Paleoceanography and Paleoclimatology

RESEARCH ARTICLE
10.1029/2024PA004880

Bayesian Errors-in-Variables Estimation of Specific Climate Sensitivity

D. Heslop[1], E. J. Rohling[1,2], G. L. Foster[2], and J. Yu[1,3]

[1]Research School of Earth Sciences, Australian National University, Canberra, ACT, Australia, [2]School of Ocean and Earth Science, National Oceanography Centre, University of Southampton, Southampton, UK, [3]Laoshan Laboratory, Qingdao, China

Key Points:

- We demonstrate that regression-based estimates of specific climate sensitivity may be biased toward zero because of data uncertainties
- A Bayesian error-in-variables approach is developed that accounts for data uncertainties in regression-based climate sensitivity estimates
- Eocene specific climate sensitivity is estimated to demonstrate the utility of Bayesian error-in-variables regression

Correspondence to:
D. Heslop,
david.heslop@anu.edu.au

Citation:
Heslop, D., Rohling, E. J., Foster, G. L., & Yu, J. (2024). Bayesian errors-in-variables estimation of specific climate sensitivity. *Paleoceanography and Paleoclimatology*, *39*, e2024PA004880. https://doi.org/10.1029/2024PA004880

Received 14 FEB 2024
Accepted 20 SEP 2024

Author Contributions:
Conceptualization: D. Heslop, E. J. Rohling, G. L. Foster, J. Yu
Funding acquisition: E. J. Rohling, G. L. Foster
Methodology: D. Heslop, E. J. Rohling
Project administration: E. J. Rohling
Software: D. Heslop
Writing – original draft: D. Heslop, E. J. Rohling, G. L. Foster, J. Yu
Writing – review & editing: D. Heslop, E. J. Rohling, G. L. Foster, J. Yu

Abstract Estimation of climate sensitivity is fundamental to assessing how global climate will warm as atmospheric CO_2 concentration increases. Geological archives of environmental change provide insights into Earth's past climate, but the incomplete nature of paleoclimate reconstructions and their inherent uncertainties make estimation of climate sensitivity challenging. Thus, quantifying climate sensitivity and assessing how it changed through geological time requires statistical frameworks that can handle data uncertainties in a principled fashion. Here we demonstrate some of the hurdles to estimating climate sensitivity, with a focus on current statistical techniques that may underestimate both climate sensitivity and its associated uncertainty. To solve these issues, we present a Bayesian error-in-variables regression model, which can yield estimates of climate sensitivity without bias. The regression model is flexible and can account for data point uncertainties with a known parametric form. The utility of this approach is demonstrated by estimating specific climate sensitivity with uncertainty for the Eocene.

Plain Language Summary As atmospheric CO_2 increases due to human activities, the Earth will warm. But how much warming can be expected? Climate sensitivity describes how much global average surface temperature will warm with a given increase in atmospheric CO_2. While this is a simple definition, estimating climate sensitivity is difficult because Earth's climate system is complex with a number of poorly understood interacting parts. One approach to estimating climate sensitivity is to quantify how Earth's climate changed as a result of variations in atmospheric CO_2 through geological time. This information is invaluable, but it is patchy and has large uncertainties that make estimating climate sensitivity challenging. In particular, existing statistical techniques may underestimate climate sensitivity and, thus, underestimate future warming. In this paper we present an alternative approach to determining climate sensitivity that overcomes the underestimation problem and demonstrate its performance using geological data from the Eocene epoch.

1. Introduction

Climate sensitivity (S) corresponds to the rise in global mean surface temperature (GMST) for a given radiative forcing (PALAEOSENS Project Members, 2012). By 2100, atmospheric CO_2 is expected to have more than doubled from its preindustrial level under a scenario described by the Intergovernmental Panel on Climate Change as "Intermediate," thus estimates of S are crucial to the understanding of future climate change (Pörtner et al., 2022). Paleoclimate reconstructions play an important role in determining how large-scale variations in atmospheric CO_2 have influenced past climates on centennial and multi-millennial time scales (Friedrich et al., 2016; Köhler et al., 2010, 2015, 2017, 2018; Martínez-Botí et al., 2015; Pagani et al., 2010; PALAEOSENS Project Members, 2012; Snyder, 2016; The Cenozoic CO2 Proxy Integration Project (CenCO2PIP) Consortium, 2023; von der Heydt et al., 2016). Furthermore, the recent study of Sherwood et al. (2020) developed a new statistical framework to constrain climate sensitivity via the combination of multiple lines of evidence. Importantly, paleoclimate reconstructions provide information across geological intervals with strongly differing climates, thus elucidating the complexities of different parts of the global climate system and their interactions (Anagnostou et al., 2020; Brown et al., 2022; Hansen et al., 2013; The Cenozoic CO2 Proxy Integration Project (CenCO2PIP) Consortium, 2023).

"Equilibrium" climate sensitivity is defined as the GMST change between two equilibrium climate states caused by a given radiative forcing (Forster et al., 2021). For example, in the case of atmospheric CO_2 doubling, this corresponds to GMST change in response to a radiative forcing of ~3.7 Wm^{-2} (Myhre et al., 1998). "Specific"

climate sensitivity is the change in GMST per 1 Wm^{-2} of radiative forcing (PALAEOSENS Project Members, 2012). If a given reference equilibrium climate state is perturbed, for example, by a change in atmospheric CO_2, while the GMST remains fixed, the change in the top of atmosphere radiative flux corresponds to a forcing, R. If ΔT is the subsequent change in GMST in response to R, specific climate sensitivity can be estimated:

$$S = \frac{\Delta T}{R}. \quad (1)$$

Equilibrium climate states are typically chosen according to specific research questions and hypotheses to be tested. For example, a preindustrial reference state is typically employed when considering potential GMST change due to anthropogenic CO_2 emissions.

Because radiative forcing is controlled by a range of connected processes, for example, atmospheric CO_2, land-ice albedo, atmospheric aerosols, etc., estimates of S will depend on which forcing parameters are considered. To avoid ambiguity, PALAEOSENS Project Members (2012) defined a nomenclature to indicate the sources of radiative forcing under consideration. For example, a reconstruction considering the radiative impacts of atmospheric CO_2 and land-ice albedo (LI) will yield an estimate of $S_{[CO2,LI]}$. We adopt $S_{[X]}$ as a generic term for the specific climate sensitivity in the following exposition and use appropriate terms when real-world examples are considered.

Equation 1 is susceptible to noise and when the denominator is small, $S_{[X]}$ can vary strongly due to data uncertainty. Alternatively, if $S_{[X]}$ is estimated based on substantially different climate states, the numerator and denominator will have large values but the estimated $S_{[X]}$ will only represent an average over potentially state-dependent values. When a limited number of observations is available, Equation 1 may be the only option to estimate $S_{[X]}$. However, when a larger number of observations is available, regression of ΔT on $R_{[X]}$ can provide a more informative estimate (Köhler et al., 2015). For example, if climate sensitivity is assumed to be constant with respect to the radiative forcing spanned by the data, $S_{[X]}$ can be estimated by fitting the straight-line relationship:

$$\Delta T = S_{[X]} R_{[X]} + \gamma, \quad (2)$$

where γ is an intercept term. Alternatively, if climate sensitivity is thought to have varied as a function of radiative forcing, $S_{[X]}$ can be estimated as the derivative of a higher-order polynomial fitted to the data (Köhler et al., 2015). Such fitting is typically performed via ordinary least-squares (OLS). Furthermore, previous studies have argued that if there is no change in radiative forcing (i.e., $R_{[X]} = 0$), there will be no corresponding change in GMST (i.e., $\Delta T = 0$), so that the intercept term in Equation 3 is not required (Köhler et al., 2015, 2017, 2018) and:

$$\Delta T = S_{[X]} R_{[X]}. \quad (3)$$

Equation 3 will only hold if $R_{[X]}$ includes complete forcing information. Typically, paleoclimate reconstructions are made via a set of incomplete observations because it is infeasible to determine all the forcing parameters, feedbacks, etc., that influence GMST (PALAEOSENS Project Members, 2012). Therefore it is expected that GMST will vary due to forcing mechanisms not quantified by a given $R_{[X]}$. As such, an intercept term will typically be required to account for the potential influence of unobserved forcing on GMST.

Observations of T and R include uncertainties due to errors in instrumental measurements, proxy calibration, etc. These uncertainties play a crucial role in estimating $S_{[X]}$ and its uncertainty. Observational uncertainties are typically propagated numerically, with the OLS regression process repeated multiple times using realizations drawn at random from data point uncertainties (i.e., parametric bootstrapping, Efron and Tibshirani (1994)). A confidence interval for $S_{[X]}$ is then estimated based on percentiles of the regression coefficients (Köhler et al., 2015). However, OLS is not symmetric and assumes that the uncertainties associated with the independent variable $\left(R_{[X]}\right)$ are negligible (Draper & Smith, 1998). When estimating $S_{[X]}$ via OLS regression, uncertainties in ΔT contribute to uncertainty in the fitted slope, and hence $S_{[X]}$, but do not introduce bias. In contrast, uncertainties in R will introduce both uncertainty and bias in the fitted slope. This is referred to as regression "dilution" and as the uncertainty in $R_{[X]}$ increases, the biased estimate of $S_{[X]}$ will approach zero (Frost & Thompson, 2000).

Previous studies have considered regression asymmetry of OLS by reversing the roles of ΔT and $R_{[X]}$, that is, regress $R_{[X]}$ on ΔT rather than ΔT on $R_{[X]}$ (Köhler et al., 2017). Such an approach; however, does not consider the uncertainties in the independent and dependent variables simultaneously, therefore yielding unreliable estimates of the fitted slope uncertainty.

Regression techniques are available that consider simultaneously uncertainty in both the independent and dependent variables. For example, major axis and scaled major axis regression correspond to the leading principal component of the data covariance and correlation matrices, respectively (Warton et al., 2006). However, these approaches do not incorporate known data uncertainties into their estimates, rather uncertainty is estimated based on the misfit of the model to the data. Similarly, total least-squares relies on assumptions concerning constant data uncertainties that will typically be inappropriate for paleoclimate data (Markovsky & Van Huffel, 2007). Brown et al. (2022) addressed dilution in paleoclimate data by estimating $S_{[X]}$ via SIMEX regression (Cook & Stefanski, 1994; Lederer & Seibold, 2019) which can incorporate Gaussian uncertainties for individual data points. High-quality paleoclimatic reconstructions now place a strong focus on full propagation of uncertainties (The Cenozoic CO2 Proxy Integration Project (CenCO2PIP) Consortium, 2023). When such quantitative information is available, it should be incorporated into the regression analysis directly to enable rigorous estimation of $S_{[X]}$. This requires a flexible regression framework that can handle specific, potentially non-Gaussian, uncertainties for individual data points.

In this paper we consider the challenges of estimating $S_{[X]}$ when uncertainties in the independent and dependent variables are known. Importantly, it should be possible to define these uncertainties fully and not rely on assumptions, such as Gaussian uncertainties or that the ratio of uncertainty variances in the independent and dependent variables is constant. To achieve this, we propose a solution based on an errors-in-variables (EIV) approach that can be placed readily into a Bayesian framework to enable estimation of the uncertainty associated with $S_{[X]}$ in a principled fashion.

2. Errors-in-Variables Regression

A likelihood-based errors-in-variables (EIV) approach developed by Sambridge (2016) enables observational uncertainties to be combined in a Bayesian regression framework. Importantly, the EIV approach of Sambridge (2016) is symmetrical, considering uncertainty in both the independent and dependent variables simultaneously. It will therefore not suffer from dilution bias.

Consider a regression model defined by a set of parameters, such as a straight-line consisting of a gradient and intercept. The "likelihood" of that set of model parameters given the observed data is a function that measures the probability of the observed data given the parameters (Bishop, 2006). Therefore, a model that fits the data poorly will have a low likelihood, and better fitting models with have higher likelihoods. Regression can therefore be performed by adjusting model parameters ($\mathbf{c}$) to maximize the likelihood given the data ($\mathbf{d}$). This will provide a maximum likelihood estimate (MLE) corresponding to the model parameters considered the most likely to have produced the observed data, that is, the model maximizes the probability; $p(\mathbf{d}|\mathbf{c})$.

Sambridge (2016) demonstrated how to determine the likelihood of a given curve with respect to a single data point with uncertainties in the independent (X) and dependent (Y) variables, which together define a joint probability density function $\left(f_{X,Y}\right)$. Specifically, Sambridge (2016) showed that the likelihood could be found by integrating $f_{X,Y}$ along the proposed regression curve (Figure 1). Formally, if θ corresponds to the position along a given curve, the likelihood of the data given the curve is:

$$p(\mathbf{d}|\mathbf{c}) = \int p(\mathbf{d}|\theta,\mathbf{c})p(\theta|\mathbf{c})d\theta, \tag{4}$$

where $p(\theta|\mathbf{c})$ is set to 1 because it is considered that all poistions along the curve could be responsible for the observed data point with equal probability (Sambridge, 2016). In the case of multiple data points it is reasonable to assume their uncertainties are independent of each other (i.e., the uncertainties associated with one data point do not depend on the uncertainties of the other data points). Then the overall likelihood considering all the data is simply the product of the integrals of the individual data points. For example, if there are N data points the likelihood is:

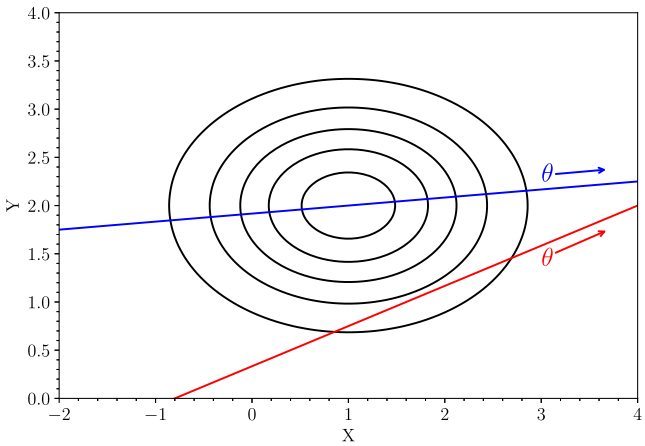

Figure 1. Demonstration of how the likelihood; $p(\mathbf{d}|\mathbf{c})$, in Equation 4 is determined. Black lines represent the $f_{X,Y}$ contours of a bivariate Gaussian distribution and two candidate regression lines are shown in blue and red. Positions along the lines are denoted by their θ values. The likelihood corresponding to each line is the integral of the bivariate Gaussian distribution with respect to θ along the lines. In this example, $p(\mathbf{d}|\mathbf{c})$ is greater for the blue line than the red line because it passes through the main body of the distribution where $f_{X,Y}$ is larger.

$$p(\mathbf{d}|\mathbf{c}) = \prod_{i=1}^{N} \int p(\mathbf{d}_i|\theta,\mathbf{c})\, p(\theta|\mathbf{c}) d\theta. \quad (5)$$

Equation 5 demonstrates that the likelihood can be determined if the data point uncertainties have a parametric form that can be used to define a joint probability function. In the case of data points with dependent uncertainties, the product of one-dimensional integrals employed in Equation 5 is inappropriate and $p(\mathbf{d}|\mathbf{c})$ must instead be found by multidimensional integration (see Sambridge (2016) for details). Previous regression-based estimates of $S_{[X]}$ have assumed independent uncertainties (Köhler et al., 2010, 2015, 2017), so Equation 5 is adopted here. Importantly, the outlined likelihood-based approach is flexible and does not assume Gaussian distributed uncertainties or even that data points have uncertainties with the same parametric form. Furthermore, for a given data point the independent and dependent parameter uncertainties can be correlated if required.

2.1. Maximum Likelihood Estimation

We first demonstrate how Equation 5 can be used to fit a polynomial to a collection of data points. 50 uniformly distributed numbers were generated in the interval (0,1) to act as X values. Values of Y were obtained as $Y = 3X$. This example relationship was selected because of its similarity to the $S_{[CO2]} = 3.08 \pm 0.96$ K W^{-1}m^2 estimated by PALAEOSENS Project Members (2012). All the points were then assigned errors based on Gaussian uncertainties with standard deviations of 0.2 and 0.1 in X and Y, respectively. Addition of these errors mean that the points no longer sit perfectly on a straight-line and can reside outside the (0,1) and (0,3) intervals for X and Y, respectively (Figure 2a).

While adaptive quadrature (Press et al., 2007) can be used to estimate the indefinite integral in Equation 4 along the entire length of a line, it is inefficient. We therefore employed the trapezoid rule (Press et al., 2007) to estimate the definite integral between limits based on the data. These limits, denoted as X_0 and X_1, are defined according to

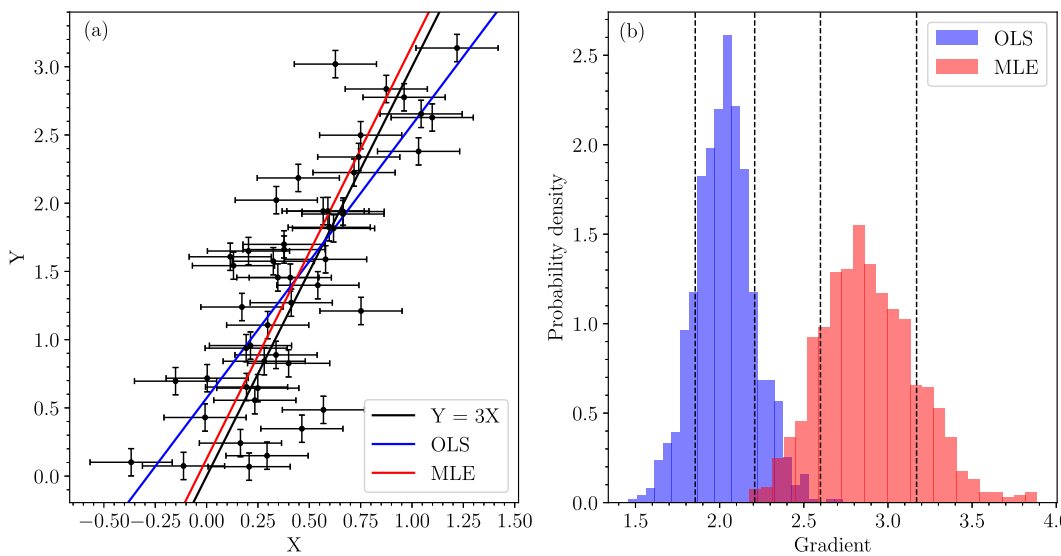

Figure 2. (a) Synthetic regression example based on $Y = 3X$ and data uncertainties with standard deviations of 0.2 and 0.1 in X and Y, respectively. While the ordinary least-squares (OLS) (blue) regression has a shallow gradient due to dilution, the maximum likelihood estimate (MLE) (red) gradient is consistent with the true gradient. (b) The fitting procedure shown in panel (a) was repeated 1,000 times with different random data. Histograms of the estimated gradients demonstrate that the OLS (blue) fits are consistently too shallow, while the MLE (red) fits are centered around the true gradient of 3. The 2.5 and 97.5 percentiles of the estimated gradients are shown as dashed lines.

X values selected to be far enough removed from the data that $f_{X,Y}$ will effectively be zero. In the case of a straight-line, the regression relationship can be defined by a line connecting two coordinates; (X_0, Y_0) and (X_1, Y_1). If θ is the distance along the line from (X_0, Y_0) and θ_1 is the distance between (X_0, Y_0) and (X_1, Y_1), then Equation 5 becomes:

$$p(\mathbf{d}|\mathbf{c}) = \prod_{i=1}^{N} \int_0^{\theta_1} p(\mathbf{d}_i|\theta, \mathbf{c})\, p(\theta|\mathbf{c}) d\theta. \tag{6}$$

The data were fitted with straight-lines using both OLS and MLE. The MLE solution was found numerically by determining the values of Y_0 and Y_1 that maximized Equation 6 when X_0 and X_1 were set to -5 and 5, respectively. This maximization was performed via sequential least squares programming (Virtanen et al., 2020) with Y_0 and Y_1 initialized based on the OLS fit (to avoid floating point underflow, numerical maximization was based on the logarithm of the likelihood). The resulting fits are compared in Figure 2a, which demonstrates that the MLE-based gradient is steeper than the OLS-based gradient, which suffers from dilution bias. This process was repeated with 1,000 random data sets to compare the estimated gradients of the OLS and MLE approaches (Figure 2b). This again demonstrates that the OLS regression suffers from dilution, while the MLE estimates are consistent with the $Y = 3X$ relationship.

2.2. Bayesian EIV Regression

Given the EIV likelihood (Equation 6), the regression process can be placed into a Bayesian framework to estimate uncertainty in $S_{[X]}$. The theorem of Bayes (1763) enables estimation of the a posteriori probability density function of the regression model parameters ($\mathbf{c}$) given the data ($\mathbf{d}$):

$$p(\mathbf{c}|\mathbf{d}) = \frac{p(\mathbf{d}|\mathbf{c})p(\mathbf{c})}{p(\mathbf{d})}. \tag{7}$$

where $p(\mathbf{c})$ and $p(\mathbf{d})$ are *prior* probability distributions for the regression model parameters and data, respectively. The $p(\mathbf{c}|\mathbf{d})$ distribution cannot be determined analytically, however, it can be approximated numerically though Markov chain Monte Carlo (MCMC) sampling (Bishop, 2006). In this study we sample the posterior distribution using a Python implementation (Foreman-Mackey et al., 2013) of the affine-invariant ensemble sampler for MCMC proposed by Goodman and Weare (2010).

When considering priors, MCMC sampling of the posterior does not require $p(\mathbf{d})$, but does require $p(\mathbf{c})$. Rather than formulating $p(\mathbf{c})$ based on knowledge of $S_{[X]}$, we instead define uniform priors for the Y_0 and Y_1 coordinates discussed in Section 2.1. For example, when considering a straight-line fit, the priors are based on one uniform distribution with user-defined limits at X_0 and a second identical uniform distribution at X_1. This is illustrated in Figure 3a using the example data from Section 2.1, where the X_0 and X_1 integration limits were set as -5 and 5, respectively. The Y variable limits for the priors are set as $(-30,30)$. A number is then drawn at random from the uniform distribution at X_0 and denoted as Y_0. Then a number is drawn from the uniform distribution at X_1 and denoted as Y_1. The equation of the resulting straight-line between $[X_0, Y_0]$ and $[X_1, Y_1]$ is one realization from $p(\mathbf{c})$ (Figure 3a). If Y_0 and Y_1 are drawn from identical uniform distributions with limits $[a,b]$, the prior on the intercept of the line (Figure 3b) will be a symmetrical triangular distribution with mean $(a+b)/2$ and variance $(a-b)^2/24$. The prior on the gradient of the line (Figure 3c) will also be a symmetrical triangular distribution with a mean of 0 and variance $(a-b)^2/6(X_1-X_0)^2$. Similarly, to fit a quadratic regression, uniform distributions can be defined at X_0, $(X_0+X_1)/2$, and X_1. Three random numbers drawn from these distributions then define coordinates along a quadratic. This process can be extended readily to higher-order polynomials as required. In the specific case of estimating climate sensitivity, the limits of the uniform distributions employed in defining the priors described above can be selected based on an expected range of ΔT or set to be sufficiently broad that the intercept and gradient triangular distributions are effectively flat. Importantly, the MCMC sampler of Foreman-Mackey et al. (2013) is flexible and it is straightforward to define alternative priors for the regression line intercept and gradient if desired.

To demonstrate the proposed approach we return to the $Y = 3X$ example employed in Section 2.1 (Figure 4). An OLS regression line was determined, with uncertainties estimated by 10,000 parametric bootstraps of the data

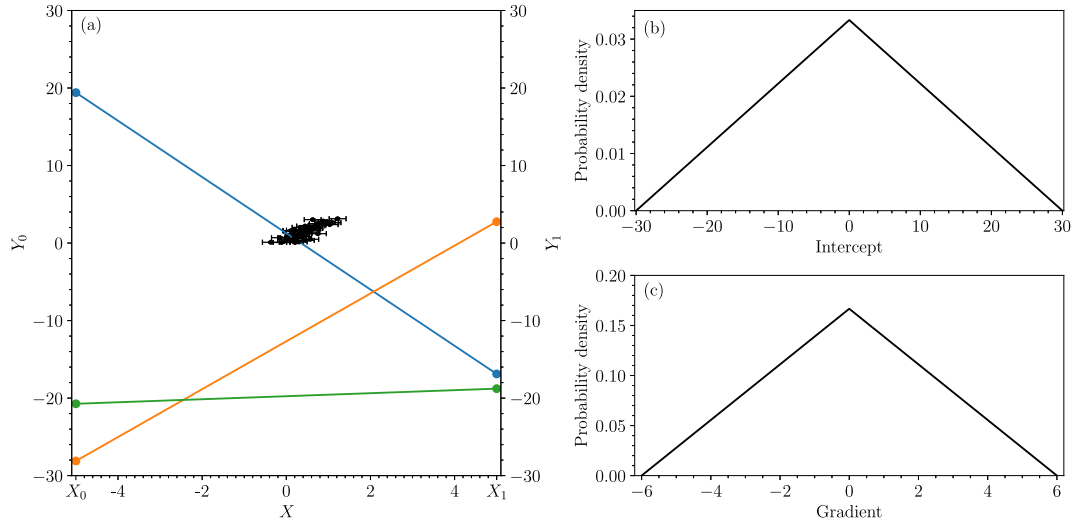

Figure 3. Demonstration of the scheme used to define a regression line prior. (a) The black points are the example data used in Figure 2, plotted between the X_0 and X_1 integration limits of −5 and 5, respectively. Two random numbers, Y_0 and Y_1, are then drawn from a uniform distribution with limits (−30,30) to define the points $[X_0, Y_0]$ and $[X_1, Y_1]$. These points are connected with a straight-line, which represents one realization of the regression line prior. The blue, green, and orange lines represent three example realizations from the prior. (b) The scheme illustrated in panel (a) yields a symmetrical triangular distribution for the prior on the intercept. (c) The scheme illustrated in panel (a) yields a symmetrical triangular distribution centered on zero for the prior on the gradient (see main text for details).

uncertainties (see Section 1). The OLS gradient estimates are centered around 1.5 and demonstrate bias resulting from regression dilution. Furthermore, the OLS 95% confidence interval of 1.21–1.77 for the gradient does not contain the true value, therefore it must be inferred that the gradient uncertainty is underestimated. The likelihood function for the Bayesian EIV was integrated between $X = -5$ and 5. The prior on the regression model was defined by broad uniform Y distributions between −30 and 30. The posterior distribution was sampled with the first 1,000 samples discarded as "burnin" (to minimize the influence of the sampler initialization) and the

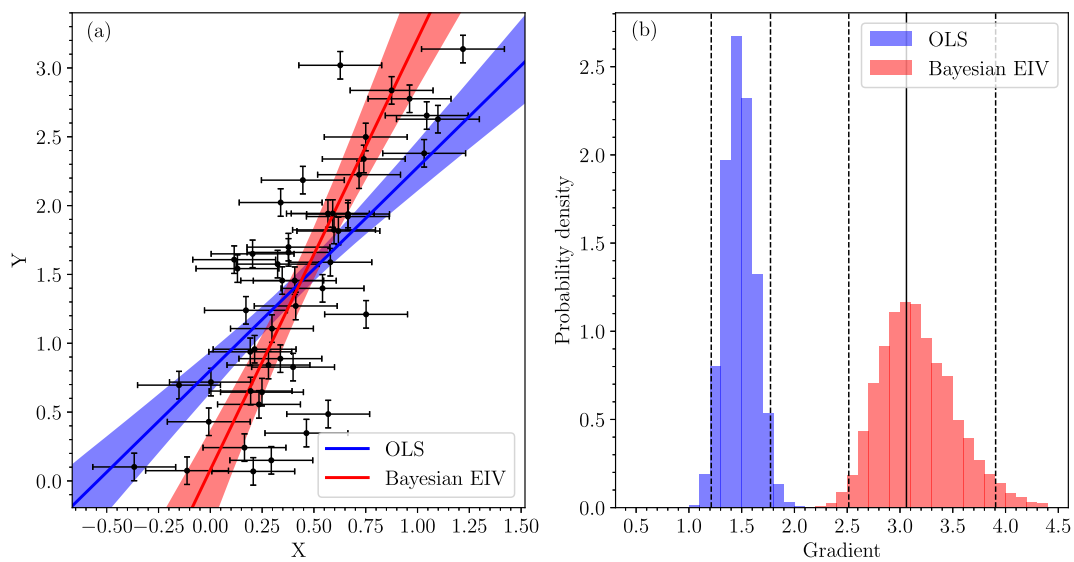

Figure 4. (a) Synthetic data set identical to that shown in Figure 2(a). The 95% confidence intervals for the ordinary least-squares (OLS) (blue) and Bayesian EIV (red) regressions are shown by shading, with thick lines indicating the median solutions. (b) Histograms of the estimated regression gradients with dashed lines showing the 95% confidence and credible intervals for the OLS and Bayesian EIV methods, respectively. The Bayesian EIV maximum a posteriori gradient estimate is indicated by a solid line.

subsequent samples "thinned" (to reduce autocorrelation) with every fiftieth sample being retained. The sampler was run until 10,000 post-burnin and post-thinning samples were available. The Bayesian gradient estimates have a 95% credible interval of 2.51–3.91 (unlike confidence intervals, there is a 95% probability that the true gradient lies within the credible interval, given the data and prior). Additionally, the maximum a posteriori (MAP, corresponding to the posterior distribution mode) gradient is 3.06, which compares well with the true gradient of 3.00 (Figure 4). This demonstrates the Bayesian EIV can reliably estimate the regression gradient and does not suffer from regression dilution.

3. Case Study

To further demonstrate the effect of OLS dilution and how Bayesian EIV can remedy this issue, we present a case study based on the Cenozoic atmospheric CO_2 reconstruction of The Cenozoic CO2 Proxy Integration Project (CenCO2PIP) Consortium (2023). Importantly, the CenCO2PIP reconstruction was formulated by the integration of critically evaluated CO_2 paleoproxies spanning the last 66 Ma (million years ago) with a focus on uncertainty propagation. The reconstruction considers published CO_2 estimates from both marine and terrestrial records, which were vetted to identify "top-tier" data. CenCO2PIP combined these data into 500-kyr windows using a hierarchical Bayesian model to determine posterior distributions of CO_2 concentration that account for uncertainties in both proxy-based estimates and their associated ages (The Cenozoic CO2 Proxy Integration Project (CenCO2PIP) Consortium, 2023).

Accompanying their CO_2 reconstruction, CenCO2PIP also derived a ΔGMST record (GMST relative to preindustrial) based on the benthic foraminiferal $\delta^{18}O$ compilation of Westerhold et al. (2020) and the $\delta^{18}O$ to temperature conversion of Hansen et al. (2013). The ΔGMST record was then segmented into 500-kyr posterior distributions with fully propagated uncertainties following the same Bayesian procedure employed for the CO_2 record; however, in the case of ΔGMST the age model of Westerhold et al. (2020) was assumed to have zero uncertainty. Readers are referred to The Cenozoic CO2 Proxy Integration Project (CenCO2PIP) Consortium (2023) for full details of how the 500-kyr windowed CO_2 and ΔGMST posterior distributions were estimated.

This case study involves estimating climate sensitivity for the Eocene (56–33.9 Ma, Gradstein and Ogg (2020)), which was a period of high CO_2 and GMST (Hansen et al., 2013). Because of the coarse temporal resolution of the CenCO2PIP data windows it is not feasible to estimate equilibrium climate sensitivity, which focuses on feedback processes over shorter time scales (The Cenozoic CO2 Proxy Integration Project (CenCO2PIP) Consortium, 2023). Rather, estimates of Eocene $S_{[CO2]}$ correspond to Earth system sensitivity, which includes longer geological scale feedback processes (The Cenozoic CO2 Proxy Integration Project (CenCO2PIP) Consortium, 2023). CenCO2PIP data were selected for 35–54 Ma, which omits the Eocene–Oligocene transition (~33.9 Ma), when permanent ice sheets started to form in Antarctica (see Hutchison et al. (2021) for a review) and the rapid CO_2 rise and fall of the Paleocene–Eocene Thermal Maximum at ~56 Ma (Röhl et al., 2007). Thus, the selected interval represents a greenhouse climate before the surface albedo of large continental ice sheets began to influence the energy balance in Earth's atmosphere.

The CenCO2PIP data products include the 2.5, 25, 50, 75, and 97.5 percentiles of the $\ln(CO_2)$ and ΔGMST posterior distributions for each 500-kyr window. The mean parameter value and uncertainty standard deviation for each $\ln(CO_2)$ and ΔGMST window was obtained by quantile-matching estimation under the assumption of Gaussian uncertainties (Delignette-Muller & Dutang, 2015). Figure 5 compares the CenCO2PIP 500-kyr window $\ln(CO_2)$ and ΔGMST posterior distribution percentile values to the corresponding percentiles from the quantile-matched Gaussian distributions. The similarities of the percentiles demonstrate that both the $\ln(CO_2)$ and ΔGMST uncertainties are almost Gaussian and can be readily approximated by the quantile-matched distributions.

Based on the relationship of Myhre et al. (1998), $\ln(CO_2)$ was converted to R relative to a preindustial atmospheric CO_2 concentration of 280 ppm:

$$R = 5.35(\ln(CO_2) - \ln(280)). \quad (8)$$

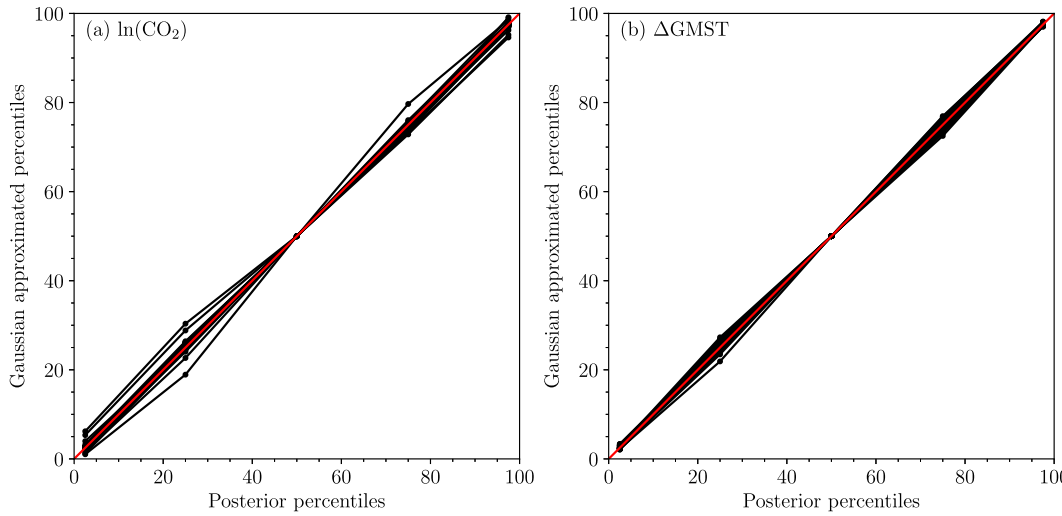

Figure 5. (a) Each CenCO2PIP 500-kyr window posterior distribution is represented by values corresponding to the 2.5, 25, 50, 75, and 97.5 percentiles. Given these values, each posterior was approximated by a best-fit Gaussian distribution obtained by quantile-matching. For each 500-kyr window the percentiles of the posterior distribution values were then found for the fitted Gaussian distribution. If the posterior percentiles and Gaussian approximated percentiles fall close to a 1:1 line (red) it implies the posterior distribution can be represented closely by a Gaussian distribution. Each black line represents the posterior versus Gaussian approximated percentiles for a given 500-kyr window. (a) Analysis of the $\ln(CO_2)$ 500-kyr window posterior percentiles reveals only small deviations from a Gaussian distribution. panel (b) Same as panel (a) for the ΔGMST 500-kyr window posterior percentiles.

Given the form of Equation 8, if $\ln(CO_2)$ has Gaussian uncertainties, then R will also have Gaussian uncertainties. Specifically, the $\ln(CO_2)$ Gaussian distribution defined by a mean $(\mu_{\ln(CO2)})$ and standard deviation $(\sigma_{\ln(CO2)})$ can be converted into a radiative forcing mean (μ_R) and standard deviation (σ_R) as.

$$\mu_R = 5.35(\mu_{\ln(CO2)} - \ln(280)) \tag{9}$$

$$\sigma_R = 5.35\sigma_{\ln(CO2)} \tag{10}$$

Eocene $S_{[CO2]}$ was estimated via straight-line OLS and Bayesian EIV fits. Uncertainties for the OLS regression were estimated from 10,000 parametric bootstrap samples of the data uncertainties as described in Section 1. The likelihood function for the Bayesian EIV was integrated between $R = 0$ and 15 W m^{-2}. Uniform ΔGMST distributions between −1,000 and 1,000 K were selected to define priors on Y_0 and $Y1$. Such broad distributions ensure that the resulting triangular distributions for the regression line intercept and gradient (Section 2.2) are effectively flat. The posterior distribution was then sampled using MCMC and the integrated autocorrelation length (Goodman & Weare, 2010) of the resulting chain was estimated to be ~33. This represents the number of steps needed for the chain to "forget" its previous position. Based on the autocorrelation length, 1,000 samples were discarded as burnin (representing ~30 times the autocorrelation length) and the subsequent samples thinned with every fiftieth sample being retained (representing ~1.5 times the autocorrelation length). Such thinning means that the MCMC samples representing the posterior distribution can be assumed to be effectively independent of each other. The sampler was run until 10,000 post-burnin and post-thinning samples were available. Based on the OLS and Bayesian EIV samples, 95% confidence intervals were determined for the regression lines (Figure 6a) and estimates of $S_{[CO2]}$ are shown in Figure 6b.

The OLS and Bayesian EIV techniques yield significantly different estimates of $S_{[CO2]}$, with 95% confidence/credible intervals that do not overlap. Based on these results we infer that OLS underestimates $S_{[CO2]}$ because of regression dilution and as in Section 2.2 the corresponding uncertainty can be assumed to be too narrow. The Bayesian EIV posterior $S_{[CO2]}$ distribution has a 95% credible interval of 2.02–2.87 K W^{-1} m^2 and an MAP of 2.37 K W^{-1} m^2. These estimates correspond to an Earth system sensitivity of 7.48–10.63 K per doubling of CO_2, with an MAP of 8.78 K per doubling of CO_2.

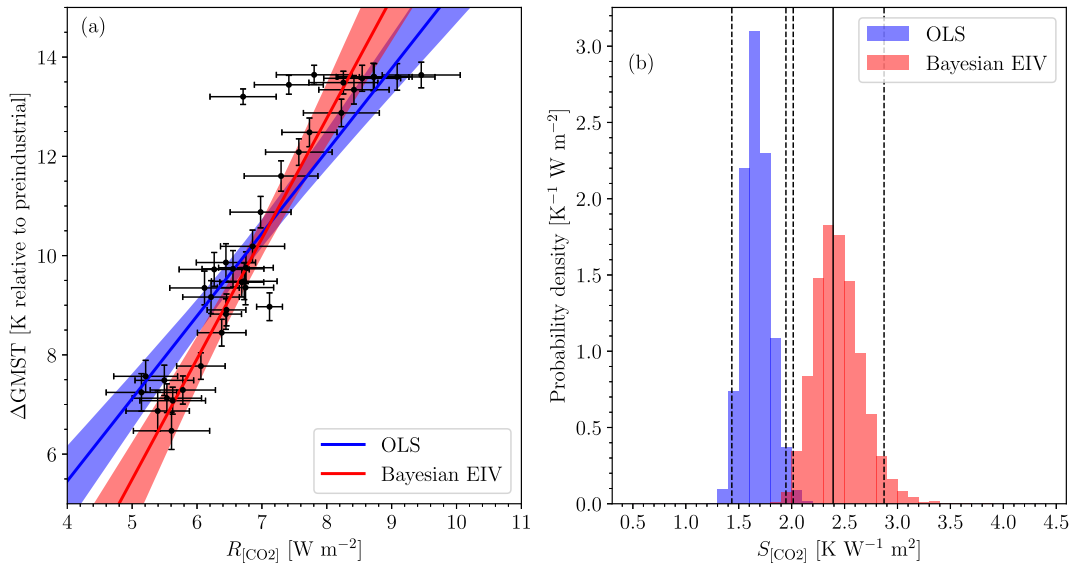

Figure 6. (a, b) same format as Figure 4 for the CenCO2PIP Eocene atmospheric CO_2 and ΔGMST records between 54 and 35 Ma.

4. Discussion

Estimates of climate sensitivity based on paleoclimate reconstructions are crucial to understanding both past and future climate change (Sherwood et al., 2020; The Cenozoic CO2 Proxy Integration Project (CenCO2PIP) Consortium, 2023). Obtaining paleoclimate data is challenging; therefore it must be ensured that statistical frameworks used to infer climate sensitivity can handle uncertainties in a principled fashion.

The proposed Bayesian EIV approach allows uncertainty in both the independent and dependent variables to be combined in a regression estimate. Although Bayesian EIV is more challenging to implement than OLS, it produces estimates of specific climate sensitivity that should not suffer from dilution bias. As discussed in Section 1, existing regression techniques used to estimate climate sensitivity rely on a number of assumptions concerning the form of data uncertainty. The proposed Bayesian EIV approach can accommodate data point specific errors in both $R_{[X]}$ and ΔT with any given parametric form. Furthermore, additional scatter not represented in the data point uncertainty could be readily incorporated into the Bayesian EIV formulation. For example, beyond their individual uncertainties, temperature observations may exhibit additional scatter resulting from unobserved processes. If such uncertainties can be formulated in terms of probability distributions, they can be incorporated into the Bayesian EIV model. We only discuss this concept in general-terms because adjustments to the model will depend specifically on the data under consideration. As an example, consider GMST estimates with known Gaussian distributed uncertainties (σ_T) that exhibit additional Gaussian scatter (σ_ϵ) resulting from unobserved forcing. If these two sources of uncertainty are independent they can be combined in quadrature for each data point (Taylor, 1997):

$$\xi_i = \sqrt{\sigma_{T_i}^2 + \sigma_\epsilon^2}. \tag{11}$$

The σ_ϵ term can then be estimated as part of the Bayesian EIV regression model with the likelihood modified to include ξ_i, rather than σ_T, and selection of an appropriate prior for $p(\sigma_\epsilon)$.

Thus, the strength of Bayesian EIV regression is an ability to define the uncertainty structure of the data fully, which in turn forms the basis of a rigorous estimate of climate sensitivity. This does limit the application of the Bayesian EIV approach to data with properly quantified uncertainties. As demonstrated by CenCO2PIP; however, large-scale paleoclimate reconstructions are placing a priority on uncertainty quantification. Therefore, Bayesian EIV is a timely addition to the statistical estimation of climate sensitivity.

Bayesian EIV produced a posterior distribution for $S_{[CO2]}$ with a 95% credible interval of 2.02–2.87 K W^{-1} m^2. The Bayesian EIV credible interval is consistent with previous estimates of $S_{[CO2]}$ for an ice-free world. For example, PALAEOSENS Project Members (2012) estimated $S_{[CO2]}$ with a 95% confidence interval of 1.20–4.96 K W^{-1} m^2. Furthermore, based on a plot of ln(CO_2) and ΔGMST, CenCO2PIP inferred Eocene $S_{[CO2]}$ to be ~2.2 K W^{-1} m^2, which is consistent with the Bayesian EIV MAP of 2.37 K W^{-1} m^2. Our Bayesian EIV analysis of the Eocene therefore strengthens paleoproxy-based evidence for higher Earth system sensitivity in past warm periods even when the large data uncertainties are taken into consideration.

5. Conclusions

Statistical frameworks for regression-based specific climate sensitivity estimates must consider uncertainties in a principled fashion. OLS regression assumes that uncertainties in estimated radiative forcing are negligible; if this assumption is not met the estimated sensitivity will be biased toward zero. This implies that paleoclimate-based sensitivity could be underestimated, and the temperature response to future greenhouse gas emission pathways could in turn be underestimated. The Bayesian EIV method presented here provides a symmetrical likelihood-based approach to regression, which can consider uncertainties of any given parametric form in both radiative forcing and GMST. The development of rigorous statistical approaches has been identified as a challenge for climate sensitivity estimation (von der Heydt et al., 2016). Bayesian EIV represents a step toward addressing this challenge in regression-based estimates of specific climate sensitivity.

Applying Bayesian EIV to the estimation of Eocene $S_{[CO2]}$ yielded an MAP climate sensitivity of 2.37 K W^{-1} m^2, with a 95% credible interval of 2.02–2.87 K W^{-1} m^2. This corresponds to an MAP of 8.78 K per doubling of CO_2 and supports previous studies that hypothesized that climate sensitivity was higher during the Eocene than the present day.

Data Availability Statement

Python code and Jupyter notebooks to recreate the calculations presented in this work are available on Zenodo (Heslop, 2024).

Acknowledgments
We thank the Editor and three reviewers for their detailed and constructive comments, which helped us to improve the quality of our manuscript. D.H. and E.J.R. were supported by the Australian Research Council Grants FL1201000050 and DP2000101157 and the Australian Research Council Special Research Initiative, Australian Centre for Excellence in Antarctic Science (project number SR200100008). Open access publishing facilitated by Australian National University, as part of the Wiley - Australian National University agreement via the Council of Australian University Librarians.

References

Anagnostou, E., John, E. H., Babila, T. L., Sexton, P. F., Ridgewell, A., Lunt, D. J., et al. (2020). Proxy evidence for state-dependence of climate sensitivity in the Eocene greenhouse. *Nature Communications*, *11*(1), 4436. https://doi.org/10.1038/s41467-020-17887-x

Bayes, T. (1763). A essay towards solving a problem in the doctrine of chances. *Philosophical Transactions of the Royal Society of London*, *53*, 370–418. https://doi.org/10.1098/rstl.1763.0053

Bishop, C. M. (2006). *Pattern recognition and machine learning*. Springer.

Brown, R. M., Chalk, T. B., Crocker, A. J., Wilson, P. A., & Foster, G. L. (2022). Late Miocene cooling coupled to carbon dioxide with Pleistocene-like climate sensitivity. *Nature Geoscience*, *15*(8), 664–670. https://doi.org/10.1038/s41561-022-00982-7

Cook, J. R., & Stefanski, L. A. (1994). Simulation-extrapolation estimation in parametric measurement error models. *Journal of the American Statistical Association*, *89*(428), 1314–1328. https://doi.org/10.2307/2290994

Delignette-Muller, M. L., & Dutang, C. (2015). fitdistrplus: An R package for fitting distributions. *Journal of Statistical Software*, *64*(4), 1–34. https://doi.org/10.18637/jss.v064.i04

Draper, N. R., & Smith, H. (1998). *Applied regression analysis*. John Wiley.

Efron, B., & Tibshirani, R. J. (1994). *An introduction to the bootstrap*. Chapman and Hall.

Foreman-Mackey, D., Hogg, D. W., Lang, D., & Goodman, J. (2013). Emcee: The MCMC hammer. *Publications of the Astronomical Society of the Pacific*, *125*(925), 306–312. https://doi.org/10.1086/670067

Forster, P., Storelvmo, T., Armour, K., Collins, W., Dufresne, J.-L., Frame, D., & Zhang, H. (2021). The Earth's energy budget, climate feedbacks, and climate sensitivity. In V. Masson-Delmotte, et al. (Eds.), *Climate change 2021: The physical science basis. Contribution of working group I to the sixth assessment report of the Intergovernmental Panel on Climate Change* (pp. 923–1054). Cambridge University Press. https://doi.org/10.1017/9781009157896.009

Friedrich, T., Timmermann, A., Tigchelaar, M., Timm, O. E., & Ganopolski, A. (2016). Nonlinear climate sensitivity and its implications for future greenhouse warming. *Science Advances*, *2*(11), e1501923. https://doi.org/10.1126/sciadv.1501923

Frost, C., & Thompson, S. G. (2000). Correcting for regression dilution bias: Comparison of methods for a single predictor variable. *Journal of the Royal Statistical Society*, *163*(2), 173–189. https://doi.org/10.1111/1467-985X.00164

Goodman, J., & Weare, J. (2010). Ensemble samplers with affine invariance. *Applied Mathematics and Computational Science*, *5*(1), 65–80. https://doi.org/10.2140/camcos.2010.5.65

Gradstein, F., & Ogg, J. (2020). Chapter 2 - the chronostratigraphic scale. In F. M. Gradstein, J. G. Ogg, M. D. Schmitz, & G. M. Ogg (Eds.), *Geologic time scale 2020* (pp. 21–32). Elsevier. https://doi.org/10.1016/B978-0-12-824360-2.00002-4

Hansen, J., Sato, M., Russell, G., & Kharecha, P. (2013). Climate sensitivity, sea level and atmospheric carbon dioxide. *Philosophical Transactions of the Royal Society A*, *371*(2001), 20120294. https://doi.org/10.1098/rsta.2012.0294

Heslop, D. (2024). Dave-heslop74/BEIV: Bayesian errors-in-variables regression estimation of specific climate sensitivity (v1.0.0) [Software]. *Zenodo*. https://doi.org/10.5281/zenodo.13800676

Hönisch, B., Royer, D. L., Breecker, D. O., Polissar, P. J., Bowen, G. J., Henehan, M. J., et al. (2023). Toward a Cenozoic history of atmospheric CO_2. *Science*, *382*(6675), eadi5177. https://doi.org/10.1126/science.adi5177

Hutchison, D. K., Coxall, H. K., Lunt, D. J., Steinthorsdottir, M., de Boer, A. M., Baatsen, M., et al. (2021). The Eocene-Oligocene transition: A review of marine and terrestrial proxy data, models and model-data comparisons. *Climate of the Past*, *17*(1), 269–315. https://doi.org/10.5194/cp-17-269-2021

Köhler, P., Bintanja, R., Fischer, H., Joos, F., Knutti, R., Lohmann, G., & Masson-Delmotte, V. (2010). What caused Earth's temperature variations during the last 800,000 years? Data-Based evidence on radiative forcing and constraints on climate sensitivity. *Quaternary Science Reviews*, *29*(1–2), 129–145. https://doi.org/10.1016/j.quascirev.2009.09.026

Köhler, P., de Boer, B., von der Heydt, A. S., Stap, L. B., & van de Wal, R. S. W. (2015). On the state dependency of the equilibrium climate sensitivity during the last 5 million years. *Climate of the Past*, *11*(12), 1801–1832. https://doi.org/10.5194/cp-11-1801-2015

Köhler, P., Knorr, G., Stap, L. B., Ganopolski, A., de Boer, B., van de Wal, R. S. W., et al. (2018). The effect of obliquity-driven changes on paleoclimate sensitivity during the Late Pleistocene. *Geophysical Research Letters*, *45*(13), 6661–6671. https://doi.org/10.1029/2018GL077717

Köhler, P., Stap, L. B., von der Heydt, A. S., de Boer, B., van de Wal, R. S. W., & Bloch-Johnson, J. (2017). A state-dependent quantification of climate sensitivity based on paleodata of the last 2.1 million years. *Paleoceanography*, *32*(11), 1102–1114. https://doi.org/10.1002/2017PA003190

Lederer, W., & Seibold, H. (2019). SIMEX- and MCSIMEX-algorithm for measurement error models (R package version 1.8). Retrieved from https://CRAN.R-project.org/package=simex

Markovsky, I., & Van Huffel, S. (2007). Overview of total least-squares methods. *Signal Processing*, *87*(10), 2283–2302. https://doi.org/10.1016/j.sigpro.2007.04.004

Martínez-Botí, M. A., Foster, G. L., Chalk, T. B., Rohling, E. J., Sexton, P. F., Lunt, D. J., et al. (2015). Plio-Pleistocene climate sensitivity evaluated using high-resolution CO_2 records. *Nature*, *518*(7537), 49–54. https://doi.org/10.1038/nature14145

Myhre, G., Highwood, E. J., Shine, K. P., & Stordal, F. (1998). New estimates of radiative forcing due to well mixed greenhouse gases. *Geophysical Research Letters*, *25*(14), 2715–2718. https://doi.org/10.1029/98GL01908

Pagani, M., Liu, Z., LaRiviere, J., & Ravelo, A. C. (2010). High Earth-system climate sensitivity determined from Pliocene carbon dioxide concentrations. *Nature Geoscience*, *3*(1), 27–30. https://doi.org/10.1038/ngeo724

PALAEOSENS Project Members. (2012). Making sense of palaeoclimate sensitivity. *Nature*, *491*(7426), 683–691. https://doi.org/10.1038/nature11574

Pörtner, H.-O., Roberts, D. C., Poloczanska, E. S., Mintenbeck, K., Tignor, M., Alegría, A., & Oken, A. (2022). Summary for policymakers. In H.-O. Pörtner, et al. (Eds.), *Climate change 2022: Impacts, adaption and vulnerability. Contribution of working group II to the sixth assessment report of the Intergovernmental Panel on Climate Change* (pp. 3–33). Cambridge University Press. https://doi.org/10.1017/9781009325844.001

Press, W. H., Teukolsky, S. A., Vetterling, W. T., & Flannery, B. P. (2007). *Numerical recipes 3rd edition: The art of scientific computing*. Cambridge University Press.

Röhl, U., Westerhold, T., Bralower, T. J., & Zachos, J. C. (2007). On the duration of the Paleocene-Eocene thermal maximum (PETM). *Geochemistry, Geophysics, Geosystems*, *8*(12), Q12002. https://doi.org/10.1029/2007GC001784

Sambridge, M. (2016). Reconstructing time series and their uncertainty from observations with universal noise. *Journal of Geophysical Research: Solid Earth*, *121*(7), 4990–5012. https://doi.org/10.1002/2016JB012901

Sherwood, S. C., Webb, M. J., Annan, J. D., Armour, K. C., Forster, P. M., Hargreaves, J. C., et al. (2020). An assessment of Earth's climate sensitivity using multiple lines of evidence. *Reviews of Geophysics*, *58*(4), e2019RG000678. https://doi.org/10.1029/2019RG000678

Snyder, C. W. (2016). Evolution of global temperature over the past two million years. *Nature*, *538*(7624), 226–228. https://doi.org/10.1038/nature19798

Taylor, J. R. (1997). *An introduction to error analysis: The study of uncertainties in physical measurements*. University Science Books.

The Cenozoic CO2 Proxy Integration Project (CenCO2PIP) Consortium. (2023). Toward a Cenozoic history of atmospheric CO_2. *Science*, *382*, eadi5177. https://doi.org/10.1126/science.adi5177

Virtanen, P., Oliphant, R. G. T. E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., et al. (2020). SciPy 1.0: Fundamental algorithms for scientific computing in Python. *Nature Methods*, *17*(3), 261–272. https://doi.org/10.1038/s41592-019-0686-2

von der Heydt, A. S., Dijkstra, H. A., van de Wal, R. S., Caballero, R., Crucifix, M., Foster, G. L., et al. (2016). Lessons on climate sensitivity from past climate changes. *Current Climate Change Reports*, *2*(4), 148–158. https://doi.org/10.1007/s40641-016-0049-3

Warton, D. I., wright, I. J., Falster, D. S., & Westoby, M. (2006). Bivariate line-fitting methods for allometry. *Biological Reviews*, *81*(2), 259–291. https://doi.org/10.1017/S1464793106007007

Westerhold, T., Marwan, N., Drury, A. J., Liebrand, D., Agnini, C., Anagnostou, E., et al. (2020). An astronomically dated record of Earth's climate and its predictability over the last 66 million years. *Science*, *369*(6509), 1383–1387. https://doi.org/10.1126/science.aba6853