Contents lists available at ScienceDirect

Palaeogeography, Palaeoclimatology, Palaeoecology

Improved reconstruction of palaeo-environments through unravelling of preserved vegetation biomarker patterns

Boris Jansen [a,*], E. Emiel van Loon [b], Henry Hooghiemstra [c], Jacobus M. Verstraten [a]

[a] Institute for Biodiversity and Ecosystem Dynamics (IBED) - Earth Surface Sciences, Universiteit van Amsterdam, Nieuwe Achtergracht 166, NL1018WV Amsterdam, The Netherlands
[b] Institute for Biodiversity and Ecosystem Dynamics (IBED) - Computational Geo-Ecology, Nieuwe Achtergracht 166, NL1018WV Amsterdam, The Netherlands
[c] Institute for Biodiversity and Ecosystem Dynamics (IBED) - Paleoecology and Landscape Ecology, Universiteit van Amsterdam, P.O. Box 94248, NL1090GE Amsterdam, The Netherlands

ARTICLE INFO

ABSTRACT

Montane forest composition and specifically the position of the upper forest line (UFL) is very sensitive to climate change and human interference. As a consequence, reconstructions of past altitudinal UFL dynamics and forest species composition are crucial instruments to infer past climate change and assess the impact of (pre)historic human settlement. One of the most detailed methods available to date to reconstruct past vegetation dynamics is the analysis of fossil pollen. Unfortunately, fossil pollen analysis does not distinguish beyond family or generic level in most cases, while its spatial resolution is limited amongst others by wind-blown dispersal of pollen, affecting the accuracy of pollen-based reconstructions of UFL positions. To overcome these limitations, we developed a new method based on the analysis of plant-specific groups of biomarkers preserved in suitable archives, such as peat deposits, that are unravelled into the plant species of origin by the newly developed VERHIB model. Here we present this new method of biomarker analysis and describe its first application in a peat sequence from a biodiversity hotspot of montane rainforest in the Ecuadorian Andes. We show how a combination of the new biomarker application with conventional pollen analysis from the same peat sequence yields a reconstruction of past forest compositions, including UFL dynamics, with previously unattainable detail.

© 2009 Elsevier B.V. All rights reserved.

1. Introduction

Reconstructions of altitudinal shifts in the upper forest line (UFL) position in mountainous areas as well as species composition of ecotone forests are instrumental to appreciate past forest dynamics and to document impact of (pre)historic human settlement (e.g. Danby and Hik, 2007; Jolly and Haxeltine, 1997; Kullman, 2007; Pope et al., 2001; Roberts, 1998; Still et al., 1999). Traditionally, analysis of fossil pollen spectra and/or stable carbon isotope ratios of organic matter preserved in peat or sediment deposits are used (Clark and McLachlan, 2003; Mackay et al., 2003; Mayle et al., 2000; Street Perrott et al., 1997). However, stable carbon isotopes do not yield a distinction further than C3 vs. C4 plant metabolism and are blurred by plants with a Crassulacean Acid Metabolism (CAM) (Boom et al., 2001). Pollen analysis gives a more detailed indication of past vegetation change, but still does not distinguish beyond family or generic level in most cases. In addition, the spatial resolution of reconstructions through fossil pollen is limited to the availability of sedimentary archives, while within pollen the representation of vegetation is restricted to plant taxa dispersed by water and wind

(Hicks, 2006; Ortu et al., 2006). In addition, wind-blown dispersal limits the spatial resolution of vegetation reconstructions based on fossil pollen (Hicks, 2006; Moscol Olivera et al., 2009; Ortu et al., 2006).

The last decades have seen increasing interest in reconstructions of past montane vegetation compositions, including UFL positions, owing to the exceptional sensitivity of such vegetation to past and present climate change (Bakker et al., 2008; Beniston et al., 1997; Birks and Ammann, 2000). Unfortunately, it is in high altitude ecosystems, such as the Ecuadorian Andes, that the spatial uncertainty introduced by wind-blown distribution of pollen is the greatest (Bakker et al., 2008; Moore et al., 1991). An additional inaccuracy in the Ecuadorian Andes is the abundance of asteraceous pollen that cannot be distinguished beyond family level, and contain species occurring in montane cloud forest as well as in the páramo grasslands above the UFL (Bakker et al., 2008). Both problems complicate the interpretation of pollen-based vegetation reconstructions and have fueled the debate over the natural extension and composition of the forest in the Ecuadorian Andes during time periods predating the current massive human influence (Laegaard, 1992; Wille et al., 2002). This in turn hinders efforts to restore montane cloud forest in the area in the frame of Kyoto protocol driven activities to fix carbon dioxide. To resolve this issue, we explored a new biomarker application as proxy to use in conjunction with existing ones in a peat sequence in

* Corresponding author. Tel.: +31 20 525 7444.
E-mail address: B.Jansen@uva.nl (B. Jansen).

the Guandera Biological Reserve in what constitutes the last remaining stretch of forest of appreciable size within the Ecuadorian inter-Andean valley (Bakker et al., 2008).

The biomarker application is based on plant-specific concentration patterns of *n*-alkanes and *n*-alcohols with chain lengths of 20–36 carbon atoms that originate exclusively from the epicuticular wax layers on leaves and roots of terrestrial higher plants (e.g. Ficken et al., 1998; Ishiwatari et al., 2005; Kolattukudy et al., 1976; Rieley et al., 1991; Van Bergen et al., 1997). Several studies have shown these compounds to occur in plant-specific patterns and tested their potential use as vegetation records through organic matter preserved in peat or lake sediment sequences (e.g. Ficken et al., 1998; Hughen et al., 2004; Nott et al., 2000; Pancost et al., 2002). However, two hurdles have prevented unequivocal success to date. Firstly, there was a lack of databases of characteristic straight-chain lipid patterns in plants. We overcame this problem in a previous study where we analyzed the straight-chain lipid patterns present in the plant species responsible for the dominant biomass input in soils and peat sediments in the Ecuadorian study area (Jansen et al., 2006a). Specifically, we sampled the leaves and roots of 19 species responsible for the dominant biomass input into soils and peat deposits in the area, making sure to mix material from different specimens from the same species to overcome potential specimen related heterogeneity (Jansen et al., 2006a). We found the concentration patterns of particularly the *n*-alkanes and *n*-alcohols to be plant-species specific in most instances and thus show great potential for application as biomarkers (Jansen et al., 2006a). Concentration patterns are defined as unique concentration ratios of several *n*-alkanes and of several *n*-alcohols of different chain lengths. Additional research confirmed that the specific concentration patterns of at least the *n*-alkanes are well preserved in soils and peat sediments in the area for several millennia (Jansen and Nierop, 2009).

However, the second and most important challenge is that it is concentration patterns and not the individual *n*-alkanes or *n*-alcohols themselves that constitute a biomarker (Jansen et al., 2006a). Obviously, *n*-alkanes and *n*-alcohols from many different plants enter the soil or sediment deposits at the same time. Unravelling the resulting mixed signal into the original combination of plant-specific suites of *n*-alkane and *n*-alcohol concentration ratios is a large challenge. The simple observation of shifts in concentration ratios of a limited number of individual compounds may help to get an idea of drastic shifts in the occurrence of larger vegetation groups, such as a rapid deforestation (Jansen et al., 2008). However, to document more subtle changes in vegetation patterns a more sophisticated method is needed that is able to unravel the entire preserved mixed signal, and thus use all information stored therein. The purpose of the present study was to develop such a method in the form of the VEgetation Reconstruction with the Help of Inverse Modeling and Biomarkers (VERHIB) model. Subsequently, to apply it for the first time to unravel and interpret straight-chain lipid patterns from a peat deposit in the Ecuadorian setting that was previously analyzed for fossil pollen, non-pollen palynomorphs and plant macrofossils (Bakker et al., 2008).

2. Methods

2.1. Description of the study area

The study area is located in the Guandera Biological Reserve in northern Ecuador, close to the Colombian border. It lies on the inner flanks of the eastern Cordillera at approximately 11 km from the town of San Gabriel (Fig. 1). The coring site (G15) consists of a small peat bog of ca. 30 m in diameter located at an altitude of 3400 m.a.s.l. The site is situated at 200 m below the present-day UFL and some 100 m above the biological station of the Reserve (0°36′N/77°42′E in WGS 1984), within a *Clusia* dominated forest that is part of the upper montane rain forest.
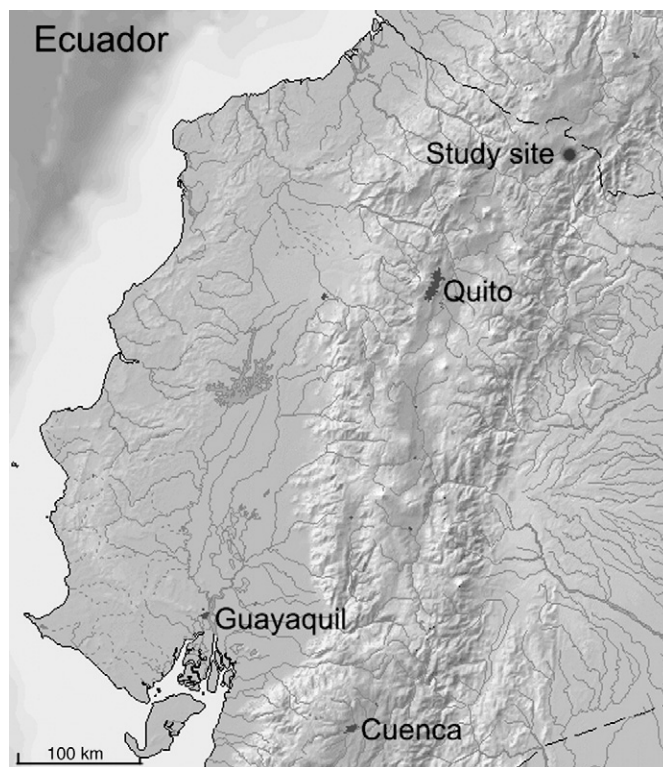

Fig. 1. Geographical map of Ecuador, indicating the location of the study area.

The study site currently has a humid tropical alpine climate with an annual precipitation of ca. 1700 mm. Strong diurnal temperature fluctuations range from 4° to 15 °C but annual temperature fluctuations are low (monthly means of maximum temperature vary between 12 and 15 °C) (Di Pasquale et al., 2008). Annual variations in temperature and precipitation are mainly forced by the annual migration of the Intertropical Convergence Zone (ITCZ).

For a detailed description of the present-day vegetation in the study area we refer to Moscol Olivera and Cleef (2009a, b). In summary, the Guandera Biological Reserve protects approximately 1000 ha of high altitude páramo grassland as well as areas of relatively undisturbed montane cloud forest. Most of this Andean forest is located between 3300 and 3640 m.a.s.l and consists of upper montane rain forest at lower altitudes, changing into a small band of sub-alpine rain forest found as dwarf forest at higher altitudes along the current UFL as well as in isolated patches above the UFL. Around an altitude of 3550 m isolated patches of páramo vegetation within the forest occur, whereas above 3640 m.a.s.l grass páramo dominates the landscape. Nevertheless, some sub-alpine rain forest patches occur up to 3700 m.a.s.l. The highest altitude in the study area is approximately 4100 m.a.s.l.

From a biomarker point of view, the most important taxa are those that are expected to be responsible for the dominant biomass input into peat and soil records in the area. These species were selected and identified during previous fieldwork in the Guandera Biological Reserve (Jansen et al., 2006a) and are presented in Table 1.

2.2. Sampling procedure

The sediment core used in the present study is core G15-II, which is one of two parallel cores taken at a vertical distance of 20 cm from one another from the deepest part of the basin (Bakker et al., 2008). Core G15-II was previously sampled, ^{14}C dated and analyzed for fossil pollen and non-pollen palynomorphs, while core G15-I was previously used exclusively for the analysis of plant macrofossils given the amount of material needed for that method (Bakker et al., 2008). The cores cover the time period of 7150 cal. yr. BP to present (Bakker

Table 1
Dominant biomass forming species identified in the Guandera Biological Station in the Eastern Cordillera.

Biotope[a]	Growth form	Family	Genus, species and identification
Upper montane rain forest	Evergreen tree[b]	Clusiaceae	*Clusia flaviflora* Engl.
	Epyphyte	Bromeliaceae	*Tilandsia* sp.2
	Fern	Blechnaceae	*Blechnum schomburgkii* (Klotzch) C. Chr.
Sub-alpine rain forest and	Shrub	Loranthaceae	*Gaiadendron punctatum* (Ruiz & Pav.) G. Don
extrazonal forest patches	Fern	Blechnaceae	*Blechnum schomburgkii* (Klotzsch) C. Chr.
	Evergreen tree[a]	Melastomataceae	*Miconia tinifolia* Naudin
	Evergreen tree[a]	Cunoniaceae	*Weinmannia cochensis* Hieron.
	Bamboo	Poaceae	*Neurolepis aristata* (Munro)Hitchc.
Grass páramo	Grass	Poaceae	*Calamagrostis effusa* (Kunth) Steud.
	Sedge	Cyperaceae	*Rhynchospora ruiziana* Boeck.
	Stem rosette	Asteraceae	*Espeletia pycnophylla* Cuatrec.
	Sedge	Cyperaceae	*Oreobolus goeppingeri* Suess.

[a] Species present in multiple biotopes were only sampled in one biotope.
[b] Trees never exceeded 10 m in height.

et al., 2008). For biomarker analyses, sub-samples of approximately 2 cm^3 were taken from the 90 cm long core G-15-II at alternating depth intervals of 2 or 3 cm throughout its entire depth. Care was taken to ensure the sample depths corresponded exactly with depths at which sub-samples for pollen analysis and non-pollen palynomorph analysis were previously taken from the same core. To avoid lipid contamination, hand contact of the sub-samples was carefully avoided during sampling and sample treatment. Upon sampling the sub-samples were freeze-dried and ground with a mortar and pestle and stored in clean glass vials awaiting extraction and analysis.

2.3. Extraction, clean-up and derivatization

Approximately 0.01 g of each of the ground sub-samples was subsequently extracted by Accelerated Solvent Extraction (ASE) using a Dionex 200 ASE extractor with 11 ml extraction cells and dichloromethane/methanol (DCM/MeOH) (93:7 v/v) as the extractant (Jansen et al., 2006b). Extractions were carried out at a temperature of 75 °C and a pressure of 17×10^6 Pa employing a heating phase of 5 min and a static extraction time of 20 min (Jansen et al., 2006b).

Upon extraction, the DCM/MeOH phase was rotary evaporated to complete dryness after which the dry extract was re-dissolved in approximately 2–5 ml DCM/2-propanol (2:1 v/v). Next, the extract was filtered using a Pasteur pipette packed with defatted cotton wool, 0.5 cm $Na_2SO_4(s)$ as a drying agent and 2 cm $SiO_2(s)$ to remove very polar constituents. To the filtered extracts, we added known amounts of an internal standard containing d_{42}-n-C_{20} alkane and d_{41}-n-C_{20} alcohol, after which we dried the extracts under a gentle stream of $N_2(g)$.

To the dried extracts we added 100 µl of cyclohexane and 50 µl of BSTFA (N,O-bis(trimethylsilyl) trifluoroacetamide) containing 1% TMCS (trimethylchlorosilane). Subsequently, the mixture was heated for 1 h at 70 °C to derivatize all free hydroxyl groups on the n-alcohols to their corresponding trimethylsilyl (TMS) ethers. After derivatization, the solutions were dried once more under $N_2(g)$ to remove the excess BSTFA, and subsequently re-dissolved in 200–1000 µl of cyclohexane depending on the extraction yields.

2.4. Quantification and identification of the n-alkanes and n-alcohols

Concentrations of n-alkanes and n-alcohols were determined on a ThermoQuest Trace GC 2000 gas chromatograph connected to a Finnigan Trace quadrupole mass spectrometer (MS). Separation took place by on-column injection of 1.0 µl of the derivatized extracts on a 30 m Rtx-5Sil MS column (Restek) with an internal diameter of 0.25 mm and film thickness of 0.1 µm, using He as a carrier gas. Temperature programming was: 50 °C (hold 2 min); 40 °C/min to 80 °C (hold 2 min); 20 °C/min to 130 °C; 4 °C/min to 350 °C (hold 10 min). Subsequent MS detection in full scan mode used a mass-to-

charge ratio (m/z) of 50–650 with a cycle time of 0.65 s and followed electron impact ionization (70 eV).

From the chromatograms, n-alkanes and n-alcohols were identified by their mass spectra and retention times. The dominant fragment ions (base peaks) were represented by: m/z 57 for the n-alkanes and m/z 75 for the n-alcohols. For absolute quantification, the base peak areas were compared to the peak areas from the corresponding base peak of the deuterated internal standard: m/z 66 for d_{42}-n-C_{20} alkane and m/z 76 for d_{41}-n-C_{20} alcohol. The variance in MS response to n-alkanes and n-alcohols with chain lengths 11–44, and to the deuterated standards was previously tested and found to be negligibly small (coefficients of variation between 3% and 6%) (Jansen et al., 2008). The variance introduced by the extraction, sample preparation and analysis procedure was also previously examined. The reproducibility of both the absolute concentrations of the individual n-alkanes and n-alcohols as well as the ratio of the various lipids within a component class was tested by comparison of replicates of soil samples as part of a previous study. The coefficients of variation were always ≤8% for the n-alkanes and ≤5% for the n-alcohols (Jansen et al., 2008).

2.5. Description of the VERHIB model

The VERHIB model consists of a linear regression model to describe the way in which a certain vegetation development over time at a certain location results in accumulation of biomarkers (in the present study n-alkanes and n-alcohols) in a suitable archive (in the present study the previously described peat sediment core). An inversion of the forward model is used to reconstruct palaeo-vegetation on the basis of the observed accumulated biomarker signal.

2.6. The forward model in VERHIB

In VERHIB the forward model uses annual plant biomass production per species as model input. The partitioning over plant leaf and root parts, the distribution of leaf and root parts over different depth layers, and the biomarker composition of the leaf and root parts are model parameters. VERHIB gives accumulated mass of biomarkers per pre-defined depth layer in the soil/peat/sediment record as output. Eq. (1) gives the discrete version of this model.

$$\begin{bmatrix} b_1(d,t) \\ b_2(d,t) \\ \vdots \\ b_I(d,t) \end{bmatrix} = \begin{bmatrix} lf_1(d)lc_{1,1} & \cdots & lf_J(d)lc_{1,J} & rf_1(d)rc_{1,1} & \cdots & rf_J(d)rc_{1,J} \\ lf_1(d)lc_{2,1} & \cdots & lf_J(d)lc_{2,J} & rf_1(d)rc_{2,1} & \cdots & rf_J(d)rc_{2,J} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ lf_1(d)lc_{I,1} & \cdots & lf_J(d)lc_{I,J} & rf_1(d)rc_{I,1} & \cdots & rf_J(d)rc_{I,J} \end{bmatrix} \begin{bmatrix} lm_1(t) \\ lm_2(t) \\ \vdots \\ lm_J(t) \\ rm_1(t) \\ rm_2(t) \\ \vdots \\ rm_J(t) \end{bmatrix}$$

(1)

where $b_i(d,t)$ is the mass of biomarker i that accumulates at depth interval d during time period t ($d=1,...,D$ and $t=1,...,T$); $lc_{i,j}$ is the concentration of biomarker i in the leaf of plant j, and $rc_{i,j}$ is the concentration of biomarker i in the root of plant j; $lf_j(d)$ is the fraction of the leaf mass of plant j that will accumulate in depth interval d (and $rf_j(d)$ refers to the same information for the root mass); $lm_j(t)$ is the leaf mass of plant j during time t, and $rm_j(t)$ is the root mass. Implicit in Eq. (1) is the independence of $lf_j(d)$, $lc_{i,j}$, $rf_j(d)$ and $rc_{i,j}$ from time. The overall partitioning of biomarkers over leaves and roots is fixed per plant species and applied as constraint to Eq. (1). Eq. (2a) defines the leaf fraction of a plant species j as the sum of the parameter $lf_j(d)$, and the root fraction of that species as the sum of the parameter $rf_j(d)$. Eq. (2b) specifies the fixed ratio that applies to each $lm_j(t)$ and $rm_j(t)$.

$$lf_j = \sum_{d=1}^{D} (lf_j(d))$$
$$rf_j = \sum_{d=1}^{D} (rf_j(d)) \tag{2a}$$

$$\begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 & \cdots & 0 & -lf_1/rf_1 & 0 & \cdots & 0 \\ 0 & 1 & & 0 & 0 & -lf_2/rf_2 & & 0 \\ \vdots & & \ddots & \vdots & \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & 1 & 0 & 0 & \cdots & -lf_J/rf_J \end{bmatrix} \begin{bmatrix} lm_1(t) \\ lm_2(t) \\ \vdots \\ lm_J(t) \\ rm_1(t) \\ rm_2(t) \\ \vdots \\ rm_J(t) \end{bmatrix} \tag{2b}$$

Another property of the vegetation is that the biomass of all plants (dry matter per unit area and time) cannot exceed a certain maximum. In addition, both leaf and root mass of each individual plant should be greater than or equal to zero (Eq. (3)).

$$\begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 0 \\ 0 \\ \vdots \\ 0 \\ tb(t) \end{bmatrix} \geq \begin{bmatrix} -1 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 \\ 0 & -1 & & 0 & 0 & 0 & & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & & -1 & 0 & 0 & & 0 \\ 0 & 0 & & 0 & -1 & 0 & & 0 \\ 0 & 0 & & 0 & 0 & -1 & & 0 \\ \vdots & \vdots & & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & & 0 & 0 & 0 & & -1 \\ 1 & 1 & \cdots & 1 & 1 & 1 & \cdots & 1 \end{bmatrix} \begin{bmatrix} lm_1(t) \\ lm_2(t) \\ \vdots \\ lm_J(t) \\ rm_1(t) \\ rm_2(t) \\ \vdots \\ rm_J(t) \end{bmatrix} \tag{3}$$

Here $tb(t)$ is the total biomass production $(\mathrm{g\,m^{-2}\,y^{-1}})$ during time period t. When defining the mass of a plant species j during time period t as the sum of leaf mass and root mass: $p_j(t) = lm_j(t) + rm_j(t)$, the temporal change of vegetation over time is described by the following autoregressive process:

$$p_j(t) = p_j(t+1) + v_j\, e(t) \tag{4a}$$

where time is defined backward so that $(t+1)$ refers to one time period prior to t; $e(t)$ is a zero mean and unit variance truncated Gaussian error such that $0 \leq p_j(t) \leq pmax$; v_j is a coefficient that describes the sensitivity of plant j to unknown external influences, i.e. climate-, succession- and/or anthropogenic-driven, and effectively scales the randomness in the autoregressive equation. A small value for v_j means that the system is deterministic, so that vegetation at t

correlates strongly with that found at $t+1$. The matrix notation for Eq. (4a) is as follows

$$\begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} = \begin{bmatrix} s1/v_1 & 0 & \cdots & 0 \\ 0 & s1/v_2 & & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & s1/v_J \end{bmatrix} \begin{bmatrix} 1 & 0 & \cdots & 0 & -1 & 0 & \cdots & 0 \\ 0 & 1 & & 0 & 0 & -1 & & 0 \\ \vdots & & \ddots & \vdots & \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & 1 & 0 & 0 & \cdots & -1 \end{bmatrix} \begin{bmatrix} p_1(t) \\ p_2(t) \\ \vdots \\ p_J(t) \\ p_1(t+1) \\ p_2(t+1) \\ \vdots \\ p_J(t+1) \end{bmatrix} \tag{4b}$$

where the value $1/v_j$ can now be interpreted as a parameter specifying the importance of autocorrelation for plant j. Furthermore, there is a relation between certain plant species, i.e. some plants are likely to occur together. This is described by Eq. (5a).

$$\frac{1}{m_n} \sum_{i=j}^{l} u_i = \frac{u_j}{m_n} p_j(t) + ... + \frac{u_l}{m_n} p_l(t) + w_n\, e(t) \tag{5a}$$

where m_n is the number of plant species that belong to group n, u_j is the relative presence (on a weight basis) of a plant species j in plant group n; $e(t)$ is a zero mean and unit variance Gaussian error and w_n is a coefficient to scale the random input for plant group n. A small value for w_n means that the system is deterministic.

And the matrix equation for this process is given by

$$s2 \begin{bmatrix} \bar{u}_1/w_1 \\ \bar{u}_2/w_2 \\ \vdots \\ \bar{u}_N/w_N \end{bmatrix} = \begin{bmatrix} s2/w_1 & 0 & \cdots & 0 \\ 0 & s2/w_2 & & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & s2/w_J \end{bmatrix} \begin{bmatrix} u_1/m_1 & u_2/m_1 & 0 & \cdots & 0 \\ 0 & u_2/m_2 & u_3/m_2 & \cdots & 0 \\ 0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 0 & \cdots & u_J/m_N \end{bmatrix}$$
$$\times \begin{bmatrix} p_1(t) \\ p_2(t) \\ p_3(t) \\ \vdots \\ p_J(t) \end{bmatrix} \tag{5b}$$

where the parameter $\bar{u}_n$ replaces the term $\frac{1}{m_n} \sum_{i=j}^{l} u_i$. The term $1/w_n$ is a weight that indicates the importance of a certain group (there are N vegetation groups in total). A large value of $1/w_n$ tells that if one species p_j is found in a certain time period, it is very likely that species belonging to that same vegetation group are also present. Based on exclusive occurrence of species, a larger value for this parameter can be chosen. In the system described by Eqs. (1)–(5a, 5b), the coefficients $lc_{i,j}$, $rc_{i,j}$, $lf_j(d)$ and $rf_j(d)$ are known constants that are tabulated. These values are established on the basis of measurements of current plant material (e.g. Jansen et al., 2006a). The coefficients $tb(t)$ are known from literature and biophysical calculations of primary production (Leuschner et al., 2007; Moser et al., 2007; Ramsay and Oxley, 2001; Soethe et al., 2006). The autocorrelation of species occurrence (the parameter v_n in Eq. (5a, 5b)), the strength of group associations (parameter w_n in Eq. (5a, 5b)) and relative abundance of individuals species within groups (the values u_j/m_n and $\bar{u}_n$ in Eq. (5a, 5b)) are known from literature and field observations (Jansen et al., 2006a). The parameters $s1$, and $s2$ are not known a-priori and derived by cross-validation as described below.

2.7. The inverse model in VERHIB

If the depth of a certain layer in an archive (e.g. sediment or soil) (d) and time period (t) where the archive at that depth was at the surface are closely related and this relation is known, Eqs. (1), (4b) and (5b) can be combined into one matrix equation containing all the depth intervals and time periods. This matrix equation to be solved has the form

$$\boldsymbol{b} = \mathbf{A}\boldsymbol{p} \tag{6}$$

where the vector $\boldsymbol{b}$ contains observed mass of the biomarkers per soil layer and some constants (the left-hand vectors of Eqs. (1), (4b) and (5b) in a different arrangement), the matrix $\mathbf{A}$ contains known constants (the matrices of Eqs. (1), (4b) and (5b), in a different arrangement) and the vector $\boldsymbol{p}$ contains the parameter values to be estimated.

The system would not have a unique solution (in the least-squares sense) on the basis of Eq. (1) alone, given a realistic number of plant species (>10), the number of biomarkers (<20) and typical observation resolution (5 layers) and observation errors. This problem is solved by application of Tikhonov regularization (Tikhonov and Arsenin, 1977). The coefficients $s1$ and $s2$ (both part of the vector $\boldsymbol{b}$) are parameters to specify the importance of the smoothness constraints and called regularization coefficients. Provided that the matrix $\mathbf{A}$ is of full rank due to regularization, the solution to Eq. (7) in a least squares sense is given by

$$p = \left(\mathbf{A}^T\mathbf{A}\right)^{-1}\mathbf{A}^T\boldsymbol{b} \tag{7}$$

To avoid unrealistic values in $\boldsymbol{p}$, equality constraints (the fixed leaf-root ratio per plant species, Eq. (2a, 2b)), as well as a set of inequality constraints (maximum primary production per unit area and time period, Eq. (3)) are added to Eq. (7):

Equality constraints: $\mathbf{d} = \mathbf{Cp}$ (8a)

Inequality constraints: $\mathbf{h} \geq \mathbf{Gp}$. (8b)

This least squares problem with non-negativity constraints is solved with the block principal pivoting algorithm that was implemented and tested previously (Portugal et al., 1994). Having a solution algorithm for the inverse problem, the only remaining problem is to select appropriate values for $s1$ and $s2$. We find the appropriate values by systematic variation of these two parameters within a specified domain, and finding minimum error in predicting biomarker accumulation (the vector $\boldsymbol{b}$ in previous equations) by leave-one-out cross-validation. This minimum prediction error is defined using the root mean square error (RMSE) as

$$RMSE_b(x,y) = \sqrt{\left(\sum_{i,d}(\hat{b}_{i,d}(s1_x,s2_y) - b_{i,d})^2 / (ID) \right)} \tag{9a}$$

$$RMSE_{b,\min} = \min_{x,y}[RMSE_b(x,y)] \tag{9b}$$

where $\hat{b}_{i,d}$ ($s1_x$, $s2_y$) refers to the estimated value of biomarker i at depth d while this observation was omitted from the data when estimating model parameters and when using values $s1_x$ and $s2_y$ as regularization parameters; $b_{i,d}$ is the observed value of biomarker i at depth d. In total (ID) combinations of I biomarkers and D depths are evaluated. The minimization $\min_{x,y}[\,]$ refers to the fact that those values of $s1_x$ and $s2_y$ are selected that lead to a minimum $RMSE$ value.

2.8. Testing the VERHIB model with artificial data

Prior to its first real-world application to reconstruct vegetation patterns based on biomarkers preserved in the previously described peat sediment core from the study area, we tested the robustness of the VERHIB model through a set of synthetic experiments.

In these synthetic experiments a forward model was specified and run to create a given set of input–output data. The input consisted of the mass of different plant species over time and the output of the biomarkers accumulated in several hypothetical layers of sediment/soil. To obtain a realistic, application-driven test, we used the biomarker patterns in the 19 selected plant species from the study

area (Table 1) as well as carbon accumulation rates previously established in the study area (Tonneijck et al. 2006). Using this data, the forward model was run 100 times, thus generating 100 input–output data sets.

In a second step, the model coefficients and/or the output data produced by running the forward model were perturbed, after which the inverse modeling procedure was applied to each of the 100 data sets. For different levels of perturbation the goodness of the reconstructed vegetation was established. Before reconstructing the vegetation data sets, first the appropriate values of the regularization parameters ($s1$ and $s2$) were selected via the objective function of Eq. (9b). This was done on the basis of a sample of 10 out of the 100 data sets. After selecting the optimal regularization parameters for a given perturbation, the inverse modeling procedure was applied to all 100 data sets. The applied perturbations consisted of: i) increasing levels of Gaussian error, ii) decreasing observation resolution over depth, and iii) the effect of omitting species. The applied levels of error and the specific resolutions are listed in Table 2.

We measured the goodness of the VERHIB model results after different levels and types of perturbation by: i) comparing the reconstructed vegetation with the artificially created vegetation ensemble that formed the basis of the synthetic data set, and ii) comparing the predicted biomarker data with the biomarker data in the synthetic data. In both cases we used an $RMSE$. For one particular data set k this is represented by:

$$RMSE_p(k) = \sqrt{\left(\sum_{j,t}(\hat{p}_{j,t} - p_{j,t})^2 / (JT) \right)} \tag{10}$$

$$RMSE_b(k) = \sqrt{\left(\sum_{i,d}(\hat{b}_{i,d} - b_{i,d})^2 / (ID) \right)}. \tag{11}$$

Here $\hat{p}_{j,t}$ refers to the predicted mass of plant species j at time period t, with J species and T time periods in total; and (analogous) $\hat{b}_{i,d}$ refers to the predicted mass of biomarker i in sediment/soil layer d, with I biomarkers and D sediment/soil layers in total. The 100 values for $RMSE_p(k)$ and $RMSE_b(k)$ (one pair for each data set) were averaged to $RMSE_p$ and $RMSE_b$.

2.9. Real-world application of the VERHIB model

The observed C_{20} – C_{36} n-alkane and n-alcohol signals at 2–3 cm intervals along the peat sediment core served as input for the first real-world application of the VERHIB model. In addition, the previously compiled database of n-alkanes and n-alcohol concentration patterns from the dominant plants (Table 1) was used (Jansen et al., 2006a). Root input was disabled in the model calculations because input from species other than the peat itself will not have been in the form of terrestrial plant roots. The resulting predicted temporal distribution of the 19 plant species were compared to the previously reconstructed vegetation changes from pollen data from the same sediment core (Bakker et al., 2008; Jansen and Nierop, 2009).

3. Results and discussion

3.1. The observed biomarker signal in the peat sediment core

The n-alkanes and n-alcohol signal in the peat sediment core with depth is provided in Tables 3 and 4 and displayed the expected characteristic higher terrestrial plant patterns with odd-over-even chain-length predominance for the n-alkanes and even-over-odd chain-length predominance for the n-alcohols (Kolattukudy et al., 1976). In addition, we noted an excellent preservation of straight-chain lipids

Table 2
Overview of the different perturbations applied to the results from the forward model.

Error imposed on biomarker data	
0%	No error is added to the biomarker data.
5%	Gaussian error is added to the biomarker data, with a standard deviation equaling 5% of the average biomarker values
10%	Gaussian error is added to the biomarker data, with a standard deviation equaling 10% of the average biomarker values
20%	Gaussian error is added the biomarker data, with a standard deviation equaling 20% of the average biomarker values
Observational resolution	
1/1	Inverse model is defined at same resolution of 5 cm as the forward model
1/2	Inverse model is defined at a resolution that is two times as coarse (instead of two layers of 5 cm, only one mixed layer of 10 cm is observed)
1/3	Inverse model is defined at a resolution that is three times as coarse (instead of three layers of 5 cm, only one mixed layer of 15 cm is observed)
1/5	Inverse model is defined at a resolution that is two times as coarse (instead of five layers of 5 cm, only one mixed layer of 25 cm is observed)
Species omission	
0	Including the same species in the inverse model as in the forward model
1	Omitting plant species 1 (out of 19) in the inverse model
2	Omitting plant species 2 (out of 19) in the inverse model
⋮	⋮
19	Omitting plant species 19 (out of 19) in the inverse model

All combinations of the perturbations are evaluated, leading to 30400 model runs in total (4 * biomaker errors * 4 resolution misspecifications * 19 species omissions * 100 synthetic data sets).

over the full length of the core that dates back to 7150 cal yr BP at 90 cm core depth (Bakker et al., 2008; Jansen and Nierop, 2009). The data in Tables 3 and 4 served as input for the VERHIB model.

3.2. The robustness of VERHIB as tested with artificial data

Table 5 presents the main observed model errors after testing the model with artificial data as described in paragraph 2.5, using the perturbations described in Table 2. The results show that that the inverse modeling procedure succeeded in recovering the original vegetation pattern. While the error incremented in a linear way, the resulting uncertainty was limited and the combined effect of the different perturbations was additive rather than multiplicative. As a result the VERHIB model appears to be robust with respect to reconstructing vegetation patterns based on imperfect data, as will be available in field studies.

Table 3
Measured concentrations of C_{20}–C_{36} *n*-alkanes with depth in the peat core in μg/g of dried peat material.

Depth (cm)	C_{20}	C_{21}	C_{22}	C_{23}	C_{24}	C_{25}	C_{26}	C_{27}	C_{28}	C_{29}	C_{30}	C_{31}	C_{32}	C_{33}	C_{34}	C_{35}	C_{36}
3	87.5	16.7	70.1	103.4	177.6	307.3	366.4	420.5	390.1	374.8	224.1	226.1	71.4	44.5	12.7	10.1	0.0
5	3.4	0.7	2.0	3.4	0.0	8.3	3.3	11.8	3.3	27.5	2.2	39.9	2.4	21.9	0.0	0.9	0.0
7	0.0	3.0	6.1	12.8	12.1	22.6	12.8	25.9	14.4	51.2	7.5	55.3	3.9	12.1	0.0	0.0	0.0
10	3.2	0.8	1.5	2.2	1.9	6.5	2.2	9.7	2.7	39.6	2.6	76.9	0.0	23.6	0.0	0.0	0.0
12	3.7	1.2	2.5	4.0	2.6	7.4	2.5	10.6	3.6	33.6	3.0	54.5	2.6	14.3	0.0	1.0	0.0
15	5.1	2.3	4.3	13.7	5.9	20.6	3.7	15.7	3.9	48.4	3.4	77.4	0.0	28.9	0.0	2.7	0.0
17	6.3	2.7	5.4	15.1	5.6	20.9	3.9	23.5	5.1	87.6	3.9	118.4	3.1	30.0	0.0	2.6	0.0
20	2.5	0.8	1.5	2.7	0.0	8.4	2.6	8.8	2.7	26.6	2.3	47.6	0.0	13.7	0.0	1.4	0.0
23	8.4	1.0	3.7	4.3	3.2	10.1	2.7	8.5	2.8	18.7	1.9	33.7	0.0	4.7	0.0	0.6	0.0
25	3.1	0.6	1.7	1.7	2.3	4.7	2.0	4.7	1.5	6.2	0.7	6.1	0.8	1.8	0.1	0.2	0.0
27	6.3	0.8	3.4	4.9	3.2	8.8	3.7	12.5	4.4	21.0	2.4	19.2	1.1	4.2	0.0	0.6	0.0
30	4.7	0.6	2.3	3.8	2.9	5.8	2.9	7.7	2.6	13.2	1.8	15.4	1.3	5.0	0.2	0.6	0.0
32	56.0	5.7	28.1	44.2	26.5	70.7	25.3	108.2	26.9	172.9	15.3	130.1	4.8	24.7	0.0	0.0	0.0
35	3.6	0.5	1.7	2.6	2.3	5.1	2.5	7.4	2.2	11.6	1.0	6.7	0.4	1.3	0.0	0.0	0.0
37	5.4	0.1	3.2	5.4	3.3	7.6	2.0	12.5	3.0	20.0	1.7	14.2	0.5	2.9	0.0	0.0	0.0
40	3.5	0.0	1.8	3.1	3.3	5.3	1.5	6.5	1.4	9.8	0.8	8.4	0.0	0.0	0.0	0.0	0.0
42	6.4	0.8	3.1	4.4	3.0	7.1	3.0	11.7	3.2	18.3	1.8	12.4	0.9	2.6	0.0	0.6	0.0
45	2.4	0.4	1.2	2.1	2.1	4.5	1.3	7.3	2.2	11.9	1.1	8.9	0.6	2.8	0.0	0.2	0.0
47	5.9	0.5	2.8	2.4	2.3	4.0	2.2	6.2	2.1	9.8	1.1	5.6	0.5	1.0	0.0	0.0	0.0
50	3.9	0.7	2.4	3.7	4.5	6.6	3.6	9.1	3.3	13.2	1.5	8.4	0.6	1.8	0.0	0.1	0.0
52	8.7	1.6	5.5	8.0	5.4	12.8	5.4	21.9	5.7	36.9	3.2	24.5	1.2	5.2	0.0	0.8	0.0
55	4.4	0.8	2.9	5.9	8.9	13.4	9.1	16.4	7.4	21.8	3.7	13.0	1.2	3.1	0.4	0.3	0.0
57	49.2	0.0	25.6	29.7	20.2	46.7	20.9	86.1	2.7	149.9	12.2	88.7	3.9	17.1	0.0	0.0	0.0
60	3.5	0.8	2.7	6.0	9.9	14.1	12.1	16.6	10.0	17.9	5.4	10.7	1.9	2.8	0.4	0.3	0.0
62	4.5	0.8	2.4	4.1	2.2	6.7	2.0	10.0	2.4	17.8	1.1	10.4	0.4	2.1	0.0	0.3	0.0
65	8.2	1.0	3.7	6.8	6.4	15.8	9.2	28.9	11.2	48.6	5.7	32.5	2.2	7.6	0.0	0.0	0.0
67	5.8	0.9	3.2	5.1	2.8	7.7	2.5	11.0	2.7	18.6	1.4	11.4	0.5	2.5	0.0	0.5	0.0
70	3.9	0.0	2.5	4.2	2.0	8.0	3.0	14.0	3.8	24.0	1.2	16.2	1.1	3.7	0.0	0.0	0.0
72	9.0	1.7	5.4	8.8	5.1	12.2	4.4	18.7	6.4	32.6	3.2	21.5	1.4	5.1	0.0	0.6	0.0
75	10.9	1.8	6.0	6.6	6.1	9.2	4.2	14.7	5.6	26.5	2.5	17.1	2.6	5.0	0.0	0.7	0.0
77	12.0	1.9	7.1	9.5	5.8	13.7	4.3	18.3	5.0	31.0	2.5	22.2	1.1	5.4	0.0	1.1	0.0
80	5.6	0.0	1.5	2.5	1.7	4.7	1.3	7.4	1.9	13.7	0.7	9.9	0.5	2.1	0.0	0.1	0.0
82	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
85	10.5	3.0	5.5	5.0	4.4	8.8	3.6	15.9	5.1	29.6	2.4	23.4	2.0	6.2	0.0	0.0	0.0
87	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
89	11.8	2.6	5.8	6.2	4.4	11.9	6.4	9.8	7.8	23.9	3.9	21.6	1.7	6.1	0.0	0.0	0.0

3.3. Overall vegetation change in the study area reconstructed with VERHIB

The most likely relative species composition with time calculated by the VERHIB model using a database of the previously established biomarkers (Jansen et al., 2006a), henceforth called 'biomarker analysis', was grouped into taxa characteristic of forest and of páramo grassland (Fig. 2). A similar temporal plot was prepared for the species composition derived from the previously performed pollen analysis (Fig. 3) (Bakker et al., 2008).

The susceptibility of pollen in the study area for wind-blown dispersal of all except arboreal pollen was confirmed by analysis of the modern pollen rain at the coring site and within the surrounding forest (Moscol Olivera et al., 2009). Based on this, the substantial páramo contribution throughout the pollen record (Fig. 3) was largely attributed to regional wind-blown pollen entering the forest (Moore et al., 1991; Moscol Olivera et al., 2009). Only within the core intervals of 89–83 cm and the lower part of 29–5 cm was the pollen record interpreted as yielding an UFL position below the coring site, with the UFL shifting upwards past the coring site in the upper part of the latter interval (Bakker et al., 2008). In all other intervals, the coring site was interpreted as being located within the forest, albeit at times close to the UFL (Bakker et al., 2008).

Biomarker analysis (Fig. 2) indicated larger fluctuations and higher overall proportions of forest cover than pollen analysis (Fig. 3), consistent with a more local image due to a much smaller influence of wind-blown dispersal. The general trend of altitudinal shifts of forest/páramo transition over time was remarkably similar to that from pollen analysis (Figs. 2 and 3), including an increased influence of páramo in the lower part of the 29–5 cm core interval, changing back

to forest dominance in the upper part. An important fact, as it offers independent support for the tentative conclusion from pollen analysis that the UFL in the area was not significantly depressed by human interference within the last few centuries (Bakker et al., 2008).

Apart from some single-point discrepancies (e.g. at 25 cm) the only significant difference in trends between biomarker analysis and pollen analysis was the 89–83 cm interval where biomarker analysis

Table 5
Summary of the resulting model errors upon perturbing one factor at a time (e.g. when adding 5% error to the biomarker data while the observation resolution is 1/1 and species omission is set to 0).

	$RMSE_p^a$	$RMSE_b^b$
No errors	0.04	12.30
Error imposed on biomarker data		
5%	0.13	12.45
10%	0.17	10.32
20%	0.53	13.14
Observation resolution		
1/2	0.15	12.41
1/3	0.31	21.23
1/5	0.75	23.53
Species omission		
Average for all species	0.32	19.29

[a] $RMSE_p$ is the root mean squared error of the predicted plant species in comparison to the species composition in the synthetic data.
[b] $RMSE_b$ is the root mean squared error of the predicted biomarker composition in comparison to the biomarker composition of the sediment/soil in the synthetic data (see Eqs. (10) and (11) in the text).

Table 4
Measured concentrations of C_{20}–C_{36} n-alcohols with depth in the peat core in µg/g of dried peat material.

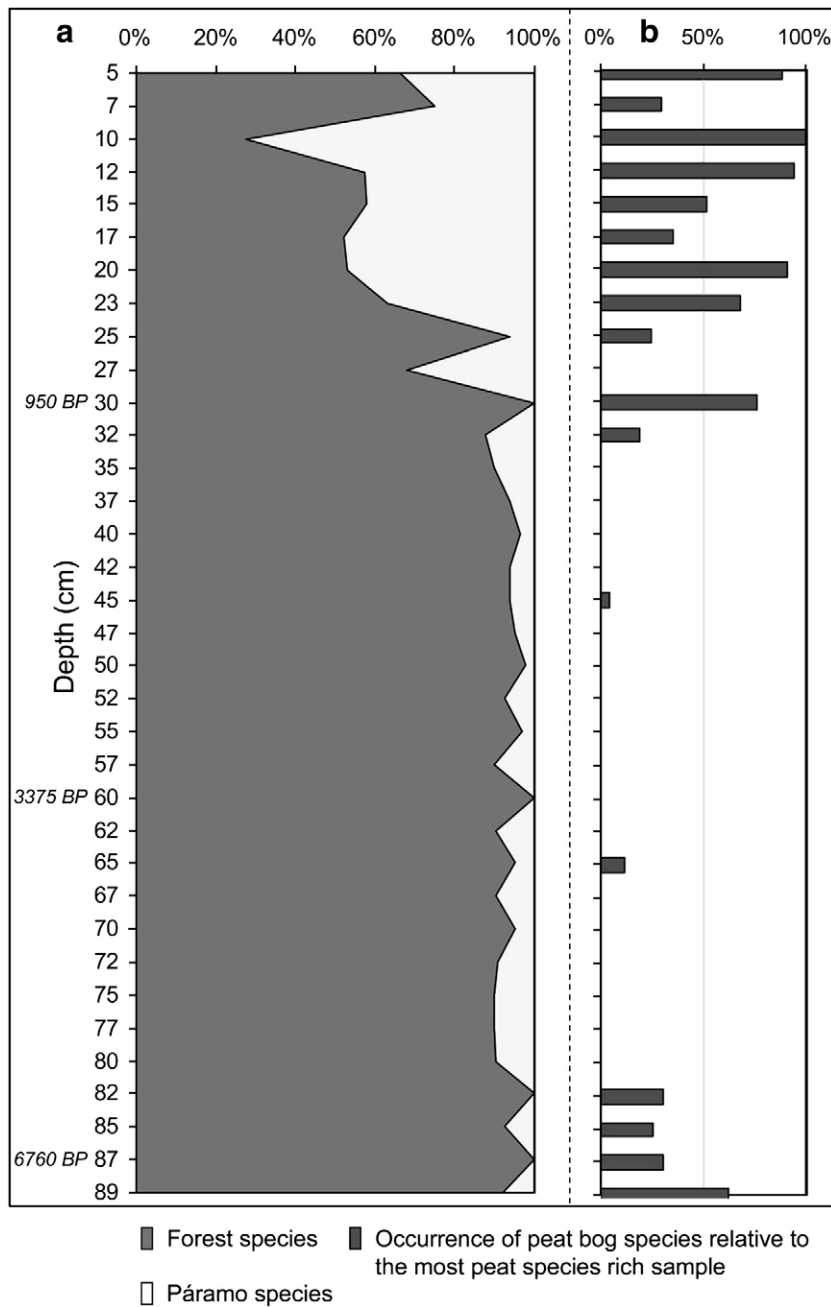| Depth (cm)	C_{20}	C_{21}	C_{22}	C_{23}	C_{24}	C_{25}	C_{26}	C_{27}	C_{28}	C_{29}	C_{30}	C_{31}	C_{32}	C_{33}	C_{34}	C_{35}	C_{36}
3	62.0	0.0	87.9	26.5	155.8	20.2	179.2	28.1	186.1	0.0	177.5	0.0	139.9	0.0	33.6	0.0	0.0
5	0.0	0.0	19.4	1.6	17.5	2.0	19.4	1.4	18.4	1.8	9.4	0.0	2.0	0.0	0.4	0.0	0.0
7	3.0	1.4	23.4	2.4	24.6	2.6	18.8	1.7	12.6	2.0	9.0	0.0	2.6	0.0	0.0	0.0	0.0
10	1.7	0.0	16.1	0.9	13.4	1.3	11.5	0.9	8.3	1.9	7.7	0.0	9.8	0.0	0.0	0.0	0.0
12	2.6	0.5	21.1	1.4	17.1	2.0	12.5	1.2	7.9	1.6	6.1	0.0	3.1	0.0	0.0	0.0	0.0
15	3.0	0.0	27.9	3.6	77.4	5.9	68.7	3.0	28.1	3.2	10.6	0.0	7.5	0.0	0.0	0.0	0.0
17	5.3	2.0	42.2	4.4	83.6	6.1	73.6	3.4	34.0	3.8	14.1	0.0	5.3	0.0	0.0	0.0	0.0
20	2.4	0.0	15.5	0.6	13.2	1.1	12.4	0.7	7.2	1.6	6.0	0.0	2.4	0.0	0.0	0.0	0.0
23	2.7	0.3	14.1	0.9	12.8	1.4	14.6	1.1	5.4	0.7	5.7	0.0	1.1	0.0	0.0	0.0	0.0
25	1.7	0.3	8.0	0.5	7.5	0.6	4.6	0.0	2.2	0.2	1.4	0.0	0.0	0.0	0.0	0.0	0.0
27	25.6	6.2	134.5	9.9	157.4	13.1	115.4	8.8	68.2	7.0	46.4	0.0	8.2	0.0	0.0	0.0	0.0
30	1.5	0.2	8.1	0.6	8.7	0.8	0.7	0.6	4.5	0.6	3.5	0.2	1.0	0.0	0.0	0.0	0.0
32	3.1	0.8	15.5	1.2	18.3	1.3	11.2	0.9	6.5	0.9	4.1	0.2	1.0	0.0	0.0	0.0	0.0
35	1.5	0.0	10.8	0.8	15.3	0.9	10.0	0.7	5.7	0.4	2.4	0.0	0.4	0.0	0.0	0.0	0.0
37	2.8	0.4	14.1	1.2	17.8	1.5	12.1	1.3	8.6	0.9	6.0	0.3	0.1	0.0	0.0	0.0	0.0
40	0.0	0.0	7.5	0.0	7.3	0.0	4.2	0.0	2.4	1.0	1.3	0.0	0.0	0.0	0.0	0.0	0.0
42	3.1	0.5	16.1	1.3	18.9	1.5	11.8	1.2	7.5	0.9	4.8	0.0	0.7	0.0	0.0	0.0	0.0
45	1.8	0.0	8.0	0.5	8.9	0.5	5.4	0.4	3.1	0.3	1.8	0.0	0.3	0.0	0.0	0.0	0.0
47	0.2	0.0	8.3	0.6	9.6	0.7	5.5	0.5	3.2	0.7	1.9	0.0	0.5	0.0	0.0	0.0	0.0
50	1.3	0.3	8.5	0.6	10.9	0.7	6.5	0.6	3.9	0.5	2.5	0.0	0.5	0.0	0.0	0.0	0.0
52	5.1	0.8	26.9	2.4	34.5	3.1	23.7	2.6	17.0	1.8	11.2	1.0	2.1	0.0	0.0	0.0	0.0
55	3.1	0.5	17.0	1.3	22.1	1.3	13.5	1.1	8.2	0.8	4.6	0.0	1.1	0.0	0.0	0.0	0.0
57	19.4	3.0	122.6	9.9	154.3	12.4	105.1	9.8	67.3	6.1	36.0	0.0	6.1	0.0	0.0	0.0	0.0
60	2.0	0.5	10.9	0.8	13.8	0.8	8.2	0.6	4.5	0.4	2.4	0.0	0.8	0.0	0.0	0.0	0.0
62	3.1	0.5	17.8	1.5	22.9	1.8	15.1	1.4	10.0	1.0	5.2	0.0	0.9	0.0	0.2	0.0	0.0
65	1.8	0.0	11.7	0.7	14.1	1.2	8.9	1.2	5.6	0.0	3.6	0.0	0.9	0.0	0.0	0.0	0.0
67	3.2	0.5	18.6	1.6	23.9	1.8	16.1	1.4	9.9	0.9	4.9	0.0	0.8	0.0	0.0	0.0	0.0
70	2.3	0.0	14.4	1.0	16.4	1.4	9.4	0.0	4.9	0.0	2.1	0.0	0.0	0.0	0.0	0.0	0.0
72	4.7	0.7	26.6	2.5	37.1	3.3	25.0	2.5	16.5	1.6	8.9	0.0	1.5	0.0	0.0	0.0	0.0
75	4.3	0.0	22.4	1.4	28.2	1.8	19.8	1.2	12.1	0.9	7.0	0.0	1.0	0.0	0.0	0.0	0.0
77	5.3	0.7	32.2	2.8	43.4	3.3	30.1	2.5	20.7	1.8	13.5	0.9	1.7	0.0	0.2	0.0	0.0
80	1.1	0.0	9.1	0.8	11.7	1.0	7.5	0.8	4.1	0.7	2.0	0.0	0.0	0.0	0.0	0.0	0.0
82	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
85	1.7	0.0	15.5	1.4	15.7	1.5	11.6	1.2	7.4	0.0	6.0	0.0	0.0	0.0	0.0	0.0	0.0
87	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
89	1.6	0.0	11.0	1.2	7.2	0.0	5.3	0.0	3.6	0.0	4.3	0.0	0.0	0.0	0.0	0.0	0.0

Fig. 2. Down-core proportions (%) of forest and páramo vegetation in a peat core from the northern Ecuadorian Andes according to biomarker analysis. (a) Temporal changes in the proportions of the total of forest species (dark) and páramo species (light) with selected ages indicated in cal. yr BP. (b) Temporal changes in the presence of peat forming species relative to the sample most rich in peat species (sample at 10 cm = 100%).

indicated presence of forest vegetation, as well as a relatively large input of peat bog taxa (Figs. 2 and 3). Previous cluster analysis revealed much greater similarities between the biomarker signal of páramo and peat bog species, than between páramo and forest species (Jansen et al., 2006a). Therefore, most likely some páramo species were dismissed as peat bog contributions by the model, leading the VERHIB model to underestimate the páramo extension in this oldest part of the core.

3.4. Detailed interpretation of changes in forest species composition

To take a more detailed look at vegetation change during the period of time covered by the core, we examined the most likely individual forest species composition along the core as calculated by the VERHIB model without grouping the results as in the previous paragraph

(Fig. 4). Combining Figs. 2 and 4 allowed for a subdivision of the peat sequence into intervals that turn out being very similar to the ones previously identified through pollen analysis (Bakker et al., 2008).

From 89–83 cm biomarker analysis indicated the forest consisted predominantly of the species *Miconia tinifolia*, *Blechnum schomburgkii* and *Neurolepis aristata* (Fig. 4) that occur mainly at the UFL or even in the lowermost páramo (Bakker et al., 2008). This is consistent with pollen analysis indicating the site to have been in the páramo at this time but close to the UFL (Bakker et al., 2008), and supports the assumption that páramo species influence was underestimated by the VERHIB model in this interval.

From 83–72 cm the abundance of the species *Hedyosmum cumbalense*, found in the lower integral upper montaine rain forest, drastically increased according to biomarker analysis (Fig. 4),
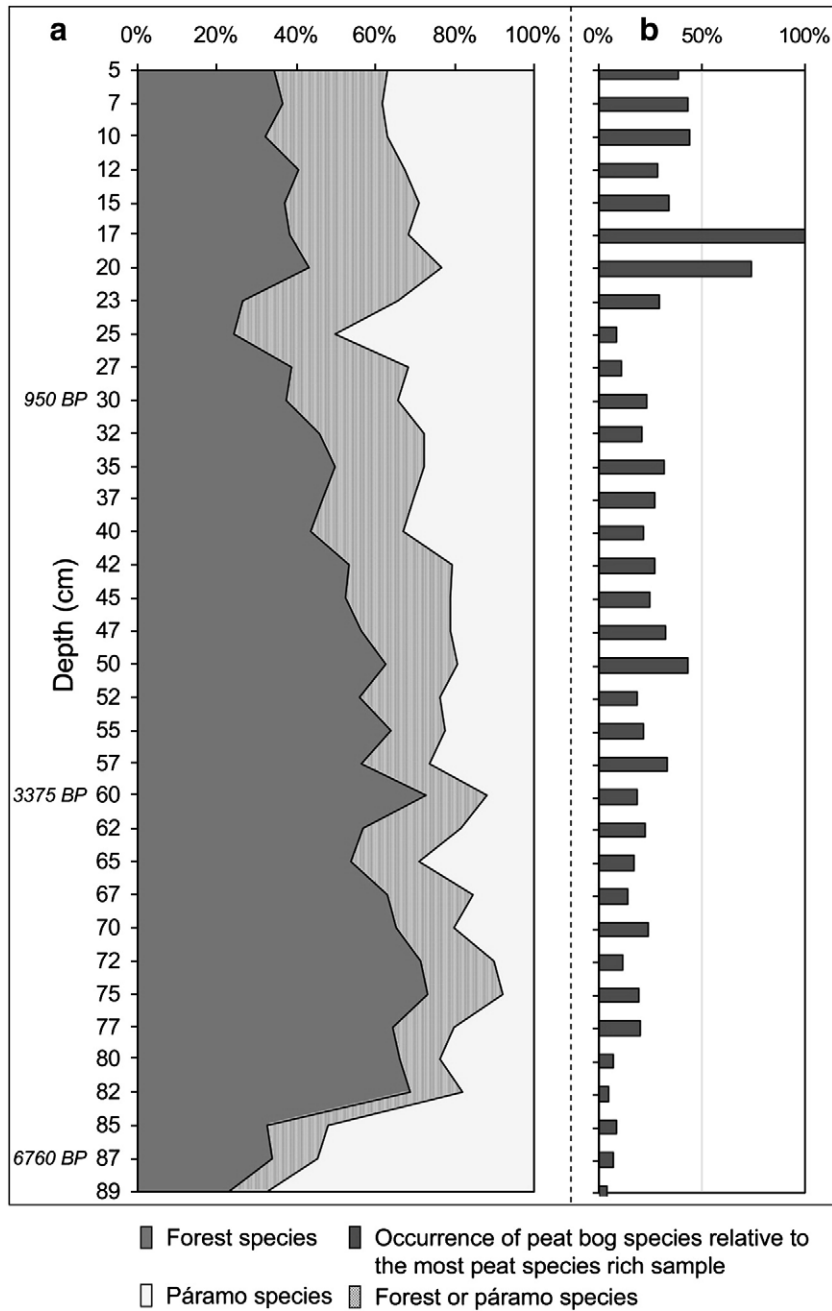
Fig. 3. Down-core proportions (%) of forest and páramo vegetation in a peat core from the northern Ecuadorian Andes according to fossil pollen analysis. (a) Previously determined percentage of the total of forest species (dark), páramo species (light) and species from the Asteraceae family that may belong to either forest or páramo vegetation (chequered) with selected ages indicated in cal. yr BP (Bakker et al., 2008). (b) Previously determined down-core occurrence of peat forming species relative to the most peat species-rich sample (sample at 17 cm = 100%) (Bakker et al., 2008).

indicating an upslope shift of the UFL. Pollen analysis found a similar shift and observed an increase of similar magnitude of the contribution of *Hedyosmum* trees, without identifying the species.

Biomarker analysis indicated the core interval 72–55 cm was marked by three episodes of intrusion of *Blechnum schomburgkii* (Fig. 4), nowadays found in patches with open páramo vegetation within the integral forest (Bakker et al., 2008). Thereby, the more local image of biomarker analysis supports the tentative conclusions from pollen analysis that such open patches occurred close to the coring site during this interval (Bakker et al., 2008).

From 55–35 cm biomarker analysis indicated a gradual decline in total forest representation (Fig. 2), and within the forest composition

a decline in contribution of *H. cumbalense* and initially an increase in the abundance of UFL species *Miconia tinifolia* (Fig. 4). Similarly, pollen analysis showed a decline of *Hedyosmum* and an increasing representation of melastomataceous pollen, speculated to represent *Miconia* (Bakker et al., 2008).

From 35–23 cm biomarker analysis showed a prominent contribution of first *Blechnum schomburgkii* to the forest, then changing to *Neurolepis aristata* and *Gaiadendron punctatum*, and finally *Miconia tinifolia* was prominent in the forest. This, together with the earlier mentioned general decline of total contribution of forest species (Fig. 2) indicate the UFL shifting downslope towards the coring site, in agreement with the pollen results (Bakker et al., 2008).
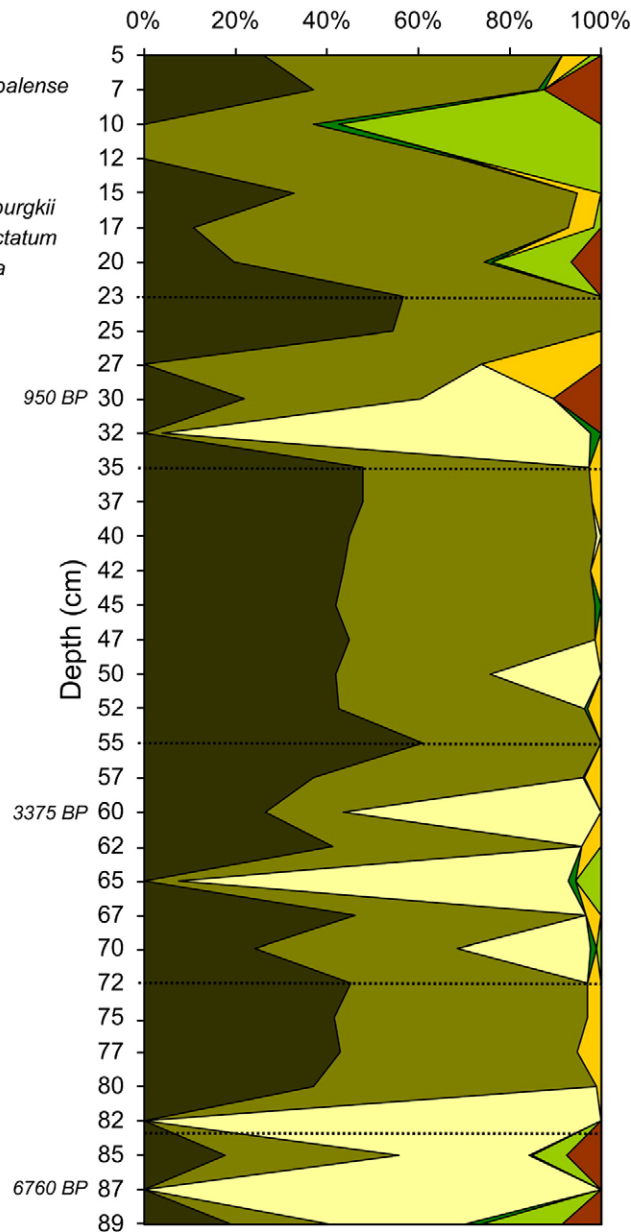
Fig. 4. Down-core proportions (%) of forest species in a peat sequence from the northern Ecuadorian Andes according to biomarker analysis. The following taxa were identified: *Hedyosmum cumbalense, Miconia tinifolia, Gynoxys buxifolia, Clusia flaviflora, Blechnum Schomburgkii, Gaiadendron punctatum* and *Neurolepis aristata*. Selected ages are indicated in cal. yr BP.

From 23–5 cm the most important observation from the forest species composition through biomarker analysis is the significant contribution of *Clusia flaviflora* to the forest for the first time in the record. This species currently grows in the vicinity of the coring site, but was virtually absent from the pollen record as well as the modern pollen rain because it produces little pollen (Bakker et al., 2008). Biomarker analysis gives us an estimate of the timing of its appearance near the coring site that proved beyond the possibilities of pollen analysis.

The previous comparisons between the results from pollen analysis and biomarker analysis using the VERHIB model showed the great value of the latter for vegetation reconstructions in the Ecuadorian study area, and the potential of applying a combination of both methods for vegetation reconstructions in other areas. However, one should note that the possibility of applying biomarker analyses in other areas with different vegetation depends also on the quality of

the biomarkers present in the vegetation there. More research aimed at identifying biomarkers, i.e. concentration patterns of *n*-alkanes and *n*-alcohols, in plant species worldwide is needed for a broader application of the method. Such an assessment should take into account possible heterogeneity of biomarkers over different specimen of the same species growing for instance in different regions under different growth conditions (e.g. Vogts et al., 2007).

4. Conclusions

From its first application, we conclude that biomarker analysis in peat sequences using the VERHIB model shows great potential to develop into a fully fledged, independent proxy for past vegetation dynamics. Potential strengths are: i) provision of a local reconstruction of past vegetation, enabling a high spatial resolution when more records are used, and ii) potential identification of past vegetation at

the species level. Potential weaknesses are: i) difficulty to identify some species, such as *Calamagrostis effusa* and *Rhynchospora ruiziana*, with very similar biomarker compositions in their leaves and/or roots, and ii) the still limited species database as a result of the novelty of the method, including a lack of information about the consistency of biomarkers in the same plant species growing in different regions and under different growth conditions. The greatest strength of biomarker analysis lies in it providing information complementary to that obtained by the existing method of pollen analysis. The local vegetation reconstruction through biomarker analysis supplements the regional vegetation reconstruction through pollen analysis that reflects a larger area. Furthermore, biomarker analysis helps to identify species from species-rich families present in more than one plant community, such as the Asteraceae. The latter group of species cannot adequately be further identified by pollen analysis. Similarly, pollen analysis helps pin-point interference by species with similar biomarker signals and identify taxa not (yet) included in the biomarker database. The implications for vegetation reconstructions in general, and in particular reconstructions in montane areas are far reaching. Applying biomarker analysis together with pollen analysis in an altitudinal transect will allow a reconstruction of UFL dynamics and temporal changes in taxonomic composition of ecotone forest with unprecedented detail. This will yield crucial information to calibrate and test models reconstructing the effects of past or present climate change on the altitudinal distribution of vegetation. In addition, it will provide more detailed insight into impact of human interference on forest distribution, helping direct sustainable replanting efforts.

Acknowledgements

References

Bakker, J., Moscol, M., Hooghiemstra, H., 2008. Holocene environmental change at the upper forest line in northern Ecuador. The Holocene 18, 877–893.

Beniston, M., Diaz, H.F., Bradley, R.S., 1997. Climatic change at high elevation sites: An overview. Climatic Change 36, 233–251.

Birks, H.H., Ammann, B., 2000. Two terrestrial records of rapid climatic change during the glacial-Holocene transition (14, 000–9, 000 calendar years BP) from Europe. Proceedings Of The National Academy Of Sciences Of The United States Of America 97, 1390–1394.

Boom, A., Marchant, R., Hooghiemstra, H., Sinninghe Damsté, J.S., 2001. CO_2- and temperature-controlled altitudinal shifts of C_4- and C_3-dominated grasslands allow reconstruction of palaeoatmospheric pCO2. Palaeogeography, Palaeoclimatology, Palaeoecology 177, 151–168.

Clark, J.S., McLachlan, J.S., 2003. Stability of forest biodiversity. Nature 423, 635–638.

Danby, R.K., Hik, D.S., 2007. Variability, contingency and rapid change in recent subarctic alpine tree line dynamics. Journal Of Ecology 95, 352–363.

Di Pasquale, G., Marziano, M., Impagliazzo, S., Lubritto, C., De Natale, S., Bader, M.Y., 2008. The Holocene treeline in the northern Andes (Ecuador): first evidence from soil charcoal. Palaeogeography, Palaeoclimatology, Palaeoecology 259, 17–34.

Ficken, K.J., Barber, K.E., Eglinton, G., 1998. Lipid biomarker, $\delta^{13}C$ and plant macrofossil stratigraphy of a Scottish montane peat bog over the last two millenia. Organic Geochemistry 28, 217–237.

Hicks, S., 2006. When no pollen does not mean no trees. Vegetation History and Archaeobotany 15, 253–261.

Hughen, K.A., Eglinton, T.I., Xu, L., Makou, M., 2004. Abrupt tropical vegetation response to rapid climate changes. Science 304, 1955–1959.

Ishiwatari, I., Yamamoto, S., Uemura, H., 2005. Lipid and lignin/cutin compounds in Lake Baikal sediments over the last 37 kyr: implications for glacial–interglacial palaeoenvironmental change. Organic Geochemistry 36, 327–347.

Jansen, B., Nierop, K.G.J., 2009. Me-ketones in high altitude Ecuadorian andisols confirm excellent conservation of plant-specific n-alkane patterns. Organic Geochemistry 40, 61–69.

Jansen, B., Nierop, K.G.J., Hageman, J.A., Cleef, A., Verstraten, J.M., 2006a. The straight-chain lipid biomarker composition of plant species responsible for the dominant biomass production along two altitudinal transects in the Ecuadorian Andes. Organic Geochemistry 37, 1514–1536.

Jansen, B., Nierop, K.G.J., Kotte, M.C., De Voogt, P., Verstraten, J.M., 2006b. The application of Accelerated Solvent Extraction (ASE) to extract lipid biomarkers from soils. Applied Geochemistry 21, 1006–1015.

Jansen, B., Haussmann, N.S., Tonneijck, F.H., Verstraten, J.M., De Voogt, P., 2008. Characteristic straight-chain lipid ratios as a quick method to assess past forest – páramo transitions in the Ecuadorian Andes. Palaeogeography Palaeoclimatology Palaeoecology 262, 129–139.

Jolly, D., Haxeltine, A., 1997. Effect of low glacial atmospheric CO_2 on tropical African montane vegetation. Science 276, 786–788.

Kolattukudy, P.E., Croteau, R., Buckner, J.S., 1976. Biochemistry of plant waxes. In: Kolattukudy, P.E. (Ed.), Chemistry and biochemistry of natural waxes. Elsevier, Amsterdam, pp. 289–347.

Kullman, L., 2007. Tree line population monitoring of Pinus sylvestris in the Swedish Scandes, 1973–2005: implications for tree line theory and climate change ecology. Journal Of Ecology 95, 41–52.

Laegaard, S., 1992. Influence of fire in the grass páramo vegetation of Ecuador. In: Balslev, H., Luteyn, J.L. (Eds.), Páramo, an Andean ecosystem under human influence. Academic Press, London, pp. 151–170.

Leuschner, C., Moser, G., Bertsch, C., Roderstein, M., Hertel, D., 2007. Large altitudinal increase in tree root/shoot ratio in tropical mountain forests of Ecuador. Basic and Applied Ecology 8, 219–230.

Mackay, A., Battarbee, R., Birks, J., Oldfield, F. (Eds.), 2003. Global Change in the Holocene. Arnold, London. 528 pp.

Mayle, F.E., Burbridge, R., Killeen, T.J., 2000. Millennial-scale dynamics of southern Amazonian rain forests. Science 290, 2291–2294.

Moore, P.D., Webb, J.A., Collinson, M.E., 1991. Pollen analysis. Blackwell, Oxford. 216 pp.

Moscol Olivera, M.C., Cleef, A.M., 2009a. Vegetation composition and altitudinal distribution of montane rain forests in northern Ecuador. Phytocoenologia 39, 175–204.

Moscol Olivera, M.C., Cleef, A.M., 2009b. A phytocoenological study of the páramo along two altitudinal transects in El Carchi province, northern Ecuador. Phytocoenologia 39, 79–107.

Moscol Olivera, M.C., Duivenvoorden, J.F., Hooghiemstra, H., 2009. Pollen rain and pollen representation across a forest-páramo ecotone in northern Ecuador. Review of Palaeobotany and Palynology (submitted).

Moser, G., Hertel, D., Leuschner, C., 2007. Altitudinal change in LAI and stand leaf biomass in tropical montane forests: a transect study in Ecuador and a pan-tropical meta-analysis. Ecosystems 10, 924–935.

Nott, C.J., Xie, S.C., Avsejs, L.A., Maddy, D., Chambers, F.M., Evershed, R.P., 2000. n-Alkane distributions in ombrotrophic mires as indicators of vegetation change related to climatic variation. Organic Geochemistry 31, 231–235.

Ortu, E., Brewer, S., Peyron, O., 2006. Pollen-inferred palaeoclimate reconstructions in mountain areas: problems and perspectives. Journal Of Quaternary Science 21, 615–627.

Pancost, R.D., Baas, M., van Geel, B., Damste, J.S.S., 2002. Biomarkers as proxies for plant inputs to peats: an example from a sub-boreal ombrotrophic bog. Organic Geochemistry 33, 675–690.

Pope, K.O., Pohl, M.E.D., Jones, J.G., Lentz, D.L., Von Nagy, C., Vega, F.J., Quitmyer, I.R., 2001. Origin and environmental setting of ancient agriculture in the lowlands of Mesoamerica. Science 292, 1370–1373.

Portugal, L.F., Judice, J.J., Vicente, L.N., 1994. A comparison of block pivoting and interior-point algorithms for linear least-squares problems with nonnegative variables. Mathematics of Computation 63, 625–643.

Ramsay, P.M., Oxley, E.R.B., 2001. An assessment of aboveground net primary productivity in Andean grasslands of central Ecuador. Mountain Research and Development 21, 161–167.

Rieley, G., Collier, R.J., Jones, D.M., Eglinton, G., 1991. The biogeochemistry of Ellesmere Lake, U.K. -I: source correlation of leaf wax inputs to the sedimentary lipid record. Organic Geochemistry 17, 901–912.

Roberts, N., 1998. The Holocene; An environmental history. Blackwell, Oxford. 316 pp.

Soethe, N., Lehmann, J., Engels, C., 2006. The vertical pattern of rooting and nutrient uptake at different altitudes of a south Ecuadorian montane forest. Plant and Soil 286, 287–299.

Still, C.J., Foster, P.N., Schneider, S.H., 1999. Simulating the effects of climate change on tropical montane cloud forests. Nature 398, 608–610.

Street Perrott, F.A., Huang, Y.S., Perrott, R.A., Eglinton, G., Barker, P., BenKhelifa, L., Harkness, D.D., Olago, D.O., 1997. Impact of lower atmospheric carbon dioxide on tropical mountain ecosystems. Science 278, 1422–1426.

Tikhonov, A.N., Arsenin, V.A., 1977. Solutions of Ill-posed Problems. Winston & Sons, Washington. 272 pp.

Tonneijck, F.H., Van der Plicht, J., Jansen, B., Verstraten, J.M., Hooghiemstra, H., 2006. Radiocarbon dating of soil organic matter fractions in Andosols in northern Ecuador. Radiocarbon 48, 337–353.

Van Bergen, P.F., Bull, I.D., Poulton, P.R., Evershed, R.P., 1997. Organic geochemical studies of soils from the Rothamsted Classical Experiments - I. Total lipid extracts, solvent insoluble residues and humic acids from Broadbalk Wilderness. Organic Geochemistry 26, 117–135.

Vogts, A., Moossen, H., Rommerskirchen, F., Rullkotter, J., 2007. Molecular delta C-13 values of leaf wax components from plants growing in different tropical habitats. Geochimica et Cosmochimica Acta 71 A1072–A1072.

Wille, M., Hooghiemstra, H., Hofstede, R., Fehse, J., Sevink, J., 2002. Upper forest line reconstruction in a deforested area in northern Ecuador based on pollen and vegetation analysis. Journal of Tropical Ecology 18, 409–440.