

# A method for climate and vegetation reconstruction through the inversion of a dynamic vegetation model

Vincent Garreta · Paul A. Miller · Joël Guiot ·  
Christelle Hély · Simon Brewer · Martin T. Sykes ·  
Thomas Litt

Received: 3 December 2008 / Accepted: 7 July 2009 / Published online: 26 July 2009  
© The Author(s) 2009. This article is published with open access at Springerlink.com

**Abstract** Climate reconstructions from data sensitive to past climates provide estimates of what these climates were like. Comparing these reconstructions with simulations from climate models allows to validate the models used for future climate prediction. It has been shown that for fossil pollen data, gaining estimates by inverting a vegetation model allows inclusion of past changes in carbon dioxide values. As a new generation of dynamic vegetation model is available we have developed an inversion method for one model, LPJ-GUESS. When this novel method is used with high-resolution sediment it allows us to bypass the classic assumptions of (1) climate and pollen independence between samples and (2) equilibrium between the vegetation, represented as pollen, and climate. Our dynamic inversion method is based on a statistical model to describe the links among climate, simulated vegetation and pollen samples. The inversion is realised thanks to a particle filter algorithm. We perform a validation on 30 modern European sites and then apply the method to the sediment core

of Meerfelder Maar (Germany), which covers the Holocene at a temporal resolution of approximately one sample per 30 years. We demonstrate that reconstructed temperatures are constrained. The reconstructed precipitation is less well constrained, due to the dimension considered (one precipitation by season), and the low sensitivity of LPJ-GUESS to precipitation changes.

**Keywords** Palaeoclimate reconstruction · Model inversion · Particle filter · Dynamic vegetation model · Pollen sample · LPJ-GUESS

## 1 Introduction

Numerous studies have produced statistical palaeoclimate estimates by using the modern relationship between pollen and climatic data (e.g. the pioneering works of Webb and Bryson (1972) and Prentice et al. (1991) or a recent review in Guiot and De Vernal (2007)). These studies have substantially improved our knowledge of past climates and have been used as benchmark to evaluate robustness of climate models (e.g. from COHMAP Members (1988) to Jost et al. (2005)).

The existing reconstruction methods are based on the assumption that plant-climate interactions remain the same through time, and implicitly assume that these interactions are independent of forcings such as changes in atmospheric CO<sub>2</sub>. Guiot et al. (2000) showed that this assumption could produce significant biases in the results and that by using a vegetation model inversion, it was possible to evaluate these biases and to correct them. Wu et al. (2007) applied the method to European, African and Asian data for two periods of the past when atmospheric CO<sub>2</sub> concentration was significantly different from the present one. They

---

V. Garreta (✉) · J. Guiot · C. Hély  
CEREGE, UMR 6635, CNRS, Université Aix-Marseille,  
Europôle de l'Arbois, 13545 Aix-en-Provence, France  
e-mail: garreta@cerege.fr

P. A. Miller · M. T. Sykes  
Department of Physical Geography and Ecosystems Analysis,  
Geobiosphere Science Centre, Lund University, Sölvegatan 12,  
223 62 Lund, Sweden

S. Brewer  
Institut d'Astrophysique et de Géophysique, Université de Liège,  
Bat. B5c, 17 Allée du Six Août, Liège, Belgium

T. Litt  
Paleontological Institute, University of Bonn,  
Nussallee 8, 53115 Bonn, Germany

showed that biases could reach up to 10°C for winter temperature in Europe during the Last Glacial Maximum. These papers used an equilibrium vegetation model (BIOME4) which accounts for processes related to carbon and water cycles, but not for those related to plant competition and mortality. A more recent and sophisticated dynamic model, LPJ-GUESS (Smith et al. 2001), takes these processes into account.

We propose a method for the inversion of a dynamic vegetation model and argue that in palaeoclimatology this method is an improvement compared to the inversion of static models. Indeed, changing the static vegetation models used previously for a up-to-date dynamic model updates the transfer function defined by inversion. Second, when the dynamic inversion (read inversion of a dynamic model) is applied to a high time-resolution sediment core it provides a way to bypass the classic assumptions of:

- *Independence between samples.* With the exception of the Haslett et al. (2006) method, classic reconstruction methods ignore temporal correlations even if pollen data are sampled in sediment cores, which provide temporal records of the pollen. The dynamic vegetation model simulates vegetation histories that can be used as a natural link for temporal reconstruction.
- *Equilibrium between climate and vegetation.* Classic transfer functions are calibrated with modern pollen and climate data which are necessarily spatial. The absence of any temporal information means that we cannot calibrate a dynamic link or disequilibrium. Under a changing climate (modern or past) this is a simplification because the vegetation response may be delayed. Using LPJ-GUESS to simulate vegetation dynamics allows us to include a delay between climate change and vegetation change, by taking into account growth, mortality and competition processes.

Both of these assumptions are admissible when working with a low time resolution because the expected reconstruction is of low resolution and samples are nearly independent when there is a long time interval between samples. When a high resolution core is used, the expectation is a high quality reconstruction, i.e. including and properly quantifying all possible sources of uncertainty. In this case, the noise associated with the independent reconstruction of samples should be reduced or properly quantified by modelling a link between samples. The equilibrium hypothesis must also be considered, as this potentially induces error in the timing of climate changes.

The inversion of the dynamic model LPJ-GUESS, compared to static model inversion, is complicated in two ways.

First, LPJ-GUESS is stochastic, which means that any two simulations realised with the same forcings (climate, CO<sub>2</sub>, soil, etc) are not exactly identical. This is due to fire,

establishment and mortality processes which are represented in LPJ-GUESS as stochastic. The vegetation must therefore be considered as a random “hidden” variable instead of a deterministic function of climate as in Guiot et al. (2000).

Second, we want to use the temporal aspect provided by the vegetation model. In the reconstruction algorithm, this would require a high-dimensional climate space to be tested and induces the classic problem of the “curse of dimensionality”. For a static model, we can run the model for a single point in time. Here several possible climates are proposed, the model is run with each scenario and the simulated vegetations are compared to a single pollen sample to retain the more coherent simulations (Guiot et al. 2000). With the same method for temporal inversion, we need to propose several high-dimensional climate histories, simulate several vegetation histories and compare them to the pollen history. This algorithm is inefficient and requires massive simulations because almost no vegetation chronology will fit the entire pollen chronology. With this kind of algorithm (a global stochastic search in the entire time-climate space) computing time for simulation is prohibitive, at least due to the use of a vegetation model.

To overcome both challenges and perform the temporal inversion, we have developed, and present here, a hierarchical Bayesian model and a particle filter algorithm (Doucet et al. 2001) for inference. The hierarchical Bayesian approach facilitates the probabilistic formalisation of the inversion process. It has generally attractive features in paleoclimatology (see the discussion in Haslett et al. 2006) and we mainly use its concept of organisation of the variables into a hierarchy, prior and posterior described in the modelling section. The particle filter algorithm is mainly a Bayesian tool used for inference in real-time (or on-line) problems. In our context it allows to bypass the curse of dimensionality because it considers the reconstruction date after date.

The paper is structured as follows: (1) The vegetation model, climate data and pollen data are presented. (2) We then describe the statistical model and inference algorithm. (3) The method is validated using 30 modern pollen samples distributed across Europe. (4) The temporal feature of the approach is fully exploited by reconstructing Holocene climate from the high resolution Meerfelder Maar sediment core (data from Litt et al. 2009).

## 2 Materials and methods

### 2.1 The LPJ-GUESS dynamic global vegetation model

LPJ-GUESS simulates the dynamics of vegetation stands, accounting for competition between tree individuals and

populations as a forest gap model (Shugart 1984; Bugmann 2001). A full description of the model can be found in Smith et al. (2001). Biophysical and physiological processes are represented mechanistically, and are based on the same formulations as the well-evaluated Lund-Potsdam-Jena dynamic global vegetation model (LPJ-DGVM; Sitch et al. 2003). Updates to the model's hydrological processes were described by Gerten et al. (2004).

In LPJ-GUESS, cohorts of trees of different species, age and structure compete for light and soil resources on a number of replicate patches (15 patches of 1,000 m<sup>2</sup> in the present study). Either plant functional types (PFTs) (Sitch et al. 2003) or species (Hickler et al. 2004; Koca et al. 2006) may be simulated.

Typical model output consists of leaf area index (LAI), net primary production (NPP), biomass, tree density, carbon fluxes and runoff. Values are averaged over the replicate patches to give stand averages of the relevant variables.

Using a very similar model set-up to that used here, Miller et al. (2008) showed that LPJ-GUESS could successfully model the Holocene dynamics of the main tree species at four sites in Fennoscandia where vegetation reconstructions using pollen accumulation rate data were possible.

### 2.1.1 Species description

In Table 1, we list the seventeen tree and shrub species, and the single grass taxon, used in the model, as well as their plant characteristics and bioclimatic limits. Further changes to the model parameters described by Smith et al. (2001), Hickler et al. (2004) and Miller et al. (2008), are listed in Tables 4 and 5, Appendix 1.

The species are trees (T) or shrubs (S), boreal (Bo) or temperate (Te), broadleaf summergreen (BS), broadleaf evergreen (BE) or needleleaf evergreen (NE), and shade tolerant (St), intermediately shade tolerant (ISt) or shade intolerant (Si) (Smith et al. 2001). Trees and shrubs have different allometric relationships, and summergreen species require varying periods of chilling to induce budburst (Murray et al. 1989). The generic C3 grass PFT is intended to represent the numerous understorey species that are not considered in this paper, but nevertheless compete with trees for water and nutrients.

The maximum range limits of the tree species are defined in LPJ-GUESS by four key, species-specific bioclimatic constraints (Prentice et al. 1992; Sykes et al. 1996):  $GDD_{5\text{ min}}$  (minimum growing degree-day sum (5°C base)),  $T_{c\text{ min}}$  (minimum temperature of the coldest month),  $T_{c\text{ max}}$  (maximum temperature of the coldest month) and drought tolerance (DT). Drought intolerant species (DT = 0) require an average growing season available water

content of 30 mm for establishment. The values in Table 1 were taken from the literature (Prentice and Helmsaari 1991; Sykes et al. 1996), with minor adjustments prompted by comparison with European species distributions. The use of this minimal set of bioclimatic constraints, each of which represents a known or likely physiological limiting mechanism (Woodward 1987; Miller et al. 2008), is more robust through time than simple correlations between climatic variables and species ranges.

For a species within its bioclimatic limits, cohort establishment and mortality are modelled yearly in LPJ-GUESS as stochastic processes within each replicate patch of the stand (Smith et al. 2001). Two additional stochastic processes are also considered in LPJ-GUESS. First, patch-destroying disturbances, representing destructive processes such as herbivory and storm damage, result in all vegetation in a patch being transferred to the patch's litter pool with a certain annual probability that is the inverse of the average disturbance interval of 100 years. Second, the yearly probability of a fire disturbance is modelled as in Thonicke et al. (2001).

The species listed in Table 1 are clearly a small subset of the full range of species seen in Holocene pollen diagrams. However, use of a restricted set was a necessary compromise. A larger species set would have increased the computational time required for model inversion. By choosing a restricted set containing a representative sample of the diversity of vegetation and functional types seen in sub-Arctic Europe today, we expect to capture the main variability seen in the Holocene pollen records. Our choice was also restricted by the relatively small set of species with bioclimatic limits used by LPJ-GUESS that are known with any great degree of certainty.

### 2.1.2 Vegetation

From the different vegetation outputs of LPJ-GUESS: NPP, LAI, biomass, we summarise vegetation by using an average of simulated NPP over 30 years. This choice is driven by the need for maximum coherence between pollen samples and the vegetation simulated at the same sites. Preliminary attempts to link pollen and these outputs convinced us that LAI and NPP are nearly equivalent and perform better than biomass which represents an accumulation of carbon mass in time. Thirty-year means for NPP correspond approximately to the accumulation period for pollen in modern samples.

We denote the simulated vegetation at the  $N$  modern sites as  $V_{s=1:N}$ , represented by the NPP averaged by species over 30 years.  $V_t$  is the mean of simulated vegetation, for the past, during the years  $t - 30$  to  $t$ . Note that all these elements are positive or null and that they represent absolute or "raw" production values.

**Table 1** Selected species with their characteristics and bioclimatic limits as specified in the model

Species	Description	GDD <sub>5 min</sub> (°C d)	$T_{c_{\min}}$ (°C)	$T_{c_{\max}}$ (°C)	DT
<i>Abies alba</i>	T, Te, NE, St	1,800	-4.5	-1	0
<i>Alnus incana</i>	T, Bo, BS, Ist	500	-30	-2.5	0
<i>Betula pendula</i>	T, Te, BS, Si	700	-30	7	0
<i>Betula pubescens</i>	T, Bo, BS, Si	300	-	6	0
<i>Carpinus betula</i>	T, Te, BS, Ist	1,100	-8	5	1
<i>Corylus avellana</i>	T, Te, BS, Ist	700	-13	10	1
<i>Fagus sylvatica</i>	T, Te, BS, St	1,300	-3.5	6	0
<i>Fraxinus excelsior</i>	T, Te, BS, Ist	1,100	-10	6	0
<i>Picea abies</i>	T, Bo, NE, St	650	-30	-1.5	0
<i>Pinus sylvestris</i>	T, Bo, NE, Ist	450	-30	-1.0	1
<i>Pinus halepensis</i>	T, Te, NE, Ist	3,000	3	9	1
<i>Populus tremula</i>	T, Te, BS, Si	500	-30	6	0
<i>Quercus coccifera</i>	S, Te, BE, Ist	3,100	3.5	11	1
<i>Quercus ilex</i>	T, Te, BE, Ist	2,000	0	10	1
<i>Quercus robur</i>	T, Te, BS, Ist	1,100	-9	7	1
<i>Tilia cordata</i>	T, Te, BS, Ist	1,100	-11	5	0
<i>Ulmus glabra</i>	T, Te, BS, Ist	850	-9.5	6	0
C3 grass	-, -, -, -	0	-	-	1

The plant characteristics are: either trees (T) or shrubs (S), either boreal (Bo) or temperate (Te), either broadleaf summergreen (BS), broadleaf evergreen (BE) or needleleaf evergreen (NE), and either shade tolerant (St), intermediately shade tolerant (Ist) or shade intolerant (Si). See Table 5. The bioclimatic limits are: GDD<sub>5 min</sub>, minimum growing degree-day sum (5°C base);  $T_{c_{\min}}$ , minimum temperature of the coldest month;  $T_{c_{\max}}$ , maximum temperature of the coldest month; and DT, drought tolerance

## 2.2 Climate data

LPJ-GUESS is forced with chronologies of monthly precipitation, temperature and cloudiness. For each pollen site, we interpolated precipitation and temperature time series from the CRU TS 1.2 dataset (New et al. 2002). We used an ordinary kriging method with altitudinal gradient as an external drift (e.g. Cressie 1991). For cloudiness we fitted a logit-linear regression between monthly cloudiness and both monthly precipitation and temperature per site.

The interpolated climate series are considered as a skeleton to which anomalies are applied to determine the optimal fit between model outputs and pollen data. These anomalies will be referred to as parameters or climate parameters in the following sections because they are the climate quantities which are reconstructed. We denote by  $C_s$  for modern sites and  $C_t$  for core at time  $t$ , the 6-dimensional climate parameter vector:  $C = (T_{\text{jan}}, T_{\text{jul}}, P_{\text{win}}, P_{\text{spr}}, P_{\text{sum}}, P_{\text{aut}})$ . The first parameters are absolute temperature anomalies (in °C) from January and July 20th century series. The precipitation parameters are relative anomalies (in %). Let  $T_{(i,j)}$  be the original temperature of year  $i$  and month  $j$ ,  $T_{(.,j)}$  the 100-year mean temperature of month  $j$ , where  $j = 1$  denotes January. Then the

transformed temperature  $\tilde{T}_{(i,j)}(C)$  is defined as a function of parameters  $T_{\text{jan}}$  and  $T_{\text{jul}}$ :

$$\tilde{T}_{(i,j)}(C) = T_{(i,j)} - T_{(.,j)} + (T_{(.,j)} - T_{(.,1)}) * \left( \frac{T_{\text{jul}} + T_{(.,7)} - (T_{\text{jan}} + T_{(.,1)})}{T_{(.,7)} - T_{(.,1)}} \right) + T_{\text{jan}} + T_{(.,1)}$$

This transformation modifies the monthly mean temperature signal ( $T_{(.,j)}$ ) by scaling it to match new January and July specified temperatures  $T_{\text{jan}} + T_{(.,1)}$  and  $T_{\text{jul}} + T_{(.,7)}$ . The interannual variability of the transformed series is exactly the same as in the original skeleton.

Precipitation parameters are seasonal percentages which are added to the original skeleton. Let  $P_{(i,j)}$  be the original precipitation of year  $i$  and month  $j$ . Then the  $P_{\text{win}}$  anomaly is applied by multiplying each winter month (January, February and March) by  $(1 + P_{\text{win}}/100)$  to obtain  $\tilde{P}(C)$  modified precipitation

$$\tilde{P}_{(i,j=1:3)}(C) = P_{(i,j=1:3)} * (1 + P_{\text{win}}/100)$$

Here the positivity constraint of precipitation is respected, but the interannual variability of the original chronologies is modified.

Once the modified  $\tilde{T}(C)$  and  $\tilde{P}(C)$  have been created, a modified sunshine  $\tilde{S}$  is computed by regression using  $\tilde{T}(C)$  and  $\tilde{P}(C)$  as regressor variables.

## 2.3 Pollen data

### 2.3.1 Pollen surface samples

The pollen surface sample database has been compiled by Bordon (2008) from data taken from Bottema (1974), Brugiapaglia (1996), Peyron et al. (1998) and Sanchez Goni and Hannon (1999).

The database was initially composed of 1,512 different modern sites covering Europe and Morocco with more than 150 different pollen taxa. A subselection of taxa was made to correspond to the output of the model. We computed 14 groups by summing taxa corresponding to each of the following 14 arboreal taxa: Abies, Alnus, Betula, Carpinus, Corylus, Fagus, Fraxinus, Picea, Pinus, Quercus Evergreen, Quercus Deciduous, Tilia, Ulmus and Populus. A “grasses and shrubs” (GrSh) group was made by summing all non-arboreal and non-aquatic taxa. This selection preserves a maximum of coherence with the 18 species (or groups of species) defined in the version of LPJ-GUESS that we use. See Table 2 for the correspondence between pollen groups and vegetation model species.

We first filtered the sites by removing all non-terrestrial sample sites due to spurious coordinates or offshore core tops (e.g. in Danube estuaries). Offshore pollen samples are representative of pollen production of a whole watershed and are not coherent with other more local records. As a second selection criterion we removed all samples where the taxa subselection resulted in a loss of more than 25% of the original pollen count. These cases occurred when more than 25% of the sample consisted of arboreal taxa other than those simulated by LPJ-GUESS. We consider that the removal of

such a large part of the original pollen spectra would result in a distorted image of the surrounding vegetation.

Since most of the pollen samples were available as percentages (more than 70%), we converted counted samples to percentages.

The final dataset is a matrix containing  $N = 1,209$  modern sites (rows) and 15 taxa per site (columns). The spatial distribution of this dataset is shown in Fig. 1. A modern pollen sample  $Y_s = (Y_{s,1}, Y_{s,2}, \dots, Y_{s,15})$  is a 15-dimensional vector representing the pollen proportion per group, where  $\sum_{j=1}^{15} Y_{s,j} = 1$ .

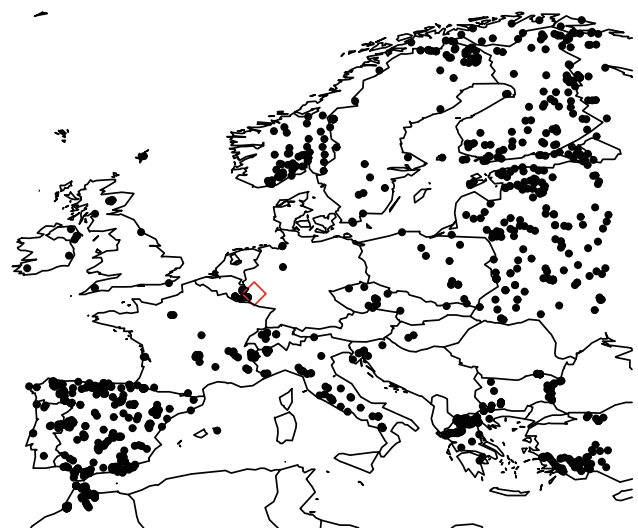
### 2.3.2 Meerfelder maar sediment core

The sediment core (Litt et al. 2009) was taken from the lake Meerfelder Maar (50.1°N, 6.75°E, see Figure 1) located within the Westeifel Volcanic Field. The uppermost 180 cm of the core, corresponding to approximately the last 1.6 cal ky BP (calendar kilo-years Before Present; which refers to the number of years before 1950), are not continuously varved and were dated using two AMS  $^{14}\text{C}$  dates and extrapolated sedimentation rates based on varve data. The other part of the core is varved and the endpoint of the core has been linked to a calendar-year chronology by using a tephra dated at 11 cal ky BP. In total 406 samples have been collected and analysed from this core. The number of pollen grains counted in each sample is between 500 and 1,000 and we transform it to percentages to agree with modern data.

The pollen diagram is presented Fig. 2. For the interpretation of climate reconstruction results we divide the 11 to 0 cal ky BP chronology in four periods. A more

**Table 2** Correspondence table between pollen types and species defined in LPJ-GUESS

$i$	Pollen type: $y_i$	Vegetation species: $v_{j(i)}$
1	Abies	Abi_alb
2	Alnus	Aln_inc
3	Betula	Bet_pen + Bet_pub
4	Carpinus	Car_bet
5	Corylus	Cor_ave
6	Fagus	Fag_syl
7	Fraxinus	Fra_exc
8	Picea	Pic_abi
9	Pinus	Pin_syl + Pin_hal
10	QuercusE	Que_coc + Que_ile
11	QuercusD	Que_rob
12	Tilia	Til_cor
13	Ulmus	Ulm_gla
14	Populus	Pop_tre
15	GrSh	C3_gr



**Fig. 1** Distribution of the 1,209 modern pollen samples (black dots) and location of the Meerfelder Maar site (red diamond)



comprehensive discussion can be found in Litt et al. (2009). The earliest period (11 to 10 cal ky BP) corresponds to the end of the glacial period and is characterised by a rapid decrease of Grass-Shrub group, Pinus and Betula. This decline is matched by marked increases in Corylus. The 10 to 6.3 cal ky BP phase shows a Corylus decrease and the arrival of Ulmus, followed by Tilia, Fraxinus and Alnus and, finally, Fagus. The 6.3 to 3.7 phase is a very stable one with high levels of Alnus, but less Ulmus and Tilia than during the previous period. The last period 3.7 to 0 cal ky BP is characterized by a number of changes, due to increasing anthropogenic pressure and climate changes. There is a global but non-monotonic increase of the Grass-Shrub group and a non-monotonic decrease of Alnus, Ulmus and Tilia.

Sediment core samples are denoted  $y_{t=t_1:t_n}$ . The core contains  $n = 406$  samples from  $t_1 = 10,988$  to  $t_{406} = 0$  cal year BP. As for the modern samples, each sample  $Y_t$  is a 15-dimensional vector representing the pollen proportion per pollen group.

#### 2.4 A statistical model to link climate, vegetation and pollen

We build a statistical model which embeds the vegetation model and describes the relations between the variables climate  $C$ , vegetation  $V$  and pollen  $Y$ . This is designed for pollen samples taken from a single sediment core. Each date  $t = t_1:t_n$  is considered known without uncertainty. This statistical model is *hierarchical Bayesian*.

*Hierarchical* means that it is based on a conditional “split” of the model. For comparison, a transfer function (TF) models a direct link between climate and pollen using the conditional distribution of pollen given climate  $p(Y|C)$ . In this work, we model  $p(Y|C)$  hierarchically by specifying a distribution of the vegetation conditional on climate  $p(V|C)$  and a distribution of the pollen conditional on vegetation  $p(Y|V)$ .

*Bayesian* theory is a framework for inference (Young and Smith 2005). In the context of this applied work we use the main concepts of “prior” and “posterior”. The prior is the information, summarised under the form of a distribution, which is available prior to the data analysis. For example we will use a prior on climate at time  $t$ ,  $p(C_t)$ . This is the information on ancient (time  $t$ ) climate we have before running the inversion, and may be estimated by climate reconstructions already available before inversion. After the choice of a prior on climate  $p(C)$  and a hierarchical model  $p(V|C), p(Y|V)$ , the Bayesian inference consists in obtaining the posterior distribution of climate and vegetation given pollen  $p(C, V|Y)$ . The Bayes theorem gives the link between the prior, the structure and the posterior:  $p(C, V|Y) = p(C) \cdot p(V|C) \cdot p(Y|V) / p(Y)$ .

The structure of the hierarchical model is illustrated in the graphic Fig. 3. It is based on the basic elements  $p(C)$ ,  $p(V|C)$  and  $p(Y|V)$  described above, and each individual part is described in more detail below. The choice of prior distribution  $p(C)$  is discussed below in each reconstruction exercise. In the next section we define and calibrate the distribution of pollen given vegetation  $p(Y|V)$ . Section 4 completes the definition by describing  $p(V|C)$  as the vegetation model.

##### 2.4.1 Calibration of the pollen/vegetation distribution

A key element of the inversion is the relationship between simulated vegetation and pollen data  $p(Y|V)$ . In statistics this is called the pollen likelihood and can be compared to a transfer function between vegetation and pollen. We model it using non-parametric kernel smoothed surfaces. These surfaces are calibrated using the modern pollen dataset and modern simulations of the vegetation.

The distribution  $p(Y|V)$  models the relationships between 15 pollen proportions  $Y$  and the simulated NPP of 18 species,  $V$ . Thus its dimension is  $15 + 18 = 33$  and it contains information about  $15 \times 18 = 270$  variable crossings. We reduce this dimension as follows:

$$p(Y|V) \approx \prod_{i=1}^{15} p_i(Y_i|V) \quad (1)$$

$$\approx \prod_{i=1}^{15} p_i(Y_i|V_{j(i)}) \quad (2)$$

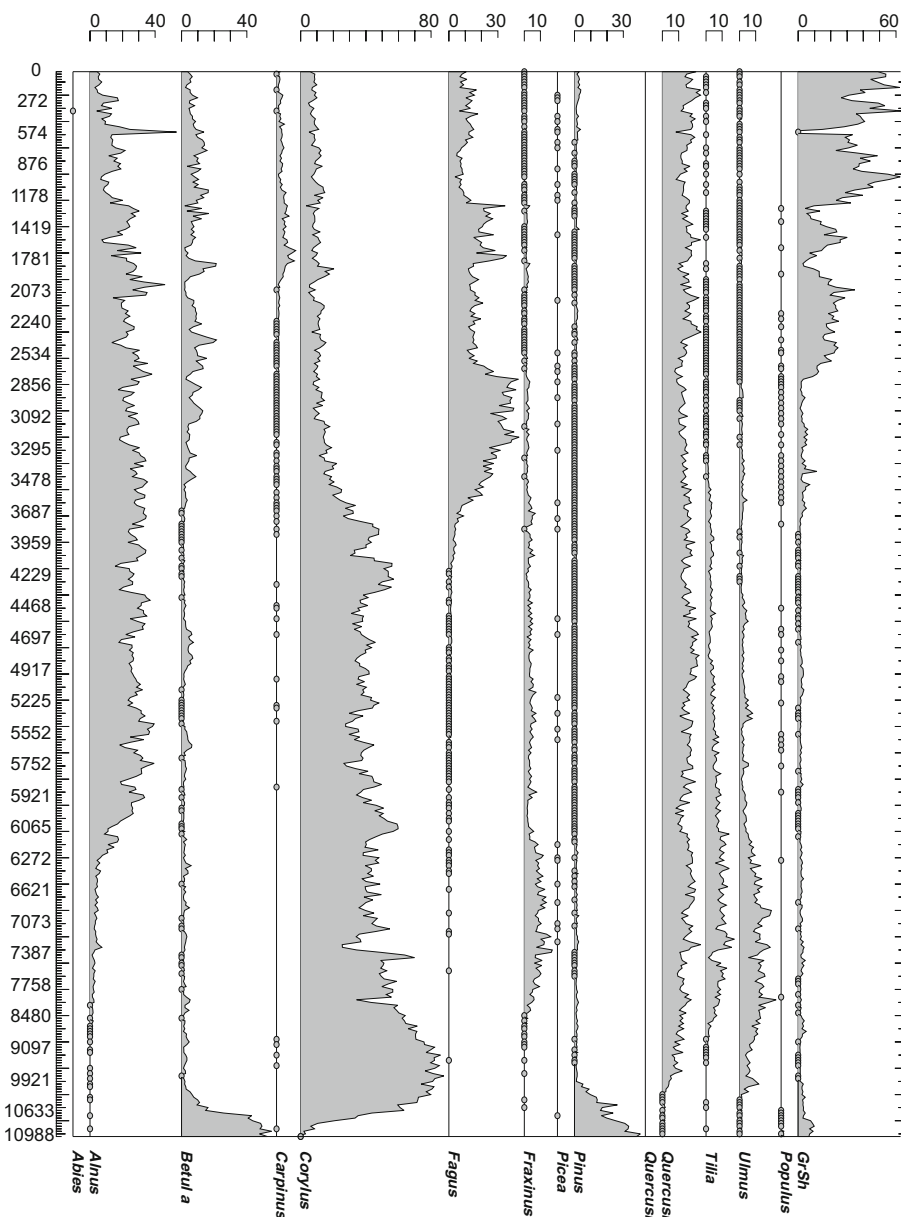
$$= \prod_{i=1}^{14} q_i(Z_i|V_{j(i)}) \quad (3)$$

$$\approx \prod_{i=1}^{14} \tilde{q}_i(Z_i|V_{j(i)}) \quad (4)$$

subject to the following assumptions:

1. Conditional on vegetation, all pollen abundances  $Y_i$ ,  $Y_j$  for  $i \neq j$  are independent. This is acceptable since, with a pollen time resolution of 20–30 years, given the vegetation, pollen production of one group can be considered independent of the production of all other pollen group.
2. All information about  $Y_i$  is carried by only one  $V_{j(i)}$  species of vegetation. The subscript  $j(i)$  refers to the  $j$ th vegetation species corresponding (i.e. as a function of) to the  $i$ th pollen group. This is acceptable if there is a good agreement between pollen groups and the species simulated by LPJ-GUESS. These correspondences are specified in Table 2.
3. Using the variable transformation  $Z = \text{alr}(Y)$  (Aitchison 1982) there always exists a relationship, called  $q_i$ ,

**Fig. 2** Pollen diagram from the Meerfelder Maar sediment core. (x-axis) The fifteen pollen groups defined in Table 2 are given as percentages (over the fifteen groups) and (y-axis) the age is in calendar years Before Present (BP, refers to before 1950) from 0 (top) to 10,988 (bottom). The total number of pollen grains counted per sample ranges between 500 and 1,000. The uppermost 180 cm of the core, corresponding to approximately the last 1.6 cal ky BP, are not continuously varved and were dated using two AMS <sup>14</sup>C dates and extrapolated sedimentation rates based on varve data. The other part of the core is varved and the endpoint of the core has been linked to a calendar-year chronology by using a tephra dated at 11 cal ky BP (Litt et al. 2009). Therefore the uncertainties associated to the chronology are not constant along the core and very hard to quantify. For the period of time between 11 cal ky BP and around 1.6 cal ky BP, varved sediments imply that there is no uncertainty between sample dates; but this is a floating chronology implying a constant uncertainty for the overall time-period. Uncertainties after 1.6 cal ky BP and for the whole floating chronology had a magnitude of around 100 year but were corrected by stratigraphic alignment with the well dated core of the Holzmaar Maar



between vegetation  $V$  and the transformed pollen variable  $Z$ . Pollen variables are probabilities implying that their sum is  $\sum_{i=1}^{15} Y_i = 1$ . This constraint reduces the dimension to 14 since the fifteenth proportion is defined by  $1 - \sum_{i=1}^{15} Y_i$ . First we set  $Y_{i,j} = 10^{-4}$  for all  $(i,j)$  where  $Y_{i,j} = 0$ . Then we apply the Aitchison (1982) transformation  $Z = alr(Y)$  defined as:

$$Z_{i=1:14} = \log(Y_i/Y_{15})$$

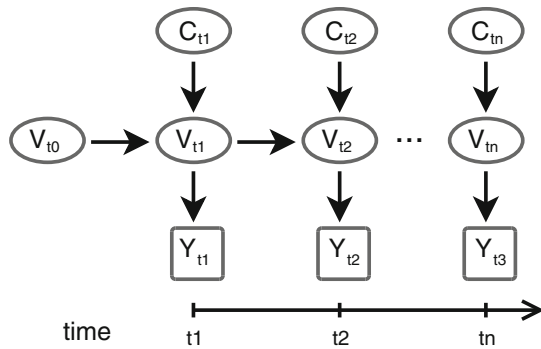
This reduces the dimension to 14 and lets us model the unconstrained variables  $Z_i$ .

- Each  $q_i$  is correctly approximated by a  $\tilde{q}_i$  obtained by kernel smoothing (e.g. Loader 1999).

In practice, the fourteen surfaces  $q_i$  are fitted on the modern dataset  $(C,V)_{s=1:N}$ . For this purpose we use a gaussian kernel whose parameters are fitted by cross validation (Loader 1999). Figure 4 shows the obtained surfaces.

#### 2.4.2 LPJ-GUESS is a vegetation/climate distribution

Let  $p(V_{t_j}, C_{t_j} | V_{t_i})$  be the distribution for the temporal transition of climate and vegetation from time  $t_i$  to time  $t_j$  where  $t_i$  and  $t_j$  are dates for consecutive samples of the core



**Fig. 3** Graphical representation of the hierarchical model for reconstruction. Considered variables are:  $C$  the climate,  $V$  the vegetation simulated by LPJ-GUESS and  $Y$  the pollen data. Subscripts indicate a sample date ( $t$ ) for the analysed core. Known variables are in a square and variables to be reconstructed in a circle. Arrows represent the conditioning between variables. The times  $t_1$  to  $t_n$  are the core point dates.  $t_1$  is the oldest age sampled in the core,  $t_n$  the most recent. Following the arrows we see that vegetation at time  $t_2$  ( $V_{t_2}$ ) depends on climate at time  $t_2$  ( $C_{t_2}$ ) and vegetation at time  $t_1$  ( $V_{t_1}$ ). This is modelled using LPJ-GUESS forced with climate  $C_{t_2}$ , starting from  $V_{t_1}$  and run during  $t_2 - t_1$  years. Pollen at time  $t_2$  ( $y_{t_2}$ ) depends only on  $V_{t_2}$  through a  $p(Y|V)$  distribution

$$p(V_{t_j}, C_{t_j} | V_{t_i}) = p_{\text{LPJ}}(V_{t_j} | V_{t_i}, C_{t_j}) \cdot p(C_{t_j}) \quad (5)$$

with  $p(C_{t_j})$  the prior climate distribution at  $t_j$ .  $p_{\text{LPJ}}(V_{t_j} | V_{t_i}, C_{t_j})$  is (defined by) the randomness of  $V_{t_j}$  when we run LPJ-GUESS for years starting from vegetation  $V_{t_i}$ , with climate  $C_{t_j}$ .

Vegetation simulated by LPJ-GUESS is stochastic in the sense that several runs of the vegetation model with the same forcings give slightly different vegetation values. We therefore consider vegetation as a random variable and LPJ-GUESS as a distribution:  $p_{\text{LPJ}}(V_{t_j} | V_{t_i}, C_{t_j})$ . This distribution output is the variable  $V_{t_j}$  and has parameters  $(V_{t_i}, C_{t_j})$  but also the climate chronologies and numerous forcings like soil,  $\text{CO}_2$  etc. We can sample from this distribution by running the model, but since it is a complex computer code we cannot compute its probability value for any given set of variables and parameters. This represents a major change from deterministic vegetation models such as BIOME3 (Haxeltine and Prentice 1996) or the LPJ-DGVM (Sitch et al. 2003).

The temporal link of the hierarchical model is given by LPJ-GUESS and arises from the later definition. To simulate vegetation at time  $t_j$  younger than (after)  $t_i$ , where  $t_i$  and  $t_j$  are the dates of consecutive core samples, the vegetation model starts with  $V_{t_i}$  and runs for  $t_j - t_i$  years. If  $t_j - t_i$  is short, the vegetation simulated at  $t_j$  is strongly forced by vegetation  $V_{t_i}$  and then, implicitly, by climate  $C_{t_i}$ . This constraint gives a time-coherence to the reconstructed vegetation.

## 2.5 Inference using a particle filter algorithm

In the Bayesian context, reconstruction of climate and vegetation involves the computation of the joint posterior distribution  $p(C_{t_1:t_n}, V_{t_1:t_n} | Y_{t_1:t_n})$ . This represents the distribution of climate and vegetation “histories” from time  $t_1$  to  $t_n$  knowing all  $Y_{t_1:t_n}$  pollen data.

Particle filters provide a reconstruction based on importance sampling (IS) which is sequential, i.e. done sample after sample. The sequential aspect solves the curse of dimensionality because it slices the climate space. A simple explanation follows: at time  $t_j$  the algorithm has a reconstruction obtained for the preceding point  $t_i$ . A set of 1,000 possible climates  $C_{t_j}^{(l)}$  is proposed from the prior  $p(C_{t_j})$ . LPJ-GUESS is then run with each of these climates starting from the reconstructed vegetation at  $t_i$  and for the years  $t_i$  to  $t_j$ . It produces couples of climate and associated vegetation  $(C_{t_j}, V_{t_j})^{(l)}$  that we call “particles”. In this set of particles, a selection is done by comparison of each vegetation simulated and the pollen  $Y_{t_j}$ . This selection consists in computing  $\omega^{(l)}$  equal to the likelihood  $p(Y|V)$  of the pollen  $Y_{t_j}$  for each simulated vegetation  $V_{t_j}^{(l)}$ . A high  $\omega^{(l)}$  score means that the couple  $(C_{t_j}, V_{t_j})^{(l)}$  is highly probable and a null score means that the couple is not coherent.

A full comprehensive description of the algorithm is given Appendix 2. We just give here a summary of the algorithm.

### 1. INITIALISATION

- Generate  $N_p$  couples  $(V_{t_1}, C_{t_1})^{(l=1:N_p)}$ , by sampling  $(C_{t_1}, V_{t_1})^{(l)}$  from  $p(C_{t_1})$  and running LPJ-GUESS until equilibrium is reached with climate  $C_{t_1}^{(l)}$  to obtain  $V_{t_1}^{(l)}$ ,
- Compute for each particle  $(V_{t_1}, C_{t_1})^{(l)}$ , the weight  $\omega_{t_1}^{(l)} = p(Y_{t_1} | V_{t_1}^{(l)})$ , using the kernel smoothed surfaces,
- Compute each normalised weights  $\tilde{\omega}_{t_1}^{(l)} = \omega_{t_1}^{(l)} / \sum_{k=1}^{N_p} \omega_{t_1}^{(k)}$

### 2. RESAMPLING

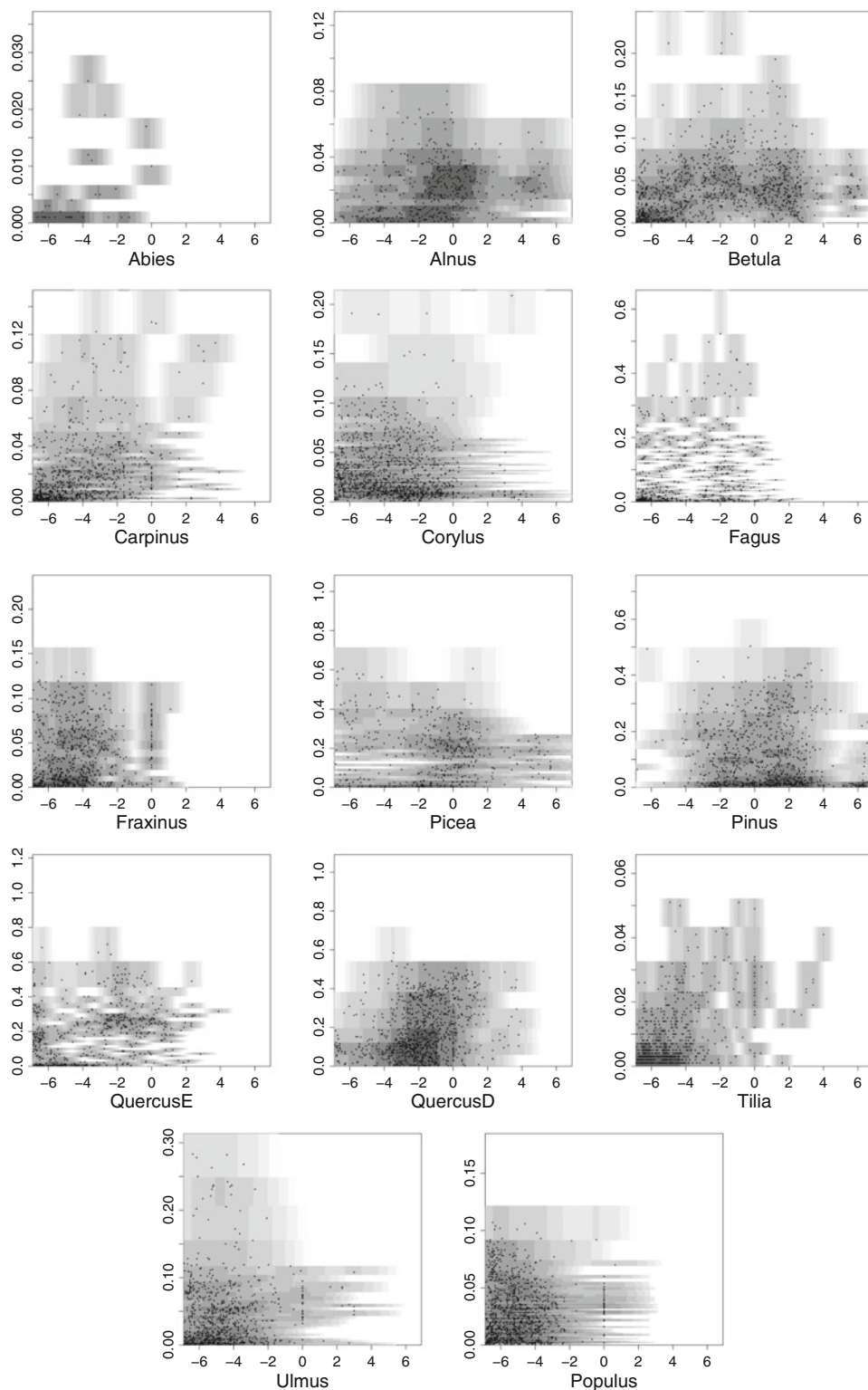
- Compute the criterion  $\text{ESS}_t = \left( \sum_{l=1}^{N_p} (\tilde{\omega}_t^{(l)})^2 \right)^{-1}$
- If  $\text{ESS}_t < N_p/2$  randomly sample the particles by residual resampling and set all weights  $\tilde{\omega}_t^{(l=1:N_p)} = 1/N_p$ .

### 3. SAMPLING

- For current time  $t_j$  immediately consecutive to  $t_i$ ,
- Sample  $N_p$  particles  $(V_{t_j}, C_{t_j})^{(l=1:N_p)}$  by sampling  $C_{t_j}^{(l)}$  from  $p(C_{t_j})$  and running LPJ-GUESS starting from  $V_{t_i}$  for  $t_j$  to  $t_i$  years with climate  $C_{t_j}$  to obtain  $V_{t_j}^{(l)}$ ,
- Compute for each particle  $V_{t_j}^{(l)}$  the weights  $\omega_{t_j}^{(l)} = \tilde{\omega}_{t_i}^{(l)} \cdot p(Y_{t_j} | V_{t_j}^{(l)})$  using the kernel smoothed surfaces,



**Fig. 4** Kernel smoothed surfaces for the 14 groups. The joint smoothings  $q_i(Z_i, V_{j(i)})$  are derived from the conditional smoothings used for inference:  $q_i(Z_i|V_{j(i)})$ . Graphics are in the same order as in Table 2. Each plot presents (*x*-axis) a pollen group transformed following Aitchison (1982) ( $Z_i$  without unit) versus (*y*-axis) its corresponding simulated annual net primary production (NPP, in  $\text{kg carbon m}^{-2} \text{ year}^{-1}$ ). The *dots* are the modern data and the shading shows the density obtained by kernel smoothing (darker means higher density)



- Compute each normalised weights  $\tilde{\omega}_j^{(l)} = \omega_j^{(l)} / \sum_{k=1}^{N_p} \omega_j^{(k)}$
- If  $t_j < t_n$ , then set  $t_j = t_k$  ( $k = j + 1$ ) and go to the RESAMPLING step *else stop*.

### 3 Results

#### 3.1 Validation using modern pollen samples

In this section our goal is to validate the statistical framework, and not the vegetation model. Since the climate-pollen relationship may be locally biased at any given site, due to local errors induced by pollen production changes, non-homogeneous transport, different accumulation processes, etc, the method must be tested at a series of sites. We have therefore reconstructed modern climate at 30 sites randomly chosen from the modern pollen dataset.

The validation at each site  $s = 1:30$  was performed as follows:

- Repeat for each particle  $l=1:1,000$
- Sample  $C_s^{(l)}$ , one 6-dimensional climate parameter following the prior defined in the next section,
- Spin-up LPJ-GUESS for 500 years with repeated 1901–1930 monthly climate to which is added  $C_s^{(l)}$ , using the 1901 value of CO<sub>2</sub> atmospheric concentration of 296.3 ppmv,
- Simulate the 1901–2000 vegetation using 1901–2000 monthly climate to which is added  $C_s^{(l)}$ , under evolving CO<sub>2</sub> atmospheric concentration as obtained from the Carbon Cycle Model Linkage Project (McGuire et al. 2001).
- Retain  $V_s^{(l)}$ , the mean of NPP over the years 1961–1990,
- Weight the couple  $(C_s^{(l)}, V_s^{(l)})$  by  $\omega_s^{(l)} = p(Y_s | V_s^{(l)})$ .

##### 3.1.1 Definition of climate prior

As each modern site represents a single point in time, priors were chosen based on the CRU 1.2 (New et al. 2002) gridded set of climatological data for the European continent. For each validation site  $s = 1:30$ , the climate prior

$$p(C_s) = p(T_{jan}, T_{jul}) \cdot p(P_{win}) \cdot p(P_{spr}) \cdot p(P_{win}) \cdot p(P_{aut}) \\ = N \left( \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} V_{jan} & C_{jan,jul} \\ C_{jan,jul} & V_{jul} \end{pmatrix} \right) \cdot (N_{t=0}(0, \sigma_{Prec}^2))^4$$

is composed of a bivariate Gaussian distribution for  $T_{jan}$  and  $T_{jul}$  and 4 times the same 0-truncated independent Gaussian distribution for each precipitation parameter. Each distribution is centred on 0, the null anomaly equal to

expected climate. Variance and covariance parameters of the bivariate Gaussian law are derived from the CRU TS 1.2 (New et al. 2002) means for months January ( $V_{jan} = 6.6^2$ ) and July ( $V_{jul} = 4.1^2$ ). The covariance ( $C_{jan,jul} = 18.67$ ) or correlation ( $\rho = 0.69$ ) represent the modern European seasonal link between these 2 months. In doing so we allow temperature parameters for each site to be distributed over the whole modern European temperature set. The standard deviation for the precipitation parameter  $\sigma_{Prec}^2$  was arbitrarily chosen as 35 (in %) giving a probability of 0.005 to exceed an 100% precipitation increase.

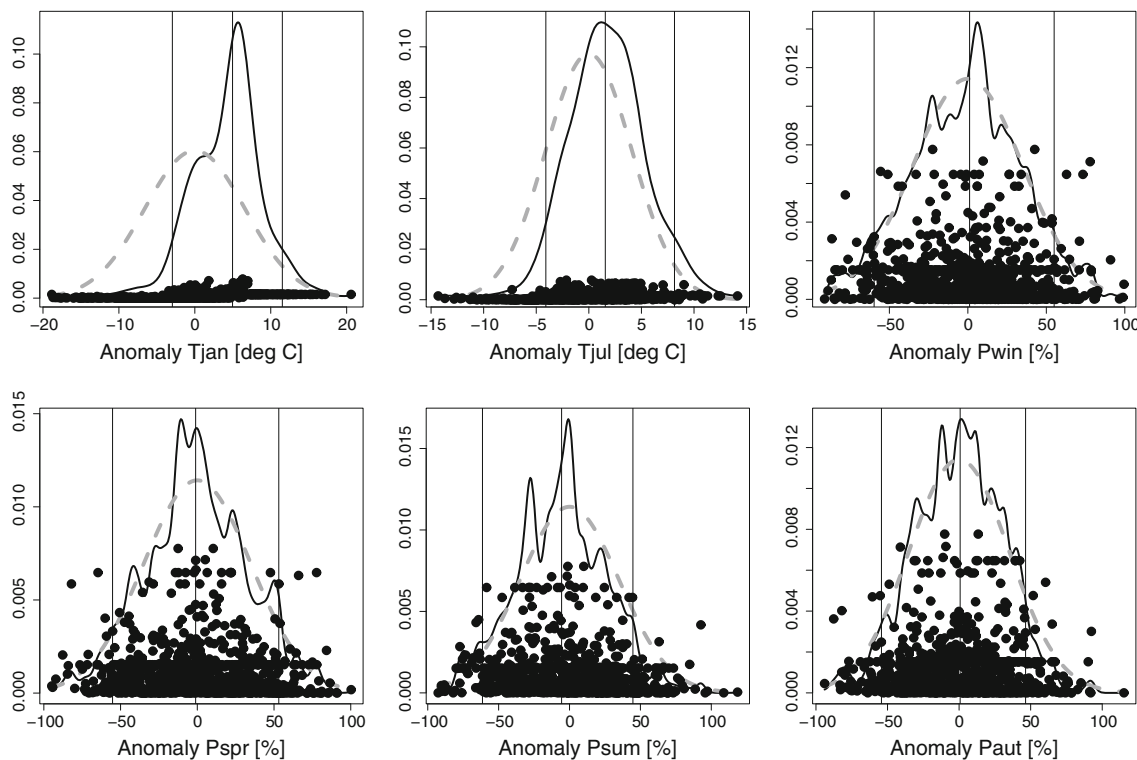
##### 3.1.2 Validation results

For each site  $s$  we obtain a set of 1,000 tested climates (the particles)  $C_s^{(l)}$ , associated to weights  $\omega_s^{(l)}$ . At each site, we summarise, the set of weighted climates by computing quantiles  $q_{0.025}$ ,  $q_{0.5}$  (median) and  $q_{0.975}$ . For visual representation we smooth the particles and obtain graphics showing the distribution of tested climates, for example Fig. 5 obtained for a Spanish site (41.39°N, 0.11°W). For the validation, however, we are interested in the global result obtained for the whole set of 30 climate reconstructions. Figure 6 presents the observed January and July temperatures versus their reconstructions. Table 3 summarises the reconstruction results.

The mean discrepancies between posterior medians and expected values of the 6 reconstructed parameters are negligible by comparison with interval widths (see Table 3). All the confidence intervals of the reconstructed temperatures contain the observed values (see Fig. 6). Thus, the method seems to be unbiased, at least at the continental level. Further, the temperature posteriors distributions are narrower than priors (see Table 3; Fig. 6). In other words, the inversion process is able to constrain both temperature variables from the specified prior.

Precipitation posteriors are not narrower than their priors. This shows that the inversion process is unable to constrain four precipitation variables (at a time) more than what has been specified as prior. That the vegetation model does not show a precipitation constraint is surprising. It may be assumed that in a non water-stressed region such as a European temperate forest an increase or a moderate decrease in precipitation would not change vegetation composition dramatically. However, in a highly water-stressed region, such as the Mediterranean dry region, water availability is one of the main vegetation drivers. We have therefore further investigated the reasons for this lack of constraint and found that it is due to a combination of the vegetation model and the prior chosen for precipitation.

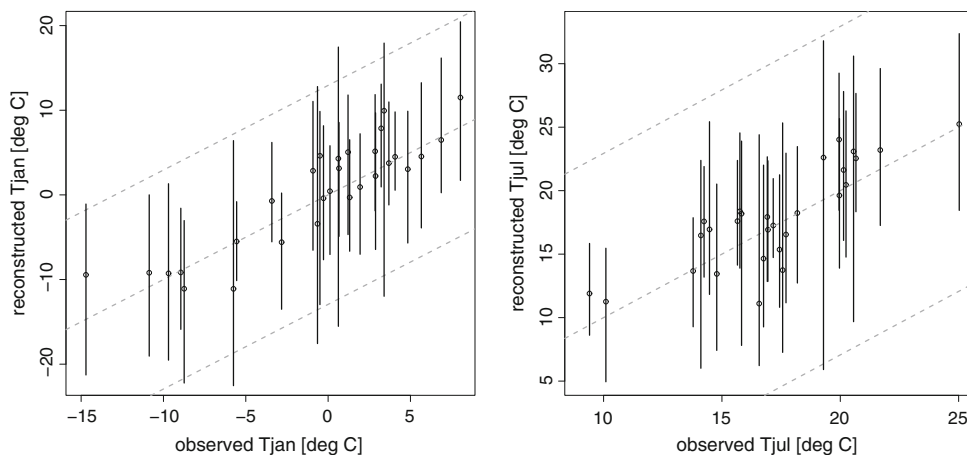
*Vegetation model:* We performed a sensitivity analysis using the inversion algorithm. A second validation was



**Fig. 5** Prior (grey dashed lines) and posterior (black lines) smoothed distributions of the 6 climate anomalies: ( $T_{jan}$ ,  $T_{jul}$ ,  $P_{win}$ ,  $P_{spr}$ ,  $P_{sum}$ ,  $P_{aut}$ ). The “particles” of climate proposed by the particle filter are the

black dots. Black thin vertical lines show the 2.5%, median and 97.5% quantiles of each posterior distribution. This is an example for a dry Mediterranean site from Spain (41.39°N, 0.11°W)

**Fig. 6** Observed values versus reconstructed (posterior) means and 95% confidence intervals for January and July temperatures in Celsius degrees for the 30 validation sites. The grey dashed lines represent the prior mean and quantiles. Note that prior mean is the expected (equal to observed) value. The black lines give the posterior interval range and the point is the reconstructed mean. (left) January temperature and (right) July temperature



performed at 20 modern Mediterranean dry points. These points were randomly selected from the set of points below latitude 42°N with less than 600 mm of rain per year, and with less than 70% of tree taxa pollen in their samples. For each climate proposed by the algorithm,  $C_s^{(l)}$ , we computed at a drought stress indicator (DSI),  $R_s^{(l)}$  based on quantities calculated by the vegetation model. This DSI is defined as the ratio between annual mean actual evapotranspiration (AET) and annual mean potential evapotranspiration (PET, e.g. Sykes et al. 1996). We found that, for a varying climate  $C_s^{(l)}$ , the DSI varies in [0.10; 0.50] which corresponds

to a xerophytic vegetation (Prentice et al. 1992). Thus, the vegetation simulated for these dry sites agrees with pollen data. This indicates two things. First, the link vegetation/pollen  $p(Y|V)$  performs well since we obtain a coherent match between simulated vegetation and sampled pollen. Second, the vegetation model seems to underestimate DSI variation as a function of precipitation change since the most humid  $C_s^{(l)}$  (proposed climates) should result in higher DSI than 0.5 and non-xerophytic vegetation.

*The priors for precipitation:* When specifying independent priors for the precipitation per season we implicitly

**Table 3** Means of the results obtained for 30 European points for the validation using present-day pollen samples

	Expected $q_{0.5}$	Prior $q_{0.975} - q_{0.025}$	Posterior $q_{0.5}$	Posterior $q_{0.975} - q_{0.025}$
$T_{\text{jan}}$ (°C)	0	25.9	0.85	16.8
$T_{\text{jul}}$ (°C)	0	16.1	0.82	12.3
$P_{\text{win}}$ (%)	0	137	-1.6	148
$P_{\text{spr}}$ (%)	0	137	-0.2	142
$P_{\text{sum}}$ (%)	0	137	0.7	143
$P_{\text{aut}}$ (%)	0	137	-2.6	138

Rows are the 6 reconstructed climate parameters. Columns are: (1) the expected median (0.5 quantile), (2) the prior 0.95 bilateral interval width (equal to 0.975 quantile minus 0.025 quantile), (3) the posterior median and (4) the posterior 0.95 bilateral interval width

specify a small relative range for the annual precipitation. This is due to compensation between seasons. In other words, to sample extreme annual precipitation it is necessary to sample extreme precipitation for the majority of seasons, which has a very small probability of occurring. A simple way to scan a large range for annual precipitation is to define a single annual precipitation parameter, but this would fix seasonality. For the Meerfelder Maar reconstruction we chose to let seasonality vary by using four precipitation parameters.

### 3.2 Temporal model inversion on the Meerfelder Maar pollen sediment core

#### 3.2.1 Prior definition

In an application to samples from a sediment core, the climate priors have to be elicited (obtained from expert knowledge) or obtained using other data than those used for reconstruction. For example one can use the Modern Analogue Method (MAT, Guiot and De Vernal 2007) and data for other sites obtained from the European Pollen Database (EPD, <http://www.europeanpollendatabase.net>) to build a prior distribution of climate for the studied European site during the Holocene. For Meerfelder Maar, since our goal is to present the method, we used an empirical prior based on MAT reconstruction of the climate using the same core. This allows us to assess how much the inversion approach modifies the standard MAT estimates.

We applied the MAT with the Meerfelder Maar sediment core to reconstruct the six climate anomalies ( $T_{\text{jan}}$ ,  $T_{\text{jul}}$ ,  $P_{\text{win}}$ ,  $P_{\text{spr}}$ ,  $P_{\text{sum}}$ ,  $P_{\text{aut}}$ ). For the analogue dataset, we used the modern pollen data and climate described above. We used the classic chord distance between pollen samples. Following Guiot and De Vernal (2007) we computed by cross validation a discriminant distance for the analogue selection and we selected a maximum number of 7 analogues if this distance is not reached. We used the means reconstructed by MAT as prior means. We used the same

temperature standard deviations as in the modern validation exercise ( $sd(T_{\text{jan}}) = 6.6$  and  $sd(T_{\text{jul}}) = 4.1$ ). The correlation between them was reduced (to 0.5), however, to relax the constraint on temperature seasonality during the Holocene. Standard deviation of the precipitation is 35 (in %).

#### 3.2.2 Reconstruction results

For atmospheric CO<sub>2</sub> input, we used a composite record composed of the ice core record from Indermuhle et al. (1999) for the period 11 cal ky BP to cal 990 cal year BP and the one from Siegenthaler et al. (2005) for the period 990 to 0 cal year BP.

Posterior reconstructions of January and July temperatures and annual precipitation are presented Fig. 7. The main events appearing in the pollen diagram (Fig. 2), compared to the climate reconstruction in Fig. 7 are:

- From about 10.6 cal ky BP, *Betula* and *Pinus* are replaced first by *Corylus*: this is the major event of the sequence indicating a warming of more than 10°C in winter and 5°C in summer and a precipitation increase of more than 500 mm/year.
- After 10 cal ky BP, first *Ulmus* and *Quercus* deciduous, second *Tilia* and finally *Fraxinus* appear according to the classical succession in Europe; this does not translate into any significant change in our climate reconstruction.
- Just before 6 cal ky BP, *Alnus* becomes dominant over *Fraxinus*, *Ulmus* and *Tilia*: this seems to indicate a slight increase of temperature (a few degrees Celsius) and of precipitation (100–200 mm/year).
- The second major event is at about 3.5 cal ky BP, *Corylus* is replaced by *Fagus* and secondly by *Betula*; this seems to indicate a slight decrease of January temperature (a few degrees Celsius) and an important precipitation decrease of about 300 mm/year.
- At about 2.5 cal ky BP, grass and shrub group (GrSh) becomes important and is dominant after 1 cal ky BP. The anthropogenic deforestation should be translated

into a reconstructed increase of drought, but this is not seen here. This seems to indicate that the method is robust against human disturbance.

This comparison of both figures show that the climate variations, as reconstructed by model inversion, are coherent with the pollen curves and seems to be robust against anthropogenic disturbance. This is a major argument in favor of these results.

The confidence intervals of January and July reconstructed temperatures have widths ranging from 10 to 20°C. This is coherent with, in the Gaussian case, a standard deviation of 2.5–5°C. These large posterior intervals are partly due to the large prior intervals. We have chosen to show means and quantiles here, but since the posterior is highly non-Gaussian the median or mode(s) of the posterior distribution may be preferred. A discussion of the results obtained follows:

In the earliest part of the sequence, 11–10.5 cal ky BP, corresponding to the transition to the Holocene period, there is a good agreement between our reconstructions and the MAT estimations.

During the period 10.5–7.5 the  $T_{\text{jan}}$  reconstruction shows two sets of possible values (high probability in dark red). In the first set, temperatures of around 10°C are reconstructed, which is higher than the MAT means. In the second set, the temperatures are lower at around –2°C. We note that these sets are non-continuous; the most probable climate jumps between paths. This results from the lack of an explicit climate link between samples in our statistical model.

At 7.5 cal ky BP the  $T_{\text{jan}}$  confidence interval becomes tighter. This is likely to be an artifact resulting from the spike in *Corylus* pollen at this date. This causes a distortion in the diagram and therefore in the model as it tries to reproduce this abrupt peak.

The period 6.3–3.7 cal ky BP is a stable period in the pollen diagram, however, both January and July reconstructed temperatures show high frequency variability. This is again due to the absence of climate correlation in our model and may be further associated to an overfitting of the  $p(Y|V)$  distribution. This would cause the vegetation model to reproduce non-significant changes in pollen samples, and give these a climatic interpretation. As there is no constraint in climate change through time, we cannot limit the reconstructed climate by reference to the value obtained for the previous sample. This results in non-significant fluctuations in the reconstruction.

The last 3.7 to 0 cal ky BP period is characterised by more marked changes in pollen composition. Here, however, the high variability in the reconstruction disappears, mean variability is more coherent in time and varies around the MAT estimates. This suggests that the changes in

pollen composition during this period are sufficiently large to adequately constrain the reconstructed climate in adjacent samples, and result in a smoother curve.

The prior and posterior distributions of annual precipitations are similar except for a short period around 8 cal ky BP. This is unsurprising, as the validation exercise indicated that there is little or no constraint on precipitation. However, as Litt et al. (2009) found a quite different pattern for precipitation with a value close to 600 mm/year at the top of the core instead of our value of around 1,000 mm/year, we devised a second test to check the precipitation constraint. Note that precipitation values recorded for the period 1961–1990 at the nearest meteorological station (approximately 30 km from Meerfelder Maar) show a mean annual precipitation of 908 mm/year (Litt et al. 2009). We specified annual precipitation priors following a linear relationship between 500 mm/year at 11 cal ky BP and 600 mm/year at 0 cal ky BP, with precipitation split amongst seasons following the modern seasonal distribution. The posterior January temperature and annual precipitation reconstruction obtained with this second test are shown in Fig. 8.

For the period 0–3 cal ky BP, reconstructed temperature (Fig. 8 top) is in good agreement with the first test (Fig. 7 top) and prior and posterior precipitation are nearly the same. This implies that precipitation is not a constraining parameter for this period. For earlier periods January reconstructions for different experiments differ and prior and posterior precipitation for this second experiment differ too. This indicates that precipitation may have been somewhat higher than the values reconstructed by Litt et al. using the Bayesian Indicator Taxa method (Neumann et al. 2007) which is a Bayesian tuning of the Probability Density Function (pdf) method of Köhl et al. (2002).

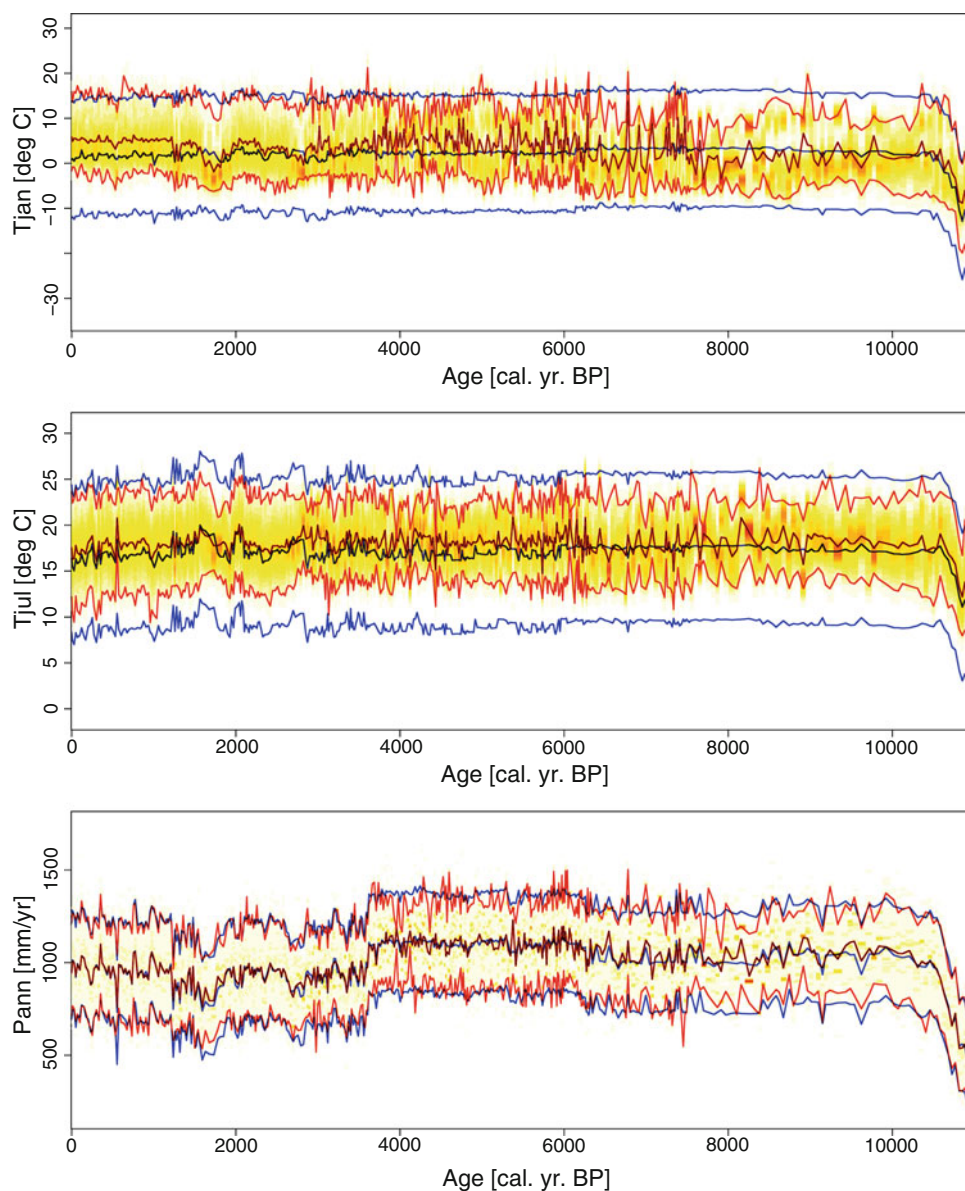
## 4 Conclusion and discussion

Climate reconstruction by static inversion initiated by Guiot et al. (2000) is now used to take CO<sub>2</sub> variations into account (e.g. Wu et al. 2007) and provides a method to reconstruct climate using different proxies (e.g. Hatté et al. 2009). The climate reconstruction by dynamic inversion retains these advantages and integrates a new generation of vegetation models. It is achieved using a Bayesian hierarchical model which sets the basis for causative modelling. We hope that this will encourage other work to use and extend this framework, and that the technical tools (statistical model and algorithm) presented in this article will help.

The use of a dynamic vegetation model has allowed an improvement of the “vegetation model inversion method”



**Fig. 7** Climate reconstruction for Meerfelder Maar sequence during the Holocene using prior based on MAT estimations. Prior (*blue lines*): mean and 95% confidence interval. Posterior (*red lines*) mean (*highlighted with black*) and 95% confidence interval. The background colour (*yellow to red*) shows the posterior smoothed (low to high) density. For the three plots, *x*-axis shows the age in cal yr BP, *y*-axis shows: (*top*) January reconstructed temperature in °C, (*center*) July reconstructed temperature in °C and (*bottom*) annual reconstructed precipitation in mm/year. Values recorded for the period 1961–1990 at the nearest meteorological station show a January mean temperature of  $-0.3$  and  $16.3^{\circ}\text{C}$  for July. Mean annual precipitation is 908 mm/year (Litt et al. 2009)



by including a temporal link. This allows a better exploitation of the information available from the fossil record. Further, the dynamic aspect of the model allows us to relax the assumption of equilibrium between vegetation and climate. In this paper, we focused on climate, but other variables are available as output of the vegetation model, e.g. primary production and carbon storage. These are useful for carbon cycle studies (see Wu et al. 2008) or could be used for paleo-fire modelling and more generally, studies of the past needing vegetation as input. The current study has, however, identified a number of problems listed below, and future work will concentrate on resolving these.

The validation tests showed that precipitation cannot be reconstructed as a four dimensional space. This seems to be partly due to the low sensitivity of the vegetation model to precipitation changes, and a sensitivity study is required.

Until this problem has been addressed, we suggest that future use of this algorithm should use a single annual precipitation anomaly.

The results obtained using this method with the Meerfelder Maar sequence showed that the stability of the reconstructed values are linked to the pollen signal. When the pollen signal is nearly constant our reconstruction method over-amplifies small variations in the pollen signal. When the pollen indicates two possible climates, the reconstruction may jump between two states, forming a non-continuous path of reconstructed values. However, when the changes occur in the pollen assemblages, the reconstructed variability is more coherent (low-frequency), for example during the period 3.7 to 0 cal ky BP. These differences are probably due to, (a) an over-fitting of the  $p(Y|V)$  surfaces which force the vegetation model to follow

the non-significant noise in pollen as if it were caused by climate change; (b) the absence of any direct time correlation in climate, which would result in a smoother reconstructed climate; (c) the use of a particle filter algorithm.

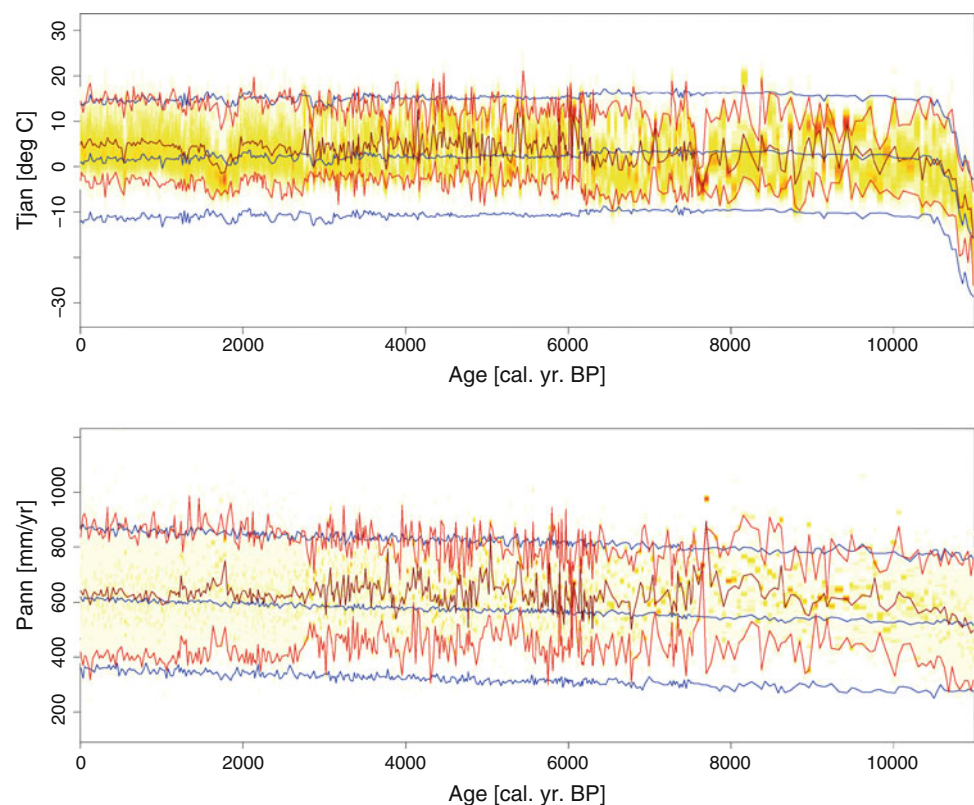
*The  $p(Y|V)$  model:* We modelled the link between simulated vegetation and pollen using non-parametric surfaces. These surfaces are fitted using modern pollen and simulated vegetation. They fail to include the uncertainty that exists on the link between vegetation and pollen. The inclusion of this uncertainty would summarise and transfer the incomplete knowledge of the link between vegetation and pollen, from the calibration to the reconstruction. This is a major departure from Bayesian modelling, and the next goal to improve the inversion method lies in using a parametric  $p(Y|V)$  which allows the full propagation of uncertainty between calibration and reconstruction. The Bayesian framework for calibration and reconstruction in two separated steps has been presented in Haslett et al. (2006).

*Temporal correlation in the climate:* In contrast to the MAT reconstructions, which need to assume independence between samples, we model a vegetation link and obtain reconstructions that seem to be “noisier”. As there is a dependency between samples, the vegetation model may require a larger climate change to fit both points than in the independent scheme. For the inversion of a dynamic

model, it therefore seems essential to define a temporal climate correlation to counterbalance the effect of vegetation correlation, at least, when the resolution of the core is high. When the resolution of the core is low, correlation of the vegetation between samples is low and the reconstruction is nearly independent. However, the particle filter proposed here remains computationally more efficient and theoretically safer when using a dynamic vegetation model. Indeed, for a the static (or “at equilibrium”) inversion of a dynamic model, the experimenter has to run the model from nothing to the equilibrium (between vegetation and climate). This is called the burn-in phase and the diagnosis of convergence to equilibrium is always critical. With our method he/she can start from the vegetation reconstructed for the previous core point and has a fixed length of time to run the algorithm for, namely the time separating the two core points.

*Filtering algorithm:* The simple particle filter algorithm infers climate sequentially along the sedimentary sequence. However, it only optimises for coherence between the previous sample and the current one, and does not take into account the following sample. This feature can result in a chaotic path around the real climate. While some more complex algorithms attempt to minimise this problem, this remains an intrinsic problem of particle filtering, despite recent advances in the field (see for example <http://www-sigproc.eng.cam.ac.uk/smc/>).

**Fig. 8** Climate reconstruction for Meerfelder Maar sequence during the Holocene using forced precipitation. Prior (blue lines): mean and 95% bilateral interval. Posterior (red lines) mean (highlighted with black) and 95% bilateral interval. The background colour (yellow to red) shows the posterior smoothed (low to high) density. For both plots, x-axis shows the age in cal year BP, y-axis shows: (top) January reconstructed temperature in °C and (bottom) annual reconstructed precipitation in mm/year



The use of the filtering algorithm arises from the need to reduce the dimension of the climate space (equal to the number of samples times the number of climate variables). Since we cannot use another algorithm for the dynamic inversion, this is potentially an obstacle for: a. taking into account radiocarbon and other age errors and, b. the inference of any parameter that is dependent on the whole core. As the particle filter needs fixed stopping times (here the dates of the samples in the core) to simulate and weight particles, major changes and theoretical work would be required to integrate a possible error on this stopping time. Second, since the filter handles data sequentially, any parameter dependent on the whole core can only be estimated at the end of filtering. For example, this is a problem when considering a parameter  $\theta$  of temporal correlation for the climate. At time  $t_2$ , the second point of the core, the algorithm has only the prior information to link  $t_1$  and  $t_2$  samples. The information about  $\theta$  is updated along the core. This means that the reconstruction quality varies between points of the same core.

Despite the limits mentioned, the particle filter algorithm remains a promising tool for inversion of dynamic models or interpolation using such dynamic models. While work is required to further adapt this method, we believe

that it will be a useful tool in the field of climatology, in which a number of dynamic models are currently used or have been proposed.

**Acknowledgments** The first author would like to thank Pascal Monestiez and Joel Chadoeuf from BSP laboratory in INRA Avignon for fruitful discussions and statistical coaching. He was supported in computing work by Cyrille Blanpain and Philippe Dussouillez and by Judicael Lebamba for the pollen diagram. The manuscript has been greatly improved thanks to the comments from three anonymous reviewers. The authors are grateful for financial support for project DECVEG from the European Science Foundation (ESF) and Institut National des Sciences de l'Univers (INSU) under the EUROCORES Programme EuroCLIMATE. This work is also a contribution to the project PICC funded by the French Agence Nationale de la Recherche (programme blanc).

**Open Access** This article is distributed under the terms of the Creative Commons Attribution Noncommercial License which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

## Appendix 1: Additional vegetation parameters

Tables of additional model parameters.

**Table 4** Additional species parameters and bioclimatic limits used in the LPJ-GUESS model

Species	$k_{\text{la,sa}}$ ( $\text{m}^2 \text{m}^{-2}$ )	Leaf long. (y)	$R_{\text{fire}}$	Chilling ( $b,k$ )	Longevity (y)
<i>Abies alba</i>	6,000	6	0.2	(100, 0.05)	350
<i>Alnus incana</i>	6,000	0.5	0.2	(100, 0.05)	200
<i>Betula pendula</i>	6,000	0.5	0.2	(500, 0.02)	300
<i>Betula pubescens</i>	6,000	0.5	0.1	(100, 0.05)	300
<i>Carpinus betula</i>	6,000	0.5	0.1	(1,000, 0.025)	150
<i>Corylus avellana</i>	5,000	0.5	0.2	(200, 0.05)	100
<i>Fagus sylvatica</i>	6,000	0.5	0.1	(220, 0.03)	400
<i>Fraxinus excelsior</i>	6,000	0.5	0.1	(100, 0.05)	400
<i>Picea abies</i>	6,000	6	0.1	(100, 0.05)	400
<i>Pinus sylvestris</i>	3,500	2	0.4	(100, 0.05)	500
<i>Pinus halepensis</i>	4,000	2	0.4	(100, 0.05)	350
<i>Populus tremula</i>	6,000	0.5	0.2	(100, 0.05)	160
<i>Quercus coccifera</i>	3,200	3	0.5	(100, 0.05)	350
<i>Quercus ilex</i>	4,000	3	0.3	(100, 0.05)	350
<i>Quercus robur</i>	6,000	0.5	0.2	(100, 0.05)	500
<i>Tilia cordata</i>	6,000	0.5	0.1	(1,000, 0.025)	500
<i>Ulmus glabra</i>	6,000	0.5	0.1	(100, 0.05)	400
C3 grass	–	1	1.0	–	1

$k_{\text{la,sa}}$  ratio leaf area to sapwood cross-sectional area, *Leaf long* leaf longevity,  $R_{\text{fire}}$  fraction of a species' patch population and litter that survives a fire, *Chilling* ( $b,k$ ) chilling parameters, as described by Sykes et al. (1996), *Longevity* tree species longevity

**Table 5** Shade tolerance parameters used in the LPJ-GUESS model

Shade class	Parff <sub>min</sub> (10 <sup>5</sup> J m <sup>-2</sup> d <sup>-1</sup> )	Greff <sub>min</sub> (kg C <sub>leaf</sub> m <sup>-2</sup> y <sup>-1</sup> )	Est <sub>max</sub> (m <sup>-2</sup> y <sup>-1</sup> )	α	Conv <sub>sap</sub> (y <sup>-1</sup> )
St	3.50	0.05	0.05	3	0.10
Ist	5.75	0.06	0.10	6	0.15
Si	8.00	0.07	0.30	9	0.20

See Smith et al. (2001) for full details. Parff<sub>min</sub> minimum photosynthetically active radiation at the forest floor for establishment, greff<sub>min</sub>: Growth efficiency threshold. est<sub>max</sub> maximum sapling establishment rate, α recruitment shape parameter, conv<sub>sap</sub> sapwood to hardwood conversion rate. Relative to Si species, St tree species require less photosynthetically active radiation at the forest floor to establish, produce fewer saplings under full light conditions, have a lower threshold growth efficiency for stress mortality, have less suppression of establishment at low forest-floor NPP, and convert proportionally less sapwood to hardwood annually. IST species have intermediate characteristics

**Appendix 2: Comprehensive description of the particle filter**

The aim of this appendix is to fully describe the particle filter algorithm used for inference. We start from the model’s equations and show how to obtain the posterior distribution.

Initialisation of the algorithm

For time  $t_1$  of the first pollen core sample, according to the Bayes theorem and a development from Eq. 5

$$p(V_{t_1}, C_{t_1} | Y_{t_1}) \propto p(Y_{t_1} | V_{t_1}) \cdot p_{LPJ}(V_{t_1} | V_{t_0}, C_{t_1}) \cdot p(C_{t_1})$$

Time  $t_0$  (see Fig. 3) is a notation to show that the time relationship between vegetation always exists. As we do not have data to reconstruct  $t_0$  vegetation we assume an equilibrium hypothesis between climate and vegetation at time  $t_1$  and then use:

$$p(V_{t_1}, C_{t_1} | Y_{t_1}) \propto p(Y_{t_1} | V_{t_1}) \cdot p_{LPJ}(V_{t_1} | C_{t_1}) \cdot p(C_{t_1})$$

From these equations the algorithm will be: Sample  $N_p$  “particles”  $C_{t_1}^{(l=1:N_p)}$  from  $p(C_{t_1})$ . “Particles” describe the simulated climates and later the simulated couples of climate-vegetation. For each particle we run LPJ-GUESS for 500 years to reach equilibrium and obtain  $N_p$  “particles”  $(C_{t_1}, V_{t_1})^{(l=1:N_p)}$  following  $p_{LPJ}(V_{t_1} | C_{t_1}) \cdot p(C_{t_1})$ . For each particle we then compute the non-normalized importance weights:

$$\omega_{t_1}^{(l)} = \frac{p(Y_{t_1} | V_{t_1}^{(l)}) \cdot p_{LPJ}(V_{t_1}^{(l)} | C_{t_1}^{(l)}) \cdot p(C_{t_1}^{(l)})}{p_{LPJ}(V_{t_1}^{(l)} | C_{t_1}^{(l)}) \cdot p(C_{t_1}^{(l)})} = p(Y_{t_1} | V_{t_1}^{(l)})$$

normalizing the weights (so their sum is 1) we obtain

$$\tilde{\omega}_{t_1}^{(l)} = \frac{\omega_{t_1}^{(l)}}{\sum_{k=1}^{N_p} \omega_{t_1}^{(k)}}$$

A discrete approximation of  $p(V_{t_1}, C_{t_1} | Y_{t_1})$  is therefore

$$\tilde{p}(V_{t_1}, C_{t_1} | Y_{t_1}) = \sum_{l=1}^{N_p} \tilde{\omega}_{t_1}^{(l)} \cdot \delta_{(V_{t_1}, C_{t_1})^{(l)}}$$

where  $\delta_{(V_{t_1}, C_{t_1})}$  is the Dirac mass applied at  $(V_{t_1}, C_{t_1})$ .

Step  $t_j$  of the algorithm

Let  $t_i$  and  $t_j$  be two consecutive core times. Starting at the  $t_j$  step of the algorithm we have  $(V_{t_1:t_i}, C_{t_1:t_i})^{(l=1:N_p)}$ ,  $N_p$  “histories” of vegetation and climate weighted by  $\tilde{\omega}_{t_i}^{(l=1:N_p)}$ . These series and weights define the discrete approximation of  $p(V_{t_1:t_i}, C_{t_1:t_i} | Y_{t_1:t_i})$ . We want to add a coherent “particle”  $(V_{t_j}, C_{t_j})^{(l)}$  to each history. Each new history obtained by concatenation of  $(V_{t_1:t_j}, C_{t_1:t_j})^{(l)}$  and  $(V_{t_j}, C_{t_j})^{(l)}$  with their associated weights  $\tilde{\omega}_{t_j}^{(l)}$  must define the discrete approximation of  $p(V_{t_1:t_j}, C_{t_1:t_j} | Y_{t_1:t_j})$ .

By Bayes theorem and the model definition  $p(V_{t_1:t_j}, C_{t_1:t_j} | Y_{t_1:t_j})$  can be developed

$$p(V_{t_1:t_j}, C_{t_1:t_j} | Y_{t_1:t_j}) \propto p(Y_{t_j} | V_{t_j}) \cdot p(V_{t_j}, C_{t_j} | V_{t_i}) \cdot p(V_{t_1:t_i}, C_{t_1:t_i} | Y_{t_1:t_i}) \propto p(Y_{t_j} | V_{t_j}) \cdot p_{LPJ}(V_{t_j} | V_{t_i}, C_{t_j}) \cdot p(C_{t_j}) \cdot p(V_{t_1:t_i}, C_{t_1:t_i} | Y_{t_1:t_i})$$

We simply have to sample  $N_p$  “particles”  $C_{t_j}^{(l=1:N_p)}$  of climate parameters from  $p(C_{t_j})$ . For each particle we run LPJ-GUESS for  $t_j - t_i$  years and obtain  $N_p$  “particles”  $(C_{t_j}, V_{t_j})^{(l=1:N_p)}$  following  $p_{LPJ}(V_{t_j} | V_{t_i}, C_{t_j}) \cdot p(C_{t_j})$ . For each particle we then recompute non-normalized importance weights:

$$\begin{aligned} \omega_{t_j}^{(l)} &= \frac{p(Y_{t_j} | V_{t_j}^{(l)}) \cdot p_{LPJ}(V_{t_j}^{(l)} | V_{t_i}^{(l)}, C_{t_j}^{(l)}) \cdot p(C_{t_j}^{(l)}) \cdot p(V_{t_1:t_i}, C_{t_1:t_i} | Y_{t_1:t_i})}{p_{LPJ}(V_{t_j}^{(l)} | V_{t_i}^{(l)}, C_{t_j}^{(l)}) \cdot p(C_{t_j}^{(l)})} \\ &= p(Y_{t_j} | V_{t_j}^{(l)}) \cdot \sum_{k=1}^{N_p} \tilde{\omega}_{t_i}^{(k)} \cdot \delta_{(V_{t_1:t_i}, C_{t_1:t_i})^{(k)}} \\ &= p(Y_{t_j} | V_{t_j}^{(l)}) \cdot \tilde{\omega}_{t_i}^{(l)} \end{aligned}$$

and then renormalize the weights according to

$$\tilde{\omega}_{t_j}^{(l)} = \frac{\omega_{t_j}^{(l)}}{\sum_{k=1}^{N_p} \omega_{t_j}^{(k)}}$$



## Regeneration

The sequential importance sampling algorithm presented above is theoretically valid, but its efficiency decreases with time, i.e. after a number of time steps the discrete approximation of the posterior distribution will be reduced to one single particle with weight equal to 1. The solution is to add a regeneration step making the algorithm a particle filter algorithm (Doucet et al. 2001).

The regeneration step consists of sampling with replacement particles according to their weights. It implies that a particle with a weight of 0 will not be sampled further and is removed but that those with high weights will be sampled many times and are therefore multiplied. Since the regeneration step introduces a Monte Carlo error in the estimation, we do not have to do it if the particles are well distributed (Doucet et al. 2001). The criterion used to determine the need to resample is called the Effective Sample Size (ESS) criterion.

$$\text{ESS}_t = \left( \sum_{l=1}^{N_p} \left( \tilde{\omega}_t^{(l)} \right)^2 \right)^{-1}$$

The ESS criterion takes its values in the range 0 to  $N_p$ . If the degeneracy of particles is too high (and thus the ESS lies under an arbitrary threshold of  $N_p/2$ ) we apply the regeneration step (resample) and we reset all weights to  $1/N_p$ , otherwise we keep all particles and weights.

Different methods are available for the resampling step. We use the efficient residual sampling technique from (Liu and Chen 1998): At step  $t$  we have  $N_p$  particles  $(V_t, C_t)^{(l=1:N_p)}$  weighted by  $\tilde{\omega}_t^{(l=1:N_p)}$ . In a first step, for each particle  $(l)$ , we keep  $n^{1,(l)} = \lfloor N_p \tilde{\omega}_t^{(l)} \rfloor$  copies of the particle. In a second step we randomly sample  $m = N - \sum_{l=1}^{N_p} n^{1,(l)}$  particles in the set of all particles weighted by  $\omega_t^{1,(l)} \propto N_p \cdot \tilde{\omega}_t^{(l)} - n^{1,(l)}$ .

## References

- Aitchison J (1982) The statistical analysis of compositional data. *J Roy Stat Soc Ser B* 44(2):139–177
- Bordon A (2008) Dynamique de la végétation et variations climatiques dans les Balkans au cours du dernier cycle climatique à partir des séquences polliniques des lacs Maliq et Ochrid (Albanie). PhD thesis, Université de Franche-Comté
- Bottema S (1974) Late quaternary vegetation history of Northwestern Greece. PhD thesis, Rijksuniversiteit Groningen
- Bругиапaglia E (1996) Dynamique de la végétation tardiglaciaire et holocène dans les Alpes italiennes nord-occidentales. PhD thesis, Université Aix-Marseille III
- Bugmann H (2001) A review of forest gap models. *Clim Change* 51:259–305
- COHMAP Members (1988) Climatic changes of the last 18,000 years: observations and model simulations. *Science* 241:1043–1052
- Cressie N (1991) *Statistics for spatial data*. Wiley, New York
- Doucet A, De Freitas N, Gordon N (2001) *Sequential Monte Carlo Methods in practice: statistics for engineering and information science*. Springer, New York
- Gerten D, Schabhoff S, Haberlandt U, Lucht W, Sitch S (2004) Terrestrial vegetation and water balance an hydrological evaluation of a dynamic global vegetation model. *J Hydrol* 286:249–270
- Guiot J, De Vernal A (2007) *Transfer functions: methods for quantitative paleoceanography based on microfossils*. vol 1, chapter C. Elsevier, Amsterdam
- Guiot J, Torre F, Jolly D, Peyron O, Borreux JJ, Cheddadi R (2000) Inverse vegetation modeling by Monte Carlo sampling to reconstruct paleoclimate under changed precipitation seasonality and CO<sub>2</sub> conditions: application to glacial climate in Mediterranean region. *Ecol Model* 127(1):119–140
- Haslett J, Whitley M, Bhattacharya S, Salter Townshend M, Wilson S, Allen JRM, Huntley B, Mitchell FJG (2006) Bayesian paleoclimate reconstruction. *J R Stat Soc Ser A* 169(3):395–438
- Hatté C, Rousseau D, Guiot J (2009) Climate reconstruction from pollen and  $\delta^{13}\text{C}$  records using inverse vegetation modeling—implication for past and future climates. *Climate Past* 5:147–156
- Haxeltine A, Prentice IC (1996) BIOME3: an equilibrium terrestrial biosphere model based on ecophysiological constraints, resource availability and competition among plant functional types. *Glob Biogeochem Cycles* 10:693–709
- Hickler T, Smith B, Sykes MT, Davis M, Sugita S, Walker K (2004) Using a generalized vegetation model to simulate vegetation dynamics in northeastern USA. *Ecology* 85:519–530
- Indermuhle A, Stocker TF, Joos F, Fisher H, Smith HJ, Wahlen M, Deck B, Mastroianni D, Tshumi J, Blunier T, Meyer R, Stauffer B (1999) Holocene carbon-cycle dynamics based on CO<sub>2</sub> trapped in ice at Taylor Dome, Antarctica. *Nature* 398:121–126
- Jost A, Lunt D, Kageyama M, Abe-Ouchi A, Peyron O, Valdes PJ, Ramstein G (2005) High-resolution simulations of the last glacial maximum climate over Europe: a solution to discrepancies with continental palaeoclimatic reconstructions?. *Clim Dyn* 24:577–590
- Koca D, Smith B, Sykes MT (2006) Modelling regional climate change effects on Swedish ecosystems. *Clim Change* 78:381–406
- Kühl N, Gebhardt Litt T, Hense A (2002) Probability density function as botanical-climatological transfer functions for climate reconstruction. *Quat Res* 58:381–392
- Litt T, Schölzel C, Kühl N, Brauer A (2009) Vegetation and climate history in the westeifel volcanic field (Germany) during the past 11 years based on annually laminated lacustrine maar sediments. *Boreas*. doi:10.1111/j.1502-3885.2009.00096.x
- Liu JS, Chen R (1998) Sequential Monte Carlo methods for dynamic systems. *J Am Stat Assoc* 93:1032–1044
- Loader C (1999) *Local regression and likelihood: statistics and computing*. Springer, New York
- McGuire AD, Sitch S, Clein JS (2001) Carbon balance of the terrestrial biosphere in the twentieth century: analyses of CO<sub>2</sub>, climate and land use effects with four process-based ecosystem models. *Glob Biogeochem Cycles* 15:183–206
- Miller PA, Giesecke T, Hickler T, Bradshaw RHW, Smith B, Seppä H, Valdes PJ, Sykes MT (2008) Exploring climatic and biotic controls on holocene vegetation change in Fennoscandia. *J Ecol* 96(2):247–259
- Murray NB, Cannell MGR, Smith I (1989) Date of budburst of fifteen tree species in Britain following climatic warming. *J Appl Ecol* 26:693–700
- Neumann F, Schölzel C, Litt T, Hense A, Stein M (2007) Holocene vegetation and climate history of the northern Golan heights (Near East). *Vegetation History Archaeobotany*



- New M, Lister D, Hulme M, Makin I (2002) A high-resolution data set of surface climate over global land areas. *Clim Res* 21:1–25
- Peyron O, Guiot J, Cheddadi R, Tarasov P, Reille M, De Beaulieu JL, Bottema S, Andrieu V (1998) Climatic reconstruction in Europe for 18,000 year BP. from pollen data. *Quat Res* 49:183–196
- Prentice IC, Bartlein PJ, Webb TI (1991) Vegetation and climate change in eastern North America since the last glacial maximum. *Ecology* 72:2038–2056
- Prentice IC, Cramer W, Harrison SP, Leemans R, Monserud RA, Solomon AM (1992) A global biome model based on plant physiology and dominance, soil properties and climate. *J Biogeography* 19:117–134
- Prentice IC, Helmisaari H (1991) Silvics of north European trees: compilation, comparisons and implications for forest succession modelling. *For Ecol Manag* 42:79–93
- Sanchez Goni MF, Hannon G (1999) High altitude vegetational pattern on the Iberian Mountain Chain (north-central Spain) during the Holocene. *Holocene* 9(1):39–57
- Shugart HH (1984) A theory of forest dynamics: the ecological implications of forest succession models. Springer, New York
- Siegenthaler U, Monnin E, Kawamura K, Spahni R, Shwanda J, Stauffer B, Stocker TF, Barnola JM, Fisher H (2005) Supporting evidence from the EPICA drilling Maud Land ice core for atmospheric CO<sub>2</sub> change during the past millenium. *Tellus* 57B(7):51–57
- Sitch S, Smith B, Prentice IC, 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 Chang Biol* 9:161–185
- Smith B, Prentice IC, Sykes MT (2001) Representation of vegetation dynamics in modelling of terrestrial ecosystems: comparing two contrasting approaches within European climate space. *Glob Ecol Biogeogr* 10:621–637
- Sykes MT, Prentice IC, Cramer W (1996) A bioclimatic model for the potential distribution of northern European tree species under present and future climates. *J Biogeogr* 23:203–233
- Thonicke K, Venevsky S, Sitch S, Cramer W (2001) The role of fire disturbance for global vegetation dynamics: coupling fire into a dynamic global vegetation model. *Glob Ecol Biogeogr* 6:483–495
- Webb TJ, Bryson RA (1972) Late- and postglacial climatic change in the northern Midwest, USA: quantitative estimates derived from fossil pollen spectra by multivariate statistical analysis. *Quat Res* 2:70–115
- Woodward FI (1987) Climate and plant distribution. Cambridge University Press, Cambridge
- Wu H, Guiot J, Brewer S, Guo Z (2007) Climatic changes in Eurasia and Africa at the last glacial maximum and mid-Holocene: reconstruction from pollen data using inverse vegetation modelling. *Clim Dyn* 29(2):211–229
- Wu H, Guiot J, Peng C, Guo Z (2008) New coupled model used inversely for reconstructing past terrestrial carbon storage from pollen data: validation of model using modern data. *Glob Chang Biol* 15:82–96
- Young GA, and Smith RL (2005) Essentials of statistical inference: Cambridge series in statistical and probabilistic mathematics. Cambridge University Press, Cambridge