



**HAL**  
open science

# **BACKLAND: spatially explicit and high-resolution pollen-based BACKward LAND-cover reconstructions**

Clara Plancher, Florence Mazier, Thomas Houet, Cédric Gaucherel

► **To cite this version:**

Clara Plancher, Florence Mazier, Thomas Houet, Cédric Gaucherel. BACKLAND: spatially explicit and high-resolution pollen-based BACKward LAND-cover reconstructions. *Ecography*, 2023, 2024 (2), 10.1111/ecog.06853 . hal-04355147

**HAL Id: hal-04355147**

**<https://hal.inrae.fr/hal-04355147>**

Submitted on 20 Dec 2023

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

# ECOGRAPHY

## Research article

### BACKLAND: spatially explicit and high-resolution pollen-based BACKward LAND-cover reconstructions

Clara Plancher<sup>1</sup>  <sup>1</sup>, Florence Mazier<sup>1</sup> , Thomas Houet<sup>2</sup>  and Cédric Gaucherel<sup>3</sup>

<sup>1</sup>GEODE, UMR CNRS, Université Jean Jaurès, 5 Allées Antonio Machado, Maison de la Recherche, Toulouse, France

<sup>2</sup>LETG-Rennes, UMR CNRS, Université Rennes 2, Place du Recteur Henri le Moal, Rennes, France

<sup>3</sup>AMAP, UMR, INRAE, CIRAD, CNRS, IRD, Université de Montpellier, Montpellier, France

Correspondence: Clara Plancher ([cl.pl@live.fr](mailto:cl.pl@live.fr))

#### Ecography

2023: e06853

doi: [10.1111/ecog.06853](https://doi.org/10.1111/ecog.06853)

Subject Editor: Kate Lyons

Editor-in-Chief: Miguel Araújo

Accepted 6 October 2023



Studying the interactions between humans, land-cover and biodiversity is necessary for the sustainable management of socio-ecosystems and requires long-term reconstructions of past landscapes, improving the integration of slow processes. The main source of information on past vegetation is fossil pollen, but pollen data are biased by inter-taxonomic differential production and dispersal. The landscape reconstruction algorithm (LRA) approach is the most widely used to correct for these biases. The LOVE algorithm (LOcal Vegetation estimates), the second step in the LRA approach, also estimates the spatial extent of the local vegetation reconstruction zone (the relevant source area of pollen, RSAP). While LRA estimates have already been integrated into certain past land-cover mapping approaches, none have been designed to allow the diachronic reconstruction of a land-cover mosaic over the long term combining the following points: the direct integration of LOVE estimates as a source of variability in the composition and distribution of pollen taxa, without multiple scenarios, and the integration of spatiotemporal autocorrelation in the taxa distribution between periods. Here, we propose an innovative approach for BACKward LAND-cover reconstruction (BACKLAND), combining these points and estimating the past land-cover mosaic within a set of RSAPs. Based on three stages using parsimonious assumptions and easy-to-implement probabilistic and statistical tools, this approach requires LOVE estimates of sites close enough to each other for their RSAPs to overlap, botanical data, a digital elevation model and two recent land-cover maps. Developed and tested on a small study area within the mountain landscape of the Bassiès valley (French Pyrenees), BACKLAND achieved the reconstruction of a past land-cover map representing eight land-cover types at a spatial resolution of 20 m with a good level of accuracy. We show in this study the originality of this approach and discuss its potential for palaeoenvironmental studies, historical ecology and the management of socio-ecosystems.

Keywords: land-cover, landscape reconstruction, LOcal VEgetation estimates, multiple linear regression, relevant source area of pollen, validation, vegetation distribution



[www.ecography.org](http://www.ecography.org)

© 2023 The Authors. Ecography published by John Wiley & Sons Ltd on behalf of Nordic Society Oikos

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

## Introduction

Within the framework of landscape ecology theory, landscapes are defined as heterogeneous land-cover mosaics covering a few hectares to several square kilometres (Turner 1989, Kienast et al. 2007), and described by their composition (diversity and relative abundances of land-cover types) and configuration (shape, size and spatial arrangement of land-cover patches) (Forman 1995, Turner et al. 2001, Kienast et al. 2007). In this study, we adopt this definition without integrating its aesthetic and socio-cultural aspects.

Because of the close connections between societies, landscapes and biodiversity (Rosenzweig 1995, Duelli 1997, Wiens 2009), landscape-scale studies are particularly appropriate for implementing ecosystem management or protection measures (Turner 1989, Crooks and Sanjayan 2006, Fischer et al. 2011, Leite et al. 2013, Opdam et al. 2013, Wu 2013). Landscape management would particularly benefit from fine-scale studies of the quantity, location and frequency of changes within landscapes, as they improve the understanding of the processes behind land-cover change, and of the impact of these changes on the environment (Houet et al. 2010a). Reconstructing successions of fine-grained land-cover maps over the long term would refine our understanding of the legacy of past changes on our current environment (Pärtel et al. 2007, Bürgi et al. 2007, 2017, Gillet et al. 2016, Neumann et al. 2017, Garbarino et al. 2020, Le Provost et al. 2020, Tappeiner et al. 2020), an approach advocated by the historical, ecological and paleoecological communities (Foster et al. 2003, Seddon et al. 2014, Dearing et al. 2015, Herrault et al. 2015).

However, mapping past land-cover at landscape scale remains challenging due to the loss of both spatial and classification resolution of land-cover types beyond the time extent of remote sensing data, i.e. before the late 20th century for satellite imagery and the 1950s for panchromatic historical aerial photographs. Cadastral maps and land-cover type surveys may extend mapping back to the 19th century (e.g. the French Napoleonic Cadastre), but these are non-exhaustive in time and spatial coverage, which does not facilitate greater time-depth in landscape change studies (Fyfe et al. 2015). Moreover, historical land-cover maps focus on a limited number of land-cover types and are therefore incomplete representations of land-cover mosaics (Dahlström 2008). Historical or remote sensing land-cover maps are the result of a tradeoff between three types of resolution: classification detail, spatial resolution and temporal resolution (Zimmermann et al. 2007). Obtaining continuous multi-decadal records of land-cover maps (Dearing et al. 2015) with these three resolutions in the finest possible degree would be helpful for an array of ecosystem purposes, but this is impossible to achieve today.

Long-term data using pollen analysis offer otherwise almost unachievable possibilities to reconstruct past vegetation. When suitable sediment archives are available for coring, it is possible to work at various temporal and spatial scales with pollen as the direct link to past vegetation. Though pollen analyses provide an opportunity for reconstructing the

relative changes in vegetation composition as individual taxa or as land-cover types in a long-term perspective (Jolly et al. 1998, Tarasov et al. 2007, 2009, Fyfe and Woodbridge 2012, Joannin et al. 2012, Giesecke et al. 2017), translation into land-cover is not a trivial exercise, especially if the reconstruction aims to be quantitative and spatially referenced rather than qualitative (Bunting et al. 2018).

Over the last decades, advances in the theory of pollen analysis (Prentice 1985, 1988, Sugita 1993) coupled with an increase in computer power have led to the development of model-based reconstructions of land-cover composition such as the landscape reconstruction algorithm (LRA, Sugita 2007a, b). The LRA effectively reduces the biases caused by the non-linear pollen–vegetation relationship due to differences in sedimentary archives, basin size, inter-taxonomic differences in pollen productivity and dispersal characteristics, and spatial scales. Based on pollen extracted from sediments (lakes and bogs) and integrating pollen dispersal and deposition models (Prentice 1985, 1988, Sugita 1993, 1994), the LRA uses two models, REVEALS (regional estimates of vegetation abundance from large sites) and LOVE (local vegetation estimates), to translate pollen assemblages from a set of sedimentary sites into regional and local cover of plant taxa, respectively (Sugita 2007a, b). The LOVE model provides vegetation composition estimates expressed as distance-weighted plant abundance (DWPA) within a defined area, namely the relevant source area of pollen (RSAP, 0.1–3 km, Nielsen and Sugita 2005, Sugita et al. 2010a, b, Hjelle and Sugita 2012, Li et al. 2018) defined as the smallest area for which vegetation abundance can be modelled using fossil pollen records (Sugita 2007b). The LRA algorithm has been used widely in many parts of Europe and elsewhere over the last decade with reasonable success, in both flat and mountainous terrains (Cui et al. 2010, 2013, 2014, Gaillard et al. 2010, Nielsen and Odgaard 2010, Sugita et al. 2010a, b, Fyfe et al. 2013, Hultberg et al. 2014, Poska et al. 2014, Hjelle et al. 2015, Mazier et al. 2015, Fredh et al. 2019, Marquer et al. 2020a, b, Prösch-Danielsen et al. 2020, Plancher et al. 2022). Although LOVE estimates cannot inform on the spatial pattern of reconstructed taxa within the RSAP, they have been shown to perform better than raw pollen data at reconstructing spatial and temporal land-cover variability from neighboring sites (Overballe-Petersen et al. 2013, Plancher et al. 2022).

Several methods have been proposed to produce spatially explicit reconstructions of past local land-cover using the LRA algorithm. Recently, O'Dwyer et al. (2021) used LOVE-based land-cover estimates from multiple sites to create spatially continuous reconstructions of land cover based on interpolation techniques between point data. Alternatively, the Multiple scenario approach was developed to simulate the pollen signal from hypothetical maps of past land cover at locations with existing palynological records. The results are then compared statistically with the actual pollen assemblages in order to identify likely past vegetation mosaics (Middleton and Bunting 2004, Bunting and Middleton 2005, Bunting et al. 2018). It is however difficult to propose

plausible scenarios of land-cover dynamics (Caseldine 2008, Bunting 2018) with these methods, as they do not incorporate spatial and temporal dependency in the distribution and composition of vegetation cover between successive periods. Incorporating this spatial and temporal dependence (hereafter named spatiotemporal autocorrelation) within a method for spatializing LOVE estimates would improve the plausibility of reconstructed land cover dynamics.

To our knowledge, none of the existing approaches for mapping long-term land-cover mosaics have combined 1) the finest possible spatiotemporal and land-cover type classification resolutions and 2) the spatiotemporal autocorrelation in the distribution of vegetation over successional periods. Developing a new approach for producing plausible successions of landscape mosaics with the highest possible precision would allow us to propose successional land-cover trajectories for environmental management studies.

The Bassiès valley, in the French Pyrenees, represents a suitable case study for the development of a spatially explicit reconstruction of past land-cover mosaic based on LOVE estimates. It contains several small sedimentary sites for which pollen data at a high temporal resolution (10–20 years) have already been analysed and converted into DWPA by the LRA approach (Marquer et al. 2020a, b, Plancher et al. 2022). Plancher et al. (2022) showed that LOVE estimates from multiple sites are variable from one site to another, despite their proximity and the overlap of their RSAPs. Because their RSAPs overlap, the LOVE estimates from the Bassiès sites contain a proportion of plant cover in common. We believe it is possible to use these overlapping RSAPs and redundant plant cover to produce taxon distributions over the study area. Furthermore, the distribution and composition of land-cover types in this valley are documented by two recent land-cover maps (1993, 2008; Houet et al. 2012) and modern floristic surveys (2015–2020; Marquer et al. 2020a, b, Mazier et al. 2022).

In this paper, we propose a new approach for backward predicting past land-cover mosaic based on LOVE estimates and using probabilistic and statistical tools. We hypothesise that 1) LOVE estimates within overlapping RSAPs will enable us to access the past local distribution of individual taxa using probability density functions (PDFs, Kühl et al. 2002); that 2) autocorrelation of the taxon distribution may be backward computed with multiscale spatial analyses (Gaucherel 2007) by combining recent land-cover maps (Houet et al. 2012) and botanical surveys (Mazier et al. 2022); and that 3) linear models (McCullagh 1984) should be able to statistically explain land-cover type distributions based on taxon distributions and selected auxiliary environmental variables. Hereafter this approach named BACKward predicting past LAND-cover at fine scale will be referred to as BACKLAND.

We focus on the methodological aspects of the BACKLAND approach, its implementation and evaluation on recent time windows to generate the past Bassiès land-cover mosaic within a set of overlapping RSAPs of small sedimentary sites. The accuracy of the Bassiès land-cover estimated using BACKLAND is assessed in comparison with a

contemporary land-cover map. Finally, we draw conclusions on the possible implications of the BACKLAND approach on fossil pollen records for palaeoenvironmental studies, historical ecology, and socio-ecosystem management.

## Materials

Due to the large number of abbreviations used throughout the article, a summary of the main ones and their meanings is given in Table 1.

### Study zone characteristics

The Bassiès valley is located in the Vicdessos area within the Pyrenees mountain range (42°46'N, 01°26'E, Ariège, France, Fig. 1). It is a hanging glacial valley characterised by a flat bottom at around 1500 m of elevation on which wetlands have developed, surrounded by steep slopes culminating at 2676 m (Pique Rouge). The average annual precipitation is 1640 mm year<sup>-1</sup>, with one-third as snow from November–December to April–May, and the average annual temperature is around 7°C (Quintana-Seguí et al. 2008). The dominant winds come mostly from the west and north-west; wind speed varies from 0 to 4 m s<sup>-1</sup> (Szczypta et al. 2015).

Our study area covers 9 km<sup>2</sup> in the Bassiès valley and its vicinity: it is delimited by a 1 km radius from each sedimentary site for which LOVE estimates are available (Plancher et al. 2022). The northern part of our study area extends down the northern slopes towards the adjacent valleys (northeast, Fig. 1) to a minimum altitude of 1066 m. The southern part rises beyond Lake Sigriou (SIG) and represents the highest part, reaching an elevation of 2410 m.

The modern vegetation is dominated by heathlands (*Rhododendron ferrugineum* or *Calluna vulgaris*-dominated) with different-sized patches of grasslands *Festuca eskia*, *Nardus stricta* mainly distributed on slopes. Patches of forest are mainly found on the northern slopes, with beech-dominated forests and clumps of *Pinus uncinata* mainly below and above 1600 m respectively.

Table 1. Abbreviations frequently used in the paper.

Acronym	Meaning
BF	Broadleaved forest
CH	<i>Calluna</i> heathland
CMP	Comparison map profile (Gaucherel et al. 2008)
FG	<i>Festuca</i> grassland
MF	Mixed forest
MH	Mixed heathland
MHM	Multiscale heterogeneity map (Gaucherel 2007)
NRMSE	Normalized root mean square errors
OG	Other grassland
PB	Peatbogs (excluded from analysis)
PDFs	Probability density functions
PF	Pine forest
RH	<i>Rhododendron</i> heathland
RSAP	Relevant source area of pollen (Sugita 1998, 2007b)
TW	Time window

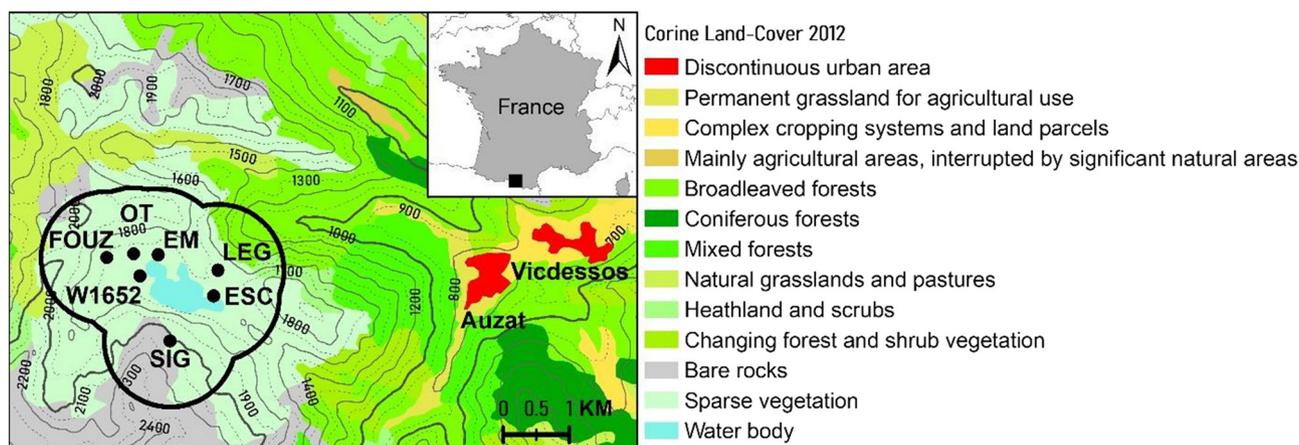


Figure 1. Location of the study zone (black line) and the targeted sedimentary sites with available TW1995 LOVE estimates of vegetation composition within 1 km RSAP (Plancher et al. 2022). Abbreviations: EM, Etang Mort, ESC: Escalé; FOUZ: Fouzès; LEG, Legunabens; OT: Orry de Théo; SIG: Sigiou. Characteristics of the sites are given in Table 2.

## Datasets

The approach to reconstruct past land-cover maps requires pollen data from several nearby sites translated into distance weighted plant abundances (DWPAs) using the LRA approach (Sugita 2007a, b), as well as cartographic (digital elevation model and land-cover maps) and botanical data (section ‘LOVE-based estimates of local vegetation composition’ and ‘Cartographic and botanical data’ below).

### LOVE-based estimates of local vegetation composition

By taking into account the dispersal and deposition of pollen grains via the Gaussian plume model, the LRA approach considers that plants closer to the sampling point contribute more grains to the pollen assemblage than plants further away (Sugita 2007b). The LOVE model thus provides DWPAs within the RSAP, the smallest spatial scale for which vegetation composition can be estimated by the model using a fossil pollen record.

Details about the chronology of the sediment cores ( $^{210}\text{Pb}$ ,  $^{137}\text{Cs}$  and  $^{14}\text{C}$  dates) and the LOVE estimates (Table 2) used in this study were published in Marquer et al. (2020a, b) and Plancher et al. (2022), including information on age-depth models, site selection, pollen sampling, pollen data, parameters used to run the LRA approach and RSAP calculation. Plancher et al. (2022) showed that the RSAP estimates varied from 250 to nearly 1000 m over the last 200 years, therefore the area within a 1 km radius (largest RSAP value) from each site was considered as appropriate and suitable for the local scale of reconstruction of vegetation in the area.

For this paper, we used the LOVE estimates for 18 major taxa available for the 1990–2000 cal. AD time window, further referred to as TW1995. Ten tree taxa (*Abies*, *Betula*, *Corylus*, *Fagus*, *Fraxinus*, *Picea*, *Pinus*, *Quercus*, *Salix* and *Tilia*), three shrub taxa (*C. vulgaris*, Ericaceae and *Juniperus*), and five grass taxa (Asteraceae Sub-Family Cichorioideae, Cyperaceae, Poaceae, *Plantago lanceolata* and *Potentilla*-type) were considered. These 18 taxa represent between 72 and

95% of the total pollen count of the targeted time window. The LOVE outputs are expressed in proportions of the total cumulative sum of distance weighted plant abundance at the 1 km-RSAP around each of the seven cores. Hereafter, these DWPAs will be referred to as LOVE estimates. LOVE estimates range from 0 to 1; 0 means that the plant species is absent, and 1 that the DWPA within the RSAP is 100%. The LOVE estimate for taxa  $i$  at site  $k$  will be written  $\text{LOVE}_{i,k}$  hereafter and the 18 taxa estimates sum up to 100% of the vegetation composition.

As the study sites are close to each other (between 352 and 1728 m), the RSAP of each site overlaps with at least one (SIG with ESC) and up to four sites (EM with OT, FOUZ, W1652 and LEG). By showing large inter-site variations despite this proximity, these LOVE estimates represent a strong potential for spatializing pollen-based vegetation reconstructions (Plancher et al. 2022). According to LOVE estimates at TW1995, the vegetation cover within 1 km radii is dominated by Ericaceae around SIG, LEG and FOUZ and W1652, by *C. vulgaris* around ESC, by Cyperaceae around EM, and by Poaceae around OT (Table 2, Plancher et al. 2022). Tree taxa represent between 9% (W1652) and 41% (LEG) of the local vegetation, shrubs between 0.1% (OT) and 91% (W1652) and grass taxa between 0% (LEG) and 88% (EM).

### Cartographic and botanical data

A digital elevation model and two recent land-cover maps (1993 and 2008, Houet et al. 2012) were used in this paper. The digital elevation model was used to estimate auxiliary environmental variables. These auxiliary variables were integrated in the models to take into account the effects of the environmental variability of the Bassiès landscape on its vegetation distribution. The 1993 and 2008 land-cover maps are based on true color (RGB) and false color (near infrared) aerial images, classified using a geographic object-based image analysis combining visual interpretation and automatic classification (Houet et al. 2012, Sheeren et al. 2012).

Table 2. Sites' characteristics and LOVE estimates of local 18 taxa-based vegetation composition (from Plancher et al. 2022).

		Sites characteristics						
Name		Legunabens	Sigriou	Etang mort	Escale	Fouzes	Orry de Théo	W1652a
Acronym		LEG	SIG	EM	ESC	FOUZ	OT	W1652
Elevation (m)		1680	2000	1670	1630	1720	1680	1660
Radius (m)		58	66	36	40	13	31	6
Type		Lake	Lake	Bog	Bog	Bog	Bog	Bog
LOVE estimates (% DWPA; Plancher et al. 2022)								
Trees	<i>Abies</i>	2.54	0.91	–	0.20	0.22	0.02	0.03
	<i>Betula</i>	1.65	–	2.20	7.47	2.15	0.79	2.29
	<i>Corylus</i>	4.35	6.36	4.36	3.89	5.03	2.97	1.91
	<i>Fagus</i>	14.32	–	0.14	10.31	1.57	0.95	1.73
	<i>Fraxinus</i>	0.08	–	–	2.90	1.19	2.25	–
	<i>Picea</i>	0.71	0.02	–	0.03	0.03	0.15	0.07
	<i>Pinus</i>	14.49	2.80	3.74	6.22	9.58	3.22	1.91
	<i>Quercus</i>	2.36	–	0.63	5.14	2.74	1.69	1.35
	<i>Salix</i>	0.04	0.21	–	0.12	0.02	0.36	–
	<i>Tilia</i>	0.61	0.71	0.30	0.29	–	–	–
Shrubs	<i>Calluna vulgaris</i>	4.78	–	–	45.92	10.50	–	41.42
	Ericaceae	49.45	51.28	0.29	6.15	41.07	–	49.29
	<i>Juniperus</i>	0.79	0.20	0.05	–	–	0.14	–
Grasses	Comp. SF Cichorioideae	2.37	11.00	3.29	–	0.16	5.20	–
	Cyperaceae	–	1.96	54.02	–	–	23.77	–
	<i>Plantago lanceolata</i>	1.46	1.31	0.95	–	0.47	4.71	–
	Poaceae	–	23.25	24.63	11.36	23.57	50.57	–
	<i>Potentilla</i> -type	–	–	5.37	–	1.69	3.22	–

These maps were resampled into a 20 m resolution raster grid by averaging the digital elevation model values and by assigning the dominant land-cover type with the largest cover area to each pixel (Fig. 2). Preliminary tests showed that a 20 m resolution is a reasonable compromise, as it enables even rare land-cover types to be integrated while having a reasonable computing time. BACKLAND is based on a simplified version of the land-cover type nomenclature, excluding from analysis and calculations both non-pollen producing areas (i.e. mineral surfaces, roads, buildings and water surfaces) and peatbogs which are assumed to be constant (PB, Fig. 2).

Eight land-cover types were used, comprising three tree forest types (Pine forest, PF; Broadleaved forest, BF; Mixed forest, MF), three heathland types (*Rhododendron* heathland, RH; *Calluna* heathland, CH; Mixed heathland, MH), and two grassland types (*Festuca* grassland, FG; Other grasslands, OG).

The proportions of each type vary slightly between 1993 and 2008. RH (27.0–28.2% of the vegetated surfaces of the study area in 1993–2008), MH (15.7–15.6%) and CH (24.8–25.2%) are distributed respectively over the southern, central and northern parts of the study area, forming the main matrix of vegetation in which other land-cover types are immersed. Types FG (4.1–3.5%) and OG (13.0–11.3%) are concentrated mainly on the slopes to the western part of the study area. PF (3.3–4.2%) forms small scattered stands on the steep northern slopes separating SIG from the

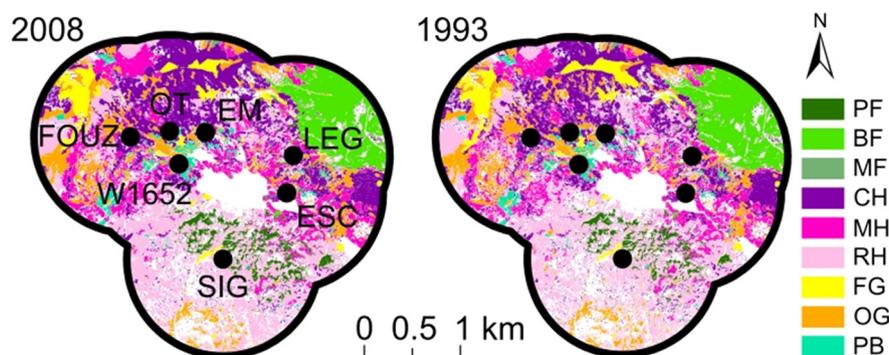


Figure 2. Land-cover maps used for backward spatiotemporal autocorrelation of taxa distributions (2008) and for models learning (1993). Adapted from Houet et al. (2012). White areas: non-vegetated areas (water, mineral). PF: Pine forest; BF: Broadleaved forest; MF: Mixed forest; CH: *Calluna* heathland; MH: Mixed heathland; RH: *Rhododendron* heathland; FG: *Festuca* grassland; OG: Other grasslands; PB: Peatbogs.

other sites. The type BF (9.5%) is essentially characterized by a homogeneous unit northwest of LEG, in the slope leading to the adjacent valley, while MF (0.30–0.34%) forms a few small patches in the north-north-east of the study area.

The vegetation composition of these land-cover types is based on modern botanical surveys including 125 plots described by 116 vascular plants (Mazier et al. 2022). Plant taxa were grouped according to the 18 pollen morphological types for which Plancher et al. (2022) computed past LOVE estimates. The vegetation data (percentage cover from field survey) were recalculated on the sum of the 18 selected taxa (Table 3). Land-cover types are characterised by seven to 15 of these 18 taxa. Grasslands and heathlands are mainly composed of ubiquitous taxa, present in at least six land-cover types. PF and heathlands are the only types containing *Pinus*, which represents less than 2% in heathlands against 32.4% in PF. BF and MF are the only types containing beech *Fagus*, representing 49.5 and 55% of the vegetation cover respectively, while BF is composed of more than 20% of taxa present exclusively in this type (*Corylus*, *Fraxinus*, *Picea*, *Quercus* and *Tilia*). The percentage cover of plant taxa for each land-cover type is assumed to be constant over time.

## Methods

The BACKLAND method for backward modelling land-cover mosaics required three stages: 1) a preliminary stage (Fig. 3a) generating two intermediate results: taxon distributions based on the 2008 map and taxon distributions based on the LOVE estimates of TW1995. These distributions were used in the second stage to 2) integrate spatiotemporal autocorrelation into the estimated taxon distribution at TW1995, constituting the explanatory variables of the TW1995 land-cover type distributions with a set of environmental variables (Fig. 3b). Finally, 3) multiple linear models were used to estimate a statistical link between the land-cover type

distributions and this set of explanatory variables, ultimately leading to the backward prediction of a land-cover map at TW1995 (Fig. 3c).

### Preliminary stage: map-inferred taxon distributions in 2008 and LOVE-inferred taxon distributions at TW1995

A general characteristic of ecological systems is that their geographical proximity is correlated with their similarity, i.e. they have a positive spatial autocorrelation (Tobler 1970, Legendre and Fortin 1989, Gaucherel et al. 2016). In the absence of a major disturbance, we can also assume a temporal autocorrelation of the vegetation composition and configuration between two successive time windows. The objective of this stage was to produce two intermediate taxa distribution maps, later used for backward integrating spatiotemporal autocorrelation into the distribution of taxa at TW1995: their recent distribution inferred from the 2008 land-cover map and their past distribution inferred from LOVE estimates at TW1995.

#### Map-inferred taxa distributions in 2008

Recent taxon distributions were obtained from the 2008 land-cover map and the 18 taxa-based botanical compositions of the land-cover types (Fig. 3a), based on the following Eq. 1:

$$MID_i = \frac{\sum_{LC=1}^{N_{LC}} D_{LC} \times P_{i,LC}}{\sum_{i=1}^I \sum_{LC=1}^{N_{LC}} D_{LC} \times P_{i,LC}} \quad (1)$$

where  $MID_i$  represents the map inferred taxa distribution of a taxon  $i$  expressed as a proportion of the total vegetation,  $N_{LC}$

Table 3. Land-cover type botanical composition based on the 18 LOVE taxa (from botanical surveys by Mazier et al. 2022).

	Pollen taxa	PF	BF	MF	CH	MH	RH	FG	OG
Trees	<i>Abies</i>		4.70	22.24		1.28 × 10 <sup>-3</sup>			
	<i>Betula</i>	3.22 × 10 <sup>-3</sup>	4.31	1.19	0.01				
	<i>Corylus</i>		12.38						
	<i>Fagus</i>		49.51	55.04					
	<i>Fraxinus</i>		4.48						
	<i>Picea</i>		0.22						
	<i>Pinus</i>	32.39			0.18	1.97	3.45 × 10 <sup>-3</sup>		
	<i>Quercus</i>		0.05						
	<i>Salix</i>					8.79 × 10 <sup>-4</sup>			
	<i>Tilia</i>			3.27					
Shrubs	<i>Calluna vulgaris</i>	15.34		2.51 × 10 <sup>-3</sup>	53.33	22.59	10.70	6.06	4.29
	Ericaceae	43.90	8.97	16.54	15.99	14.02	72.49	4.95	7.22
	<i>Juniperus</i>	6.33	0.02		4.10	0.75	0.84	0.90	
Herbs	Comp. SF. Cichorioideae	0.03	0.01	0.27	0.62	0.03	0.02	2.28	0.01
	Cyperaceae	0.06	1.65	2.68	4.45	42.16	1.98	2.75	26.76
	<i>Plantago lanceolata</i>				8.79 × 10 <sup>-4</sup>				1.45 × 10 <sup>-3</sup>
	Poaceae	1.93	8.77	2.03	19.14	17.19	12.15	79.68	60.61
	<i>Potentilla-t</i>	0.01	1.66		2.18	1.29	1.82	3.37	1.12

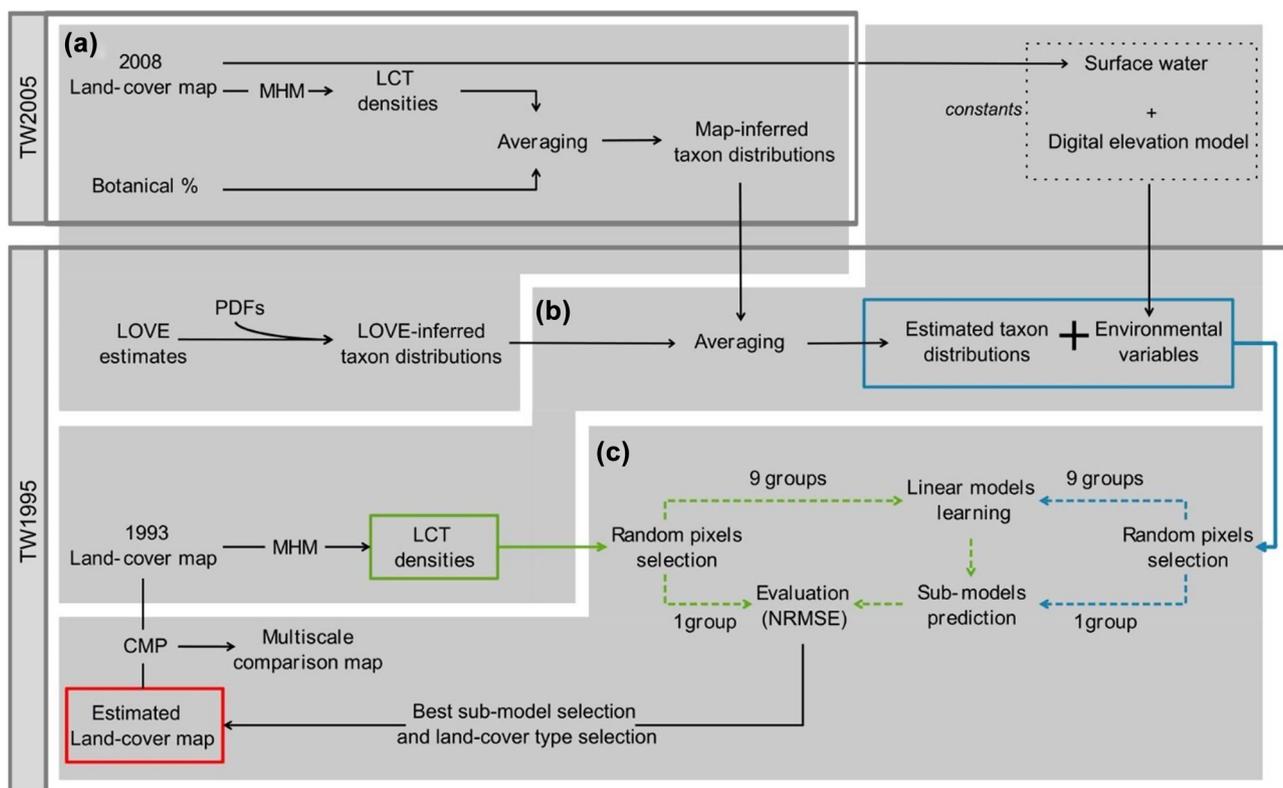


Figure 3. Backward land-cover reconstruction (BACKLAND), a three-stage approach: (a) preliminary stage for taxon distributions (section ‘Preliminary stage: map-inferred taxon distributions in 2008 and LOVE-inferred taxon distributions at TW1995’); (b) variables used by the BACKLAND approach (section ‘Variables used by the BACKLAND approach’); (c) BACKLAND modelling (section ‘Backward landscape modelling’). LCT: land-cover type; green: explained variables; blue: explanatory variables; red: final result; dotted arrows: learning and cross-validation steps, repeated for each sub-model; plain arrows: step processed once for each LCT.

is the number of land-cover types (eight),  $P_{i,LC}$  the botanical proportion of the taxon  $i$  in the land-cover type,  $I$  the total number of taxa (18), and  $D_{LC}$  their multi-scale density. Preliminary analyses showed that the use of quantitative and continuous densities is preferable to the presence/absence of land-cover types for several steps in the processing chain. In this step, the use of multiscale densities was motivated by the need to take into account both the heterogeneity of land-cover types and the fuzziness of their boundaries – due to their permeability to dispersal and colonisation by ubiquitous taxa. We used the Multiscale heterogeneity map software (MHM; Gaucherel 2007, Pavageau et al. 2017) to produce a multi-scale density map for each land-cover type. These maps represent the average proportion of each type in their neighbourhood, on observation scales ranging from 40 to 280 m, every 40 m, in order to reduce and homogenise possible biases due to scale choices.

#### LOVE-Inferred taxa distributions at TW-1995

This step corresponds to the translation of point LOVE estimates into LOVE-inferred taxa distribution maps (Fig. 3a). First, we assumed that the probability of the presence of a taxon  $i$  around each pollen site ( $P_{\text{presence}_i}$ ) follows a bivariate normal PDF (Kühl et al. 2002). This choice was consistent with the working hypotheses of the LRA algorithms,

considering that the atmospheric dispersion of pollen follows a Gaussian distribution (Prentice 1985, Sugita 1993, 1994, 2007b). Moreover, it had the advantage of requiring few parameters to be estimated, allowing a simple estimation of probability densities:

$$P_{\text{presence}_i} \sim \text{PDF}, \text{ with PDF} = N(\mu, \Sigma) \quad (2)$$

where  $\mu = [\mu_x, \mu_y]$  and  $\mu_x$  and  $\mu_y$  correspond to the longitude and latitude of the site respectively, and where  $\Sigma = (\sigma_{xx}, \sigma_{xy}, \sigma_{yx}, \sigma_{yy}) = (\sigma \ 0 \ 0 \ \sigma)$  the variance-covariance matrix, assuming directional independence and considering a reasonable isotropic bivariate normal distribution. The standard deviation  $\sigma$  of the PDF around  $\mu_x$  and  $\mu_y$  is linked with the 1 km radius of the RSAP (Sugita 1994), assumed to be constant over time (Plancher et al. 2022). As the LRA algorithm does not provide a measure of uncertainty for RSAP (Sugita 2007b), it was assumed that each taxon is present within RSAP with a 99.9% probability. Then,

$$P(\mu^* - \text{RSAP} < \text{Presence}_i < \mu^* + \text{RSAP}) = 0.999 \quad (3)$$

where  $\text{Presence}_i$  is the geographic coordinate of a pixel where taxon  $i$  presence probability is estimated. Here,  $\mu^*$  represents

$\mu_x$  or  $\mu_y$ , as Eq. 3 is valid in both directions of space. According to the tables of quantiles of the normal distribution,

$$P(\mu_* - 3.09 \times \sigma < \text{Presence}_i < \mu_* + 3.09 \times \sigma) = 0.999,$$

$$\text{and then } \sigma = \frac{\text{RSAP}}{3.09} \quad (4)$$

LOVE estimates represent the composition of vegetation as reflected by pollen assemblages, incorporating dispersal/deposition biases and thus the fact that more distant emitting plants contribute less to the pollen assemblage than nearby plants (distance-weighted plant abundance, Sugita 2007b). It therefore makes sense to combine these LOVE estimates with bivariate Gaussian PDFs of taxon presence centered on the sampling points. For each taxon, the PDF around each site was therefore weighted by its corresponding LOVE estimate. As the PDFs are continuous and thus overlapping, a taxon LOVE-inferred distribution is the weighted sum of the seven PDFs of each site, obtained as follows:

$$\text{LID}_i = \frac{\sum_{k=1}^K \text{PDF}_k \times \text{LOVE}_{i,k}}{\sum_{i=1}^I \sum_{k=1}^K \text{PDF}_k \times \text{LOVE}_{i,k}} \quad (5)$$

where  $\text{LID}_i$  is the LOVE-inferred distribution of the taxon  $i$ ,  $K$  and  $I$  are the total number of sites and taxa included, respectively (i.e. seven and 18 in this case), and  $\text{PDF}_k$  is the probability density of a taxon's presence around site  $k$ . PDFs and LOVE-inferred distribution calculations were coded in R ([www.r-project.org](http://www.r-project.org)).

## Variables used by the BACKLAND approach

### Land-cover distributions at TW1995

The explained variables, representative of the land-cover type distribution at TW1995, were the multiscale densities of each land-cover type on the 1993 map, computed with MHM with the same settings as those used for the map-inferred distributions (section 'Map-inferred taxa distributions in 2008'). Models based on presence/absence maps of each land-cover type were tested, but using the multiscale density maps improved the goodness of fit of the models (not shown).

### Explanatory variables

Two types of variables were used in the models to explain the TW1995 land-cover type densities: the estimated taxon distributions at TW1995 and environmental variables. First, the estimated taxon distributions were obtained by averaging the 2008 map-inferred distributions and the TW1995 LOVE-inferred distributions. Estimated distributions at TW1995 thus take into account backward spatiotemporal autocorrelation, yet assuming a similar influence of map-inferred

distributions and LOVE-inferred distributions. The residual autocorrelation of the models (Cruse et al. 2012) was previously considered but produced too strong a constraint in the models (not shown).

Environmental variables were included for calculating the effect of auxiliary explanatory variables on the vegetation distribution. Average values for elevation, exposure, slope and curvature were estimated from the digital elevation model. In order to obtain a north-south gradient, the exposure values were set between  $0^\circ$  (north) and  $180^\circ$  (south). The distance from the centre of each cell to the nearest water point (i.e. stream, lake) was derived from the original 2008 land-cover map (Houet et al. 2012).

## Backward landscape modelling

Finally the BACKLAND approach for landscape backward prediction involves (section 'Land-cover types multiple linear regressions') learning and applying the models to the set of variables in order to then (section 'Estimated land-cover map prediction and evaluation') backward-model a land-cover map and assess its similarity with the observed data (Fig. 3c).

### Land-cover types multiple linear regressions

The relationship between land-cover type densities and explanatory variables (estimated taxon distributions and environmental variables) was explored by ordinary multiple linear regressions, commonly used in statistical analysis of spatial data to predict distribution maps (Guisan and Zimmermann 2000). One linear regression was built for each land-cover type, assuming that their densities follow a Gaussian distribution and respond linearly to the explanatory variables. Since we considered the composition of a land-cover type to be constant, the model of a given type contains only the estimated distributions of taxa whose presence in this type was attested by the botanical data. The purpose of this supervision was to help the models produce a relevant statistical link. Then, a stepwise regression was applied to each model as a first analysis, in order to select only the estimated distributions and the environmental variables that significantly explained the variance of the studied land-cover type density ( $p < 0.05$ ). The models and stepwise regressions were computed using the 'stats' and 'MASS' packages ([www.r-project.org](http://www.r-project.org), Ripley et al. 2013).

In order to estimate the prediction errors of the models, a K-fold cross-validation was implemented for the learning of each model (Hastie et al. 2009, Bennett et al. 2013). For each land-cover type, the response variable (land-cover type density) and the selected explanatory variables were divided into ten random groups (the same for all land-cover type models), each comprising 2100 pixels. Then, ten different sub-models were run, each time removing a different group of pixels to perform the training on the other nine groups (18 900 pixels) and the prediction of the density of the land-cover type in question on the excluded group. The normalized root mean square errors (NRMSE, Shcherbakov et al. 2013) were then used to estimate the prediction error of each land-cover type

sub-model and to compare the predictive power of the different models with each other:

$$\text{NRMSE} = \frac{\text{RMSE}}{D_{\text{obs}}} = \frac{\sqrt{\frac{\sum (D_{\text{obs}} - D_{\text{pred}})^2}{N}}}{D_{\text{obs}}} \quad (6)$$

with  $D_{\text{obs}}$  the observed density of a land-cover type in a pixel,  $D_{\text{pred}}$  the density predicted by the corresponding sub-model on this pixel, and  $N$  the number of observations (2100 pixels). We chose to normalise the RMSE to avoid the scale dependency between land-cover type models (Shcherbakov et al. 2013).

#### Estimated land-cover map prediction and evaluation

Finally, the densities of each land-cover type were predicted over the whole study area using the best sub-model, selected on the basis of the minimum NRMSE, pinpointing the model with the best predictive power. Using the sub-model with the best goodness of fit (largest fitted  $R^2$ ) or the complete models before cross-validation did not significantly change the prediction of the models. At each pixel, the land-cover type corresponding to the highest predicted density value is chosen, ultimately producing pixel-by-pixel the land-cover mosaic of the past landscape. The similarity between the TW1995 predicted map and the observed 1993 map was evaluated using the comparison map profile software (CMP, Gaucherel et al. 2008, 2018), calculating the Cohen's Kappa index (Cohen 1960) at the same observation scales as those used by the MHM software to calculate land-cover type densities (section 'Map-inferred taxa distributions in 2008'). CMP produces a multi-scale comparison map, thus reducing scaling biases too, where the average Kappa value at each pixel quantifies the average similarity between the two maps in the vicinity of that pixel. An average Kappa higher than 0.6 reflects a strong similarity between observed and estimated datasets (Cohen 1960, Landis and Koch 1977). Using the same method, we also evaluated the similarity between the BACKLAND map at TW1995 and the 2008 map integrated into the BACKLAND estimation with the map-inferred taxon distributions. This was done in order to test the influence of the 2008 map on the final TW1995 BACKLAND result. These comparisons finally enabled the overall quality of the proposed method to be assessed. In order to test for a potential bias related to the different spatial configurations of the land-cover types, a Pearson correlation coefficient was calculated between their average patch density on the observed 1993 land-cover map and the average Kappa index between their observed and BACKLAND-estimated distributions.

## Results

### Intermediate maps and estimated taxa distributions

The LOVE-inferred distributions showed gradients in vegetation composition (Fig. 4b) following the variability of the LOVE estimates between the seven sedimentary sites at TW1995 (Table 2). By integrating the 2008 map-inferred

(Fig. 4a) and the TW1995 LOVE-inferred distributions with equal weights, the estimated taxa distributions at TW1995 kept the gradients associated with the LOVE-inferred distributions while being constrained by the map-inferred distributions (Fig. 4c), thus incorporating the spatiotemporal autocorrelation in a straightforward manner.

According to the estimated distributions, the areas that were the most represented by grass taxa during the TW1995 were the northern and western parts of the study area, concentrating the highest percentages of Cyperaceae and Poaceae, reaching up to 42 and 52% respectively (Fig. 4c). The southern part of the study area was mainly associated with the dominance of Ericaceae, with percentages greater than 50% over most of the studied area, and reaching 61% of the estimated vegetation. Their abundance estimated by the TW1995 LOVE-inferred distributions in the north-east part was largely mitigated by the map-inferred distributions constraint, as the broadleaved forest present at this location on the 2008 land-cover map contains only a small amount of Ericaceae (Table 3). The other two shrub taxa (*C. vulgaris* and *Juniperus*), mainly present in the central and northern parts, were more disparately distributed throughout the study area, with a strong predominance of *C. vulgaris* (up to 43%) over *Juniperus* (less than 3%). Finally, the eastern part was associated with the dominance of tree taxa, mainly *Fagus* in the northeast, whose percentages reached 32% in places, and *Pinus*, more to the southeast, representing around 15% of the vegetation (Fig. 4c).

### Land-cover models

#### Variables selection and cross-validation

The land-cover type densities from the 1993 land-cover map were related to the taxon distributions estimated at TW1995 and the environmental variables using multiple linear models. The stepwise regression did not change the composition of the linear predictors based on the land-cover type botanical composition, with the exception of the exclusion of *Quercus*, and *Tilia* estimated distributions from the broadleaved forest (BF) predictor (Table 4). All environmental variables were also used by the stepwise regression. Slope, elevation and distance to the nearest water point were significant in the linear models for all land-cover types, while only *Rhododendron* heathland (RH) was not significantly associated with exposure. Terrain curvature was only significantly explanatory for the *Festuca* grassland (FG) distribution (positive influence), while the other environmental variables were significant for all land-cover types, except distance to water for mixed heathlands (MH).

All linear models performed with K-fold cross-validation were significant ( $p < 0.05$ ). The goodness of fit of each sub-model was not greatly influenced by the random sampling, as adjusted  $R^2$  appear stable across sub-models with small standard deviations (less than 0.002) (Fig. 5a). BF had the highest average fitted  $R^2$  (0.99), followed by mixed forests (MF, 0.91). Pine forest (PF) had an average fitted  $R^2$  of 0.83, and mixed (MH), *Rhododendron* (RH) and *Calluna* heathlands (CH) had fitted  $R^2$  of 0.76, 0.88 and 0.89, respectively.

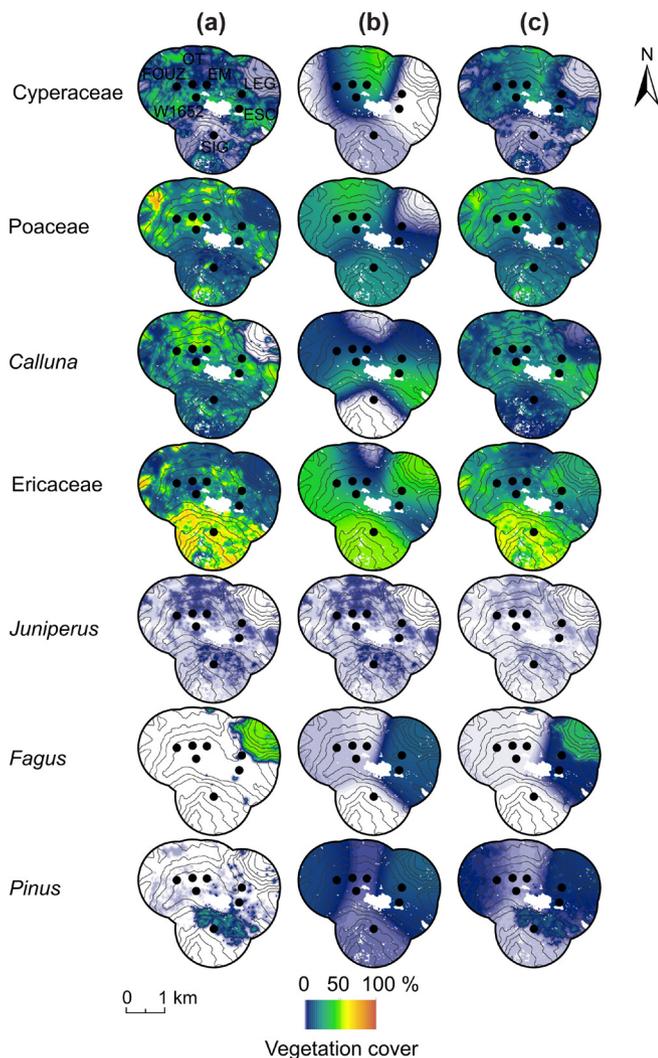


Figure 4. Taxon distribution maps. (a) Map-inferred taxon distributions based on 2008 land-cover map. (b) LOVE-inferred taxon distributions based on TW1995 LOVE estimates. (c) Estimated taxon distributions obtained by averaging (a) and (b). Only seven taxa are shown here.

Grasslands showed the lowest fitted  $R^2$ , with averages of 0.43 and 0.63 for FG and other grasslands (OG).

NRMSEs indicate differences in the quality of model predictions (Fig. 5b). The model with the highest predictive power was the BF model. The three heathlands and PF models had intermediate predictive abilities, with the CH model being the best of the three heathlands, followed by RH, MH and PF. The MF and FG models had significantly higher NRMSEs than the others, with higher standard deviations, indicating a higher sensitivity to random sampling.

#### Composition of predictors

The linear predictors of land-cover type densities contained between six (FG) and 12 (BF) estimated taxa distributions, and between three (RH) and five (FG) environmental variables. The linear predictors of the two grassland types were composed of ten common explanatory variables (six estimated

distributions and four environmental variables). But only four of these ten common variables had the same sign of coefficient, indicating opposite effects of most variables on grassland type density. The taxa that most influenced the distribution of FG were *Juniperus* and *Potentilla*-type (negatively and positively, respectively). Conversely, the density of OG was the most negatively influenced by the estimated distribution of *Potentilla*-type, and the most positively by that of *P. lanceolata*. Both grassland types are denser at lower elevations and away from water, but FG is favoured by a rather flat terrain with a northern exposure, while OG is favoured by slope and southern exposure. FG is the only land-cover type influenced by curvature, with a positive effect indicating a positive influence of a curved terrain favouring rainwater runoff. Regarding the three heathland types, the CH predictor differed from both MH and RH in terms of the influence of the estimated distributions. With seven estimated distributions having the same sign coefficient, MH and RH had the most similar predictors, MH differing only by a positive influence of *Abies* and *Cyperaceae*. The patches of CH and RH are denser on the slopes, whatever the exposure for RH but preferentially south for CH, closer to water points and at lower altitude for RH and far from water points and at higher altitude for CH. The density of MH is disadvantaged by slope, but is favoured by southern exposures and altitude.

The three tree-covered types were positively influenced by the estimated distributions of *Cyperaceae*, *Ericaceae* and *Poaceae*. *Fagus* estimated distribution was associated with the largest coefficient of BF, while MF was mostly influenced by *Abies*, and PF by *Juniperus*. Regarding the influence of the environmental variables, the densities of the three forest types are favoured by steep slopes and the proximity to water points. PF and MF are denser on northern slopes and altitude has a negative influence on the density of their patches, unlike BF.

#### Predicted land-cover map

The 1993 land-cover map and its BACKLAND estimate for TW1995 were highly similar (Fig. 6), with an average multiscale Kappa of 0.65 ( $> 0.6$ , Landis and Koch 1977). However, the mean Kappa is higher between the estimated TW1995 map and the 2008 map (0.71), integrated into the BACKLAND estimate by map-inferred distribution as a representation of spatio-temporal autocorrelation.

The average multiscale similarity differed according to the predicted land-cover type (Fig. 6) and was positively correlated with the average multiscale patch density of each type (Fig. 7). The land-cover type associated with the best prediction was Deciduous forest, which also had the densest patches, while *Festuca* grassland was the least well predicted type, although pine and mixed forest patches showed lower mean multiscale densities (Fig. 7).

#### Discussion

In this study, we succeeded in reconstructing a continuous land-cover mosaic by combining LOVE estimates around

Table 4. Multiple linear predictor composition. Only explanatory variables with significant influence on the land-cover type density are shown ( $p < 0.05$ ).

Explanatory variables	Tree-covered coefficients (x10 000)			Heathlands coefficients (x10 000)			Grasslands coefficients (x10 000)		
	PF	BF	MF	CH	MH	RH	FG	OG	
(Intercept)	-1833.79		-138.03	2959.93	-3523.91	-4331.16		382.85	
Trees									
<i>Abies</i>		-514.84	322.78		23.68				
<i>Betula</i>	212.20	-324.29	46.25	-224.61					
<i>Corylus</i>		17.76							
<i>Fagus</i>		225.85	-36.29						
<i>Fraxinus</i>		-104.14							
<i>Picea</i>		-2635.98							
<i>Pinus</i>	88.11			-287.80	304.24	202.76			
<i>Salix</i>				-4176.78					
Shrubs									
<i>Calluna vulgaris</i>	-10.19		2.13	-22.43	100.83	95.79	-1.36	-8.48	
Ericaceae	25.43	1.07	0.97	-36.13	11.61	72.27	-11.29	2.41	
<i>Juniperus</i>	498.23	19.86		1571.21	-1643.32	-1414.47	-87.77		
Herbs									
Comp. SF Cichorioideae	33.20	-17.84	-11.90	-240.27	283.70	376.16	23.16	-76.77	
Cyperaceae	29.11	1.35	1.92	-30.33	65.33	-14.20	-33.33	28.09	
<i>Plantago lanceolata</i>				239.38				73.38	
Poaceae	24.49	1.50	2.73	-15.73	-19.90	-13.70	22.32	62.52	
<i>Potentilla</i> -type	-90.92	24.37		-392.79	330.99	1077.23	222.61	-609.22	
Environmental variables									
Slope	1.87	0.47	0.36	1.62	-2.96	4.46	-2.64	2.57	
Exposure	-0.64	0.04	-0.11	0.99	0.36		-0.76	1.21	
Altitude	-0.27	0.08	-0.03	0.51	0.55	-0.12	-0.08	-0.20	
Curve							3.71		
Distance to water	-0.19	-0.12	-0.02	0.15		-0.12	0.05	0.44	

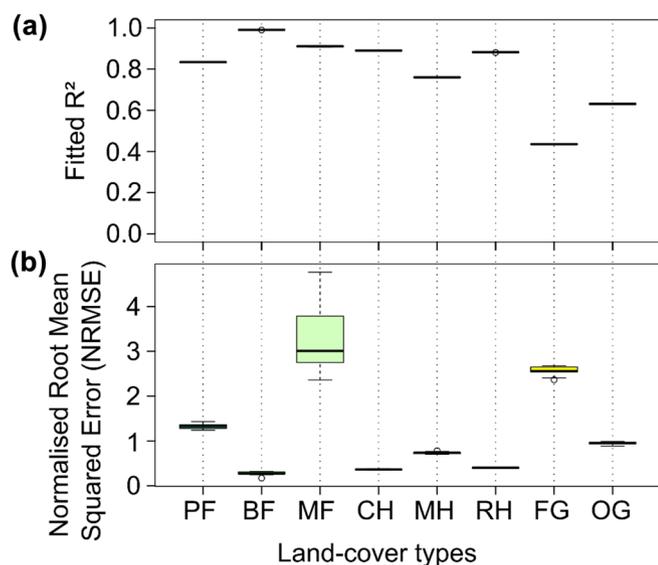


Figure 5. Estimation of model fit by K-fold cross-validation. (a) Sub-models fitted  $R^2$  and (b) sub-models normalized root mean square errors (NRMSE).

nearby sedimentary sites and recent cartographic and botanical data. The proposed BACKLAND approach 'BACKward reconstruction of LAND-cover mosaics' was implemented on a recent time-window (TW1995, 1990–2000) to test

its robustness and reliability. In the Bassiès study area, we achieved high levels of accuracy in terms of both land-cover types (eight) and spatial resolution (20 m). To our knowledge, this is the first past land-cover map reconstruction derived from pollen data with such a detailed vegetation composition and spatial resolution. BACKLAND integrates one of the most important properties of land cover and land-cover change, i.e. its strong dependence over space and time (i.e. spatial autocorrelation and temporal dependence). For this purpose, the reconstruction requires a three-stage approach (Intermediate taxon distributions, land-cover and explanatory variables estimation, and land-cover backward modelling) detailed in the following section. After a discussion of the advantages and weaknesses of the BACKLAND approach, it is compared to other existing pollen-based land-cover reconstruction approaches. Finally, we present its potential in the fields of historical ecology, landscape ecology and habitat management.

### BACKLAND approach

#### Application conditions and required data

The BACKLAND approach relies on past estimates of local vegetation composition estimated by the LRA approach (REVEALS and LOVE, Sugita 2007a, b). The use of LOVE estimates of vegetation composition is motivated by the characterisation of the spatial extent of the reconstruction using

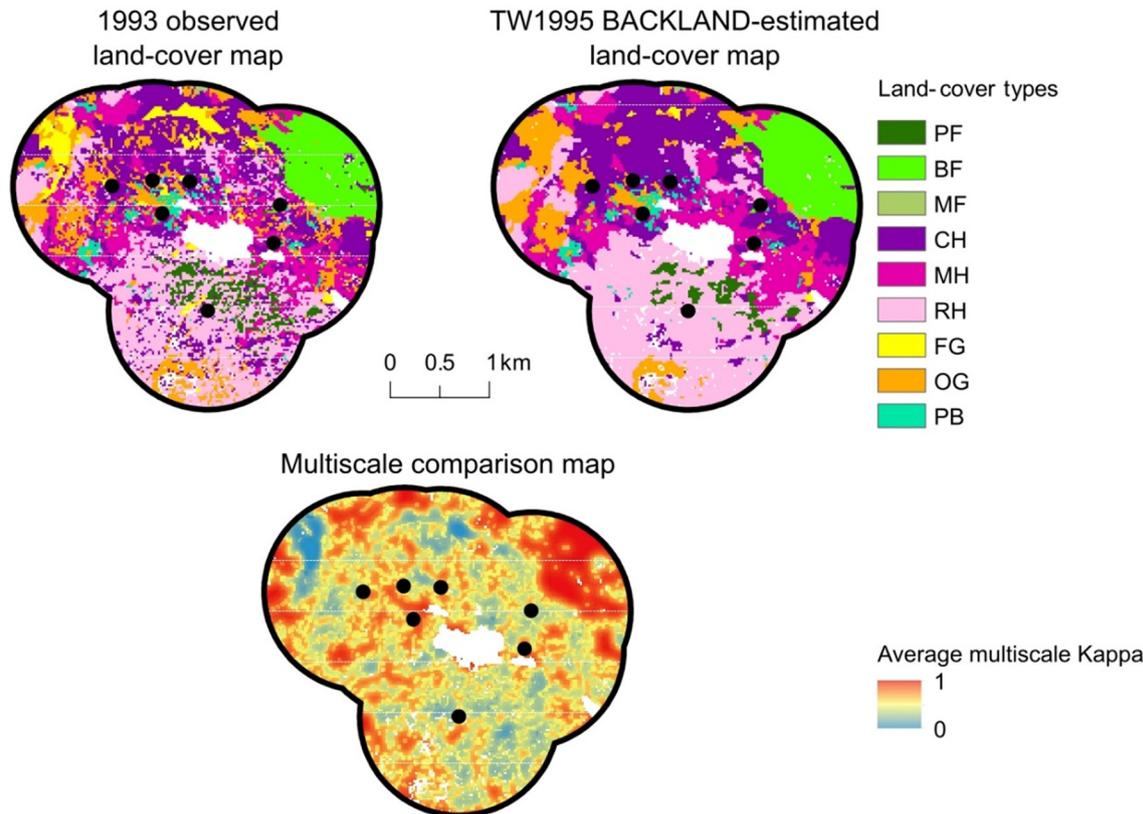


Figure 6. Observed 1993 and BACKLAND-estimated TW1995 land-cover maps and their corresponding multiscale comparison map.

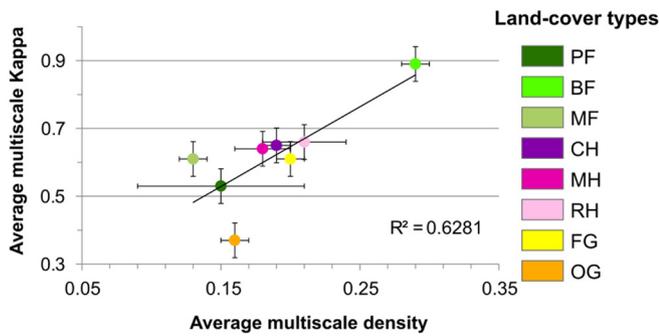


Figure 7. Correlation between the 1993 observed land-cover average densities and their BACKLAND-estimated distribution accuracy (average multiscale Kappa between the 1993 observed and the BACKLAND-estimated TW1995 land-cover maps).

the relevant source areas of pollen (RSAPs), which is essential for the BACKLAND approach and cannot be identified with raw pollen data, and because they allow a more accurate reconstruction than the latter (Hellman et al. 2008, Sugita et al. 2010b, Overballe-Petersen et al. 2013, Mazier et al. 2015, Marquer et al. 2020b, Plancher et al. 2022). However, certain conditions are necessary for the application of the LRA approach (Sugita 2007b), and therefore for the application of BACKLAND. In particular, it is necessary to have the input parameters of the LRA, including pollen fall speeds, estimates of the relative pollen productivity of key taxa and their standard errors. These values exist for a large number of African, Asiatic and European plant taxa (Duffin and Bunting 2008, Bunting et al. 2013, Li et al. 2018, Wieczorek and Herzschuh 2020, Gaillard et al. 2021, Githumbi et al. 2022), although the data are currently mainly provided for central/northern European and Chinese plant taxa (Wieczorek and Herzschuh 2020). Furthermore, the BACKLAND approach relies on the redundancy of information between LOVE estimates to provide continuous maps of LOVE-inferred taxa distribution. It is therefore necessary to target a study area comprising sedimentary sites that are sufficiently close to each other so that their RSAPs overlap as much as possible. Therefore areas with a dense network of lakes and peat bogs, such as mountain areas with a bedrock of magmatic rocks or boreal regions (Peiry 2015), should be particularly suitable if well-preserved sediments and reliable chronologies are available.

In addition, BACKLAND requires a pair of recent and past land-cover maps characterising the land-cover mosaic of the studied landscape with the same land-cover type classification. The past land-cover map must be within the time-window of the LOVE estimates. The spatial resolution and classification of land-cover types achieved by BACKLAND reconstructions depends on their definition on the land-cover maps used for training. Moreover, the more the landscape contains rare or highly fragmented land-cover types, the finer the spatial resolution required to integrate these types into the models, which can generate excessive computation times. In this study, a spatial resolution of 20 m was a reasonable compromise between number of land-cover types and computation times. Additional maps may also be needed to

extract auxiliary variables depending on the studied landscapes and the environmental constraints influencing their vegetation distribution. A digital elevation model was used here to extract altitude, exposure, slope and curvature variables. The distance to the nearest water point, a significant auxiliary variable for all land-cover types here, was extracted from the recent land-cover map, but other sources providing information on the hydrographic network can be considered.

Finally, botanical data should provide information on the recent floristic composition of the land-cover types. The species reported by these inventories must be converted according to the key pollen taxa modelled by LOVE, which are considered to represent 100% of the vegetation cover. These data are then combined in different steps by several simple methods to produce the backward landscape prediction, detailed in the next section.

#### Originality and assumptions of the BACKLAND approach

In stage A of the approach (Fig. 3a; section 'Preliminary stage: map-inferred taxon distributions in 2008 and LOVE-inferred taxon distributions at TW1995'), the recent land-cover map and botanical data are combined into map-inferred taxon distributions (Fig. 4a) with the Multiscale heterogeneity map software (MHM, Gaucherel 2007). MHM density analysis takes into account the heterogeneity and the multiscale distribution of species (Chen et al. 2005, Gaucherel et al. 2007, Lemly and Cooper 2011, Dray et al. 2012, Viers et al. 2012). In Bassiès, plant taxa are mainly ubiquitous and are found in various proportions in several land-cover types, thus justifying the use of a multiscale tool. By combining botanical data and multiscale density maps, we assume that land-cover densities influence linearly and with equal weight the taxon proportion for the production of map-inferred taxon distributions. In the absence of any contradictory evidence, this seems to be a reasonable assumption. These map-inferred distributions form one of the two categories of intermediate taxon distribution maps and are used in stage B of the BACKLAND approach to retroactively integrate spatiotemporal autocorrelation into the distribution of taxa estimated at the past time window. The other category of intermediate distribution maps is LOVE-inferred taxon distributions (Fig. 4b), which are a main originality of the BACKLAND approach. They are formed by the combination of PDFs (Kühl et al. 2002) with LOVE estimates from nearby sedimentary sites. For this step, we assumed a bivariate Gaussian distribution of taxon occurrence probability around pollen sites, with 99% probability that a taxon is present within the RSAPs. While elevation co-kriging has been advocated for the spatial interpolation of LRA estimates beyond RSAPs (O'Dwyer et al. 2021), these two assumptions in line with the Gaussian dispersion of pollen grains around the study sites (Prentice 1985) make the method proposed here both more parsimonious and more suitable for LOVE spatialisation when RSAPs overlap. Moreover, this method revealed smooth LOVE compositional gradients, which would have been difficult with kriging as the latter performs poorly when it is necessary to extrapolate estimates beyond the sedimentary

sites over the entire study area (not shown). PDFs are already used in palaeoecology at regional to continental scales to infer broad and long-term vegetation changes based on pollen percentage data (Hély and Lézine 2014, Ledru et al. 2016). Here, they allowed the construction of distribution maps of pollen taxa whose abundance gradients are determined both by variations in their LOVE estimates between each site and by their probability of occurrence within the 9 km<sup>2</sup> study area, taking into account the overlapping RSAPs. Such local gradients would not have been revealed using raw pollen percentages that show less spatial variability than the LOVE estimates (Plancher et al. 2022). The LOVE-inferred distributions are intermediate results that could also be of interest for studies on past plant distributions.

In stage B (Fig. 3b; section 'Variables used by the BACKLAND approach'), we used MHM density maps to smooth past (1993) land-cover distributions and make their spatial variability comparable to continuous explanatory variables (Fig. 3b). The land-cover explanatory variables included environmental variables as well as a set of estimated taxa distributions at TW1995, according to the botanical composition of each land-cover type. The choice of auxiliary environmental variables depends on the study area and it could be relevant to integrate others, such as pedological data not available in our area. The TW1995 estimated taxon distributions are the result of averaging TW1995 LOVE-inferred distributions and 2008 Map-inferred distributions to incorporate the land-cover spatiotemporal autocorrelation directly into explanatory variables. This straightforward and exploratory way to integrate spatiotemporal autocorrelation resulted in estimated distributions consistent with 2008 taxon distributions and that revealed TW1995 LOVE-based gradients (Fig. 4c), while using residual autocorrelation (Cruse et al. 2012) would drive the results towards the sole landscape autocorrelation, preventing LOVE-based variations from being revealed (not shown). The production of estimated taxon distributions is based on two assumptions. First, due to the integration of map-inferred distributions, it is assumed that the same land-cover types were present during the study period and that their composition remained constant. This imposes a limited temporal perspective on the application of BACKLAND, since the older the time window targeted, the less reasonable this assumption is. Then, the incorporation of spatiotemporal autocorrelation in the estimated taxon distributions assumes an influence of equal weight between 2008 map-inferred and TW1995 LOVE-inferred distributions through averaging. The relative influence of map-inferred and LOVE-inferred distributions could be estimated based on independent data indicating environmental or societal changes that may have sharply modified vegetation composition and/or patterns, and thus the influence of autocorrelation between two target periods (e.g. proxies related to local fire events or grazing activities). There is currently no way of estimating such weighting factors, so their use would require additional assumptions about the stationarity of taxon distributions over time. Our approach, although involving a strong and arbitrary hypothesis, remains parsimonious. In our case

study in particular, the recent land-cover maps used to implement BACKLAND are very similar (Fig. 2), so it is unlikely that the integration of spatial autocorrelation with the 2008 map-inferred distribution would have a strong impact on the BACKLAND estimate of TW1995. It would be interesting to implement the BACKLAND approach on study areas that have recently undergone significant landscape changes, and which are informed by land-cover maps, in order to assess the influence of the integration of map-inferred taxon distributions.

In stage C (Fig. 3c, section 'Backward landscape modelling'), traditional and easy-to-use multiple linear models were employed for the backward modelling of a TW1995 land-cover map. We therefore assumed a linear relationship between land-cover type densities and explanatory variables. Such conditions are not fully met in our datasets, but large sample sizes limit departures from this assumption. Transforming the data or using more complex and non-linear models (e.g. generalized additive models, Hastie and Tibshirani 1987) did not change the quality of the results (not shown). Untransformed land-cover type densities, as well as linear models less prone to overfitting to the training data than other less straightforward or non-linear models, were therefore preferred, especially since they revealed a good fit to the predictors and an overall high accuracy (Fig. 5). First, the stepwise regression and cross-validation training of the linear models identified statistically significant relationships between all 1993 land-cover density maps and explanatory variables (Table 4). The estimation of negative coefficients associated with some estimated taxon distributions, despite their integration on a botanical basis, is evidence of the taxon distribution heterogeneity, as most taxa are found in all types in varying proportions (Table 4). The significant influence of environmental variables on the distribution of all land-cover types reflects the environmental constraints in the Bassiès vegetation distribution. Nevertheless, the signs of these environmental influences must be interpreted with caution, as statistical relationships can be estimated by the models through interactions with other variables not included in the models. Here, the positive influence of altitude and southern exposure on the broadleaved forest (BF) patch density could be due to the presence of scree (a non-pollen-producing area excluded from analyses) located on the northern slope underlying LEG, thus reducing the density of the forest patch mainly in its lowest part. The variability in prediction accuracy between land-cover types revealed a weakness in modelling the most heterogeneous and rarest ones in the study area (here mixed forest, pine forest and *Festuca* grasslands, Fig. 5b, 6–7). The greater difference in spatial variability between explanatory variables and rare land-cover type distributions (which have low densities across the whole landscape) made it more difficult to establish a linear relationship between them than for the more abundant types in the study area. We thus expect BACKLAND to have difficulty representing the rarest and most fragmented land-cover types. This difficulty could explain why the TW1995 BACKLAND-estimated map is more similar to the 2008 map than to the 1993 map. Indeed,

the main difference between the two remote-sensed maps is the reduction in the patch of *Festuca* grassland to the north of the study area (Fig. 2). Since BACKLAND performs poorly in representing this type of land-cover, the map estimated at TW1995 is slightly more similar to the 2008 map.

The ideal landscape on which this approach could be applied would be a land-cover mosaic with high equitability and aggregation indexes, thus avoiding representation biases that disadvantage the sparsest or most fragmented land-cover types. Nevertheless, all things considered, BACKLAND produced a land-cover map representing the eight targeted land-cover types with a strong similarity with the reference map of 1993 (average Kappa = 0.65 > 0.6, Landis and Koch 1977), attesting the potential of this approach for the reconstruction of past land-cover mosaics. Such precision in terms of nomenclature and spatial resolution together with simple and parsimonious working assumptions make the BACKLAND method original and fully complementary to previous attempts at pollen-based spatially explicit land-cover reconstructions.

Finally, uncertainties arising from both the datasets used (botanical data, LOVE estimates, maps) and the methods employed (PDFs, linear models) are present at each stage of the learning process and will accumulate when applying BACKLAND on past time-windows. This methodological development should be completed by an uncertainty propagation analysis in order to be able to assess the uncertainty of the estimated land-cover maps.

### Comparison with previous pollen-based land-cover maps

Based on LOVE estimates of local vegetation composition, using a combination of simple probabilistic and statistical tools, and relying on reasonable assumptions, the BACKLAND method represents a new way of exploring past land-cover mosaics that is complementary to existing approaches. Other approaches proposed land-cover maps based on pollen data and pollen dispersal and deposition. With the multiple scenario approach (MSA, Bunting and Middleton 2009, Bunting et al. 2018, HUMPOL, Middleton and Bunting 2004, Bunting and Middleton 2005), plausible landscape scenarios are selected on the basis of the best similarities between modelled and empirical pollen data: LOVE data are therefore not used directly for vegetation spatialization, despite their potential to reflect inter-site variations (Overballe-Petersen et al. 2013, Plancher et al. 2022), and map successions do not take account of spatiotemporal autocorrelation in taxon distribution. Due to the multiplicity of plausible scenarios, the analysis of landscape dynamics proposed by the MSA is complex. In BACKLAND, LOVE estimates are used directly in the production of taxon distribution maps. LOVE estimates therefore play a role in both the estimation of land-cover composition and the distribution of land-cover types estimated by the models, thus limiting the assumptions on taxon location, and spatiotemporal autocorrelation ensures plausible continuity in the estimated landscape dynamics. Although LOVE estimates

have previously been used to directly produce taxon distribution maps via interpolation techniques (O'Dwyer et al. 2021), BACKLAND has the advantage of basing taxon distributions on Gaussian PDFs of taxon occurrence consistent with the LRA assumptions (Prentice 1985, Sugita 2007b). Compared with the smooth distributions produced by interpolation techniques (O'Dwyer et al. 2021), and because BACKLAND integrates spatiotemporal autocorrelation, taxon distribution maps estimated by BACKLAND are both more spatially heterogeneous (and thus more realistic) and less impacted by potential pollen biases persisting after LRA modelling – which may, for example, be due to changes in the structure of the vegetation close to the sites rather than in the composition (Sugita et al. 2010a). The use of multiple linear models results in the production of a single land-cover map, greatly facilitating the interpretability of the results, particularly for future BACKLAND applications for long-term land-cover mosaic successions, which is difficult when multiple scenarios are proposed (Caseldine et al. 2008, Bunting et al. 2018). Finally, using linear models does not constrain the direction of possible transitions between land-cover types from one period to another, unlike the use of Markov chain models also used for the reconstruction of a spatially explicit landscape within a lake RSAP (Poska et al. 2008).

### Implications for environmental sciences

BACKLAND has an interesting potential to reconstruct long-term land-cover mosaic dynamics in anthropogenic contexts, where reverse transitions to natural vegetation succession can occur. Indeed, once BACKLAND has been implemented over a recent period, it is intended to be easily applied retroactively and step-by-step to produce a succession of land-cover maps based on pollen data (as long as the study area meets the conditions outlined in section 'Application conditions and required data'). Retroactive application only requires 1) estimating the new map-inferred taxon distributions from the recent land-cover map and 2) calculating the LOVE-inferred taxon distributions corresponding to the targeted past time period to 3) obtain their corresponding estimated distributions. Obtaining the new estimated distributions, associated with the explanatory environmental variables, thus appears to be sufficient to predict the land-cover mosaic of the targeted past period. In theory, this process can be repeated step-by-step to produce a temporal atlas of land cover spanning several centuries (Plancher et al. unpubl.). By producing a continuous series of maps using the same approach, it offers the possibility to study long-term land-cover composition and configuration changes without the issues of nomenclature changes and resolution degradation.

The BACKLAND approach is a static modeling approach, and the integrated explanatory variables do not include socio-economic, ecological or climatic drivers whose variation can influence the land-cover spatial and temporal variability. However, by integrating and spatialising the LOVE estimates, BACKLAND indirectly integrates the effect of these drivers on vegetation composition and configuration. Future studies

would be needed to compare the outputs of this static diachronic approach with those of other types of pattern-based (Houet et al. 2012a, b) or process-based (Gaucherel and Pommereau 2019, Gaucherel et al. 2020, Cosme et al. 2022) modeling approaches, in order to evaluate and interpret the convergences and divergences of the modeled trajectories. Such comparisons will allow a better overall understanding of the Bassiès socio-ecosystem dynamics (Houet et al. 2010b, Gritti et al. 2013).

Better understanding and managing cultural landscapes, characterizing their ecosystem service dynamics, being able to predict their trajectories following land-use and land-cover changes, and improving their management strategies are research priorities involving a close connection between ecology and palaeoecology (Rull 2014, Seddon et al. 2014). Indeed, the few decades covered by ecological studies are not enough to fully integrate ecological processes (Jeffers et al. 2015). Unlike ecology, paleoecology allows the integration of slow processes, but palaeoecologists must make an effort to ensure that paleoecological data are suited to the needs of ecologists, especially in terms of spatial resolution and extent of reconstructions (Rull 2014, Birks 2019). Ecological processes evolve over time under the influence of the spatial context in which they operate (Ricklefs 1987, Leibold et al. 2004), but there are still few approaches that are capable of integrating both the temporal and spatial aspects of ecosystem dynamics over the long term (White et al. 2010). This study represents progress in the conciliation between paleoecology and general ecology. By increasing the temporal extent of landscape change studies as well as their precision in terms of nomenclature and spatial resolution of land-cover reconstructions, BACKLAND will help to improve our understanding of the legacy of land-cover change on biodiversity at several scales (alpha, beta diversities; Rosenzweig 1995, Duelli 1997, Wiens 2009, Zimmermann et al. 2010, Tschardt et al. 2012, Woodbridge et al. 2020), to assess the responses and feedbacks of vegetation to global change (Turner 1994, Turner et al. 2007), and to refine studies on species autecology (Abraham et al. 2023).

## Conclusion

Maps of past land-cover mosaic provide essential information related to the ecological state of land areas and how they have been modified by humans. Hence, accurate information related to past land cover is essential both for managing natural resources and for understanding society's ecological, biophysical, and resource management footprint. In this paper we describe a new approach based on LOVE estimates of neighbouring sites with overlapping RSAPs, cartographic and botanical data, and parsimonious statistical tools, to backward estimate land-cover maps with a 20 m spatial resolution. The approach has been tested on a well-documented area in terms of available pollen and cartographic data. Its accuracy was assessed on a recent time window, revealing a high similarity between the observed and estimated maps.

It makes BACKLAND a promising approach to provide fine-grained reconstruction of heterogeneous land-cover mosaics. By integrating spatiotemporal autocorrelation in estimated taxon distributions, BACKLAND is suitable for exploring long-term land-cover dynamics. Moving forward, we will apply the method to fossil pollen data from consecutive time windows over the last 200 years in the Bassiès area.

*Acknowledgements* – Special thanks are due to Elizabeth Rowley-Jolivet for proofreading the article.

*Funding* – This paper forms part of C. Plancher's doctoral thesis cosupervised by F. Mazier and D. Galop and funded by the University Toulouse 2 Jean Jaures. The data were funded and provided by the ANR JCJC TRAM (ANR-15-CE01-0008) (Trace metal legacy on mountains aquatic ecogeochemistry) and ANR-10-JCJC-1804 MODE-RESPYR (Modelling Past and future land cover changes in the Pyrenees) projects coordinated respectively by G. Le Roux and T. Houet, the PEPS CNRS BIOPAST and POPEYE projects coordinated respectively by D. Galop and F. Mazier. This work was carried out within the framework of the Observatoire Homme-Milieu Pyrénées Haut Vicdessos (headed by D. Galop) and (co)funded by the ANR under the LabEx DRIIHM ANR-11-LABX0010 as part of the DYVAH, GEOSPAT and PASTSERV projects.

## Author contributions

**Clara Plancher:** Conceptualization (equal); Formal analysis (lead); Investigation (equal); Methodology (lead); Software (lead); Validation (lead); Visualization (lead); Writing – original draft (lead). **Florence Mazier:** Conceptualization (equal); Funding acquisition (lead); Investigation (equal); Project administration (lead); Resources (equal); Supervision (equal); Writing – review and editing (equal). **Thomas Houet:** Conceptualization (equal); Resources (supporting); Writing – review and editing (equal). **Cédric Gaucherel:** Conceptualization (equal); Formal analysis (supporting); Investigation (supporting); Methodology (supporting); Resources (equal); Software (supporting); Supervision (lead); Validation (supporting); Writing – review and editing (equal).

## Transparent peer review

The peer review history for this article is available at <https://publons.com/publon/10.1111/ecog.06853>.

## Data availability statement

There are no additional data for this paper.

## References

- Abraham, V., Man, M., Theuerkauf, M., Pokorný, P., Bobek, P. and Novák, J. 2023. Spatially explicit, quantitative reconstruction of past vegetation based on pollen or charcoal data as a tool for autecology of trees. – *Landscape Ecol.* 38: 1747–1763.
- Bennett, N. D., Croke, B. F. W., Guariso, G., Guillaume, J. H. A., Hamilton, S. H., Jakeman, A. J., Marsili-Libelli, S., Newham, L. T. H., Norton, J. P., Perrin, C., Pierce, S. A., Robson, B., Seppelt, R., Voinov, A. A., Fath, B. D. and Andreassian, V.

2013. Characterising performance of environmental models. – *Environ. Modell. Softw.* 40: 1–20.
- Birks, H. J. B. 2019. Contributions of quaternary botany to modern ecology and biogeography. – *Plant Ecol. Divers.* 12: 189–385.
- Bunting, M. J. and Middleton, D. 2005. Modelling pollen dispersal and deposition using HUMPOL software, including simulating windroses and irregular lakes. – *Rev. Palaeobot. Palynol.* 134: 185–196.
- Bunting, M. J. and Middleton, R. 2009. Equifinality and uncertainty in the interpretation of pollen data: the multiple scenario approach to reconstruction of past vegetation mosaics. – *Holocene* 19: 799–803.
- Bunting, M. J., Schofield, J. E. and Edwards, K. J. 2013. Estimates of relative pollen productivity (RPP) for selected taxa from southern Greenland: a pragmatic solution. – *Rev. Palaeobot. Palynol.* 190: 66–74.
- Bunting, M. J., Farrell, M., Bayliss, A., Marshall, P. and Whittle, A. 2018. Maps from mud — using the multiple scenario approach to reconstruct land cover dynamics from pollen records: a case study of two Neolithic landscapes. – *Front. Ecol. Evol.* 6: 1–20.
- Bürgi, M., Hersperger, A. M., Hall, M., Southgate, E. W. B. and Schneeberger, N. 2007. Using the past to understand the present land use and land cover. – In: Kienast, F., Wildi, O. and Ghosh, S. (eds), *A changing world: challenges for landscape research, landscape series*. Springer Netherlands, pp. 133–144.
- Bürgi, M., Östlund, L. and Mladenoff, D. J. 2017. Legacy effects of human land use: ecosystems as time-lagged systems. – *Ecosystems* 20: 94–103.
- Caseldine, C., Fyfe, R. and Hjelle, K. 2008. Pollen modelling, palaeoecology and archaeology: virtualisation and/or visualisation of the past? – *Veg. Hist. Archaeobot.* 17: 543–549.
- Chen, X., Li, B. and Collins, S. L. 2005. Multiscale monitoring of a multispecies case study: two grass species at Seville. – *Plant Ecol.* 179: 149–154.
- Cohen, J. 1960. A coefficient of agreement for nominal scales. – *Educ. Psychol. Meas.* 20: 37–46.
- Cosme, M., Hély, C., Pommereau, F., Pasquariello, P., Tiberi, C., Treydte, A. and Gaucherel, C. 2022. Qualitative modeling for bridging expert-knowledge and social-ecological dynamics of an East African savanna. – *Land* 11: 42.
- Crane, B., Liedloff, A. C. and Wintle, B. A. 2012. A new method for dealing with residual spatial autocorrelation in species distribution models. – *Ecography* 35: 879–888.
- Crooks, K. R. and Sanjayan, M. 2006. *Connectivity conservation*. – Cambridge Univ. Press
- Cui, Q. Y., Gaillard, M. J., Lemdahl, G., Olsson, F. and Sugita, S. 2010. Holocene local forest history at two sites in Småland, southern Sweden—insights from quantitative reconstructions using the landscape reconstruction algorithm. – *Geophys. Res. Abstr.* 12: 593–611.
- Cui, Q. Y., Gaillard, M. J., Lemdahl, G., Sugita, S., Greisman, A., Jacobson, G. L. and Olsson, F. 2013. The role of tree composition in Holocene fire history of the hemiboreal and southern boreal zones of southern Sweden, as revealed by the application of the Landscape Reconstruction Algorithm: implications for biodiversity and climate-change issues. – *Holocene* 23: 1747–1763.
- Cui, Q. Y., Gaillard, M. J., Lemdahl, G., Stenberg, L., Sugita, S. and Zernova, G. 2014. Historical land-use and landscape change in southern Sweden and implications for present and future biodiversity. – *Ecol. Evol.* 4: 3555–3570.
- Dahlström, A. 2008. Grazing dynamics at different spatial and temporal scales: examples from the Swedish historical record A.D. 1620–1850. – *Veg. Hist. Archaeobot.* 17: 563–572.
- Dearing, J. A. et al. 2015. Social-ecological systems in the Anthropocene: the need for integrating social and biophysical records at regional scales. – *Anthr. Rev.* 2: 220–246.
- Dray, S., Pélissier, R., Couteron, P., Fortin, M.-J., Legendre, P., Peres-Neto, P. R., Bellier, E., Bivand, R., Blanchet, F. G., De Cáceres, M., Dufour, A.-B., Heegaard, E., Jombart, T., Munoz, F., Oksanen, J., Thioulouse, J. and Wagner, H. H. 2012. Community ecology in the age of multivariate multiscale spatial analysis. – *Ecol. Monogr.* 82: 257–275.
- Duelli, P. 1997. Biodiversity evaluation in agricultural landscapes: an approach at two different scales. – *Agric. Ecosyst. Environ. Biodivers. Agric. Sustain. Future* 62: 81–91.
- Duffin, K. I. and Bunting, M. J. 2008. Relative pollen productivity and fall speed estimates for southern African savanna taxa. – *Veg. Hist. Archaeobot.* 17: 507–525.
- Fischer, J., Hanspach, J. and Hartel, T. 2011. Continental-scale ecology versus landscape-scale case studies. – *Front. Ecol. Environ.* 9: 430–430.
- Forman, R. T. T. 1995. *Land mosaics: the ecology of landscapes and regions*. – Cambridge Univ. Press
- Foster, D., Swanson, F., Aber, J., Burke, I., Brokaw, N., Tilman, D. and Knapp, A. 2003. The importance of land-use legacies to ecology and conservation. – *BioScience* 53: 77–88.
- Fredh, E. D., Lagerås, P., Mazier, F., Björkman, L., Lindbladh, M. and Broström, A. 2019. Farm establishment, abandonment and agricultural practices during the last 1300 years: a case study from southern Sweden based on pollen records and the LOVE model. – *Veg. Hist. Archaeobot.* 285: 529–544.
- Fyfe, R. M. and Woodbridge, J. 2012. Differences in time and space in vegetation patterning: analysis of pollen data from Dartmoor, UK. – *Landscape Ecol.* 27: 745–760.
- Fyfe, R. M., Twiddle, C., Sugita, S., Gaillard, M. J., Barratt, P., Caseldine, C. J., Dodson, J., Edwards, K. J., Farrell, M., Froyd, C., Grant, M. J., Huckerby, E., Innes, J. B., Shaw, H. and Waller, M. 2013. The Holocene vegetation cover of Britain and Ireland: overcoming problems of scale and discerning patterns of openness. – *Quat. Sci. Rev.* 73: 132–148.
- Fyfe, R. M., Woodbridge, J. and Roberts, N. 2015. From forest to farmland: pollen-inferred land cover change across Europe using the pseudobiomization approach. – *Global Change Biol.* 21: 1197–1212.
- Gaillard, M.-J. et al. 2010. Holocene land-cover reconstructions for studies on land cover-climate feedbacks. – *Clim. Past* 6: 483–499.
- Gaillard, M.-J., Githumbi, E., Achoundong, G., Lézine, A.-M., Hély, C., Lebamba, J., Marquer, L., Mazier, F., Li, F. and Sugita, S. 2021. The challenge of pollen-based quantitative reconstruction of Holocene plant cover in tropical regions: a pilot study in Cameroon. – *Quaternary Vegetation Dynamics – The African Pollen Database*, pp. 183–206.
- Garbarino, M., Morresi, D., Urbinati, C., Malandra, F., Motta, R., Sibona, E. M., Vitali, A. and Weisberg, P. J. 2020. Contrasting land use legacy effects on forest landscape dynamics in the Italian Alps and the Apennines. – *Landscape Ecol.* 35: 2679–2694.
- Gaucherel, C. 2007. Multiscale heterogeneity map and associated scaling profile for landscape analysis. – *Landscape Urban Plan.* 82: 95–102.
- Gaucherel, C. and Pommereau, F. 2019. Using discrete systems to exhaustively characterize the dynamics of an integrated ecosystem. – *Methods Ecol. Evol.* 10: 1615–1627.

- Gaucherel, C., Burel, F. and Baudry, J. 2007. Multiscale and surface pattern analysis of the effect of landscape pattern on carabid beetles distribution. – *Ecol. Indic.* 7: 598–609.
- Gaucherel, C., Alleaume, S. and Hély, C. 2008. The comparison map profile method : a strategy for multiscale comparison of quantitative and qualitative images. – *IEEE Trans. Geosci. Remote Sens.* 46: 2708–2719.
- Gaucherel, C., Vezy, R., Gontrand, F., Bouchet, D. and Ramesh, B. R. 2016. Spatial analysis of endemism to redefine conservation areas in Western Ghats (India). – *J. Nat. Conserv.* 34: 33–41.
- Gaucherel, C., Tramier, C., Devictor, V., Svenning, J. C. and Hély, C. 2018. Where and at which scales does the LDG fail? – *J. Biogeogr.* 45: 1905–1916.
- Gaucherel, C., Pommereau, F. and Hély, C. 2020. Understanding ecosystem complexity via application of a process-based state space rather than a potential surface. – *Complexity* 2020: 1–14.
- Giesecke, T., Brewer, S., Finsinger, W., Leydet, M. and Bradshaw, R. H. W. 2017. Patterns and dynamics of European vegetation change over the last 15 000 years. – *J. Biogeogr.* 44: 1441–1456.
- Gillet, F., Mauchamp, L., Badot, P. M. and Mouly, A. 2016. Recent changes in mountain grasslands: a vegetation resampling study. – *Ecol. Evol.* 6: 2333–2345.
- Githumbi, E., Fyfe, R., Gaillard, M.-J., Trondman, A.-K., Mazier, F., Nielsen, A.-B., Poska, A., Sugita, S., Woodbridge, J., Azuara, J., Feurdean, A., Grindean, R., Lebreton, V., Marquer, L., Nebout-Combourieu, N., Stančikaitė, M., Tanțău, I., Tonkov, S., Shumilovskikh, L. and LandClimII Data Contributors. 2022. European pollen-based REVEALS land-cover reconstructions for the Holocene: methodology, mapping and potentials. – *Earth Syst. Sci. Data* 14: 1581–1619.
- Gritti, E. S., Gaucherel, C., Crespo-Perez, M. V. and Chuine, I. 2013. How can model comparison help improving species distribution models? – *PLoS One* 8: e68823.
- Guisan, A. and Zimmermann, N. E. 2000. Predictive habitat distribution models in ecology. – *Ecol. Modell.* 135: 147–186.
- Hastie, T. and Tibshirani, R. 1987. Generalized additive models: some applications. – *J. Am. Stat. Assoc.* 82: 371–386.
- Hastie, T., Tibshirani, R. and Friedman, J. 2009. Cross-validation. – In: *The elements of statistical learning. data mining, inference, and prediction.* Springer Science & Business Media, pp. 241–259.
- Hellman, S., Gaillard, M.-J., Broström, A. and Sugita, S. 2008. The REVEALS model, a new tool to estimate past regional plant abundance from pollen data in large lakes : validation in southern Sweden. – *J. Quat. Sci.* 23: 21–42.
- Hély, C., Lézine, A.-M. and contributors, A. 2014. Holocene changes in African vegetation : tradeoff between climate and water availability. – *Clim. Past* 10: 681–686.
- Herrault, P.-A., Larrieu, L., Cordier, S., Gimmi, U., Lachat, T., Ouin, A., Sarthou, J.-P. and Sheeren, D. 2015. Combined effects of area, connectivity, history and structural heterogeneity of woodlands on the species richness of hoverflies (Diptera: Syrphidae). – *Landscape Ecol.* 31: 877–893.
- Hjelle, K. L. and Sugita, S. 2012. Estimating pollen productivity and relevant source area of pollen using lake sediments in Norway: how does lake size variation affect the estimates? – *Holocene* 22: 313–324.
- Hjelle, K. L., Mehl, I. K., Sugita, S. and Andersen, G. L. 2015. From pollen percentage to vegetation cover: evaluation of the Landscape Reconstruction Algorithm in western Norway. – *J. Quat. Sci.* 30: 312–324.
- Houet, T., Loveland, T. R., Hubert-Moy, L., Gaucherel, C., Napton, D., Barnes, C. A. and Saylor, K. 2010a. Exploring subtle land use and land cover changes: a framework for future landscape studies. – *Landscape Ecol.* 25: 249–266.
- Houet, T., Verburg, P. H. and Loveland, T. R. 2010b. Monitoring and modelling landscape dynamics. – *Landscape Ecol.* 25: 163–167.
- Houet, T., Vacquié, L., Vidal, F. and Galop, D. 2012. Caractérisation de la fermeture des paysages dans les Pyrénées depuis les années 1940. Application sur le Haut-Vicdessos. – *Sud-Ouest Eur.* 33: 41–56.
- Hultberg, T., Gaillard, M.-J., Grundmann, B. and Lindbladh, M. 2014. Reconstruction of past landscape openness using the Landscape Reconstruction Algorithm (LRA) applied on three local pollen sites in a southern Swedish biodiversity hotspot. – *Leg. Hist. Archaeobot.* 242: 253–266.
- Jeffers, E. S., Nogué, S. and Willis, K. J. 2015. The role of palaeoecological records in assessing ecosystem services. – *Quat. Sci. Rev.* 112: 17–32.
- Joannin, S., Brugiapaglia, E., De Beaulieu, J.-L., Bernardo, L., Magny, M., Peyron, O., Goring, S. and Vannièrre, B. 2012. Pollen-based reconstruction of Holocene vegetation and climate in southern Italy: the case of Lago Trifoglietti. – *Clim. Past* 8: 1973–1996.
- Jolly, D. et al. 1998. Biome reconstruction from pollen and plant macrofossil data for Africa and the Arabian Peninsula at 0 and 6000 years. – *J. Biogeogr.* 25: 1007–1027.
- Kienast, F., Wildi, O. and Ghosh, S. (eds) 2007. A changing world: challenges for landscape research, ed. landscape series. – Springer Science&Business Media
- Kühl, N., Gebhardt, C., Litt, T. and Hense, A. 2002. Probability density functions as botanical-climatological transfer functions for climate reconstruction. – *Quat. Res.* 58: 381–392.
- Landis, J. R. and Koch, G. G. 1977. The measurement of observer agreement for categorical data. – *Biometrics* 33: 159.
- Le Provost, G., Badenhauer, I., Le Bagousse-Pinguet, Y., Clough, Y., Henckel, L., Violle, C., Bretagnolle, V., Roncoroni, M., Manning, P. and Gross, N. 2020. Land-use history impacts functional diversity across multiple trophic groups. – *Proc. Natl Acad. Sci. USA* 117: 1573–1579.
- Ledru, M., Montade, V., Cedex, M., Cedex, M. and Pratique, E. 2016. Long-term spatial changes in the distribution of the Brazilian Atlantic forest. – *Biotropica* 48: 159–169.
- Legendre, P. and Fortin, M. J. 1989. Spatial pattern and ecological analysis. – *Vegetatio* 80: 107–138.
- Leibold, M. A., Holyoak, M., Mouquet, N., Amarasekare, P., Chase, J. M., Hoopes, M. F., Holt, R. D., Shurin, J. B., Law, R., Tilman, D., Loreau, M. and Gonzalez, A. 2004. The meta-community concept: a framework for multi-scale community ecology. – *Ecol. Lett.* 7: 601–613.
- Leite, M. D. S., Tambosi, L. R., Romitelli, I. and Metzger, J. P. 2013. Landscape ecology perspective in restoration projects for biodiversity conservation: a review. – *Braz. J. Nat. Conserv* 11: 108–118.
- Lemly, J. M. and Cooper, D. J. 2011. Multiscale factors control community and species distribution in mountain peatlands. – *Botany* 89: 689–713.
- Li, F., Gaillard, M. J., Xu, Q., Bunting, M. J., Li, Y., Li, J., Mu, H., Lu, J., Zhang, P., Zhang, S., Cui, Q., Zhang, Y. and Shen, W. 2018. A review of relative pollen productivity estimates from temperate China for pollen-based quantitative reconstruction of past plant cover. – *Front. Plant Sci.* 9: 1214.

- Marquer, L., Mazier, F., Sugita, S., Galop, D., Houet, T., Faure, E., Gaillard, M. J., Haunold, S., de Munnik, N., Simonneau, A., De Vleeschouwer, F. and Le Roux, G. 2020a. Reply to Theuerkauf and Couwenberg (2020) comment on: "Pollen-based reconstruction of Holocene land-cover in mountain regions: evaluation of the Landscape Reconstruction Algorithm in the Vicdessos valley, northern Pyrenees, France". – *Quat. Sci. Rev.* 244: 106463.
- Marquer, L., Mazier, F., Sugita, S., Galop, D., Houet, T., Faure, E., Gaillard, M.-J., Haunold, S., de Munnik, N. D., Simonneau, A., De Vleeschouwer, F. and Le Roux, G. 2020b. Pollen-based reconstruction of Holocene land-cover in mountain regions : evaluation of the landscape reconstruction algorithm in the Vicdessos valley, northern Pyrenees, France. – *Quat. Sci. Rev.* 228: 1–14.
- Mazier, F., Broström, A., Bragée, P., Fredh, D., Stenberg, L., Thiery, G., Sugita, S. and Hammarlund, D. 2015. Two hundred years of land-use change in the South Swedish uplands : comparison of historical map-based estimates with a pollen-based reconstruction using the landscape reconstruction algorithm. – *Veg. Hist. Archaeobot.* 24: 555–570.
- Mazier, F., Plancher, C., Haunold, S., Bonnet, C., Belly, H. and Houet, T. 2022. Vegetation surveys in a mountainous area (Pyrenees, France). – GBIF.org, <https://doi.org/10.15468/4q47d3>.
- McCullagh, P. 1984. Generalized linear models. – *Eur. J. Oper. Res.* 16: 285–292.
- Middleton, R. and Bunting, M. J. 2004. Mosaic v1.1: landscape scenario creation software for simulation of pollen dispersal and deposition. – *Rev. Palaeobot. Palynol.* 132: 61–66.
- Neumann, J. L., Holloway, G. J., Hoodless, A. and Griffiths, G. H. 2017. The legacy of 20th Century landscape change on today's woodland carabid communities. – *Divers. Distrib.* 23: 1447–1458.
- Nielsen, A. B. and Sugita, S. 2005. Estimating relevant source area of pollen for small Danish lakes around AD 1800. – *Holocene* 15: 1006–1020.
- Nielsen, A. B. and Odgaard, B. V. 2010. Quantitative landscape dynamics in Denmark through the last three millennia based on the Landscape Reconstruction Algorithm approach. – *Veg. Hist. Archaeobot.* 19: 375–387.
- O'Dwyer, R., Marquer, L., Trondman, A.-K. and Jönsson, A. M. 2021. Spatially continuous land-cover reconstructions through the Holocene in southern Sweden. – *Ecosystems* 24: 1450–1467.
- Opdam, P., Nassauer, J. I., Wang, Z., Albert, C., Bentrup, G., Castella, J.-C., McAlpine, C., Liu, J., Sheppard, S. and Swaffield, S. 2013. Science for action at the local landscape scale. – *Landscape Ecol.* 28: 1439–1445.
- Overballe-Petersen, M. V., Nielsen, A. B. and Bradshaw, R. H. W. 2013. Quantitative vegetation reconstruction from pollen analysis and historical inventory data around a Danish small forest hollow. – *J. Veg. Sci.* 24: 755–771.
- Pärtel, M., Helm, A., Reitalu, T., Liira, J. and Zobel, M. 2007. Grassland diversity related to the Late Iron Age human population density. – *J. Ecol.* 95: 574–582.
- Pavageau, C., Gaucherel, C., Garcia, C. and Ghazoul, J. 2017. Nesting sites of giant honeybees modulated by landscape patterns. – *J. Appl. Ecol.* 55: 1230–1240.
- Peiry, J.-L. 2015. Zones humides et marais d'eau douce. – In: Euzen, A., Jeandel, C. and Mossery, R. (eds), *L'eau à Découvert*. CNRS Editions, pp. 108–109.
- Plancher, C., Galop, D., Houet, T., Lerigoleur, E., Marquer, L., Sugita, S. and Mazier, F. 2022. Spatial and temporal patterns of upland vegetation over the last 200 years in the northern Pyrenees: example from the Bassiès valley, Ariège, France. – *Quat. Sci. Rev.* 294: 107753.
- Poska, A., Sepp, E., Veski, S. and Koppel, K. 2008. Using quantitative pollen-based land-cover estimations and a spatial CA-Markov model to reconstruct the development of cultural landscape at Rouge, south Estonia. – *Veg. Hist. Archaeobot.* 17: 527–541.
- Poska, A., Saarse, L., Koppel, K., Nielsen, A. B., Avel, E., Vassiljev, J. and Väli, V. 2014. The Verijärv area, South Estonia over the last millennium: a high resolution quantitative land-cover reconstruction based on pollen and historical data. – *Rev. Palaeobot. Palynol.* 207: 5–17.
- Prentice, I. C. 1985. Pollen representation, source area, and basin size: toward a unified theory of pollen analysis. – *Quat. Res.* 23: 76–86.
- Prentice, I. C. 1988. Records of vegetation in time and space: the principles of pollen analysis. – In: Huntley, B. and Webb, T. III. (eds), *Vegetation history*, Kluwer Academic Publishers, pp. 17–42.
- Prøsch-Danielsen, L., Prescott, C. and Fredh, E. D. 2020. Land cover and exploitation of upland resources on the Høg-Jæren Plateau, southwestern Norway, over the last 6500 years. – *J. Archaeol. Sci. Rep.* 32: 102443.
- Quintana-Seguí, P., Le Moigne, P., Durand, Y., Martin, E., Habets, F., Baillon, M., Canellas, C., Franchisteguy, L. and Morel, S. 2008. Analysis of near-surface atmospheric variables: validation of the SAFRAN analysis over France. – *J. Appl. Meteorol. Climatol.* 47: 92–107.
- Ricklefs, R. E. 1987. Community diversity: relative roles of local and regional processes. – *Science* 235: 167–171.
- Ripley, B., Venables, B., Bates, D. M., Hornik, K., Gebhardt, A., Firth, D. and Ripley, M. B. 2013. Package "mass". – *Cran R* 538: 113–120.
- Rosenzweig, M. L. 1995. *Species diversity in space and time*. – Cambridge Univ. Press.
- Rull, V. 2014. Time continuum and true long-term ecology: from theory to practice. – *Front. Ecol. Evol.* 2: 75.
- Seddon, A. W. R. et al. 2014. Looking forward through the past: identification of 50 priority research questions in palaeoecology. – *J. Ecol.* 102: 256–267.
- Shcherbakov, M. V., Brebels, A., Shcherbakova, N. L., Tyukov, A. P., Janovsky, T. A. and Kamaev, V. A. 2013. A survey of forecast error measures. – *World Appl. Sci. J.* 24: 171–176.
- Sheeren, D., Ladet, S., Ribiere, O., Raynaud, B., Paegelow, M. and Houet, T. 2012. Assessing land cover changes in the French Pyrenees since the 1940s: a semi-automatic GEOBIA approach using aerial photographs. – In: *Proceedings of AGILE 2012 international conference on geographic information science*, pp. 1940–1942.
- Sugita, S. 1993. A model of pollen source area for an entire basin. – *Quat. Res.* 39: 239–244.
- Sugita, S. 1994. Pollen representation of vegetation in quaternary sediments: theory and method in patchy vegetation. – *J. Ecol.* 82: 881–897.
- Sugita, S. 1998. Quantification of land surfaces cleared of forests during the Holocene - Modern pollen / vegetation / landscape relationships as an aid to the interpretation of pollen data. – *Palaeoclimatic Res., Acad. Sci. Literature* 27: 1–15.
- Sugita, S. 2007a. Theory of quantitative reconstruction of vegetation I: pollen from large sites REVEALS regional vegetation composition. – *Holocene* 17: 229–241.

- Sugita, S. 2007b. Theory of quantitative reconstruction of vegetation II: all you need is LOVE. – *Holocene* 17: 243–257.
- Sugita, S., Hicks, S. and Sormunen, H. 2010a. Absolute pollen productivity and pollen-vegetation relationships in northern Finland. – *J. Quat. Sci.* 25: 724–736.
- Sugita, S., Parshall, T., Calcote, R. and Walker, K. 2010b. Testing the Landscape Reconstruction Algorithm for spatially explicit reconstruction of vegetation in northern Michigan and Wisconsin. – *Quat. Res.* 74: 289–300.
- Szczypta, C., Gascoïn, S., Houet, T., Hagolle, O., Dejoux, J.-F., Vigneau, C. and Fanise, P. 2015. Impact of climate and land cover changes on snow cover in a small Pyrenean catchment. – *J. Hydrol.* 521: 84–99.
- Tappeiner, U., Leitinger, G., Zariņa, A. and Bürgi, M. 2020. How to consider history in landscape ecology: patterns, processes, and pathways. – *Landscape Ecol.* 36: 1–12.
- Tarasov, P. E., Williams, J. W., Andreev, A., Nakagawa, T., Bezrukova, E., Herzschuh, U., Igarashi, Y., Müller, S., Werner, K. and Zheng, Z. 2007. Satellite- and pollen-based quantitative woody cover reconstructions for northern Asia: verification and application to late-Quaternary pollen data. – *Earth Planet. Sci. Lett.* 264: 284–298.
- Tarasov, P. E., Bezrukova, E. V. and Krivonogov, S. K. 2009. Late Glacial and Holocene changes in vegetation cover and climate in southern Siberia derived from a 15 kyr long pollen record from Lake Kotokel. – *Clim. Past Discuss.* 5: 127–151.
- Tobler, W. R. 1970. A computer movie simulating urban growth in the Detroit region. – *Econ. Geogr.* 46: 234–240.
- Tscharntke, T. et al. 2012. Landscape moderation of biodiversity patterns and processes - eight hypotheses. – *Biol. Rev.* 87: 661–685.
- Turner, M. G. 1989. Landscape ecology: the effect of pattern on process. – *Annu. Rev. Ecol. Syst.* 20: 171–197.
- Turner, B. L. 1994. Local faces, global flows: the role of land use and land cover in global environmental change. – *Land Degrad. Dev.* 5: 71–78.
- Turner, M. G., Gardner, R. H. and O'Neill, R. V. 2001. *Landscape ecology in theory and practice: pattern and process.* – Springer
- Turner, B. L., Lambin, E. F. and Reenberg, A. 2007. The emergence of land change science for global environmental change and sustainability. – *Proc. Natl Acad. Sci. USA* 104: 20666–20671.
- Viers, J. H., Fremier, A. K., Hutchinson, R. A., Quinn, J. F., Thorne, J. H. and Vaghti, M. G. 2012. Multiscale patterns of riparian plant diversity and implications for restoration. – *Restor. Ecol.* 20: 160–169.
- White, E. P., Ernest, S. K., Adler, P. B., Hurlbert, A. H. and Lyons, S. K. 2010. Integrating spatial and temporal approaches to understanding species richness. – *Phil. Trans. R. Soc. B* 365: 3633–3643.
- Wieczorek, M. and Herzschuh, U. 2020. Compilation of relative pollen productivity (RPP) estimates and taxonomically harmonised RPP datasets for single continents and Northern Hemisphere extratropics. – *Earth Syst. Sci. Data* 12: 3515–3528.
- Wiens, J. A. 2009. Landscape ecology as a foundation for sustainable conservation. – *Landscape Ecol.* 24: 1053–1065.
- Woodbridge, J., Fyfe, R., Smith, D., Pelling, R., Vareilles, A., Batchelor, R., Bevan, A. and Davies, A. L. 2020. What drives biodiversity patterns? Using long-term multidisciplinary data to discern centennial-scale change. – *J. Ecol.* 109: 1365.
- Wu, J. 2013. Key concepts and research topics in landscape ecology revisited: 30 years after the Allerton Park workshop. – *Landscape Ecol.* 28: 1–11.
- Zimmermann, N. E., Washington-Allen, R. A., Ramsey, R. D., Schaeppman, M. E., Mathys, L., Kötz, B., Kneubühler, M. and Edwards, T. C. 2007. Modern remote sensing for environmental monitoring of landscape states and trajectories. – In: Kienast, F., Wildi, O., Ghosh, S. (eds), *A changing world, challenges for landscape research, landscape series*, pp. 65–91.
- Zimmermann, P., Tasser, E., Leitinger, G. and Tappeiner, U. 2010. Effects of land-use and land-cover pattern on landscape-scale biodiversity in the European Alps. – *Ecosyst. Environ.* 139: 13–22.