An Emulator-Based Modelling Framework for Studying Astronomical Controls on Ocean Anoxia with an Application to the Devonian

Loïc Sablon[1], Pierre Maffre[2,3], Yves Goddéris[2], Paul J. Valdes[4], Justin Gérard[1], Jarno J.C. Huygh[5], Anne-Christine Da Silva[5], and Michel Crucifix[1]

[1]Earth and Life Institute, UCLouvain, Louvain-la-Neuve, Belgium
[2]Géosciences Environnement Toulouse, CNRS–Université Paul Sabatier, Toulouse, France
[3]Aix-Marseille Université, CNRS, IRD, INRA, Coll. France, CEREGE, Aix-en-Provence, France
[4]School of Geographical Sciences, University of Bristol, Bristol BS8 1SS, United Kingdom
[5]SediCClim, ULiège, Liège, Belgium

Correspondence: Loïc Sablon (loic.sablon@uclouvain.be)

Abstract. We present a modelling framework to study the response of continental flux dynamics and ocean anoxia to astronomical forcing. The GEOCLIM model is coupled with a Gaussian Process-based climate emulator, designed to efficiently capture the global distribution of temperature and precipitation as simulated by the general circulation model HadCM3. The emulator employs principal component analysis for dimensionality reduction. Compared to earlier approaches, our emulator features an additive kernel function that better captures the spatial complexity of ocean responses and accounts for ocean heat transport. This setup facilitates interactive coupling between CO_2 levels through an iterative procedure involving GEOCLIM and the emulator, enabling systematic exploration of various orbital and pCO_2 configurations. We demonstrate the model's capabilities with a Devonian case study, revealing an emergent relationship between high eccentricity periods and enhanced regolith development, though with limited impact on phosphorus fluxes to the ocean.

1 Introduction

Modelling long-term climate dynamics and related environmental phenomena presents significant challenges, particularly when investigating events from deep geological time. These challenges stem from uncertainties in boundary conditions, forcing mechanisms, and geological reconstructions, which complicate parameter tuning, as well as from the long timescales involved. A fundamental computational bottleneck emerges when attempting to couple detailed climate modelling with biogeochemical processes over geological timescales: while spatially resolved climate information is essential for understanding continental weathering and nutrient transport, the computational demands of General Circulation Models (GCMs) make multimillion-year integrations impractical.

The GEOCLIM framework (Donnadieu et al., 2006; Arndt et al., 2011; Goddéris and Donnadieu, 2019; Maffre et al., 2021) has been developed as a versatile suite of models designed to dynamically simulate geochemical cycles and regolith (i.e., the upper layer of the continental crust where chemical weathering occurs) dynamics over long timeframes. Its computational

efficiency stems from a box model approach for ocean biogeochemistry, combined with a spatially resolved continental module that captures weathering and erosion processes. However, GEOCLIM faces a critical limitation: realistically estimating the response of atmospheric conditions, particularly the impact of astronomical forcing on rainfall patterns that control nutrient mobilization, requires detailed climate inputs that are computationally prohibitive to generate over geological timescales using traditional GCM approaches.

Climate emulation provides an elegant solution to this incompatibility. By training statistical models on carefully designed ensembles of GCM experiments, emulators can capture the spatial complexity of the ocean-atmosphere response to astronomical forcing as simulated by the GCM, while requiring minimal computational resources once trained. This efficiency enables systematic exploration of multiple astronomical scenarios and their impact on nutrient fluxes and biogeochemical cycles, bridging the gap between detailed climate modelling and long-term Earth system investigations.

In this work, we advance the emulation paradigm by developing a dynamically coupled system that introduces several key methodological innovations. Building upon the framework proposed by Bounceur et al. (2015); Lord et al. (2017); Van Breedam et al. (2021) and the theory of Gaussian Processes (GPs, MacKay et al., 1998; Rasmussen, 2006), our approach features an additive kernel function that better captures spatial complexity in climate responses, a two-tier approach using both coupled and slab ocean models to account for ocean heat transport, and full bidirectional coupling that enables interactive feedback between atmospheric CO_2 levels and climate patterns. The geochemical model determines CO_2 levels, which the emulator then translates into climate predictions affecting continental weathering and erosion, and consequently nutrient fluxes. These, in turn, influence subsequent geochemical calculations, allowing for an iterative feedback process that can reveal how astronomical forcing ultimately impacts continental fluxes and ocean biogeochemistry.

The astronomical forcing of interest involves changes in the rotation of the Earth's axis and the geometry of its orbit, which impact seasonal variations and the level of incoming solar radiation, thereby affecting global climate (see Berger and Yin (2021) or Zeebe and Kocken (2024) for an introduction on astronomical forcing). These orbital cycles operate on timescales ranging from tens of thousands to millions of years, creating long-term modulations in climate patterns that can influence continental weathering, nutrient storage and release, and ultimately ocean biogeochemistry. Understanding these connections requires modelling approaches capable of capturing both the spatial complexity of climate responses and the temporal scales of biogeochemical processes.

To validate our methodological advances, we demonstrate the framework's capabilities with an application to the Devonian period (419 to 359 million years ago). The Devonian serves as an excellent test case, as it was characterised by significant climatic and environmental transformations, including 29 instances of ocean anoxic or hypoxic conditions (Becker et al., 2020). Climate modelling studies have shown this period to be sensitive to changes in atmospheric CO_2 concentrations and orbital configuration (Brugger et al., 2019), making it ideal for testing our emulator framework. It has been hypothesized that astronomical cycles may influence the recurrence of anoxia through their effects on regional climate patterns, particularly tropical precipitation, which controls nutrient mobilization and ocean biogeochemistry (De Vleeschouwer et al., 2017; Wichern et al., 2024). Testing these hypotheses requires precisely the type of spatially and temporally resolved modelling that our emulator-GEOCLIM coupling provides.

The primary objective of this paper is to provide a comprehensive description of our emulator-based modelling framework, along with the justification for our technical choices. This serves as a methodological foundation for investigating how astronomical forcing impacts continental fluxes and ocean biogeochemistry across geological timescales. It is essential to ensure that the GP emulator is robust enough for integration into a coupled soil-biogeochemical model. Though we illustrate its application with the Devonian case study, the methodology remains applicable to other geological periods and research questions requiring long-term Earth system modelling.

2 Methods

Following the definition by Kennedy and O'Hagan (2000); Oakley and O'Hagan (2002); Wilkinson (2010), an emulator is a statistical model that predicts the output of climate conditions provided by a 'simulator', here a GCM, in response to input parameters. The emulator must first be trained with a set of experiments with the GCM; the experiment set is designed such as to efficiently probe an input parameter space. Once trained, the emulator enables swift approximations of climate outputs, thereby allowing us to explore scenarios that would be practically impossible to produce through a traditional GCM usage.

To construct the climate database used to train the emulator, we decided to employ two types of GCMs: a coupled ocean-atmosphere model and a lighter, computationally less expensive model where the ocean is represented by a flat layer called a slab (we hereafter call it the slab model). We follow a two-tier approach by which we calibrate the slab model with the coupled one, and then produce a larger set of experiments with the slab model to generate the data needed to train the emulator. This double-layer strategy enables us to capture essential atmospheric dynamics accounting for changes in ocean heat transport while mitigating the computational cost. We now give the technical details of this procedure.

2.1 Geophysical models setup and boundary conditions

2.1.1 HadCM3L and HadSM3

The Hadley Centre Coupled Model version 3 Low Resolution (HadCM3L) serves as our primary reference GCM for simulating Devonian climate conditions. The spatial resolution of $3.75°$ longitude $\times$ $2.5°$ latitude for both the atmosphere and oceanic grid can adequately capture the essential mechanics of monsoonal systems. Albedo, snow accumulation, vegetation and soil processes were computed using the MOSES-II land-surface scheme (Cox et al., 1999, 2000; Cox, 2001; Essery et al., 2001). For a comprehensive description of the model's technical aspects, the reader is directed to its descriptive paper (Valdes et al., 2017).

HadSM3 is the slab version of HadCM3 where the atmosphere is coupled to a simple non-dynamic mixed layer ocean (Williams et al., 2005). This simpler model is computationally efficient in two ways: ocean dynamics are by-passed, and a stationary solution can be found after an integration of ca. 30 years, compared to 5000+ years with HadCM3 (Valdes et al., 2021). Hence, it will be used here to perform a larger number of experiments.
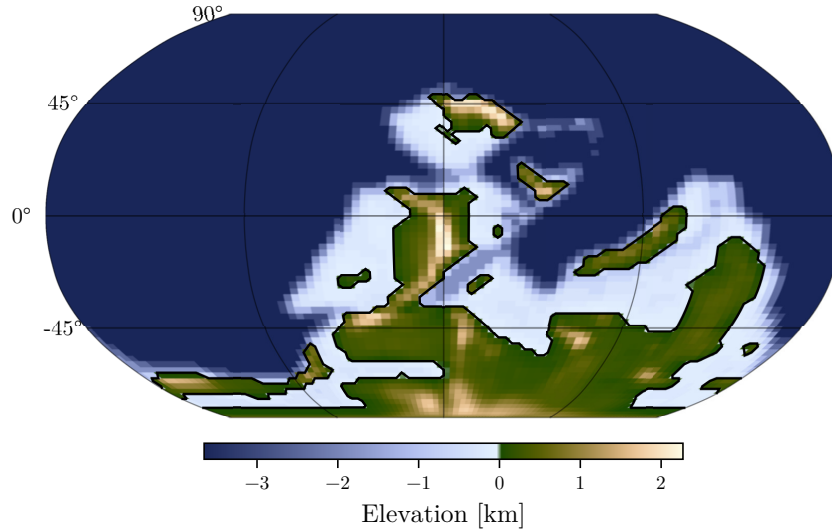
Figure 1. Topography of the Late Devonian (370 Ma) according to Scotese and Wright (2018) palaeo-digital elevation model, and downscaled to match the grid resolution of HadCM3L.

The model uses large input files called *dumps*, which contain initial and boundary conditions. Those used here are adapted from Valdes et al. (2021). The 370 Ma palaeogeography is obtained from the dataset of Scotese and Wright (2018), as depicted in Fig. 1. It was chosen with the intention to best represent the Earth's geometry during the Frasnian-Famennian boundary, coinciding with an OAE that triggered a mass extinction around 372 million years ago (Percival et al., 2018; Carmichael et al., 90   2019).

Given the slab ocean model's deficiency in ocean dynamics, the computation of a corrective flux, known as the Anomalous Heat Convergence (AHC), is required to accurately represent sea-surface temperatures (SSTs) and sea-ice concentration.

2.1.2 Anomalous Heat Convergence

In a slab model, the atmosphere is coupled to a much simpler resolution of the ocean, having a single layer. This coupling 95   has the advantage of still capturing the first order interaction of the ocean without the need of running experiments for several thousands of years, as the response is much faster. However, the model must be calibrated on realistic ocean to be able to capture its dynamic.

To transition from the coupled model (HadCM3L) to the slab model (HadSM3), we follow a two-step approach. First, a series of 15 experiments were conducted using the coupled version of the model with varying orbital parameters. These 100   experiments were run for 250 years, sufficient to achieve a near-equilibrium state in the upper ocean. For each experiment, the resulting SST and sea-ice concentration fields from HadCM3L were then prescribed to the slab model to diagnose the AHC.

This heat convergence field represents the heat transport that would have been provided by ocean dynamics in the coupled model, allowing the slab model to replicate similar climate patterns despite its simplified ocean representation. Second, we use an Inverse Distance Weighting interpolation (IDW) to obtain appropriate heat convergence fields for our full set of 81 experiments with different orbital configurations (cf. Fig. 4).

The 15 experiments using HadCM3 were designed with varying orbital parameters (see Fig. 4). After 250 years, the upper layer of the ocean was close to equilibrium, and the simulation was stopped (the deep ocean could take more than 3000 years of model run). The obtained SST and sea-ice concentration fields were then prescribed to the slab model to diagnose the AHC. At steady state, the temporal and spatial mean of the field should be equal to $0\ \mathrm{W\,m^{-2}}$. A non-zero value is equivalent to applying a constant forcing to the model. It appears that our set of experiments exhibits this non-desired behaviour, in part coming from the fact that the whole system was not exactly at equilibrium after the time period. In order to remove that parameter-dependent bias, a constant value is added to each such that the resulting average forcing is the same for each experiment, with a value of $-0.486\ \mathrm{W\,m^{-2}}$. Fig. 3 illustrates the average of the annual AHC across the set of 15 experiments.

Each of our 81 HadSM3 experiments should have a AHC field prescribed that is consistent with the information obtained from the 15 HadCM3 experiments. With 15 experiments only, it is more parsimonious and practical to estimate AHC using IDW interpolation, rather than with a more complex Gaussian Process Regression.

We designate $\boldsymbol{x}_i = [\tilde{e}_i \sin \varpi_i, \tilde{e}_i \cos \varpi_i, \tilde{\varepsilon}_i]^\top \in [-1, 1]^3$ as the vector that encapsulates the orbital parameters. These parameters are linearly rescaled to ensure that each one falls within the same range.

Subsequently, an interpolated value of the AHC can be computed as the weighted sum of the entire set. In this context, each weight is determined by the inverse square distance between the new point and the data, indicating that the weight is inversely proportional to the square of the distance.

$$
\mathbf{Y}(\boldsymbol{x}) = \begin{cases} \mathbf{Y}_i & \text{if} \quad \boldsymbol{x} = \boldsymbol{x}_i \\ \dfrac{\sum_{i=1}^{N} (\boldsymbol{x} - \boldsymbol{x}_i)^{-2} \mathbf{Y}_i}{\sum_{k=1}^{N} (\boldsymbol{x} - \boldsymbol{x}_k)^{-2}} & \text{otherwise} \end{cases} \tag{1}
$$

2.1.3 Simulation setup

For all simulations, the palaeogeography and the solar constant ($1322\ \mathrm{W\,m^{-2}}$, Bahcall et al., 2001) are kept identical. The orbital parameters (eccentricity, longitude of the perihelion, and obliquity), the atmospheric CO_2 concentration and the monthly heat-correction flux are constant in time within a given experiment but prescribed to different values across experiments. Once these boundary conditions and parameters are established, HadSM3 is run for 80 model years, with the last 30 years averaged to minimise interannual variability and accurately represent annual climatic averages of key variables, such as temperature, evaporation, and precipitation.
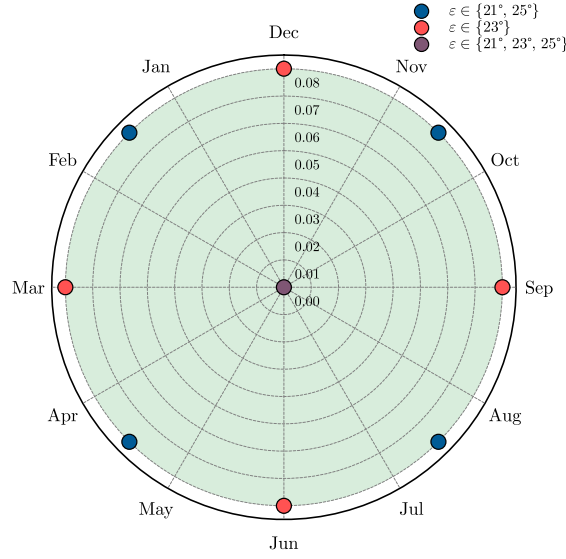
Figure 2. Slices of the cylinder constituting our HadCM3 experimental plan, with the eccentricity as the radius and the longitude of perihelion as the azimuthal angle. The angle is expressed in the month when the Earth is closest to the Sun, such that the vertical axis is the climatic precession $e \sin \varpi$ and the horizontal axis is the coprecession $e \cos \varpi$. The colour is associated to obliquity ε being the third dimension. The green area illustrates the possible values used during the interpolation process.

2.1.4 GEOCLIM

GEOCLIM (Donnadieu et al., 2006; Arndt et al., 2011; Goddéris and Donnadieu, 2019; Maffre et al., 2021) is a versatile suite of models designed to dynamically simulate geochemical cycles over geological timescales. The core of the model structure is COMBINE, an ocean-atmosphere chemistry model based on advection-reaction principles, using a discretisation into 10 boxes representing nine oceanic and one atmospheric compartments. More precisely, we use version 6.1.0, which differs slightly from its newest version GEOCLIM7 (Maffre et al., 2024). However, the continental modules of both versions are the same, which is the key point of this study. The coupling with the emulator developed in this work creates the integrated framework we designate as `GEOCLIM6.1-EMUL`.

The nine oceanic boxes represent the following spatial and depth discretisation: polar deep and surface (including thermocline) for both northern and southern hemispheres, mid-latitude deep, intermediate (thermocline), and surface, and epicontinental deep and surface (receiving the continental fluxes). This box structure enables the representation of the main oceanic water masses and their biogeochemical characteristics whilst maintaining computational efficiency for long-term simulations.

This model is integrated with an early diagenesis module responsible for calculating output fluxes, such as the burial of elements in marine sediments across each box. Additionally, GEOCLIM features a continental module, which computes input
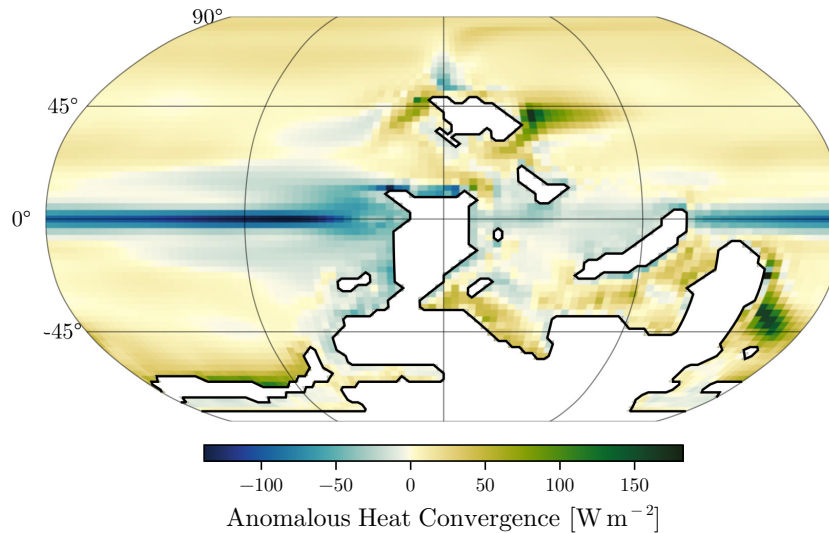
Figure 3. Ensemble mean of the annual Anomalous Heat Convergence, expressed in $\mathrm{W\,m^{-2}}$.

fluxes and is spatially resolved using a geographic mesh grid. For this study, the continental module is configured to use
the same Late Devonian paleogeography (370 Ma) as employed in the HadCM3/HadSM3 simulations, ensuring consistency
between the climate emulator and biogeochemical model. This module is adaptable and includes a soil model based on previous
work by Gabet and Mudd (2009) and West (2012).

GEOCLIM computes various geochemical cycles, including those of carbon, oxygen, alkalinity, phosphorus, and sulfur,
within each reservoir. It dynamically simulates these cycles without assuming steady-state conditions, incorporating rapid
processes like ocean mixing and water column sedimentation under parametrised forms. However, it necessitates inputs from
a climate model.

Within the continental module of GEOCLIM, Dynsoil accounts for physical erosion and silicate weathering. The other fluxes
computed by the continental module are carbonate weathering, petrogenic organic carbon weathering, phosphorus weathering,
and terrestrial organic carbon export. Surface bedrock is categorized into six lithological classes, following Park et al. (2020),
utilising the data compilation of Hartmann and Moosdorf (2012). The erosion and silicate weathering components solve the
governing equations dynamically instead of assuming a steady-state regolith, enabling the model to capture transient responses
to climate perturbations and the coupling between physical erosion and chemical weathering rates. Regarding petrogenic
organic carbon and sulfide weathering, these fluxes are considered proportional to the modelled erosion rate with prescribed
organic matter content.

A key aspect of Dynsoil is that it resolves the regolith thickness and computes the weathering fluxes accordingly. Here, we
call *regolith* the interface between unweathered bedrock and the Earth's surface where chemical weathering reactions occur.

It can be thought of as the layer of unconsolidated material that covers the bedrock. The regolith model describes a vertical profile of the abundance of primary minerals, beginning with an abundance of 1 at the regolith/bedrock transition and decaying towards the surface as a result of dissolution.

165 In GEOCLIM6.1.0, phosphorus weathering represents a key biogeochemical flux, as being the only nutrient represented, delivered to the ocean proportionally through three distinct weathering sources: silicate, carbonate, and petrogenic organic carbon. Critically, only silicate weathering is dynamically computed within Dynsoil and therefore directly affected by regolith thickness variations. Carbonate weathering and petrogenic organic carbon weathering contribute to phosphorus delivery but are computed independently of regolith dynamics. This coupling between regolith thickness and silicate weathering explains

170 why changes in regolith dynamics may not always translate directly to proportional changes in total phosphorus fluxes, as the regolith-dependent component represents only one of the three phosphorus sources.

The Dynsoil module needs inputs of temperature and runoff (or precipitation and evaporation) over a spatial grid, which are typically supplied by a GCM or, here, the emulator of the GCM.

In summary, assuming adequate climate forcing, GEOCLIM's integration of continental processes, climate dynamics, and

175 ocean biogeochemistry enables comprehensive investigation of Earth system dynamics over multimillion-year timescales as needed for our objective.

2.2 Experimental design

2.2.1 Parameter space

The inner solar system's dynamics exhibit chaotic behaviour, making it impossible to pinpoint an exact astronomical solution

180 for the Late Devonian period. Only a handful of constraints on cycle frequencies, stabilities and amplitudes remain (Laskar et al., 2004, 2011; Mogavero and Laskar, 2022; Zeebe and Kocken, 2024; Zeebe and Lantink, 2024).

Since the objective of this study is to construct climatologies for a plausible Late Devonian orbital configuration, it requires an extensive exploration of a broad domain. We focus on the response to four parameters: eccentricity e (which characterises the shape of the Earth's orbit), longitude of the perihelion ϖ (the angular distance between the September equinox and the

185 point in Earth's orbit where it is closest to the Sun – heliocentric convention), obliquity ε (the tilt of the ecliptic relative to the celestial equator), and the partial pressure of carbon dioxide in the atmosphere $p\mathrm{CO}_2$.

The following intervals for each parameter were chosen to make sure that no extrapolation outside the domain will be needed:

$$e \in [0, 0.075] \tag{2}$$

190
$$\varpi \in [0, 2\pi[\tag{3}$$

$$\varepsilon \in [20.75°, 23.75°] \tag{4}$$

$$p\mathrm{CO}_2 \in [1, 16] \times 280 \text{ ppm} \tag{5}$$

We propose the hypothesis that the prior basis $\{e\sin\varpi, e\cos\varpi, \varepsilon, \log_2(p\mathrm{CO}_2)\}$, where $e\sin\varpi$ denotes the climatic precession and $p\mathrm{CO}_2$ is expressed relative to pre-industrial levels (i.e., in units of 280 ppm), optimally encodes the initial distance metric among different climatologies. A more refined metric will be learned during the training phase of the GP emulator. This hypothesis is anchored in the interpretation of the climatic precession, where a large positive value signifies an excess of insolation in the Southern Hemisphere during the austral summer, and a large negative value indicates the same in the Northern Hemisphere during the boreal summer. Moreover, our prior judgement is that the globally averaged temperature responds by the same variation to every doubling of carbon dioxide concentration, hence the logarithm. Accordingly, we also define the rescaled CO_2 parameter:

$$m = \log_2\left(\frac{p\mathrm{CO}_2}{280\text{ ppm}}\right) \tag{6}$$

2.3 Construction and optimisation of the training set

To construct the training set for the GP emulator, it is essential to sample the parameter space effectively to cover all regions of interest. For this reason, we use Latin Hypercube Sampling (LHS, Eglajs and Audze, 1977; Iman et al., 1981; McKay et al., 2000). This method divides the range of each input parameter into equally probable intervals and selects one sample from each interval. Unlike direct random sampling, LHS provides a more uniform and comprehensive exploration of the parameter space, reducing redundancy and improving the representativeness of sampled points. This method is widely recognized for enhancing emulator accuracy and robustness (Urban and Fricker, 2010).

In this study, the LHS was further optimised using the Audze-Eglājs (AE) criterion (Eglajs and Audze, 1977; Eliáš and Vořechovský, 2016). The AE criterion was developed to optimise the spatial arrangement of points in a hypercube. This is achieved by expressing the potential energy of a system of particles where each pair of particles exerts a repulsive force on each other. These forces are functions of the distance between the pairs of points. By minimising this potential energy, the AE criterion ensures a uniform and space-filling distribution of experimental points.

The optimisation was executed in Julia, a high-level general-purpose dynamic programming language designed for high-performance numerical analysis and computational science (Bezanson et al., 2012), using a genetic algorithm (Bates et al.; Urquhart et al., 2020).

For our application, we must also account for some specificities of the input parameters. Indeed, uniformly sampling the climatic precession and coprecession would lead to unwanted high eccentricity experiments. This is resolved by making a change of variables and sampling uniformly $\{e^2, \varpi, \varepsilon, m\}$. Doing so also requires adapting the distance function in the optimisation algorithm from Euclidean to

$$d_{ij} = \tilde{e}_i^2 + \tilde{e}_j^2 - 2\tilde{e}_i\tilde{e}_j\cos(\varpi_i - \varpi_j) + (\tilde{\varepsilon}_i - \tilde{\varepsilon}_j)^2 + (\tilde{m}_i - \tilde{m}_j)^2 \tag{7}$$

where the tilde signifies that the value has been linearly rescaled to $[0, 1]$ interval for the eccentricity and $[-1, 1]$ for the obliquity and CO_2. This is necessary because the relative importances of eccentricity-modulated climate precession and obliquity for our variables of interest are a priori unknown. Ultimately, our input space will be optimally filled; climatic precession and coprecession will follow a semicircular distribution, while obliquity and m will be uniformly distributed.
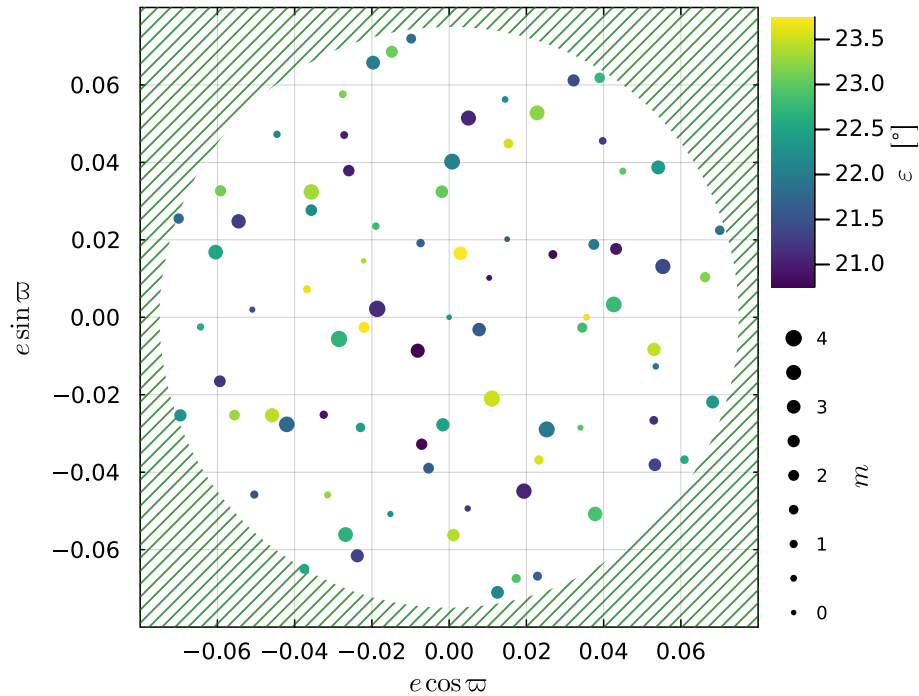
Figure 4. Projection of the experimental design onto the $\{e\sin\varpi,\, e\cos\varpi\}$ plane. Each point represents a GCM experiment. The colour is associated to obliquity ε (in degrees) and the marker size to carbon dioxide $m = \log_2\left(\frac{pCO_2}{280\text{ ppm}}\right)$. The dashed area is outside the range of our maximal eccentricity $e_{\max} = 0.075$.

In this work, based on the computational cost of HadSM3 and previous similar studies (Bounceur et al., 2015; Lord et al., 2017; Van Breedam et al., 2021), the number of samples has been arbitrarily fixed to 81.

3 Emulator development

The function of the emulator is to predict the output of HadSM3 for any input in the range spanned by the set produced with the procedure described in Sect. 2.3. Fundamentally, it is an interpolator: if it is well-designed and successfully validated, there is no need for further experiments using the GCM.

3.1 Theoretical basis of a Gaussian Process regression

We follow (MacKay et al., 1998; Rasmussen, 2006) Gaussian Processes (GPs), and specifically, Gaussian Process Regression (GPR) to execute the interpolation procedure.

A GP (MacKay et al., 1998; Rasmussen, 2006) is a stochastic process, generally comprising random variables indexed by time or space. Its unique property is that any finite collection of these variables adheres to a multivariate Gaussian distribution.

Therefore, a GP is a distribution over functions with a continuous domain, effectively describing a probability distribution over an infinite-dimensional vector space.

Let $\mathcal{Z} \subseteq \mathbb{R}^{n_z}$ be the index set, which is the number of input variables (in our case, $n_z = 4$). The function $f_{\mathrm{GP}}(\boldsymbol{z})$ with $\boldsymbol{z} \in \mathcal{Z}$ is a random variable. A GP is a stochastic process completely described by a mean function $m : \mathcal{Z} \to \mathbb{R}$ and a covariance function $k : \mathcal{Z} \times \mathcal{Z} \to \mathbb{R}$ such that

$$f_{\mathrm{GP}}(\boldsymbol{z}) \sim \mathcal{GP}\left(m(\boldsymbol{z}), k(\boldsymbol{z}, \boldsymbol{z}')\right) \tag{8}$$

$$m(\boldsymbol{z}) = \mathbb{E}\left[f_{\mathrm{GP}}(\boldsymbol{z})\right] \tag{9}$$

$$k(\boldsymbol{z}, \boldsymbol{z}') = \mathbb{E}\left[\left(f_{\mathrm{GP}}(\boldsymbol{z}) - m(\boldsymbol{z})\right)\left(f_{\mathrm{GP}}(\boldsymbol{z}') - m(\boldsymbol{z}')\right)\right] \tag{10}$$

The covariance function serves as a measure of the correlation between two states $(\boldsymbol{z}, \boldsymbol{z}')$, and within the framework of GPs, is known as the kernel.

In Bayesian inference, the GP serves as the prior probability distribution, which can be used for function regression. This Bayesian approach merges new data with existing knowledge. Specifically, Bayes' theorem is employed to integrate the prior with new data, resulting in a posterior distribution. The new data is represented as a training dataset $\mathcal{D} = \{\mathbf{X}, \mathbf{Y}\}$ where $\mathbf{X} = \left[\boldsymbol{x}_{\mathrm{dat}}^{\{1\}}, \ldots, \boldsymbol{x}_{\mathrm{dat}}^{\{n_{\mathcal{D}}\}}\right] \in \mathcal{Z}^{1 \times n_{\mathcal{D}}}$ (in our case $n_{\mathcal{D}} = 81$) is the input matrix and $\mathbf{Y} = \left[\tilde{y}_{\mathrm{dat}}^{\{1\}}, \ldots, \tilde{y}_{\mathrm{dat}}^{\{n_{\mathcal{D}}\}}\right]^{\top} \in \mathbb{R}^{n_{\mathcal{D}} \times 1}$ contains the output values. The one-dimensional regression problem takes the following form

$$\tilde{y}_{\mathrm{dat}}^{\{i\}} = f_{\mathrm{GP}}(\boldsymbol{x}_{\mathrm{dat}}^{\{i\}}) + \epsilon \tag{11}$$

for all $i \in \{1, \ldots, n_{\mathcal{D}}\}$ and $\epsilon \sim \mathcal{N}(0, \nu^2)$ is i.i.d. Gaussian measurement noise.

Considering that each finite subset of a GP follows a multivariate Gaussian distribution, we can construct the joint distribution for a random test point $\boldsymbol{z}^* \in \mathcal{Z}$:

$$\begin{bmatrix} \mathbf{Y} \\ f_{\mathrm{GP}}(\boldsymbol{z}^*) \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} m(\boldsymbol{x}_{\mathrm{dat}}^{\{1\}}) \\ \vdots \\ m(\boldsymbol{x}_{\mathrm{dat}}^{\{n_{\mathcal{D}}\}}) \\ m(\boldsymbol{z}^*) \end{bmatrix}, \begin{bmatrix} \mathbf{K}(\mathbf{X}, \mathbf{X}) + \nu^2 \mathbf{I}_{n_{\mathcal{D}}} & \mathbf{k}(\boldsymbol{z}^*, \mathbf{X}) \\ \mathbf{k}(\boldsymbol{z}^*, \mathbf{X})^{\top} & k(\boldsymbol{z}^*, \boldsymbol{z}^*) \end{bmatrix} \right) \tag{12}$$

In this expression, the matrix function $\mathbf{K} : \mathcal{Z}^{1 \times n_{\mathcal{D}}} \times \mathcal{Z}^{1 \times n_{\mathcal{D}}} \to \mathbb{R}^{n_{\mathcal{D}} \times n_{\mathcal{D}}}$, is called the covariance or Gram matrix and is related to the kernel via $K_{j,l} = k(\boldsymbol{X}_{:,l}, \boldsymbol{X}_{:,j})$ for all $j, l \in \{1, \ldots, n_{\mathcal{D}}\}$.

The posterior predictive distribution of $f_{\mathrm{GP}}(\boldsymbol{z}^*)$ is obtained by conditioning on the test point and the dataset $\mathcal{D}$, using Bayes' theorem:

$$\mathrm{p}(f_{\mathrm{GP}}(\boldsymbol{z}^*)|\boldsymbol{z}^*, \mathcal{D}) = \frac{\mathrm{p}(f_{\mathrm{GP}}(\boldsymbol{z}^*), \mathbf{Y}|\mathbf{X}, \boldsymbol{z}^*)}{\mathrm{p}(\mathbf{Y}|\mathbf{X})} \tag{13}$$

The conditional posterior is a Gaussian distribution described by the mean and variance:

$$\mu(f_{\mathrm{GP}}(\boldsymbol{z}^*)|\boldsymbol{z}^*, \mathcal{D}) = m(\boldsymbol{z}^*) + \mathbf{k}(\boldsymbol{z}^*, \mathbf{X})^{\top}\left(\mathbf{K}(\mathbf{X}, \mathbf{X}) + \nu^2 \mathbf{I}_{n_{\mathcal{D}}}\right)^{-1}\left(\mathbf{Y} - [m(\boldsymbol{X}_{:,1}), \ldots, m(\boldsymbol{X}_{:,n_{\mathcal{D}}})]^{\top}\right) \tag{14}$$

$$\mathbb{V}\mathrm{ar}(f_{\mathrm{GP}}(\boldsymbol{z}^*)|\boldsymbol{z}^*, \mathcal{D}) = k(\boldsymbol{z}^*, \boldsymbol{z}^*) - \mathbf{k}(\boldsymbol{z}^*, \mathbf{X})^{\top}\left(\mathbf{K}(\mathbf{X}, \mathbf{X}) + \nu^2 \mathbf{I}_{n_{\mathcal{D}}}\right)^{-1}\mathbf{k}(\boldsymbol{z}^*, \mathbf{X}) \tag{15}$$

265   A comprehensive detail of the derivation of these formulae can be found in Appendix A of Beckers (2021).

The prior mean function $m(\cdot)$ characterises the average function under the GP distribution before any data is observed. This function presents a simple approach to include prior knowledge about the function we desire to model. An idealized physics model can hence be used as a prior. The GP then models the difference between this prior and the empirical data. When there is no such prior knowledge, a typical choice is to set the prior mean function to zero, i.e., $m(\cdot) \equiv 0$, but following our

270   considerations about the physical climate response discussed in Sect. 2.3, we rather adopt an affine mean function of the form

$$m(\boldsymbol{x}) = \beta_0 + \boldsymbol{\beta} \cdot \boldsymbol{x} + \beta_e \sqrt{x_1^2 + x_2^2} \tag{16}$$

where the last term accounts for a possible effect of the eccentricity. The coefficients β_i are chosen to minimise the sum of the squares of the residuals.

3.2   Application to climate fields

275   The regression approach discussed in the previous section is only applicable for sampling one-dimensional functions. However, when we model the GCM outcomes similarly to $\mathbf{f}_{\mathrm{GCM}}(\boldsymbol{z}) + \boldsymbol{\Sigma}$, we are confronted with a dimensionality challenge as the function's image is now of high dimension (a dimension equal to the number of grid points, or pixels). There are several strategies to manage this: using a Multi-Output Gaussian Process (MOGP) or an *outer-product* emulator (Rougier et al., 2009), creating a different GP for each dimension, and/or applying a dimensionality reduction technique (Holden and Edwards, 2010;

280   Wilkinson, 2010). Although a MOGP could theoretically yield better results, its implementation for a GCM longitude-latitude grid is computationally challenging. The design of an appropriate kernel function would be highly complex, and the computational cost of GPR grows cubically with the number of samples ($\mathcal{O}(n_{\mathcal{D}}^3)$), making this approach impractical. In our study, we chose the third strategy and used Principal Component Analysis (PCA) as the dimensionality reduction technique. This choice was driven by our goal of preserving as many teleconnections within the data as possible. Constructing a different GP

285   for each pixel would eliminate many of these correlations. Moreover, the statistical interpretation of the principal components is advantageous, as they can be interpreted as either signal or noise. Furthermore, as the principal components are designed to be uncorrelated with each other, the use of a MOGP appears superfluous.

3.3   Kernel function

Equation 14 shows that the selection of the kernel function may significantly influence the posterior mean. The kernel is crucial

290   in expressing how the output associated with two neighbouring inputs are related. Hence, an inappropriate kernel choice or design can result in less accurate, or even incorrect, output (Rasmussen, 2006).

Selecting the appropriate function design with the correct number of hyperparameters is a complex task, but some studies offer an automated method for this search (Duvenaud et al., 2013). The most popular covariance function is the squared exponential kernel (SE-Kernel), as favoured by Kennedy and O'Hagan and Higdon et al. (2008).

295   In this study, we initially faced some issues when applying existing models to the Devonian dataset and decided to search for a more general kernel function compared to Bounceur et al. (2015); Lord et al. (2017); Van Breedam et al. (2021). We

evaluated two kernel approaches and found performance differences that are quantified in Sect. 3.6. We settled on the work of Duvenaud et al. (2011) and employed an Additive Kernel, that presented slightly better performance. The kernel works as follows:

To each dimension $i \in \{1, \ldots, n_z\}$, we can assign a *base kernel* $k_i(\cdot, \cdot)$. An Additive Kernel is then the sum

$$k_{\text{add}}(\boldsymbol{z}, \boldsymbol{z}') = \sum_{n=1}^{n_z} k_n(\boldsymbol{z}, \boldsymbol{z}') \tag{17}$$

where the first, second and n^{th} terms are given by

$$k_1(\boldsymbol{z}, \boldsymbol{z}') = \sigma_1^2 \sum_{i=1}^{n_z} k_i(z_i, z_i') \tag{18}$$

$$k_2(\boldsymbol{z}, \boldsymbol{z}') = \sigma_2^2 \sum_{i_1=1}^{n_z} \sum_{i_2=i_1+1}^{n_z} k_{i_1}(z_{i_1}, z_{i_1}') k_{i_2}(z_{i_2}, z_{i_2}') \tag{19}$$

$$k_n(\boldsymbol{z}, \boldsymbol{z}') = \sigma_n^2 \sum_{1 \le i_1 < i_2 < \cdots < i_n \le n_z} \prod_{d=1}^{n} \tilde{k}_{i_d}(z_{i_d}, z_{i_d}') \tag{20}$$

where n_z is the dimension of the input space and $\sigma_n \in \mathbb{R}_0^+$. Specifically, the nth order interaction kernel k_n is the sum of $\binom{n_z}{n}$ terms. The relative magnitude of each σ_n informs about the importance of each order of interaction.

In this research, we have selected the Square Exponential Kernel (SE-Kernel) as the one-dimensional basis function:

$$\tilde{k}_{\text{SE}}(x, y; l) = \exp\left(-\frac{(x-y)^2}{2l^2}\right) \tag{21}$$

with $l \in \mathbb{R}_{>0}$ the lengthscale hyperparameter.

This kernel exhibits notable properties: it is stationary, meaning that it depends only on the distance between each point, and it is universal, implying that a GPR with such a kernel can approximate any continuous function with arbitrary precision on a compact set. Moreover, a SE-Kernel of high dimension can be seen as the product of multiple one-dimensional SE-Kernels.

The entire set of hyperparameters for the Additive Kernel is then defined by the $n_z = 4$ order variance $\{\sigma_1, \ldots, \sigma_{n_z}\}$ and lengthscales $\{l_1, \ldots, l_{n_z}\}$. These define a vector $\boldsymbol{\theta}$.

The first and last order terms in the sum are given by:

$$k_1(\boldsymbol{z}, \boldsymbol{z}') = \sigma_1^2 \sum_{i=1}^{n_z} \exp\left(-\frac{(z_i - z_i')^2}{2l_i^2}\right) \tag{22}$$

$$k_{n_z}(\boldsymbol{z}, \boldsymbol{z}') = \sigma_{n_z}^2 \exp\left(-\frac{1}{2} \sum_{i=1}^{n_z} \frac{(z_i - z_i')^2}{l_i^2}\right) \tag{23}$$

This Additive Kernel can be seen as an augmented SE-Kernel, as used in Lord et al. (2017). It maintains its properties and computational advantages while adding an optimal level of complexity. This complexity is enough to grasp useful information at a greater distance (see Fig. 5), but still small enough to avoid over-fitting, classically known to appear when parameters to fit are too numerous.

(a)	(b)
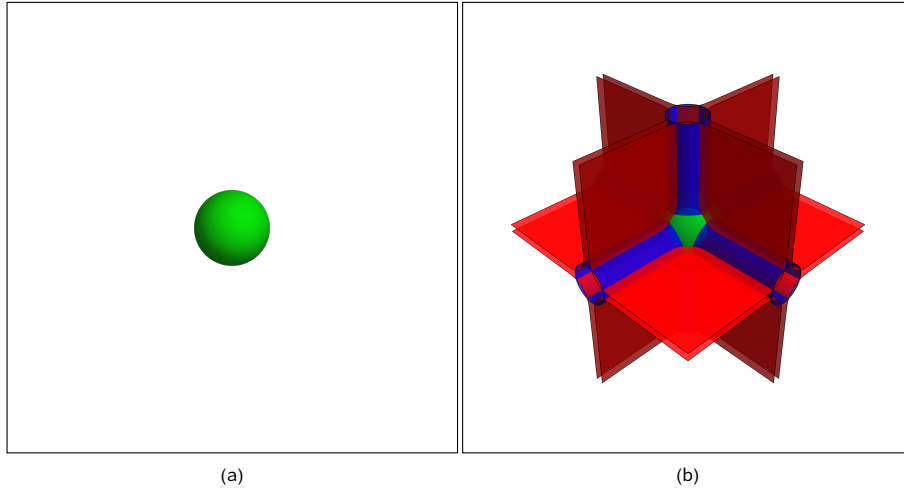
Figure 5. Isocontours of (a) a Square Exponential Kernel and (b) an Additive Kernel in a 3-dimensional space. The SE-Kernel is designed to give importance only to points in close proximity. In contrast, the lower-order kernels (in red and blue) are more inclusive. Reproduced from Duvenaud et al. (2011).

3.4 Objective function and optimisation

The GPR being non-parametric but the kernel function possesses free parameters, known as hyperparameters, that must be optimised to best fit the data.

The negative log marginal likelihood function (Eq. 24), as fully derived in Rasmussen (2006), serves as an objective measure of model fit. It quantifies the likelihood of observing the given data under specific hyperparameters. The goal is to minimise this function to find the hyperparameters that maximise the model's likelihood of generating the observed data.

$$\log \mathrm{p}(\mathbf{Y}|\mathbf{X},\boldsymbol{\theta}) = -\frac{1}{2}\tilde{\mathbf{Y}}^{\top}\left(\mathbf{K}(\mathbf{X},\mathbf{X};\boldsymbol{\theta})+\nu^2\mathbf{I}_{n_{\mathcal{D}}}\right)^{-1}\tilde{\mathbf{Y}} - \frac{1}{2}\log\left|\mathbf{K}(\mathbf{X},\mathbf{X};\boldsymbol{\theta})+\nu^2\mathbf{I}_{n_{\mathcal{D}}}\right| - \frac{1}{2}n_{\mathcal{D}}\log 2\pi \tag{24}$$

where we defined $\tilde{\mathbf{Y}} = \mathbf{Y} - [m(\boldsymbol{X}_{:,\mathbf{1}}),\ldots,m(\boldsymbol{X}_{:,\boldsymbol{n_{\mathcal{D}}}})]^{\top}$ to take into account a non-zero mean-function.

$$\begin{bmatrix}\boldsymbol{\theta}_{\mathrm{opt}} \\ \nu_{\mathrm{opt}}\end{bmatrix} = \arg\min_{\boldsymbol{\theta},\nu} -\log \mathrm{p}(\mathbf{Y}|\mathbf{X},\boldsymbol{\theta}) \tag{25}$$

The negative log marginal likelihood function effectively balances model complexity and goodness of fit.

Optimising the negative log marginal likelihood provides an approach to hyperparameter tuning with a theoretical justification. Furthermore, by design, it includes a complexity penalty that discourages overly complex models, effectively preventing overfitting and thus promoting the selection of parsimonious models.

Yet, its optimisation can be computationally intensive, and complicated by the presence of multiple local optima.

With these difficulties in mind, we implemented the GPR in the Julia language using the versatile AbstractGP.jl package (Widmann et al., 2024) and Zygote.jl (Innes, 2019) to efficiently compute the gradients. The optimisation scheme is based on a bounded probabilistic descent and a local L-BFGS gradient descent algorithm, both available within the Optimization.jl package (Dixit and Rackauckas, 2023). This method showed the best performance amongst common algorithms. We initiate the algorithm from a wide range of initial conditions to favour a global minimum. Since the values to emulate are rescaled to be of order 1, we expect the lengthscale and order-variance parameters to be of the same order. We first run 1024 optimisations at low tolerance from points lying in a Latin hypercube of bounds $[10^{-16}, 150]$ for each l_i, $[0, 2]$ for each σ_i and $[10^{-6}, 2]$ for ν^2, where a non-zero lower bound ensures numerical stability. The 64 lowest different minima are kept and used in a second optimisation round with a much lower tolerance. This procedure strongly favours a global optimum.

3.5 Data preprocessing

As discussed in Sect. 3.2, the GPR is not directly applied to the climate fields but rather to the PCA loading coefficients, building an independent GP for each component. PCA condenses the variability of multivariate observations into a set of orthogonal modes, known as principal components (PCs). These PCs encapsulate the most significant patterns in the data. The mathematical foundation of PCA involves the computation of eigenvalues and eigenvectors from the data's covariance matrix. The eigenvectors, also termed as loadings, signify the directions of maximum variance, while the eigenvalues quantify the variance explained along each eigenvector. These loadings offer insights into the relationships amongst the original variables and assist in identifying the primary modes of variability within the dataset.

The decomposition of the output in PCs is a linear operation (although the PCs themselves are non-linear), as illustrated by the following equations:

$$\boldsymbol{y} = \mathbf{P}_{\mathrm{GCM}}^{\top}(\boldsymbol{x} - \boldsymbol{\mu}_{\mathrm{GCM}}) \tag{26}$$

$$\tilde{\boldsymbol{x}} = \mathbf{P}_{\mathrm{GCM}}\boldsymbol{y} + \boldsymbol{\mu}_{\mathrm{GCM}} \tag{27}$$

In Eq. 26, $\mathbf{P}_{\mathrm{GCM}}$ denotes the matrix of loadings, whose columns are the orthonormal eigenvectors of the data. The term $\boldsymbol{x}$ signifies a climate field that has been reshaped into a vector and imputed for missing values, while $\boldsymbol{\mu}_{\mathrm{GCM}}$ represents the mean value of each grid cell, also reshaped into a vector. The term $\boldsymbol{y}$ is the projection of $\boldsymbol{x}$ onto the lower-dimensional vector space spanned by the PCs, the components of $\boldsymbol{y}$ are often referred to as the PCA scores. These scores are the scalar values interpolated using the GPR.

Eq. 27 demonstrates how to reconstruct the climate field from the lower-dimensional representation (denoted by $\tilde{\boldsymbol{x}}$). This reconstruction is approximate and will only be exact if the true value $\boldsymbol{x}$ resides within the vector space spanned by the data used to generate the projection matrix. This equation is employed to reconstruct the GCM outcomes using the results from the GPR.

For optimal performance of PCA, certain data assumptions must be satisfied. Specifically, the data should follow a Gaussian distribution to ensure the validity of statistical measures like covariance. Standardizing (i.e., subtracting the mean and dividing

by the standard deviation) is also a common procedure for ensuring comparability in scale across pixels, preventing dominance by points with larger magnitudes.

Standardisation of each pixel is typically applied to data that follows a Gaussian distribution, which is the case for our temperature data. Precipitation minus evaporation (or *runoff*, which often contains many zeros in arid grid cells) is conversely highly asymmetric and skewed. In that case, standardisation does not yield a Gaussian distribution via affine transformation of the data.

In this study, we chose to standardise temperature but not runoff before computing the PCA. This proved to be the most effective approach. Additionally, we opted to emulate the total annual water volume (i.e., runoff multiplied by area) rather than runoff directly. This decision is motivated by the unequal areas of grid cells. As a result, this procedure effectively assigns greater weight to tropical regions, which are of particular interest in our study.

The PCA scores themselves are also rescaled. Indeed, the magnitude of these scores is proportional to $\sqrt{\lambda}$, where λ is the variance associated with the principal component of interest (i.e., an eigenvalue). To ensure the stability of the optimisation algorithm and the interpretability of the objective function, we divide each coefficient by $2\sqrt{\lambda}$. This transformation ensures that the values to be interpolated typically lie between -1 and 1 and have the same scale across different principal components.

In principle, some experiments may exhibit a highly non-linear behaviour, e.g. by crossing some tipping points, and effectively behave as outliers. When such outliers occur, they deserve particular attention because they may complicate the learning process and justify a specific treatment.

The adopted strategy to detect these points is based on the PCA decomposition of the set. We assume that each component of $\boldsymbol{y}$ follows a Gaussian distribution. The mean and standard deviation of each principal component are computed while excluding a specific GCM outcome. This outcome is then projected onto the lower-dimensional vectorial space spanned by the remaining data, and its z-score is computed as follows:

$$z_{\mathrm{score},k} = \left| \frac{\tilde{y}_k^{\{i\}} - \mu_k^{\{-i\}}}{\sigma_k^{\{-i\}}} \right| \tag{28}$$

Here, the superscript $\{i\}$ is the experiment index, $\{-i\}$ signifies that μ and σ have been computed excluding the specific experiment, and the subscript k is the PC dimension. If the value of this score is higher than a specific threshold (e.g., 3 for a 99.73% accuracy) for the main contributing PCs, then the probability that the behaviour of this GCM run is an outlier is high and should be excluded, or at least specifically examined. In our set of 81 HadSM3 runs, though, none of them had to be excluded and no specific treatment was thus necessary.

3.6 Validation

The last step in our process is to determine the best number of PCs to emulate. Insufficient PCs could result in information loss, while an excess might introduce noise, given that the least significant components often resemble random fields attributable to the GCM internal variability.
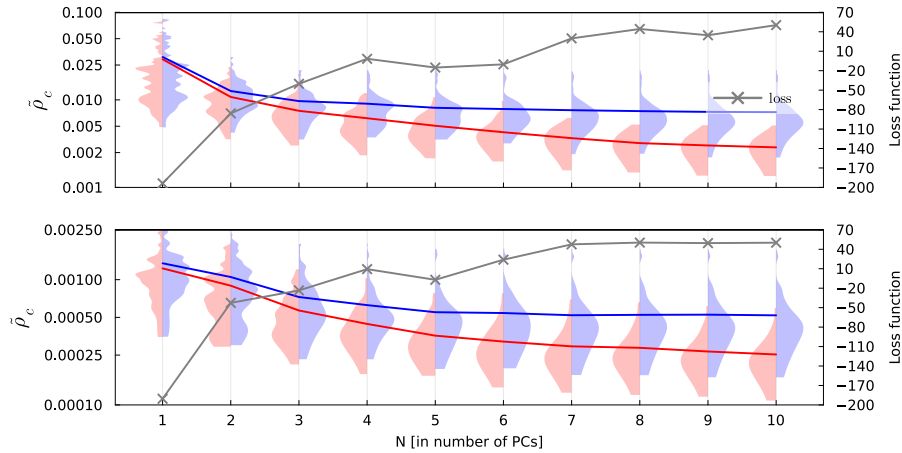
Figure 6. Shifted concordance correlation coefficient ($\tilde{\rho}_c = 1 - \rho_c$) distributions (left axis, in log-scale) and individual loss function value (right axis) for runoff (top) and surface temperature (bottom). In blue, the CCC has been computed on a sample while having excluded it during the learning process (*leave-one-out*; LOO). In red are depicted the CCC using the full set of experiments. Plain blue and red lines represent the average value across the whole set of 81 GCM runs, while the grey line indicates the objective function value for the specific N^{th} PC.

In this study, we opted for a dimensionless metric: the weighted Concordance Correlation Coefficient (CCC) (Lin, 1989), to quantify the accuracy of emulator-based reconstructions.

$$\rho_c(\boldsymbol{y}, \hat{\boldsymbol{y}}) = \frac{2\sum_i w_i(y_i - \mu_y)(\hat{y}_i - \hat{\mu}_y)}{\sum_i w_i(y_i - \mu_y)^2 + \sum_i w_i(\hat{y}_i - \hat{\mu}_y)^2 + (\mu_y - \hat{\mu}_y)^2} \tag{29}$$

where $\hat{y}_i$ denotes the prediction, and y_i the GCM output. This approach involves excluding a single data point during training and using it to validate the emulator's prediction. The calculated CCC values, combined with the loss function, guide the selection of the optimal number of PCs to retain for each variable.

As indicated by Fig. 6, retaining the first 7 PCs for both runoff and surface temperature achieves an effective balance between minimising information loss and avoiding noise amplification. This compromise ensures the emulator's robustness and reliability for climate field predictions.

We can also use the same metric to evaluate the effectiveness of a more complex kernel function. To do so, we compare two emulators designed to be as similar as possible; their only difference lies in the correlation function $k(\boldsymbol{z}, \boldsymbol{z}')$. For one emulator, this function is an *Additive Kernel* as described earlier, while for the other it is the more conventional *Square Exponential kernel*, as used in previous studies (Bounceur et al., 2015; Lord et al., 2017; Van Breedam et al., 2021). Fig. 7 presents the results of this comparison for continental runoff – the variable that initially motivated the development of this improved emulator because
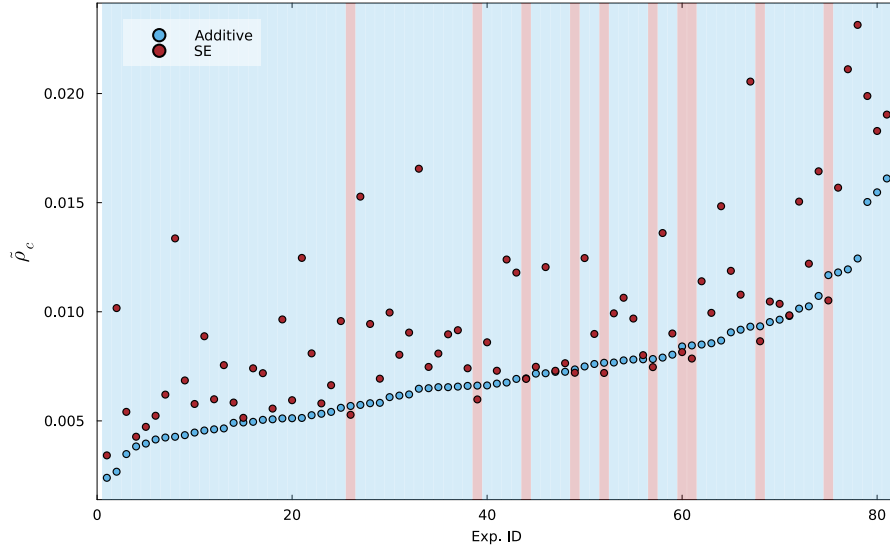
Figure 7. Comparison of the performance of the two kernel functions: Additive and Square Exponential (SE). Each point represents the shifted CCC value computed between the GCM climate field and the prediction produced by an emulator trained on the dataset with the specific experiment omitted (LOO). A blue background indicates that the Additive Kernel outperforms the alternative in that experiment, whereas a red background signifies that the SE kernel achieves better performance. The 81 experiments were sorted by $\tilde{\rho}_{c,\text{Additive}}$ for clarity.

of its initial poor predictability. We observe that employing the Additive Kernel function results in superior performance for almost all the test experiments.

3.7 Interpretation and sensitivity

Fig. 8 shows the first six PCs and the associated normalised variance $\lambda_i/\lambda_{\text{tot}}$, which indicates the proportion of variability each

420 PC explains. As these are variations relative to the mean value and can be multiplied by either a positive or negative factor, the sign is not inherently meaningful. Consequently, the terms 'cooling' and 'warming' can be interchanged without loss of generality in the subsequent discussion.

The first PC, accounting for approximately 97% of the total variance, represents a global anomaly, where the sign is consistent across locations, but the intensity varies. As inferred from Fig. 9 (a), this component primarily results from changes in

425 pCO_2. However, as depicted in Fig. 10 (a), eccentricity has a similar signature, as expected, since this parameter modulates the total amount of radiation received from the Sun. The second PC depicts a polar effect, with cooling at the poles and warming in the mid-latitudes. As shown in Fig. 9 (b), this effect is mostly attributable to changes in obliquity.

The subsequent PCs reveal more localised variations. Notably, the sixth PC is highly localised and shows most of its effect in the polar south region. This specific temperature anomaly can be attributed to the presence of sea ice. To capture this particular

430 pattern, it may be beneficial to design a focused emulation process for this region. This approach would increase the variance
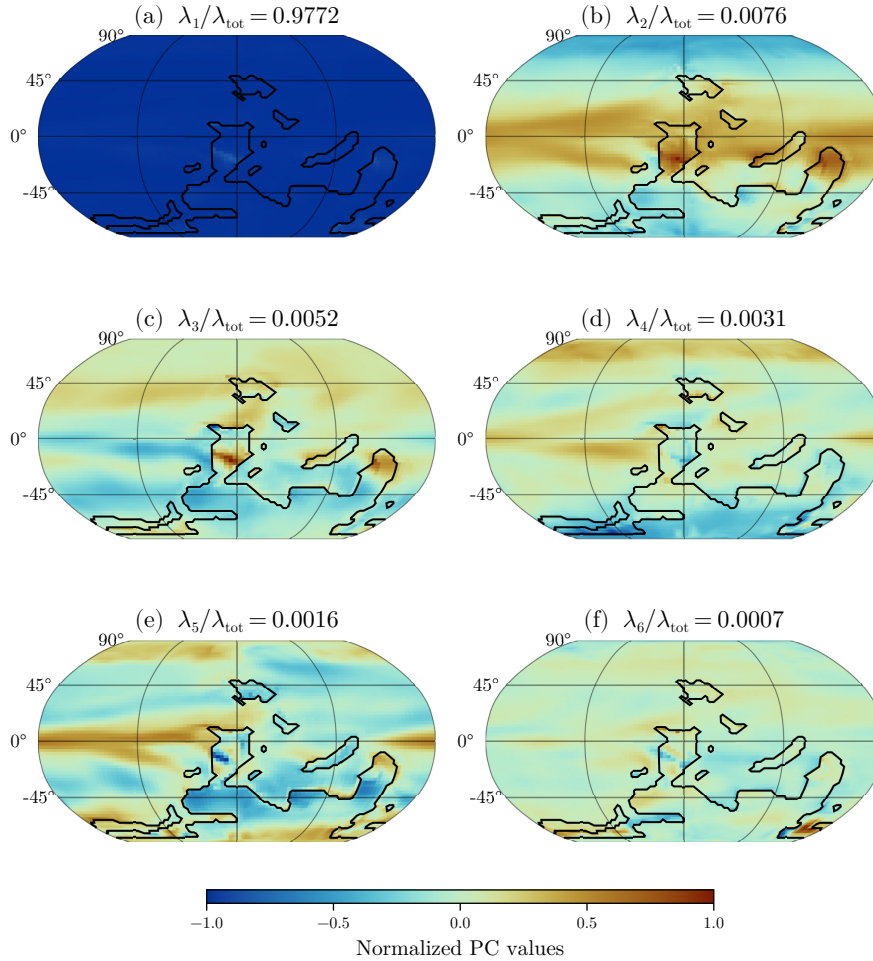
Figure 8. The figure presents the first six principal components of the surface temperature, ordered by their eigenvalues and rescaled by a multiplicative factor. The colour scale is indicative of the amplitude and its sign within a single PC, but is independent across different PCs. The black lines delineate the coastlines of the continental configuration.

of this specific component, enabling better handling by the GPR. Visualisation of the remaining PCs shows that this particular area is the only one with a significant anomaly, suggesting that the other PCs are less relevant for the emulation process. A specific design where different GPRs are trained for continental and sea-surface temperature has also been considered, but the performance improvement – it showed slightly better results amongst 55% of the training set – was not deemed significant enough to justify the additional computational cost when merged inside GEOCLIM. The corresponding principal components for runoff are shown in Fig. S2 from the supplementary material.
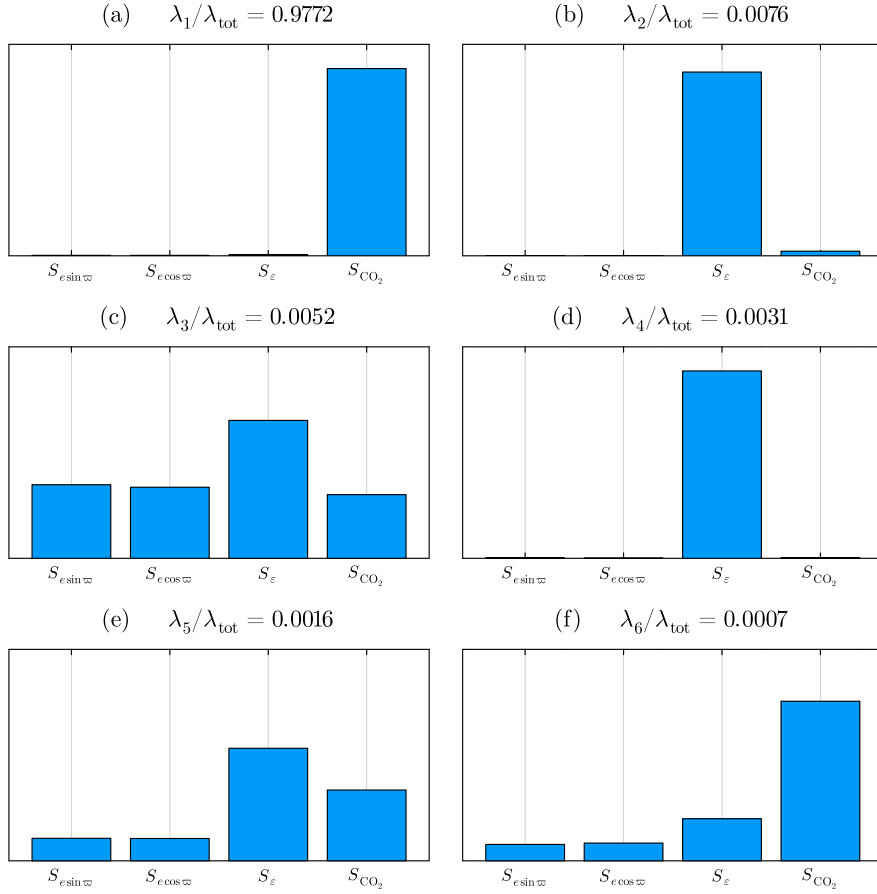
Figure 9. Magnitudes of the total Sobol index (Sobol′, 2001) for each input variable. A superior total-order Sobol index signifies the parameter's comprehensive influence on the output's variability, integrating both direct impacts and interaction effects. Due to a larger domain explored for the carbon dioxide and obliquity, the effect of climatic precession is screened (cf. Fig. 10).

Maps comparing high vs. low climate precession and high vs. low obliquity for temperature and runoff are provided in Figs. S3–S8.

As described in the previous text, most of the variance arises from changes in pCO_2. Indeed, the experimental design allows for significantly more variance in this direction than in climate precession. However, one could restrict the parameter space accordingly and compute, for instance, the global mean surface temperature (GMST) to test whether the results are consistent with the astronomical forcing or not. If we assume that a doubling of CO_2 decreases outgoing radiation by $3.7 \, \mathrm{W \, m^{-2}}$, and that the transition from zero eccentricity to e_{max} is equivalent (in terms of total mean annual radiative forcing) to (Milankovitch, 1941; Berger and Loutre, 1994):

$$\left(\left(1 - e_{\mathrm{max}}^2 \right)^{-\frac{1}{2}} - 1 \right) \frac{S}{4} \approx 0.93 \, \mathrm{W \, m^{-2}}$$
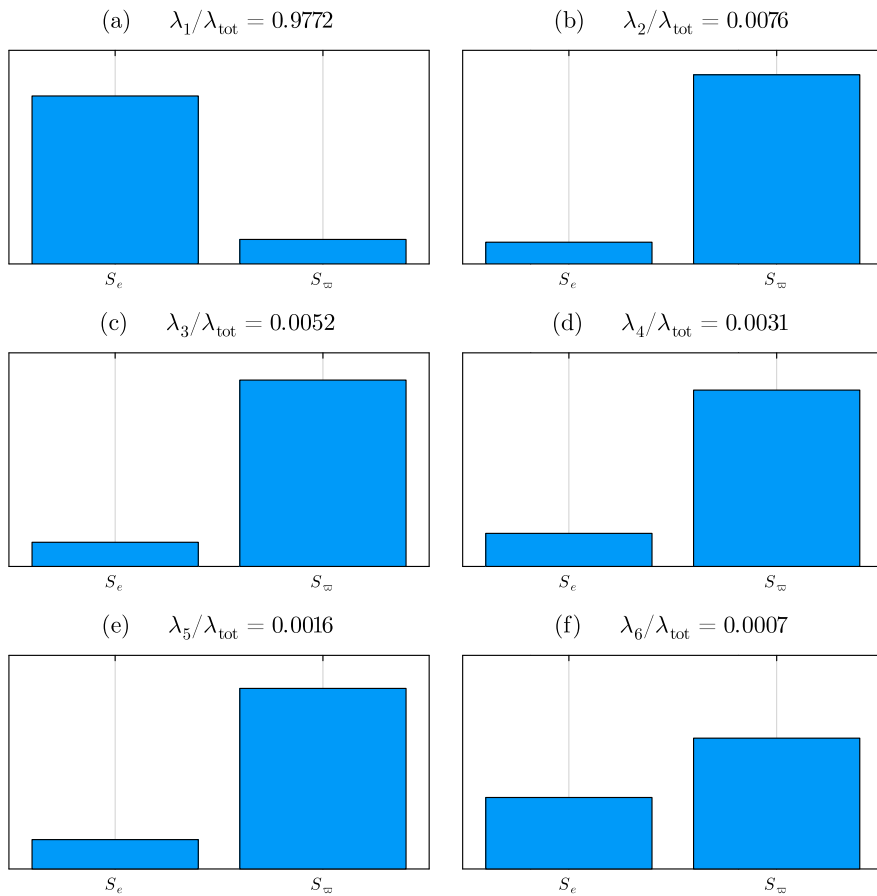
Figure 10. Magnitudes of the total Sobol index for eccentricity e and longitude of the perihelion ϖ. To obtain this figure, the values of obliquity and pCO_2 were held constant respectively at $\varepsilon = 22.23°$ and 5×280 ppm.

for $e_{\max} = 0.075$ and $S = 1322 \ \mathrm{W\,m^{-2}}$, then we can estimate that a multiplicative factor of $2^{\frac{1.06}{3.7}} \approx 1.2$ for pCO_2 would produce similar effects on GMST for both astronomical forcing and carbon dioxide.

As shown in Fig. 11, a 20% variation in pCO_2 results in a GMST change equivalent to the effect of increasing eccentricity from $e = 0$ to $e = 0.075$, corresponding to approximately $+0.7$ °C. This suggests that orbital variability is not overshadowed by variations in atmospheric carbon dioxide.

3.8 Coupling to GEOCLIM

The GP calibration is done in a high-level computing language, namely Julia, for easier tuning and understanding of the code. However, most of the historical models in climate science are written in Fortran, and this is the case for GEOCLIM.
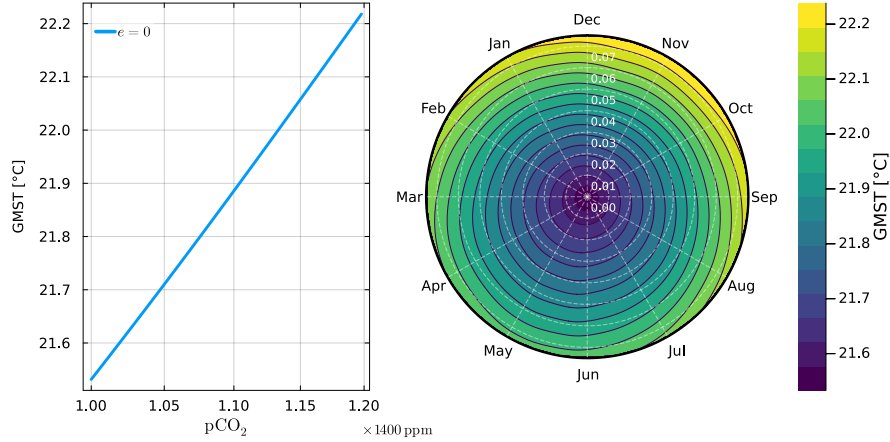
Figure 11. The left panel shows the global mean surface temperature (GMST) for a fixed astronomical configuration ($e = 0$, $\varepsilon = 22.23°$) with pCO_2 varying from 5×280 ppm to $1.2 \times 5 \times 280$ ppm. On the right is displayed a phase contour plot of the GMST, with the eccentricity as the radius and the longitude of perihelion as the azimuthal angle. The angle is expressed in the month when the Earth is closest to the Sun, such that the vertical axis is the climatic precession $e \sin \varpi$ and the horizontal axis is the coprecession $e \cos \varpi$. The polar representation is shown for a fixed obliquity ($\varepsilon = 22.23°$) and a constant pCO_2 of 5×280 ppm.

455 Once trained, making a prediction with the emulator boils down to multiplying matrices, operations that are well optimised in Fortran using the OpenBLAS library. A script to automatically generate the Fortran routine, given a trained emulator, has been written for this coupling purpose.

In our setup, we want the pCO_2 to be interactive and, therefore, cannot precompute the whole climate time series, as we need input from the biogeochemical model to predict the next climate state. The coupling time is an arbitrary parameter that depends

460 on the time resolution desired and the magnitude of climatic perturbations. Depending on the situation, this parameter can be easily adjusted to a different value. Given the processes featured by GEOCLIM and the timescale of astronomical forcing, it is conservative to update the climate fields every 25 years and synchronise it with the calls of the continental weathering module. We do not expect any process to operate on a faster timescale.

We emphasise that our emulator predicts annual-mean temperature and runoff fields, which GEOCLIM uses as inputs.

465 Indeed, GEOCLIM is not designed to simulate seasonal variability. Yet, the seasonal cycle is captured in the HadSM3 simulations used to train the emulator. In particular, orbital forcing effects on the seasonal cycle effectively influence the annual-mean fields that the emulator produces – for example, this is how monsoons respond to precession. In principle, the emulator could be adapted to account for seasonal effects by emulating monthly mean climate fields, but this would require a significant decrease in the model timestep and a corresponding large increase in computational cost.

470 The complete emulator development workflow described in this section is summarized in Fig. 12, which illustrates the seven key methodological steps from parameter space design through final integration into GEOCLIM.
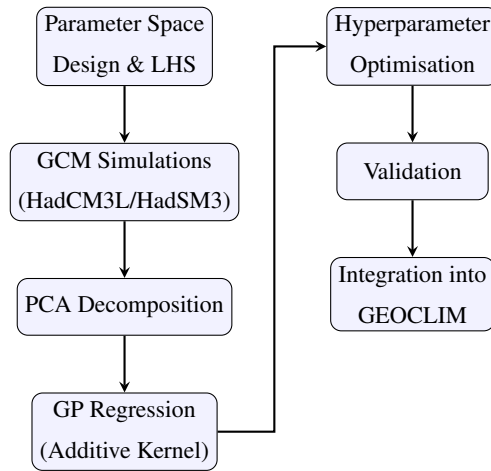
Figure 12. Methodological flowchart showing the seven key steps in the emulator development process: (1) Parameter space design using Latin Hypercube Sampling, (2) GCM simulations with the two-tier HadCM3L/HadSM3 approach, (3) PCA decomposition of climate fields, (4) GP regression using additive kernel for each PC, (5) Hyperparameter optimisation, (6) Validation, and (7) Integration into GEOCLIM.

4 Case study: orbital forcing of the Late Devonian climate

4.1 Devonian Ocean Anoxic Events and astronomical forcing

The dynamics of Ocean Anoxic Events (OAEs) and their responsible mechanisms have been investigated over the last decades,
but failed to yield consensus (see for instance Haddad et al., 2018). It has been hypothesized that the cyclical recurrence of anoxia in geological records can be influenced by astronomical cycles within a broader context of background environmental conditions (cf., De Vleeschouwer et al., 2017; Whalen et al., 2017; Da Silva et al., 2020; Lu et al., 2021; Ma et al., 2022, 2025; Wichern et al., 2024). Specific configurations of the astronomical forcing would help triggering regional or even possibly global anoxia in a context of already low-oxygen content. Furthermore, orbital forcing would pace anoxia bursts. One suggested
mechanism involves the influence of astronomical forcing on the position of tropical precipitation. Such changes would affect the dynamics of nutrient storage on the continent, as well as the rate and time at which they are released in the oceans. Finally, the modulation of nutrient release would impact ocean biogeochemistry (De Vleeschouwer et al., 2017, 2024; Wichern et al., 2024).

In the specific case of the Devonian, De Vleeschouwer et al. (2017) and Wichern et al. (2024) suggest that conditions for
triggering and sustaining anoxia may best be met after passage through a so-called node of the very-long eccentricity cycle: a configuration met approximately every 2.4 million years (Laskar, 2020). However, as the precise timing of astronomical forcing during the Paleozoic is uncertain due to the chaotic nature of planetary dynamics, the investigation strategy needs to consider different scenarios. Furthermore, the anoxia itself may have a duration ranging from several tens of thousands to approximately one hundred thousand years, which poses additional questions about sustaining mechanisms. A possible explanation would be a

490 positive feedback loop involving nutrient recycling (Reershemius and Planavsky, 2021; De Vleeschouwer et al., 2024; Wichern et al., 2024).

4.2 Calibration of GEOCLIM

GEOCLIM version 6.1.0 was selected for long-term carbon cycle modelling. While the primary modification is the implementation of the emulator for dynamic climate computation, several other adjustments were necessary for our model setup.

495 The temperature of each oceanic box is determined through a linear relationship with the global sea-surface temperature (GSST):

$$T_i = \gamma_{0,i} + \gamma_{1,i} \times \text{GSST} \tag{30}$$

where the GSST is derived from the emulator output. Since our HadSM3 setup employs a slab ocean model without vertical resolution, we require a method to specify oceanic box temperatures for GEOCLIM. The coefficient vectors γ were fitted using 500 least-squares regression on cGENIE simulations with a similar continental configuration (Gérard et al., 2025), following the same experimental design as our 81 HadSM3 runs. This parameterisation ensures that oceanic box temperatures incorporate both orbital forcing and $p\text{CO}_2$ fluctuations through the emulated GSST, with updates synchronised to the 25-year climate field intervals.

Water mass exchanges between neighbouring boxes were also updated accordingly, though kept constant in time and across 505 experiments as this version of GEOCLIM is not designed to include a dynamic ocean circulation.

Following the methodology outlined in the supplementary material of Maffre et al. (2021), we calibrated the uncertain parameters using a present-day configuration. Box geometries (volumes and surfaces) were computed using Scotese's reconstruction. The surfaces and volumes of epicontinental boxes, which carry significant uncertainty, were adjusted to achieve an equilibrium concentration of 10 mol m^{-3} dissolved SO_4 in the ocean (Cai et al., 2022), using mean-orbital parameters ($e = 0$ 510 and $\varepsilon = 22.23°$).

Given the substantial uncertainties in reconstructing solid Earth degassing rates during the Devonian period (Marcilly et al., 2021; Müller et al., 2024), we adopted a fixed value in this study. This value was determined by calculating the degassing rate necessary to balance the total silicate weathering consumption in a present-day configuration, corresponding to the rate required to maintain a steady atmospheric $p\text{CO}_2$ level of 280 ppm.

4.3 Fourier analysis of the response

The development of this emulator and modelling strategy was motivated by investigating the potential connection between astronomical forcing and the timing of Devonian OAEs, as discussed in the introduction. Previous studies by De Vleeschouwer et al. (2024), and Wichern et al. (2024) hypothesized that reduced seasonal contrasts during eccentricity minima within 2.4 Myr nodes facilitate gradual regolith accumulation on continents. This accumulated regolith is subsequently mobilized and eroded 520 during the first pronounced precession maximum following the node. Our modelling framework enables the investigation of potential long-term effects within geological soil systems.

In this part, we present an integrated result of the emulator into GEOCLIM6.1.0. The astronomical scenario, which outlines the values of the three orbital parameters over time, is a plausible but hypothetical scenario, given that no precise solution for the Devonian can be produced. We use the planetary solution La10a provided by Laskar et al. (2011), and selected the period 125 − 100 Ma, during which 2.4 Myr eccentricity nodes are well-marked. We coupled this solution with the precession model of Sharaf and Boudnikova (1967), following the procedure outlined in Berger and Loutre (1991). This procedure implies, first, a harmonic decomposition of the planetary solution (using R-code based on Šidlichovský and Nesvorný (1997)) and then specification of reference obliquity and general precession rate. The latter are given by Farhat et al. (2022) for the period of 370 Ma. With this construction, the solution has obliquity frequencies of 30555 and 37335 years, and climatic precession frequencies of 16470 and 19386 years considered to be plausible for the Devonian. The R-code used to generate those time series is available in Crucifix (2025).

Starting from its calibration value, the model was run for 25 Myr. The last 22.5 Myr were selected for detailed analysis because the static equilibrium state, which has a slightly higher $p\text{CO}_2$, differs from the dynamic equilibrium state. The initial 2.5 Myr were used to stabilise the model and avoid an unwanted artefact trend.

Fig. 13 shows the normalised Fourier power spectral density of the forcing (i.e., the three orbital parameters) and of a spatial average and latitudinal band of the climate fields, specifically continental temperature and runoff. These fields are internally computed by the emulator based on the orbital parameters and carbon dioxide concentration. Globally, the surface temperature shows a strong influence from the obliquity and eccentricity cycles. Conversely, runoff is primarily affected by climatic precession ($e \sin \varpi$). Locally, the long eccentricity cycle (± 405 kyr) seems to be dominant, but its influence is reduced when considering the global average. The 2.4 Myr cycle is present in the forcing, but its influence on the climatic fields is very weak.

From Fig. 14 (a) & (c), we can observe that the atmospheric $p\text{CO}_2$ and average regolith thickness exhibits strong 405 kyr and 2.4 Myr cycles, suggesting the presence of non-linearities in the subsystem. The atmospheric $p\text{CO}_2$ variations shown in the spectral analysis exhibit amplitude variations of up to 50 ppm, which are comparable to the temporal variations displayed in Fig. 17. As depicted in Fig. 15, the amplification of the 2.4 Myr cycle is a spatially localised feature, with a maximal amplitude around the tropics. Time series of GEOCLIM outputs are presented in the following section to further investigate this local phenomenon. However, these cycles are almost absent from the total reactive phosphorus flux, as shown in Fig. 14 (b). Under the parametrisation of GEOCLIM, this result is inconsistent with the hypothesis that the regolith thickness is a major driver of the nutrient weathering flux.

An illustrative time series of oceanic dissolved oxygen for one scenario is shown in Fig. S9.

4.4 Regolith dynamics

Here, we examine the variations in regolith thickness within the model. Regolith thickness is a crucial parameter in GEOCLIM, controlling weathering rates and governing the efficiency of continental weathering processes. As such, it is central to many hypotheses regarding Devonian OAEs.
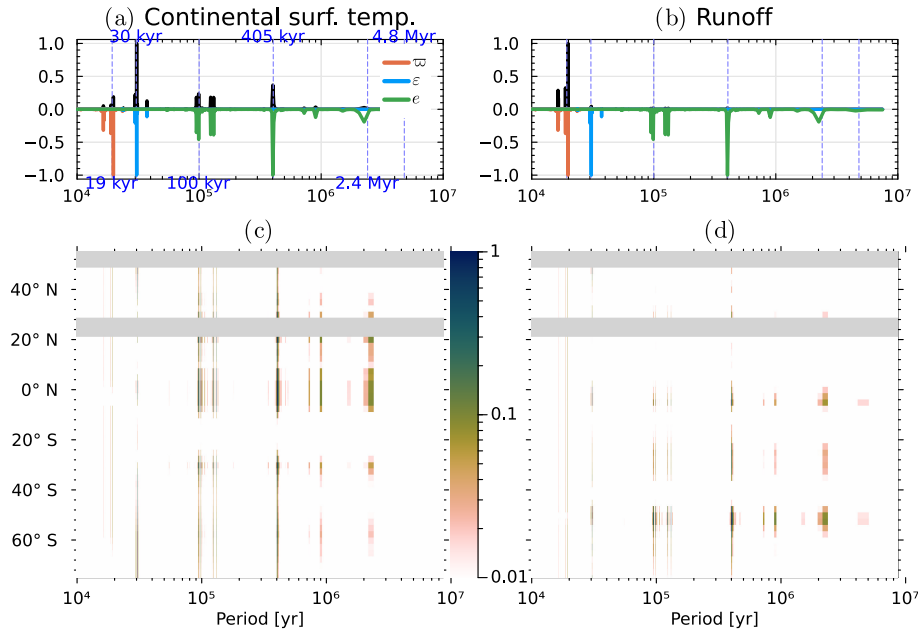
Figure 13. Normalised Fourier power spectral density of (a) & (c) continental surface temperature and (b) & (d) runoff. The top panels show the spectrum of the spatial average in black and the spectra of the orbital parameters flipped over the x-axis (for better visibility) and the bottom panels display the latitudinal averages. Various important cycles are shown in vertical dashed blue lines as landmarks.

The simulation employs the same astronomical scenario as previously described, but with obliquity held constant at $22.23°$ to isolate the effects of eccentricity and climatic precession. Fig. 17 presents a 2 Myr segment of this scenario, displaying three model outcomes: atmospheric carbon dioxide concentration, average runoff, and surface temperature in two distinct regions illustrated in Fig. 16.

Fig. 17 shows a clear precession signal on the runoff and surface temperatures. Nonetheless, the amplitude and duration of these perturbations alone are insufficient to independently trigger a global OAE. As depicted in Fig. 18, the regolith in the green region predominantly maintains equilibrium and does not exhibit significant thickening paced by eccentricity. Conversely, the blue region demonstrates growth paced by this parameter. The green region can be characterized as an ever-wet tropical region, whereas the blue region experiences aridity paced by precession minimum.

In the blue region, when eccentricity is high, the moisture available during a precession maximum amplifies regolith production more than it augments erosion, resulting in a thickening of the regolith. In the ensuing precession minimum, the absence of water for production or erosion renders the regolith static until the next precession maximum for further growth. This mechanism operates as long as the eccentricity increases. However, under conditions of decreasing eccentricity, the subsequent precession maximum lacks sufficient wetness to sustain the thickness of the regolith, leading to production becoming lower than erosion. A remarkable phenomenon can be observed during an eccentricity node. During this period, the eccentricity is low for a complete 405 kyr cycle, and the regolith thickness at the beginning, resulting from previous high eccentricity cycles,
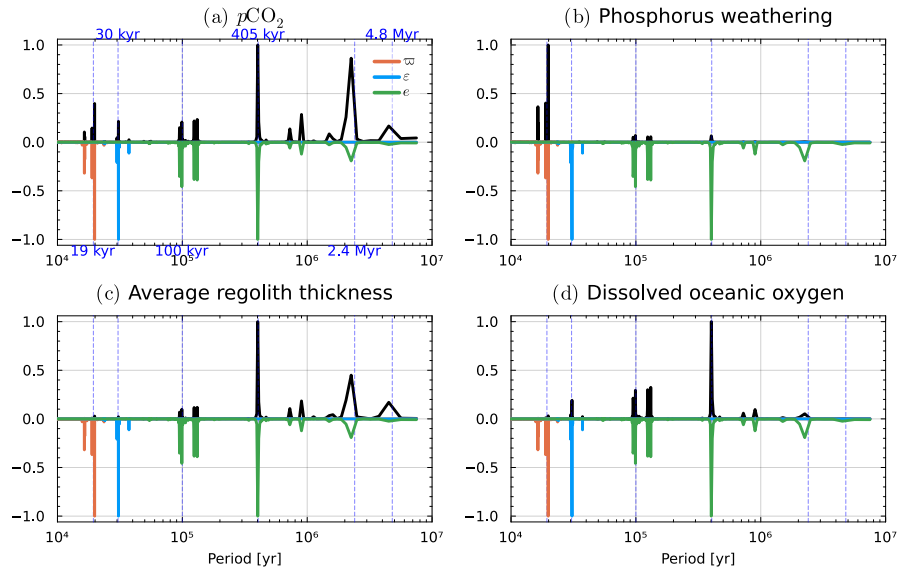
Figure 14. Normalised Fourier power spectral density of (a) atmospheric pCO_2, (b) total reactive phosphorus weathering flux, (c) global average regolith thickness and (d) global oceanic dissolved oxygen. The spectra of the orbital parameters are shown flipped over the x-axis. Various important cycles are shown in vertical dashed blue lines as landmarks.
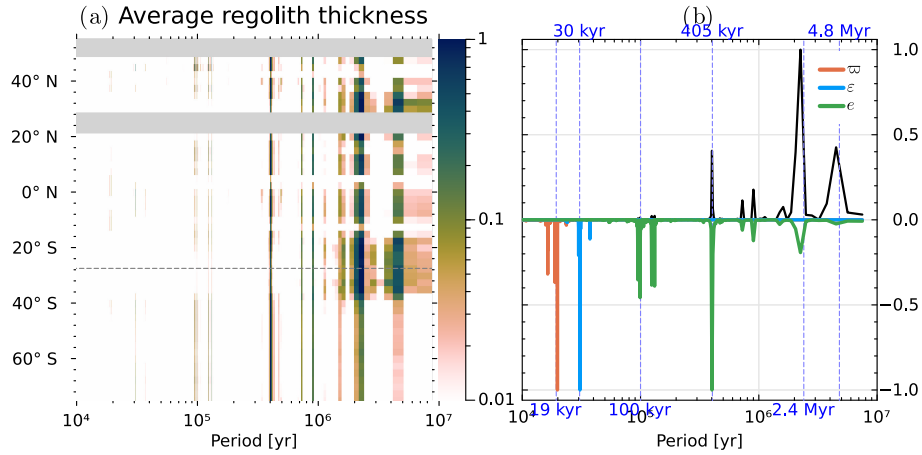
Figure 15. Normalised Fourier power spectral density of (a) longitudinal average of the regolith thickness and (b) a section at $27.5°$ S, represented by the dashed line in (a). The spectra of the orbital parameters are shown flipped over the x-axis. Various important cycles are shown in vertical dashed blue lines as landmarks.

is more important than its equilibrium value. This results in a significantly higher erosion than production of the regolith during the node, creating an important 2.4 Myr cycle in the signal thickness.
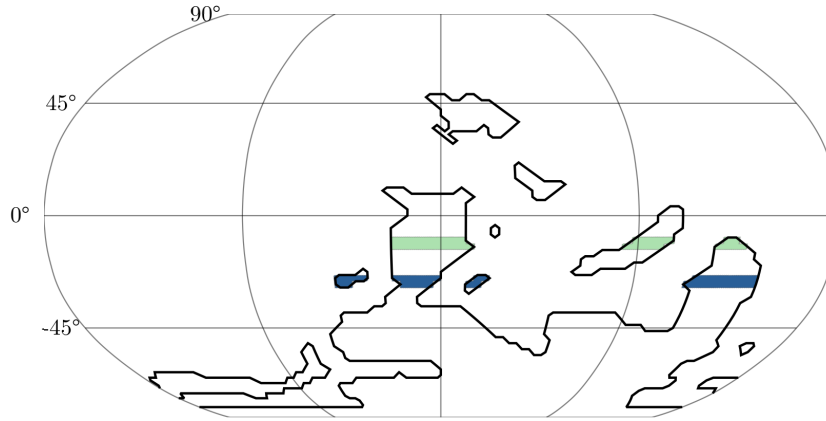
Figure 16. The two regions under consideration for the spatial averages in Fig. 17. The green region is consistently wet, while the blue region, located at the edge of the intertropical convergence zone (ITCZ), can experience arid conditions under precession minima.

In the green region, both production and erosion are influenced to a comparable degree by precession, resulting in the regolith nearing a steady-state.

From these findings, our model suggests an alternative narrative to the eccentricity node hypothesis. Globally, the regolith (controlling weathering-driven nutrient delivery) demonstrates a positive correlation with eccentricity, indicating that soil development occurs predominantly during high eccentricity periods rather than during nodes. Furthermore, the impact of regolith thickness variations on oceanic nutrient fluxes appears minimal.

As illustrated in Fig. 18, these variations are relatively modest, with runoff – primarily controlled by climatic precession $e\sin\varpi$ – emerging as the dominant driver of weathering and erosion processes. Moreover, the majority of nutrients originate from regions analogous to the green zone, where significant long-term response effects appear absent.

4.5 Discussion of the case study

With this case study, we now have a framework for studying the astronomical forcing on regolith dynamics during the Devonian period. With the hypotheses adopted here (fixed palaeography, solid Earth degassing, vegetation and constant lithology), results indicate that regolith build-up is more closely linked to periods of high eccentricity rather than to eccentricity nodes.

More specifically, we found that climatic precession ($e\sin\varpi$) primarily controls runoff and weathering, while the direct impact of regolith dynamics on nutrient fluxes appears minimal. This confirms that, in the Devonian context, additional environmental factors are necessary to trigger widespread OAEs. For example, obliquity may play a role in sea-level variation through effects from ice sheets (Ma et al., 2022, 2025).

In the context of our study, it appears impossible to create an OAE from orbital forcing alone. The link between nutrient fluxes and anoxia operates through increased phosphorus delivery to the ocean, which stimulates primary productivity and subsequent oxygen consumption during organic matter decomposition, potentially leading to oxygen depletion. However, the
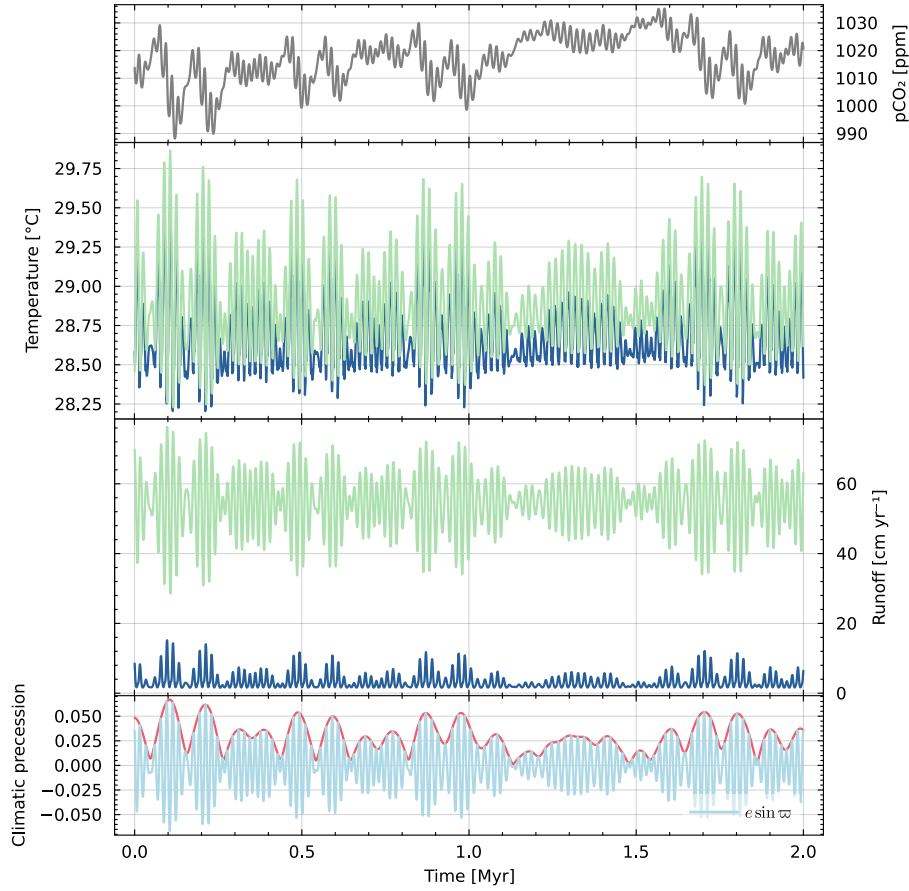
Figure 17. This figure displays a 2 Myr long segment of a plausible Devonian orbital solution with constant obliquity. The top line represents the atmospheric $p\mathrm{CO_2}$, the blue and green curves are spatial averages on the respective regions of the same colour as depicted in Fig. 16. The eccentricity is shown in red on the bottom graph. Except for climatic precession, which is the forcing here, all other curves are outcomes of the model. This section has been chosen to specifically show a 2.4 node, approximately located between 1.1 and 1.5 Myr, where the 100 kyr eccentricity signal is almost absent.

small amplitude of flux variations induced by orbital forcing in our simulations is insufficient to trigger changes in oceanic dissolved oxygen concentrations which would be consistent with widespread anoxic events. While no strong conclusions regarding the precise timing and pacing of anoxic events can be drawn from these results due to the model design limitations, we have established a solid basis for exploring the dynamics of Devonian anoxia. Future work directions should include vegetation feedbacks and an improved ocean discretisation (Algeo and Scheckler, 2010; Maffre et al., 2022; Smart et al., 2023; Maffre et al., 2024).
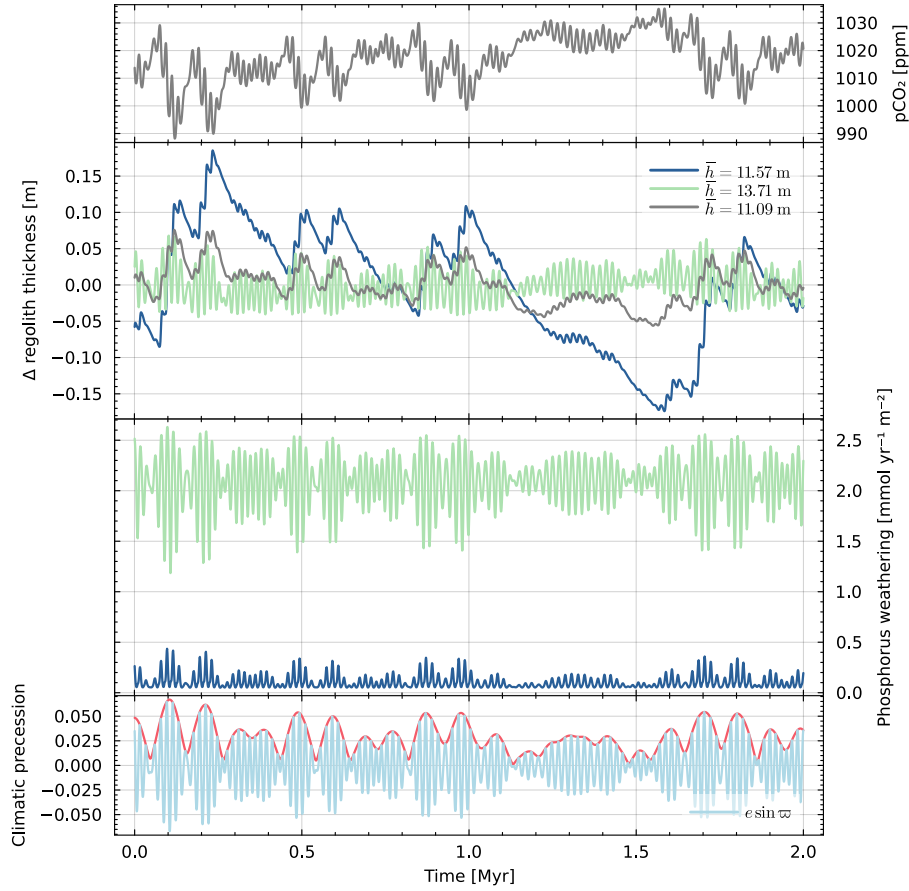
Figure 18. Displayed alongside the pCO_2 and orbital forcing are the average regolith thickness variation with respect to the temporal mean $\bar{h}$ and phosphorus weathering, following the colour convention for each region defined in Fig. 16. The third grey line of the second panel depicts the global behaviour of the regolith thickness variation.

5 Conclusions

600   In this study, we presented a dynamically coupled Gaussian Process emulator designed to bridge the gap between complex General Circulation Models and geochemical simulations over geological timescales. Our approach leverages an optimised training dataset, advanced kernel functions, and robust hyperparameter tuning to capture essential climate dynamics while significantly reducing computational costs. The additive kernel framework, inspired by Duvenaud et al. (2011), allowed us to account for both individual and interacting effects of key input parameters, and the integration of PCA ensured that the 605   emulator maintained crucial spatial teleconnections. Furthermore, the development of the emulator has taken into account the importance of a careful treatment of the Anomalous Heat Convergence flux to avoid possible global unwanted parameter-

dependent forcing, and has underscored that Principal Component Analysis is effective for dimensionality reduction while a specific pre-processing of the input variables can slightly impact the emulator's performance.

The proposed method not only reproduces key features observed in the HadSM3 simulations but also provides a flexible tool for exploring the parameter space robustly and efficiently. As a case study, the setup has been applied to a 25 Myr Devonian orbital scenario to test the relationship between orbital forcing, regolith dynamics and nutrient fluxes. Our preliminary findings indicate that high eccentricity, rather than low eccentricity, promotes regolith growth, though the variations are not significant enough to have a substantial impact on the nutrient fluxes. Therefore, it is very challenging to produce an OAE by the effect of astronomical forcing only on the atmospheric components in this setup.

By enabling the investigation of multiple astronomical scenarios at minimal computational cost, our emulator provides valuable insights into the complex relationship between orbital forcing and palaeoenvironmental changes.

Our application on the Devonian shows that we have a modelling framework stimulating research directions for understanding the mechanisms of anoxia during that period. Yet, the methodology is also applicable to other geological intervals throughout the Phanerozoic. More generally, it contributes to addressing the crucial challenge of studying Earth's climate system across a wide range of timescales in biogeochemical models, from yearly variations to million-year-long geochemical processes. We believe this work lays a solid foundation for enhancing our understanding of palaeoenvironmental variations and their driving mechanisms across the Phanerozoic.

Code and data availability. The code and data for training the emulator are archived on Zenodo: https://doi.org/10.5281/zenodo.15113858 (Sablon, 2025a). The vanilla version of GEOCLIM6.1.0 can be found on the public GEOCLIM GitHub repository and the modified version, including the emulator and outputs of the climate simulations presented in the article, can be found on a separated Zenodo archive: https://doi.org/10.5281/zenodo.15114774 (Sablon, 2025b). The code and data used to generate the orbital solutions are also available on Zenodo: https://doi.org/10.5281/zenodo.14894811 (Crucifix, 2025).

Author contributions. PJV provided the starting dumps for the HadCM3 experiments. YG and PM provided the code of GEOCLIM6.1.0. MC and LS conceptualized the emulator, LS wrote the code, trained the model and conducted the HadCM3, HadSM3 and GEOCLIM simulations with astronomical scenarios generated by MC. LS drafted the manuscript and all authors contributed to reviewing and editing.

Competing interests. The authors declare no competing interests.

Acknowledgements. We are grateful to Prof. David De Vleeschouwer (University of Münster) for his involvement in the project as scientific advisor, and specifically for his suggestion to develop plausible Devonian astronomical solutions at a meeting held in Louvain-la-Neuve in

May 2023. Computational resources have been provided by the supercomputing facilities of the UCLouvain (CISM) and the Consortium des Équipements de Calcul Intensif en Fédération Wallonie Bruxelles (CÉCI).

635

References

Algeo, T. J. and Scheckler, S. E.: Land Plant Evolution and Weathering Rate Changes in the Devonian, Journal of Earth Science, 21, 75–78, https://doi.org/10.1007/s12583-010-0173-2, 2010.

Arndt, S., Regnier, P., Goddéris, Y., and Donnadieu, Y.: GEOCLIM *reloaded* (v 1.0): a new coupled earth system model for past climate change, Geoscientific Model Development, 4, 451–481, https://doi.org/10.5194/gmd-4-451-2011, publisher: Copernicus GmbH, 2011.

Bahcall, J. N., Pinsonneault, M. H., and Basu, S.: Solar Models: Current Epoch and Time Dependences, Neutrinos, and Helioseismological Properties, The Astrophysical Journal, 555, 990, https://doi.org/10.1086/321493, publisher: IOP Publishing, 2001.

Bates, S., Sienz, J., and Toropov, V.: Formulation of the Optimal Latin Hypercube Design of Experiments Using a Permutation Genetic Algorithm, https://doi.org/10.2514/6.2004-2011.

Becker, R. T., Marshall, J. E. A., Da Silva, A. C., Agterberg, F. P., Gradstein, F. M., and Ogg, J. G.: Chapter 22 - The Devonian Period, in: Geologic Time Scale 2020, edited by Gradstein, F. M., Ogg, J. G., Schmitz, M. D., and Ogg, G. M., pp. 733–810, Elsevier, ISBN 978-0-12-824360-2, https://doi.org/10.1016/B978-0-12-824360-2.00022-X, 2020.

Beckers, T.: An Introduction to Gaussian Process Models, https://doi.org/10.48550/arXiv.2102.05497, arXiv:2102.05497 [cs, eess], 2021.

Berger, A. and Loutre, M. F.: Insolation values for the climate of the last 10 million years, Quaternary Science Reviews, 10, 297 – 317, https://doi.org/10.1016/0277-3791(91)90033-Q, 1991.

Berger, A. and Loutre, M. F.: Precession, Eccentricity, Obliquity, Insolation and Paleoclimates, in: Long-Term Climatic Variations, edited by Duplessy, J.-C. and Spyridakis, M.-T., pp. 107–151, Springer, Berlin, Heidelberg, ISBN 978-3-642-79066-9, https://doi.org/10.1007/978-3-642-79066-9_5, 1994.

Berger, A. and Yin, Q.: Orbital Forcing (Astronomical Theory of Paleoclimates), in: Encyclopedia of Geology (Second Edition), edited by Alderton, D. and Elias, S. A., pp. 435–443, Academic Press, Oxford, second edition edn., ISBN 978-0-08-102909-1, https://doi.org/https://doi.org/10.1016/B978-0-12-409548-9.11990-6, 2021.

Bezanson, J., Karpinski, S., Shah, V. B., and Edelman, A.: Julia: A fast dynamic language for technical computing, arXiv preprint arXiv:1209.5145, 2012.

Bounceur, N., Crucifix, M., and Wilkinson, R. D.: Global sensitivity analysis of the climate–vegetation system to astronomical forcing: an emulator-based approach, Earth System Dynamics, 6, 205–224, https://doi.org/10.5194/esd-6-205-2015, publisher: Copernicus GmbH, 2015.

Brugger, J., Hofmann, M., Petri, S., and Feulner, G.: On the Sensitivity of the Devonian Climate to Continental Configuration, Vegetation Cover, Orbital Configuration, CO 2 Concentration, and Insolation, Paleoceanography and Paleoclimatology, 34, 1375–1398, https://doi.org/10.1029/2019PA003562, 2019.

Cai, C., Xu, C., Fakhraee, M., Chen, D., and Peng, Y.: Significant fluctuation in the global sulfate reservoir and oceanic redox state during the Late Devonian event, PNAS Nexus, 1, pgac122, https://doi.org/10.1093/pnasnexus/pgac122, 2022.

Carmichael, S. K., Waters, J. A., Königshof, P., Suttner, T. J., and Kido, E.: Paleogeography and Paleoenvironments of the Late Devonian Kellwasser Event: A Review of Its Sedimentological and Geochemical Expression, Global and Planetary Change, 183, 102 984, https://doi.org/10.1016/j.gloplacha.2019.102984, 2019.

Cox, P., Betts, R., Bunton, C., Essery, R., Rowntree, P., and Smith, J.: The impact of new land surface physics on the GCM simulation of climate and climate sensitivity, Climate Dynamics, 15, 183–203, 1999.

Cox, P. M.: Description of the" TRIFFID" dynamic global vegetation model, 2001.

Cox, P. M., Betts, R. A., Jones, C. D., Spall, S. A., and Totterdell, I. J.: Acceleration of global warming due to carbon-cycle feedbacks in a coupled climate model, Nature, 408, 184–187, 2000.

675 Crucifix, M.: Generation of sample climatic precession and obliquity solutions for running climate experiments over the Devonian, https://doi.org/10.5281/zenodo.14894811, 2025.

Da Silva, A.-C., Sinnesael, M., Claeys, P., Davies, J. H. F. L., de Winter, N. J., Percival, L. M. E., Schaltegger, U., and De Vleeschouwer, D.: Anchoring the Late Devonian mass extinction in absolute time by integrating climatic controls and radio-isotopic dating, Scientific Reports, 10, 12 940, https://doi.org/10.1038/s41598-020-69097-6, publisher: Nature Publishing Group, 2020.

680 De Vleeschouwer, D., Da Silva, A.-C., Sinnesael, M., Chen, D., Day, J. E., Whalen, M. T., Guo, Z., and Claeys, P.: Timing and pacing of the Late Devonian mass extinction event regulated by eccentricity and obliquity, Nature Communications, 8, 2268, https://doi.org/10.1038/s41467-017-02407-1, publisher: Nature Publishing Group, 2017.

De Vleeschouwer, D., Percival, L. M. E., Wichern, N. M. A., and Batenburg, S. J.: Pre-Cenozoic Cyclostratigraphy and Palaeoclimate Responses to Astronomical Forcing, Nature Reviews Earth & Environment, 5, 59–74, https://doi.org/10.1038/s43017-023-00505-x, 2024.

685 Dixit, V. K. and Rackauckas, C.: Optimization.jl: A Unified Optimization Package, https://doi.org/10.5281/zenodo.7738525, 2023.

Donnadieu, Y., Goddéris, Y., Pierrehumbert, R., Dromart, G., Fluteau, F., and Jacob, R.: A GEOCLIM simulation of climatic and biogeo-chemical consequences of Pangea breakup, Geochemistry, Geophysics, Geosystems, 7, https://doi.org/10.1029/2006GC001278, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1029/2006GC001278, 2006.

Duvenaud, D., Nickisch, H., and Rasmussen, C. E.: Additive Gaussian Processes, 24, 226–234, mAG ID: 2130283669, 2011.

690 Duvenaud, D., Lloyd, J., Grosse, R., Tenenbaum, J., and Zoubin, G.: Structure Discovery in Nonparametric Regression through Com-positional Kernel Search, in: Proceedings of the 30th International Conference on Machine Learning, pp. 1166–1174, PMLR, iSSN: 1938-7228, 2013.

Eglajs, V. and Audze, P.: New approach to the design of multifactor experiments, Problems of Dynamics and Strengths, 35, 104–107, 1977.

Eliáš, J. and Vořechovský, M.: Modification of the Audze–Eglājs criterion to achieve a uniform distribution of sampling points, Advances in 695 Engineering Software, 100, 82–96, https://doi.org/10.1016/j.advengsoft.2016.07.004, 2016.

Essery, R., Best, M., and Cox, P.: MOSES 2.2 technical documentation, vol. 30, Hadley Centre for Climate Prediction and Research, 2001.

Farhat, M., Auclair-Desrotour, P., Boué, G., and Laskar, J.: The resonant tidal evolution of the Earth-Moon distance, Astronomy & Astro-physics, 665, L1, https://doi.org/10.1051/0004-6361/202243445, publisher: EDP Sciences, 2022.

Gabet, E. J. and Mudd, S. M.: A theoretical model coupling chemical weathering rates with denudation rates, Geology, 37, 151–154, 700 https://doi.org/10.1130/G25270A.1, 2009.

Gérard, J., Sablon, L., Huygh, J. J. C., Da Silva, A.-C., Pohl, A., Vérard, C., and Crucifix, M.: Exploring the Mechanisms of Devonian Oceanic Anoxia: Impact of Ocean Dynamics, Palaeogeography, and Orbital Forcing, Climate of the Past, 21, 239–260, https://doi.org/10.5194/cp-21-239-2025, 2025.

Goddéris, Y. and Donnadieu, Y.: A sink- or a source-driven carbon cycle at the geological timescale? Relative importance of 705 palaeogeography versus solid Earth degassing rate in the Phanerozoic climatic evolution, Geological Magazine, 156, 355–365, https://doi.org/10.1017/S0016756817001054, 2019.

Haddad, E. E., Boyer, D. L., Droser, M. L., Lee, B. K., Lyons, T. W., and Love, G. D.: Ichnofabrics and Chemostratigraphy Argue against Persistent Anoxia during the Upper Kellwasser Event in New York State, Palaeogeography, Palaeoclimatology, Palaeoecology, 490, 178–190, https://doi.org/10.1016/j.palaeo.2017.10.025, 2018.

Hartmann, J. and Moosdorf, N.: The New Global Lithological Map Database GLiM: A Representation of Rock Properties at the Earth Surface, Geochemistry, Geophysics, Geosystems, 13, https://doi.org/10.1029/2012GC004370, 2012.

Higdon, D., Gattiker, J., Williams, B., and Rightley, M.: Computer Model Calibration Using High-Dimensional Output, Journal of the American Statistical Association, 103, 570–583, https://doi.org/10.1198/016214507000000888, 2008.

Holden, P. B. and Edwards, N. R.: Dimensionally Reduced Emulation of an AOGCM for Application to Integrated Assessment Modelling, Geophysical Research Letters, 37, https://doi.org/10.1029/2010GL045137, 2010.

Iman, R. L., Helton, J. C., and Campbell, J. E.: An approach to sensitivity analysis of computer models: Part I—Introduction, input variable selection and preliminary variable assessment, Journal of quality technology, 13, 174–183, 1981.

Innes, M.: Don't Unroll Adjoint: Differentiating SSA-Form Programs, 2019.

Kennedy, M. and O'Hagan, A.: Predicting the output from a complex computer code when fast approximations are available, Biometrika, 87, 1–13, https://doi.org/10.1093/biomet/87.1.1, 2000.

Laskar, J.: Chapter 4 - Astrochronology, in: Geologic Time Scale 2020, edited by Gradstein, F. M., Ogg, J. G., Schmitz, M. D., and Ogg, G. M., pp. 139–158, Elsevier, ISBN 978-0-12-824360-2, https://doi.org/10.1016/B978-0-12-824360-2.00004-8, 2020.

Laskar, J., Robutel, P., Joutel, F., Gastineau, M., Correia, A. C. M., and Levrard, B.: A long-term numerical solution for the insolation quantities of the Earth, Astronomy and Astrophysics, 428, 261–285, https://doi.org/10.1051/0004-6361:20041335, mAG ID: 2140477756, 2004.

Laskar, J., Fienga, A., Gastineau, M., and Manche, H.: La2010: a new orbital solution for the long-term motion of the Earth, Astronomy and Astrophysics, 532, https://doi.org/10.1051/0004-6361/201116836, mAG ID: 2096334509, 2011.

Lin, L. I.-K.: A Concordance Correlation Coefficient to Evaluate Reproducibility, Biometrics, 45, 255–268, https://doi.org/10.2307/2532051, 1989.

Lord, N. S., Crucifix, M., Lunt, D. J., Thorne, M. C., Bounceur, N., Dowsett, H., O'Brien, C. L., and Ridgwell, A.: Emulation of long-term changes in global climate: application to the late Pliocene and future, Climate of the Past, 13, 1539–1571, https://doi.org/10.5194/cp-13-1539-2017, publisher: Copernicus GmbH, 2017.

Lu, M., Lu, Y., Ikejiri, T., Sun, D., Carroll, R. E., Blair, E. H., Algeo, T. J., and Sun, Y.: Periodic oceanic euxinia and terrestrial fluxes linked to astronomical forcing during the Late Devonian Frasnian-Famennian mass extinction, Earth and Planetary Science Letters, 562, 116 839, https://doi.org/10.1016/j.epsl.2021.116839, mAG ID: 3134992909, 2021.

Ma, K., Hinnov, L., Zhang, X., and Gong, Y.: Astronomical climate changes trigger Late Devonian bio- and environmental events in South China, Global and Planetary Change, 215, 103 874, https://doi.org/10.1016/j.gloplacha.2022.103874, 2022.

Ma, K., Hinnov, L., Wang, Z., Wang, K., Zong, R., Zhang, X., Song, J., Bai, Y., and Gong, Y.: Astronomically Forced Dynamics of Late Devonian (Famennian) Sea Level and Biotic Recovery in Western Junggar, Northwest China, Global and Planetary Change, 245, 104 677, https://doi.org/10.1016/j.gloplacha.2024.104677, 2025.

MacKay, D. J. et al.: Introduction to Gaussian processes, NATO ASI series F computer and systems sciences, 168, 133–166, 1998.

Maffre, P., Swanson-Hysell, N. L., and Goddéris, Y.: Limited Carbon Cycle Response to Increased Sulfide Weathering Due to Oxygen Feedback, Geophysical Research Letters, 48, e2021GL094 589, https://doi.org/10.1029/2021GL094589, 2021.

Maffre, P., Godderis, Y., Pohl, A., Donnadieu, Y., Carretier, S., and Hir, G. L.: The Complex Response of Continental Silicate Rock Weathering to the Colonization of the Continents by Vascular Plants in the Devonian, American Journal of Science, 322, 461–492, https://doi.org/10.2475/03.2022.02, 2022.

Maffre, P., Goddéris, Y., Le Hir, G., Nardin, É., Sarr, A.-C., and Donnadieu, Y.: GEOCLIM7, an Earth System Model for Multi-Million Years Evolution of the Geochemical Cycles and Climate, Geoscientific Model Development Discussions, pp. 1–66, https://doi.org/10.5194/gmd-2024-220, 2024.

750 Marcilly, C. M., Torsvik, T. H., Domeier, M., and Royer, D. L.: New paleogeographic and degassing parameters for long-term carbon cycle models, Gondwana Research, 97, 176–203, https://doi.org/10.1016/j.gr.2021.05.016, 2021.

McKay, M. D., Beckman, R. J., and Conover, W. J.: A comparison of three methods for selecting values of input variables in the analysis of output from a computer code, Technometrics, 42, 55–61, 2000.

Milankovitch, M.: Kanon der Erdbestrahlung und seine Anwendung auf das Eiszeitenproblem, Royal Serbian Academy Special Publication, 755 133, 1–633, 1941.

Mogavero, F. and Laskar, J.: The origin of chaos in the Solar System through computer algebra, Astronomy & Astrophysics, 662, L3, https://doi.org/10.1051/0004-6361/202243327, publisher: EDP Sciences, 2022.

Müller, R. D., Dutkiewicz, A., Zahirovic, S., Merdith, A. S., Scotese, C. R., Mills, B. J. W., Ilano, L., and Mather, B.: Solid Earth Carbon Degassing and Sequestration Since 1 Billion Years Ago, Geochemistry, Geophysics, Geosystems, 25, e2024GC011713, 760 https://doi.org/10.1029/2024GC011713, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1029/2024GC011713, 2024.

Oakley, J. and O'Hagan, A.: Bayesian inference for the uncertainty distribution of computer model outputs, Biometrika, 89, 769–784, https://doi.org/10.1093/biomet/89.4.769, 2002.

Park, Y., Maffre, P., Goddéris, Y., Macdonald, F. A., Anttila, E. S. C., and Swanson-Hysell, N. L.: Emergence of the Southeast Asian Islands as a Driver for Neogene Cooling, Proceedings of the National Academy of Sciences, 117, 25319–25326, 765 https://doi.org/10.1073/pnas.2011033117, 2020.

Percival, L. M. E., Davies, J. H. F. L., Schaltegger, U., De Vleeschouwer, D., Da Silva, A.-C., and Föllmi, K. B.: Precisely dating the Frasnian–Famennian boundary: implications for the cause of the Late Devonian mass extinction, Scientific Reports, 8, 9578, https://doi.org/10.1038/s41598-018-27847-7, publisher: Nature Publishing Group, 2018.

Rasmussen, C. E.: Gaussian processes for machine learning, Adaptive computation and machine learning, MIT Press, Cambridge, Mass., 770 ISBN 0-262-18253-X, https://ils.bib.uclouvain.be/global/documents/1896326, 2006.

Reershemius, T. and Planavsky, N. J.: What Controls the Duration and Intensity of Ocean Anoxic Events in the Paleozoic and the Mesozoic?, Earth-Science Reviews, 221, 103787, https://doi.org/10.1016/j.earscirev.2021.103787, 2021.

Rougier, J., Guillas, S., Maute, A., and Richmond, A. D.: Expert Knowledge and Multivariate Emulation: The Thermosphere–Ionosphere Electrodynamics General Circulation Model (TIE-GCM), Technometrics, 51, 414–424, https://doi.org/10.1198/TECH.2009.07123, pub-775 lisher: ASA Website _eprint: https://doi.org/10.1198/TECH.2009.07123, 2009.

Sablon, L.: HadSM3 Emulator for the Devonian: Dataset and Code, https://doi.org/10.5281/zenodo.15113858, 2025a.

Sablon, L.: GEOCLIM6.1-EMUL, https://doi.org/10.5281/zenodo.15114774, 2025b.

Scotese, C. R. and Wright, N. M.: PALEOMAP Paleodigital Elevation Models (PaleoDEMS) for the Phanerozoic, https://doi.org/10.5281/zenodo.5460860, 2018.

780 Sharaf, S. G. and Boudnikova, N. A.: Secular variations of elements of the Earth's orbit which influences the climates of the geological past, Tr. Inst. Theor. Astron. Leningrad, 11, 231–261, 1967.

Šidlichovský, M. and Nesvorný, D.: Frequency Modified Fourier Transform and Its Application to Asteroids, in: The Dynamical Behaviour of Our Planetary System, edited by Dvorak, R. and Henrard, J., pp. 137–148, Springer Netherlands, Dordrecht, ISBN 978-94-011-5510-6, https://doi.org/10.1007/978-94-011-5510-6_9, 1997.

785 Smart, M. S., Filippelli, G., Gilhooly, W. P., Ozaki, K., Reinhard, C. T., Marshall, J. E. A., and Whiteside, J. H.: The Expansion of Land Plants during the Late Devonian Contributed to the Marine Mass Extinction, Communications Earth & Environment, 4, 1–13, https://doi.org/10.1038/s43247-023-01087-8, 2023.

Sobolʹ, I. M.: Global Sensitivity Indices for Nonlinear Mathematical Models and Their Monte Carlo Estimates, Mathematics and Computers in Simulation, 55, 271–280, https://doi.org/10.1016/S0378-4754(00)00270-6, 2001.

790 Urban, N. M. and Fricker, T. E.: A Comparison of Latin Hypercube and Grid Ensemble Designs for the Multivariate Emulation of an Earth System Model, Computers & Geosciences, 36, 746–755, https://doi.org/10.1016/j.cageo.2009.11.004, 2010.

Urquhart, M., Ljungskog, E., and Sebben, S.: Surrogate-based optimisation using adaptively scaled radial basis functions, Applied Soft Computing, 88, 106 050, https://doi.org/10.1016/j.asoc.2019.106050, 2020.

Valdes, P. J., Armstrong, E., Badger, M. P. S., Bradshaw, C. D., Bragg, F., Crucifix, M., Davies-Barnard, T., Day, J. J., Farnsworth, A.,
795 Gordon, C., Hopcroft, P. O., Kennedy, A. T., Lord, N. S., Lunt, D. J., Marzocchi, A., Parry, L. M., Pope, V., Roberts, W. H. G., Stone, E. J., Tourte, G. J. L., and Williams, J. H. T.: The BRIDGE HadCM3 family of climate models: HadCM3@Bristol v1.0, Geoscientific Model Development, 10, 3715–3743, https://doi.org/10.5194/gmd-10-3715-2017, publisher: Copernicus GmbH, 2017.

Valdes, P. J., Scotese, C. R., and Lunt, D. J.: Deep ocean temperatures through time, Climate of the Past, 17, 1483–1506, https://doi.org/10.5194/cp-17-1483-2021, publisher: Copernicus GmbH, 2021.

800 Van Breedam, J., Huybrechts, P., and Crucifix, M.: A Gaussian process emulator for simulating ice sheet–climate interactions on a multi-million-year timescale: CLISEMv1.0, Geoscientific Model Development, 14, 6373–6401, https://doi.org/10.5194/gmd-14-6373-2021, publisher: Copernicus GmbH, 2021.

West, A. J.: Thickness of the Chemical Weathering Zone and Implications for Erosional and Climatic Drivers of Weathering and for Carbon-Cycle Feedbacks, Geology, 40, 811–814, https://doi.org/10.1130/G33041.1, 2012.

805 Whalen, M. T., De Vleeschouwer, D., Payne, J. H., Day, J. E. J., Over, D. J., and Claeys, P.: Pattern and timing of the Late Devonian biotic crisis in Western Canada: Insights from carbon isotopes and astronomical calibration of magnetic susceptibility data, in: New advances in Devonian carbonates: Outcrop analogs, reservoirs and chronostratigraphy, edited by Playton, T. E., Kerans, C., and Weissenberger, J. A., vol. 107, p. 0, Society for Sedimentary Geology, ISBN 978-1-56576-344-9, https://doi.org/10.2110/sepmsp.107.02, 2017.

Wichern, N. M. A., Bialik, O. M., Nohl, T., Percival, L. M. E., Becker, R. T., Kaskes, P., Claeys, P., and De Vleeschouwer, D.: Astronomically
810 paced climate and carbon cycle feedbacks in the lead-up to the Late Devonian Kellwasser Crisis, Climate of the Past, 20, 415–448, https://doi.org/10.5194/cp-20-415-2024, publisher: Copernicus GmbH, 2024.

Widmann, D., Tebbutt, W., st–, Galy-Fajou, T., Yalburgi, S., Ge, H., Surace, S. C., david-vicente, Vaverka, J., Skovbekk, J., Bosch, N., Schmitz, N., Viljoen, R., Wright, T., Vikram, and andreaskoher: JuliaGaussianProcesses/AbstractGPs.jl: v0.5.21, https://doi.org/10.5281/zenodo.10801764, 2024.

815 Wilkinson, R. D.: Bayesian Calibration of Expensive Multivariate Computer Experiments, in: Large-Scale Inverse Problems and Quantification of Uncertainty, chap. 10, pp. 195–215, John Wiley & Sons, Ltd, ISBN 978-0-470-68585-3, https://doi.org/10.1002/9780470685853.ch10, 2010.

Williams, K. D., Senior, C. A., Slingo, A., and Mitchell, J. F. B.: Towards evaluating cloud response to climate change using clustering technique identification of cloud regimes, Climate Dynamics, 24, 701–719, https://doi.org/10.1007/s00382-004-0512-z, 2005.

820 Zeebe, R. E. and Kocken, I. J.: Applying astronomical solutions and Milanković forcing in the Earth sciences, Earth-Science Reviews, p. 104959, https://doi.org/10.1016/j.earscirev.2024.104959, 2024.

Zeebe, R. E. and Lantink, M. L.: Milanković Forcing in Deep Time, Paleoceanography and Paleoclimatology, 39, e2024PA004 861, https://doi.org/10.1029/2024PA004861, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1029/2024PA004861, 2024.