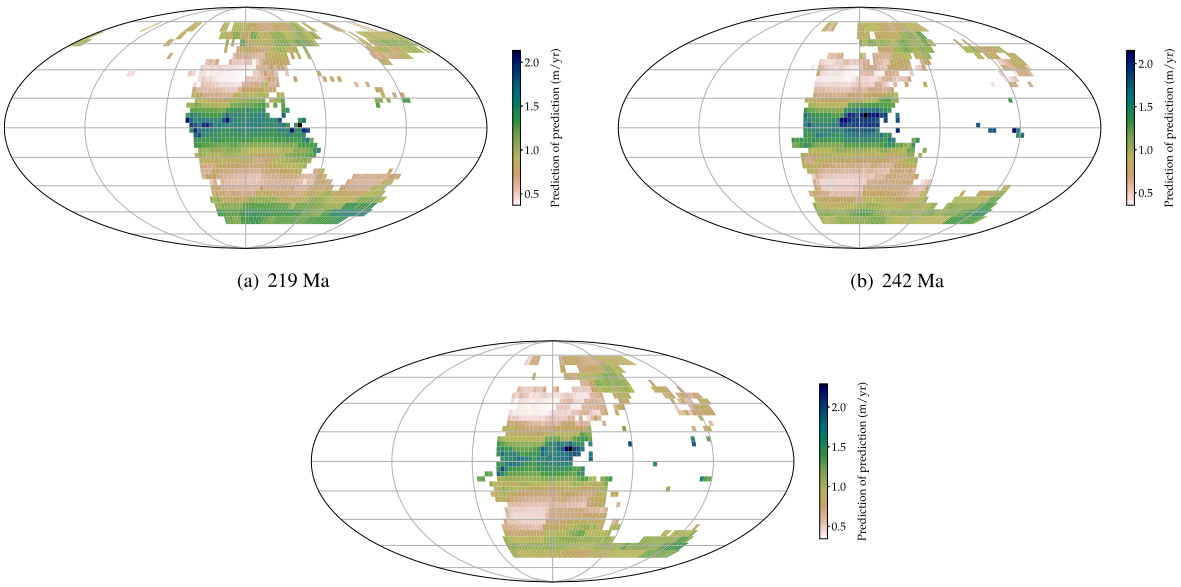


Fig. 14. Maps of the posterior median precipitation from the mid-Miocene (14 Ma) to the early Jurassic (182 Ma).

intervals for the posterior median of precipitation by the model, assuming this prior and using latitude as the only covariate (shaded function regions in Figs. 12 and 13).

One of the model assumptions is that the relationship between latitude and precipitation is constant across all epochs. Figs. 12 and 13 show that this assumption is reasonable; both plots show local minima at $30^o N/S$, and local maxima at about $55^o N/S$. Fig. 12, shows the global maxima of median precipitation occurs at the equator; however, it is



(a) 219 Ma (b) 242 Ma

(c) 249 Ma

Fig. 15. Maps of the posterior median precipitation for three reconstruction times in the Triassic (219–249 Ma).

interesting to note that this maximum is shifted North by about 5^o for the Eocene data.

In estimating the time varying posterior distribution of precipitation, spatial data such as *D2S*, *elevation*, *latitude* and paleo-shoreline orientation *angle* and geological data, such as the deposit information from all epochs are used, as well as precipitation data from both the Miocene and Eocene. We account for the uncertainty in the missing deposit information by integrating over all possible values where the integration is with respect to the posterior distribution of the deposits.

Figs. 14 and 15, show the posterior median annual precipitation for time slices $t = 1, \ldots, 10$ and $t = 11, \ldots, 13$ respectively.

Figs. 16 and 17, show the posterior uncertainty surrounding the annual precipitation for time slices $t = 1, \ldots, 10$ and $t = 11, \ldots, 13$; respectively. The uncertainty is computed by the difference of 5th and 95th percentile in predictions.

Fig. 18 shows the posterior distribution of the regression coefficients contained in $\boldsymbol{\beta}^*$ in Equation (3). As expected the presence of a coal deposit is positively related to precipitation while the presence of an evaporite deposit is negatively related to precipitation.

Our predicted precipitation maps (Figs. 14 and 15) illustrate how the distribution of continents and oceans combined with latitude, reflecting paleo-temperature gradients, control precipitation. Moving back in time from the Late (14 Ma) to the Early Cenozoic (61 Ma), changing latitudinal positions of Africa results in northern Africa becoming wetter while central Africa south of the equator becomes progressive drier (Fig. 14). In contrast, there is less of a change in broad precipitation patterns observed in the Americas, because their latitudinal positions have changed little during the Cenozoic Period. India moves through the wet equatorial belt in the early-mid Cenozoic, resulting in intense precipitation throughout most of the sub-continent at 38 and 51 Ma, while at 61 Ma most of India is located south of this belt, leading to a progressive drying that intensifies in the Cretaceous (Fig. 14). As Australia moves south back in geological time, gradually closing the southeast Indian Ocean, its southern half experiences a greening due to its movement into southern hemisphere mid-latitudes. This is marked by moderate rainfall reflecting eastward traveling depressions and fronts that yield rain throughout the year (Fig. 14).

Throughout the Cretaceous Period, patterns of annual precipitation are predicted to change more significantly back in time as the South Atlantic closes, creating a large landmass by joining South America and Africa, as seen at 129 Ma (Fig. 14). The elimination of western African coasts and the creation of a vast continental interior region at 129 Ma results in a marked change in precipitation, moderating the precipitation highs seen later at 101 and 77 Ma along the coast of Western Africa between what is now Nigeria and Senegal (Fig. 14). The large equatorial Tethys gateway that existed in the Early Cretaceous between Africa/South America and Eurasia results in an increase in precipitation along the northeastern coastline of Africa, seen at 129 Ma. Moving back in time from the Late into the Early Cretaceous, India experiences a greening again as it moves into the southern mid-latitude belt of moderate rainfall (Fig. 14).

Moving into the Jurassic a large southern continent, Gondwana, forms, whose interior is relatively dry, but whose northern and southern portions straddle high to medium rainfall belts, as visible at 154 Ma (Fig. 14). The Early Jurassic (182 Ma) exhibits a reduction in northern Gondwana rainfall compared to the Late Jurassic (154 Ma), reflecting the elimination of central Atlantic coastlines by closing the central North Atlantic Ocean and forming the supercontinent Pangea, which persists throughout the Triassic (219–249 Ma) (Fig. 15). Pangea in the Triassic shows a relatively stable precipitation pattern with two prominent rainfall lows centered around 35° north and south, with an equatorial high and moderate mid-latitude highs (Fig. 15). Southern hemisphere mid-latitude rainfall is most intense in the Late Triassic at 219 Ma, while the Early Triassic is somewhat drier in Gondwana's (southern Pangea's) interior and along the southern perimeter of Pangea (Fig. 15).

The main difficulty in comparing our results with other published

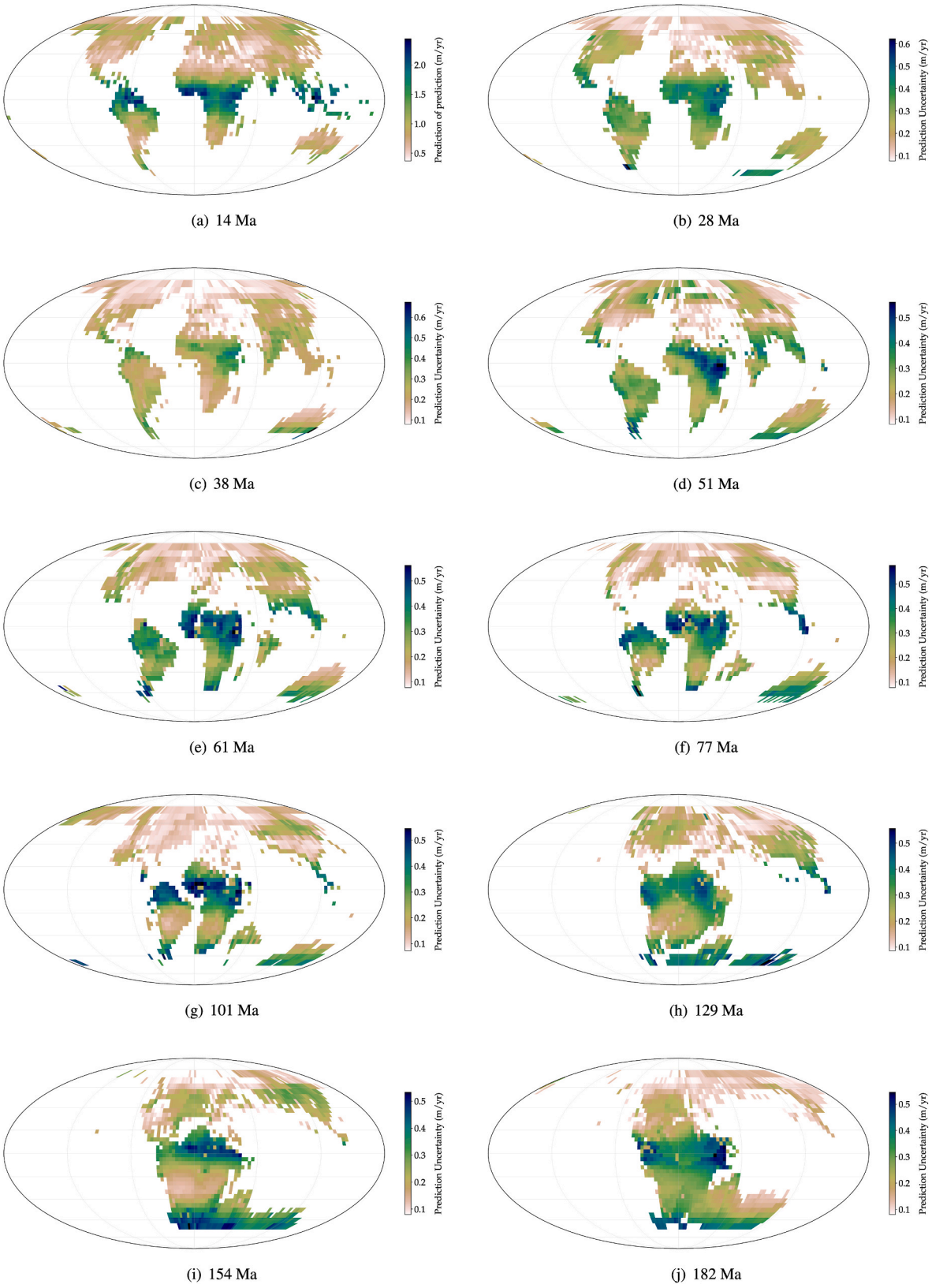
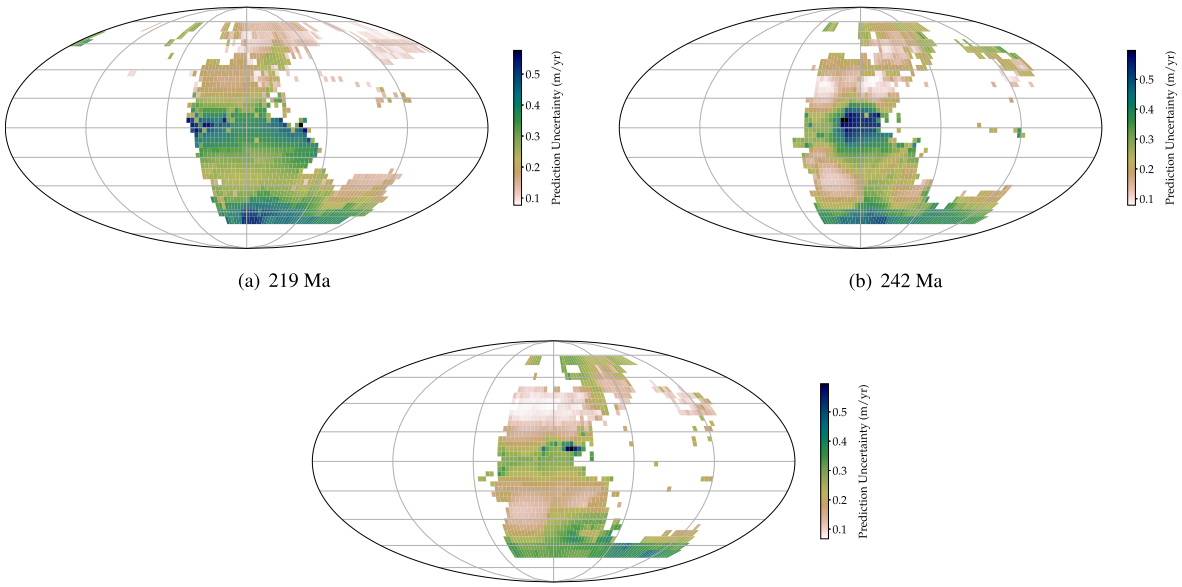


Fig. 16. Posterior uncertainty of annual precipitation for the epochs corresponding to $t = 1, \ldots, 10$. The uncertainty is computed by the difference of 5th and 95th percentile in predictions.



(a) 219 Ma (b) 242 Ma

(c) 249 Ma

Fig. 17. Posterior uncertainty of annual precipitation for the epochs corresponding to $t = 11, \ldots, 13$. The uncertainty is computed by the difference of 5th and 95th percentile in predictions.

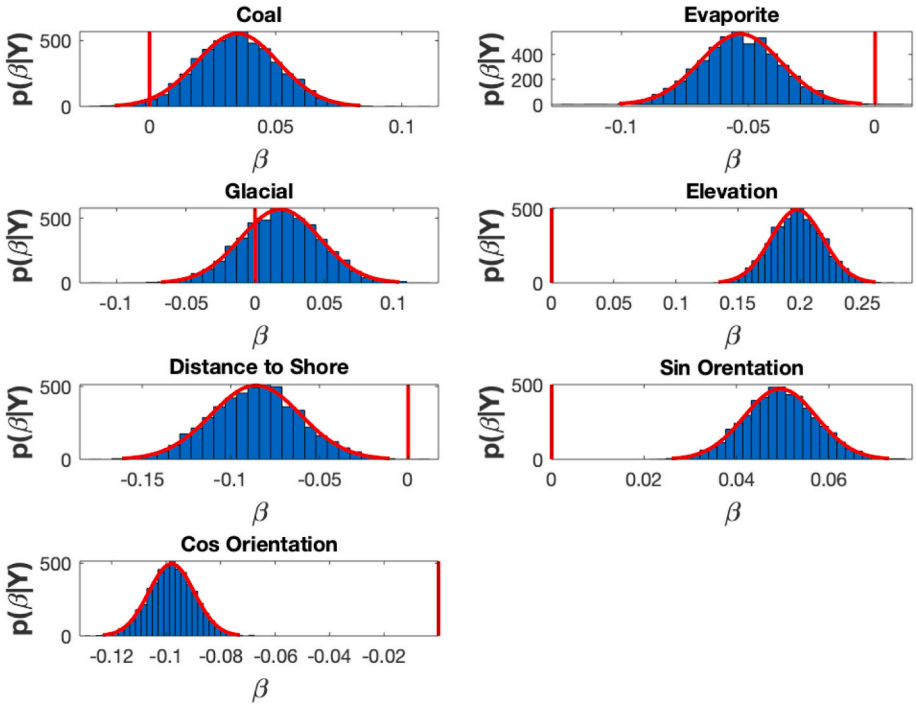


Fig. 18. Regression coefficients of the respective input features used for predicting precipitation using data from the Miocene and Eocene as training. These regression coefficient represent the marginal impact of that variable on precipitation after controlling for the other variables.

Table 3
Root-mean squared error (RMSE) taking into account the actual and predicted values for Miocene and Eocene precipitation (m/yr).

	Timeslice (Ma)	RMSE
Miocene	14	0.5439
Eocene	38	0.7636

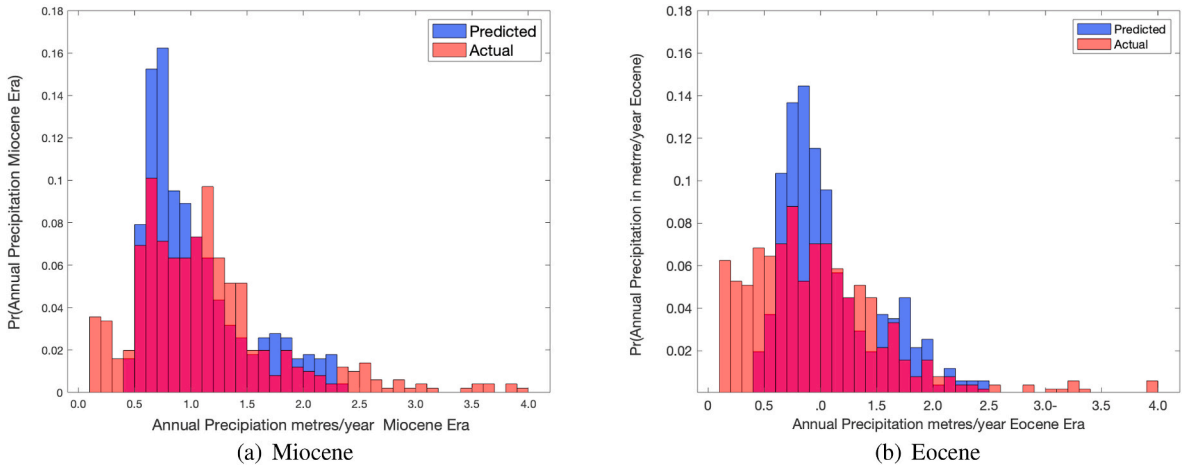


Fig. 19. The distribution of actual and predicted annual precipitation (metres/year) for the Miocene era, panel (a) and the Eocene era, panel (b).

work (i.e. climate simulations) is that precipitation as output of climate simulations is not frequently made available. In addition, the timeframes for which these models are run would not generally coincide with the timeframes for which we have predictions. However, we used the Eocene simulation dataset as a test set where the training model was based on Miocene data only.

We show the accuracy of the predictions by the root-mean squared-error (RMSE) shown in Table 3 and Fig. 19. The RMSE takes into account the actual and predicted values for Miocene and Eocene precipitation (m/yr). We notice that the RMSE values become somewhat poorer for the Eocene when compared to the Miocene and from the histogram in Fig. 19. It is clear that the range of predictions for the Eocene range from 0.25 to 2.5, while the actual data range from 0 to 4.5. Hence, the model does not seem to do well for extreme cases, i.e anomalously high precipitation values.

6. Discussion

6.1. Mesozoic era

The quality of our precipitation maps can be compared to published reconstructions and models of paleo-climate belts and rainfall. Our oldest map centered on 249 Ma (Fig. 16c) represents the earliest part of the Triassic Period. The Triassic is known for the driest climate in the last 500 million years, driven by a large contiguous land area, a small area of tropical ocean and expansive deserts, lacking vegetation and a number of other factors (Hay and Wold, 1998). The precipitation model (Fig. 16) agrees very well with a recent climate simulation for this time (Montenegro et al., 2011). We have two prominent regions north and south of the equator with minimal precipitation, leading to evaporation exceeding precipitation (Montenegro et al., 2011) and the widespread occurrence of evaporites (Fig. 8). Our maps reveal a number of changes in the spatial extent of the equatorial humid belt while Pangea was assembled in the Triassic Period (roughly from 249 to 200 Ma) (Fig. 16a and b) and throughout Pangea's initial breakup in the Jurassic Period postdating 200 Ma, but it is difficult to verify our models with other independent climate reconstructions, as mid-late Triassic and Jurassic climate models are sparse. The Cretaceous Period, commencing around 145 Ma has been the subject of a number of climate models and proxy reconstructions (e.g. (Ziegler et al., 1987; Bush and Philander, 1997). A Cretaceous simulation of global precipitation (Bush and Philander, 1997) suggests that it was approximately 10 percent greater than at present, with the only region of reduced precipitation occurring in southern central Eurasia. Such a pattern of pronounced low-latitude rainfall, especially in regions that are deserts today, such as northern Africa, while southern central Eurasia is relatively dry, is well reflected in our Cretaceous precipitation reconstructions (Fig. 15f–h). The existence of an equatorial seaway in the Cretaceous Period produces a large body of warm tropical water, much unlike Pangea's configuration, leading to a relatively wet northern Africa, northern South America and Southeast Asia.

6.2. Cenozoic era

The early Cenozoic (Fig. 15 d and e) still resembles the Cretaceous in terms of the persisting presence of an equatorial seaway, relatively warm climate and a distribution of precipitation with only minor changes relative to the Cretaceous. A major change can be seen in the late Eocene (38 Ma) (Fig. 1 5c), a period approaching the initial formation of Antarctic ice sheets in the Oligocene at 34 Ma, associated with gradual global cooling (Zachos et al., 2001). This transition is associated with reduced global precipitation (Fig. 15 c) as compared to the relatively wet Cretaceous and early Cenozoic climate. The initial expansion of Antarctic ice sheets in the early Oligocene is followed by a transient warming period, climaxing at the mid Miocene climate optimum. This warming is associated with increased equatorial precipitation relative to 38 Ma (Fig. 15 a and b); however, the continuing dispersal of the continents following the breakup of Pangea leads to northern Africa moving out of the low-latitude region characterised by high precipitation. This results in a distinct reduction in rainfall in northern Africa at 14 Ma (Fig. 15 a). Both the Eocene and Miocene precipitation reconstructions are well constrained via the two GCMs we are using to train our model.

7. Conclusions and future work

We present a Bayesian machine learning framework for modelling the joint posterior distribution of precipitation and deposit presence across the global for 14 timeslices ranging from 14 Ma to 249 Ma. In this challenging spatiotemporal paleoclimate reconstruction problem, the posterior distribution accounts for the uncertainty in the missing deposit information by integrating over possible deposit locations using MCMC sampling to perform the required multidimensional integration. Our work represents one of the first attempts to couple physical process models with observational data in a fully probabilistic framework. Our precipitation maps can be used as inputs for surface process models that depend on rainfall through time as a boundary condition driving erosion and landscape evolution. Our reconstructions of climate-sensitive lithologies provide a link between changes in long-term climate and forest

cover, and also open the opportunity for a more differentiated reconstruction of different soil types, including calcrete and laterite, in the future.

Data and software

Data and open source software from this research is available online.[1]

All the model outputs (precipitation and predicted lithology grids with separate uncertainty grids) are made available on github and included with the paper.[2]:

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

We would like to thank Madhura Killedar, Sebastian Haan and David Kohn from the Sydney Informatics Hub of the University of Sydney for providing informatics support. We would like to thank Simon Williams and John Cannon for providing the paleo-elevation dataset.

R. Dietmar Muller acknowledges support from the Australian Research Council through grant IH130200012. S. Cripps and R. Chandra acknowledge support from the Australian Research Council through grant IC190100031.

Appendix

Table 4 provides a summary of key terms used in the paper.

Table 4
Summary of key terms

Term	Description
Prior	Prior probability distribution expresses belief, expert or prior knowledge about a quantity before taking into account evidence or data.
Likelihood function	A probabilistic measure of goodness of fit that considers data and model output.
Posterior	Posterior probability distribution of an unknown quality. is determined after taking into account prior and likelihood.
MCMC	Markov chain Monte Carlo used for sampling the posterior using likelihood and prior distribution.
Lithology	Geological indicators such as coal, evaporates, and glacial deposits.
Epoch	Geological time-span in millions of years (Ma).
Distance to shoreline	Gives the distance from the grid where the deposit was found to the nearest shoreline
Gaussian process model	Statistical model based on Gaussian process.
Precipitation	Condensation of atmospheric water vapor from clouds that includes rainfall and snow which are measured in meters per annum (m/a).
Impute	To estimate the value of a missing datapoint using statistical inference.

References

Albert, J.H., Chib, S., 1993. Bayesian analysis of binary and polychotomous response data. J. Am. Stat. Assoc. 88 (422), 669–679.

Arıkan, B., 2015. Modeling the paleoclimate (ca. 6000–3200 cal bp) in eastern anatolia: the method of macrophysical climate model and comparisons with proxy data. J. Archaeol. Sci. 57, 158–167. https://doi.org/10.1016/j.jas.2015.02.016. http://www.sciencedirect.com/science/article/pii/S0305440315000540.

Baatsen, M., von der Heydt, A.S., Huber, M., Kliphuis, M.A., Bijl, P.K., Sluijs, A., Dijkstra, H.A., 2020. The middle-to-late eocene greenhouse climate, modelled using the cesm 1.0.5. Clim. Past Discuss 2020, 1–44. https://doi.org/10.5194/cp-2020-29. https://cp.copernicus.org/preprints/cp-2020-29/.

Bernardo, J., Berger, J., Dawid, A., Smith, A., et al., 1998. Regression and classification using Gaussian process priors. Bayesian statistics 6, 475.

Birchfield, G., Weertman, J., Lunde, A.T., 1981. A paleoclimate model of northern hemisphere ice sheets. Quat. Res. 15 (2), 126–142.

Boucot, A.J., Xu, C., Scotese, C.R., Morley, R.J., 2013a. Phanerozoic Paleoclimate: an Atlas of Lithologic Indicators of Climate. SEPM (Society for Sedimentary Geology), Tulsa, OK.

Boucot, A.J., Xu, C., Scotese, C.R., Morley, R.J., 2013b. Phanerozoic paleoclimate: an atlas of lithologic indicators of climate. In: Nichols, G.J., Ricketts, B. (Eds.), Concepts in Sedimentology and Paleontology, no. 11, 478. Society for Sedimentary Geology (SEPM), Tulsa, Oklahoma, USA, pp. 1–30.

Bradley, R.S., 1999. Paleoclimatology: Reconstructing Climates of the Quaternary. Elsevier.

Bush, A.B., Philander, S.G.H., 1997. The late cretaceous: simulation with a coupled atmosphere-ocean general circulation model. Paleoceanography 12 (3), 495–516.

Cao, W., Williams, S., Flament, N., Zahirovic, S., Scotese, C., Müller, R.D., 2019. Palaeolatitudinal distribution of lithologic indicators of climate in a palaeogeographic framework. Geol. Mag. 156 (2), 331–354.

Carson, J., Crucifix, M., Preston, S., Wilkinson, R.D., 2018. Bayesian model selection for the glacial–interglacial cycle. J. Roy. Stat. Soc.: Ser. C (Appl. Stat.) 67 (1), 25–54.

Chandra, R., Azam, D., Müller, R.D., Salles, T., Cripps, S., 2019. Bayeslands: A bayesian inference approach for parameter uncertainty quantification in Badlands. Comput. Geosci. 131, 89–101.

Chiles, J.-P., Delfiner, P., 2009. Geostatistics: Modeling Spatial Uncertainty, vol. 497. John Wiley & Sons.

Contreras, D.A., Bondeau, A., Guiot, J., Kirman, A., Hiriart, E., Bernard, L., Suarez, R., Fader, M., 2019. From paleoclimate variables to prehistoric agriculture: using a process-based agro-ecosystem model to simulate the impacts of holocene climate change on potential agricultural productivity in provence, France. Quat. Int. 501, 303–316. https://doi.org/10.1016/j.quaint.2018.02.019. http://www.sciencedirect.com/science/article/pii/S1040618216313568.

Crowley, T.J., 1988. Paleoclimate modelling. In: Physically-Based Modelling and Simulation of Climate and Climatic Change, pp. 883–949. Springer.

Crowley, T.J., North, G.R., 1991. Paleoclimatology. Oxford University Press, New York, NY (United States).

Deutsch, C.A., Tewksbury, J.J., Huey, R.B., Sheldon, K.S., Ghalambor, C.K., Haak, D.C., Martin, P.R., 2008. Impacts of climate warming on terrestrial ectotherms across latitude. Proc. Natl. Acad. Sci. Unit. States Am. 105 (18), 6668–6672.

Diggle, P.J., Tawn, J.A., Moyeed, R.A., 1998. Model-based geostatistics. J. Roy. Stat. Soc.: Ser. C (Appl. Stat.) 47 (3), 299–350.

Gelfand, A.E., Smith, A.F., 1990. Sampling-based approaches to calculating marginal densities. J. Am. Stat. Assoc. 85 (410), 398–409.

Gelman, A., et al., 2006. Prior distributions for variance parameters in hierarchical models (comment on article by browne and draper). Bayesian Anal. 1 (3), 515–534.

[1] https://github.com/EarthByte/paleoclimate-reconstruction.

[2] https://github.com/EarthByte/paleoclimate-reconstruction/tree/master/reconstruction_prediction/results_depositsprecip.

Geman, S., Geman, D., 1984. Stochastic relaxation, gibbs distributions, and the bayesian restoration of images. IEEE Trans. Pattern Anal. Mach. Intell. (6), 721–741.
Glancy, T., Arthur, M.A., Barron, E., Kauffman, E., 1993. A paleoclimate model for the north american cretaceous (cenomanian-turonian) epicontinental sea. In: Geological Association of Canada Special Paper, vol. 39, pp. 219–241.
Hansen, J.E., Sato, M., 2012. Paleoclimate implications for human-made climate change. In: Climate Change, pp. 21–47. Springer.
Haslett, J., Whiley, M., Bhattacharya, S., Salter-Townshend, M., Wilson, S.P., Allen, J., Huntley, B., Mitchell, F., 2006. Bayesian palaeoclimate reconstruction. J. Roy. Stat. Soc. 169 (3), 395–438.
Hastings, W.K., 1970. Monte Carlo sampling methods using Markov chains and their applications. Biometrica 57, 97–109.
Hay, W.W., Wold, C.N., 1998. The Role of Mountains and Plateaus in a Triassic Climate Model. Tectonic Boundary Conditions for Climate Reconstructions. In: Crowley, T.J., Burke, K.C. (Eds.), In: Oxford Monographs on Geology and Geophysics, 39. Oxford University Press, New York, pp. 116–143.
Herold, N., Huber, M., Müller, R., 2011. Modeling the miocene climatic optimum. part i: land and atmosphere. J. Clim. 24 (24), 6353–6372.
Hutchinson, D.K., Boer, A.M.D., Coxall, H.K., Caballero, R., Nilsson, J., Baatsen, M., 2018. Climate sensitivity and meridional overturning circulation in the late Eocene using GFDL CM2. 1. Clim. Past 14 (6), 789–810.
Ilvonen, L., Holmström, L., Seppä, H., Veski, S., 2016. A Bayesian multinomial regression model for palaeoclimate reconstruction with time uncertainty. Environmetrics 27 (7), 409–422.
Jordan, M.I., Ghahramani, Z., Jaakkola, T.S., Saul, L.K., 1999. An introduction to variational methods for graphical models. Mach. Learn. 37 (2), 183–233.
Li, B., Nychka, D.W., Ammann, C.M., 2010. The value of multiproxy reconstruction of past climate. J. Am. Stat. Assoc. 105 (491), 883–895.
Lunt, D.J., Huber, M., Anagnostou, E., Baatsen, M.L.J., Caballero, R., DeConto, R., Dijkstra, H.A., Donnadieu, Y., Evans, D., Feng, R., Foster, G.L., Gasson, E., von der Heydt, A.S., Hollis, C.J., Inglis, G.N., Jones, S.M., Kiehl, J., Kirtland Turner, S., Korty, R.L., Kozdon, R., Krishnan, S., Ladant, J.-B., Langebroek, P., Lear, C.H., LeGrande, A.N., Littler, K., Markwick, P., Otto-Bliesner, B., Pearson, P., Poulsen, C.J., Salzmann, U., Shields, C., Snell, K., Stärz, M., Super, J., Tabor, C., Tierney, J.E., Tourte, G.J.L., Tripati, A., Upchurch, G.R., Wade, B.S., Wing, S.L., Winguth, A.M.E., Wright, N.M., Zachos, J.C., Zeebe, R.E., 2017. The deepmip contribution to pmip4: experimental design for model simulations of the eeco, petm, and pre-petm (version 1.0). Geosci. Model Dev. (GMD) 10 (2), 889–901.
Mann, M.E., Rutherford, S., 2002. Climate reconstruction using 'pseudoproxies'. Geophys. Res. Lett. 29 (10), 139–1–139–4.
Marchant, R., Haan, S., Clancey, G., Cripps, S., 2018. Applying machine learning to criminology: semi-parametric spatial-demographic Bayesian regression. Secur. Inf. 7 (1), 1–19.
Matthews, K.J., Maloney, K.T., Zahirovic, S., Williams, S.E., Seton, M., Muller, R.D., 2016. Global plate boundary evolution and kinematics since the late paleozoic. Global Planet. Change 146, 226–250.
McGehee, R., Lehman, C., 2012. A paleoclimate model of ice-albedo feedback forced by variations in earth's orbit. SIAM J. Appl. Dyn. Syst. 11 (2), 684–707.
McIntyre, S., McKitrick, R., 2009. Proxy inconsistency and other problems in millennial paleoclimate reconstructions. Proc. Natl. Acad. Sci. Unit. States Am. 106 (6). E10–E10.
Montenegro, A., Spence, P., Meissner, K.J., Eby, M., Melchin, M.J., Johnston, S.T., 2011. Climate simulations of the Permian-Triassic boundary: Ocean acidification and the extinction event. Paleoceanography 26 (3). https://doi.org/10.1029/2010PA002058.
Monterrubio-Velasco, M., Rodríguez-Pérez, Q., Zúñiga, R., Scholz, D., Aguilar-Meléndez, A., Puente, J.D.L., 2019. A stochastic rupture earthquake code based on the fiber bundle model (TREMOL v0. 1): application to Mexican subduction earthquakes. Geosci. Model Dev. (GMD) 12 (5), 1809–1831.
Nychka, D., 1988. Bayesian confidence intervals for smoothing splines. J. Am. Stat. Assoc. 83 (404), 1134–1143.
Pall, J., Chandra, R., Azam, D., Salles, T., Webster, J.M., Scalzo, R., Cripps, S., 2020. Bayesreef: a bayesian inference framework for modelling reef growth in response to environmental change and biological dynamics. Environ. Model. Software 125, 104610.
Patzkowsky, M.E., Smith, L.H., Markwick, P.J., Engberts, C.J., Gyllenhaal, E.D., 1991. Application of the fujita-ziegler paleoclimate model: early permian and late cretaceous examples. Palaeogeogr. Palaeoclimatol. Palaeoecol. 86 (1–2), 67–85.
Phipps, S.J., McGregor, H.V., Gergis, J., Gallant, A.J., Neukom, R., Stevenson, S., Ackerley, D., Brown, J.R., Fischer, M.J., Van Ommen, T.D., 2013. Paleoclimate data–model comparison and the role of climate forcings over the past 1500 years. J. Clim. 26 (18), 6915–6936.
Reichstein, M., Camps-Valls, G., Stevens, B., Jung, M., Denzler, J., Carvalhais, N., et al., 2019. Deep learning and process understanding for data-driven earth system science. Nature 566 (7743), 195–204.
Reid, A.S., O'Callaghan, S., Bonilla, E., McCalman, L., Rawling, T., Ramos, F., 2013. Bayesian joint inversions for the exploration of earth resources. In: Twenty-Third International Joint Conference on Artificial Intelligence, pp. 2877–2884.
Ritz, S.P., Stocker, T.F., Joos, F., 2011. A coupled dynamical ocean–energy balance atmosphere model for paleoclimate studies. J. Clim. 24 (2), 349–375.
Sanso, B., Guenni, L., 2000. A nonstationary multisite model for rainfall. J. Am. Stat. Assoc. 95 (452), 1089–1100.
Scalzo, R., Kohn, D., Olierook, H., Houseman, G., Chandra, R., Girolami, M., Cripps, S., 2019. Efficiency and robustness in Monte Carlo sampling for 3-d geophysical inversions with obsidian v0.1.2: setting up for success. Geosci. Model Dev. (GMD) 12 (7), 2941–2960. https://doi.org/10.5194/gmd-12-2941-2019. https://www.geosci-model-dev.net/12/2941/2019/.
Sellwood, B.W., Valdes, P.J., 2006. Mesozoic climates: general circulation models and the rock record. Sediment. Geol. 190 (1–4), 269–287.
Sen, M.K., Stoffa, P.L., 1996. Bayesian inference, Gibbs' sampler and uncertainty estimation in geophysical inversion. Geophys. Prospect. 44 (2), 313–350.
Steiger, N.J., Hakim, G.J., Steig, E.J., Battisti, D.S., Roe, G.H., 2014. Assimilation of time-averaged pseudoproxies for climate reconstruction. J. Clim. 27 (1), 426–441.
Stidd, C., 1953. Cube-root-normal precipitation distributions. Eos Trans. Am. Geophys. Union 34 (1), 31–35. https://doi.org/10.1029/TR034i001p00031.
Stocker, T.F., Mysak, L.A., Wright, D.G., 1992. A zonally averaged, coupled ocean-atmosphere model for paleoclimate studies. J. Clim. 5 (8), 773–797.
Tingley, M.P., Huybers, P., 2010. A bayesian algorithm for reconstructing climate anomalies in space and time. part i: development and applications to paleoclimate reconstruction problems. J. Clim. 23 (10), 2759–2781.
Tingley, M.P., Craigmile, P.F., Haran, M., Li, B., Mannshardt, E., Rajaratnam, B., 2012. Piecing together the past: statistical insights into paleoclimatic reconstructions. Quat. Sci. Rev. 35, 1–22. https://doi.org/10.1016/j.quascirev.2012.01.012.
Wahba, G., 1990. Spline models for observational data. Society for Industrial and Applied Mathematics.
Wainwright, M.J., Jordan, M.I., 2008. Graphical models, exponential families, and variational inference. Found. Trends Mach. Learn. 1 (1–2), 1–305.
Wang, Z., Mysak, L.A., 2000. A simple coupled atmosphere–ocean–sea ice–land surface model for climate and paleoclimate studies. J. Clim. 13 (6), 1150–1172.
Wood, S., 2013. Applications of Bayesian smoothing splines. In: Damien, P., Dellaportas, P., Polson, N.G., Stephens, D.A. (Eds.), Bayesian Theory and Applications. OUP Oxford, pp. 1–10.
Zachos, J., Pagani, M., Sloan, L., Thomas, E., Billups, K., 2001. Trends, rhythms, and aberrations in global climate 65 ma to present. Science 292 (5517), 686–693.
Ziegler, A., Rowley, D.B., Lottes, A.L., Sahagian, D.L., Hulver, M.L., Gierlowski, T.C., 1985. Paleogeographic interpretation: with an example from the mid-cretaceous. Annu. Rev. Earth Planet Sci. 13 (1), 385–428.
Ziegler, A., Raymond, A., Gierlowski, T., Horrell, M., Rowley, D., Lottes, A., 1987. Coal, climate and terrestrial productivity: the present and early cretaceous compared. Geol. Soc. Lond. Spec. Publ. 32 (1), 25–49.
Ziegler, A., Eshel, G., Rees, P.M., Rothfus, T., Rowley, D., Sunderlin, D., 2003. Tracing the tropics across land and sea: permian to present. Lethaia 36 (3), 227–254.