LUND UNIVERSITY

Creating spatially continuous maps of past land cover from point estimates: A new statistical approach applied to pollen data

Pirzamanbein, Behnaz; Lindström, Johan; Poska, Anneli; Sugita, Shinya; Trondman, Anna-Kari; Fyfe, Ralph; Mazier, Florence; Nielsen, Anne Birgitte; Kaplan, Jed O.; Bjune, Anne E.; B. Birks, H. John; Giesecke, Thomas; Kangur, Mikhel; Latałowa, Małgorzata; Marquer, Laurent; Smith, Benjamin; Gaillard, Marie-José

Link to publication

LUND UNIVERSITY

PO Box 117
221 00 Lund
+46 46-222 00 00

Creating spatially continuous maps of past land cover from point estimates: A new statistical approach applied to pollen data

Behnaz Pirzamanbein[a,b,*], Johan Lindström[a], Anneli Poska[c,d], Shinya Sugita[e], Anna-Kari Trondman[f], Ralph Fyfe[g], Florence Mazier[h], Anne B. Nielsen[i], Jed O. Kaplan[j], Anne E. Bjune[k], H. John B. Birks[l,m,n], Thomas Giesecke[o], Mikhel Kangur[e], Małgorzata Latałowa[p], Laurent Marquer[f], Benjamin Smith[c], Marie-José Gaillard[f]

[a]*Centre for Mathematical Sciences, Lund University, Sweden*
[b]*ClimBEco Graduate Research School, CEC, Lund University, Sweden*
[c]*Department of Physical Geography and Ecosystems Analysis, Lund University, Sweden*
[d]*Institute of Geology, Tallinn University of Technology, Estonia*
[e]*Institute of Ecology, Tallinn University, Estonia*
[f]*Department of Biology and Environmental Sciences, Linnaeus University, Sweden*
[g]*School of Geography, Earth and Environmental Sciences, University of Plymouth, UK*
[h]*GEODE, UMR 5602, University of Toulouse, France*
[i]*Department of Geology, Lund University, Sweden*
[j]*Institute for Environmental Sciences, University of Geneva, Switzerland*
[k]*Uni Climate, Uni Research AS and Bjerknes Centre for Climate Research, Norway*
[l]*Department of Biology, University of Bergen, Norway*
[m]*Environmental Change Research Centre, University College London, UK*
[n]*School of Geography and Environment, University of Oxford, UK*
[o]*Department of Palynology and Climate Dynamics, University of Göttingen, Germany*
[p]*Laboratory of Palaeoecology and Archaeobotany, Department of Plant Ecology, University of Gdańsk, Poland*

Abstract

Reliable estimates of past land cover are critical for assessing potential effects of anthropogenic land-cover changes on past earth surface-climate feedbacks and landscape complexity. Fossil pollen records from lakes and bogs have provided important information on past natural and human-induced vegetation cover. However, those records provide only point estimates of past land cover, and not the spatially continuous maps at regional and sub-continental scales needed for climate modelling.

We propose a set of statistical models that create spatially continuous maps of past land cover by combining two data sets: 1) pollen-based point estimates of past land cover (from the REVEALS model), and 2) spatially continuous estimates of past land cover, obtained by combining simulated potential vegetation (from LPJ-GUESS) with an anthropogenic land-cover change scenario (KK10). The proposed models rely on statistical methodology for compositional data and use Gaussian Markov Random Fields to model spatial dependencies in the data.

Land-cover reconstructions are presented for three time windows in Europe: 0.05 ka, 0.2 ka, and 6 ka years before present (BP). The models are evaluated through cross-validation, deviance information criteria and by comparing the reconstruction of the 0.05 ka time window to the present-day land-cover data compiled by the European Forest Institute (EFI). For 0.05 ka, the proposed models provide reconstructions that are closer to the EFI data than either the REVEALS- or LPJ-GUESS/KK10-based estimates; thus the statistical combination of the two estimates improves the reconstruction. The reconstruction by the proposed models for 0.2 ka is also good. For 6 ka, however, the large differences between the REVEALS- and LPJ-GUESS/KK10-based estimates reduce the reliability of the proposed models. Possible reasons for the increased differences between REVEALS and LPJ-GUESS/KK10 for older time periods and further improvement of the proposed models are discussed.

Keywords: Land cover, Spatial modeling, Paleoecology, Pollen, Compositional data, Gaussian Markov random fields

1. Introduction

Anthropogenic impacts on past land cover have potentially influenced the climate system more significantly than previously assumed (e.g. Ruddiman, 2005). Many simulation studies have evaluated the biogeophysical effects of vegetation and land-use changes on past climate at the global scale (e.g. Claussen et al., 2001; Brovkin et al., 2006; Pitman et al., 2009; Pongratz et al., 2010; de Noblet-Ducoudré et al., 2012; Christidis et al., 2013). However, descriptions of past land cover vary considerably among studies, including: static present-day land cover (Strandberg et al., 2011), dynamic (or static) potential land cover simulated by dynamic vegetation models (DVMs) (e.g. Brovkin et al., 2002; Hickler et al., 2012),

and land-cover estimates combining DVMs and anthropogenic land-cover change (ALCC) scenarios (Pongratz et al., 2009; de Noblet-Ducoudré et al., 2012). The existing ALCC scenarios (e.g. Kaplan et al., 2009; Pongratz et al., 2009; Klein Goldewijk et al., 2011) also differ significantly from each other (Gaillard et al., 2010) and their reliability still needs to be evaluated.

Palaeoecology has provided important information on past vegetation and land cover using fossil pollen and plant macrore-mains deposited and preserved in lake and bog sediments over thousands of years. Although those palaeorecords provide insights into the past vegetation that modelling approaches cannot, the interpretation of palaeorecords, particularly quantification of land-cover changes in specific spatiotemporal scales, remains difficult. In addition palaeorecords are point estimates of land cover around study sites. Therefore, the gaps between estimates at study points need to be filled if palaeorecords of land-cover changes are to be useful in climate modelling and other simulation studies that require quantitative and spatially continuous input datasets. To achieve this interpolation process we propose a new statistical approach based on statistical spatial models and methods developed in Tjelmeland and Lund (2003), Lindgren et al. (2011) and Rue et al. (2009). Our approach takes spatially continuous estimates of past land cover from a DVM and an ALCC scenario as covariates and then constrains those using the point estimates of pollen-based land cover; thus it can potentially avoid problems that conventional interpolation methods using fossil pollen records have. The DVMs and ALCC scenarios provide a way of capturing land-cover changes due to the non-stationary environmental conditions in Europe over areas with few or no pollen-based observations.

This paper aims at reconstructing the land cover in Europe at 6.0 ka, 0.2 ka and 0.05 ka (calibrated year BP) using the methods developed in this study. The work is part of the LAND-CLIM project (LAND cover - CLIMate interactions in Europe during the Holocene; Gaillard et al., 2010) that assesses the possible effects of long-term changes in anthropogenic land cover on the Holocene climate (Strandberg et al., 2014). Our objective is also to provide methods and reconstructions that can be used in the evaluation of ecological complexity of European landscapes in the past, i.e. give us new insights on the respective roles played by climate, soils, geography, geology and human impact in landscape dynamics at the spatio-temporal resolutions we are working with. Here is a brief roadmap of this paper to explain and help sort out the complex web of different models and datasets used in the analysis:

Section 2 describes a statistical approach for compositional data (Aitchison, 1986) such as land-cover estimates in proportion. To avoid the time-consuming inference in Tjelmeland and Lund (2003), the spatial dependence is modelled using a Gaussian Markov Random Field (GMRF) (Lindgren et al., 2011) with fast inference obtained through R-INLA (Rue et al., 2009; Lindgren and Rue, 2013). Two standard linear regression models and two GMRF-based models are developed to explain RE-VEALS land-cover by various sets of covariates (i.e. estimates from a DVM and an ALCC scenario, elevation, longitude and latitude).

Section 3 describes models and databases used for re-

construction of past and recent (0.05 ka) land cover with the new statistical approach and for data-model comparison. Pollen-based estimates of three land-cover types (coniferous, broadleaved and unforested) at $1° × 1°$ resolution are obtained using the REVEALS model (Sugita, 2007); hereafter those estimates are referred to as grid-based REVEALS (GB-REVEALS). Potential natural vegetation is simulated by a process-based dynamic ecosystem model LPJ-GUESS (Smith et al., 2001), and anthropogenic land cover is extracted from the ALCC KK10 scenario of Kaplan et al. (2009) based on human population history and technology development. KK10 is the existing ALCC scenario that is closest to the pollen-based GB-REVEALS in terms of degree of past deforestation (Trondman et al., 2012; Strandberg et al., 2014; Kaplan et al., 2014). Combined estimates of model-based potential vegetation and ALCC, hereafter referred to as LPJ-GUESS$_{KK10}$, are used as one of the main covariates in the data analysis. In addition, the present-day land cover is obtained from the land-cover database of the European Forest Institute (EFI).

Section 4 describes the results and section 5 discusses the significance and implications of the approach developed in this study. The reconstruction of recent land cover is compared to the EFI forest map for evaluation, and pros and cons of the new statistical approach are assessed in detail.

2. Development of the Statistical Model

2.1. Methods for compositional data

In each grid cell three land-cover types (LCTs) - coniferous forest, broadleaved forest, and unforested land - are expressed as proportions. To account for the restrictions inherent to compositional data we apply logratio transformation (Aitchison, 1986) for the LCT data.

Letting $y_i(\mathbf{s})$ denote the fraction of the ith LCT at grid cell location $s \in R^2$; the values have to sum to one and be non-negative, i.e.

$$\sum_{i=1}^{D} y_i(\mathbf{s}) = 1, \quad \text{and} \quad 0 \leq y_i(\mathbf{s}) \leq 1, \forall i. \tag{1}$$

These conditions complicate any statistical analysis. A common solution (Aitchison, 1986; Tjelmeland and Lund, 2003) is to transform the data, allowing modelling to proceed without being encumbered by the restrictions in Eq. (1). Several possible transformations exist. Here we use the additive logratio (alr) following Tjelmeland and Lund (2003);

$$u_i(\mathbf{s}) = \log \frac{y_i(\mathbf{s})}{y_D(\mathbf{s})}, \quad i = 1, \dots, D-1, \tag{2}$$

with D denoting the number of components ($D = 3$ for our three LCTs). The alr takes the set of D compositional values in [0, 1] and transforms them into $D - 1$ real valued (i.e. unrestricted) data, $u_i(\mathbf{s})$. The original fractions can be recovered from $u_i(\mathbf{s})$

through the inverse transformation:

$$y_i(\mathbf{s}) = \frac{\exp(u_i(\mathbf{s}))}{1 + \sum_i^{D-1} \exp(u_i(\mathbf{s}))}, \quad i = 1, \ldots, D-1,$$
$$y_D(\mathbf{s}) = \frac{1}{1 + \sum_i^{D-1} \exp(u_i(\mathbf{s}))}, \tag{3}$$

where it is easy to see that the $y_i(\mathbf{s})$ obeys the restrictions in Eq. (1).

The alr transformation has its own limitations. It requires proportions to be $y_i(\mathbf{s}) > 0$ and $y_i(\mathbf{s}) < 1$ eliminating the possibility of an equality in Eq. (1). This limitation is not an issue for the data used in this paper, since all LCTs are present in all grid cells. Note that increasing y_1 implies a lowering of y_2 and y_3 through the sum to one constraint; thus u_1 and u_2 are dependent, an important fact for the modelling that is further discussed in Section 2.2.

To compute the difference between two compositions we use the compositional distance (Aitchison, 1986):

$$\Delta(\mathbf{u}, \mathbf{v}) = [(\mathbf{u} - \mathbf{v})^T \mathbf{H}^{-1} (\mathbf{u} - \mathbf{v})]^{1/2} \tag{4}$$

where u and v are alr transforms of compositions and $\mathbf{H}$ is a $d \times d$-matrix ($d = D - 1$) with elements $h_{ij} = 2$ if $i = j$, and $h_{ij} = 1$ if $i \neq j$.

A convenient way of illustrating the variability of D-compositional data is a ternary diagram (see Aitchison, 1986, ch. 1.4). Figure 1 illustrates the concept using four compositional data points, each containing coniferous forest (C), broadleaved forest (B), and unforested land (U). In a ternary diagram, a point close to a vertex (e.g. $y(3)$ close to U) has large proportion of the corresponding vertex and a point close to each edge (e.g. $y(4)$ close to U-C) has a low proportion of the opposite vertex (B).
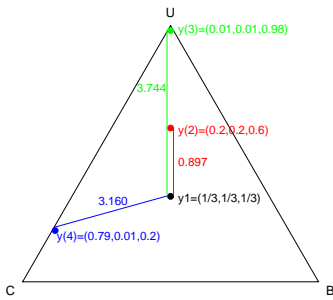

Figure 1: A ternary diagram containing four 3-compositional data points, each consisting of coniferous forest (C), broadleaved forest (B), and unforested land (U). The points correspond to compositions of $y(s) = (y_C(s), y_B(s), y_U(s))$. Numbers along the lines between points indicate their distances according to Eq. (4) from $y(1)$. The figure is inspired by Fig. 1. in Billheimer et al. (2001).

2.2. Statistical model

The transformed data, $(u_1, u_2) = \text{alr}(y_C, y_B, y_U)$, is modelled as a multivariate Gaussian process (see Tjelmeland and Lund, 2003),

$$\begin{bmatrix} \boldsymbol{u}_1 \\ \boldsymbol{u}_2 \end{bmatrix} = \begin{bmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{bmatrix} + A \begin{bmatrix} z_1 \\ z_2 \end{bmatrix} + \boldsymbol{\epsilon}, \quad \begin{bmatrix} z_1 \\ z_2 \end{bmatrix} \in \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_z), \quad \boldsymbol{\epsilon} \in \mathcal{N}(\mathbf{0}, \boldsymbol{I}\sigma_\epsilon^2). \tag{5}$$

Here $\boldsymbol{u}_1$ and $\boldsymbol{u}_2$ are column vectors containing the alr transformed (Eq. (2)) compositional values for the N_{obs} observed locations; $\boldsymbol{\mu}_i$ are column vectors of mean values for each location; $\boldsymbol{\epsilon}$ are independent Gaussian residuals with variance σ_ϵ^2; and z_i are spatially dependent residual fields modelling any remaining dependence in the observations, the $N_{\text{obs}} \times N$ matrix $\boldsymbol{A}$ is a sparse matrix that extracts elements corresponding to the observed locations from z_i. The multivariate Normal model contains two main components: a mean field, $\boldsymbol{\mu}$, and a spatially dependent residual field, z. Those two components are described below.

2.2.1. Mean field

The mean field is modelled as a linear combination of covariates

$$\boldsymbol{\mu}_i = \mathbf{1}\beta_{0,i} + \sum_p \boldsymbol{B}_p(s)\beta_{p,i} \tag{6}$$

where $\boldsymbol{B}_p$ is a column vector containing the pth covariate, $\beta_{p,i}$ are unknown regression coefficients and $\mathbf{1}$ is a column vector of ones (the intercept). Two sets of different covariates are used for vegetation reconstruction in this study: $\boldsymbol{B}$ — contains the alr transformation of LPJ-GUESS$_{\text{KK10}}$ (LPJ-GUESS estimates adjusted for human impact with the KK10 scenario; see next section) and $\boldsymbol{B}_{\text{geo}}$ — contains LPJ-GUESS$_{\text{KK10}}$ and the geographical coordinates, i.e. longitude and latitude. Both $\boldsymbol{B}$ and $\boldsymbol{B}_{\text{geo}}$ also take elevation as a covariate. The geographical covariates are fixed over the different time windows, thus possibly adjusting reconstructions for geographically consistent biases in the potential vegetation.

As an alternative to potential vegetation, some bioclimatic covariates (i.e. temperature, precipitation, and soil suitability), used as drivers for LPJ-GUESS, were also included directly in the mean field. Those alternatives did not improve the reconstructions and, for brevity, the results are neither shown nor discussed in this paper.

Using only the mean field, Eq. (6) and *without* spatially dependent residual fields (i.e. $z_i = 0$), the full model Eq. (5) reduces to a standard linear regression. We construct two Regression Models **RM** and **RM$_{\text{geo}}$** with B and B_{geo}, respectively.

2.2.2. Residual field

The inclusion of coordinates in the mean field only implies a linear dependence between the transformed composition in each grid cell and the corresponding coordinates; no other dependence among neighbouring locations is implied. Any remaining spatial structure can be accounted for by imposing a more complex model for the covariance matrix, $\boldsymbol{\Sigma}_z$, of the residual field z in Eq. (5). Tjelmeland and Lund (2003) used a Gaussian field (GF) specified through the covariance function,

3

and Paciorek and McLachlan (2009) used a thin plate spline to model the spatial structure. Here we replace the GF with a Gaussian Markov Random Field (GMRF) (Lindgren et al., 2011); this has two main benefits:

1. GMRF has computational benefits over the covariance formulations, and

2. it allows the use of standard software (the R-INLA package Lindgren and Rue, 2013; Rue et al., 2009) for inference.

We now briefly present the GMRF model. According to Whittle (1954, 1963) GFs with Matérn covariance

$$\text{cov}(z(\mathbf{0}), z(\mathbf{s})) = \sigma^2 \frac{(\kappa\|\mathbf{s}\|)^\nu K_\nu(\kappa\|\mathbf{s}\|)}{\Gamma(\nu)2^{\nu-1}} \qquad (7)$$

are the solutions to Stochastic Partial Differential Equation (SPDE)

$$(\kappa^2 - \Delta)^{\frac{\alpha}{2}} z(\mathbf{s}) = \tau \mathcal{W}(\mathbf{s}), \qquad \mathbf{s} \in \mathbb{R}^d, \quad \alpha = \nu + d/2, \qquad (8)$$

where $\mathcal{W}(\mathbf{s})$ is Gaussian white noise, $\Delta = \dfrac{\partial^2}{\partial s_x^2} + \dfrac{\partial^2}{\partial s_y^2}$ is the Laplacian, κ is the spatial scale parameter, ν controls the smoothness, τ controls the variance of z and is linked to σ^2 (see Lindgren et al., 2011, ch. 2.1), and K_ν is the modified Bessel function of the second kind. Lindgren et al. (2011) showed that a GMRF representation of a Matérn GF can be explicitly constructed with precision matrix $\mathbf{Q}$. Let the z's in Eq. (5) be a GMRF defined on a regular lattice, then in case of $\alpha = 2$ the appropriate precision matrix is obtained by

$$\mathbf{Q} = \frac{1}{\tau^2}(\kappa^4 \mathbf{C} + 2\kappa^2 \mathbf{G} + \underbrace{\mathbf{G}\mathbf{C}^{-1}\mathbf{G}}_{\mathbf{G2}}) \qquad (9)$$

where $\mathbf{C}, \mathbf{G}$, and $\mathbf{G_2}$ are sparse matrices (see Lindgren et al., 2011, for details). A special case of Eq. (8) is the intrinsic Matérn model with $\kappa = 0$, given by

$$(-\Delta)^{\frac{\alpha}{2}} z(\mathbf{s}) = \tau \mathcal{W}(\mathbf{s}), \qquad (10)$$

this is a spline smoothing model (see Duchon, 1976; Kimeldorf and Wahba, 1970; Nychka, 2000; Wahba, 1981). In this case, $\mathbf{Q}$ is the precision matrix of an intrinsic GMRF.

To obtain a suitable dependence model for z, we assume the same precision $\mathbf{Q}$ for both fields, z_1 and z_2, in Eq. (5), but allow the fields to be correlated. The result is a separable precision (inverse covariance) matrix that can be expressed as a Kronecker product,

$$\mathbf{\Sigma_z}^{-1} = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}^{-1} \otimes \mathbf{Q}, \qquad (11)$$

where ρ is the correlation between the fields.

2.2.3. Interpretation of the statistical model

The purpose of the residual field is to capture spatial structure not explained by the mean field; if GB-REVEALS in several nearby grid cells deviate in a similar fashion from the regression model then predictions at adjacent grid cells should take this information into account. The intrinsic GMRF model (IGMRF) used here can be interpreted as either universal Kriging using a Matérn-covariance with very large range (Lindgren et al., 2011), or as spline smoothing of the residuals (Nychka, 2000; Wahba, 1981). Predictions at an unobserved grid location are essentially given by

$$\hat{u}_i(\mathbf{s}_0) = \hat{\beta}_{0,i} + \sum_p B_p(\mathbf{s}_0)\hat{\beta}_{p,i}$$
$$+ \sum_{j=1}^{2} \sum_{\mathbf{s} \in \text{observed}} c(\mathbf{s} - \mathbf{s}_0, j) \left(u_j(\mathbf{s}) - \hat{\beta}_{0,j} - \sum_p B_p(\mathbf{s})\hat{\beta}_{p,j} \right). \qquad (12)$$

Here $c(\mathbf{s}, i)$ are weighting coefficients that depend on τ, σ_ϵ, and ρ in Eq. (A.2) in Appendix A and decay as the distance from $\mathbf{s}_0$ increases (see Rue and Held, 2004; Lindgren et al., 2011, for details). Note that the predictions include residuals from both fields; this is due to the correlation, ρ, introduced in Eq. (11).

For prediction and cross-validation both the regression parameters (β) and the parameters describing the spatial structure ($\tau, \sigma_\epsilon^2, \rho$) are calibrated based on a validation set (i.e. selected grid cells with GB-REVEALS). These parameters are then used to compute predictions at unobserved sites and sites left-out for the cross-validation.

For details regarding the calibration and reconstruction see Appendix A.

2.2.4. Models used for data analysis

For reconstruction of land-cover types using the GB-REVEALS and LPJ-GUESS$_{\text{KK10}}$ data we consider a total of four models.

The two regression models, RM and RM$_{\text{geo}}$, without spatial dependencies and the two models, IGMRF and IGMRF$_{\text{geo}}$, with spatial dependencies. The two spatially dependent models are created by adding spatial residual fields, z, to the same mean fields, Eq. (6), as in the regression models. All four models have been implemented in R (R Core Team, 2014) using the R-INLA package (Rue et al., 2009; Lindgren and Rue, 2013).

3. Land-cover type and auxiliary data

The target region of the LANDCLIM project is Europe (Fig. 2; Gaillard et al., 2010). The data used for validation and application of the statistical models proposed in the previous section consist of 1) GB-REVEALS extracted from the LANDCLIM-REVEALS database, 2) DVM LPJ-GUESS estimates of potential natural vegetation and ALCC KK10 scenario estimates, and 3) the present-day forest map of Europe, geographical coordinates (longitude and latitude), and elevation.

The two selected time windows of the past are characterized by contrasting human impact on land cover (Gaillard et al.,
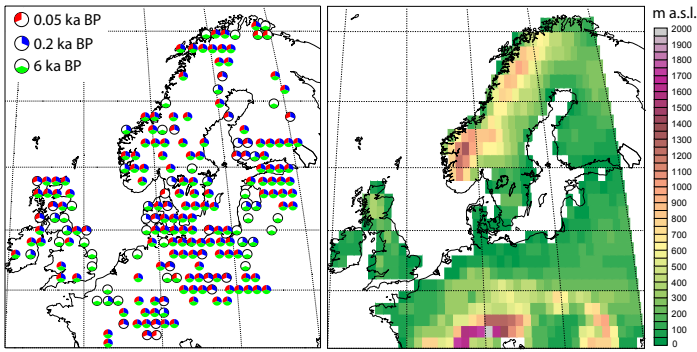
Figure 2: The left plot shows the availability of GB-REVEALS for all three time-windows. The right plot shows the elevation for each grid cell (truncated to ≥ 0).

2010; Strandberg et al., 2014) and recent land cover is used to validate the methods:

6 ka (5.7 - 6.2 ka) — the mid-Holocene warm period characterized by low human impact, often used as a baseline time-window to assess the effects of orbital forcing and pre-industrial greenhouse gases on climate (e.g. Harrison et al., 1998; Kohfeld and Harrison, 2000; Braconnot et al., 2012).

0.2 ka (0.1 - 0.35 ka) — the Little Ice Age (AD 1550 1850), a cool period in Europe with substantial anthropogenic land cover but low levels of human-induced greenhouse gases; AD 1850 and AD 1750 were used as pre-industrial baselines in the two last IPCC (Intergovernmental Panel on Climate Change) reports (Pachauri and Reisinger, 2007; Stocker et al., 2013).

0.05 ka (x[1]-0.1 ka) — recent land cover is characterized by afforestation of large areas of Europe in mountainous areas and other regions, such as southern Sweden, northern Germany and Poland, and the Baltic states (Krzywinski et al., 2009).

3.1. Land-cover reconstruction using fossil pollen

The LCT data used for analysis are calculated from mean REVEALS estimates for 25 major plant taxa in $1° \times 1°$ grid cells following the LANDCLIM protocols described in Gaillard et al. (2010); Mazier et al. (2012); Trondman et al. (2014). REVEALS (Sugita, 2007) is a mechanistic model for regional vegetation reconstruction that takes into account the inter-taxonomic differences in pollen productivity and dispersal, and size of sedimentary basins from which pollen data are obtained. It has been tested in several regions of Europe and North America (Hellman et al., 2008a,b; Soepboer et al., 2010; Sugita et al., 2010). Hellman et al. (2008b) showed that the spatial scale of REVEALS estimates of land cover was generally closer to 100 km × 100 km than 50 km × 50 km; however, the difference in fit between the REVEALS estimates and the actual land cover was

[1]x = date of the core surface, e.g. AD 2005-100 BP if x= AD 2005

small between the two spatial scales. Therefore, the $1° \times 1°$ spatial scale chosen for the LANDCLIM project is adequate for the REVEALS model. The LANDCLIM database includes more than 600 Holocene pollen records that are compiled from the European Pollen Database, various national pollen databases and archives, and individual contributors. In our study, the number of $1° \times 1°$ grid cells with GB-REVEALS are 184, 179, and 168 (out of a total of 644, 675, and 658 grid cells over the study region) for the 6 ka, 0.2 ka, and 0.05 ka time windows, respectively (Fig. 2). The total number of grid cells and the number of grid cells with GB-REVEALS differ among the time periods because of the differences in the coastline of the ALCC scenario used, and the differences in availability of pollen data among the time windows, respectively.

3.2. Estimates of potential vegetation and anthropogenic land cover

For estimating changes in the distribution and cover of LCTs in each $1° \times 1°$ grid cell, we use a combination of simulated potential vegetation and estimates from ALCC scenario as follows.

3.2.1. Potential natural vegetation

Potential vegetation in the study region is simulated by a process-based dynamic ecosystem model LPJ-GUESS 2.1 (Lund-Potsdam-Jena-General Ecosystem Simulator, Smith et al., 2001; Sitch et al., 2003). For the three, above specified, time windows LPJ-GUESS was run using climate input data provided by 1) the SMHI Rossby Centre Regional Climate Model (RCA3) (Samuelsson et al., 2011) with a 0.44° spatial resolution over Europe for 6 ka (5859–5811) BP and 0.2 ka (AD 1700–1800), (Strandberg et al., 2014) and 2) the Climatic Research Unit with a 0.5° resolution for 0.05 ka (AD 1901–2006) (Mitchell and Jones, 2005). The modern soil-texture data as described in Sitch et al. (2003) were used in all simulations. Percentage covers of plant functional types simulated by LPJ-GUESS are averaged over the modelled periods and converted to the three LCTs in proportion. All the estimates are upscaled, by averaging, to $1° \times 1°$ resolution. See Appendix B for a more detailed description of mechanisms and parameterizations in LPJ-GUESS.

3.2.2. Anthropogenic deforestation

Anthropogenic deforestation in the study region is extracted from the standard scenario of the ALCC KK10 (Kaplan et al., 2009). The KK10 scenario is based upon estimates of past human population density and the land requirement per capita to estimate the area of land needed for sustaining the assumed population. The spatial distribution of anthropogenic land cover is determined by environmental suitability estimates based mainly on climate conditions and soil type. We chose the KK10 scenario to represent anthropogenic deforestation, because there is a good correlation between GB-REVEALS and KK10 (Trondman et al., 2012; Strandberg et al., 2014; Kaplan et al., 2014). The KK10 scenario provides estimates of the fraction of land used for agrarian activities (i.e. deforested land) at

5′ spatial resolution. These estimates are averaged for 100-year time windows around 6 ka and 0.2 ka BP and upscaled to $1° \times 1°$ resolution.

3.2.3. LCT estimates based on the LPJ-GUESS and KK10 simulations

Because LPJ-GUESS does not account for the increase in unforested area due to human impact (Fig. 3), this study uses the following adjustment to estimate land cover:

$$\mathbb{P}(\text{Coniferous}_{\text{adj.}}) = \mathbb{P}(\text{Coniferous}) \cdot (1 - \mathbb{P}(\text{HLU}_{\text{KK10}})),$$
$$\mathbb{P}(\text{Broadleaved}_{\text{adj.}}) = \mathbb{P}(\text{Broadleaved}) \cdot (1 - \mathbb{P}(\text{HLU}_{\text{KK10}})),$$
$$\mathbb{P}(\text{Unforested}_{\text{adj.}}) = \mathbb{P}(\text{Unforested}) \cdot (1 - \mathbb{P}(\text{HLU}_{\text{KK10}})) + \mathbb{P}(\text{HLU}_{\text{KK10}}).$$
(13)

In each grid cell, the LPJ-GUESS-based proportions of the area occupied by coniferous forest, broadleaved forest and unforested land are expressed as $\mathbb{P}(\text{Coniferous})$, $\mathbb{P}(\text{Broadleaved})$ and $\mathbb{P}(\text{Unforested})$, respectively; the estimated proportion of human induced deforestation from the KK10 scenario is $\mathbb{P}(\text{HLU}_{\text{KK10}})$. Adjusted land cover proportions — $\mathbb{P}(\text{Coniferous}_{\text{adj.}})$, $\mathbb{P}(\text{Broadleaved}_{\text{adj.}})$ and $\mathbb{P}(\text{Unforested}_{\text{adj.}})$ — are calculated in such a way that the cover fractions of all the potential vegetation components are uniformly reduced by the predicted anthropogenic land-cover proportion. The anthropogenic land-cover proportions, $\mathbb{P}(\text{HLU}_{\text{KK10}})$, are then added to the unforested fraction, $\mathbb{P}(\text{Unforested})$. The resulting adjusted land-cover proportions are then used as explanatory variables in the mean field, Eq. (6).

3.3. Present-day land cover, elevation, longitude and latitude

Data on the present-day land cover in the study region were obtained from the forest map of Europe compiled by the European Forest Institute (EFI). Raster maps based on a combination of satellite data (NOAA-AVHRR) and national forest-inventory statistics from 1990–2005 (Päivinen et al., 2001; Schuck et al., 2002) were downloaded from the EFI webpage (`http://www.efi.int/portal/virtual_library/information_services/mapping_services/forest_map_of_europe`). The forest maps (with proportions of coniferous- and broadleaved-forest cover) were upscaled, by averaging, from 1 km × 1 km to $1° \times 1°$ resolution. The proportion of unforested area were calculated by subtracting the total sum of forested covers from 1.0.

The elevation data were obtained from the Shuttle Radar Topography Mission (SRTM) (Becker et al., 2009) downloaded from `ftp://topex.ucsd.edu/pub/srtm30_plus/` on 2011–09–03, averaged over each GB-REVEALS grid-cell, and truncated to ≥ 0 to avoid a few grid cells along Norway's coast with elevation down to -1000. The geographical coordinates consist of the longitude and latitude of the central point of each GB-REVEALS grid cell.

4. Results

4.1. Evaluation of the statistical models

To evaluate and validate the four statistical models, we compared the differences between the reconstructed values for the
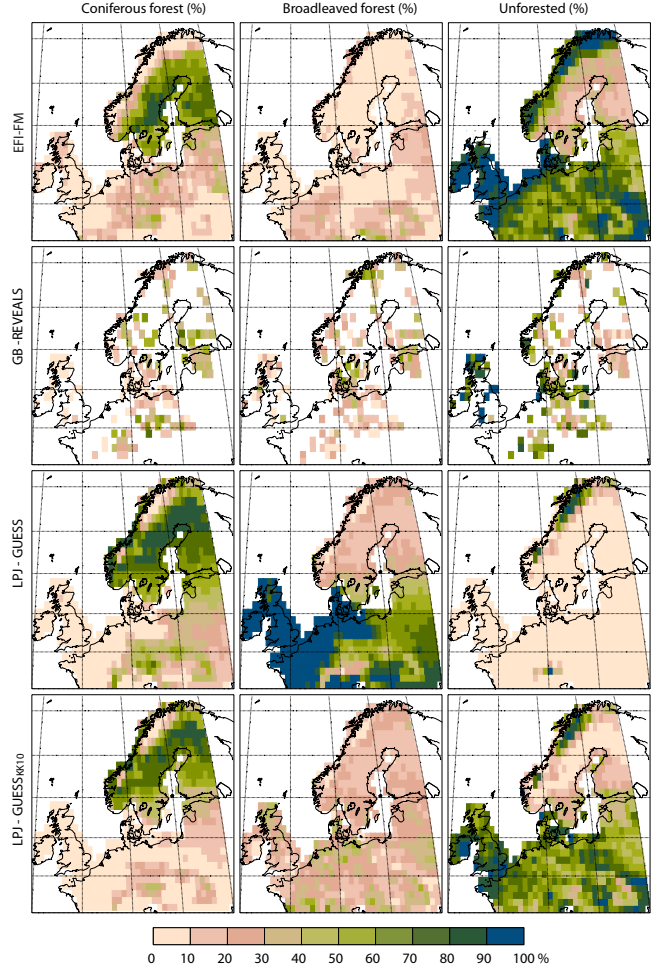

Figure 3: Available data for the 0.05 ka period, showing the proportion of LCTs. From top to bottom, data for the present-day window from EFI-FM, GB-REVEALS, LPJ-GUESS, and LPJ-GUESS$_{\text{KK10}}$ (Eq. (13)).

0.05 ka and the data from the EFI forest map (EFI-FM) using compositional distances (Eq. (4)) for individual grid cells. The average distances (i.e. the mean difference between the model-based reconstruction and EFI-FM) are 1.711, 1.520, 1.782, and 1.517 for IGMRF$_{\text{geo}}$, IGMRF, RM$_{\text{geo}}$, and RM, respectively. Thus the models without geographic coordinates (RM and IGMRF) provide estimates closer to those from EFI-FM than the other two models.

We also adopted a 6-fold cross-validation scheme for each of the three time windows (see Hastie et al., 2001, ch. 7.10). To assess the possible variability due to the selection of different groupings, the cross-validation is run for 10 different, randomly selected, 6 folds. Average compositional errors and standard deviations are shown in Table 1. For the cross-validation (CV) the RM model is consistently best for all time windows followed by RM$_{\text{geo}}$, IGMRF, and IGMRF$_{\text{geo}}$. However, longitudinal effects introduced by the geographic coordinates result in unsatisfactory reconstructions from the RM$_{\text{geo}}$ and IGMRF$_{\text{geo}}$ model for areas of eastern Europe with few GB-REVEALS. Due to the scarcity of the GB-REVEALS data, these longitudinal effects are not penalised by the cross-validation, but show

up in the comparisons with EFI-FM data; this leads us to prefer IGMRF over the RM$_{geo}$.

In addition, we computed the deviance information criteria (DIC; see Ch. 7.2 Gelman et al., 2014) for each of the models (Table 2); the DIC is a generalization of the Akaike information criterion (AIC; Akaike, 1969). The DIC suggests that the IGMRF models outperform the regression models for all time windows, indicating the need for spatial dependency. The results of CV and DIC lead us to choose RM and IGMRF as the best model.

Accordingly we proceeded with the data analysis using the RM and IGMRF only. The reconstructions from the RM and IGMRF are shown in Fig. 4.

Model	0.05 ka		0.2 ka		6 ka	
	CV$_{error}$	(sd)	CV$_{error}$	(sd)	CV$_{error}$	(sd)
RM	1.565	(0.033)	1.843	(0.057)	1.761	(0.041)
RM$_{geo}$	1.631	(0.051)	1.985	(0.068)	1.825	(0.049)
IGMRF	1.679	(0.046)	2.020	(0.071)	1.928	(0.048)
IGRMF$_{geo}$	1.705	(0.049)	2.060	(0.077)	1.956	(0.053)

Table 1: Average compositional error (and standard deviation) from 10 different 6-fold cross-validations for each of the 4 different models, and 3 time windows.

DIC	0.05 ka	0.2 ka	6 ka
IGMRF$_{geo}$	750.13	891.84	-2123.6
IGMRF	664.05	839.51	-2128.82
RM$_{geo}$	1058.59	1194.41	1319.18
RM	1142.96	1348.56	1372.9

Table 2: Deviance information criteria (DIC) for each of the 4 different models, and 3 time windows.

4.2. Assessment of the data quality

To gain an overall understanding of data quality and to detect possible method-inherent biases, the compositional distances are calculated between the EFI-FM and either LPJ-GUESS$_{KK10}$, GB-REVEALS, or the statistical reconstructions, RM and IGMRF, for the 0.05 ka. The distances between LPJ-GUESS$_{KK10}$ and GB-REVEALS are also computed for each of the three time windows, allowing us to investigate how much these two datasets differ in each time window. The discrepancy between LPJ-GUESS$_{KK10}$ and GB-REVEALS increases from the 0.05 ka to 6 ka (average distance of 1.644, 1.701, and 2.054 for the 0.05 ka, 0.2 ka, and 6 ka windows, respectively; Fig. 5). Reasons behind these discrepancies are presented in Sec. 5.2.

For the 0.05 ka time window the RM- and IGMRF-based reconstructions are closer to EFI-FM (average distance: 1.517 and 1.519, respectively) than either LPJ-GUESS$_{KK10}$ (2.081) or GB-REVEALS (1.675) alone. Further, the RM- and IGMRF-based reconstructions, when averaged *only over grid cells* where we have GB-REVEALS, are closer to the EFI-FM data than GB-REVEALS — RM 1.499, IGMRF 1.592, and GB-REVEALS 1.675. Thus, the statistical modelling approach reduces the compositional error of the land-cover reconstructions.
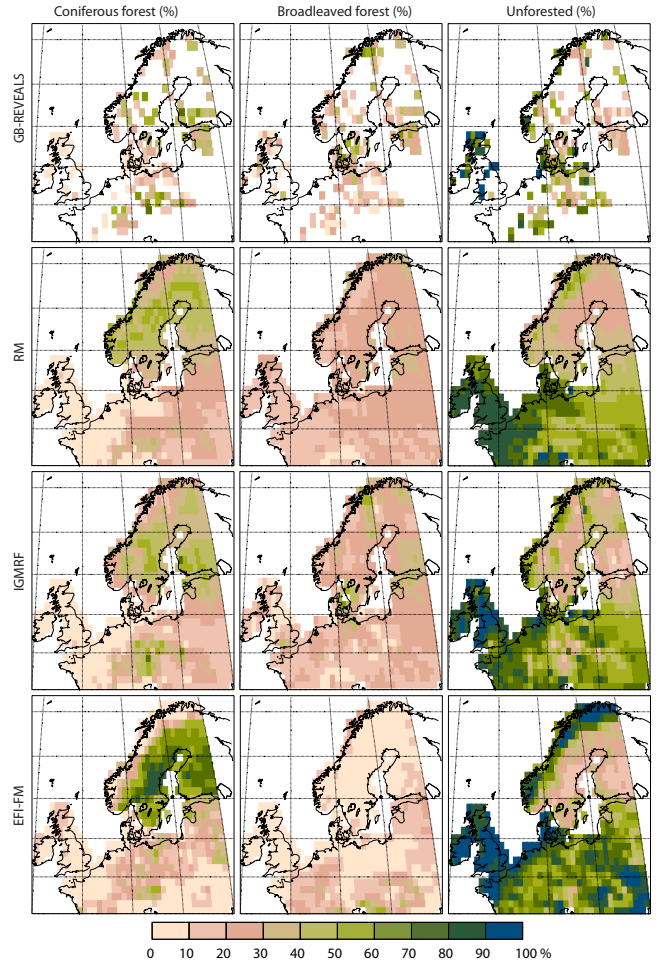

Figure 4: Reconstructions of proportion of LCTs for the 0.05 ka time window. From top to bottom, the REVEALS data, the RM reconstruction, the IGMRF reconstruction and the EFI-FM data.

Spatially, discrepancies between EFI-FM and GB-REVEALS or LPJ-GUESS$_{KK10}$ are noticeable along the coastal areas of western Europe, in central Sweden and southern Finland, with the largest disagreement at the northern British Isles. The discrepancies are more pronounced for LPJ-GUESS$_{KK10}$ than either GB-REVEALS or one of the proposed statistical model (Fig. 5). When comparing LPJ-GUESS$_{KK10}$ and GB-REVEALS, the patchy nature of GB-REVEALS makes it hard to distinguish any specific spatial patterns among the discrepancies.

4.3. Qualitative differences among the models

In general, IGMRF captures more of the local variability in GB-REVEALS than RM, while RM smooths GB-REVEALS. For example, the high variation of unforested land in Britain and the Alps is well captured by IGMRF while RM is capable of capturing the gradual changes in vegetation abundances observable along the western coast of Norway and around the northern Baltic (Fig. 4).

Both statistical models overestimate the abundance of broadleaved forest relative to the EFI-FM data. Figure 6 shows that the low abundance of broadleaved forest in EFI-FM can
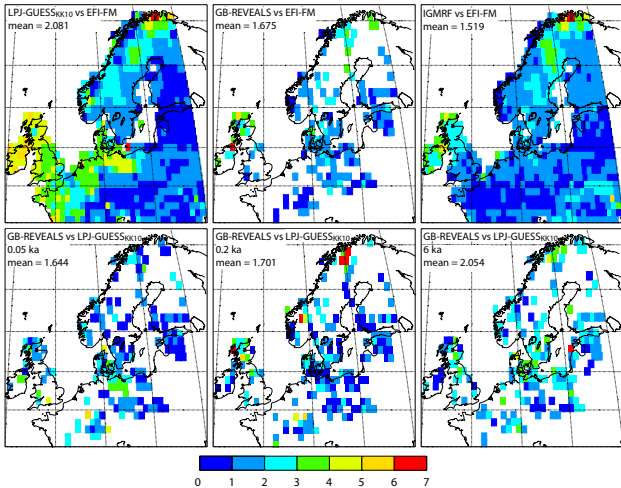
Figure 5: The compositional distances among different datasets. The first row shows the distances between EFI-FM, and (from left to right) 1) LPJ-GUESS$_{KK10}$, 2) GB-REVEALS, and 3) one of the proposed models, IGMRF for the 0.05 ka time window. The second row shows average distances between GB-REVEALS and LPJ-GUESS$_{KK10}$ for three time-windows, 0.05 ka, 0.2 ka, and 6 ka.
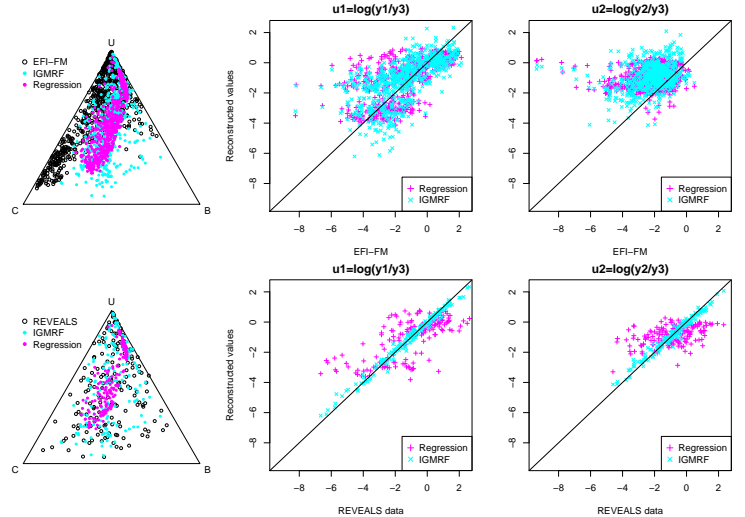

Figure 6: The reconstructions from RM and IGMRF against the EFI-FM data on the top row, and against the 0.05 ka GB-REVEALS on the bottom. The first column shows the ternary diagram for three compositions (C: Coniferous forest, B: Broadleaved forest, and U: Unforested land). Note the cluster of data points along the U-C edge, representing the low abundance of broadleaved forest in the EFI-FM data which does not exist in the 0.05 ka GB-REVEALS. The remaining columns show the scatter plots of alr-transformed, (u_1, u_2), reconstructed values for IGMRF and RM.

be seen as a cluster of EFI-FM data along the U-C edge (corresponding to $\sim 0\%$ of broadleaved cover) for which no matching GB-REVEALS exists. This is due to GB-REVEALS having a higher abundance of broadleaved forest than EFI-FM (Fig. 6).

4.4. Statistical reconstruction of past land cover

Figures 7 and 8 show the reconstructed land cover at 0.2 ka and 6 ka using RM and IGMRF. In northern Europe, there is a general shift from largely unforested to more coniferous-dominated land cover between 0.2 ka and 0.05 ka (Fig. 4). The shift reflects a considerable decrease in agrarian land use in favour of modern forestry with conifer species in many regions (e.g. Poska et al., 2008; Fredh et al., 2013). Both models capture the shift (Figs. 4 and 7). Further, IGMRF is capable of capturing a number of sub-regional structures such as the high abundance of unforested land in Britain, especially in its northernmost regions (Fyfe et al., 2013) and the high abundance of coniferous forest around the Alps, central Sweden, and southern Finland; these patterns are also present in GB-REVEALS for 0.2 ka.

Broadleaved forest is a major constituent of the land cover at 6 ka over Europe (Fig. 8). IGMRF captures the GB-REVEALS structure with a locally highly varying (between 20% and 80%) abundance of unforested land in western Europe and the Carpathians, while the RM produces a smoother reconstruction with an averaged (around 40% to 50%) and regionally smooth abundance of unforested land. IGMRF also captures the higher than average abundance of coniferous forest in the south-eastern Baltic states and the Alps, while RM only captures these features around the Alps.

Statistical reconstructions of land cover for the 0.05 ka and 0.2 ka time windows show a good fit of the statistical models to the EFI-FM data (for 0.05 ka) and to GB-REVEALS (for both times). For the 6 ka time window the increasing discrepancies

between GB-REVEALS and LPJ-GUESS$_{KK10}$ makes it hard for the statistical models to combine the two data sets; resulting in either over-smoothing (RM) or exaggeration (IGMRF) of local structures in the GB-REVEALS.

5. Discussion

5.1. Advantage of the new approach over previously proposed methods

The statistical method developed and used in this paper utilizes GMRFs (Lindgren et al., 2011) and the R-INLA package (Rue et al., 2009; Lindgren and Rue, 2013) to obtain fast inference for a complex statistical model using standard tools. Spatially dependent compositional data has previously been modelled using similar approaches with different specifications of the latent field, different applications, and more time consuming calibration methods (e.g. Billheimer et al., 2001; Tjelmeland and Lund, 2003; Paciorek and McLachlan, 2009).

In addition to the statistical methods used, the inclusion of estimates from LPJ-GUESS and KK10 provides a way of capturing the non-linear effects of bioclimatic variables on land cover by combining data from a DVM (LPJ-GUESS) and an ALCC scenario (KK10) with grid-based pollen estimates (GB-REVEALS). All the suggested models rely, to some extent, on covariates and to obtain a good reconstruction it is important to identify and use covariates with strong explanatory power. For regions with few GB-REVEALS the statistical modelling will essentially extrapolate the covariate behaviour observed in other regions with more data. For our study the DVM based covariates produce good reconstructions, while the geographic
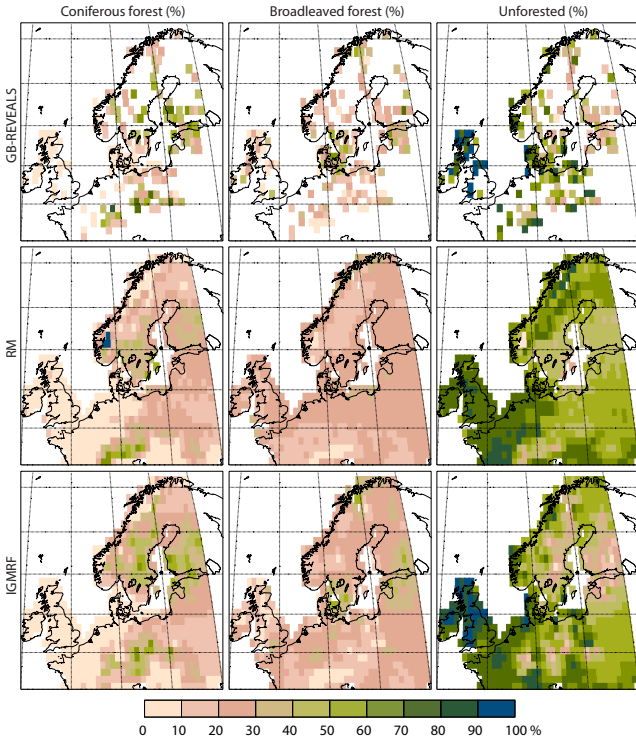
Figure 7: Reconstructions for the 0.2 ka time window of proportion of LCTs. From top to bottom, GB-REVEALS, the RM reconstruction, and the IGMRF reconstruction.
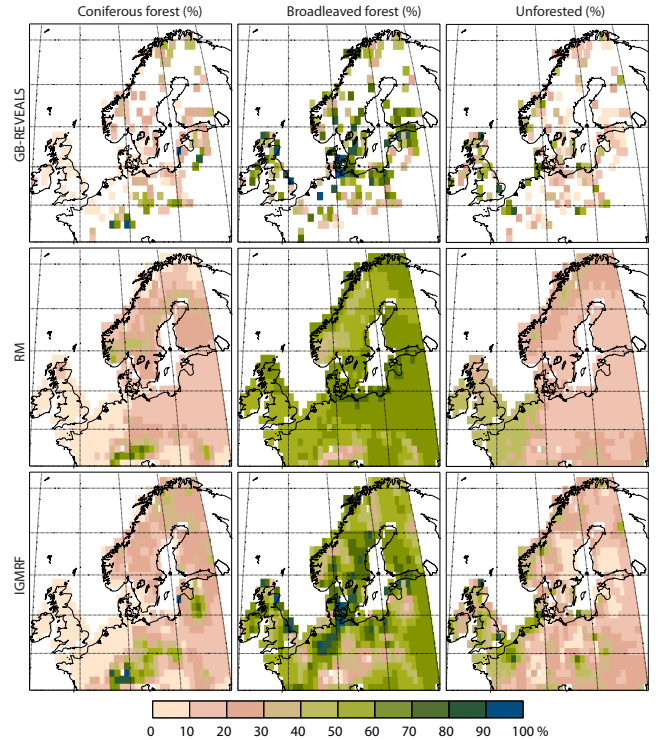

Figure 8: Reconstructions for the 6 ka time window of proportion of LCTs. From top to bottom, GB-REVEALS, the RM reconstruction, and the IGMRF reconstruction.

covariates in the RM_{geo} model lead to a reconstruction exhibiting longitudinal effects that do not correspond to patterns in the EFI-FM data or in the GB-REVEALS.

Pollen-inferred land cover has also been studied by Paciorek and McLachlan (2009), but over a much smaller geographic area (roughly equivalent to one of the grid cells in our approach) than our continent-wide focus. Paciorek and McLachlan (2009) combined pollen data with maps of vegetation abundances to produce a combined estimate of pollen productivity and land cover. Thus the focus and geographic extent of their paper is closer to that of REVEALS (Sugita, 2007) than to ours. Another potential use of the pollen data (Garreta et al., 2010) is to attempt recovery of past climate by the inversion of a dynamic vegetation model; this approach provides past climate, but does not provide spatial reconstructs of the past land-use.

Although the model provides good results, it may be improved, for instance by introducing a more elaborate method of adjusting potential vegetation for human deforestation, Eq. (13) than that used in this study. Currently, we assume that human deforestation affects the three land cover classes equally. An interesting extension would be to include, and estimate, the differential impact of deforestation on the natural vegetation types. Another potential improvement would be the inclusion of the estimated uncertainties in GB-REVEALS; this might allow the model to disregard uncertain GB-REVEALS values, possibly improving the reconstruction.

5.2. Biases of land-cover databases and their effects on RM and IGMRF applications

5.2.1. GB-REVEALS

Possible sources of errors and biases in the GB-REVEALS are discussed in details in Trondman et al. (2014). Here we mention the most important ones. The pollen productivity estimates (PPEs) used to obtain the GB-REVEALS are based on pollen and vegetation data from low-land areas of NW and W Europe (Broström et al., 2008; Mazier et al., 2012). This might lead to biased estimates of the regional vegetation in areas where region-specific PPEs are not available, such as the high mountain areas of Norway, the Czech Republic and the Alps. Also, the use of all pollen records available in each grid cell, i.e. pollen data from both lakes and bogs (small or large in size) in the LANDCLIM-REVEALS database, may create errors caused by variations in the number, type and size of the sites used to calculate the GB-REVEALS (Mazier et al., 2012; Trondman et al., 2014). The best pollen records for the application of REVEALS are those from large lakes (one or several, > 20–30 ha) or multiple small lakes (≤ 20–30 ha; see Sugita, 2007; Hellman et al., 2008a, for details). If REVEALS is applied on pollen data from small sites only, a small number of sites may result in very large error estimates (Sugita, 2007; Mazier et al., 2012; Trondman et al., 2014). Further, the REVEALS model assumes that no vegetation is growing on the surface of the basin where pollen is deposited (Sugita, 2007), which applies to lakes only. Although REVEALS includes two versions of the pollen dispersal and deposition model (one for

9

lakes and one for bogs) the assumption mentioned above is violated. Therefore, pollen data from large bogs in particular might bias the GB-REVEALS due to the local vegetation on the bog (Trondman et al., 2014).

In addition, locally grown shade-intolerant deciduous trees, such as *Alnus* and *Betula*, on the wetland and along the shores of the lakes tend to be over-represented in pollen records. This may bias the REVEALS reconstructions because the model cannot fully correct for the over-representation. Although broadleaved trees are over-represented compared to EFI-FM in southern Sweden in this study, the validation of the RE-VEALS model in the same areas do not exhibit any such over-representation (Hellman et al., 2008b,a).

When comparing EFI-FM with the 0.05 ka GB-REVEALS it is important to note that: i) due to the spectral reflectance of broadleaved forest in combination with that of water resembling quite closely the spectral reflectance of coniferous forest, EFI tends to underestimate broadleaved forest along water courses (Schuck et al., 2002); ii) EFI-FM had the lowest accuracy for broadleaved forest (c.f. Section 5.2.4); and iii) the GB-REVEALS cover a much longer time interval (from AD 1850 to the year of coring at each site) than the EFI-FM (inventory and satellite data from 1990–2005) and, during the past century, parts of our study region were characterized by the abandonment of traditional agriculture in favour of sylviculture with plantations of *Picea* and *Pinus* (e.g. Poska et al., 2008; Krzywinski et al., 2009; Fredh et al., 2013; Cui et al., 2014).

Any biases in GB-REVEALS will be propagated into the statistical reconstructions and it is not possible to assess the detailed effects of such biases on the RM and IGMRF applications neither in qualitative nor in quantitative terms. But it is important to have these possible biases in mind when the RM and IGMRF reconstructions are discussed and validated against present-day data. Nevertheless, GB-REVEALS used in this study are mostly credible for the time interval they represent (Trondman et al., 2014) and any biases should be small.

5.2.2. LPJ-GUESS

The largest discrepancies between LPJ-GUESS simulated land cover and present-day EFI-FM or palaeo-based GB-REVEALS of past land cover coincide with areas characterized by high rainfall and/or long-term anthropogenic land cover and inherent specific land-cover types such as heathlands or blanket-bogs that are difficult to model using a natural terrestrial vegetation models. For instance, the LPJ-GUESS standard soil biogeochemistry used in this study, which excludes the nutrient cycle (Sitch et al., 2003), can lead to imprecise estimates of vegetation composition in nutrient-limited environments, for example at high latitudes (Wårlind, 2013). Moreover, the absence of dispersal and migratory processes in the LPJ-GUESS standard setup (Smith et al., 2001) leads to an overrepresentation of coniferous forest (spruce in particular) in central and northern Europe, especially during the early and mid Holocene (Lehsten et al., 2014), which may affect the land-cover reconstruction at 6 ka. Post-processing of the LPJ-GUESS simulated natural vegetation for migration processes of taxa such as *Fagus* (beech) may decrease the difference between LPJ-GUESS

and GB-REVEALS at 6k (Poska et al., 2012).

At 0.05 ka (Fig. 3) there is a similar discrepancy between LPJ-GUESS$_{KK10}$ and GB-REVEALS along the coasts of Norway. Here, LPJ-GUESS simulates high cover of coniferous forest while GB-REVEALS exhibit higher cover of broadleaved forest. This is a consequence of the bias in LPJ-GUESS mentioned above, but also of long-term human impact in these regions with the development of grazed heaths from the Neolithic time that still cover large areas (e.g. Gaillard et al., 2009).

Further, the LPJ-GUESS estimates of land-cover composition are highly dependent on the climate input data (RCA3 simulations of past climate) used to force LPJ-GUESS. Strandberg et al. (2014) showed that there are some discrepancies between proxy-based reconstructions of past climate and RCA3 simulations; although both climate palaeo-proxies and RCA3 simulations show higher temperatures at 6 ka than at 0.2 ka, the difference in magnitude between the two time windows and the geographical/spatial patterns of reconstructed versus simulated temperature and precipitation can be very large. Biases are seen in particular in the Scandinavian mountains and in eastern and north-eastern Europe. The latter will in turn bias the LPJ-GUESS simulated vegetation and, therefore, the RM and IGMRF applications.

5.2.3. ALCC KK10

The ALCC scenario used for adjustment of LPJ-GUESS is based on the following assumptions: 1) the parameters that drive deforestation are similar in different population regions, 2) the areas with highest suitability for farming are deforested first, and 3) agricultural products were the major food source for human populations (Kaplan et al., 2009). As the extent and intensity of population pressure on the landscape may be characterized by strong regional to local-scale spatial and temporal differences in terms of technology development and usage of non-agricultural food resources, these assumptions might cause over- or underestimations of deforestation, especially for the far past (here 6 ka). For instance, the high fractions of deforested land in the ALCC scenario at 6 ka in southern Sweden and Belgium do not seem reasonable when compared to GB-REVEALS at individual sites in e.g. southern Sweden (Gaillard et al., 2010; Cui et al., 2013).

Further, the low fraction of deforested land along the coasts of Norway in the ALCC scenarios at 0.05 ka does not agree with the cover of unforested land in EFI-FM and GB-REVEALS. The ALCC scenario underestimates unforested land in these areas because the geographical and geological characteristics do not correspond to conditions associated with good suitability for farming. As a consequence, the LPJ-GUESS$_{KK10}$ estimates of deforested land may bias the RM and IGMRF applications. Moreover, the correction of the LPJ-GUESS estimates with the ALCC scenario assumes that all three LCTs are equally suitable for human land use, which is not necessarily the case. Many archaeological and palaeoecological studies in Europe have shown that the areas covered by deciduous forests tended to be deforested first for cultivation and grazing because of the favourable soil conditions (e.g. Gaillard and Göransson,

1991; Poska et al., 2004). The latter could also bias the RM and IGMRF results.

5.2.4. EFI-FM

The quality assessment of the EFI-FM by Kempeneers et al. (2012) shows 88% overall accuracy of the dataset, with accuracy for broadleaved forest being the lowest at 58%. The mapping performance was found to be spatially varying, with the best fit to ground observations in central Europe and an underestimation of tree cover in areas of sparse forest cover in Spain, Ireland and parts of Finland. This, together with the temporal miss-alignment between EFI-FM and GB-REVEALS discussed in Sec. 5.2.1, implies that model comparisons at the 0.05 ka time window needs some caution.

5.3. Implications of the results

The RM and IGMRF models show a potential to provide spatially more explicit and realistic reconstruction of the Holocene land cover than LPJ-GUESS, ALCC KK10 or REVEALS do alone.

The balance between relying on covariates (i.e. RM and mean field in IGMRF) or on nearby observations (i.e. spatial dependency part in IGMRF) is an issue in spatial statistical reconstructions. The RM model essentially consists of a regression of GB-REVEALS onto covariates (LPJ-GUESS$_{KK10}$ and elevation). Large spatially varying discrepancies between GB-REVEALS and the covariates can result in an inadequate mean field, which needs to be compensated through spatial dependencies. For the 6 ka time window this is evident in the very smooth reconstructions from the RM, and the overfitting of IGMRF to GB-REVEALS. It is important to note that the RM primarily captures the large-scale variability in land cover, while IGMRF mainly captures details on a regional scale. Credible IGMRF reconstructions obviously require that the GB-REVEALS point data are reliable and that deviations from the mean model, Eq. (6), are spatially smooth. For areas with few GB-REVEALS (e.g. the northern Baltic region at 6 ka in our study, Fig. 8) the scarce data may provide a too strong local influence on the IGMRF reconstruction. In such cases, the RM reconstructions will be safer to use because individual GB-REVEALS play a less important role in the local statistical reconstruction.

6. Conclusions

The results presented here suggest that it is possible to statistically combine pollen-based reconstructions of land cover with simulated potential land cover and ALCC scenario to create spatially-explicit estimates of past land cover over large areas, such as Europe. Accurate estimates of past land cover is important, allowing for the assessment of biogeophysical effects of vegetation and land-use changes on past climate.

The proposed best models provide good reconstructions for the 0.05 ka and 0.2 ka time windows, although highlighting

slightly different features. The larger differences among GB-REVEALS, LPJ-GUESS$_{KK10}$, and the statistical reconstructions at 6 ka suggest that further modifications and developments of the models are necessary to improve the estimates of land cover in older time periods. Future improvements may be possible by: 1) using a more flexible way of combining an ALCC scenario with estimates from a DVM, i.e. accounting for the varying suitability of land-cover types for agrarian activities, 2) including the error estimates of GB-REVEALS in the statistical modelling.

These pollen-based, spatially continuous land-cover reconstructions can then be used in the analysis of landscape ecological complexity in time and space (particularly the IGMRF) and in climate simulations (preferably RM) following e.g. the same scheme as Strandberg et al. (2014). A similar approach can be applied in other parts of the world, such as China, India and Africa, where long and extensive human activities have modified the earth surface significantly.

References

Aitchison, J., 1986. The statistical analysis of compositional data. Chapman & Hall, Ltd.

Akaike, H., 1969. Fitting autoregressive models for prediction 21, 243–247.

Becker, J. J., Sandwell, D. T., Smith, W. H. F., Braud, J., Binder, B., Depner, J., Fabre, D., Factor, J., Ingalls, S., Kim, S. H., Ladner, R., Marks, K., Nelson, S., Pharaoh, A., Sharman, G., Trimmer, R., VonRosenburg, J., Wallace, G., Weatherall, P., 2009. Global bathymetry and elevation data at 30 arc seconds resolution: SRTM30_PLUS. Mar. Geod. 32 (4), 355–371.

Billheimer, D., Guttorp, P., Fagan, W. F., 2001. Statistical interpretation of species composition. J. Am. Statist. Assoc. 96 (456), 1205–1214.

Braconnot, P., Harrison, S. P., Kageyama, M., Bartlein, P. J., Masson-Delmotte, V., Abe-Ouchi, A., Otto-Bliesner, B., Zhao, Y., 2012. Evaluation of climate models using palaeoclimatic data. Nature Clim. Change 2 (6), 417–424.

Broström, A., Nielsen, A. B., Gaillard, M.-J., Hjelle, K., Mazier, F., Binney, H., Bunting, J., Fyfe, R., Meltsov, V., Poska, A., et al., 2008. Pollen productivity estimates of key European plant taxa for quantitative reconstruction of past vegetation: a review. Veg. Hist. Archaeobot. 17 (5), 461–478.

Brovkin, V., Bendtsen, J., Claussen, M., Ganopolski, A., Kubatzki, C., Petoukhov, V., Andreev, A., 2002. Carbon cycle, vegetation, and climate dynamics in the Holocene: Experiments with the CLIMBER-2 model. Global. Biogeochem. Cy. 16 (4), 1139.

Brovkin, V., Claussen, M., Driesschaert, E., Fichefet, T., Kicklighter, D., Loutre, M., Matthews, H., Ramankutty, N., Schaeffer, M., Sokolov, A., 2006. Biogeophysical effects of historical land cover changes simulated by six Earth system models of intermediate complexity. Clim. Dynam. 26 (6), 587–600.

Christidis, N., Stott, P. A., Hegerl, G. C., Betts, R. A., 2013. The role of land use change in the recent warming of daily extreme temperatures. Geophys. Res. Lett. 40 (3), 589–594.

Claussen, M., Brovkin, V., Ganopolski, A., 2001. Biogeophysical versus biogeochemical feedbacks of large-scale land cover change. Geophys. Res. Lett. 28 (6), 1011–1014.

Cui, Q., Gaillard, M.-J., Lemdahl, G., Stenberg, L., Sugita, S., 2014. Historical land-use and landscape chaneg in southern Sweden and implications for present and future biodiversity, submitted to Ecology and Evolution.

Cui, Q.-Y., Gaillard, M.-J., Lemdahl, G., Sugita, S., Greisman, A., Jacobson, G. L., Olsson, F., 2013. The role of tree composition in Holocene fire history of the hemiboreal and southern boreal zones of southern Sweden, as revealed by the application of the landscape reconstruction algorithm: Implications for biodiversity and climate-change issues. The Holocene 23 (12), 1747–1763.

de Noblet-Ducoudré, N., Boisier, J.-P., Pitman, A., Bonan, G., Brovkin, V., Cruz, F., Delire, C., Gayler, V., van den Hurk, B., Lawrence, P., van der Molen, M. K., Müller, C., Reick, C. H., Strengers, B. J., , Voldoire, A., 2012. Determining robust impacts of land-use-induced land cover changes on surface climate over North America and Eurasia: results from the first set of LUCID experiments. J. Climate 25 (9), 3261–3281.

Duchon, J., 1976. Splines minimizing rotation invariant seminorms in Sobolev spaces. In: Schempp, W., Zeller, K. (Eds.), Constructive Theory of Functions of Several Variables. Springer-Verlag, pp. 85–100.

Fredh, D., Broström, A., Rundgren, M., Lagerås, P., Mazier, F., Zillén, L., 2013. The impact of land-use change on floristic diversity at regional scale in southern Sweden 600 BC—AD 2008. Biogeosciences 10 (5), 3159–3173.
URL http://www.biogeosciences.net/10/3159/2013/

Fyfe, R. M., Twiddle, C., Sugita, S., Gaillard, M.-J., Barratt, P., Caseldine, C. J., Dodson, J., Edwards, K. J., Farrell, M., Froyd, C., et al., 2013. The Holocene vegetation cover of Britain and Ireland: overcoming problems of scale and discerning patterns of openness. Quaternary. Sci. Rev. 73, 132–148.

Gaillard, M.-J., Dutoit, T., Hjelle, K., Koff, T., O'Connell, M., 2009. European cultural landscape — insights into origins and development. In: Krzywinski, K., Küster, H., O'Connell, M. (Eds.), Cultural Landscapes of Europe: Fields of Demeter, Haunts of Pan. Aschenbeck Media, pp. 35–44.

Gaillard, M.-J., Göransson, H., 1991. The bjäresjö area –vegetation and landscape through time. In: Berglund, B. E. (Ed.), The cultural landscape during 6000 years in southern Sweden: the Ystad project. Vol. 41 of Ecological Bulletins. Wiley-Blackwell, pp. 167–173.

Gaillard, M.-J., Sugita, S., Mazier, F., Trondman, A.-K., Brostrom, A., Hickler, T., Kaplan, J. O., Kjellström, E., Kokfelt, U., Kuneš, P., , Lemmen, C., Miller, P., Olofsson, J., Poska, A., Rundgren, M., Smith, B., Strandberg, G., Fyfe, R., Nielsen, A., Alenius, T., Balakauskas, L., Barnekow, L., Birks, H., Bjune, A., Björkman, L., Giesecke, T., Hjelle, K., Kalnina, L., Kangur, M., van der Knaap, W., Koff, T., Lagerås, P., Latałowa, M., Leydet, M., Lechterbeck, J., Lindbladh, M., Odgaard, B., Peglar, S., Segerström, U., von Stedingk, H., Seppä, H., 2010. Holocene land-cover reconstructions for studies on land cover-climate feedbacks. Clim. Past. 6, 483–499.

Garreta, V., Miller, P., Guiot, J., Hély, C., Brewer, S., Sykes, M., Litt, T., 2010. A method for climate and vegetation reconstruction through the inversion of a dynamic vegetation model. Clim. Dynam. 35 (2-3), 371–389.
URL http://dx.doi.org/10.1007/s00382-009-0629-1

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., Rubin, D. B., 2014. Bayesian data analysis. CRC press.

Gerten, D., Schaphoff, S., Haberlandt, U., Lucht, W., Sitch, S., 2004. Terrestrial vegetation and water balancehydrological evaluation of a dynamic global vegetation model. J. Hydrol 286 (1), 249–270.

Harrison, S., Jolly, D., Laarif, F., Abe-Ouchi, A., Dong, B., Herterich, K., Hewitt, C., Joussaume, S., Kutzbach, J., Mitchell, J., de Noblet, N., , Valdes, P., 1998. Intercomparison of simulated global vegetation distributions in response to 6 kyr BP orbital forcing. J. Climate 11 (11), 2721–2742.

Hastie, T., Tibshirani, R., Friedman, J., 2001. The Elements of Statistical Learning. Springer Series in Statistics. Springer New York Inc., New York, NY, USA.

Hellman, S., Gaillard, M.-J., Broström, A., Sugita, S., 2008a. The REVEALS model, a new tool to estimate past regional plant abundance from pollen data in large lakes: validation in southern Sweden. J. Quaternary. Sci. 23 (1), 21–42.

Hellman, S. E., Gaillard, M.-j., Broström, A., Sugita, S., 2008b. Effects of the sampling design and selection of parameter values on pollen-based quantitative reconstructions of regional vegetation: a case study in southern Sweden using the REVEALS model. Veg. Hist. Archaeobot. 17 (5), 445–459.

Hickler, T., Vohland, K., Feehan, J., Miller, P. A., Smith, B., Costa, L., Giesecke, T., Fronzek, S., Carter, T. R., Cramer, W., Kühn, I., Sykes, M. T., 2012. Projecting the future distribution of European potential natural vegetation zones with a generalized, tree species-based dynamic vegetation model. Global. Ecol. Biogeogr. 21 (1), 50–63.

Kaplan, J. O., Krumhardt, K. M., Gaillard, M.-J., Sugita, S., Trondman, A.-K., Mazier, F., 2014. The deforestation history of northwest Europe: Evaluating anthropogenic land cover change scenarios with pollen-based landscape reconstructions, in preparation.

Kaplan, J. O., Krumhardt, K. M., Zimmermann, N., 2009. The prehistoric and preindustrial deforestation of Europe. Quaternary. Sci. Rev. 28 (27), 3016–3034.

Kempeneers, P., Sedano, F., Pekkarinen, A., Seebach, L., Strobl, P., San-Miguel-Ayanz, J., Jul. 2012. Pan-european forest maps derived from optical satellite imagery. IEEE Earthzine 5.

Kimeldorf, G. S., Wahba, G., 1970. A correspondence between bayesian estimation on stochastic processes and smoothing by splines. Ann. Math. Statist. 41 (2), 495–502.

Klein Goldewijk, K., Beusen, A., Van Drecht, G., De Vos, M., 2011. The HYDE 3.1 spatially explicit database of human-induced global land-use change over the past 12,000 years. Global. Ecol. Biogeogr. 20 (1), 73–86.

Kohfeld, K., Harrison, S., 2000. How well can we simulate past climates? evaluating the models using global palaeoenvironmental datasets. Quaternary. Sci. Rev. 19 (1), 321–346.

Krzywinski, K., O'Connell, M., Küster, H., 2009. Cultural Landscapes of Europe: Fields of Demeter, Haunts of Pan. Aschenbeck Media.

Lehsten, D., Dullinger, S., Hülber, K., Schurgers, G., Cheddadi, R., Laborde, H., Lehsten, V., François, L., Dury, M., Sykes, M. T., 2014. Modelling the Holocene migrational dynamics of fagus sylvatica l. and picea abies (l.) h. karst. Global. Ecol. Biogeogr.

Lindgren, F., Rue, H., 2013. Bayesian spatial and spatio-temporal modelling with R-INLA, submitted to J. Stat. Softw.
URL http://www.math.ntnu.no/inla/r-inla.org/papers/jss/lindgren.pdf

Lindgren, F., Rue, H., Lindström, J., 2011. An explicit link between Gaussian fields and Gaussian Markov random fields: the stochastic partial differential equation approach. J. Roy. Statist. Soc. Ser. B 73 (4), 423–498.

Mazier, F., Gaillard, M.-J., Kuneš, P., Sugita, S., Trondman, A.-K., Broström, A., 2012. Testing the effect of site selection and parameter setting on REVEALS-model estimates of plant abundance using the Czech quaternary palynological database. Rev. Palaeobot. Palyno. 187 (1), 38–49.

Mitchell, T. D., Jones, P. D., 2005. An improved method of constructing a

database of monthly climate observations and associated high-resolution grids. Int. J. Climatol. 25 (6), 693–712.

Monsi, M., Saeki, T., 1953. The light factor in plant communities and its significance for dry matter production. Jpn. J. Bot. 14, 22–52.

Nychka, D. W., 2000. Spatial-process estimates as smoothers. In: Schimek, M. G. A. (Ed.), Smoothing and Regression: Approaches, Computation, and Application. Wiley, New York, USA, pp. 393–424.

Pachauri, R., Reisinger, A. (Eds.), 2007. Contribution of Working Groups I, II and III to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change. IPCC, Geneva, Switzerland.
URL http://www.ipcc.ch/publications_and_data/ar4/syr/en/contents.html

Paciorek, C. J., McLachlan, J. S., 2009. Mapping ancient forests: Bayesian inference for spatio-temporal trends in forest composition using the fossil pollen proxy record. J. Am. Statist. Assoc. 104 (486), 608–622.

Päivinen, R., Lehikoinen, M., Schuck, A., Häme, T., Väätäinen, S., Kennedy, P., Folving, S., 2001. Combining earth observation data and forest statistics. EuroForIns.

Pitman, A., de Noblet-Ducoudré, N., Cruz, F., Davin, E., Bonan, G., Brovkin, V., Claussen, M., Delire, C., Ganzeveld, L., Gayler, V., van den Hurk, B. J. J. M., Lawrence, P. J., van der Molen, M. K., Mller, C., Reick, C. H., Seneviratne, S. I., Strengers, B. J., Voldoire, A., 2009. Uncertainties in climate responses to past land cover change: First results from the LUCID intercomparison study. Geophys. Res. Lett. 36 (14).

Pongratz, J., Reick, C., Raddatz, T., Claussen, M., 2009. Effects of anthropogenic land cover change on the carbon cycle of the last millennium. Global. Biogeochem. Cy. 23 (4), GB4001.

Pongratz, J., Reick, C., Raddatz, T., Claussen, M., 2010. Biogeophysical versus biogeochemical climate response to historical anthropogenic land cover change. Geophys. Res. Lett. 37 (8), L08702.

Poska, A., Saarse, L., Veski, S., 2004. Reflections of pre- and early-agrarian human impact in the pollen diagrams of Estonia. Palaeogeogr. Palaeocl. 209 (14), 37 – 50.
URL http://www.sciencedirect.com/science/article/pii/S0031018204001142

Poska, A., Sepp, E., Veski, S., Koppel, K., 2008. Using quantitative pollen-based land-cover estimations and a spatial CA_Markov model to reconstruct the development of cultural landscape at Rouge, South Estonia. Veg. Hist. Archaeobot. 17 (5), 527–541.

Poska, A., Smith, B., Lehsten, D., Marquer, L., Sugita, S., Gaillard, M.-J., 2012. Pollen-inferred quantitative reconstructions of past plant abundance for evaluation and development of dynamic vegetation models. In: IPC/IOPC 2012; SS07-O16 (411).

R Core Team, 2014. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.
URL http://www.R-project.org

Ruddiman, W. F., 2005. How did humans first alter global climate? SciAm March 2005, 34–41.

Rue, H., Held, L., 2004. Gaussian Markov random fields: theory and applications. CRC Press.

Rue, H., Martino, S., Chopin, N., 2009. Approximate bayesian inference for latent gaussian models by using integrated nested Laplace approximations. J. Roy. Statist. Soc. Ser. B 71 (2), 319–392.

Samuelsson, P., Jones, C. G., Willén, U., Ullerstig, A., Gollvik, S., Hansson, U., Jansson, C., Kjellström, E., Nikulin, G., Wyser, K., 2011. The rossby centre regional climate model RCA3: model description and performance. Tellus. A 63 (1), 4–23.

Schuck, A., van Brusselen, J., Päivinen, R., Häme, T., Kennedy, P., Folving, S., 2002. Compilation of a calibrated European forest map derived from NOAA-AVHRR data. EFI Internal Report 13, EuroForIns.
URL http://www.efi.int/portal/virtual_library/publications/technical_reports/13/

Sitch, S., Smith, B., Prentice, I. C., Arneth, A., Bondeau, A., Cramer, W., Kaplan, J., Levis, S., Lucht, W., Sykes, M., Thonicke, K., Venevsky, S., 2003. Evaluation of ecosystem dynamics, plant geography and terrestrial carbon cycling in the LPJ dynamic global vegetation model. Glob. Change Biol. 9 (2), 161–185.

Smith, B., Prentice, I. C., Sykes, M. T., 2001. Representation of vegetation dynamics in the modelling of terrestrial ecosystems: Comparing two contrasting approaches within European climate space. Global. Ecol. Biogeogr. 10 (6), 621–637.

Soepboer, W., Sugita, S., Lotter, A. F., 2010. Regional vegetation-cover changes on the Swiss Plateau during the past two millennia: A pollen-based reconstruction using the REVEALS model. Quaternary. Sci. Rev. 29 (34), 472 – 483.
URL http://www.sciencedirect.com/science/article/pii/S0277379109003308

Stocker, T. F., Qin, D., Plattner, G.-K., Tignor, M. M., Allen, S. K., Boschung, J., Nauels, A., Xia, Y., Bex, V., Midgley, P. M. (Eds.), 2013. IPCC, 2013: Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.
URL https://www.ipcc.ch/report/ar5/wg1/

Strandberg, G., Brandefelt, J., Kjellström, E., Smith, B., 2011. High-resolution regional simulation of last glacial maximum climate in Europe. Tellus. A 63 (1), 107–125.

Strandberg, G., Kjellström, E., Poska, A., Wagner, S., Gaillard, M.-J., Trondman, A.-K., Mauri, A., Davis, B. A. S., Kaplan, J. O., Birks, H. J. B., Bjune, A. E., Fyfe, R., Giesecke, T., Kalnina, L., Kangur, M., van der Knaap, W. O., Kokfelt, U., Kuneš, P., Latał owa, M., Marquer, L., Mazier, F., Nielsen, A. B., Smith, B., Seppä, H., Sugita, S., 2014. Regional climate model simulations for Europe at 6 and 0.2 k bp: sensitivity to changes in anthropogenic deforestation. Clim. Past. 10 (2), 661–680.
URL http://www.clim-past.net/10/661/2014/

Sugita, S., 2007. Theory of quantitative reconstruction of vegetation I: pollen from large sites REVEALS regional vegetation composition. The Holocene 17 (2), 229–241.

Sugita, S., Parshall, T., Calcote, R., Walker, K., 2010. Testing the landscape reconstruction algorithm for spatially explicit reconstruction of vegetation in northern Michigan and Wisconsin. Quaternary. Res. 74 (2), 289–300.

Tjelmeland, H., Lund, K. V., 2003. Bayesian modelling of spatial compositional data. J. Appl. Stat. 30 (1), 87–100.

Trondman, A.-K., Gaillard, M.-J., Sugita, S., Fyfe, R., Kaplan, J., Nielsen, A.-B., Marquer, L., Mazier, F., Poska, A., Strandberg, G., 2012. Land cover-climate interactions in NW Europe, 6000 BP and 200 BP – first results of the Swedish LANDCLIM project. In: IPC/IOPC 2012; SS07-O14 (530).

Trondman, A.-K., Gaillard, M.-J., Sugita, S., Mazier, F., Fyfe, R., Lechterbeck, J., Marquer, L., Nielsen, A., Twiddle, C., Barratt, P., Birks, H., Bjune, A., Caseldine, C., David, R., Dodson, J., Dörfler, W., Fischer, E., Giesecke, T., Hultberg, T., Kangur, M., Kuneš, P., Latałowa, M., Leydet, M., Lindbaldh, M., Mitchell, F., Odgaard, B., Peglar, S., Persson, T., Rösch, M., van der Knaap, P., van Geel, B., Smith, A., Wick, L., 2014. Pollen-based quantitative reconstructions of past land-cover in NW Europe between 6k years BP and present for climate modelling, submitted to Global Change Bio.

Wahba, G., 1981. Spline interpolation and smoothing on the sphere. SIAM J. Sci. Stat. Comp. 2 (1), 5–16.

Wårlind, D., 2013. The role of carbon-nitrogen interactions for terrestrial ecosystem dynamics under global change-a modelling perspective. Ph.D. thesis, Lund University.

Whittle, P., 1954. On stationary processes in the plane. Biometrika 41, 434–449.

Whittle, P., 1963. Stochastic processes in several dimensions. B. Int. Statist. Inst. 40 (2), 975–994.

Appendix A. Calibration and reconstruction

Parameter calibration for the models is either accomplished through standard linear regression (the RM and RM_{geo} models), or by maximising the resulting Gaussian likelihood using the R-INLA package (Rue et al., 2009; Lindgren and Rue, 2013) (the IGMRF and $IGMRF_{geo}$ models). Given calibrated parameters the transformed compositions u at unobserved locations are reconstructed and back-transformed Eq. (3) to obtain compositional values at all locations. Both parameter calibration and reconstruction uses the same calibration set.

For the linear regression cases the reconstruction of the alr transformed compositions at an unobserved location, s_0, is obtain as

$$\hat{u}_i(s_0) = \hat{\mu}_i(s_0) = \hat{\beta}_{0,i} + \sum_p B_p(s_0)\hat{\beta}_{p,i}, \qquad (A.1)$$

where $\hat{\beta}_{p,i}$ are standard linear regression estimates (i.e. parameter calibration). For the IGMRF models the reconstruction at all locations is given by

$$\begin{bmatrix} \hat{u}_1 \\ \hat{u}_2 \end{bmatrix} = \begin{bmatrix} \hat{\mu}_1 \\ \hat{\mu}_2 \end{bmatrix} + \left(\Sigma_z^{-1} + \sigma_\epsilon^{-2} A^\top A \right)^{-1} \sigma_\epsilon^{-2} A^\top \left(\begin{bmatrix} u_1 \\ u_2 \end{bmatrix} - A \begin{bmatrix} \hat{\mu}_1 \\ \hat{\mu}_2 \end{bmatrix} \right),$$
$$(A.2)$$

where $\hat{\mu}_1$ is the reconstruction due to the mean field in Eq. (6) and the second term adjusts nearby locations for deviations between observations and mean-model (recall that the A-matrix extracts the observed locations).

Appendix B. LPJ-GUESS

The vegetation is simulated as plant functional types (PFTs) discriminated in terms of bioclimatic limits, growth form, phenology, life-history strategy, and various aspects of physiology. The bioclimatic niche parameterization is based on current vegetation distribution (Hickler et al., 2012). The model was run in cohort mode, in which all individuals belonging to the same age class of a PFT within a patch (local neighbourhood of individuals) are assumed to be identical in size, form, and response to the microenvironment. Multiple patches are simulated to encompass variability across the landscape of a grid cell in stand history, depending on disturbances, which recur stochastically with an expected local return time of 100 years, and stand demography. Competition for resources (light, water, etc.) among individuals is defined by the prescribed characteristics of the PFTs in combination with the emergent vegetation structure of a patch and its effect on the microenvironment and resource availability experienced by plants. A full description of LPJ-GUESS is provided in Smith et al. (2001) and references therein. Plant physiological and ecosystem biogeochemical processes are modelled as in LPJ-DGVM (Sitch et al., 2003). The current version includes the updates described in Gerten et al. (2004) and Hickler et al. (2012).

The simulated PFT-specific leaf-area index (LAI (PTF)) output was averaged over the modelled period and converted to fractional plant cover (FPC (PTF)). The LAI (PTF) to FPC (PTF) conversion was performed by applying the Lambert-Beer law (Monsi and Saeki, 1953) to the area of ground covered by foliage directly above it (Sitch et al., 2003):

$$FPC(PFT) = 1 - \exp\left(- k \cdot LAI(PFT) \right)$$