



The Measurement Error Proxy System Model: MEPSM v0.2

Matt J. Fischer

Environment Research & Technology Group, ANSTO, Lucas Heights, Australia

Correspondence: Matt J. Fischer (mjf@ansto.gov.au)

Received: 3 April 2023 – Discussion started: 6 April 2023

Revised: 6 July 2024 – Accepted: 15 July 2024 – Published: 12 September 2024

Abstract. Proxy system models (PSMs) are an essential component of paleoclimate data assimilation and for testing climate field reconstruction methods. Generally, current statistical PSMs consider the noise in the output (proxy) variable only and ignore the noise in the input (environmental) variables. This problem is exacerbated when there are several input variables. Here we develop a new PSM, the Measurement Error Proxy System Model (MEPSM), which includes noise in all variables, including noise auto- and cross-correlation. The MEPSM is calibrated using a quasi-Bayesian solution, which leverages Gaussian conjugacy to produce a fast solution. Another advantage of MEPSM is that the prior can be used to stabilize the solution between an informative prior (e.g., with a non-zero mean) and the maximum likelihood solution. MEPSM is illustrated by calibrating a proxy model for $\delta^{18}\text{O}_{\text{coral}}$ with multiple inputs (marine temperature and salinity), including noise in all variables. MEPSM is applicable to many different climate proxies and will improve our understanding of the effects of predictor noise on PSMs, data assimilation, and climate reconstruction.

1 Introduction

Proxy system models (PSMs) describe how biological, geological, or chemical archives are imprinted with environmental signals (Evans et al., 2013; Dee, 2015). They are an essential component of paleoclimate data assimilation (Steiger et al., 2018; Tardif et al., 2019; Sanchez et al., 2021; King et al., 2023) and of pseudoproxy experiments, e.g., testing the fidelity of climate reconstruction methods (Loope et al., 2020a, b). For PSMs, input variables include one or more environmental variables (e.g., temperature, water salinity, rainfall), and output variables include variables which can be

read from natural archives, e.g., the abundance of trace elements or isotopes in carbonate archives. The processes between input and output may be linear or nonlinear fitted relationships (statistical PSMs) or more detailed physiochemical models (physiochemical PSMs). Statistical PSMs can be frequentist (i.e., the PSM is calibrated using ordinary least squares, OLS) or Bayesian (PSMs that make use of Bayes' rule). Current paleoclimate data assimilation projects often use frequentist PSMs (e.g., Steiger et al., 2018; Tardif et al., 2019) (see also example 1 in King et al., 2023), but some paleodata assimilation projects have begun to incorporate Bayesian PSMs (King et al., 2023). Current Bayesian PSMs (e.g., Tierney and Tingley, 2014, 2018; Tierney et al., 2019; Malevich et al., 2019) typically have the following form:

$$\zeta \sim \mathcal{N}(b_0 + f(\mathbf{x}, \theta), \sigma_\zeta^2), \quad (1)$$

where ζ is sampled from the normal distribution $\mathcal{N}(\mu, \sigma^2)$, where b_0 is the regression intercept, \mathbf{x} and ζ are the input and output variables respectively (note \mathbf{x} is a row vector of predictor matrix \mathbf{X}), and θ is a set of parameters. These Bayesian models are based on Bayesian OLS rather than Bayesian TLS (total least squares). In comparison, an errors-in-variables (EIV) model (containing error in both \mathbf{x} and y) may have the following form:

$$y \sim \mathcal{N}(b_0 + f(\mathbf{x}, \mathbf{B}^*), df(\mathbf{x}, \mathbf{B}^*)' \Sigma_{aa} df(\mathbf{x}, \mathbf{B}^*) + \varepsilon_y^2), \quad (2)$$

where b_0 is the regression intercept, $df(\mathbf{x}, \mathbf{B}^*)$ is the derivative of $f(\mathbf{x}, \mathbf{B}^*)$ for each observation vector $[y \ \mathbf{x}]$, and the prime symbol ($'$) is the transpose operator. Appendix A contains further information about Eq. (2). The covariance matrix Σ_{aa} is the covariance of the noise \mathbf{a} associated with $\mathbf{x} = \mathbf{x}^* + \mathbf{a}$, where \mathbf{x}^* denotes the true but unobservable part of \mathbf{x} (so \mathbf{x}^* is unobservable because it cannot be measured without noise). See Table 1 for notation information (in-

cluding \mathbf{B}^* explained under \mathbf{B}). The term measurement error is used here generally to describe any uncertainty associated with the value of the predictors and the response variable, which could include both measurement uncertainty (e.g., from a thermometer) and methodological uncertainty (from the method used to “combine” several measurements, which may be spatially and temporally apart). The model in Eq. (2) is not identifiable, meaning the parameters cannot be uniquely determined without additional information, such as information about Σ_{aa} . A full Bayesian solution to the problem may also require reformulating the model in terms of a subspace of $[\mathbf{Y} \ \mathbf{X}]$; see, e.g., Florens et al. (1974). The focus of this paper is on a quasi-Bayesian solution to Eq. (2), which leverages Gaussian conjugacy instead of a Markov chain Monte Carlo (MCMC) solution.

In this paper, EIV and WLS (weighted least squares) refer to general approaches. York (1966, 1968) introduced a root-finding solution for one-predictor WLS regression. Ludwig and Titterton (1994) presented a maximum likelihood (ML) solution for a straight line in 3-dimensional space, with heterogenous noise in all variables. Schneider (2001) briefly discussed the multivariate method of total least squares, which assumes homogenous and identically distributed noise for both predictor and response fields. Hannart et al. (2014) introduced a ML algorithm for weighted TLS (WTLS) that accounts for heterogeneity and temporal autocorrelation in the predictor noise and the response noise but lacks noise cross-correlation. Also ML solutions, by their very nature, do not incorporate prior information about the regression coefficients. This study expands the model of Hannart et al. (2014), to include both prior information and generalized noise (auto- and cross-correlation within and between the noise series), in the Measurement Error Proxy System Model (MEPSM). The general EIV model is discussed in Sect. 2.1, MEPSM is formulated in Sect. 2.2, and steps for practical implementation are given in Sect. 2.3. MEPSM is applied to a real example in Sect. 3. Further, Sect. 3.3 investigates the differences that PSM calibration using MEPSM can make to paleodata assimilation.

2 Methods

2.1 The EIV model

The general EIV model can be expressed as

$$\mathbf{z}_t = \mathbf{z}_t^* + \mathbf{q}_t \tag{3a}$$

$$\mathbf{z}_t^* = [y_t^* \ \mathbf{x}_t^*] \tag{3b}$$

$$\mathbf{q}_t = [v_t \ \mathbf{a}_t] \tag{3c}$$

$$y_t^* = b_0 + \mathbf{x}_t^* \mathbf{B}^* + \epsilon_t. \tag{3d}$$

Note that, for \mathbf{X}^* and \mathbf{B}^* , the superscript star also has a second use: it denotes that these entities do not include an intercept term, e.g., $\mathbf{B} = [b_0 \ \mathbf{B}^*]'$. From Eq. (3), the relevant

covariance matrices are

$$\Sigma_{z^*z^*} = \begin{bmatrix} \sigma_{Y^*Y^*}^2 & \Sigma_{Y^*X^*} \\ \Sigma_{X^*Y^*} & \Sigma_{X^*X^*} \end{bmatrix}, \Sigma_Q = \begin{bmatrix} \Omega_V & \mathbf{0} \\ \mathbf{0} & \Sigma_A \end{bmatrix}, \tag{4}$$

and

$$\Sigma_A = \begin{bmatrix} \Omega_{A_1} & \Omega_{A_1A_2} & \dots & \Omega_{A_1A_p} \\ \Omega_{A_2A_1} & \Omega_{A_2} & \dots & \Omega_{A_2A_p} \\ \vdots & \vdots & \ddots & \vdots \\ \Omega_{A_pA_1} & \Omega_{A_pA_2} & \dots & \Omega_{A_p} \end{bmatrix}. \tag{5}$$

See Appendix B for further information on covariance notation. These matrices describe a statistical stationary model. For example, the covariance matrix Σ_A fully specifies the lag autocovariance and lag cross-covariances of the noise variables in \mathbf{A} , but $(\Sigma_A)_t = \Sigma_A$. Note this does not mean within Σ_A that $\Sigma_{a_i a_i} = \Sigma_{a_j a_j}$ or $\Omega_{A_i} = \Omega_{A_j}$ (those are specific cases of Σ_A). Also note that in this paper $\Omega_V = \Sigma_V$. More will be said about Σ_Q in Sect. 3.1.

Table 2 summarizes the main differences between previous solutions of the EIV model. First, these methods all assume that the equation noise ϵ_t (Eq. 3d) is either zero or not separate from v_t . Earlier work by Fuller (1987) shows a way of including the equation noise separate from the response noise, but the predictor noise is limited to the $p \times p$ matrices $\Sigma_{a_i a_i}$. The solutions in Table 2 relate to generalized least squares (GLS), weighted total least squares (WTLS), maximum likelihood (ML), and maximum a posteriori (MAP) methods. A general estimator is of the following form:

$$\Sigma_{ZZ} = (\mathbf{Z} - \mu_Z)' \mathbb{J}_n' \begin{bmatrix} \Sigma_Y & \Sigma_{YX} \\ \Sigma_{XY} & \Sigma_X \end{bmatrix}^{-1} \mathbb{J}_n (\mathbf{Z} - \mu_Z) \tag{6a}$$

$$\Sigma_{Z^*Z^*} = \Sigma_{ZZ} - \Sigma_{QQ} = \begin{bmatrix} \sigma_{Y^*Y^*}^2 & \Sigma_{Y^*X^*} \\ \Sigma_{X^*Y^*} & \Sigma_{X^*X^*} \end{bmatrix} \tag{6b}$$

$$\mathbf{B}^* = \Sigma_{X^*X^*}^{-1} \Sigma_{X^*Y^*} \tag{6c}$$

$$b_0 = \mu_Z \cdot [1 \ -\mathbf{B}^*]', \tag{6d}$$

where μ_Z is the GLS mean, $\Sigma_Y = \Sigma_{Y^*} + \Omega_V$, and $\Sigma_X = \Sigma_{X^*} + \Sigma_A$. The covariance matrix $\Sigma_{X^*X^*}$ is a weighted matrix, e.g., $\mathbf{X}^* \Sigma^{-1} \mathbf{X}^*$, but different ways of solving for \mathbf{B} allow for different weightings. The assumption is then made that $\Sigma_{B^*} = \Sigma_{X^*X^*}^{-1}$ (e.g., Amiri-Simkooei and Jazaeri, 2012; Fang et al., 2017), despite the fact that different possible weightings exist. The ML method can be used to obtain a theoretically better estimation of Σ_B , but it is computationally expensive if Σ_A is not constrained. For example, in Hannart et al. (2014) Σ_A is a block diagonal matrix: $\Sigma_A = \oplus_{i=1}^p \Omega_{A_i}$, where \oplus is the direct sum operator for matrices (see Table 2 footnotes). The aim of the present study is to provide a solution for the EIV model that includes a fast theory-derived approximation of Σ_B , a prior for Σ_B , and a full Σ_A , including a way to estimate Σ_A with limited initial information.

Table 1. Notation used.

Symbol	Definition
$\mathbf{0}_n, \mathbf{1}_n$	n -length vectors of zeros or ones (or $\mathbf{0}$ for undefined size)
\mathbf{A}	$n \times p$ matrix of predictor noise, p predictors with n time points, $\mathbf{A} = \mathbf{X} - \mathbf{X}^*$
\mathbf{a} or \mathbf{a}_t	$1 \times p$ vector, a row of \mathbf{A}
\mathbf{B}	$(p + 1)$ -length vector of regression weights, $\mathbf{B}' = [b_0 \ \mathbf{B}^{*'}]$
$d\mathbf{B}$	$\frac{\partial}{\partial \varepsilon_\theta} \varepsilon_B$, where, e.g., ε_θ is the noise associated with parameters θ
\mathbf{e}	n -length vector of noise, $\mathbf{Y} - \mathbf{X}\mathbf{B}$
ϵ	Equation noise
\mathbb{I}_d	Identity matrix of rank d
\mathbb{J}_n	$[\mathbb{I}_n \ \mathbf{0}_{n \times np}]'$
$L(\cdot), \ell(\cdot), p(\cdot)$	Likelihood, log-likelihood, and probability density function
$\mathcal{N}(\mu, \Sigma)$	Normal distribution with mean μ and covariance Σ
\mathbf{Q}	$n \times (p + 1)$ matrix of noise for all variables, $\mathbf{Q} = [\mathbf{V} \ \mathbf{A}]$
\mathbf{R}	$n \times n$ auto- or cross-correlation matrix
\mathbf{V}	n -length vector of response noise, $\mathbf{V} = \mathbf{Y} - \mathbf{Y}^*$
\mathbf{W}	$n \times n$ weight matrix
\mathbf{X}	$n \times (p \text{ or } p + 1)$ matrix, p predictors with n time points, may include an intercept term
\mathbf{X}^*	$n \times p$ matrix of unobserved values underlying \mathbf{X} , no intercept term
\mathbf{Y}	n -length response vector
\mathbf{Z}	$n \times (p + 1)$ matrix, $\mathbf{Z} = [\mathbf{Y} \ \mathbf{X}]$
Σ	General covariance matrix
Σ_0	Prior covariance matrix for Σ_B
Σ_A	$np \times np$ covariance matrix of $\text{vec}(\mathbf{A})^a$
Σ_{aa}	$p \times p$ covariance matrix of noise vector \mathbf{a}
Σ_{ZZ}	$(p + 1) \times (p + 1)$ covariance matrix of \mathbf{Z}
Ω	$n \times n$ general lag-covariance matrix of noise
Ω_{A_i}	$n \times n$ lag-covariance matrix of error variable \mathbf{A}_i
$\Omega_{A_i A_j}$	$n \times n$ lag cross-covariance matrix of error variables \mathbf{A}_i and \mathbf{A}_j
$\Psi_{zz}^{(i)}$	$\Sigma_{z_i z_i}^{-1}$ is a $(p + 1) \times (p + 1)$ inverse covariance matrix ^b
Φ	$p^2 \times (\frac{p}{2}(p + 1))$ duplication matrix ^c
θ, Θ	Generic parameter vector and parameter space

^a The operator $\text{vec}(\mathbf{A})$ means to column stack the matrix \mathbf{A} (Henderson and Searle, 1979). ^b Superscript (i) refers to an index, rather than subscripting (e.g., $z_i z_i$). ^c For the duplication matrix, see Magnus and Neudecker (1986).

Table 2. Solutions for the EIV model. Σ_A and Σ_B are given in my notation.

Study	Σ_A	Σ_B	Sol. method
Fuller (1990)	$\oplus_{i=1}^n \Sigma_{a_i a_i}$ ^{a,b}	$n \cdot d\mathbf{B}' \left[\begin{matrix} \Sigma \Psi_{zz}^{(i)} & \mathbf{0} \\ \mathbf{0} & \frac{1}{2} \sum \Phi' (\Psi_{zz}^{(i)} \otimes \Psi_{zz}^{(i)}) \Phi \end{matrix} \right]^{-1} d\mathbf{B}$	GLS
Amiri-Simkooei and Jazaeri (2012)	Full ^c	$\frac{\mathbf{e}' \mathbf{W}^{-1} \mathbf{e}}{n-p} (\mathbf{X}^{*'} \mathbf{W}^{-1} \mathbf{X}^*)^{-1}$	WTLS
Hannart et al. (2014)	$\oplus_{i=1}^p \Omega_{A_i}$	$\max_{\theta_{-i} \in \Theta_{-i} \theta_i} \ell(\theta_i, \theta_{-i})$	ML
Fang et al. (2017)	Full	$(\mathbf{X}^{*'} \Omega_V^{-1} \mathbf{X}^* + \Sigma_0^{-1})^{-1}$	MAP
This study	Full, can also be estimated	$(\Sigma_B^{\text{wtls}}{}^{-1} + \Sigma_0^{-1})^{-1}$	MEPSM

^a The matrix direct sum operator \oplus is notation to build a block diagonal matrix, e.g., $\text{Diag}(\{\Sigma_{a_1 a_1}, \dots, \Sigma_{a_n a_n}\})$. ^b Note though that technically $\oplus_i \Sigma_{a_i a_i} = \Sigma_{A'}$, not Σ_A , so \oplus is used abstractly in this case. ^c Full: Σ_A is not constrained to a particular type.

2.2 MEPSM

As in previous studies, I also ignore the equation error, so the model in Eq. (3) can be re-expressed:

$$\begin{bmatrix} Z_{11} \\ Z_{21} \\ \vdots \\ Z_{n1} \end{bmatrix} = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix} \begin{bmatrix} b_0 \\ \mathbf{B}^* \end{bmatrix} + [\mathbb{I}_n \quad -(\mathbf{B}^{*'} \otimes \mathbb{I}_n)] \begin{bmatrix} \mathbf{V} \\ \text{vec}(\mathbf{A}) \end{bmatrix}, \tag{7}$$

where $\mathbf{V} \sim \mathcal{N}(\mathbf{0}_n, \Omega_V)$, and $\text{vec}(\mathbf{A}) \sim \mathcal{N}(\mathbf{0}_{n \cdot p}, \Sigma_A)$ (see Table 1 footnotes for the operator vec). The response variable is the first column of \mathbf{Z} , denoted as \mathbf{Z}_1 or more generally as \mathbf{Y} . Note also the two possible definitions of response noise for the model: $\mathbf{V} = \mathbf{Y} - b_0 - \mathbf{X}^* \mathbf{B}^*$ and $\mathbf{e} = \mathbf{Y} - \mathbf{X} \mathbf{B}$ (where \mathbf{X} includes an intercept term, to match with \mathbf{B}). The difference is that

$$\mathbf{V} = \mathbf{Y} - \mathbf{X} \mathbf{B} + \mathbf{A} \mathbf{B}^* = \mathbf{e} + \mathbf{A} \mathbf{B}^*. \tag{8}$$

From $\mathbf{e} = \mathbf{V} - \mathbf{A} \mathbf{B}^*$, the covariance matrix of model prediction can be expressed as

$$\mathbf{W} = \Omega_V + (\mathbf{B}^{*'} \otimes \mathbb{I}_n) \Sigma_A (\mathbf{B}^* \otimes \mathbb{I}_n). \tag{9}$$

Bayes' rule states that the posterior density function $p(\boldsymbol{\theta}|\mathcal{D})$ is proportional to the product of the likelihood function $L(\boldsymbol{\theta})$ and the prior density function $p(\boldsymbol{\theta})$:

$$p(\boldsymbol{\theta}|\mathcal{D}) \propto L(\boldsymbol{\theta}) \times p(\boldsymbol{\theta}). \tag{10}$$

Note that it does not matter if the likelihood function is a scaled distribution, e.g., a normal density function multiplied by a scalar (so it integrates to a number other than unity).

Assuming that the right-side distributions are Gaussian (or scaled Gaussian), then

$$\mathcal{N}(\mathbf{B}, \Sigma_B) \propto \mathcal{N}(\mathbf{B}_{ml}, \Sigma_B^{ml}) \times \mathcal{N}(\mathbf{B}_0, \Sigma_0), \tag{11}$$

where ml refers to the maximum likelihood solution. The posterior $\mathcal{N}(\mathbf{B}, \Sigma_B)$ integrates to unity, as expected. First I show an example by using the GLS likelihood. Let the GLS likelihood be

$$\mathcal{N}(\mathbf{B}_{ml}, \Sigma_B^{ml}) = \mathcal{N}\left((\mathbf{X}' \Sigma^{-1} \mathbf{X})^{-1} \mathbf{X}' \Sigma^{-1} \mathbf{Y}, (\mathbf{X}' \Sigma^{-1} \mathbf{X})^{-1}\right). \tag{12}$$

Now apply the Gaussian multiplication identity:

$$\begin{aligned} \mathcal{N}(\mathbf{C}(\Sigma_B^{ml-1} \mathbf{B}_{ml} + \Sigma_0^{-1} \mathbf{B}_0), \mathbf{C}) &\propto \mathcal{N}(\mathbf{B}_{ml}, \Sigma_B^{ml}) \\ &\times \mathcal{N}(\mathbf{B}_0, \Sigma_0), \end{aligned} \tag{13}$$

where $\mathbf{C} = \left(\Sigma_B^{ml-1} + \Sigma_0^{-1}\right)^{-1}$. Thus the mean and variance of $p(\boldsymbol{\theta}|\mathcal{D})$ are

$$\mathbf{B} = (\mathbf{X}' \Sigma^{-1} \mathbf{X} + \Sigma_0^{-1})^{-1} (\mathbf{X}' \Sigma^{-1} \mathbf{Y} + \Sigma_0^{-1} \mathbf{B}_0) \tag{14a}$$

$$\Sigma_B = (\mathbf{X}' \Sigma^{-1} \mathbf{X} + \Sigma_0^{-1})^{-1}. \tag{14b}$$

The above solution is a simple example using the GLS likelihood. To obtain the solution for the EIV problem,

the GLS likelihood is replaced with the WTLS likelihood $\mathcal{N}(\mathbf{B}_{wtls}, \Sigma_B^{wtls})$, where \mathbf{B}_{wtls} and Σ_B^{wtls} are derived in Appendix C and D. Using the WTLS likelihood, simplifying the product $\Sigma_B^{ml-1} \mathbf{B}_{ml}$ (from Eq. 13) is no longer feasible, so instead we write the mean and variance of $p(\boldsymbol{\theta}|\mathcal{D})$ simply as

$$\mathbf{B} = \left(\Sigma_B^{wtls-1} + \Sigma_0^{-1}\right)^{-1} \left(\Sigma_B^{wtls-1} \mathbf{B}_{wtls} + \Sigma_0^{-1} \mathbf{B}_0\right) \tag{15}$$

$$\Sigma_B = \left(\Sigma_B^{wtls-1} + \Sigma_0^{-1}\right)^{-1}. \tag{16}$$

2.3 Implementation

An application to calibrate a real PSM is provided in Sect. 3. Here general aspects of the solution and implementation are discussed.

The predictor noise and response noise, \mathbf{A} and \mathbf{V} , can be estimated as

$$\mathbf{A} = -(\mathbf{B}^{*'} \otimes \mathbb{I}_n) \Sigma_A (\mathbb{I}_p \otimes \mathbf{W}^{-1} \mathbf{e}) \tag{17}$$

$$\mathbf{V} = \Omega_V \mathbf{W}^{-1} \mathbf{e}. \tag{18}$$

Matrix \mathbf{A} is needed for the calculation of both \mathbf{B}_{wtls} and Σ_B^{wtls} , while \mathbf{V} is needed to re-estimate Ω_V if relevant (see next paragraph).

The $n \times n$ covariance matrices Ω_{A_i} (or $\Omega_{A_i A_j}$) and Ω_V ideally need information about the cross-correlation within \mathbf{A} and autocorrelation of \mathbf{A} and \mathbf{V} . If the ACF (autocorrelation function) and CCF (cross-correlation function) are not available, these can potentially be estimated using an iterative procedure. The procedure begins with vectors σ_A^2 and σ_V^2 , which are the heteroscedastic error variances (the diagonals of Σ_A and Ω_V).

1. Initialize $\Sigma_A = \text{Diag}(\sigma_A \odot \sigma_A)$ and $\Omega_V = \text{Diag}(\sigma_V \odot \sigma_V)$, where \odot denotes the Hadamard product.
2. Estimate the first posterior, $\mathcal{N}(\mathbf{B}, \Sigma_B)$, and the predictor and response noise, \mathbf{A} and \mathbf{V} (Eqs. 15–18).
3. Calculate the ACF (for \mathbf{A} and \mathbf{V}) and the CCF (for \mathbf{A}).
4. Estimate Σ_A and Ω_V as in Appendix E, and recalculate the posterior distribution.

3 Application

3.1 A proxy system model, uncertainties, and the prior

A bivariate (i.e., two-predictor) PSM for coral $\delta^{18}\text{O}$ is

$$\delta^{18}\text{O}_{\text{coral}} = b_0 + b_1 \text{SST} + b_2 \text{SSS}, \tag{19}$$

where SST is sea surface temperature and SSS is sea surface salinity. Note that in the coral literature, the coefficients b_1 and b_2 are commonly referred to as a_1 and a_2 , but here

I use the former symbols for generality. This bivariate PSM has been used in data assimilation (Sanchez et al., 2021) and Monte Carlo experiments (Thompson et al., 2022). The bivariate PSM can be extended by rewriting it in the form of Eq. (7). The MEPSM requires estimates of the time-varying uncertainty of all variables and a prior distribution for \mathbf{B} (see below). Here the time-varying uncertainties for the predictors are obtained from recent SST and SSS products that provide the complete uncertainty field $\varepsilon^2(\text{long, lat, } t)$, i.e., the uncertainty variance for each grid cell, at each time point.

For ERSSTv5, for each grid cell and time point, the SST total uncertainty $\varepsilon_T^2(\text{long, lat, } t)$ is the sum of the parametric uncertainty and reconstruction uncertainty. These uncertainties are derived from a large reconstruction ensemble (Huang et al., 2020). The parametric uncertainty is based on the difference between the ensemble average and each ensemble member, given various internal parameters which affect SST uncertainty (including the uncertainty of the ship, buoy, or float measurement). The reconstruction uncertainty is based on the difference between each ensemble member and a pseudo-observation dataset on which the reconstruction method is trained, e.g., OISST (optimally interpolated SST). Note that the calculation of reconstruction uncertainty includes grid cells that have observational or no observational data, while the calculation of parametric uncertainty includes only grid cells that have observational data.

For HadEN4, the SSS total uncertainty $\varepsilon_S^2(\text{long, lat, } t)$ is a combination of background uncertainty and observational uncertainty (Good et al., 2013). The observational uncertainty is obtained from previous studies, while the background uncertainty is the uncertainty associated with a persistent-based forecast.

For $\delta^{18}\text{O}_{\text{coral}}$, the main known uncertainty is the analytical uncertainty, which is $\sim 0.05\text{‰}$ for $\delta^{18}\text{O}_{\text{coral}}$ (Osborne et al., 2013). Time variation could be included by expressing the analytical uncertainty as a relative standard deviation of the mean $\delta^{18}\text{O}$ (the effect of this will be tested elsewhere). Other uncertainties, which are unknown at many sites, include the intra- and inter-colony noise (Sayani et al., 2019). Note also that detrending adds time-varying uncertainty, because the trend uncertainty is larger at the ends of the time series (this is applicable to all detrended variables).

For MEPSM, the matrix Σ_A can be initially constructed using $\text{Diag}(\sigma_A \odot \sigma_A)$, where $\sigma'_A = [\varepsilon_T(\text{site}, \cdot), \varepsilon_S(\text{site}, \cdot)]$, “site” refers to the grid cell of a particular site, and the (site, \cdot) notation means over all time points. In the SST and SSS products, no information is given on the noise serial correlation, and no information exists on the noise cross-correlation between SST and SSS products. Seasonal peaks in the autocorrelation function (ACF) of SST noise could arise from, for example, seasonal changes in shipping tracks, which would affect the parametric uncertainty. Cross-correlation between SST and SSS noise might arise from sampling these variables at the same points within a particular grid cell (i.e., from the same ship) or due

to seasonal variation in shipping tracks. Future work should investigate the ACF and CCF (cross-correlation function) for the SST and SSS ensembles (this is beyond the scope of the current paper). Finally, since the uncertainties in the predictors and response are fundamentally different, then Σ_Q is assumed to be a block diagonal matrix, with Ω_V and Σ_A on the block diagonal and zero-filled blocks on the block anti-diagonal (Eq. 4).

For Bayesian methods, informative prior distributions are thought to be more useful than noninformative priors (Lemoine, 2019). For example, when using a fully noninformative prior distribution, the posterior distribution will be close to the maximum likelihood distribution, which would make using the fully noninformative prior somewhat obsolete.

The prior distribution for \mathbf{B} is obtained from previous values that were used in Monte Carlo PSM experiments (Thompson et al., 2022; Watanbe and Pfeiffer, 2022). Those previous studies did not seek to calculate Bayesian posterior distributions, but the values in those studies represent current beliefs about the coefficients and their uncertainty and can be used as a prior. The prior distribution obtained from those studies is

$$p(\mathbf{B}) = \mathcal{N}(\mathbf{B}_0, \Sigma_0) = \mathcal{N}\left(\begin{bmatrix} b_0 \\ -0.22 \\ 0.27s \end{bmatrix}, \begin{bmatrix} \sigma_{b_0}^2 & 0 & 0 \\ 0 & 0.02^2 & 0 \\ 0 & 0 & 0.15^2 \end{bmatrix}\right), \tag{20}$$

where $s = 0.97$ is a scaling factor between $\delta^{18}\text{O}_{\text{seawater}}$ (in VSMOW) and $\delta^{18}\text{O}_{\text{coral}}$ (in VPDB). The values of $b_1 = -0.22$ and $\sigma_{b_1}^2 = 0.02^2$ are from Thompson et al. (2022) and Watanbe and Pfeiffer (2022) respectively. The value of $s^{-1}b_2 = 0.27\text{‰}$ psu^{-1} originates from LeGrande and Schmidt (2006) and is an average ‰ per salinity value for the tropical Pacific Ocean. Regional variations in this value will probably be revealed by new datasets (DeLong et al., 2022), but for the purpose of this example of MEPSM, the value of 0.27‰ psu^{-1} is used. Thompson et al. (2022) considered two values for $\sigma_{b_2}^2$ (i.e., $\sigma_{b_2}^2 = \{0.1^2, 0.2^2\}$), so for this example an “average” value is used (0.15^2). The value of the intercept b_0 and its uncertainty $\sigma_{b_0}^2$ are unknown in the sense that they are site-dependent values. A prior for these unknown values can be estimated by using the other prior coefficients and SST and SSS data for a particular site. For a particular site, the prior value of b_0 can be calculated as

$$b_0 = \mu_y - \mu_X \begin{bmatrix} -0.22 \\ 0.27s \end{bmatrix}, \tag{21}$$

where $\mu_X = [\mu_{X_1} \ \mu_{X_2}]$ and μ_y are the mean values of the predictors and response. From ordinary least squares theory, the diagonal elements of the covariance of \mathbf{B} are typically

given as

$$\sigma_{b_j}^2 = \sigma_e^2 / \text{RSS}_j, \tag{22}$$

where σ_e^2 is the error variance, and RSS_j is the residual sum of squares after regressing \mathbf{X}_j on \mathbf{X}_{-j} , where \mathbf{X}_{-j} is the matrix \mathbf{X} omitting column j . So for $\sigma_{b_0}^2$,

$$\text{RSS}_0 = n - n^2 \mu'_X (\mathbf{X}'_0 \mathbf{X}_0)^{-1} \mu_X, \tag{23}$$

where \mathbf{X}_0 is the observed part of \mathbf{X} not including an intercept term (different than \mathbf{X}^* , which is the unobserved part of \mathbf{X}). Preliminary analysis showed that the prior value for $\sigma_{b_0}^2$ had a large influence on the posterior distribution of \mathbf{B} , so $\sigma_{b_0}^2$ only was scaled to be uninformative, i.e., $10^3 \sigma_{b_0}^2$.

3.2 Coral example

Rock Islands (7.27° N, 134.38° E) is a site on the western barrier reef of Palau, western Pacific Ocean. A number of coral cores were sampled from Rock Islands (RI) and the surrounding area, but the focus here is on one core labeled RI6, which is a monthly resolved record spanning 1899–2008. The core was sampled from a narrow-entrance lagoon, in shallow water (~2 m depth). Osborne et al. (2013) compared the coral $\delta^{18}\text{O}$ record with temperature data from the NCEP/NCAR Reanalysis (Kalnay et al., 1996) and salinity data from LEGOS (Delcroix et al., 2011). Using the raw (not detrended or deseasonalized) RI record from 1970–2008 and ordinary least squares, Osborne et al. (2013) regressed the coral $\delta^{18}\text{O}$ against surface temperature and salinity, producing the following equation:

$$\delta^{18}\text{O}_{\text{coral}} = -0.06\text{SST} + 0.43\text{SSS} - 18.70; \tag{24}$$

see Table A4 in Osborne et al. (2013). Osborne et al. (2013) attributed the small SST coefficient to the small seasonal SST range at Palau (~1.5 °C), presumably because the small SST range makes the data scatter with $\delta^{18}\text{O}$ more spherical than ellipsoid. They did not directly compare their raw SSS coefficient with any other data. Hence the RI $\delta^{18}\text{O}$ record was chosen for this example, in order to further investigate the SST and SSS effects on coral $\delta^{18}\text{O}$ at this site. In the following analysis the main differences with the analysis of Osborne et al. (2013) are that here the SST and SSS data are extracted from ERSSTv5 and HadEN4, because these products provide the uncertainty fields (Sect. 3.1). Secondly, the time period examined is from 1950–2008. Thirdly, for this study, all variables were detrended (using linear detrending and without removing the intercept), in order to ensure stationarity in the mean. For all variables, the trends at this site were weak. For example, for $\delta^{18}\text{O}_{\text{coral}}$ the trend from 1950–2008 was $-0.00135 \pm 0.004 \text{‰}$ per year. A Jupyter Notebook with these preprocessing steps is available in the Supplement. Using this updated (and detrended) data, the OLS regression was recalculated as

$$\delta^{18}\text{O}_{\text{coral}} = -0.13\text{SST} + 0.42\text{SSS} - 16.25. \tag{25}$$

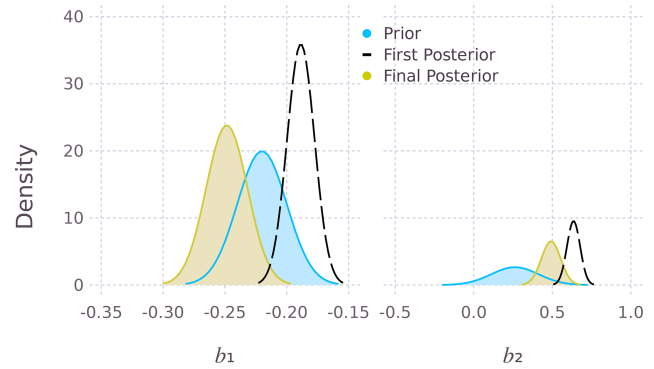


Figure 1. The prior and posterior distributions for the SST (b_1) and SSS (b_2) coefficients. The final posterior is wider than the first posterior, because the final posterior accounts for auto- and cross-correlation in Σ_A . In Figs. 1–3 the first posterior is the dashed line.

Next MEPSM was applied to the same data, following the steps in Sect. 2.3. Except, for the application presented here, only Σ_A was updated in step 4. Ω_V was set as $\Omega_V = \text{Diag}(\sigma_V \odot \sigma_V)$ throughout, because here σ_V is thought to be mostly analytical noise, which should be approximately white (the uncertainty due to detrending is time-varying as well as white). For Σ_A , for the auto- and cross-correlation functions of the predictor noise, up to 100 lags (months) were retained to construct the submatrices \mathbf{R}_{A_i} and $\mathbf{R}_{A_i A_j}$ (see Appendix E). Experiments with 50 and 150 lags showed little difference in the final posterior. In any application of MEPSM, the user should consider the source of the predictor and response noise in constructing Σ_A and Ω_V (as in Sect. 3.1).

Figure 1 shows the prior and the first and final posterior distributions for the SST and SSS coefficients. The first posterior distribution (dashed line) is obtained after step 2 (Sect. 2.3), while the final posterior is obtained after updating Σ_A for auto- and cross-correlation. The (final) mean SST and SSS coefficients are given in Fig. 2: $b_1 = -0.25 \text{‰} \text{°C}^{-1}$ and $b_2 = 0.49 \text{‰} \text{psu}^{-1}$. Figure 2 shows the marginal plots of the 3-dimensional scatter plot (SST, SSS, $\delta^{18}\text{O}$), together with the prior and posterior regression lines. For Rock Island, the SST and SSS coefficients are both steeper for the final posterior, compared to the OLS equation (Eq. 25).

Another advantage of Bayesian analysis is that attaching a ridge parameter λ to the prior variances means that the method can be used to regularize the solution, in a way that is different than traditional methods of regularization, such as principal component analysis, where components are retained by hard thresholding. Here, the ridge parameter λ is attached to the prior and allowed to vary over the range 10^{-4} to 10^2 :

$$p(\mathbf{B}) = \mathcal{N} \left(\begin{bmatrix} b_0 \\ -0.22 \\ 0.27s \end{bmatrix}, \begin{bmatrix} 10^3 \sigma_{b_0}^2 & 0 & 0 \\ 0 & \lambda 0.02^2 & 0 \\ 0 & 0 & \lambda 0.15^2 \end{bmatrix} \right). \tag{26}$$

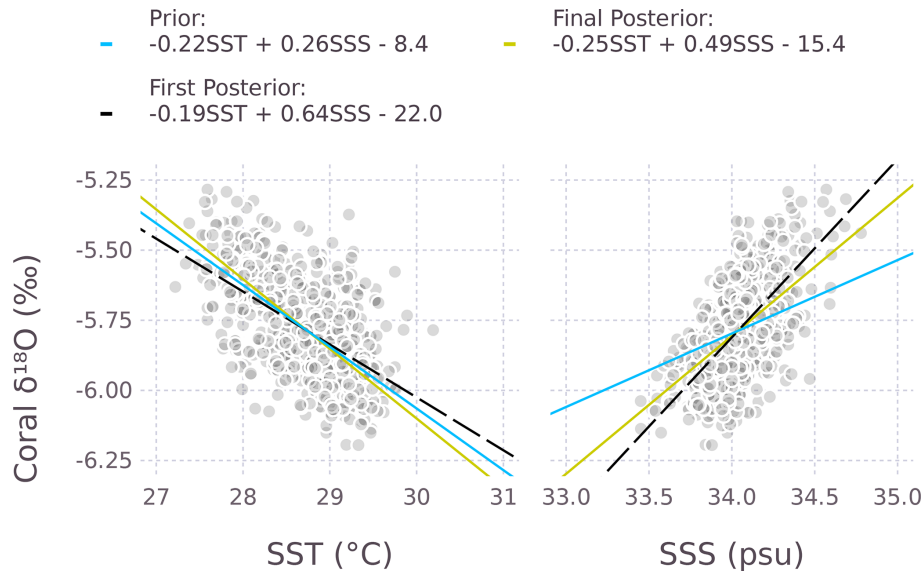


Figure 2. The regression lines of the prior and posterior distributions. In 3 dimensions, these regression lines are the sides of a plane. Each shaded point corresponds to a monthly value from 1950–2008.

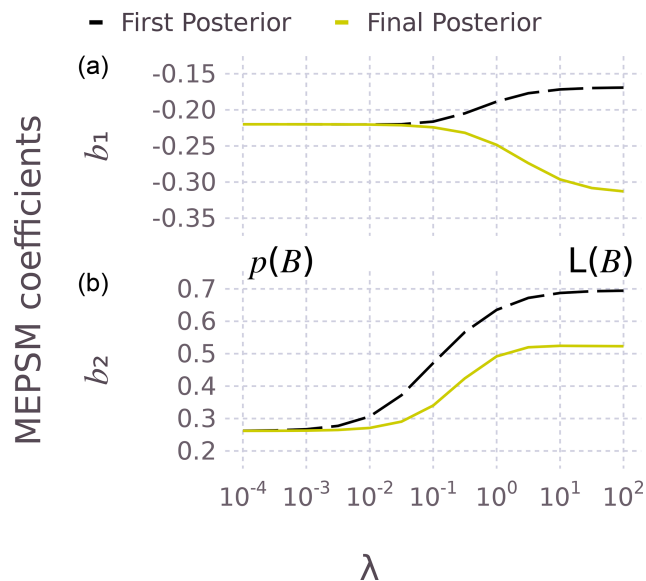


Figure 3. The effect of the regularization parameter λ on the SST (b_1 , top) and SSS (b_2 , bottom) coefficients. The left and right side of the plots correspond to the prior distribution $p(\mathbf{B})$ and likelihood distribution $L(\mathbf{B})$ respectively.

Figure 3 shows the effect of λ for both the first and final posterior. As λ increases, \mathbf{B} moves from the prior distribution to the likelihood distribution. The difference between the solid and dashed lines (for each coefficient) is because of the effect of adjusting Σ_A for auto- and cross-correlation, which becomes less relevant for small λ ($\lambda < 10^{-2}$). More will be said about serial dependence in \mathbf{A} below. This type of prior regularization will likely be useful for sites where the $(p+1)$ -

dimensional scatter of data (predictors plus response) is more spherical than ellipsoid. For example, a more spherical scatter of data could occur at sites where the SST range is small. In practice, the value of λ could be estimated using cross-validation methods.

Table 3 compares the regression coefficients (and variance) from four different solutions: a ML solution (Appendix F), the WTLS method (Appendix C and D), and the MEPSM solution for the first and final posterior. The maximum likelihood estimate (MLE) and WTLS solution are basic implementations, i.e., no prior and no adjustment for autocorrelation or cross-correlation. The MLE and WTLS solutions are similar, with the variance being generally more conservative for WTLS, owing either to assumptions in the two methods or differences in method implementation, e.g., optimization for MLE versus analytical solution (with Gaussian assumptions) for WTLS. For the first posterior of MEPSM, both the mean and the variance move away from the WTLS solution and towards a tight prior: in Fig. 3 the MEPSM first posterior (on the dashed line) is the same as the $\lambda = 10^0$ solution, and the basic WTLS solution is the same as $\lambda \geq 10^2$. For the final posterior (Table 3), the SST and SSS coefficients change steepness: b_1 becomes steeper, while b_2 becomes a little less steep compared to the WTLS solution and the first posterior (also seen in Fig. 3 as the difference between the solid and dashed lines). For both coefficients (Table 3), the variance increases for the final posterior (relative to the first posterior), because the final posterior adjusts for the auto- and cross-correlation in Σ_A .

Figure 4 shows the autocorrelation of the noise in each variable (\mathbf{A}_1 is the SST noise, and \mathbf{A}_2 is the SSS noise), for the first and final posterior distribution. For the first posterior

Table 3. Regression parameters and their standard deviation, from four different solutions.

Method	b_0	b_1	b_2	Equations
MLE	-24.6 ± 1.2	-0.169 ± 0.014	0.696 ± 0.027	Eq. (F3)
WTLS	-24.6 ± 1.8	-0.169 ± 0.014	0.696 ± 0.046	Eq. (C8) & (D3)
First posterior	-22.0 ± 1.6	-0.189 ± 0.011	0.635 ± 0.042	Eqs. (15) & (16)
Final posterior	-15.4 ± 2.2	-0.248 ± 0.017	0.492 ± 0.061	Eqs. (15) & (16)

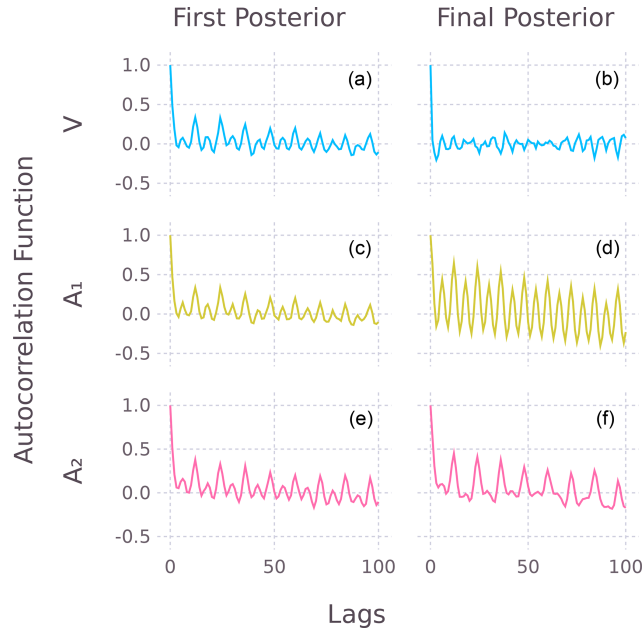


Figure 4. The autocorrelation function of the response noise (V), the SST noise A_1 , and the SSS noise A_2 , for the first posterior (left column) and the final posterior (right column). The predictor and response noise were calculated using Eqs. (17)–(18).

solution, the noise covariance in Σ_A and Ω_V is assumed to be heteroscedastic but white. This assumption is clearly not true, because all variables appear to have seasonally varying noise as shown by their autocorrelation functions (Fig. 4, left column). However, when this seasonally varying noise is included in Σ_A (but not in Ω_V), the response noise V essentially becomes white, as all the seasonal dependence moves into the predictor noise (Fig. 4, right column). Whether or not the response noise V should have seasonal variation (see Sect. 3.1) requires expert knowledge. This example is simply to illustrate the effect of including serially correlated noise in a multipredictor model. The implications of seasonally varying noise for coral PSMs will be addressed in another paper, with the application of MEPSM to a large coral database (Walter et al., 2023).

3.3 Data assimilation experiments

Two data assimilation (DA) experiments were performed using the Palau coral example. The purpose of these experi-

ments is to investigate the effect of different \mathbf{H} on data assimilation (Appendix G introduces DA and describes potential biases caused by misspecified \mathbf{H}). For the experiments, the model prior covariance $\Sigma_{\varepsilon\varepsilon}$ was obtained from iCESM1 (Brady et al., 2019), using a preindustrial run, and data from the last 400 years (1450–1850 CE, which is “closer” to the coral example period 1950–2007). Model SST and SSS data were extracted from a line of five grid cells, centered on the RI6 location. Although there is only one observational coral record, for these experiments, the same observational record is repeated for the five grid cells (\mathbf{Y}^o in DA). All data were deseasonalized but kept at monthly resolution. A line of grid cells was used, so the covariance $\Sigma_{\varepsilon\varepsilon}$ includes some spatial correlation. The reasons for using covariance matrices in these experiments will become apparent. The covariances $\Sigma_{\varepsilon\varepsilon}$ and Σ_{rr} were the same for both experiments and were constructed as follows:

- $\Sigma_{\varepsilon\varepsilon} = \text{Var}(\mathbf{X} - \tilde{\mathbf{U}}_k \mathbf{D}_k \mathbf{V}'_k)$, where $\mathbf{X} = \mathbf{U}\mathbf{D}\mathbf{V}'$ is the singular value decomposition of \mathbf{X} , $k=1:3$ (means the three leading eigencomponents), and $\tilde{\mathbf{U}}$ denotes components filtered with a bandpass filter (2–20 years). Here interannual variability (as in \mathbf{U}_k) is considered to be the state of interest. Filtering \mathbf{U} better separates signal and noise.
- $\Sigma_{rr} = \mathbf{F}\mathbf{H}\Sigma_{\varepsilon\varepsilon}\mathbf{H}'\mathbf{F}$, where $\mathbf{F} = \text{Diag}\left(\left[\frac{\sigma_V}{\sigma_{h_1}}, \dots, \frac{\sigma_V}{\sigma_{h_p}}\right]\right)$ and $\text{Diag}(\mathbf{H}\Sigma_{\varepsilon\varepsilon}\mathbf{H}') = [\sigma_{h_1}^2, \dots, \sigma_{h_p}^2]$. This means that Σ_{rr} is a covariance matrix with the same correlation structure as $\mathbf{H}\Sigma_{\varepsilon\varepsilon}\mathbf{H}'$ but with variance scaling provided by $\sigma_V^2 = \text{Var}(\mathbf{V})$, where \mathbf{V} is from Eq. (18) (for DA, $\text{Var}(\mathbf{V})$ is just the total variance of \mathbf{V}). Also here $\mathbf{H} = \mathbf{B}'_{\text{post}2} \otimes \mathbb{I}_p$, where $\mathbf{B}_{\text{post}2}$ denotes the values of \mathbf{B} for the final posterior in Table 3. In the experiments below, keeping Σ_{rr} constant simplifies the comparisons.

To understand the experiment results, it is useful to cast data assimilation into observation space:

$$\begin{aligned} \mathbf{y}^a &= \mathbf{H}\mathbf{x}^a = \mathbf{H}\mathbf{K}\mathbf{y}^o + (\mathbb{I}_p - \mathbf{H}\mathbf{K})\mathbf{H}\mathbf{x}^b \\ &= \frac{(\mathbf{H}\Sigma_{\varepsilon\varepsilon}\mathbf{H}')\mathbf{y}^o + \Sigma_{rr}\mathbf{y}^b}{(\mathbf{H}\Sigma_{\varepsilon\varepsilon}\mathbf{H}' + \Sigma_{rr})}, \end{aligned} \tag{27}$$

where $\mathbf{K} = \Sigma_{\varepsilon\varepsilon}\mathbf{H}'(\mathbf{H}\Sigma_{\varepsilon\varepsilon}\mathbf{H}' + \Sigma_{rr})^{-1}$, and \mathbf{x}^b , \mathbf{y}^o , and \mathbf{y}^b are treated like column vectors (to simplify the syntax). As expected, the analysis \mathbf{y}^a is a weighting between the prior $\mathcal{N}(\mathbf{y}^b, \mathbf{H}\Sigma_{\varepsilon\varepsilon}\mathbf{H}')$ and some likelihood $\mathcal{N}(\mathbf{y}^o, \Sigma_{rr})$.

Table 4. Notation for Sect. 3.3 (unless otherwise stated in text).

Symbol	Definition
\mathbf{C}	Generic real symmetric square matrix
\mathbf{D}^2, \mathbf{V}	Matrices of eigenvalues and eigenvectors
\mathbf{U}	Left singular vectors
$\Lambda(\cdot)$	Eigenvalue function, e.g., $\Lambda(\mathbf{C}) = [d_1^2, \dots, d_p^2]$
\mathbf{F}	Diagonal scaling matrix
\mathbf{H}	$p \times m$ proxy model matrix*, $\mathbf{H} = \mathbf{B}^{*'} \otimes \mathbb{I}_p$
\mathbf{K}	$m \times p$ Kalman gain matrix
$\mathbf{X} = [\text{SST} \ \text{SSS}]$	$n \times m$ matrix with n time points
\mathbf{Y}^b	An n -length column vector of \mathbf{Y}^b
$\mathbf{x}^a, \mathbf{x}^b$	m -length vectors, the analysis and background state
$\mathbf{y}^a, \mathbf{y}^b, \mathbf{y}^o$	Similarly, the p -length vectors for \mathbf{y} (and the observation vector) Also, e.g., $\mathbf{x}^b, \mathbf{y}^b$ are row vectors of $\mathbf{X}^b, \mathbf{Y}^b$
$\lambda \Sigma_{\varepsilon\varepsilon}$	$m \times m$ (prior) covariance matrix of state noise, with scaling parameter λ
Σ_{rr}	$p \times p$ covariance matrix of observation noise

* In the experiments, $p = 5$ and $m = 10$ because there are two state variables and five grid points.

In both experiments that follow, the value of \mathbf{H} in Eq. (27) changes (but not the value of Σ_{rr} , as defined above). Three different $\mathbf{H} = \mathbf{B}^{*'} \otimes \mathbb{I}_p$ are tested, which are constructed from three values of \mathbf{B}^* : $\mathbf{B}_{\text{ols}}^* = [-0.13, 0.42]$, $\mathbf{B}_{\text{post1}}^* = [-0.189, 0.635]$, and $\mathbf{B}_{\text{post2}}^* = [-0.248, 0.492]$ (see Eq. 25 and Table 3). The difference between the two experiments is the value of $\mathbf{y}^b = \mathbf{H}\mathbf{x}^b$.

The data assimilation was implemented using a reduced-rank Kalman filter (Elbern et al., 2014) (rank reduction of $\Sigma_{\varepsilon\varepsilon}$ is only used in Experiment 2). A notebook and code module for the experiments is available (Fischer, 2024). Note in the following subsections that \mathbf{V} only refers to eigenvectors (see Table 4) and not the noise from MEPSM.

3.3.1 Experiment 1

In Experiment 1, \mathbf{x}^b (and thus \mathbf{y}^b) is set to zero. This mimics an offline DA situation where the model prior might be centered (i.e., zero mean) and constant. In addition, a scaling parameter λ is added to the prior, i.e., $\mathcal{N}(\mathbf{y}^b, \mathbf{H}(\lambda \Sigma_{\varepsilon\varepsilon})\mathbf{H}')$. If the prior covariance is large, then the prior is noninformative, and $\mathbf{y}^a \approx \mathbf{y}^o$. Here a tighter prior is of interest, so λ is set to $\lambda = 0.1$. Henceforth $\Sigma_{\varepsilon\varepsilon}$ refers to a scaled prior covariance matrix, but the λ term is dropped.

The DA posterior time series \mathbf{Y}^a is shown for the three different values of \mathbf{H} in Fig. 5a, and the eigenvalue shrinkage (explained below) is shown in Fig. 5b. For $\lambda = 0.1$, the \mathbf{Y}^a resulting from \mathbf{H}_{ols} shows the most shrinkage towards the zero-mean prior (Fig. 5a).

To understand the results from theory, two pieces of information are needed (see Table 4 for notation):

- $\Lambda(\Sigma\mathbf{C}) = \Lambda(\Sigma^{0.5}\mathbf{C}\Sigma^{0.5})$, where Σ and \mathbf{C} are real symmetric square matrices.

- $\mathbf{H}\Sigma_{\varepsilon\varepsilon}\mathbf{H}'$ and Σ_{rr} have similar eigenvectors (because here $\text{Diag}(\mathbf{F})$ is near constant) but different eigenvalues.

Setting \mathbf{y}^b to zero removes the $\Sigma_{rr}\mathbf{y}^b$ term in Eq. (27), leaving $\mathbf{y}^a = \frac{(\mathbf{H}\Sigma_{\varepsilon\varepsilon}\mathbf{H}')}{(\mathbf{H}\Sigma_{\varepsilon\varepsilon}\mathbf{H}' + \Sigma_{rr})}\mathbf{y}^o$. The eigenvalues of the $p \times p$ projection matrix $\frac{(\mathbf{H}\Sigma_{\varepsilon\varepsilon}\mathbf{H}')}{(\mathbf{H}\Sigma_{\varepsilon\varepsilon}\mathbf{H}' + \Sigma_{rr})}$ are of interest here. It follows (from the points above) that the projection matrix has the same eigenvalues as

$$\begin{aligned} & (\mathbf{H}\Sigma_{\varepsilon\varepsilon}\mathbf{H}')^{0.5}(\mathbf{H}\Sigma_{\varepsilon\varepsilon}\mathbf{H}' + \Sigma_{rr})^{-1}(\mathbf{H}\Sigma_{\varepsilon\varepsilon}\mathbf{H}')^{0.5} \\ &= \mathbf{VDV}'(\mathbf{VD}^2\mathbf{V}' + \mathbf{VS}^2\mathbf{V}')^{-1}\mathbf{VDV}' \\ &= \mathbf{VD}(\mathbf{D}^2 + \mathbf{S}^2)^{-1}\mathbf{DV}', \end{aligned} \tag{28}$$

where \mathbf{D}^2 and \mathbf{S}^2 are the eigenvalue matrices of $\mathbf{H}\Sigma_{\varepsilon\varepsilon}\mathbf{H}'$ and Σ_{rr} respectively. So in this example, the shrinkage factors of the projection matrix can be expressed as $d_j^2/(d_j^2 + s_j^2)$. Figure 5b shows that using \mathbf{H}_{ols} shrinks the eigenvalues more than for \mathbf{H} from MEPSM. In this introductory experiment, the results have been examined in observation space; the next experiment will consider both \mathbf{x}^a and \mathbf{y}^a .

3.3.2 Experiment 2

In Experiment 2, model SST and SSS data were obtained for 1950–2005 from the iCESM historical run (which ended in 2005), for \mathbf{x}^b . The model data were deseasonalized and filtered using a low-pass (> 2 -year) filter (so here \mathbf{x}^b is a smooth prior, not a zero vector). This setup mimics data assimilation with a transient prior, e.g., offline data assimilation with time windows (windows are a type of prior smoothing). Note, for this experiment, that the model trajectory does not need to strongly match the observed historic trajectory. Further, $\Sigma_{\varepsilon\varepsilon}$ is not a full rank matrix (the first three eigenvalues span 98% of the total variance), so for Experiment 2 only

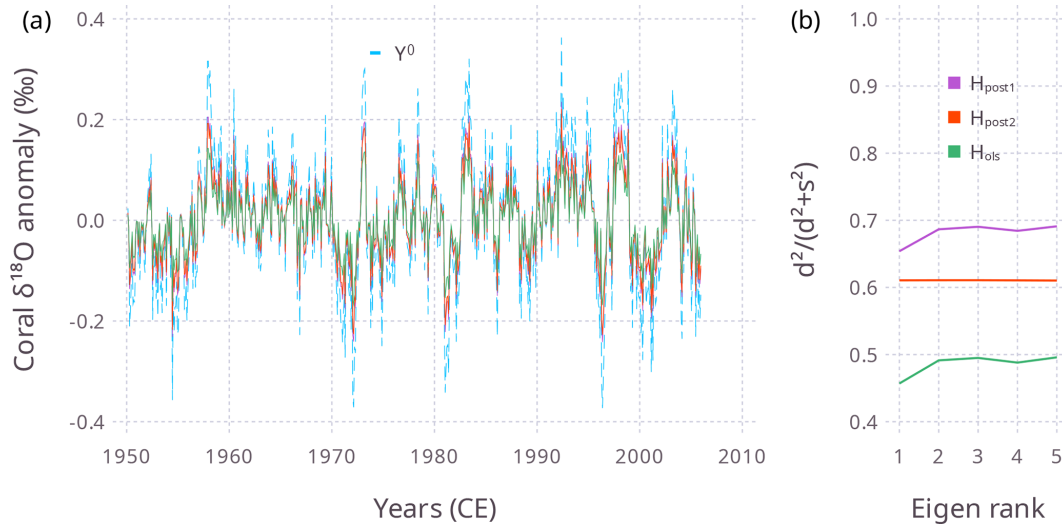


Figure 5. For Experiment 1 ($\lambda = 0.1, \mathbf{y}^b = \mathbf{0}_p$), (a) the time series \mathbf{Y}^a (solid lines, color key as in b) and (b) the eigenvalue shrinkage of the projection matrix (see text), for three different values of \mathbf{H} . These experiments used five model grid points, but here \mathbf{Y}^a is plotted for the central grid point. In (a) the dashed line is \mathbf{Y}^o .

the first three leading components of $\Sigma_{\epsilon\epsilon}$ are used (which helps to regularize \mathbf{X}^a). Figure 6 shows, for $\lambda = 0.1$, the three time series $\mathbf{Y}^o, \mathbf{Y}^b$, and \mathbf{Y}^a for each different \mathbf{H} (named $\mathbf{H}_{ols}, \mathbf{H}_{post1}, \mathbf{H}_{post2}$, where the latter two refer to the first and final posterior for MEPSM). Figure 6 shows the advantage of \mathbf{H} from MEPSM is that, for example, $\mathbf{y}^b = \mathbf{H}_{post1}\mathbf{x}^b$ has a greater range and thus allows \mathbf{y}^a to better capture the peaks and troughs in \mathbf{y}^o than can $\mathbf{y}^b = \mathbf{H}_{ols}\mathbf{x}^b$. To further compare the results for the different \mathbf{H} , ($\mathbf{X}^a, \mathbf{Y}^a$) was calculated for $\lambda = [10^{-2}, \dots, 10^2]$ and compared with $(\mathbf{X}^*, \mathbf{Y}^*)$ using the mean absolute error metric:

$$mae(\mathbf{Y}^a, \mathbf{Y}^*) = n^{-1} \sum_{i=1}^n |y_i^a - y_i^*|. \tag{29}$$

This was done similarly for $mae(\mathbf{X}^a, \mathbf{X}^*)$. Note that $(\mathbf{X}^*, \mathbf{Y}^*)$, the estimated state from MEPSM, has no direct part in the DA.

Figure 7 shows that for this site, for \mathbf{H}_{post2} (and \mathbf{H}_{post1}) the analysis \mathbf{X}_{sst}^a and \mathbf{Y}^a typically have a smaller mae compared to \mathbf{H}_{ols} . Also, mae decreases as λ increases. For $mae(\mathbf{Y}^a, \mathbf{Y}^*)$, the decrease flattens when $\mathbf{y}^a \approx \mathbf{y}^o$ (for $\lambda > 1$, Fig. 7b), but note that \mathbf{Y}^o is not \mathbf{Y}^* . So in this example, it seems that a solution (\mathbf{X}^a) better than the likelihood ($\mathbf{y}^o = \mathbf{H}\mathbf{x}^a$) cannot be achieved, which means that \mathbf{H} from MEPSM wins (lower mae for \mathbf{X}_{sst}^a), because \mathbf{H}_{ols} biases \mathbf{X}_{sst}^a (Fig. 7a). For \mathbf{H}_{ols} , there is less bias for \mathbf{X}_{sst}^a , perhaps owing to the relative difference between OLS and MEPSM with respect to b_2 (the SSS coefficient). Performing similar data assimilation experiments on global climate proxy datasets will better inform such issues and regional variation.

4 Conclusions

MEPSM is a new type of proxy system model which incorporates both prior information and generalized noise in all variables, such as cross-correlated noise within the predictor variables. DA theory and two experiments suggest that, in data assimilation, the MEPSM should have an advantage over conventional OLS-calibrated PSMs (the latter are widely employed in current paleo-DA projects). In the work presented here for MEPSM, the response noise is assumed to be independent of the predictor noise (the zero matrices in Σ_Q), but cross-correlation between the predictor and response noise is possible (if needed), by complete quadratic multiplication of Σ_Q , rather than treating Ω_V and Σ_A separately, in Eq. (9). However, for many PSMs the predictor and response noise are fundamentally different and therefore should be independent. The next step is to apply MEPSM to calibrate multiple-input proxy system models for different climate proxies (e.g., isotopes in corals, tree rings, and speleothems) and to incorporate MEPSM into current data assimilation projects.

Appendix A: Basis of Eq. (2)

Let

$$y = f(x_1 - a_1, \dots, x_p - a_p). \tag{A1}$$

A first-order Taylor expansion gives

$$y \approx f(x_1, \dots, x_p) - \sum_{k=1}^p \left[\frac{\partial f(x_1, \dots, x_p)}{\partial x_k} a_k \right] \tag{A2}$$

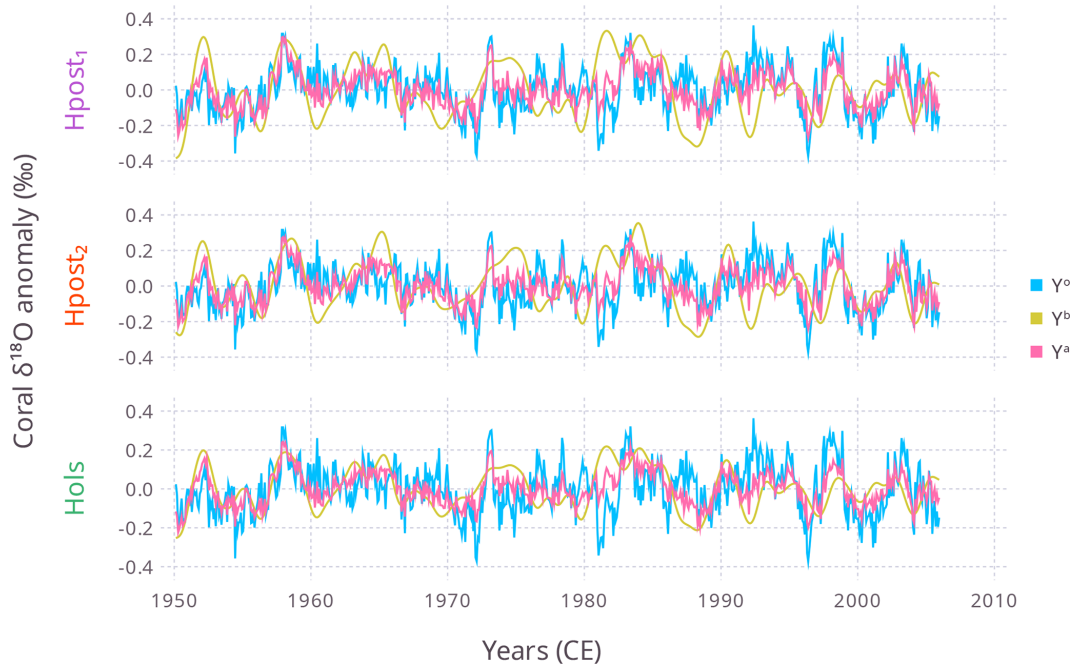


Figure 6. For Experiment 2 ($\lambda = 0.1$, $\mathbf{y}^b = \mathbf{H}\mathbf{x}^b$), the time series of \mathbf{Y}^o , \mathbf{Y}^b , and \mathbf{Y}^a (color key right) for three different \mathbf{H} (on left axis). As in Fig. 5 \mathbf{Y}^a and \mathbf{Y}^b refer to the central grid point. In the panels, \mathbf{Y}^o is the same, but \mathbf{Y}^b and \mathbf{Y}^a depend on \mathbf{H} . Note how \mathbf{y}^a is always between \mathbf{y}^b and \mathbf{y}^o .

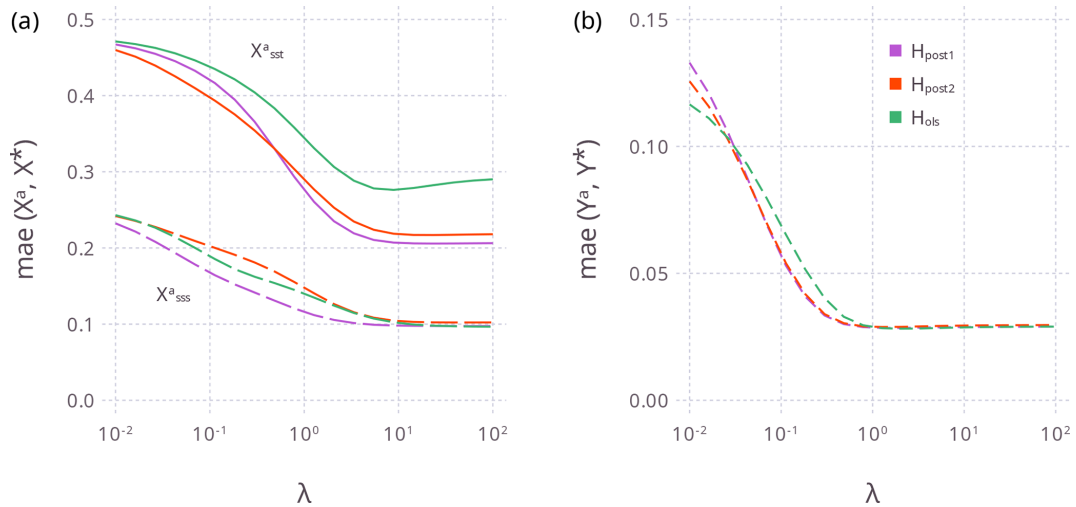


Figure 7. For Experiment 2 ($\mathbf{y}^b = \mathbf{H}\mathbf{x}^b$), the mean absolute error for different λ for (a) \mathbf{X}^a and (b) \mathbf{Y}^a (using the central grid point series) and for the three different \mathbf{H} (color key as in b). Panel (a) shows the mae for \mathbf{X}^a_{sst} (solid lines) and \mathbf{X}^a_{sss} (long dashed lines), while panel (b) is the mae for \mathbf{Y}^a (see Eq. 29). Small values of λ mean $\mathbf{y}^a \approx \mathbf{y}^b$, while large λ values mean $\mathbf{y}^a \approx \mathbf{y}^o$.

such that the variance of $f(\mathbf{x})$ is $\text{Var}(f(\mathbf{x})) \approx df(\mathbf{x})' \Sigma_{aa} df(\mathbf{x})$. The first-order Taylor expansion will be exact when $f(\mathbf{x})$ is a linear function.

Appendix B: Notation for covariance

Covariance can be given in plain or scaled format, e.g.,

$$\Sigma = \sigma^2 \mathbf{S}. \tag{B1}$$

So, just as an example, a naive covariance matrix for \mathbf{B} could be written as follows:

$$\Sigma_B = (\mathbf{X}^{*'} \Sigma^{-1} \mathbf{X}^*)^{-1} = \sigma^2 (\mathbf{X}^{*'} \mathbf{S}^{-1} \mathbf{X}^*)^{-1}. \tag{B2}$$

In this paper I use the plain format because it simplifies many expressions.

Appendix C: The WTLS likelihood and its mean B_{wtls}

The target function for WTLS essentially consists of the quadratic component of the log-likelihood function:

$$\phi_{wtls}(\mathbf{B}, \mathbf{X}^*) = \left(\text{vec}(\mathbf{Z}) - \begin{bmatrix} \mathbf{Y}^* \\ \text{vec}(\mathbf{X}^*) \end{bmatrix} \right)' \cdot \Sigma_Q^{-1} \left(\text{vec}(\mathbf{Z}) - \begin{bmatrix} \mathbf{Y}^* \\ \text{vec}(\mathbf{X}^*) \end{bmatrix} \right), \quad (C1)$$

where $\mathbf{Y}^* = [\mathbf{1}_n \quad \mathbf{X}^*] \mathbf{B}$. For $\partial\phi/\partial\mathbf{B}$, Eq. (C1) reduces to

$$\frac{\partial\phi_{wtls}}{\partial\mathbf{B}} = \frac{\partial[(\mathbf{Y} - b_0 - \mathbf{X}^* \mathbf{B}^*)' \Omega_V^{-1} (\mathbf{Y} - b_0 - \mathbf{X}^* \mathbf{B}^*)]}{\partial\mathbf{B}} \quad (C2a)$$

$$= 2\tilde{\mathbf{X}}' \Omega_V^{-1} \mathbf{V}, \quad (C2b)$$

where $\tilde{\mathbf{X}} = [\mathbf{1}_n \quad \mathbf{X}^*]$. Next substitute \mathbf{V} with Eq. (18):

$$\frac{\partial\phi_{wtls}}{\partial\mathbf{B}} = 2\tilde{\mathbf{X}}' \mathbf{W}^{-1} \mathbf{e}. \quad (C3)$$

The residual vector \mathbf{e} can be written as

$$\mathbf{e} = (\mathbf{Y} - b_0 - \mathbf{A}\mathbf{B}^*) - \mathbf{X}^* \mathbf{B}^* \quad (C4)$$

or

$$\mathbf{e} = \mathbf{Y} - \mathbf{X}\mathbf{B}. \quad (C5)$$

Then setting Eq. (C3) to zero, substituting for \mathbf{e} , and rearranging gives

$$\mathbf{B}_{wtls} = \left[(\mathbf{X} - \mathbf{A})' \mathbf{W}^{-1} (\mathbf{X} - \mathbf{A}) \right]^{-1} \cdot \left[(\mathbf{X} - \mathbf{A})' \mathbf{W}^{-1} (\mathbf{Y} - \mathbf{A}\mathbf{B}) \right] \quad (C6)$$

or

$$\mathbf{B}_{wtls} = \left[(\mathbf{X} - \mathbf{A})' \mathbf{W}^{-1} \mathbf{X} \right]^{-1} \left[(\mathbf{X} - \mathbf{A}) \mathbf{W}^{-1} \mathbf{Y} \right], \quad (C7)$$

where \mathbf{A} now may include a column of zeros (corresponding to noiseless predictors in \mathbf{X}).

Equation (C7) expands to

$$\mathbf{B}_{wtls} = \left[\mathbf{X}' \mathbf{W}^{-1} \mathbf{X} - \mathbf{A}' \mathbf{W}^{-1} \mathbf{X} \right]^{-1} \left[\mathbf{X}' \mathbf{W}^{-1} \mathbf{Y} - \mathbf{A}' \mathbf{W}^{-1} \mathbf{Y} \right], \quad (C8)$$

where $(\mathbf{A}' \mathbf{W}^{-1} \mathbf{X})$ is the weighted covariance between \mathbf{A} and the predictors, and $(\mathbf{A}' \mathbf{W}^{-1} \mathbf{Y})$ is the weighted covariance between \mathbf{A} and the response vector. This shows that WTLS can be expressed as an adjusted least squares problem. Hence, it can also be expressed as an augmented linear regression:

$$\mathbf{Y} - \mathbf{A}\mathbf{B} = (\mathbf{X} - \mathbf{A})\mathbf{B} + (\mathbf{V} - \mathbf{A}\mathbf{B}), \quad (C9)$$

which is equivalent to Eq. (7).

Appendix D: Covariance of B_{wtls}

One approximation for Σ_B is as follows. Let

$$\mathbf{B} = \mathbf{K}\mathbf{Y} = \mathbf{K}(\mathbf{X}\mathbf{B} + \mathbf{e}). \quad (D1)$$

If $\mathbf{K}\mathbf{X} = \mathbb{I}$, then

$$\Sigma_B = \mathbf{K}\Sigma\mathbf{K}', \quad (D2)$$

where $\Sigma = \text{Diag}(\mathbf{e} \otimes \mathbf{e})$. For \mathbf{B}_{wtls} , from Eqs. (D1) and (C7) it follows that $\mathbf{K} = [(\mathbf{X} - \mathbf{A})' \mathbf{W}^{-1} \mathbf{X}]^{-1} (\mathbf{X} - \mathbf{A})' \mathbf{W}^{-1}$. Therefore,

$$\Sigma_B^{wtls} = \left[(\mathbf{X} - \mathbf{A})' \mathbf{W}^{-1} \mathbf{X} \right]^{-1} \mathbf{M} \left[(\mathbf{X} - \mathbf{A})' \mathbf{W}^{-1} \mathbf{X} \right]^{-1'}, \quad (D3)$$

where $\mathbf{M} = (\mathbf{X} - \mathbf{A})' \mathbf{W}^{-1} \Sigma \mathbf{W}^{-1} (\mathbf{X} - \mathbf{A})$.

Appendix E: Construction of Σ_A and Ω_V

For two predictors, Σ_A is

$$\Sigma_A = \begin{bmatrix} \Omega_{A_1} & \Omega_{A_1 A_2} \\ \Omega_{A_2 A_1} & \Omega_{A_2} \end{bmatrix}. \quad (E1)$$

Σ_A can be constructed as

$$\Sigma_A = \text{Diag}(\sigma_A)' \begin{bmatrix} \mathbf{R}_{A_1} & \mathbf{R}_{A_1 A_2} \\ \mathbf{R}_{A_2 A_1} & \mathbf{R}_{A_2} \end{bmatrix} \text{Diag}(\sigma_A), \quad (E2)$$

where $\text{Diag}(\cdot)$ is a diagonal matrix, and σ_A is a $(n \cdot p)$ -length vector containing the heteroscedastic ‘‘variance’’ of the predictor noise for each predictor (stacked vertically). The central block matrix, made up of submatrices \mathbf{R} , will be labeled $\tilde{\mathbf{R}}$. The submatrices of $\tilde{\mathbf{R}}$ are $n \times n$ correlation matrices, containing the autocorrelation or cross-correlation information for the predictor noise \mathbf{A} :

$$\mathbf{R} = \begin{bmatrix} \rho_0 & \rho_1 & \rho_2 & \rho_3 & \dots & \rho_{n-1} \\ \rho_1 & \rho_0 & \rho_1 & \rho_2 & \dots & \rho_{n-2} \\ \rho_2 & \rho_1 & \rho_0 & \rho_1 & \dots & \rho_{n-3} \\ \rho_3 & \rho_2 & \rho_1 & \rho_0 & \dots & \rho_{n-3} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho_{n-1} & \rho_{n-2} & \rho_{n-3} & \rho_{n-4} & \dots & \rho_0 \end{bmatrix}, \quad (E3)$$

where $(\rho_0, \rho_1, \dots, \rho_{n-1})$ is the auto- or cross-correlation function.

If \mathbf{R} is an autocorrelation matrix, then $\rho_0 = 1$. Also, because the auto- or cross-correlation may become ~ 0 at long lags, then $\rho_{>k}$ may be set to 0, for a chosen lag k . To ensure that \mathbf{R} is positive definite, the algorithm of Rebonato and Jaekel (1999) is computationally fast.

Ω_V can be similarly constructed:

$$\Omega_V = \text{Diag}(\sigma_V)' \mathbf{R}_V \text{Diag}(\sigma_V). \quad (E4)$$

Appendix F: Maximum likelihood solution of Hannart et al. (2014)

Hannart et al. (2014) defined the log-likelihood of the EIV model as (here written using the notation in this paper, rather than the notation of Hannart et al., 2014)

$$\ell(\mathbf{B}^*, \mathbf{X}^*) = (\mathbf{Y} - b_0 - \mathbf{X}^* \mathbf{B}^*)' \Omega_V^{-1} (\mathbf{Y} - b_0 - \mathbf{X}^* \mathbf{B}^*) + \sum_i (\mathbf{X}_i - \mathbf{X}_i^*)' \Omega_{A_i}^{-1} (\mathbf{X}_i - \mathbf{X}_i^*), \tag{F1}$$

which is a reduced form of Eq. (C1), because the off-diagonal matrices $\Omega_{A_i A_j}$ in Σ_A are not included. Hannart et al. (2014) wrote an iterative (fixed-point) solution in terms of \mathbf{X}^* and \mathbf{B}^* .

$$\mathbf{X}_i^* = (\Omega_{A_i}^{-1} + b_i^2 \Omega_V^{-1})^{-1} (b_i \Omega_V^{-1} \bar{\mathbf{y}}_i + \Omega_{A_i}^{-1} \mathbf{X}_i) \tag{F2a}$$

$$\mathbf{B}^* = (\mathbf{X}^{*'} \Omega_V^{-1} \mathbf{X}^*)^{-1} \mathbf{X}^{*'} \Omega_V^{-1} (\mathbf{Y} - b_0), \tag{F2b}$$

where here $\mathbf{B}^* = [b_1 \ b_2]'$, and $\bar{\mathbf{y}}_i = \mathbf{Y} - b_0 - \mathbf{X}_{-i}^* \mathbf{B}_{-i}^*$. (I have included the intercept b_0 explicitly.) Confidence intervals for each b_i were calculated from the likelihood profiles (see Sect. 3 in Hannart et al., 2014).

The ML solution I provide here differs from Hannart et al. (2014), but I adopt a similar form (what follows is only for the ML estimate in the first row of Table 3). I express the log-likelihood in Eq. (F1) as a function of \mathbf{V} and \mathbf{A} :

$$\ell(\mathbf{B}) = \mathbf{V}' \Omega_V^{-1} \mathbf{V} + \sum_i \mathbf{A}_i' \Omega_{A_i}^{-1} \mathbf{A}_i, \tag{F3}$$

where \mathbf{A} and \mathbf{V} are calculated using Eqs. (17)–(18). The difference with Hannart et al. (2014) is that the likelihood here is rewritten with respect to $\mathbf{e} = \mathbf{Y} - \mathbf{X}\mathbf{B}$ (through Eqs. 17–18), whereas the procedure of Hannart et al. (2014) is expressed with respect to $\mathbf{Y} - \mathbf{X}^* \mathbf{B}^*$, i.e., the main difference being \mathbf{X} or \mathbf{X}^* . The point here is merely to provide a basic ML solution in Table 3, in order to compare with the other solutions. Equation (F3) should work with many standard ML packages, because there is no iterative dependence on \mathbf{X}^* .

The first row in Table 3 was calculated using Eq. (F3), and the Julia package ProfileLikelihoods.jl (VandenHeuvel, 2022).

Appendix G: MEPSM and paleodata assimilation

This appendix explores the potential advantages of MEPSM for paleodata assimilation (PDA). This is a stand-alone appendix, and notation may differ from the main text. The following notation is used for the data assimilation components:

- \mathbf{x}^b is the climate model data, at m grid points.
- \mathbf{y} is the proxy data, at p grid points.
- θ is the unobserved state vector, with m grid points.

– \mathbf{H} is the operator which converts model space to proxy space.

Matrix \mathbf{H} is $p \times m$. (Note that if a proxy system model has multiple inputs, then \mathbf{H} can be formed by concatenation, e.g., $\mathbf{H} = [\mathbf{H}_{\text{sst}} \ \mathbf{H}_{\text{sss}}]$ in which case m above becomes $2m$.)

Data assimilation can be expressed in GLS format:

$$\mathbf{y} = \mathbf{H}\theta + \mathbf{r} \tag{G1a}$$

$$\mathbf{x}^b = \theta + \boldsymbol{\varepsilon}, \tag{G1b}$$

where $\mathbf{r} \sim \mathcal{N}(\mathbf{0}, \Sigma_{rr})$ and $\boldsymbol{\varepsilon} \sim \mathcal{N}(\mathbf{0}, \Sigma_{\varepsilon\varepsilon})$. In the typical of-fine PDA, $\Sigma_{\varepsilon\varepsilon}$ is an $m \times m$ background covariance matrix. In MEPSM, the predictor noise matrix is Σ_A . Out of interest, a time-constant noise matrix could be specified as $\Sigma_A = \Sigma_{aa} \otimes \mathbb{I}_n$. Note the difference between $\Sigma_{\varepsilon\varepsilon}$ and Σ_{aa} : $\Sigma_{\varepsilon\varepsilon}$ is estimated from climate model ensembles, whereas Σ_{aa} is from ensemble climate data reconstructions (e.g., an ensemble SST product). Also, PDA is about estimating θ , whereas MEPSM is about calibrating \mathbf{H} (in MEPSM \mathbf{B} contains the non-zero elements of \mathbf{H}). Note that \mathbf{H} is typically calibrated using observed modern climate data and proxies, over a coeval period (e.g., the 20th century). I will say more about \mathbf{H} below.

Next concatenate the two equations above into one:

$$\mathbf{z} = \begin{bmatrix} \mathbf{y} \\ \mathbf{x}^b \end{bmatrix} = \tilde{\mathbf{H}}\theta + \mathbf{q}, \tag{G2}$$

where $\tilde{\mathbf{H}} = \begin{bmatrix} \mathbf{H} \\ \mathbb{I}_m \end{bmatrix}$, and let $\Sigma_{qq} = \begin{bmatrix} \Sigma_{rr} & 0 \\ 0 & \Sigma_{\varepsilon\varepsilon} \end{bmatrix}$. The GLS solution for θ is

$$\begin{aligned} \theta^a &= (\tilde{\mathbf{H}}' \Sigma_{qq}^{-1} \tilde{\mathbf{H}})^{-1} \tilde{\mathbf{H}}' \Sigma_{qq}^{-1} \mathbf{z} \\ &= (\mathbf{H}' \Sigma_{rr}^{-1} \mathbf{H} + \Sigma_{\varepsilon\varepsilon}^{-1})^{-1} (\mathbf{H}' \Sigma_{rr}^{-1} \mathbf{y} + \Sigma_{\varepsilon\varepsilon}^{-1} \mathbf{x}^b). \end{aligned} \tag{G3}$$

For the analysis error covariance Σ^a , it is

$$\Sigma^a = (\mathbf{H}' \Sigma_{rr}^{-1} \mathbf{H} + \Sigma_{\varepsilon\varepsilon}^{-1})^{-1}. \tag{G4}$$

Equation (G4) can also be expanded as

$$\begin{aligned} \Sigma^a &= \Sigma^a (\Sigma^a)^{-1} \Sigma^a = \Sigma^a (\mathbf{H}' \Sigma_{rr}^{-1} \mathbf{H} + \Sigma_{\varepsilon\varepsilon}^{-1}) \Sigma^a \\ &= \mathbf{K} \Sigma_{rr} \mathbf{K}' + \Sigma^a \Sigma_{\varepsilon\varepsilon}^{-1} \Sigma^a, \end{aligned} \tag{G5}$$

where $\mathbf{K} = \Sigma^a \mathbf{H}' \Sigma_{rr}^{-1}$. The question here is what happens if \mathbf{H} is calibrated using a biased method (e.g., OLS). Equation (G4) suggests that Σ^a will also be biased. Ensemble filtering methods are used for their computational advantage: replacing a large $m \times m$ $\Sigma_{\varepsilon\varepsilon}$ with a smaller ensemble of model \mathbf{x}^b vectors. But the model \mathbf{x}^b vectors are filtered in a way that ensures the resulting ensemble of $\delta\theta^a$ (i.e., deviation from $\bar{\theta}^a$) agrees with Eq. (G5). So ensemble filtering

methods do not help if \mathbf{H} is biased. Equation (G1) shows that the operator \mathbf{H} is applied to the unobserved state θ , not to the “observed” \mathbf{x}^b , so using OLS for calibrating \mathbf{H} seems flawed. OLS calibrates a PSM on observed (but not true) data, whereas MEPSM calibrates a PSM on the unobserved state \mathbf{x}^* .

Code and data availability. Intuitive Julia code for MEPSM is available on GitHub (<https://github.com/Mattriks/MeasurementErrorModels.jl>, last access: 2 April 2023) and Zenodo (<https://doi.org/10.5281/zenodo.7793741>, Fischer, 2023a). A manual for MEPSM is available at <https://mattriks.github.io/MeasurementErrorModels.jl/dev/> (Fischer, 2023b). The manual contains three examples which reproduce results from this paper.

The ERSSTv5 sea surface temperature data are available at <https://psl.noaa.gov/data/gridded/data.noaa.ersst.v5.html> (last access: 16 December 2019) (metadata: <https://doi.org/10.7289/V5T72FNM>, Huang et al., 2017). The ERSSTv5 uncertainty data were obtained from Boyin Huang, NOAA (22 October 2020).

The HadEN4 ocean salinity data (and uncertainty) are available at <https://apdrc.soest.hawaii.edu/las/v6/dataset?catitem=16652> (Gouretski and Reseghetti, 2020). These data are from version EN.4.2.1 and are the G10 analyses, i.e., with the corrections of Gouretski and Reseghetti (2010).

The coral $\delta^{18}\text{O}$ data from core RI6 were extracted from the CoralHydro2k database (Walter et al., 2023). A Jupyter Notebook (as a PDF) showing the preprocessing of these data is available in the Supplement. The dataset stored for the code examples in the GitHub repository contains all three detrended variables (detrended without removing the intercept).

Some input variables for PSMs, from the iCESM1 last millennium simulation (Brady et al., 2019), are available at the University of Washington (<https://atmos.washington.edu/~rtardif/LMR/prior>, Tardif, 2024). The variables ocean surface temperature (tos_sfc) and ocean surface salinity (sos_sfc) are used in Sect. 3.3.

A notebook and code module for Sect. 3.3 are available on Zenodo (<https://doi.org/10.5281/zenodo.12660485>, Fischer, 2024).

Supplement. The supplement related to this article is available online at: <https://doi.org/10.5194/gmd-17-6745-2024-supplement>.

Competing interests. The author has declared that there are no competing interests.

Disclaimer. Publisher’s note: Copernicus Publications remains neutral with regard to jurisdictional claims made in the text, published maps, institutional affiliations, or any other geographical representation in this paper. While Copernicus Publications makes every effort to include appropriate place names, the final responsibility lies with the authors.

Acknowledgements. I thank the reviewers for useful suggestions that helped to improve the manuscript.

Review statement. This paper was edited by Marko Scholze and reviewed by two anonymous referees.

References

- Amiri-Simkooei, A. and Jazaeri, S.: Weighted total least squares formulated by standard least squares theory, *Journal of Geodetic Science*, 2, 113–124, <https://doi.org/10.2478/v10156-011-0036-5>, 2012.
- Brady, E., Stevenson, S., Bailey, D., Liu, Z., Noone, D., Nusbaumer, J., Otto-Bliesner, B. L., Tabor, C., Tomas, R., Wong, T., Zhang, J., and Zhu, J.: The Connected Isotopic Water Cycle in the Community Earth System Model Version 1, *J. Adv. Model. Earth Sy.*, 11, 2547–2566, <https://doi.org/10.1029/2019MS001663>, 2019.
- Dee, S.: PRYSM: An open-source framework for PRoxY System Modeling, with applications to oxygen-isotope systems, *Adv. Model. Earth Sy.*, 7, 1220–1247, <https://doi.org/10.1002/2015MS000447>, 2015.
- Delcroix, T., Alory, G., Cravatte, S., Corrège, T., and McPhaden, M. J.: A gridded sea surface salinity data set for the tropical Pacific with sample applications (1950–2008), *Deep-Sea Res. Pt. I*, 58, 38–48, <https://doi.org/10.1016/j.dsr.2010.11.002>, 2011.
- DeLong, K. L., Atwood, A., Moore, A., and Sanchez, S.: Clues from the sea paint a picture of Earth’s water cycle, *Eos*, 103, <https://doi.org/10.1029/2022EO220231>, 2022.
- Elbern, H., Friese, E., Nieradzick, L., and Schwinger, J.: Data assimilation in atmospheric chemistry and air quality, in: *Advanced Data Assimilation for Geosciences: Lecture Notes of the Les Houches School of Physics: Special Issue, June 2012*, Oxford University Press, <https://doi.org/10.1093/acprof:oso/9780198723844.003.0022>, 2014.
- Evans, M. N., Tolwinski-Ward, S. E., Thompson, D. M., and Anchukaitis, K. J.: Applications of proxy system modeling in high resolution paleoclimatology, *Quaternary Sci. Rev.*, 76, 16–28, <https://doi.org/10.1016/j.quascirev.2013.05.024>, 2013.
- Fang, X., Li, B., Alkhatib, H., Zeng, W., and Yao, Y.: Bayesian Inference for the Errors-In-Variables model, *Stud. Geophys. Geod.*, 61, 35–52, <https://doi.org/10.1007/s11200-015-6107-9>, 2017.
- Fischer, M.: Mattriks/MeasurementErrorModels.jl: MEPSM v0.2.0, Zenodo [code], <https://doi.org/10.5281/zenodo.7793741>, 2023a.
- Fischer, M.: Measurement Error Proxy System Models: MEPSM v0.2, Model documentation, <https://mattriks.github.io/MeasurementErrorModels.jl/dev/> (last access: 2 April 2023), 2023b.
- Fischer, M.: Data assimilation notebook and code module, Zenodo [code], <https://doi.org/10.5281/zenodo.12660485>, 2024.
- Florens, J.-P., Mouchart, M., and Richard, J.-F.: Bayesian inference in Errors-in-Variables Models, *J. Multivariate Anal.*, 4, 419–452, <https://doi.org/10.1007/s11200-015-6107-9>, 1974.
- Fuller, W. A.: Chapter 3: Extensions of the single relation model, in: *Measurement Error Models*, John Wiley & Sons, <https://doi.org/10.1002/9780470316665.ch3>, 1987.

- Fuller, W. A.: Prediction of true values for the measurement error model, in: *Statistical Analysis of Measurement Error Models and Applications*, edited by: Brown, P. J. and Fuller, W. A., American Mathematical Society, 41–58, <https://doi.org/10.1090/conm/112/1087098>, 1990.
- Good, S. A., Martin, M. J., and Rayner, N. A.: EN4: Quality controlled ocean temperature and salinity profiles and monthly objective analyses with uncertainty estimates, *J. Geophys. Res.-Oceans*, 118, 6704–6716, <https://doi.org/10.1002/2013JC009067>, 2013.
- Gouretski, V. and Reseghetti, F.: On depth and temperature biases in bathythermograph data: Development of a new correction scheme based on analysis of a global ocean database, *Deep-Sea Res. Pt. I*, 57, 812–833, <https://doi.org/10.1016/j.dsr.2010.03.011>, 2010.
- Gouretski, V. and Reseghetti, F.: HadEN4 salinity and salinity error standard deviation data, Asia-Pacific Data-Research Center [data set], <https://apdrc.soest.hawaii.edu/las/v6/dataset?catitem=16652>, last access: 11 November 2020.
- Hannart, A., Ribes, A., and Naveau, P.: Optimal fingerprinting under multiple sources of uncertainty, *Geophys. Res. Lett.*, 41, 1261–1268, <https://doi.org/10.1002/2013GL058653>, 2014.
- Henderson, H. V. and Searle, S. R.: Vec and vech operators for matrices, with some uses in jacobians and multivariate statistics, *Can. J. Stat.*, 7, 65–81, <https://doi.org/10.2307/3315017>, 1979.
- Huang, B., Thorne, P. W., Banzon, V. F., Boyer, T., Chepurin, G., Lawrimore, J. H., Menne, M. J., Smith, T. M., Vose, R. S., and Zhang, H.-M.: NOAA Extended Reconstructed Sea Surface Temperature (ERSST), Version 5, NOAA National Centers for Environmental Information [data set], <https://doi.org/10.7289/V5T72FNM>, 2017.
- Huang, B., Menne, M. J., Boyer, T., Freeman, E., Gleason, B. E., Lawrimore, J. H., Liu, C., Rennie III, J. J., C. J. S., Sun, F., Vose, R., Williams, C. N., Yin, X., and Zhang, H.-M.: Uncertainty Estimates for Sea Surface Temperature and Land Surface Air Temperature in NOAA GlobalTemp Version 5, *J. Climate*, 33, 1351–1379, <https://doi.org/10.1175/JCLI-D-19-0395.1>, 2020.
- Kalnay, E., Kanamitsu, M., Kistler, R., Collins, W., Deaven, D., Gandin, L., Iredell, M., Saha, S., White, G., Woollen, J., Zhu, Y., Chelliah, M., Ebisuzaki, W., Higgins, W., Janowiak, J., Mo, K. C., Ropelewski, C., Wang, J., Leetmaa, A., Reynolds, R., Jenne, R., and Joseph, D.: The NCEP/NCAR 40-Year Reanalysis Project, *B. Am. Meteorol. Soc.*, 77, 437–471, [https://doi.org/10.1175/1520-0477\(1996\)077<0437:TNYRP>2.0.CO;2](https://doi.org/10.1175/1520-0477(1996)077<0437:TNYRP>2.0.CO;2), 1996.
- King, J., Tierney, J., Osman, M., Judd, E. J., and Anchukaitis, K. J.: DASH: a MATLAB toolbox for paleoclimate data assimilation, *Geosci. Model Dev.*, 16, 5653–5683, <https://doi.org/10.5194/gmd-16-5653-2023>, 2023.
- LeGrande, A. N. and Schmidt, G. A.: Global gridded data set of the oxygen isotopic composition in seawater, *Geophys. Res. Lett.*, 33, L12604, <https://doi.org/10.1029/2006GL026011>, 2006.
- Lemoine, N. P.: Moving beyond noninformative priors: why and how to choose weakly informative priors in Bayesian analyses, *Oikos*, 128, 912–928, <https://doi.org/10.1111/oik.05985>, 2019.
- Loope, G., Thompson, D. M., Cole, J., and Overpeck, J.: Is there a low-frequency bias in multiproxy reconstructions of tropical pacific SST variability?, *Quaternary Sci. Rev.*, 246, 106530, <https://doi.org/10.1016/j.quascirev.2020.106530>, 2020a.
- Loope, G., Thompson, D. M., and Overpeck, J.: The spectrum of Asian Monsoon variability: a proxy system model approach to the hydroclimate scaling mismatch, *Quaternary Sci. Rev.*, 240, 106362, <https://doi.org/10.1016/j.quascirev.2020.106362>, 2020b.
- Ludwig, K. R. and Titterton, D. M.: Calculation of ^{230}ThU isochrons, ages, and errors, *Geochim. Cosmochim. Ac.*, 58, 5031–5042, [https://doi.org/10.1016/0016-7037\(94\)90229-1](https://doi.org/10.1016/0016-7037(94)90229-1), 1994.
- Magnus, J. R. and Neudecker, H.: Symmetry, 0-1 Matrices and Jacobians, *Econometric Theory*, 2, 157–190, <https://doi.org/10.1017/S0266466600011476>, 1986.
- Malevich, S. B., Vetter, L., and Tierney, J. E.: Global Core Top Calibration of $\delta^{18}\text{O}$ in Planktic Foraminifera to Sea Surface Temperature, *Paleoceanogr. Paleocl.*, 34, 1292–1315, <https://doi.org/10.1029/2019PA003576>, 2019.
- Osborne, M. C., Dunbar, R. B., and Mucciarone, D. A.: Regional calibration of coral-based climate reconstructions from Palau, West Pacific Warm Pool, *Palaeogeogr. Palaeocl.*, 386, 308–320, <https://doi.org/10.1016/j.palaeo.2013.06.001>, 2013.
- Rebonato, R. and Jaekel, P.: The most general methodology to create a valid correlation matrix for risk management and option pricing purposes, *J. Risk*, 2, 17–27, <https://doi.org/10.21314/JOR.2000.023>, 1999.
- Sanchez, S. C., Hakim, G. J., and Sanger, C. P.: Climate Model Teleconnection Patterns Govern the Niño-3.4 Response to Early Nineteenth-Century Volcanism in Coral-Based Data Assimilation Reconstructions, *J. Climate*, 34, 1863–1880, <https://doi.org/10.1175/JCLI-D-20-0549.1>, 2021.
- Sayani, H. R., Cobb, K. M., DeLong, K., Hitt, N. T., and Druffel, E. R. M.: Intercolony $\delta^{18}\text{O}$ and Sr/Ca variability among Porites spp. corals at Palmyra Atoll: Toward more robust coral-based estimates of climate, *Geochem. Geophys. Geosyst.*, 20, 5270–5284, <https://doi.org/10.1029/2019GC008420>, 2019.
- Schneider, T.: Analysis of Incomplete Climate Data: Estimation of Mean Values and Covariance Matrices and Imputation of Missing Values, *J. Climate*, 14, 853–871, [https://doi.org/10.1175/1520-0442\(2001\)014<0853:AOICDE>2.0.CO;2](https://doi.org/10.1175/1520-0442(2001)014<0853:AOICDE>2.0.CO;2), 2001.
- Steiger, N. J., Smerdon, J. E., Cook, E. R., and Cook, B. I.: A reconstruction of global hydroclimate and dynamical variables over the Common Era, *Sci. Data*, 5, 180086, <https://doi.org/10.1038/sdata.2018.86>, 2018.
- Tardif, R.: The post-processed “iCESM1” last millennium simulation, *atmos.washington* [data set], <https://atmos.washington.edu/~rtardif/LMR/prior>, last access: 14 June 2024.
- Tardif, R., Hakim, G. J., Perkins, W. A., Horlick, K. A., Erb, M. P., Emile-Geay, J., Anderson, D. M., Steig, E. J., and Noone, D.: Last Millennium Reanalysis with an expanded proxy database and seasonal proxy modeling, *Clim. Past*, 15, 1251–1273, <https://doi.org/10.5194/cp-15-1251-2019>, 2019.
- Thompson, D. M., Conroy, J. L., Konecky, B. L., Stevenson, S., DeLong, K. L., McKay, N., Reed, E., Jonkers, L., and Carré, M.: Identifying Hydro-Sensitive Coral $\delta^{18}\text{O}$ Records for Improved High-Resolution Temperature and Salinity Reconstructions, *Geophys. Res. Lett.*, 49, e2021GL096153, <https://doi.org/10.1029/2021GL096153>, 2022.

- Tierney, J. E. and Tingley, M. P.: A Bayesian, spatially-varying calibration model for the TEX₈₆ proxy, *Geochim. Cosmochim. Ac.*, 127, 83–106, <https://doi.org/10.1016/j.gca.2013.11.026>, 2014.
- Tierney, J. E. and Tingley, M. P.: BAYSPLINE: A New Calibration for the Alkenone Paleothermometer, *Paleoceanogr. Paleocl.*, 33, 281–301, <https://doi.org/10.1002/2017PA003201>, 2018.
- Tierney, J. E., Malevich, S. B., Gray, W., Vetter, L., and Thirumalai, K.: Bayesian Calibration of the Mg/Ca Paleothermometer in Planktic Foraminifera, *Paleoceanogr. Paleocl.*, 34, 2005–2030, <https://doi.org/10.1029/2019PA003744>, 2019.
- VandenHeuvel, D.: DanielVandH/ProfileLikelihood.jl: v0.2.1, Zenodo [code], <https://doi.org/10.5281/zenodo.7493449>, 2022.
- Walter, R. M., Sayani, H. R., Felis, T., Cobb, K. M., Abram, N. J., Arzey, A. K., Atwood, A. R., Brenner, L. D., Dassié, É. P., DeLong, K. L., Ellis, B., Emile-Geay, J., Fischer, M. J., Goodkin, N. F., Hargreaves, J. A., Kilbourne, K. H., Krawczyk, H., McKay, N. P., Moore, A. L., Murty, S. A., Ong, M. R., Ramos, R. D., Reed, E. V., Samanta, D., Sanchez, S. C., Zinke, J., and the PAGES CoralHydro2k Project Members: The CoralHydro2k database: a global, actively curated compilation of coral $\delta^{18}\text{O}$ and Sr/Ca proxy records of tropical ocean hydrology and temperature for the Common Era, *Earth Syst. Sci. Data*, 15, 2081–2116, <https://doi.org/10.5194/essd-15-2081-2023>, 2023.
- Watanabe, T. K. and Pfeiffer, M.: A Simple Monte Carlo Approach to Estimate the Uncertainties of SST and $\delta^{18}\text{O}_{\text{sw}}$ Inferred From Coral Proxies, *Geochem. Geophys. Geosyst.*, 23, e2021GC009813, <https://doi.org/10.1029/2021GC009813>, 2022.
- York, D.: Least-squares fitting of a straight line, *Can. J. Phys.*, 44, 1079–1086, <https://doi.org/10.1139/p66-090>, 1966.
- York, D.: Least-squares fitting of a straight line with correlated errors, *Earth Planet. Sci. Lett.*, 5, 320–324, [https://doi.org/10.1016/S0012-821X\(68\)80059-7](https://doi.org/10.1016/S0012-821X(68)80059-7), 1968.