Clim. Past Discuss., https://doi.org/10.5194/cp-2018-87
Manuscript under review for journal Clim. Past
Discussion started: 16 August 2018

Climate of the Past Discussions
Open Access
EGU

Combining a pollen synthesis and climate simulations for spatial reconstructions of European climate using Bayesian modelling

Nils Weitzel[1], Andreas Hense[1], and Christian Ohlwein[1]

[1]Meteorological Institute, University of Bonn, Auf dem Hügel 20, 53121 Bonn, Germany

Correspondence: Nils Weitzel (nils.weitzel@uni-bonn.de)

Abstract. Probabilistic spatial reconstructions of past climate states are valuable to quantitatively study the climate system under different forcing conditions because they combine the information contained in a proxy synthesis in a comprehensible product. Unfortunately, they are subject to a complex uncertainty structure due to complicated proxy-climate relations and sparse data, which makes interpolation between samples difficult. Bayesian hierarchical models feature promising properties to handle these issues like the possibility to include multiple sources of information and to quantify uncertainties in a statistically rigorous way.

We present a Bayesian framework that combines a network of pollen samples with a spatial prior distribution estimated from a multi-model ensemble of climate simulations. The use of climate simulation output aims at a physically reasonable spatial interpolation of proxy data on a regional scale. To transfer the pollen data into (local) climate information, we apply a forward version of the probabilistic indicator taxa model. The Bayesian inference is performed using Markov chain Monte Carlo methods following a Metropolis-within-Gibbs strategy.

We reconstruct mean temperature of the warmest and mean temperature of the coldest month during the mid-Holocene in Europe using a published pollen and macrofossil synthesis in combination with the Paleoclimate Modelling Intercomparison Project Phase III mid-Holocene ensemble. The output of our Bayesian model is a spatially distributed probability distribution that facilitates quantitative analyses which account for uncertainties. Our reconstruction performs well in cross-validation experiments and shows a reasonable degree of spatial smoothing.

1 Introduction

Spatial or climate field reconstructions of past near surface climate states combine information from proxy samples, which are mostly localized, with a model for interpolation between those samples. They are valuable for comparisons of the state of the climate system under different external forcing conditions, because they produce a comprehensible product containing the joint information in a proxy synthesis. Thereby, spatial reconstructions are more suitable for many quantitative analyses of past climate than individual proxy records. Unfortunately, spatial reconstructions are subject to a complex uncertainty structure due to uncertainties in the proxy-climate relation and the sparseness of available proxy data which leads to additional interpolation uncertainties. Therefore, a meaningful reconstruction has to include these uncertainties (Tingley et al., 2012). A natural way to represent uncertainties in the proxy-climate relation are so called probabilistic transferfunctions (Ohlwein and Wahl, 2012). To

Clim. Past Discuss., https://doi.org/10.5194/cp-2018-87
Manuscript under review for journal Clim. Past
Discussion started: 16 August 2018

account for the uncertainties due to sparseness of proxy data, we suggest the use of stochastic interpolation techniques. Most standard geostatistical methods like kriging or Gaussian modelling with Matérn covariances are designed for interpolation in data rich situations, while in paleoclimatology we deal with sparse data. Therefore, their direct application to paleo situations is not suitable. Instead, we propose to use interpolation schemes that contain additional physical knowledge, such that the resulting product combines the information from a proxy network in a physically reasonable way (Gebhardt et al., 2008). As will be shown, our approach can in addition be used for structural extrapolation.

We use Bayesian hierarchical modelling to combine the two modules mentioned above: The (local) proxy-climate relation and spatial interpolation. The Bayesian framework allows the combination of multiple data types. In our case, these are pollen records to constrain the local climate, and climate simulations, which produce physically consistent spatial fields for a given set of large scale external forcings. In addition, our framework accounts for several sources of uncertainty in a statistically rigorous way by estimating and inferring a multivariate probability distribution, the so-called posterior distribution (Gelman et al., 2013).

Pollen are the terrestrial proxy with the highest spatial coverage (Bradley, 2015), and there is a long tradition of using them for inferring past climate by applying statistical transferfunctions (Birks et al., 2010). In recent years, several traditional transferfunctions like indicator taxa, modern analogues, and weighted averaging have been translated to Bayesian frameworks (e.g., Kühl et al., 2002; Haslett et al., 2006; Holden et al., 2017). Pollen records contain information on the local climate during a time slice, where the spatial scale is constrained by the influx domain of horizontal pollen transport. Equilibrium simulations with earth system models (ESMs) produce a physically consistent estimate of the atmospheric and oceanic circulation and the regional energy balance given a set of forcings (boundary conditions). Important boundary conditions, for which information are available from proxy data and physical models, are insolation determined by the earth orbital parameters, greenhouse gas concentrations, ice sheet configurations, and land-sea masks. We use an ensemble of simulations from different ESMs to estimate a prior distribution, which contains a wide range of physically reasonable climate states. The combination of these two sources of information can be interpreted as a downscaling of forcing conditions via ESMs and an upscaling of local information contained in pollen records via spatial covariance matrices. The result is a spatially distributed and physically reasonable probabilistic climate reconstruction on continental domains.

We apply our framework to a mid-Holocene (MH, around 6ka) example for two reasons. First, compared with other time slices before the common era, the MH has a high proxy data coverage, particularly for Europe. Therefore, we can use raw pollen and macrofossil data with a sparse but relatively uniform spatial coverage over Europe as input for probabilistic transferfunctions, while still having other reconstructions available, that can be compared with our results. Second, a multi-model ensemble of climate simulations with boundary conditions adjusted to the MH was produced in the Paleoclimate Modelling Intercomparison Project Phase III (PMIP3) project (Braconnot et al., 2011). This ensemble is used to estimate the spatial prior distribution. The posterior distribution, which we estimate, is a multivariate probability distribution, with marginal distributions for each grid box, as well as spatial correlations and correlations between two climate variables, the mean temperature of the warmest month (MTWA) and the mean temperature of the coldest month (MTCO). For further analyses, we create samples from this distribution, such that each sample is an equally probable estimate of the bivariate spatial field. In the context of

Clim. Past Discuss., https://doi.org/10.5194/cp-2018-87
Manuscript under review for journal Clim. Past
Discussion started: 16 August 2018

temporal reconstructions these samples were called "climate histories" by Parnell et al. (2016). From the samples, quantitative properties of the climate state during the MH, which account for uncertainties, are computed. In addition, our framework can be used to compare the model-data mismatch of multiple ESMs, to analyse the consistency of a given proxy network, and to help in the identification of potential outliers.

This work is related to several concepts that were developed for applications in paleoclimatology. In recent years, several authors constructed Bayesian hierarchical models (BHMs) for paleoclimate reconstructions: Tingley and Huybers (2010) introduced a spatio-temporal BHM for reconstructions of the last millennium with an underlying structure that is stationary, linear, and Gaussian. Other authors developed temporal (Li et al., 2010; Parnell et al., 2015) or small-scale spatio-temporal BHMs (Holmström et al., 2015). All of these approaches differ from our model in being purely proxy data driven. Additional information on orbital configurations were incorporated by Gebhardt et al. (2008) and Simonis et al. (2012) via an advection-diffusion model which is combined with proxy data using a variational inference approach. Li et al. (2010) included information on solar, greenhouse gas, and volcanic forcing for spatially averaged reconstructions of the last millennium via linear regression. Annan and Hargreaves (2013) combined Paleoclimate Modelling Intercomparison Project Phase II (PMIP2) simulations with the syntheses of Bartlein et al. (2011) and MARGO Project Members (2009) in a multi-linear regression model. We build on these approaches by incorporating fields that are simulated from a set of MH forcing conditions in a fully Bayesian framework. A different approach to combine proxy data and climate simulations for spatio-temporal reconstructions of the common era was developed by Steiger et al. (2014) and Dee et al. (2016) using so called off-line data assimilation methods. They apply an ensemble Kalman filter, where the observation operators are forward models for proxy data, and the prior covariance is estimated from a database of transient climate simulations. Our purely spatial reconstructions can be interpreted as an off-line data assimilation with only one time step. This reduced dimensionality permits the exploration of the full posterior distribution despite incorporating non-linear and non-Gaussian elements in the observation operator and a multi-modal spatial prior distribution estimated from a multi-model ensemble.

The structure of the paper is as follows. In Sect. 2, we describe the pollen synthesis and climate simulations which we use. This is followed by a detailed description of our proposed Bayesian framework in Sect. 3. Results from applying our methodology to the data are presented in Sect. 4. Finally, we discuss and summarize our methodology and results in Sect. 5 and 6.

2 Data

2.1 Proxy data

The pollen and macrofossil synthesis, that we use in this study, stems from Simonis et al. (2012) as part of the European Science Foundation project DECVeg (Dynamic European Climate-Vegetation impacts and interactions). In the following, we refer only to pollen instead of pollen and macrofossils for linguistic simplicity. Out of the four time slices (6ka, 8ka, 12ka, 13ka), which were compiled, we only use the 6ka dataset because there is no ensemble of climate simulations available for the other three time slices. For 50 paleosites, presence and absence information for 59 taxa are provided. The data is

Clim. Past Discuss., https://doi.org/10.5194/cp-2018-87
Manuscript under review for journal Clim. Past
Discussion started: 16 August 2018

already statistically preprocessed to remove co-occurring taxa that lead to an underestimation of uncertainty (Kühl et al., 2002; Gebhardt et al., 2008). The synthesis is optimised for the use in local climate reconstructions with different versions of the probabilistic indicator taxa model (PITM, originally called "pdf method"; Kühl et al., 2002), and is therefore well-suited for the implementation in a BHM that combines probabilistic transferfunctions with stochastic interpolation methods. The PITM model is described in detail in Sect. 3.3.

The 50 paleosites are sparsely but relatively uniformly distributed over Europe. Their locations are delimited by 6.5° W, 26.5° E, 37.5° N and 69.5° N. Compared with other recent syntheses like Bartlein et al. (2011), less records are included due to high quality control criteria. The raw pollen data and radiocarbon measurements, from which at least one was supposed to be close to 6ka, had to be available to recalculate age-depth models and ensure the use of calibrated radiocarbon dates as common time scale. Each site is assigned to the corresponding cell of a 2° by 2° grid which we use for our reconstructions. The locations of the proxy samples are depicted by black dots in Fig. 1. The full list of sites included in the synthesis can be found in Simonis et al. (2012). The list of taxa, which remain for each site after statistical preprocessing, is given in Simonis (2009).

2.2 Climate simulations

We use a multi-model ensemble of climate simulations which were run within PMIP3 with forcings adjusted to the MH. This includes changed orbital configurations and greenhouse gas concentrations (Braconnot et al., 2011). The ensemble contains all available MH simulations in the PMIP3 database (downloaded from the German Climate Computing Center (DKRZ) long term archive, available under https://cera-www.dkrz.de), which have a grid spacing of at most 2°. This constraint, which retains only the models with the smallest grid spacings, is chosen to better match the resolutions of pollen samples and simulations. The condition results in using seven model runs performed with the CCSM4, CNRM-CM5, CSIRO-Mk2-6-0, EC-Earth-2-2, HadGEM2-CC, MPI-ESM-P, and MRI-CGCM3. Properties of the included simulations are given in Table 1. The ensemble is a multi-model ensemble with common boundary conditions. The models are run to an equilibrium state (spin-up), followed by typically around 100 simulated years to minimize noise due to high frequency internal variability. Therefore, the differences within the ensemble can be interpreted as modelling uncertainties (epistemic uncertainty).

The mean summer climate expressed as MTWA (Fig. 1a) from the MH ensemble is warmer than the University of East Anglia Climatic Research Unit (CRU) 1961 to 1990 reference climatology (CRU TS v.4.01 over land, Harris et al. (2014), Harris and Jones (2017), and HadCRUT absolute over sea, Jones et al. (1999)) in large parts of Europe, especially eastern Europe and the Norwegian Sea. These areas coincide predominantly with areas of large ensemble spreads, expressed as the size of point-wise 90% credible intervals (CIs) of the prior distribution in Fig. 1c. The construction of the prior distribution is described in Sect. 3.2. The spread increases up to 10K in some areas of southern and eastern Europe, which might originate from varying change patterns of the general circulation over Europe in the models. In contrast, the MH mean winter climate measured by MTCO in Fig. 1b shows a more dispersed structure with cooling in Fennoscandia, warming in the Mediterranean and Balkan peninsula, and mixed patterns in western and central Europe. The ensemble spread is predominantly small (Fig. 1d), but increases towards northern Europe with very large inter-model differences for the Norwegian Sea and eastern Fennoscandia. Most of the grid box anomalies for MTWA as well as MTCO are not significant on a 5% level. Here, a positive anomaly is called significant if

Clim. Past Discuss., https://doi.org/10.5194/cp-2018-87
Manuscript under review for journal Clim. Past
Discussion started: 16 August 2018

the probability of the prior temperature to exceed the reference climatology is at least 0.95. Significant negative anomalies are defined accordingly. The significance estimates are calculated point-wise.

2.3 Reconstruction variables

The spatial distribution of taxa is limited by three climatic factors: Temperature during growing season and in winter, and moisture availability (Huntley, 2012). Therefore, these three factors, translated into quantitative variables, are important for climate reconstructions from pollen. For large parts of Europe, it was shown in Simonis (2009) that the PITM model (see Sect. 3.3) is well-suited for joint reconstructions of July and January temperature as measures for the warmth of growing season and cold of winter, because at least one of these two variables is a limiting factor for most taxa growing in the mid and high latitudes of Eurasia during the Holocene. Instead, testing various climate variables as indicators for moisture availability was less promising since the moisture availability is rarely a limiting factor for European taxa (Simonis, 2009). Hence, in this study, we choose MTWA and MTCO as the target variables for our climate reconstructions. This is a compromise between variables that are bioclimatically meaningful and variables for which accurate data is available to calibrate the transferfunctions against modern vegetation and climate data. In the mid to high latitudes, MTWA and MTCO are highly correlated with July and January temperature, respectively, which is why they are also described as "functionally equivalent" (Bartlein et al., 2011).

To calculate MTWA and MTCO from time series of monthly averages, the data is first interpolated to the desired spatial grid. Then, for each hydrological year (October to September), the warmest and coldest month are extracted. We choose the hydrological instead of the calendar year to ensure that the months are taken from connected seasons. Afterwards, the climatological mean is calculated by averaging over the values for each year.

3 Methods

3.1 Bayesian framework

We use Bayesian statistics to combine a network of pollen samples with an ensemble of PMIP3 simulations because in this approach each source of information has an associated uncertainty that is naturally included in the inference process. In this section, we specify the quantities that are combined in our reconstruction, and describe the inference algorithm that is used to create the results presented below as well as the statistical measures which we use to assess the reconstructions.

In the following, we denote fossil pollen data by P_p, past climate by C_p, modern vegetation and climate data for the calibration of transferfunctions by P_m and C_m, respectively, and additional model parameters by $\Theta := (\omega, z, \theta)$. Here, ω and z represent weights of the PMIP3 ensemble members as defined in Sect. 3.2 and 3.4, and θ are transferfunction parameters, which are specified in Sect. 3.3. We are interested in the conditional distribution of C_p and Θ given fossil pollen, modern vegetation, and modern climate data, i.e. we want to estimate the posterior distribution $\mathbb{P}(C_p, \Theta \,|\, P_p, C_m, P_m)$.

Clim. Past Discuss., https://doi.org/10.5194/cp-2018-87
Manuscript under review for journal Clim. Past
Discussion started: 16 August 2018

Climate of the Past Discussions

Applying Bayes' theorem to $\mathbb{P}(C_p, \Theta \,|\, P_p, C_m, P_m)$ (in the following, we omit normalizing constants), we get:

$$\underbrace{\mathbb{P}(C_p, \Theta \,|\, P_p, C_m, P_m)}_{\text{Posterior}} \quad \propto \quad \underbrace{\mathbb{P}(P_p, P_m \,|\, C_p, C_m, \Theta)}_{\text{Data Stage / Likelihood}} \cdot \underbrace{\mathbb{P}(C_p, C_m \,|\, \Theta)}_{\text{Process Stage}} \cdot \underbrace{\mathbb{P}(\Theta)}_{\text{Prior Stage}} . \tag{1}$$

Following Tingley and Huybers (2010), we call $\mathbb{P}(P_p, P_m \,|\, C_p, C_m, \Theta)$ the data stage, $\mathbb{P}(C_p, C_m \,|\, \Theta)$ the process stage, and $\mathbb{P}(\Theta)$ the prior stage. The structure of the Bayesian model can be expressed by a directed acyclic graph as shown in Fig. 2. In the graph, each node represents a variable and the arrows indicate dependences of variables.

We assume that the model weights ω and z are a priori independent of the transferfunction parameters θ, and that the data stage is conditionally independent of ω and z given C_p. Furthermore, by construction, P_m and C_m only contribute to the reconstruction via the transferfunction parameters, i.e. they are assumed to be independent of all other quantities. Hence, we can rewrite Eq. (1) and get

$$\mathbb{P}(C_p, \Theta \,|\, P_p, C_m, P_m) \quad \propto \quad \mathbb{P}(P_m \,|\, C_m, \theta) \quad \mathbb{P}(P_p \,|\, C_p, \theta) \quad \mathbb{P}(C_p \,|\, z) \quad \mathbb{P}(z \,|\, \omega) \quad \mathbb{P}(\omega) \quad \mathbb{P}(\theta). \tag{2}$$

In paleoclimatology, the data stage is traditionally called transferfunction, which in our case is formulated in a forward way. It probabilistically models the proxy data given climate variables and is assumed to act locally, i.e. given the climate at location x, a pollen sample at x is conditionally independent of the climate and proxy data at all other locations. A more rigorous formulation is given in Sect. 3.3.

The process stage stochastically interpolates the local climate information from the proxy data to a spatial domain. In our model, this interpolation is performed using a mixture of Gaussian distributions calculated from the ensemble of PMIP3 simulations, which is described in detail in Sect. 3.2. Note that our approach is not restricted to interpolation between proxy samples, but allows structural extrapolation through the eigenvectors of the covariance matrix and the weights of the ensemble members. Hence, we can infer climate beyond the domain of the proxy synthesis. The prior stage defines prior distributions for the model parameters Θ. These prior distributions are necessary to get a closed Bayesian model which ensures that the posterior is a valid probality distribution (Gelman et al., 2013).

3.2 Preprocessing of PMIP3 simulations

We model the process stage in Eq. (1) and the prior distributions for ω and z based on the ensemble of PMIP3 simulations described in Sect. 2.2 following the ensemble kernel dressing approach of Schölzel and Hense (2011). Six steps are applied successively:

1. All simulations are projected to a 2° by 2° grid using bilinear interpolation.
2. We calculate the climatological means for MTWA and MTCO from the full timeseries of each simulation as described in Sect. 2.3. This minimizes high frequency internal variability which cannot be resolved by the pollen data because it is a time integrated quantity (Annan and Hargreaves, 2013).
3. The resulting climatologies from step two define the means μ_k, $k = 1, ..., K$, of the mixture components, where K is the number of ensemble members. Each mixture component is assumed to be multivariate Gaussian. The dimension N of

Clim. Past Discuss., https://doi.org/10.5194/cp-2018-87
Manuscript under review for journal Clim. Past
Discussion started: 16 August 2018

each Gaussian vector is either the number of grid boxes in the case of single variable reconstructions or twice the number of grid boxes in the case of joint reconstructions of MTWA and MTCO.

4. Following the kernel approach in Silverman (1986), we calculate the covariance matrix $\tilde{\Sigma}_{\text{prior}}$ as the empirical covariance of the inter-model differences scaled by the Silverman factor

$$f := \left(\frac{4}{K \cdot (N+2)}\right)^{\frac{2}{N+4}}. \tag{3}$$

Hence, $\tilde{\Sigma}_{\text{prior}}$ is given by

$$\tilde{\Sigma}_{\text{prior}} = f \cdot \frac{1}{K-1} \sum_{k=1}^{K} (\mu_k - \bar{\mu}) (\mu_k - \bar{\mu})^t, \tag{4}$$

where $\bar{\mu}$ is the mean over all mixture components, and the superscript t denotes the matrix transpose.

5. To get a non-singular covariance matrix and avoid spurious correlations, we apply the graphical lasso algorithm (Friedman et al. (2008), implemented in the R-package glasso, https://cran.r-project.org/package=glasso) to $\tilde{\Sigma}_{\text{prior}}$. This algorithm approximates the precision matrix (inverse covariance) by a positive definite, symmetric, and sparse matrix $\Sigma_{\text{prior}}^{-1}$. Therefore, Σ_{prior} is a valid N-dimensional covariance matrix that we use as covariance matrix of the mixture components. The sparseness of the precision matrix has the additional advantage that computationally efficient Gaussian Markov random field techniques (Rue and Held, 2005) can be used, which reduces the computational burden significantly.
Glasso maximizes the penalized log-likelihood

$$\log\det \Sigma_{\text{prior}}^{-1} - \text{trace}(\tilde{\Sigma}_{\text{prior}} \Sigma_{\text{prior}}^{-1}) - \rho \|\Sigma_{\text{prior}}^{-1}\|_1, \tag{5}$$

where ρ is the penalty parameter, $\|\cdot\|_1$ is the vector L_1-norm, and the first two terms are the Gaussian log-likelihood. To determine ρ, we first recognize that for values smaller than $\rho = 0.3$ the resulting matrices become numerically unstable in our application due to the small ensemble size. Five values for ρ were tested: 0.3, 0.5, 0.7, 1.0, and 2.0. Larger values lead to sparser precision matrices and therefore to smaller spatial correlations. This in turn means that the local information from the proxy data is spread less into space. For each of the five parameters we perform a cross-validation experiment and compare the resulting Brier scores (see Sect. 3.5). While the smallest penalty parameters have the best mean Brier scores, the differences are generally small (Fig. 8). On the other hand, the influence of the penalty term in Eq. (5) on the overall regression increases from 79.5% for $\rho = 0.3$ to 98.5% for $\rho = 2.0$. Based on these diagnostics, we choose $\rho = 0.3$ for the reconstructions in Sect. 4 to get a numerically stable covariance which performs at least as good as other choices of ρ in cross-validations, and is comparably little influenced by the penalty term. The sensitivity of the reconstruction with respect to ρ is further studied in Sect. 4.4. Note that for other applications of our framework different values of ρ can produce the best results.

6. The result of the previous steps is a mixture distribution

$$\mathbb{P}(C_p \,|\, \omega, \mu_1, ..., \mu_K, \Sigma_{\text{prior}}) = \sum_{k=1}^{K} \omega_k \, \mathcal{N}(C_p \,|\, \mu_k, \Sigma_{\text{prior}}), \tag{6}$$

Clim. Past Discuss., https://doi.org/10.5194/cp-2018-87
Manuscript under review for journal Clim. Past
Discussion started: 16 August 2018

where $\omega = (\omega_1, ..., \omega_K)$ are the prior weights.

We define a Dirichlet distributed prior for ω with parameter $\frac{1}{2}$ for each of the K components. This corresponds to an objective prior (Jeffrey's prior; Gelman et al., 2013) in the Dirichlet-multinomial model. The variable z is added as a help parameter, which simplifies the inference algorithm as described in Sect. 3.4.

3.3 Transferfunctions

The Bayesian model uses probabilistic transferfunctions to model the pollen-climate relation. From all the terms in Eq. (2), the calibration layer $(P_m \,|\, C_m, \theta)$, the observation layer $\mathbb{P}(P_p \,|\, C_p, , \theta)$, and the prior distribution of the transferfunction parameters $\mathbb{P}(\theta)$ are related to the transferfunction (Parnell et al., 2015).

To reconstruct climate from the Simonis et al. (2012) synthesis, we need a transferfunction that converts presence-absence information on taxa to climate variables. As described above, our main reconstruction target is the bivariate climate $C = (C_1, C_2)$, where C_1 is MTWA and C_2 is MTCO. Previously, reconstructions from the Simonis et al. (2012) dataset were performed using PITM, orginially named "pdf method" (Kühl et al., 2002), which is an extension of the classical indicator species method (Iversen, 1944). PITM fits probability distributions to presence-absence maps of taxa (Schölzel et al., 2002), which are plotted in the bivariate climate space. The reference climatology for this mapping is derived from the CRU TS v.4.01 1961 to 1990 monthly averages following the definitions of MTWA and MTCO described in Sect. 2.3. Initially, Gaussian distributions were used for calibration (Kühl et al., 2002). Later, the model was extended to mixtures of Gaussians (Gebhardt et al., 2008). Taxa are treated as conditionally independent given climate. A statistical pre-selection of taxa, which are present in a sample, is applied to avoid over-fitting originating from violations of the independence assumption between taxa, i.e. due to co-occurrence of taxa (Kühl et al., 2002). This procedure uses the Mahalanobis distance (Mahalanobis, 1936) between the fitted distributions.

We reformulate PITM in a forward way using quadratic logistic regression similar to Stolzenberger (2017). For each taxa, we fit a regression model (response function) describing the probability of observing the taxa for a given value of C. The idea of using quadratic logistic regression stems from the BIOMOD (BIOdiversity MODelling) software which is a model to predict species distributions (Thuiller, 2003). The regression for taxa T contains linear and quadratic terms for each of the climate variables as well as an interaction term:

$$\mathbb{P}(T = 1 \,|\, C = (C_1, C_2)) \;=\; \text{logit}\left(\beta_1^T + \beta_2^T C_1 + \beta_3^T C_2 + \beta_4^T C_1 C_2 + \beta_5^T C_1^2 + \beta_6^T C_2^2\right). \tag{7}$$

Here, logit denotes the logistic function and $\beta_1^T, ..., \beta_6^T$ are regression coefficients. This regression leads to a unimodal response function which is anisotropic but has two symmetry axes, as can be seen for dwarf birch (Betula nana) and European ivy (Hedera helix) in Fig. 3.

For the calibration against the modern dataset, we use presence (T=1) as well as absence (T=0) information of the taxa which can be justified by assuming that the vegetation maps contain accurate information on the presence or absence of taxa. On the other hand, for the fossil pollen samples, we do not include absent taxa in the reconstruction, as the absence of a taxa in a pollen assemblage can have multiple non-climatic reasons like changing biologic competition or pollen transport effects

Clim. Past Discuss., https://doi.org/10.5194/cp-2018-87
Manuscript under review for journal Clim. Past
Discussion started: 16 August 2018

Climate of the Past Discussions

(Gebhardt et al., 2003). From the definition given in Sect. 2.3, it follows that at any location MTWA is larger or equal than MTCO. Formally incorporating this constraint in the inference leads to a non-linear condition on the regression parameters, which is very hard to implement. Therefore, we choose the more practical way of adding artificial absence information for combinations of MTWA and MTCO such that MTCO > MTWA. While this leads to transferfunctions, which do not preclude reconstructions of MTCO values larger than MTWA, it is at least very improbable.

To apply the response functions for individual taxa to a set of proxy data, we assume that proxy samples $P(s)$, where $s = 1,...,S$ subscripts the proxy samples, are conditionally independent given a climate field and that, conditioned on $C(x_s)$, where x_s is the location of the s-th sample, $P(s)$ is independent of the climate at all other locations. This leads to the following probabilistic model for the set of modern vegetation samples

$$\mathbb{P}(P_m \,|\, C_m, \theta) \;=\; \prod_{s=1}^{S_m} \prod_{T \in T(P)} \mathbb{P}\left(P_m^T(s) \,|\, C_m(x_s), \beta_1^T, ..., \beta_6^T\right). \tag{8}$$

Here, $P_m^T(s)$ is the presence or absence of taxa T in the s-th calibration sample, $T(P)$ is the set of all Taxa occurring in the fossil pollen samples, and $\theta := (\beta_i^T,\, i = 1,...,6, T \in T(P))$. $\mathbb{P}(P_p \,|\, C_p, \theta)$ is given by

$$\mathbb{P}(P_p \,|\, C_p, \theta) \;:=\; \prod_{s=1}^{S_p} \prod_{T \in T(s)} \mathbb{P}\left(P_p^T(s) \,|\, C_p(x_s), \beta_1^T, ..., \beta_6^T\right), \tag{9}$$

where $T(s)$ are the taxa occurring in sample s.

Finally, we define a prior distribution for θ. We use a Gaussian distribution centred at 0 and a marginal variance of 10 for each parameter β_i^T. Due to the absence of prior information on the correlation structure, we assume independence between the taxa as well as within a taxa. Hence, we get

$$\mathbb{P}(\theta = (\beta_i^T, i = 1,...,6, T \in T(P))) \;=\; \prod_{T \in T(P)} \prod_{i=1}^{6} \mathcal{N}\left(\beta_i^T \,|\, 0, 10\right). \tag{10}$$

Due to the high information content in the calibration data set, the influence of the prior (Eq. 10) on the parameter estimates is negligible. Using a flat prior for $C_p(x_s)$ and removing spatial correlations, local climate reconstructions at the locations of the proxy samples can be calculated. These reconstructions depend only on the proxy data in grid box x_s. Results of local MH reconstructions for each grid box with proxy data are shown in Fig. 4, where the local reconstruction means and the marginal 90% CIs are plotted for MTWA and MTCO.

Local reconstructions can also be used to evaluate the ability of the transferfunctions to reconstruct modern climate which provides a reference for possible regional biases. For the PITM model such evaluations have been performed by Gebhardt et al. (2008) and Stolzenberger (2011). Both evaluations show that the model tends to underestimate north-south gradients leading to positive biases in Fennoscandia, and slightly negative biases in the Mediterranean. The biases as well as the uncertainties are larger for winter temperature than for summer. Therefore, results for MTCO in northern Fennoscandia should be treated with caution, while for all other regions biases of the reconstruction means are within reconstruction uncertainties.

Clim. Past Discuss., https://doi.org/10.5194/cp-2018-87
Manuscript under review for journal Clim. Past
Discussion started: 16 August 2018

Climate of the Past Discussions

3.4 Inference and computational performance

Because the PITM model is non-Gaussian and non-linear, the posterior climate follows again a mixture distribution, but the components do not belong to a standard probability distribution, that could be used for sampling. Therefore, we use Markov chain Monte Carlo (MCMC) techniques to asymptotically sample from the correct posterior distribution. These samples allow analyses beyond summary statistics like means and standard deviations. We choose a Metropolis-within-Gibbs strategy, which means that we sample alternately from the full conditional distributions (i.e. the distribution of the respective variable given all other variables) of θ, ω, z, and C_p.

To sample the regression parameters θ in Eq. (7) to (9) efficiently, the data augmentation scheme of Polson et al. (2013) is used. For taxa T, the full conditional is only depending on C_p, C_m, P_p^T, and P_m^T, but not on other taxa. Therefore, we can sample $\beta_1^T, ..., \beta_6^T$ independently from the other taxa. Polson et al. (2013) introduce help variables $\gamma_l^T, l = 1, ..., L$, where L is the number of observations (absence and presence) of taxa T, such that $\mathbb{P}(\gamma_l^T \,|\, \beta_1^T, ..., \beta_6^T, C_m, C_p)$ is Pólya-Gamma (PG) distributed, and $\mathbb{P}(\beta_1^T, ..., \beta_6^T \,|\, P_m^T, P_p^T, \gamma_1^T, ..., \gamma_L^T)$ is Gaussian. Therefore, we sample alternately from a PG distribution using the sampler of Windle et al. (2014) and from a multivariate Gaussian. The PG sampler is implemented in the R package BayesLogit (Windle et al., 2013).

An easy way to sample from the climate mixture distribution is to introduce a multinomially distributed augmentation variable $z = (z_1, ..., z_K)$ such that

$$\mathbb{P}(\omega) = \text{Dirichlet}\left(\omega_1, ..., \omega_K \,|\, \tfrac{1}{2}, ..., \tfrac{1}{2}\right) \tag{11}$$

$$\mathbb{P}(z \,|\, \omega) = \text{Multinomial}\left(z_1, ..., z_K \,|\, n = 1, \omega_1, ..., \omega_K\right) \tag{12}$$

$$\mathbb{P}(C_p \,|\, z) = \prod_{k=1}^{K} \left(\mathcal{N}(C_p \,|\, \mu_k, \Sigma_{\text{prior}})\right)^{z_k}, \tag{13}$$

i.e. C_p, conditioned on z selecting model k, is Gaussian. Integrating out z yields the mixture distribution Eq. (6). Equations (11) to (13) lead to full conditionals for ω and z, which are again Dirichlet and multinomially distributed but with updated parameters. Therefore, we can use Gibbs sampling to update ω and z. Preliminary tests revealed that in our application the prior distribution of ω has a negligible influence on the posterior distributions of z and C_p.

To sample from the full conditional of C_p, we separate the grid boxes x_P with at least one proxy record from those without any proxy records denoted by x_Q. There is no closed form available for the full conditionals of $C_p(x_P)$. Therefore, we use a random walk Metropolis-Hastings algorithm to update $C_p(x_P)$ sequentially for all members of x_P. As these updates are two-dimensional and the target distributions are in most cases close to Gaussian, the random walk proposals are very efficient. As the transferfunctions act locally, $C_p(x_Q)$ is conditionally independent of P_p given $C_p(x_P)$ and z. Therefore, we subsequently update $C_p(x_Q)$ by sampling from $\mathbb{P}(C_p(x_Q) \,|\, C_p(x_P), z)$ which is Gaussian. Detailed formulas for the full conditional distributions are given in Appendix A.

The mixture distribution Eq. (6) leads to a multimodal posterior. Due to the high dimensionality of C_p, the likelihood of choosing a new model z_k from one MCMC step to the next one is very small. This leads to very slow mixing of the MCMC algorithm described above. To speed up the mixing, we apply a Metropolis coupled MCMC strategy (MC^3), also called parallel

Clim. Past Discuss., https://doi.org/10.5194/cp-2018-87
Manuscript under review for journal Clim. Past
Discussion started: 16 August 2018

tempering, which was developed by Geyer (1991) and adapted to parallel computer architectures by Altekar et al. (2004) and Werner and Tingley (2015). We run A MCMC chains in parallel, and after every M steps, we use an additional Metropolis-Hastings step to swap the states of the Markov chains a_1 and a_2 with probability $0 < p_{a_1,a_2} < 1$, where p_{a_1,a_2} is calculated from the Metropolis-Hastings odds ratio. The Markov chains are created by exponentiating the process stage and the data stage by constants $\nu_1 = 1 > ... > \nu_A > 0$. The first Markov chain ($\nu_1 = 1$) asymptotically retains the original posterior distribution for all variables, whereas the subsequent chains sample from a flatter posterior distribution, in which it is easier to jump from one mixture component to another. Following empirical testing, we run the European reconstructions with $A = 8$ parallel chains, levels $\nu_1 = 1, \nu_2 = 1.25^{-1}, ..., \nu_8 = 1.25^{-7}$, and swaps after every $M = 30$ steps.

As a second strategy to speed up the inference, we treat the grid boxes with proxy data and those without proxy data sequentially. Because of the Gaussian mixture components in the process stage and because the grid boxes without proxy data are not influencing the posterior model weights, we can first integrate out $C_p(x_Q)$ and get an estimate of the joint distribution of ω, z, and $C_p(x_P)$. In a second step, we sample from $C_p(x_Q)$ conditioned on $C_p(x_P)$ and z, which leads to joint samples of C_p from the asymptotically correct posterior distribution. In practice, this is done by drawing a sample of $C_p(x_Q)$ conditioned on each of the MCMC samples from $(C_p(x_P), z, \omega, \theta)$. This strategy reduces the number of MCMC chains needed in the MC^3 algorithm, and it reduces the computation time of each MCMC update due to faster matrix operations. Pseudo-code for the MC^3 algorithm is given in Appendix B.

The remaining bottleneck in computation time is the estimation of the transferfunction parameters due to the large modern calibration set. While in theory the observation layer influences the updates of θ, in practice the influence of Eq. (9) on the posterior of θ is negligible. Therefore, we use a modularization approach (Liu et al., 2009) similar to Parnell et al. (2015) in cross-validation experiments, where a sequence of reconstructions with slightly changed proxy networks is computed (see Sect. 3.5). This means that we cut feedback between Eq. (8) and Eq. (9) by first drawing as many MCMC samples as necessary from θ using only Eq. (10) and Eq. (8). Thereafter, we reconstruct C_p using these samples instead of sampling θ from its full conditional.

For a 798 dimensional climate posterior as it is the case in joint reconstructions of MTWA and MTCO, and 45 grid boxes that contain at least one proxy record, we create 75,000 MCMC samples for each of eight parallel chains, where every chain runs on one central processing unit (CPU). The first 25,000 samples are discarded as burnin. To reduce the autocorrelation of subsequent samples, we extract every fifth sample to create a set of 10,000 posterior samples which is used for further analyses. On a standard desktop computer with at least eight CPUs, reconstructions with the modularized model can be computed in approximately 30 minutes. The convergence of all MCMC variables is checked using the Gelman-Rubin-Brooks criterion (Brooks and Gelman, 1998) implemented in the R package coda (Plummer et al., 2006).

3.5 Assessing the added value of reconstructions

To analyse the added value of combining proxy data and PMIP3 simulations compared to only using PMIP3 simulations and to assess the consistency of the reconstruction, we perform leave-one-out cross validation. Due to the sparseness of the proxy

Clim. Past Discuss., https://doi.org/10.5194/cp-2018-87
Manuscript under review for journal Clim. Past
Discussion started: 16 August 2018

network, we do not leave out larger amounts of data. In addition, the cross-validations can be used to identify systematic mismatches between proxy data and simulations as well as potential outliers in the data.

The cross-validation is performed in the observation space and consists of three steps:

1. A reconstruction with all proxy samples except for those in grid box x is computed.
2. The forward model Eq. (9) is applied to the posterior and the prior distribution of $C_p(x)$ to get two probabilistic predictions for each proxy sample $P_p(s)$ with $x_s = x$.
3. Proper scores (Gneiting and Raftery, 2007) are computed for both probabilistic predictions. Then, the corresponding skill score, which is a measure of the added value from constraining the PMIP3 ensemble by pollen data, is calculated.

The PITM forward model maps a climate field to a binary field (presence or absence of a taxa), where the probabilistic prediction is represented by the probability of presence $p \in [0,1]$. A common proper score function for binary variables is the Brier score (BS; Brier, 1950) given by

$$BS(T) := (\delta_T(1) - p)^2 + (\delta_T(0) - (1-p))^2 = \begin{cases} 2 + 2p^2 - 4p & \text{if } T = 1 \\ 2p^2 & \text{if } T = 0 \end{cases} \tag{14}$$

where δ_T denotes the indicator function of taxa T. The BS takes values between zero and two, where zero corresponds to a perfect prediction and two to the worst possible prediction. In a recent study, Stolzenberger (2017) used the BS to assess the skill of PMIP3 simulations for the MH using a network of pollen records. To calculate the BS for a set of climate samples, either MCMC samples from the posterior or independent samples from the prior mixture distribution, the PITM forward model is applied to each sample which leads to a set of probabilistic predictions $p_j(T)$, $j = 1, ..., J$ for taxa T. Predictions are calculated for each taxa which occurs in sample $P(s)$. The joint score of $P(s)$ is then calculated by averaging the BS of each taxa and prediction:

$$BS(P(s)) := \frac{1}{|T(s)|\, J} \sum_{T \in T(s)} \sum_{j=1}^{J} \left(2 + 2p_j(T)^2 - 4p_j(T)\right). \tag{15}$$

If multiple samples are assigned to one grid box, the mean score of those samples is taken. Finally, the Brier skill score (BSS) is computed by comparing BS for posterior and prior:

$$\text{BSS} := \frac{\text{BS(Prior)} - \text{BS(Posterior)}}{\text{BS(Prior)}}. \tag{16}$$

4 Results

In this section, results from the MH reconstruction for Europe with the Simonis et al. (2012) synthesis and the PMIP3 MH ensemble are presented. We study the mean and uncertainty structure of the posterior distribution, the added value of the reconstruction, the performance of different ensemble members, the sensitivity with respect to the glasso penalty parameter, and the results of joint compared to separate MTWA and MTCO reconstructions. Results from all reconstructions are summarized in Table 1.

Clim. Past Discuss., https://doi.org/10.5194/cp-2018-87
Manuscript under review for journal Clim. Past
Discussion started: 16 August 2018

4.1 Posterior mean and uncertainty structure

The spatially averaged mean temperature of the reconstruction (posterior mean) is 17.77°C (90% CI: (17.44°C, 18.09°C)) for MTWA and 2.24°C (90% CI: (1.87°C, 2.62°C)) for MTCO, which is in the former case equal to the CRU reference climatology (+0.01 K) and in the latter 1.11 K warmer. Larger differences are found for subregions (Fig. 5a, 5b): For MTWA as well as MTCO, cooler temperatures than today are found in most areas south of 54° N, while north of this line the temperatures are predominantly higher than today with the exception of a cooling in parts of Fennoscandia. The on average higher MTCO anomalies compared to MTWA stem from higher anomalies in Fennoscandia. Many of the warming anomalies in the northern part as well as the cooling anomalies in the southern part of the domain are significant on a 5% level (see Sect. 2.2 for the definition of significance in the Bayesian context). In particular, using the joint information in the pollen synthesis and combining it with the spatial structure of the PMIP3 ensemble leads to more significant signals than in any of the individual data products.

Most of the taxa, which are used in the reconstruction, are stronger confined for MTWA than for MTCO because the growth of most European plants is more sensitive to conditions during the growing season. This results in more constrained local MTWA reconstructions (Fig. 4a), which is in concordance with findings from Gebhardt et al. (2008). Hence, the uncertainty in the MTWA reconstruction is smaller (spatially averaged point-wise 90% CI size of 2.90 K) than in the MTCO reconstruction (spatially averaged point-wise 90% CI size of 3.18 K) (Fig. 5c, 5d). The uncertainty is smallest for grid boxes with proxy records, and highest in the north eastern and north western parts of the domain where the PMIP3 ensemble spread is large and the constraint from proxy data is weak. Besides, the reconstruction uncertainty has small spatial variations. The ratio of the CI size of spatially averaged temperatures and the spatially averaged point-wise CIs can be interpreted as a measure for the spatial degrees of freedom in the reconstruction.

The highest reduction of uncertainty due to the inclusion of proxy data is found at grid boxes with proxy data, as quantified by a spatially averaged reduction of point-wise CI sizes from prior to posterior of 69.5% compared to 60.5% for grid boxes without proxy data (Fig. 5e, 5f). The uncertainty reduction for MTWA is higher for terrestrial grid boxes than marine ones, but the smaller PMIP3 ensemble spread over sea leads to similar CI sizes. For MTCO, the reduction of uncertainty is highest for terrestrial areas and the Norwegian Sea and lowest for the Mediterranean Sea and off the French and Iberian coast. But since the PMIP3 ensemble spread is small in these areas, the reconstruction uncertainty is still lower than in some land parts of the domain.

To study whether the degree of spatial smoothing of the reconstruction is reasonable, we calculate a measure inspired by discrete gradients. For each grid box, we calculate the mean absolute difference between the value in the box and its eight nearest neighbours. Then, we compare the spatial averages of this homogeneity measure H in the posterior, the climatologies of the PMIP3 ensemble members and the reference climatology. We expect a reconstruction with a good degree of smoothing to have similar spatial homogeneity than the PMIP3 ensemble and the reference climatology, as H depends mainly on local features like orography or land-sea contrasts, and we expect these features to affect the local climate of the MH similarly than today's climate since reconstructions suggest only small changes in topography between the MH and today. For MTWA, we

Clim. Past Discuss., https://doi.org/10.5194/cp-2018-87
Manuscript under review for journal Clim. Past
Discussion started: 16 August 2018

get a posterior mean of 1.48 K (90% CI: (1.40 K, 1.57 K)), which is in agreement with 1.39 K for the reference climatology and values between 1.08 K and 1.54 K for the PMIP3 climatologies. The heterogeneity of MTCO is higher than of MTWA, but again, the mean posterior value of 2.18 K (90% CI: (2.09 K, 2.27 K)) is in concordance with the reference climatology (2.02 K) and the PMIP3 climatologies (between 1.89 K and 2.41 K). From these results, we deduce that the posterior has a reasonable degree of spatial smoothing.

4.2 Added value of the reconstruction

The skill of the reconstruction is measured using the BSS, defined in Eq. (16), in cross-validation experiments. For positive BSS values, the posterior distribution is superior to the prior, which means that constraining the PMIP3 ensemble by proxy data adds value to the reconstruction. On the other hand, the posterior distribution is inferior to the prior for negative values. This would indicate inconsistencies in the local proxy reconstructions, a scaling mismatch of simulations and data, or the existence of spurious correlations in the prior covariance matrix.

For most left-out proxy samples the BSS is positive (80% of grid boxes with left-out pollen data) with a median of 0.39 (Fig. 6). The BSS values are predominantly positive for all regions but the British Islands and Norway. This indicates a high consistency of the reconstruction in large parts of the domain due to a good fit of the local reconstructions with the spatial correlation structure. In particular, we see a consistent MTWA cooling south of 54° N in the local reconstructions compared to the prior distribution. This leads to a mean cooling and reduction of uncertainty in the posterior which is transported to grid boxes without proxy data, including those with left-out proxy data, via spatial correlations and higher weights for cooler ensemble members. Similarly, the consistent MTWA and MTCO warming of the local reconstructions in the north-eastern part of the domain lead to positive BSS values.

The persistent negative BSS values for the British Islands and Norway cannot be explained by data outliers but warrant a more systematic issue. For these two regions, the uncertainty in the local reconstructions is larger than for other areas (Fig. 4), such that the local proxy records constrain the posterior less than the posterior ensemble member weights and some of the more distanced proxy records. The spatial correlations and degenerated weights (see Sect. 4.3) lead to a reduction of the posterior uncertainty compared to the unconstrained PMIP3 ensemble (Fig. 5e, 5f) but without improving the concordance of the mean state with the local reconstructions, which in turn results in negative BSS values. In and near the Alps, BSS values are near zero indicating no added value from constraining the PMIP3 ensemble by proxy data. This might be a result of not accounting enough for orographic effects in the different sources of information.

4.3 Posterior ensemble member weights

The posterior weights of the PMIP3 ensemble members are determined by ω and z. The posterior of ω is a combination of its prior and the distribution of model choices z. As we defined a Jeffrey's prior and choose only one ensemble member in each MCMC step, the posterior of ω is a flattened version of the posterior of z (Fig. 7). z is more relevant for the posterior of the state C_p as it determines which ensemble member is chosen in each MCMC step. The posterior of z combines the distribution

Clim. Past Discuss., https://doi.org/10.5194/cp-2018-87
Manuscript under review for journal Clim. Past
Discussion started: 16 August 2018

of ω and the likelihood of C_p for each of the ensemble members (see Appendix A). The latter term is more informative and therefore more important for determining z.

In our reconstruction, the posterior weights are dominated by the MPI-ESM-P climatology, with a posterior mean of z of 0.98 (blue diamonds in Fig. 7). The degeneracy of weights can be explained by large differences between PMIP3 climatologies. Because there is less uncertainty in the local MTWA reconstructions, it is the major variable for determining the posterior ensemble member weights. Among all included models, the MPI-ESM-P simulation is closest to the dipole structure with MTWA warming in northern and cooling in southern Europe, which explains the high model weight.

The prior climate state is changed in two ways during the inference: The ensemble member weights are updated, and each of the mixture components is modified according to the proxy data. By comparing the posterior with the prior and the local reconstructions, it can be seen that for most terrestrial areas the posterior mean resembles the local reconstructions more than the PMIP3 ensemble mean. This shows that the uncertainty in the prior distribution is large enough to lead to a reconstruction which is mostly determined by proxy data wherever available. The main exceptions are the British Islands and Norway, where the local reconstructions are not informative enough to strongly affect the joint signal from posterior ensemble member weights and farther away proxy samples. The aforementioned changes of the prior lead to a posterior mean which is cooler than the prior mean for MTWA for most of the domain except northern areas. For MTCO, the posterior mean is much warmer in northern Europe than the prior mean and slightly cooler in southern Europe.

4.4 Sensitivity with respect to the glasso penalty parameter

As described in Sect. 3.2, the glasso algorithm, which regularizes the covariance matrix estimated from the PMIP3 ensemble, requires the specification of a penalty parameter ρ. Values of at least 0.3 produce numerically stable matrices. Larger values lead to smaller correlations, and therefore higher posterior uncertainties due to less spatial transfer of information from the proxy data. To test the influence of ρ on the reconstruction results, we compute reconstructions and cross-validations for $\rho = 0.3, 0.5, 0.7, 1.0, 2.0$.

The fraction of the penalty term on the total cost given by Eq. (5) increases from 79.5% for $\rho = 0.3$ to 98.5% for $\rho = 2.0$ (Fig. 8) showing the generally high influence of the penalty term due to the small ensemble size and the increasing importance for larger ρ. While the mean BS is lowest for $\rho = 0.3$ with 0.37 and increases with ρ, the differences are small in total with 0.44 being the highest value for $\rho = 2.0$ (Fig. 8). The mean value of the posterior climate and the posterior ensemble member weights are insensitive to changes in ρ with differences in the mean spatial average of at most 0.3 K (Fig. 8) as well as no substantial regional differences, and mean z values for MPI-ESM-P of more than 0.98 in all cases. On the other hand, the posterior uncertainty grows substantially for larger ρ values. While the spatially averaged size of point-wise 90% CIs is 3.03 K for $\rho = 0.3$, the value increases to 6.34 K for $\rho = 2.0$ (Fig. 8).

4.5 Joint versus separate MTWA and MTCO reconstructions

To compare the effect of reconstructing MTWA and MTCO jointly compared to separately, additional reconstructions with only one climate variable are computed. Note that the interactions of MTWA and MTCO are twofold in the joint reconstruction:

Clim. Past Discuss., https://doi.org/10.5194/cp-2018-87
Manuscript under review for journal Clim. Past
Discussion started: 16 August 2018

(a) the response functions have an interaction term in the logistic regression Eq. (7), and (b) the process stage Eq. (6) contains joint ensemble member weights for MTWA and MTCO as well as inter-variable correlations in the covariance matrix.

The separate MTWA reconstruction is warmer than the joint reconstruction, with the spatially averaged posterior mean being 0.63 K warmer. On the other hand, the spatially averaged posterior mean of the separate MTCO reconstruction is 1.05 K cooler. Hence, the seasonal difference is smaller in the joint reconstruction, due to smoothing from the PMIP3 ensemble and slightly positive correlations between MTWA and MTCO in most of the joint local reconstructions. The MTWA only reconstruction is warmer in most land areas, with largest differences in southern Europe, but most of the differences are not significant on a 5% level (Fig. 9a). As this part of the domain is best constrained by proxy data, and because the posterior ensemble member weights are similar to the joint reconstruction (mean z of 0.93 for MPI-ESM-P, Fig. 7), it is likely that the additional warming is due to the missing interaction in the transferfunction. On the other hand, the posterior ensemble member weights are very different for the separate MTCO reconstruction, with HadGEM2-CC and MRI-CGCM3 being the models with the highest weights (mean z of 0.65 and 0.34, respectively). Together with the less constrained transferfunctions for MTCO than MTWA and the missing interaction term, this leads to a cooler reconstruction for all areas but some parts of central Europe and the Mediterranean (Fig. 9b). The cooling is strongest in Scandinavia, the British Islands, the Norwegian Sea, and the Iberian peninsula. As these are the regions which are least constrained by proxy data (Fig. 4d), choosing different PMIP3 ensemble members affects the reconstruction more than in other areas. Many of the differences in Fennoscandia and the south-western part of the domain are significant (5% level), which indicates that for these areas our method might underestimate the reconstruction uncertainty.

The missing interaction of MTWA and MTCO also leads to a higher uncertainty in the separate reconstructions, in particular in areas which are not well constrained by proxy data. For MTWA, the difference is relatively small with spatially averaged only 0.2 K larger point-wise 90% CIs. The spatial average of the point-wise 90% CIs for MTCO is 0.96 K larger than in the joint reconstruction, showing again that reconstructing the variables jointly has a larger effect on MTCO than on MTWA. The larger MTCO uncertainties are mostly due to much larger CIs at the Norwegian Sea (Fig. 9d).

The sign of the BSS in the MTWA only reconstruction is mostly the same than in the joint reconstruction, but the magnitude is smaller for most grid boxes (Fig. 9e). This shows that the added value of the joint reconstruction compared to the unconstrained PMIP3 ensemble is mainly determined by the MTWA reconstruction. In particular, the negative BSS at the British Islands and Norway, which is found in the joint reconstruction, is also present in the MTWA only reconstruction but not in the MTCO only reconstruction. For most of the domain, the BSS values of the MTCO only reconstruction are small (Fig. 9f), which indicates no or only little added value from including proxy data. Particularly in central Europe, the values are very close to zero because the comparably small PMIP3 ensemble spread (Fig. 1d) is just weakly constrained by local reconstructions.

5 Discussion and possible extensions

5.1 Robustness of the reconstruction

Our approach is designed with the goal of being more suitable for sparse data situations than standard geostatistical models. In a subsequent paper, we plan to apply the method to data from the Last Glacial Maximum, for which considerably less

Clim. Past Discuss., https://doi.org/10.5194/cp-2018-87
Manuscript under review for journal Clim. Past
Discussion started: 16 August 2018

proxy samples are available than for the MH. To understand the robustness of our method with respect to the amount of data included in a proxy synthesis, we perform experiments with only half of the samples, which are either selected to retain the spatial distribution or chosen randomly. In all of the tests, the general spatial structure of the posterior distribution, including the anomaly patterns, is preserved. Only the northern part of the MTCO reconstruction, especially the Norwegian Sea, changes substantially in some experiments. While in each experiment MPI-ESM-P remains the favoured ensemble member, other members can reach weights, which are higher than in the prior. The spatially averaged temperatures as well as the spatial homogeneity H stay within the uncertainty ranges of the reference reconstruction. The uncertainty structure of all tests is similar to the reference reconstruction but the lower number of proxy records leads to an increase of the overall uncertainty. For example, the spatially averaged point-wise 90% CIs grow by up to 0.5 K. This number is still low compared to the overall uncertainty keeping in mind that the number of proxy samples is reduced by 50%. The experiments show that our method is robust with respect to the number of proxy samples as long as the remaining samples are informative and relatively uniformly distributed across space. In our example, this is not the case for the Norwegian Sea. Combining pollen records with sea surface temperature proxies could potentially overcome this issue.

The more constrained local MTWA reconstructions lead to a higher influence on the joint reconstruction than the local MTCO reconstructions. Therefore, the MTCO only reconstruction differs more from the MTCO estimate in the joint reconstruction. Reconstructing MTWA and MTCO jointly should in theory lead to a physically more reasonable reconstruction by creating samples drawn from the same ensemble member and therefore corresponding to a physically consistent MTWA and MTCO simulation. On the other hand, Rehfeld et al. (2016) show that multi-variable reconstructions can be biased when signals from a dominant variable are transferred to a minor variable. While we believe that the PITM model is less sensitive to this issue than the weighted averaging partial least squares (WA-PLS) method used in Rehfeld et al. (2016) because it better respects the larger MTCO uncertainties, it will be subject to future work to study whether joint or separate reconstructions lead to more reliable results.

The large PMIP3 ensemble spread for most grid boxes shows that the prior distribution, which is calculated from the ensemble, contains a wide range of possible states. In areas which are well constrained by proxy data, this large total uncertainty leads to a reconstruction which depends little on the climatologies of the ensemble members. Hence, in these areas the reconstruction is not much influenced by the ensemble member weights. On the other hand, the spatial correlation structure estimated from the ensemble controls the smoothness of the reconstruction and the spread of local information into space. As the correlation structure becomes more robust for larger ensemble sizes, it would be desirable to have larger ensembles available for future applications of our method. In addition, the current ensemble covers only modelling uncertainty, while internal variability and uncertainties in forcings are not explicitly included. A possible extension of our method would use ensembles which account for all these types of uncertainty in a structured way allowing the estimation of a better constrained prior distribution. For most of terrestrial Europe, the posterior mean is more similar to the local proxy reconstructions than to the prior mean from the unconstrained ensemble. This shows that the spatial structure is the more important part of the prior distribution compared to the mean state as long as the prior uncertainty is larger than the proxy uncertainty. Therefore, our method is applicable despite well-known model-data mismatches for the MH (Mauri et al., 2014).

Clim. Past Discuss., https://doi.org/10.5194/cp-2018-87
Manuscript under review for journal Clim. Past
Discussion started: 16 August 2018

5.2 Comparison with previous reconstructions

Several reconstructions of European climate during the MH have been compiled previously. Here, we compare our reconstructions to those of Mauri et al. (2015), Simonis et al. (2012), and Bartlein et al. (2011).

Mauri et al. (2015) use a plant functional type modern analogue transferfunction and a thin splate interpolation for pollen samples stemming mostly from the European pollen database. The simpler interpolation method allows the treatment of the samples as point data and the interpolation of the local reconstructions to an arbitrary grid. Among other variables, summer and winter temperatures are reconstructed. We find a dipole anomaly structure similar to Mauri et al. (2015) in our reconstructions, with mostly positive anomalies in northern Europe and negative anomalies in southern Europe. In Mauri et al. (2015) as well as in our reconstruction, the Alps are the only region with significant warming in central and southern Europe for summer temperature. Generally, the amplitude of summer anomalies in the two reconstructions are similar, although locally there are differences with cooler anomalies at the British Islands and southern Fennoscandia in our reconstruction as well as warmer anomalies in south-eastern Europe and Finland. For winter temperatures, we do not find the cooling over the British Islands that is reported in Mauri et al. (2015). As for summer temperatures, we find smaller anomalies in southern Fennoscandia. In contrast, our reconstruction shows higher anomalies in northern Scandinavia. In all other regions, the amplitude of anomalies are similar between the two reconstructions despite some local disagreements.

The same pollen dataset and a closely related transferfunction than in our reconstruction are used in Simonis et al. (2012) to reconstruct July and January temperature, such that differences between the two reconstructions are mostly related to the different smoothing technique. Simonis et al. (2012) minimize a cost function which combines pollen samples with an advection-diffusion model that is driven by insolation changes between the MH and today. In Simonis et al. (2012), the dipole structure is not found in the same way than in our reconstruction, which might be due to the different way how regions which are not well constrained by proxy data are treated. Both reconstructions share positive summer temperature anomalies in northern Europe as well as negative anomalies in central Europe and the Iberian peninsula. Unlike our reconstruction, Simonis et al. (2012) find positive anomalies in western and south-eastern Europe. For winter temperatures, the reconstruction of Simonis et al. (2012) shows an east-west dipole in contrast to the north-east to south-west dipole in our reconstruction. This different structure might be due to the smaller proxy data control of the winter reconstructions, which leads to a higher importance of the interpolation schemes.

A reconstruction designed to evaluate the PMIP3 simulations was provided by Bartlein et al. (2011). They combine a large number of pollen based local reconstructions from the literature to produce a gridded product of six climate variables including MTWA and MTCO. In contrast to our reconstruction, the used local reconstructions are not smoothed across space but only within a grid box. Their results show a dipole structure but less pronounced than in our reconstruction and in Mauri et al. (2015). In particular, they find a cooling for eastern Fennoscandia in summer, a much smaller warming of northern Fennoscandia than our reconstruction, and a warming in Germany and France. On the other hand, the reported anomalies in Bartlein et al. (2011) for the Mediterranean and eastern Europe are similar to our results.

Clim. Past Discuss., https://doi.org/10.5194/cp-2018-87
Manuscript under review for journal Clim. Past
Discussion started: 16 August 2018

The comparisons show that patterns like the dipole type anomaly structure, which are not present in the PMIP3 ensemble, seem to be consistent across pollen transferfunctions. While some of the differences between the existing literature and our results can be explained by the used transferfunctions and proxy syntheses, the choice of an appropriate interpolation method plays an important role, too, especially in areas with very sparse and weakly informative proxy data and to determine the degree of smoothing.

To facilitate more quantitative comparisons of reconstructions and allow a fair testing of modelling assumptions, we plan to expand our current framework to different types of transferfunctions and interpolation techniques. In particular, an implementation of a parametric process stage using the model of Simonis et al. (2012), which is independent of climate simulation output, is envisaged.

5.3 Towards global and spatio-temporal reconstructions

A valuable extension of our approach would be the computation of reconstructions on hemispheric to global scales. In this case, a more flexible structure for the ensemble member weights is desirable to account for the varying skill of climate models in different regions. On the other hand, the weights should not change too much within small domains to avoid unreasonable spatial heterogeneity. A way to combine those two aspects would be a cost function for the model weights, which combines a spatial smoothing term and the local fit of the ensemble members with proxy data. The balancing of those two terms to optimize the degree of smoothing and to respect the local proxy reconstructions is a challenging task itself. An additional problem is that the penalty parameter ρ in the glasso algorithm is applied locally, such that different regions require different values of ρ to result in a numerically stable covariance matrix and to optimize the criteria described in Sect. 3.2. While glasso allows spatially varying penalty parameters, optimizing the values in an objective way is a complicated issue. An alternative would be the reformulation of the penalty term in Eq. (5), such that not only local properties are optimized but also the structure between subregions. One possibility to achieve this could be the use of an additional L_2-penalization as in Zou et al. (2006).

Computing spatio-temporal reconstructions with our approach currently faces two challenges: The increasing dimensionality due to the additional temporal component, and the small number of available transient paleosimulations with comprehensive ESMs. In particular, for time scales beyond the last millennium, there is currently no multi-model ensemble available which is comparable to the PMIP3 simulations for the MH. To deal with the higher dimension a method to reduce the complexity is necessary. One way could be a separation of time and space but this might lead to a breakup of the temporal structure in the climate simulations resulting in inconsistencies in the reconstructions. More generally, any method to reduce the degrees of freedom in the reconstruction can lead to a significant underestimation of uncertainty if either the dimensionality is reduced too much or if the retained spatio-temporal patterns are too far from reality. A possibility to deal with both problems would be a Bayesian framework that combines a flexible parametric interpolation method featuring a large number of degrees of freedom with additional constraints from transient as well as time slice simulations to increase the physical consistency.

Clim. Past Discuss., https://doi.org/10.5194/cp-2018-87
Manuscript under review for journal Clim. Past
Discussion started: 16 August 2018

6 Conclusions

We presented a new method for probabilistic spatial reconstructions of paleoclimate. The approach combines the strengths of pollen records, which provide information about the climate state in a small scale domain, and of climate simulations, which downscale forcing conditions to physically consistent regional climate patterns. Thus, we reconstruct physically reasonable spatial fields, which are consistent with a given proxy synthesis. Bayesian modelling combined with MCMC methods is well suited for paleoclimate reconstructions as it can include multiple sources of information together with the corresponding uncertainties, and facilitates the separate modelling of proxy-climate transferfunctions and stochastic interpolation between proxy samples. Our framework can deal with probabilistic transferfunctions, which are non-linear and non-Gaussian, such that an extension to a wide range of proxies and associated transferfunctions is possible.

We apply our framework to MTWA and MTCO in Europe during the MH using the proxy synthesis of Simonis et al. (2012) and the PMIP3 MH ensemble. Brier scores from cross-validations reveal that the spatial reconstruction predominantly adds value to the unconstrained PMIP3 ensemble, and analyses of the spatial homogeneity of the posterior distribution indicate a reasonable degree of smoothing through the induced correlation structure. The large scale spatial patterns of our reconstruction are in agreement with previous work (Mauri et al., 2015; Bartlein et al., 2011). As the posterior mean is more similar to the local proxy reconstructions than to the prior mean for most terrestrial areas, we see that the main role of the simulation ensemble is to provide a spatial correlation structure and that a reconstruction, which is in line with reconstructions that do not include simulation output, is possible despite well-known model-data mismatches (Mauri et al., 2014). Our framework provides a way to quantitatively test hypotheses in paleoclimatology and to assess the consistency of a given proxy synthesis. This includes the fit with large scale structures, the spatial homogeneity, and the quantitative quality control of the proxy data by identification of potential outliers.

In future work, we plan to apply our framework to new multi-proxy syntheses like the one currently compiled in the German PalMod project (www.palmod.de). This includes the incorporation of new proxies and transferfunctions as well as the use of larger spatial domains. In particular, the lower degree of uncertainty reduction for the marine parts of the domain suggests an inclusion of supplementary marine proxies. Moreover, the addition of new high resolution paleosimulations for example from the ongoing PalMod and Paleoclimate Modelling Intercomparison Project Phase 4 (PMIP4) projects (Kageyama et al., 2018) will lead to a better quantification of uncertainties in the prior mixture distribution and, subsequently, to more robust reconstructions.

The hierarchical structure of the framework permits the replacement of the spatial interpolation module by other types of stochastic interpolation, e.g. geostatistical models (Tingley et al., 2012) or simple physical models (Gebhardt et al., 2008). The comparison of our current framework with these different modelling assumptions will lead to an improved understanding of the spatial patterns of past climate.

Clim. Past Discuss., https://doi.org/10.5194/cp-2018-87
Manuscript under review for journal Clim. Past
Discussion started: 16 August 2018

Climate of the Past Discussions

Appendix A: Full conditional distributions

The Metropolis-within-Gibbs approach samples (asymptotically) from the full conditional distributions of each variable, i.e. the distribution of the variable given all other variables. Some variables are treated block-wise to account for correlations between them, while others are updated sequentially if they are independent from each other or the joint distribution is too complicated for efficient sampling. In this appendix, we detail the conditional distributions that are used for sampling.

To sample the transferfunction parameters, we introduce PG distributed augmented variables γ_l^T, where $T \in T(P)$ and $l = 1, ..., L(T)$ are the number of observations for taxa T (Polson et al., 2013). γ_l^T is PG distributed given $\beta^T = (\beta_1^T, ..., \beta_6^T)$ and climate data $C(l) = (C_1(l), C_2(l))$, where C_1 and C_2 denote MTWA and MTCO. More precisely, the full conditional distribution is given by

$$\gamma_l^T \mid \beta^T, C(l) \quad \sim \quad \text{PG}(n = 1, X_{C(l)} \cdot \beta^T), \tag{A1}$$

$$\text{where } X_{C(l)} := \left(1, C_1(l), C_2(l), C_1(l)C_2(l), C_1(l)^2, C_2(l)^2\right). \tag{A2}$$

Including the Gaussian prior defined in Sect. 3.3, the full conditional of β^T is multivariate normal distributed:

$$\beta^T \mid P_m^T, P_p^T, \gamma_1^T, ..., \gamma_{L(T)}^T \quad \sim \quad \mathcal{N}\left(V_\gamma X^t \kappa^T, V_\gamma\right), \tag{A3}$$

$$\text{where } V_\gamma := \left(X^t \Gamma X + B^{-1}\right)^{-1}. \tag{A4}$$

Here, X is a matrix with rows $X_{C(l)}$ for $l = 1, ..., L(T)$, Γ is a diagonal matrix with entries γ_l^T, B is the 6×6 prior covariance matrix of β^T, and κ^T is a vector with entries $\left(P^T(l) - \frac{1}{2}\right)$, where $P^T(l)$ is the presence or absence of taxa T in observation l. In our case, B is a diagonal matrix with all values equal to 10. Details on the definition of PG variables and the augmented Gibbs sampler can be found in Polson et al. (2013).

The full conditional of $\omega = (\omega_1, ..., \omega_K)$ is Dirichlet distributed given $z = (z_1, ...z_K)$ and its Dirichlet prior:

$$\omega \mid z \quad \sim \quad \text{Dirichlet}\left(\tfrac{1}{2} + z_1, ..., \tfrac{1}{2} + z_K\right). \tag{A5}$$

Given ω and C_p, z is multinomially distributed:

$$z \mid \omega, C_p \quad \sim \quad \text{Multinomial}\left(n = 1, \alpha_1, ..., \alpha_K\right), \tag{A6}$$

$$\text{where } \alpha_k := \frac{\omega_k \cdot \exp\left(-\frac{1}{2}(C_p - \mu_k)^t \, \Sigma_{\text{prior}}^{-1} (C_p - \mu_k)\right)}{\sum_{i=1}^{K}\left(\omega_i \cdot \exp\left(-\frac{1}{2}(C_p - \mu_i)^t \, \Sigma_{\text{prior}}^{-1} (C_p - \mu_i)\right)\right)} \tag{A7}$$

We update $C_p(x)$ for $x \in x_P$ sequentially using random walk Metropolis-Hastings sampling. The set of all grid boxes but x is denoted by Y_x, and let $\Sigma_{\text{prior}}^{-1}(a, b)$ be the block of the inverse covariance matrix containing the rows a and columns b. Then, the full conditional distribution of $C_p(x)$ depends on the pollen samples $P_p(s)$ with location $x_s = x$, the climate $C_p(Y_x)$ at the other locations, and the chosen model z with $z_k = 1$. It does not follow a standard distribution:

$$C_p(x) \mid P_p, C_p(Y_x), z \;\sim\; \mathcal{N}\left(\tilde{\mu}_k(x), \left(\Sigma_{\text{prior}}^{-1}(x, x)\right)^{-1}\right) \prod_{\substack{s \\ \text{with} \\ x_s = x}} \prod_{T \in T(s)} \text{logit}\left(X_{C_p(x)} \cdot \beta^T\right), \tag{A8}$$

$$\text{where } \tilde{\mu}_k(x) := \mu_k(x) - \left(\Sigma_{\text{prior}}^{-1}(x, x)\right)^{-1} \Sigma_{\text{prior}}^{-1}(x, Y_x)\left(C_p(Y_x) - \mu_k(Y_x)\right). \tag{A9}$$

Clim. Past Discuss., https://doi.org/10.5194/cp-2018-87
Manuscript under review for journal Clim. Past
Discussion started: 16 August 2018

Climate of the Past Discussions
Open Access
EGU

The step size of the random walk proposals is controlled by the conditional covariance $\left(\Sigma^{-1}_{\text{prior}}(x,x)\right)^{-1}$.

Conditioned on $C_p(x_P)$ and z with $z_k = 1$, $C_p(x_Q)$ follows a Gaussian distribution:

$$C_p(x_Q) \mid C_p(x_P), z \quad \sim \quad \mathcal{N}\left(\tilde{\mu}_k(x_Q), \left(\Sigma^{-1}_{\text{prior}}(x_Q,x_Q)\right)^{-1}\right), \tag{A10}$$

$$\text{where } \tilde{\mu}_k(x_Q) := \mu_k(x_Q) - \left(\Sigma^{-1}_{\text{prior}}(x_Q,x_Q)\right)^{-1} \Sigma^{-1}_{\text{prior}}(x_Q,x_P)\left(C_p(x_P) - \mu_k(x_P)\right). \tag{A11}$$

Appendix B: Pseudo-code for MC^3 algorithm

For our reconstructions, we use $J = 2500$ iterations of $M = 30$ steps to draw altogether 75,000 samples from $A = 8$ MCMC chains (see Sect. 3.4). The MC^3 algorithm follows the pseudo-code in Algorithm 1.

If the modularized version is used, 75,000 samples of the transferfunction parameters θ are drawn first. These are used for subsequent reconstructions which follow the pseudo-code without sampling the transferfunction parameters.

Competing interests. The authors declare that they have no conflict of interest.

Acknowledgements. This work was supported by German Federal Ministry of Education and Research (BMBF) as Research for Sustainability initiative (FONA, www.fona.de) through the Palmod project (FKZ: 01LP1509D). N. Weitzel was additionally supported by the National Center for Atmospheric Research (NCAR). NCAR is funded by the National Science Foundation. We acknowledge all groups involved in producing and making available the PMIP3 multi-model ensemble. We acknowledge the World Climate Research Programme's Working Group on Coupled Modelling, which is responsible for CMIP. For CMIP the US Department of Energy's Program for Climate Model Diagnosis and Intercomparison provides coordinating support and led development of software infrastructure in partnership with the Global Organization for Earth System Science Portals. We thank Douglas Nychka for helpful ideas to speed up the MCMC algorithm.

Clim. Past Discuss., https://doi.org/10.5194/cp-2018-87
Manuscript under review for journal Clim. Past
Discussion started: 16 August 2018

References

Altekar, G., Dwarkadas, S., Huelsenbeck, J. P., and Ronquist, F.: Parallel Metropolis coupled Markov chain Monte Carlo for Bayesian phylogenetic inference, Bioinformatics, 20/3, 407–415, 2004.

Annan, J. D. and Hargreaves, J. C.: A new global reconstruction of temperature changes at the Last Glacial Maximum, Climate of the Past, 9, 367–376, 2013.

Bartlein, P. J., Harrison, S. P., Brewer, S., Connor, S., Davis, B. A. S., Gajewski, K., Guiot, J., Harrison-Prentice, T. I., Henderson, A., Peyron, O., Prentice, I. C., Scholze, M., Seppä, H., Shuman, B., Sugita, S., Thompson, R. S., Viau, A. E., Williams, J., and Wu, H.: Pollen-based continental climate reconstructions at 6 and 21 ka: a global synthesis, Climate Dynamics, 37, 775–802, 2011.

Birks, H. J. B., Heiri, O., Seppä, H., and Bjune, A. E.: Strengths and Weaknesses of Quantitative Climate Reconstructions Based on Late-Quaternary Biological Proxies, The Open Ecology Journal, 3, 68–110, 2010.

Braconnot, P., Harrison, S. P., Otto-Bliesner, B., Abe-Ouchi, A., Jungclaus, J., and Peterschmitt, J.-Y.: The Paleoclimate Modeling Intercomparison Project contribution to CMIP5, CLIVAR Exchanges, 56/16/2, 15–19, 2011.

Bradley, R. S.: Paleoclimatology - Reconstructing Climates of the Quaternary, Academic Press, Oxford, 3 edn., 2015.

Brier, G. W.: Verification of forecasts expressed in terms of probability, Monthly Weather Review, 78, 1–3, 1950.

Brooks, S. P. and Gelman, A.: General Methods for Monitoring Convergence of Iterative Simulations, Journal of Computational and Graphical Statistics, 7/4, 434–455, 1998.

Dee, S. G., Steiger, N. J., Emile-Geay, J., and Hakim, G. J.: On the utility of proxy system models for estimating climate states over the common era, Journal of Advances in Modeling Earth Systems, 8, 1164–1179, 2016.

Friedman, J., Hastie, T., and Tibshirani, R.: Sparse inverse covariance estimation with the graphical lasso, Biostatistics, 9/3, 432–441, 2008.

Gebhardt, C., Kühl, N., Hense, A., and Litt, T.: Multi-Scale Processes and the Reconstruction of Palaeoclimate, pp. 325–336, In: Neugebauer, H. J. and Simmer, C. (ed), Dynamics of Multiscale Earth Systems, Springer, Berlin, 2003.

Gebhardt, C., Kühl, N., Hense, A., and Litt, T.: Reconstruction of Quaternary temperature fields by dynamically consistent smoothing, Climate Dynamics, 30, 421–437, 2008.

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., and Rubin, D. B.: Bayesian data analysis, Chapman & Hall / CRC, Boca Raton, 3 edn., 2013.

Geyer, C. J.: Markov Chain Monte Carlo Maximum Likelihood, pp. 156–163, In: Keramidas (ed.), Computing Science and Statistics: Proceedings of the 23rd Symposium on the Interface, Interface Foundation, Fairfax Station, 1991.

Gneiting, T. and Raftery, A. E.: Strictly Proper Scoring Rules, Prediction, and Estimation, Journal of the American Statistical Association, 102/477, 359–378, 2007.

Harris, I. and Jones, P.: CRU TS4.01: Climatic Research Unit (CRU) Time-Series (TS) version 4.01 of high-resolution gridded data of month-by-month variation in climate (Jan. 1901- Dec. 2016), https://doi.org/doi:10.5285/58a8802721c94c66ae45c3baa4d814d0, 2017.

Harris, I., Jones, P. D., Osborn, T. J., and Lister, D. H.: Updated high-resolution grids of monthly climatic observations - the CRU TS3.10 Dataset, International Journal of Climatology, 34/3, 623–642, 2014.

Haslett, J., Whiley, M., Bhattacharya, S., Salter-Townshend, M., Wilson, S. P., Allen, J. R. M., Huntley, B., and Mitchell, F. J. G.: Bayesian paleoclimate reconstruction, Journal of the Royal Statistical Society, Series A (Statistics in Society), 169, 395–438, 2006.

Holden, P. B., Birks, H. J. B., Brooks, S. J., Bush, M. B., Hwang, G. M., Matthews-Bird, F., Valencia, B. G., and van Woesik, R.: BUMPER v1.0: a Bayesian user-friendly model for palaeo-environmental reconstruction, Geoscientific Model Development, 10, 483–498, 2017.

Holmström, L., Ilvonen, L., Seppä, H., and Veski, S.: A Bayesian spatiotemporal model for reconstructing climate from multiple pollen records, The Annals of Applied Statistics, 9/3, 1194–1225, 2015.

Huntley, B.: Reconstructing palaeoclimates from biological proxies: Some often overlooked sources of uncertainty, Quarternary Science Reviews, 31, 1–16, 2012.

Iversen, J.: Viscum, Hedera and Ilex as climate indicators, Geologiska Föreningen i Stockholm Förhandlingar, 66, 463–483, 1944.

Jones, P. D., New, M., Parker, D. E., Martin, S., and Rigor, I. G.: Surface air temperature and its variations over the last 150 years, Reviews of Geophysics, 37, 173–199, 1999.

Kageyama, M., Braconnot, P., Harrison, S. P., Haywood, A. M., Jungclaus, J., Otto-Bliesner, B. L., Peterschmitt, J.-Y., Abe-Ouchi, A., Albani, S., Bartlein, P. J., Brierley, C., Crucifix, M., Dolan, A., Fernandez-Donado, L., Fischer, H., Hopcroft, P. O., Ivanovic, R. F., Lambert, F., Lunt, D. J., Mahowald, N. M., Peltier, W. R., Phipps, S. J., Roche, D. M., Schmidt, G. A., Tarasov, L., Valdes, P. J., Zhang, Q., and Zhou, T.: The PMIP4 contribution to CMIP6 – Part1: Overview and over-arching analysis plan, Geoscientific Model Development, pp. 1033–1057, 2018.

Kühl, N., Gebhardt, C., Litt, T., and Hense, A.: Probability Density Functions as Botanical-Climatological Transfer Functions for Climate Reconstruction, Quarternary Research, 58, 381–392, 2002.

Li, B., Nychka, D. W., and Ammann, C. M.: The Value of Multiproxy Reconstruction of Past Climate, Journal of the American Statistical Association, 105/491, 883–895, 2010.

Liu, F., Bayarri, M. J., and Berger, J. O.: Modularization in Bayesian Analysis, with Emphasis on Analysis of Computer Models, Bayesian Analysis, 4, 119–150, 2009.

Mahalanobis, P.: On the generalized distance in statistics, Proceedings of the National Institute of Sciences (Calcutta), 2, 49–55, 1936.

MARGO Project Members: Constraints on the magnitude and patterns of ocean cooling at the Last Glacial Maximum, Nature Geoscience, 2, 127–132, 2009.

Mauri, A., Davis, B. A. S., Collins, P. M., and Kaplan, J. O.: The influence of atmospheric circulation on the mid-Holocene climate of Europe: a data-model comparison, Climate of the Past, 10, 1925–1938, 2014.

Mauri, A., Davis, B. A. S., Collins, P. M., and Kaplan, J. O.: The climate of Europe during the Holocene: a gridded pollen-based reconstruction and its multi-proxy evaluation, Quaternary Science Reviews, 112, 109–127, 2015.

Ohlwein, C. and Wahl, E. R.: Review of probabilistic pollen-climate transfer methods, Quaternary Science Reviews, 31, 17–29, 2012.

Parnell, A. C., Sweeney, J., Doan, T. K., Salter-Townshend, M., Allen, J. R., Huntley, B., and Haslett, J.: Bayesian inference for palaeoclimate with time uncertainty and stochastic volatility, Journal of the Royal Statistical Society, Series C (Applied Statistics), 64/1, 115–138, 2015.

Parnell, A. C., Haslett, J., Sweeney, J., Doan, T. K., Allen, J. R., and Huntley, B.: Joint Palaeoclimate reconstruction from pollen data via forward models and climate histories, Quarternary Science Reviews, 151, 111–126, 2016.

Plummer, M., Best, N., Cowles, K., and Vines, K.: CODA: Convergence Diagnosis and Output Analysis for MCMC, R News, 6, 7–11, https://www.r-project.org/doc/Rnews/Rnews_2006-1.pdf, 2006.

Polson, N. G., Scott, J. G., and Windle, J.: Bayesian inference for logistic models using Pólya-Gamma latent variables, arXiv:1205.0310v3, pp. 1–42, 2013.

Rehfeld, K., Trachsel, M., Telford, R. J., and Laepple, T.: Assessing performance and seasonal bias of pollen-based climate reconstructions in a perfect model world, Climate of the Past, 12, 2255–2270, 2016.

Rue, H. and Held, L.: Gaussian Markov Random Fields: Theory and Applications, Chapman & Hall / CRC, Boca Raton, 2005.

Schölzel, C. and Hense, A.: Probabilistic assessment of regional climate change in Southwest Germany by ensemble dressing, Climate Dynamics, 36, 2003–2014, 2011.

Schölzel, C., Hense, A., Hübl, P., Kühl, N., and Litt, T.: Digitization and geo-referencing of botanical distribution maps, Journal of Biogeography, 29, 851–856, 2002.

Silverman, B. W.: Density Estimation for Statistics and Data Analysis, vol. 26 of *Monographs on Statistics and Applied Probability*, Chapman & Hall / CRC, Boca Raton, 1986.

Simonis, D.: Reconstruction of possible realizations of the Late Glacial and Holocene near surface climate in Central Europe, Dissertation, Meteorologisches Institut der Rheinischen Friedrich-Wilhelms-Universität Bonn, 2009.

Simonis, D., Hense, A., and Litt, T.: Reconstruction of late Glacial and Early Holocene near surface temperature anomalies in Europe and their statistical interpretation, Quaternary International, 274, 233–250, 2012.

Steiger, N. J., Hakim, G. J., Steig, E. J., Battisti, D. S., and Roe, G. H.: Assimilation of Time-Averaged Pseudoproxies for Climate Reconstruction, Journal of Climate, 27, 426–441, 2014.

Stolzenberger, S.: Untersuchungen zu botanischen Paläoklimatransferfunktionen, Diploma thesis, Meteorologisches Institut der Rheinischen Friedrich-Wilhelms-Universität Bonn, 2011.

Stolzenberger, S.: On the probabilistic evaluation of decadal and paleoclimate model predictions, Dissertation, Meteorologisches Institut der Rheinischen Friedrich-Wilhelms-Universität Bonn, 2017.

Thuiller, W.: BIOMOD – optimizing predictions of species distributions and projecting potential future shifts under global change, Global Change Biology, 9, 1353–1362, 2003.

Tingley, M. P. and Huybers, P.: A Bayesian Algorithm for Reconstructing Climate Anomalies in Space and Time. Part I: Development and Applications to Paleoclimate Reconstruction Problems, Journal of Climate, 23, 2759–2781, 2010.

Tingley, M. P., Craigmile, P. F., Haran, M., Li, B., Mannshardt, E., and Rajaratnam, B.: Piecing together the past: statistical insights into paleoclimatic reconstructions, Quaternary Science Reviews, 35, 1–22, 2012.

Werner, J. P. and Tingley, M. P.: Technical Note: Probabilistically constraining proxy age-depth models within a Bayesian hierarchical reconstruction model, Climate of the Past, 11, 533–545, 2015.

Windle, J., Polson, N. G., and Scott, J. G.: BayesLogit: Bayesian logistic regression, http://cran.r-project.org/web/packages/BayesLogit/index.html, 2013.

Windle, J., Polson, N. G., and Scott, J. G.: Sampling Pólya-Gamma random variates: alternate and approximate techniques, arXiv:1405.0506v1, pp. 1–26, 2014.

Zou, H., Hastie, T., and Tibshirani, R.: Sparse Principal Component Analysis, Journal of Computational and Graphical Statistics, 15/2, 265–286, 2006.

Clim. Past Discuss., https://doi.org/10.5194/cp-2018-87
Manuscript under review for journal Clim. Past
Discussion started: 16 August 2018

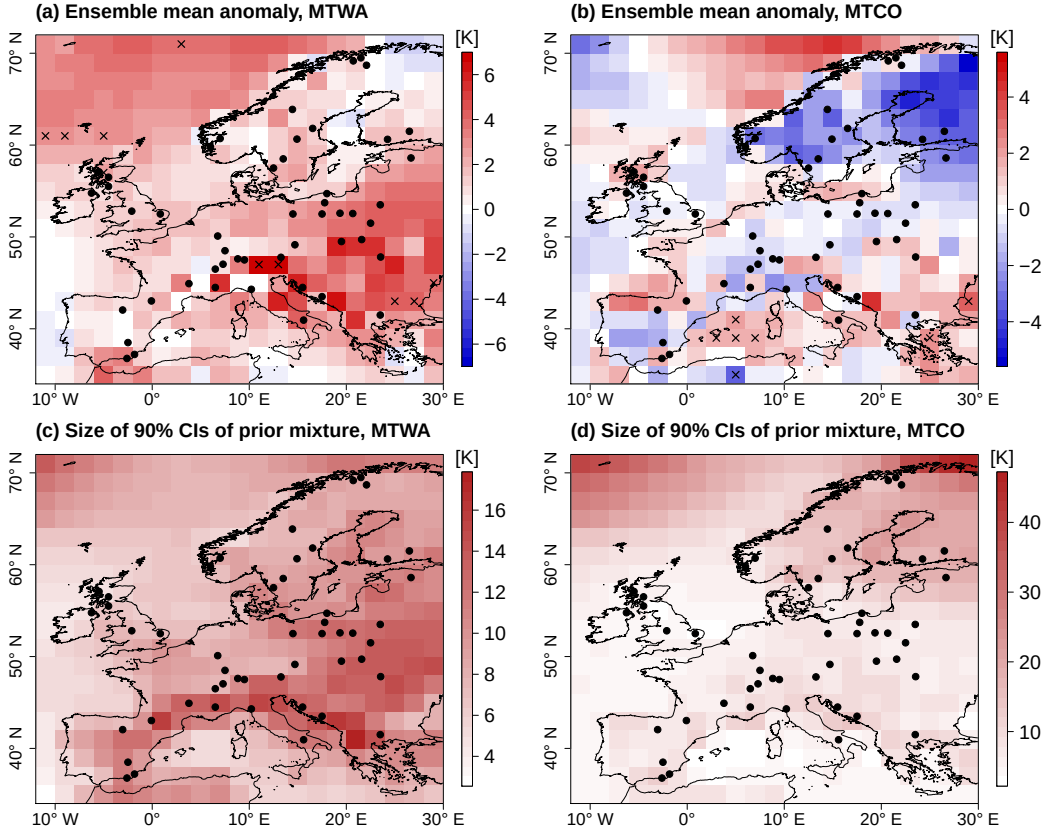

Figure 1. PMIP3 MH ensemble mean anomaly from CRU reference climatology, (a) MTWA, (b) MTCO, and ensemble spread as size of point-wise 90% CIs of the prior mixture distribution, (c) MTWA, (d) MTCO. Black dots depict proxy samples from Simonis et al. (2012). In the top row, point-wise significant anomalies (5% level) are marked by black crosses.

Clim. Past Discuss., https://doi.org/10.5194/cp-2018-87
Manuscript under review for journal Clim. Past
Discussion started: 16 August 2018

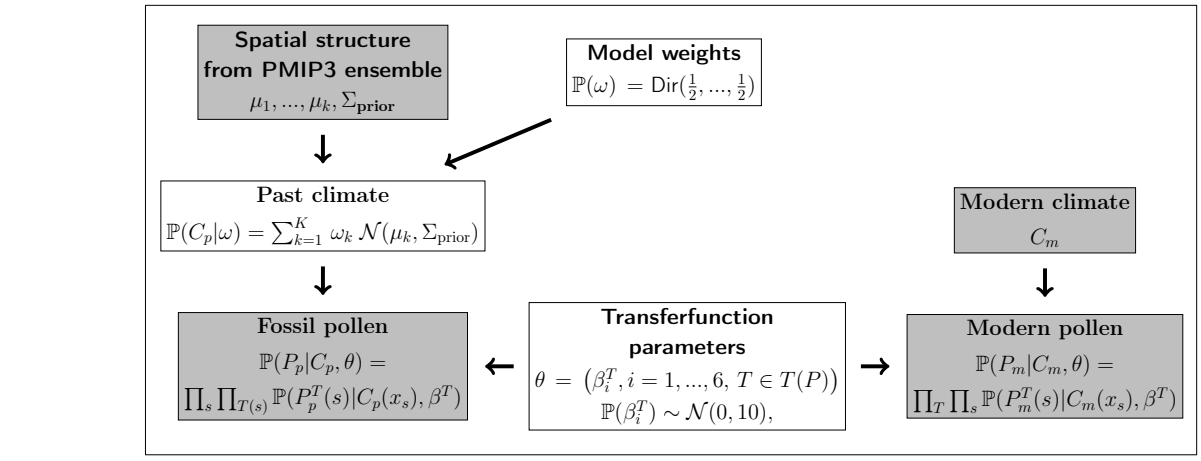

Figure 2. Directed acyclic graph corresponding to the Bayesian framework Eq. (2). Involved quantities are given by nodes and arrows indicate dependences of variables. Details of the formulas are explained in Sect. 3.2 and 3.3. White: inferred quantities; gray: input data.

Clim. Past Discuss., https://doi.org/10.5194/cp-2018-87
Manuscript under review for journal Clim. Past
Discussion started: 16 August 2018

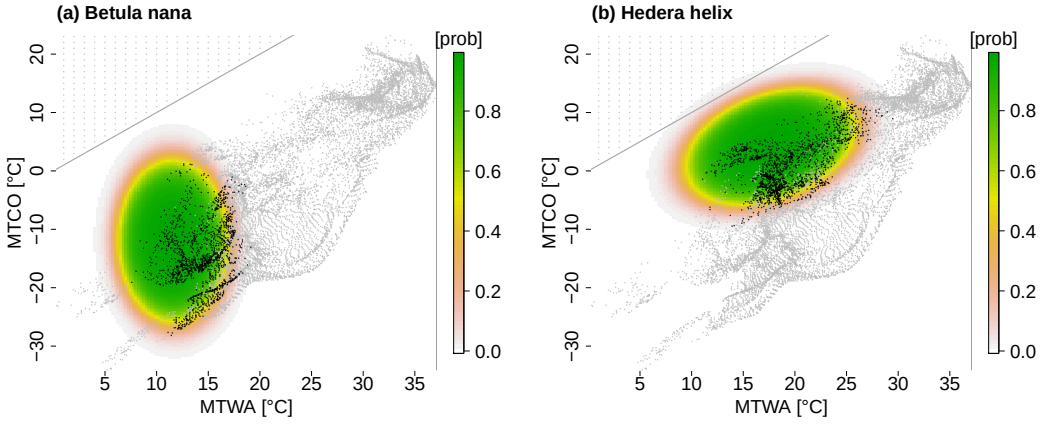

Figure 3. Response functions for Betula nana (a) and Hedera helix (b). The probability of presence of the taxa is shown in colours, black dots denote presence of the taxa, and gray dots denote absence of the taxa in the calibration dataset. In the climate space, combinations of MTWA and MTCO above the gray line at MTWA = MTCO cannot occur by definition. Gray dots above this line are artificial absence information added to account for this constraint.

Clim. Past Discuss., https://doi.org/10.5194/cp-2018-87
Manuscript under review for journal Clim. Past
Discussion started: 16 August 2018

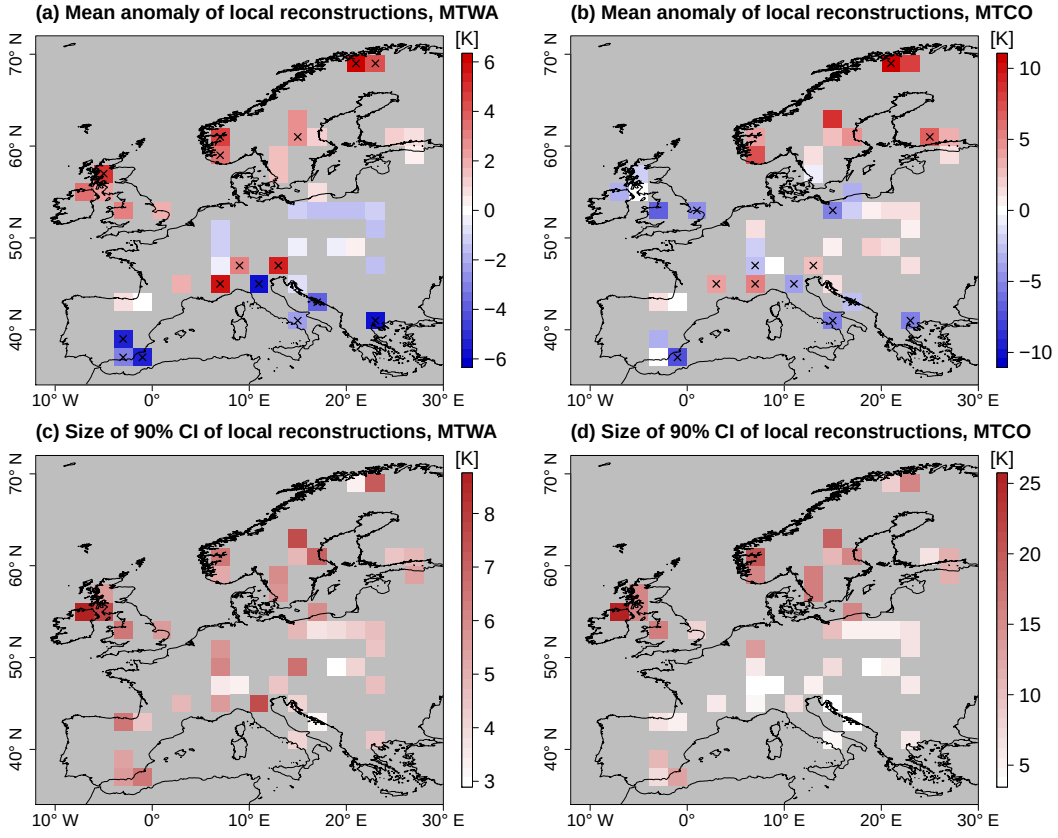

Figure 4. Summary statistics of local reconstructions using the PITM forward model. Top row: mean anomaly from CRU reference climatology (left: MTWA, right: MTCO), bottom row: uncertainty measured by the size of marginal 90% CIs (left: MTWA, right: MTCO). In the top row, significant anomalies (5% level) are marked by black crosses.

Clim. Past Discuss., https://doi.org/10.5194/cp-2018-87
Manuscript under review for journal Clim. Past
Discussion started: 16 August 2018

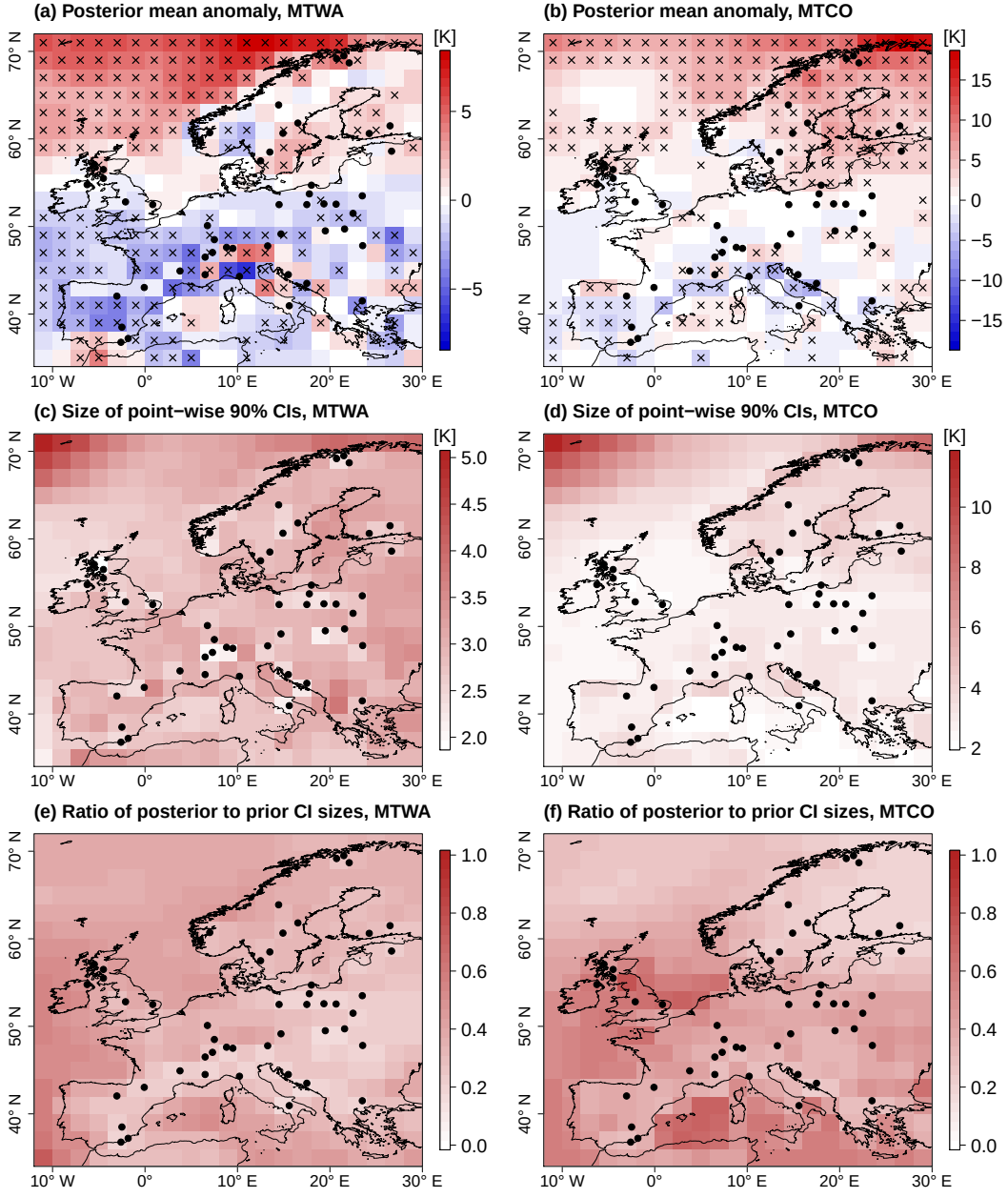

Figure 5. Spatial reconstruction for MH. Top row: Posterior mean anomaly from CRU reference climatology (left: MTWA, right: MTCO), middle row: reconstruction uncertainty plotted as size of point-wise 90% CIs (left: MTWA, right: MTCO), bottom row: reduction of uncertainty from posterior to prior measured by ratio of posterior to prior point-wise 90% CI sizes (left: MTWA, right: MTCO). Black dots depict proxy samples. In the top row, point-wise significant anomalies (5% level) are marked by black crosses.

Clim. Past Discuss., https://doi.org/10.5194/cp-2018-87
Manuscript under review for journal Clim. Past
Discussion started: 16 August 2018

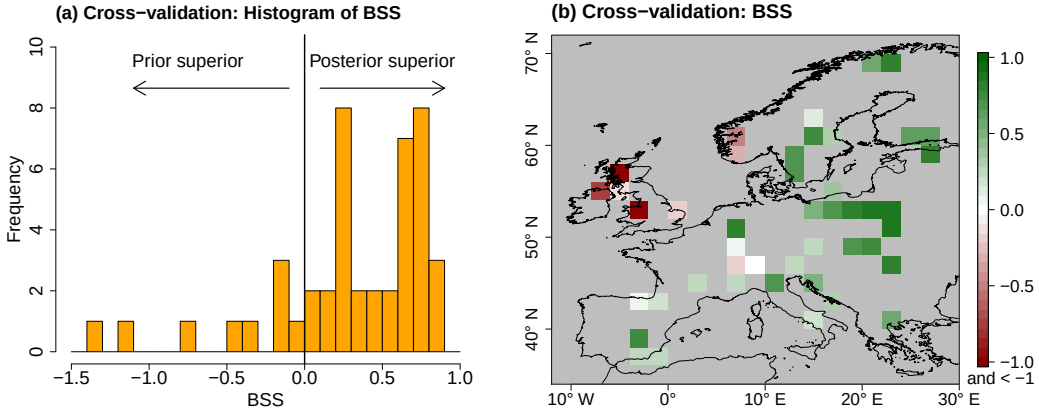

Figure 6. BSS from leave-one-out cross-validation. For positive values the posterior is superior to the prior distribution from the unconstrained PMIP3 ensemble, while for negative values the posterior is inferior to the prior. (a) histogram, (b) spatial distribution

Clim. Past Discuss., https://doi.org/10.5194/cp-2018-87
Manuscript under review for journal Clim. Past
Discussion started: 16 August 2018

Climate of the Past Discussions
Open Access
EGU

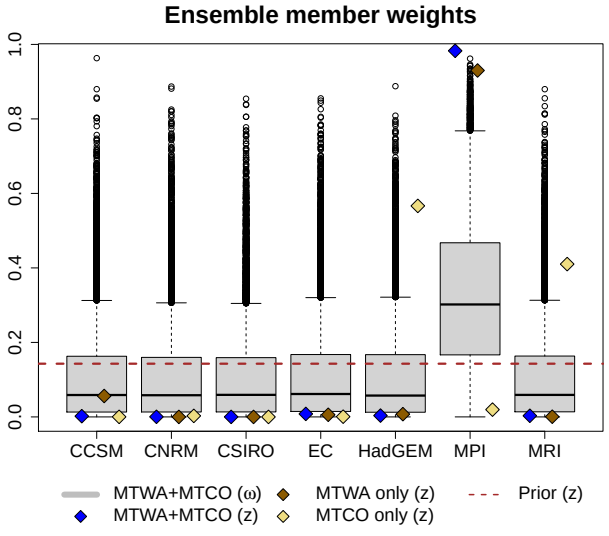

Figure 7. Posterior ensemble member weights of the reconstructions. The posterior of ω in the joint reconstruction is shown as boxplot. The diamonds represent the mean of z in the joint, the MTWA only, and the MTCO only reconstruction. Prior weights (mean of z) are denoted by the dashed line.

Clim. Past Discuss., https://doi.org/10.5194/cp-2018-87
Manuscript under review for journal Clim. Past
Discussion started: 16 August 2018

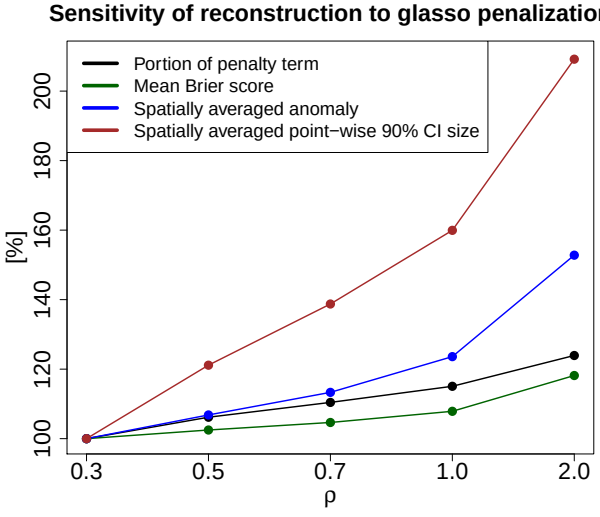

Figure 8. Sensitivity of reconstructions to changes in the glasso penalty parameter ρ. Portion of the penalty term on the total cost in Eq. (5), mean BS of cross-validations, spatially averaged posterior mean anomaly from CRU reference climatology, and spatially averaged size of point-wise 90% CIs. All quantities are plotted as percentage of the respective values for $\rho = 0.3$. Note the non-linear scaling of the x-axis.

Clim. Past Discuss., https://doi.org/10.5194/cp-2018-87
Manuscript under review for journal Clim. Past
Discussion started: 16 August 2018

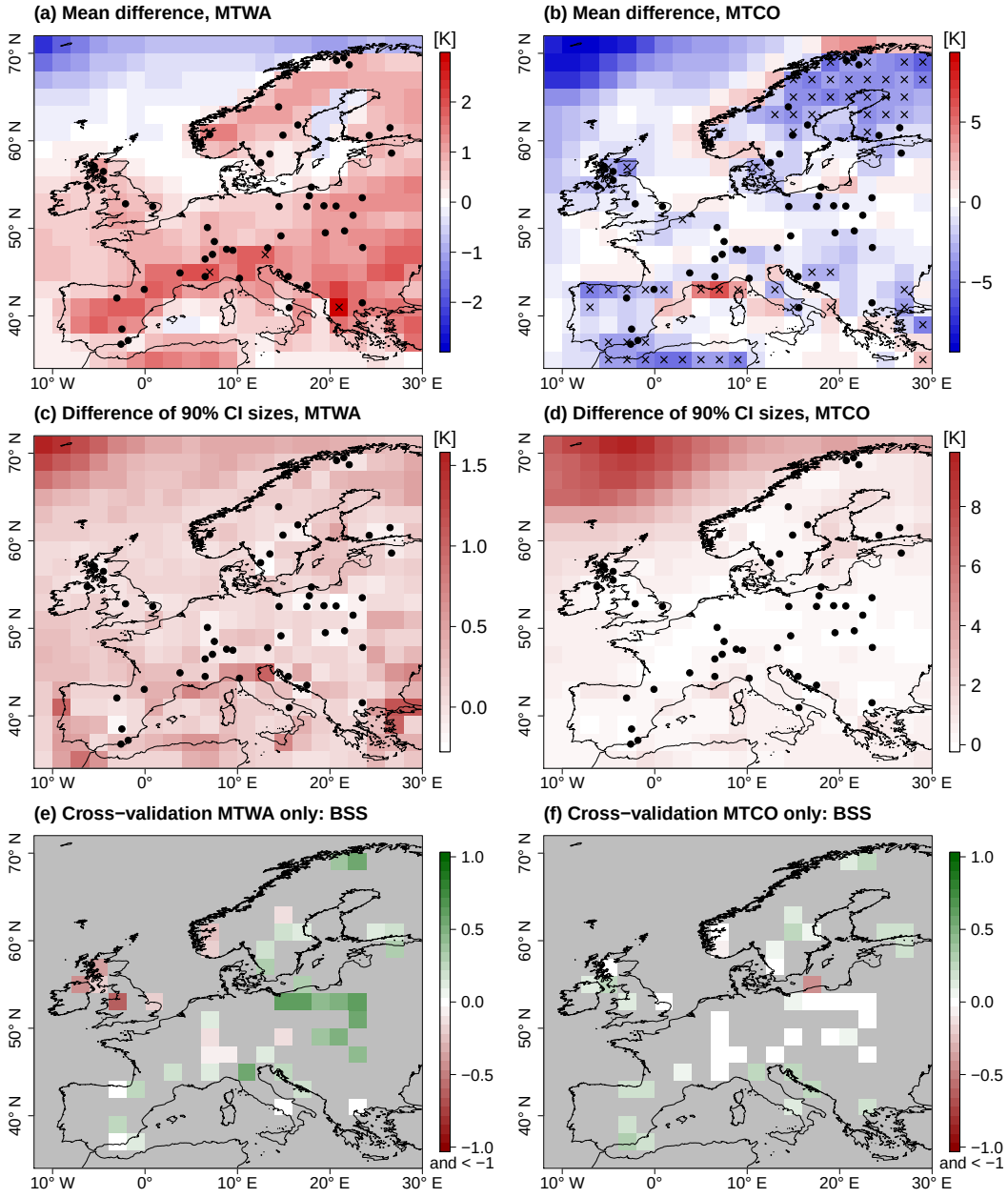

Figure 9. Differences of joint and separate reconstructions of MTWA and MTCO. Top row: posterior mean difference (left: MTWA, right: MTCO); middle row: difference of size of point-wise 90% CIs (left: MTWA, right: MTCO); bottom row: BSS of the separate reconstructions (left: MTWA, right: MTCO). Black dots depict proxy samples. In the top row, point-wise significant differences (5% level) between the separate and the joint reconstructions are marked by black crosses.

Clim. Past Discuss., https://doi.org/10.5194/cp-2018-87
Manuscript under review for journal Clim. Past
Discussion started: 16 August 2018

Climate of the Past Discussions
Open Access
EGU

Table 1. Basic information on the PMIP3 climate simulations used to construct the prior in the Bayesian framework (from https://pmip3.lsce.ipsl.fr)

Model	Institute	Atmospheric grid	Ocean grid	Model years MH
CCSM4	NCAR	288x192xL26	320x384xL60	301
CNRM-CM5	CNRM-CERFACS	256x128xL31	362x292xL42	200
CSIRO-Mk3-6-0	CSIRO-QCCCE	192x96xL18	192x195xL31	100
EC-Earth-2-2	ICHEC	320x160xL62	362x292xL42	40
HadGEM2-CC	MOHC	192x144xL60	360x216xL40	35
MPI-ESM-P	MPI-M	196x98xL47	256x220xL40	100
MRI-CGCM3	MRI	320x160xL48	364x368xL51	100

Clim. Past Discuss., https://doi.org/10.5194/cp-2018-87
Manuscript under review for journal Clim. Past
Discussion started: 16 August 2018

Climate of the Past Discussions
Open Access
EGU

Table 2. Summary measures for the joint MTWA and MTCO reconstructions (rows 1 and 2) and the separated reconstructions of MTWA (row 3) and MTCO (row 4). Numbers in brackets are minima and maxima of the corresponding 90% CIs. Additional explanations for all the columns can be found in Sect. 4.1 to 4.5.

Reconstruction name	Spatial mean anomaly	Spatially averaged 90% CI size	Point-wise uncertainty reduction	Spatial homogeneity	Median BSS	PMIP model with highest weight
Joint reconstruction (MTWA)	(0.33 K) 0.01 K (-0.32 K)	2.90 K	64.5%	(1.40 K) 1.48 K (1.57 K)	0.39	MPI-ESM-P
Joint reconstruction (MTCO)	(1.50 K) 1.11 K (0.74 K)	3.18 K	58.6%	(2.09 K) 2.18 K (2.27 K)		
Separate MTWA reconstruction	(0.96 K) 0.64 K (0.30 K)	3.10 K	62.2%	(1.38 K) 1.47 K (1.55 K)	0.16	MPI-ESM-P
Separate MTCO reconstruction	(0.92 K) 0.06 K (-0.79 K)	4.14 K	49.4%	(2.45 K) 2.66 K (2.86 K)	0.04	HadGem2-CC

Clim. Past Discuss., https://doi.org/10.5194/cp-2018-87
Manuscript under review for journal Clim. Past
Discussion started: 16 August 2018

Climate of the Past Discussions
Open Access
EGU

Algorithm 1 Pseudo-code for MC^3 algorithm

Initialize A MCMC chains
for $j = 1, ..., J$ **do**
  for $m = 1, ..., M$ **do**
    for $T \in T(P)$ **do**
      for $l = 1, ..., L(T)$ **do**
        Sample from full conditional of γ_l^T
      end for
      Sample from full condition of β^T
    end for
    Sample from full conditional of ω
    Sample from full conditional of z
    for $x \in x_P$ **do**
      Sample random walk proposal for $C_p(x)$
      Calculate Metropolis-Hastings ratio r
      Accept proposal for $C_p(x)$ with probability $p = \max(1, r)$
    end for
  end for
  for $a = 1, ..., (A - 1)$ **do**
    Calculate Metropolis-Hastings ratio r of chains a and $a + 1$
    Swap chains a and $a + 1$ with probability $p = \max(1, r)$
  end for
end for
for $i = 1, ..., JM$ **do**
  Sample from full conditional of $C_p(x_Q)$
end for