Ecological Complexity 20 (2014) 127–141

Contents lists available at ScienceDirect

Ecological Complexity

journal homepage: www.elsevier.com/locate/ecocom

Original Research Article

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*

ARTICLE INFO

Article history:
Received 31 March 2014
Received in revised form 31 August 2014
Accepted 21 September 2014
Available online 10 October 2014

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

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, 0.2, 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.

* Corresponding author at: Centre for Mathematical Sciences Mathematical Statistics, Lund University, Box 118, SE-221 00 Lund, Sweden. Tel.: +46 765836095.
E-mail addresses: behnaz@maths.lth.se, behnaz@pirzamanbin.name (B. Pirzamanbein).

http://dx.doi.org/10.1016/j.ecocom.2014.09.005
1476-945X/

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 macro-remains 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, 0.2 and 0.05 ka (calibrated year BP) using the methods developed in this study. The work is part of the LANDCLIM 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 REVEALS 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 reconstruction 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 $1^\circ \times 1^\circ$ 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$_{\mathrm{KK10}}$, 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.

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

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);

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

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:

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

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