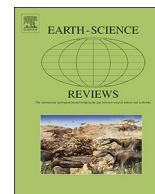




ELSEVIER

Contents lists available at ScienceDirect

## Earth-Science Reviews

journal homepage: [www.elsevier.com/locate/earscirev](http://www.elsevier.com/locate/earscirev)

## Review Article

## Pollen-based climate reconstruction techniques for late Quaternary studies

Manuel Chevalier<sup>a,\*</sup>, Basil A.S. Davis<sup>a</sup>, Oliver Heiri<sup>b</sup>, Heikki Seppä<sup>c</sup>, Brian M. Chase<sup>d</sup>, Konrad Gajewski<sup>e</sup>, Terri Lacourse<sup>f</sup>, Richard J. Telford<sup>g</sup>, Walter Finsinger<sup>h</sup>, Joël Guiot<sup>i</sup>, Norbert Kühl<sup>j</sup>, S. Yoshi Maezumi<sup>k,l</sup>, John R. Tipton<sup>m</sup>, Vachel A. Carter<sup>n</sup>, Thomas Brussel<sup>o</sup>, Leanne N. Phelps<sup>a,p,q</sup>, Andria Dawson<sup>r</sup>, Marco Zanon<sup>s</sup>, Francesca Vallé<sup>t</sup>, Connor Nolan<sup>u</sup>, Achille Mauri<sup>v</sup>, Anne de Vernal<sup>w</sup>, Kenji Izumi<sup>x,y</sup>, Lasse Holmström<sup>z</sup>, Jeremiah Marsicek<sup>aa</sup>, Simon Goring<sup>ab</sup>, Philipp S. Sommer<sup>a,ac</sup>, Michelle Chaput<sup>e</sup>, Dmitry Kupriyanov<sup>ad</sup>

<sup>a</sup> Institute of Earth Surface Dynamics, Geopolis, University of Lausanne, Switzerland

<sup>b</sup> Geocology, Department of Environmental Sciences, University of Basel, Switzerland

<sup>c</sup> Department of Geosciences and Geography, P.O. Box 64, 00014, University of Helsinki, Finland

<sup>d</sup> Institut des Sciences de l'Evolution-Montpellier (ISEM), University of Montpellier, Centre National de la Recherche Scientifique (CNRS), EPHE, IRD, Montpellier, France

<sup>e</sup> Department of Geography, Environment and Geomatics, University of Ottawa, Canada

<sup>f</sup> Department of Biology and Centre for Forest Biology, University of Victoria, Canada

<sup>g</sup> Department of Biological Sciences, University of Bergen and Bjerknes Centre for Climate Research, Postbox 7803, N-5020 Bergen, Norway

<sup>h</sup> ISEM, Univ Montpellier, CNRS, EPHE, IRD, Montpellier, France

<sup>i</sup> Aix Marseille Univ, CNRS, IRD, INRA, Coll France, CEREGE, Aix-en-Provence, France

<sup>j</sup> Institute of Forest Sciences, Chair of Silviculture, Tennenbacherstr. 4, University of Freiburg, 79106 Freiburg, Germany

<sup>k</sup> Department of Ecosystem and Landscape Dynamics, University of Amsterdam, the Netherlands

<sup>l</sup> Department of Archaeology, University of Exeter, United Kingdom

<sup>m</sup> Department of Mathematical Sciences, University of Arkansas, USA

<sup>n</sup> Department of Botany, Charles University, Prague, Czechia

<sup>o</sup> Department of Geography, University of Oregon, Eugene, Oregon 97403, USA

<sup>p</sup> School of Geosciences, University of Edinburgh, UK

<sup>q</sup> Royal Botanic Gardens Edinburgh, UK

<sup>r</sup> Mount Royal University, Calgary, Alberta, Canada

<sup>s</sup> Institute of Pre- and Protohistoric Archaeology, Christian-Albrechts-Universität zu Kiel, Germany

<sup>t</sup> Department of Earth and Environmental Sciences, University of Milano-Bicocca, Milan, Italy

<sup>u</sup> Woods Institute for the Environment, Stanford University, USA

<sup>v</sup> European Commission, Joint Research Centre, Directorate D – Sustainable Resources - Bio-Economy Unit, Italy

<sup>w</sup> Geotop & Department of Earth and atmospheric sciences, Université du Québec à Montréal, Canada

<sup>x</sup> School of Geographical Sciences, University of Bristol, UK

<sup>y</sup> Laboratoire des Sciences du Climat et de l'Environnement, CNRS-CEA-UVSQ, Université Paris-Saclay, France

<sup>z</sup> Research Unit of Mathematical Sciences, University of Oulu, Finland

<sup>aa</sup> Department of Geoscience, University of Wisconsin-, Madison, USA

<sup>ab</sup> Department of Geography, University of Wisconsin-, Madison, USA

<sup>ac</sup> Institute of Coastal Research, Helmholtz-Zentrum Geesthacht, Geesthacht, Germany

<sup>ad</sup> Department of Physical Geography and Landscape Science, Faculty of Geography, Lomonosov Moscow State University, Russia

## ARTICLE INFO

## Keywords:

Pollen  
Climate  
Palaeoclimate  
Quantitative reconstructions  
Probability density functions  
Analogues  
Transfer functions  
Community-based standards

## ABSTRACT

Fossil pollen records are well-established indicators of past vegetation changes. The prevalence of pollen across environmental settings including lakes, wetlands, and marine sediments, has made palynology one of the most ubiquitous and valuable tools for studying past environmental and climatic change globally for decades. A complementary research focus has been the development of statistical techniques to derive quantitative estimates of climatic conditions from pollen assemblages. This paper reviews the most commonly used statistical techniques and their rationale and seeks to provide a resource to facilitate their inclusion in more palaeoclimatic research. To this end, we first address the fundamental aspects of fossil pollen data that should be considered when undertaking pollen-based climate reconstructions. We then introduce the range of techniques currently available, the history of their development, and the situations in which they can be best employed. We review

\* Corresponding author.

E-mail address: [chevalier.manuel@gmail.com](mailto:chevalier.manuel@gmail.com) (M. Chevalier).

<https://doi.org/10.1016/j.earscirev.2020.103384>

Received 8 May 2020; Received in revised form 20 September 2020; Accepted 24 September 2020

Available online 28 September 2020

0012-8252/ © 2020 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license

(<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

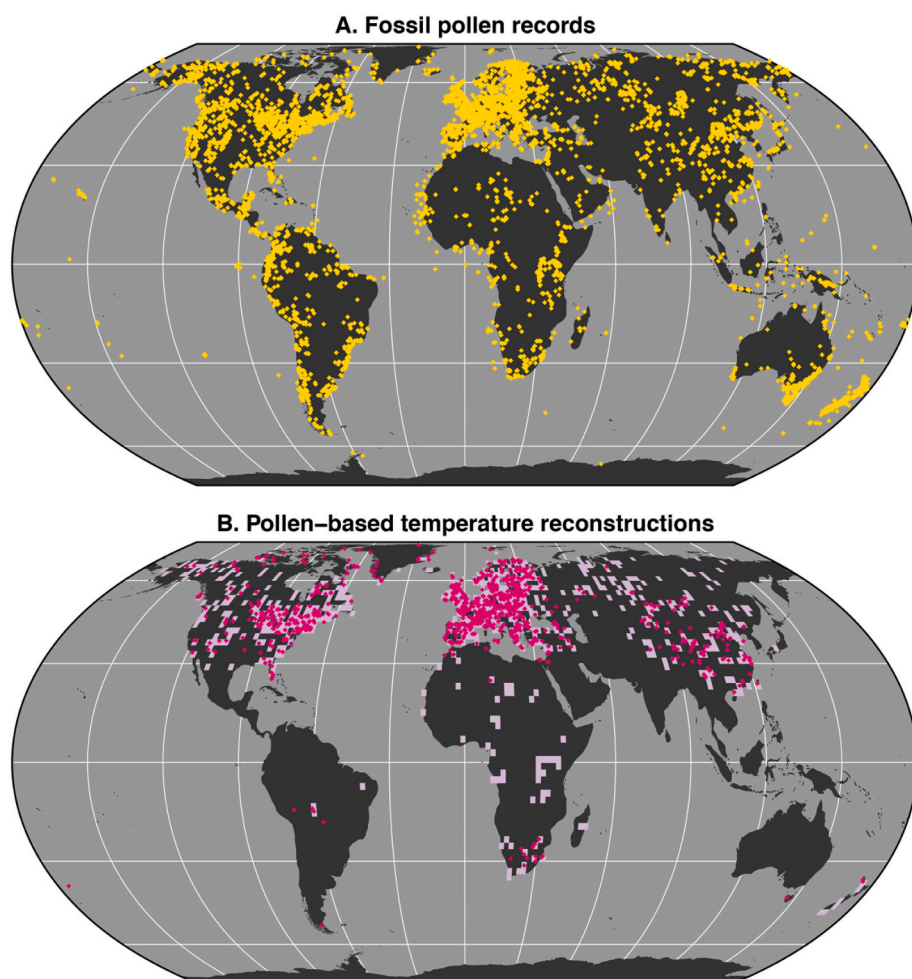
the literature on how to define robust calibration datasets, produce high-quality reconstructions, and evaluate climate reconstructions, and suggest methods and products that could be developed to facilitate accessibility and global usability. To continue to foster the development and inclusion of pollen climate reconstruction methods, we promote the development of reporting standards. When established, such standards should 1) enable broader application of climate reconstruction techniques, especially in regions where such methods are currently underused, and 2) enable the evaluation and reproduction of individual reconstructions, structuring them for the evolving open-science era, and optimising the use of fossil pollen data as a vital means for the study of past environmental and climatic variability. We also strongly encourage developers and users of palaeoclimate reconstruction methodologies to make associated programming code publicly available, which will further help disseminate these techniques to interested communities.

## 1. Introduction

Fossil pollen has been used as a palaeoecological and palaeoclimatological proxy for many decades (Birks, 2019). As a result, it is relatively well understood and intensive work at local and global scales has led to a large body of literature regarding its strengths and weaknesses as a climate indicator (e.g. Birks et al., 2010; Gajewski, 1993; Webb, 1986). Fossil pollen records have been recovered from almost all regions of the world, including ocean sediments and ice cores, spanning the Quaternary period (2.6 million years ago to present) and older (a time interval beyond the focus on this paper; see Traverse (2007) for a summary of deep-time research) and decades of palaeoecological and palaeoclimatic research have produced a wealth of fossil pollen records from thousands of locations, with new records regularly published (Fig. 1A). With its global prevalence and diverse range of climate sensitivities, fossil pollen represents a particularly powerful palaeoclimatic

proxy and a basis for quantitative climate reconstructions and improved understanding of global climate dynamics.

Fossil pollen records can be used to quantitatively reconstruct past climate variables – *i.e.* to estimate numerical climate values expressed in explicit physical units, such as degrees Celsius or millimetres of rainfall – including seasonal and annual temperature and precipitation, as well as related bioclimatic parameters, such as growing degree-days, drought stress and moisture balance (e.g. Bartlein et al., 1984; Birks et al., 2010; Prentice et al., 1991). Pollen-based climate reconstructions performed at the site level provide precise understanding of how local climate changed over time (Bush et al., 2004; Chen et al., 2015; Litt et al., 2012). As reconstructions are expressed in the same units, this advantage can be further used to produce regional to continental syntheses, extract synoptic-scale climate change patterns and, ultimately, inform past atmospheric dynamics (e.g. Davis et al., 2003; Simonis et al., 2012). In particular, numerous broad-scale compilations



**Fig. 1.** (A) Compilation of existing fossil records from the Neotoma Paleocology Database (Williams et al., 2018a), including its regional constituent databases [the European Pollen Database (Fyfe et al., 2009), the North American Pollen Database (Grimm, 2000), and the Latin American Pollen Database (Flantua et al., 2015)], the Indo-Pacific Pollen Database (Hope, 2018), the African Pollen Database (Vincens et al., 2007) and the ACER database (Sánchez Goñi et al., 2017). (B) Locations where temperature reconstructions have been generated from fossil pollen data. The Bartlein et al. (2011) data are represented as  $2^{\circ}\times 2^{\circ}$  light pink grid boxes, while the recent multi-sites studies of Mauri et al. (2015), Gajewski (2015), Marsicek et al. (2018), Shuman et al. (2018) and Herzschuh et al. (2019), as well as several site-based reconstructions (Bush et al., 2004; Chase et al., 2015; Chase et al., 2015; Chevalier and Chase, 2015; Fletcher and Thomas, 2010; Jara et al., 2017; Lim et al., 2016; Newnham et al., 2017; Punyasena et al., 2008; Schäbitz et al., 2013; Whitney et al., 2011) are represented in pink (actual coordinates of sites). This compilation of fossil records and pollen-based climate reconstructions is not exhaustive, but is nevertheless representative of the existing density of records and reconstructions in each region. The two maps are based on the 'Equal Earth' map projection to better represent the relative sizes of the different continents (Šavrič et al., 2019). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

of site-based reconstructions have been built and used in this fashion to evaluate different generations of Earth System Model (ESM) palaeo-simulations to highlight key data-model discrepancies and congruences (e.g. Brewer et al., 2007b; COHMAP Members, 1988; Harrison et al., 2015; Joussaume and Taylor, 1995; Sawada et al., 2004).

A decade ago, Bartlein et al. (2011) produced an exhaustive global synthesis of pollen-based reconstructions covering the mid-Holocene (c. 6000 years ago) and/or Last Glacial Maximum (LGM; 19–26.5 ka as defined by Clark et al. (2009)). Their gridded data highlighted regionally homogeneous climate signals for both periods and have been subsequently used to address past and future climate change issues. For example, Schmittner et al. (2011) used the LGM reconstructions to evaluate different model simulations of the LGM climate and identify the models that were performing best to refine the likely range of future temperature projections. This work also identified important gaps in the global coverage and density of quantitative pollen-based climate reconstructions and emphasised the need to obtain better data coverage in many underrepresented regions (Africa, South and Central America, India, Central and southeast Asia and Oceania, hereafter ‘quantification deserts’; Fig. 1B). However, only northern and eastern Asia (Herzschuh et al., 2019) and the Canadian Arctic (Gajewski, 2015) have been densely populated by new reconstructions since 2011 (Fig. 1B), as the main research focus of the community has continued to be climate variability in the Northern Hemisphere mid-latitudes during the last 2000 years (e.g. Rodysill et al., 2018; Shuman et al., 2018), the Holocene (e.g. Marsicek et al., 2018; Mauri et al., 2015) or the Last Glacial Maximum (e.g. Izumi and Bartlein, 2016).

As quantification deserts are not specific to pollen-based climate reconstructions and many other palaeoclimate proxies share similar spatial bias (e.g. Kaufman et al., 2020), a sharp palaeoclimate knowledge gap exist between well-studied regions and quantification deserts. Recent methodological developments together with increased data availability have started to pave the way towards broader applicability of pollen-based climate reconstruction techniques in some quantification deserts, such as Africa (Truc et al., 2013), South America (Schäbitz et al., 2013) and New Zealand (Newnham et al., 2017). Given these advances in methods and data coverage, existing fossil pollen records in these regions have become suitable candidates for quantitative climate reconstruction. Increasing the coverage and number of reconstructions from such regions will 1) refine our understanding of climate variability across a broader range of timescales, 2) contribute to shaping a more complete description of past global climate dynamics and interactions and 3) enable a more spatially complete evaluation of ESM palaeo-simulations and thus contribute to improving their capacity to predict past and future climate change (Harrison et al., 2016).

This review paper aims to facilitate access, use and distribution of pollen-based climate reconstruction techniques and calibration data to contribute to the further development and understanding of past climate variability during the late Quaternary from the local to the global scale. We seek to 1) provide a foundation for new developments in this field by surveying existing techniques and tools, 2) identify the requirements that should be met to produce high-quality pollen-based climate reconstructions, and the challenges in doing this, and 3) promote strategies and numerical solutions by which quantitative reconstruction techniques can be more broadly and easily accessed and used, especially in quantification deserts. To reach these goals, this review is divided into seven sections. In Section 2, we review the historical milestones that have marked the development of pollen-based climate reconstruction since the 19th century. In Section 3, we describe the fundamentals common among pollen-based climate reconstructions and discuss factors and processes that determine the composition of pollen assemblages, particularly those that affect pollen-climate relationships. Section 4 combines a detailed description of the state-of-the-art quantitative reconstruction techniques with a review of the historical methodological developments that have led to the most recent techniques. This section is divided in five subsections that

represent the major analytical approaches: 1) the species-indicator approach, 2) the regression approach, 3) the classification approach, 4) the process-based approach and 5) the Bayesian approach. Since this review is primarily intended for non-experts, we provide high-level descriptions of each approach, and refer readers to original publications for theoretical details. Section 5 provides general guidelines on how to avoid commonly encountered challenges and perform reliable reconstructions. Section 6 provides a list of currently available software packages, scripts and data used to apply the various techniques described. In Section 7, we discuss potential avenues for future methodological developments and propose a list of possible axes of development for the integration of pollen-based quantitative reconstruction techniques in multi-disciplinary research.

## 2. A brief historical perspective

The pioneering essay of von Humboldt and Bonpland (1807) demonstrated that climate strongly influences vegetation spatial distribution patterns. Vegetation ecologists later established that these distributions were primarily defined by species-specific climate and environmental requirements, a concept ultimately formalised as the environmental niche (Gleason, 1917; Good, 1931; Hutchinson, 1957; Pulliam, 2000; Whittaker et al., 1973). Niche theory implies that plant species are only found across specific portions of the available environmental space (i.e. within the tolerance of the species) and exhibit strong preferences along that portion (i.e. towards the environmental optimum of the species). Most pollen-based climate reconstruction techniques aim at modelling/extracting the relationships between vegetation and climate from modern observations to reconstruct past climates, under the assumption that the relationship between vegetation and climate has remained constant through time (i.e. the uniformitarian principle).

The history of the use of fossil pollen data as the basis for quantitative palaeoclimatology is characterised by four distinct phases. The first phase covers the late 19th and first half of the 20th centuries, during which numerous remains of plants (e.g. leaves, needles, seeds or nuts) preserved in lake sediments or peat were analysed and interpreted as indicators of past climate (e.g. Andersson, 1909; Grichuk, 1969; Iversen, 1944; Nathorst, 1871; von Post, 1903). It was not until the work of von Post (1918) that pollen began to be analysed as an indicator of past vegetation and climate (Birks and Seppä, 2010; Edwards et al., 2017). After this discovery, palynologists focused on developing methods to collect pollen records in the field, process and extract pollen grains, and to identify them. This work eventually led to the first quantitative reconstruction of Scandinavian January and July temperatures by Iversen (1944), who combined the presence/absence of mistletoe (*Viscum album*), ivy (*Hedera helix*) and holly (*Ilex aquifolium*) pollen from the fossil record to identify the ‘Atlantic’ (7–8 ka BP), ‘Sub-Boreal’ (2.4–5 ka BP) and modern phases of the Holocene (Fig. 2).

The work of Iversen (1944) marks the transition to the second phase (1944 to ~1990) during which quantitative palaeoclimatology became a progressively more active field of research and transitioned from qualitative interpretations to computing-intensive quantitative reconstructions. The simultaneous publication of a methodology to produce quantitative climate reconstructions from terrestrial pollen (Webb and Bryson, 1972) and ocean foraminifera (Imbrie et al., 1973; Imbrie and Kipp, 1971) led to the development of the first generation of modern and fossil pollen databases, and the development of spatially-comprehensive reconstructions of palaeoclimates during the 1970s and 1980s (e.g. Bartlein and Webb, 1985; Huntley and Prentice, 1988; Webb and Clark, 1977). Shortly thereafter, global temperature reconstructions of Holocene terrestrial (pollen) and oceanic (foraminifera) surfaces were used to evaluate the first generation of climate models (COHMAP Members, 1988; Wright, 1993).

During the third phase (~1990 – ~2010) of quantitative palaeoclimatology, site-specific reconstructions based on fossil pollen became

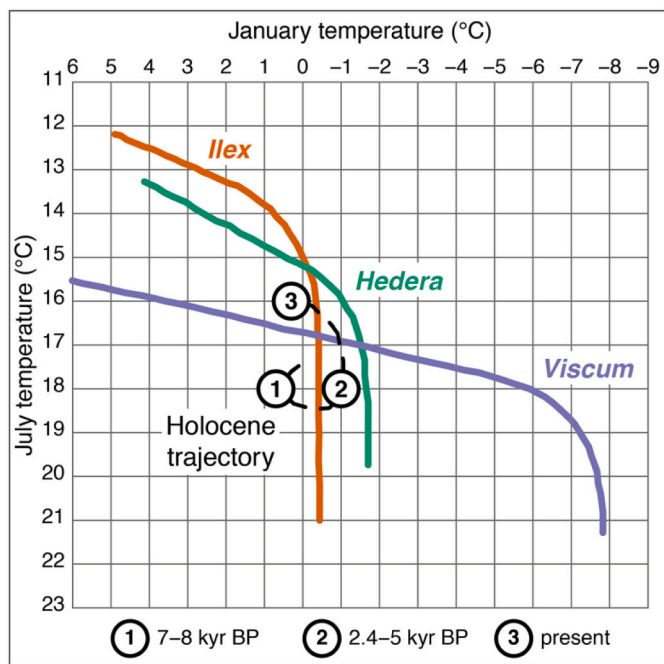


Fig. 2. Classic figure by Iversen (1944) representing the thermal niches of *Viscum*, *Hedera* and *Ilex* and how the combination of their presence (below their respective curve) or absence (above and to the right of their respective curve) in modern environments could be used to estimate past temperatures for the ‘Atlantic’ (1: 7–8 ka BP), ‘Sub-Boreal’ (2: 2.4–5 ka BP) and modern (3) phases of the Holocene.

increasingly prevalent, providing a wealth of new information regarding the evolution of past climates. This extensive use of fossil pollen has fostered 1) the development and refinement of many of the reconstruction methods in use today (see Section 4) and 2) the continuous expansion of their applicability, which have eventually led to the global synthesis of Bartlein et al. (2011) (see Fig. 1B). In parallel, the proliferation of climate simulations (e.g. the Paleoclimate Modelling Intercomparison Project (PMIP); Joussaume and Taylor, 1995) has provided the scientific community with a richer spectrum of tools to be used in conjunction with pollen-based reconstructions to investigate past climate change during different key periods of the past and decipher the main drivers of past climate change (e.g. Cheddadi et al., 1996; Kageyama et al., 2006; Prentice et al., 2001; Williams et al., 1998).

In the last decade (phase 4, ~2010 to present), new climate reconstruction techniques, based on Bayesian statistics, machine learning or process-based approaches, have been proposed to improve inferences of past climate from pollen data (see Section 4 for details; Chevalier

et al., 2014; Izumi and Bartlein, 2016; Parnell et al., 2008; Salonen et al., 2019). The emerging open-data era, illustrated by the growth of online databases and data repositories (e.g. Williams et al., 2018b) and the advance of informatics and data standards (e.g. Khider et al., 2019; McKay and Emile-geay, 2016), has also influenced how pollen data are used to reconstruct past climates. As such, many studies have recently addressed fundamental palaeoclimatic questions by making use of extensive networks of fossil pollen data, including Marsicek et al. (2018) addressing the Holocene temperature conundrum (Liu et al., 2014), Routson et al. (2019) analysing the effects of latitudinal temperature gradients on moisture transport in the Northern Hemisphere, and Kaufman et al. (2020) reconstructing mean global temperature trends during the Holocene (pollen-derived records contributing to almost 40% of records in the analysed data collection). The extension of the classic site-based analytical frameworks to networks of pollen records accessible from public resources represents an unprecedented opportunity to re-evaluate how fossil pollen data can be used to infer past climate, both in data-dense regions and in quantification deserts.

### 3. General principles to pollen-based climate quantifications

#### 3.1. Fundamental modelling assumptions

Modern pollen-based quantitative palaeoclimatology has undergone the development of a diverse array of statistical or modelling techniques to transform fossil pollen assemblage data into past climate estimates. These techniques draw information from spatial networks of modern data to represent the empirical relationship between climate and either plants, plant distributions, or pollen assemblages (‘modern calibration’ stage; Fig. 3). These relationships are then used to make quantitative inferences about past climates from fossil pollen assemblages (‘climate reconstruction’ stage; Fig. 3). Combined, these two stages are called a ‘space-for-time’ substitution approach (Blois et al., 2013; Jackson and Williams, 2004) and are based on the following five assumptions (e.g. Birks et al., 2010; Imbrie and Kipp, 1971):

1. The taxa observed in both modern and past samples are systematically related to the climate in which they live. This assumption assumes approximate dynamic equilibrium between the observed pollen assemblage and climate.
2. The climate variables to be reconstructed are important determinants, or systematically related to important determinants of the vegetation and the associated pollen. This assumption assumes that the relationship has remained stable throughout the record (i.e. the uniformitarian principle).
3. Unstudied environmental variables have had a minor influence on the fossil data over the studied period of time and at the spatial scale of the pollen record, or their relationship with the variable of

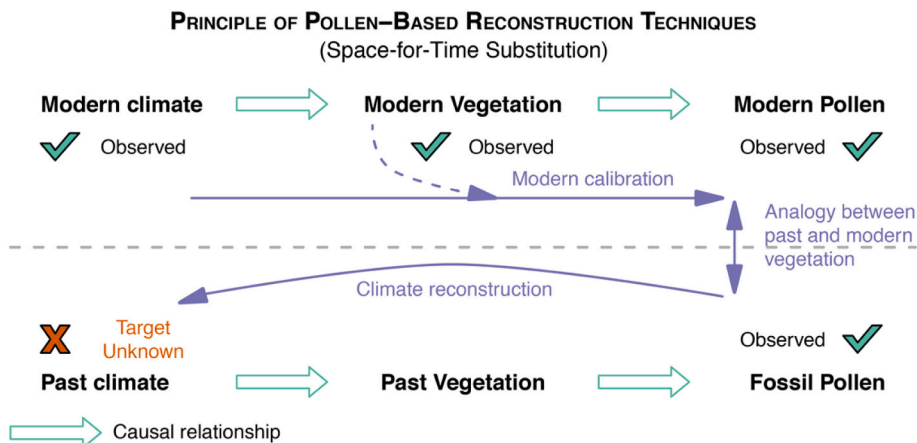


Fig. 3. Schematic representation of the space-for-time principle underlying pollen-based climate reconstruction approaches. For most models, the modern relationships between climate and pollen are used to infer past climates under the assumption that similar pollen compositions are the result of the same climatic conditions (Sections 4.2 to 4.4). Other models extract information from modern vegetation to define pollen-climate relationships (Section 4.1).

interest remained the same as in the calibration dataset.

4. The taxa in the modern calibration dataset are the same biological entities as in the fossil data, and their ecological responses have not changed over the studied period of time (uniformitarian principle).
5. The specified mathematical model has sufficient flexibility to represent the complexity of the ecological responses to the climate variables of interest.

These assumptions have long been recognised, but the difficulty of testing them has sometimes turned them into what Jackson (2012) referred to as *unknown knowns*, namely assumptions that are often taken for granted without even knowing the nature of how these factors interact. Juggins (2013) discussed how the violation of assumptions 2 and 3 could lead to spurious reconstructions. We thus re-emphasise the need to proceed with suitable care when undertaking pollen-based climate reconstructions.

### 3.2. The nature of fossil pollen data

A good comprehension of the processes that govern the composition of pollen assemblages and their relationship with climate and the environment is necessary to understand the different assumptions underlying the reconstruction techniques. In this section, we present the most defining characteristics of fossil pollen assemblages, how these characteristics relate to the five fundamental modelling assumptions and how they can influence a reconstruction.

#### 3.2.1. Pollen production, dispersal strategies and preservation

Pollen grains are produced in all environments where spermatophytes (i.e. seed-bearing plants) grow (Wodehouse, 1935). As the walls of pollen grains contain the chemically-stable polymer sporopollenin, they can also be preserved over long periods of time in anoxic environments (Faegri et al., 1989; Havinga, 1967). Each plant species adopts specific production and dispersal strategies, which can influence the representation and abundance of taxa in fossil pollen records. Anemophilous taxa depend on wind to disperse their pollen and

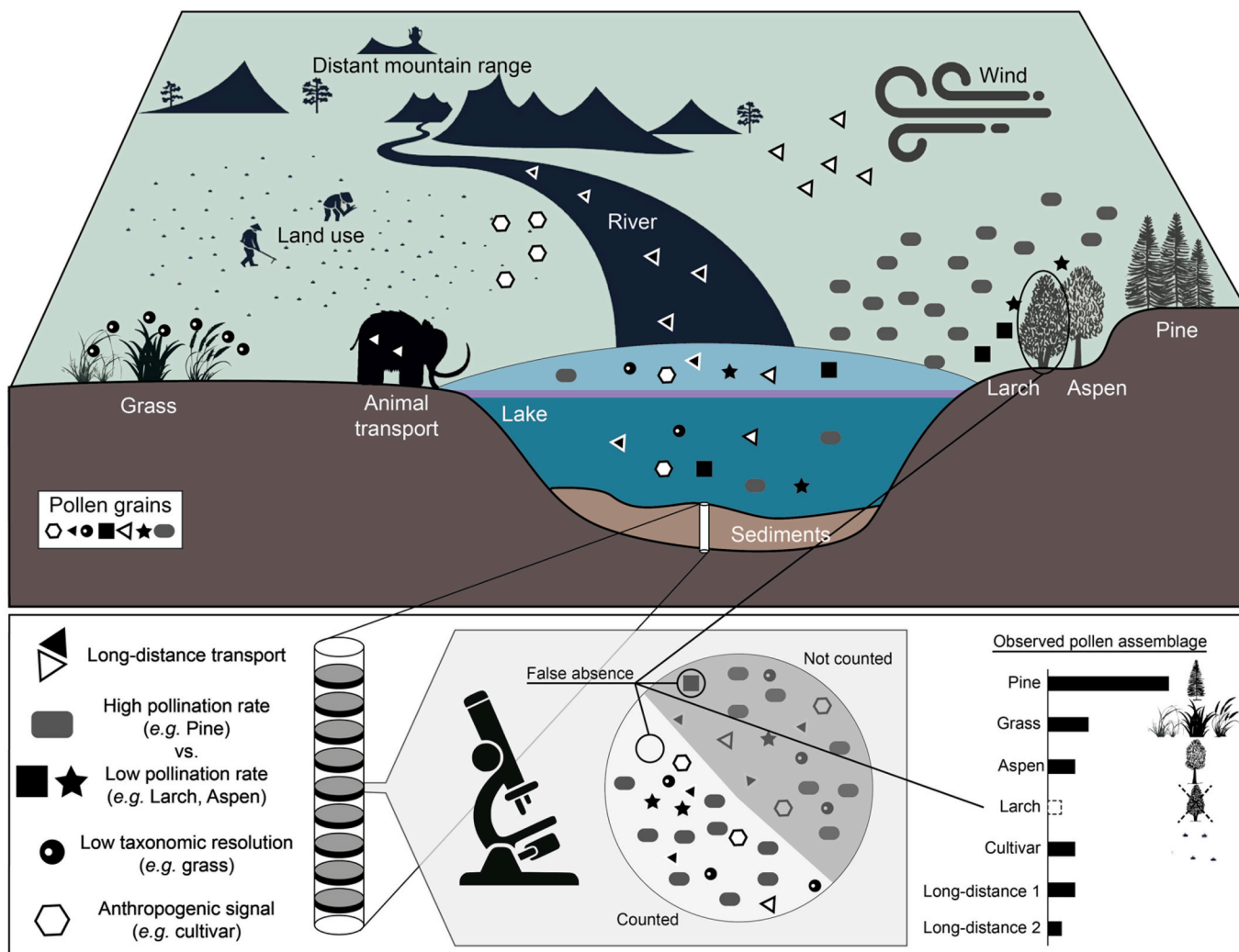


Fig. 4. Conceptual model of different processes that contribute to pollen being preserved in lake sediments. The squares, stars, rounded rectangles and black circles represent pollen grains derived from vegetation around the lake. Black and white triangles represent pollen transported over long distances by the river and wind, respectively. Each type of plant produces different quantities of pollen (cf. the trees on the right shore of the lake). The different grass species produce indistinguishable pollen grains (cf. the grasses on the left shore of the lake). The white hexagons are produced by cultivated plants (e.g. rice or maize fields) and represent anthropogenic activity. Animals on the landscape (here represented by a mammoth) can also transport pollen that deposits in the sediments. All these pollen grains accumulate in the sediment at the bottom of the lake where they are collected using coring equipment (white cylinder). When sampled and counted under a light microscope, the observed pollen assemblage represents a combination of all the processes in action in this landscape and in the lab (e.g. the number of pollen grains counted that can lead to false absences).

commonly have high pollen production rates, while entomophilous taxa that rely on animals (e.g. insects) to disperse their pollen generally produce fewer grains. For example, wind-pollinated pine (*Pinus*) species have one of the highest pollen production rates and often dominates pollen assemblages, even when they are not prominent in the local environment (e.g. Campbell et al., 1999; Tyldesley, 1973) (Fig. 4). In contrast, larch (*Larix*) and aspen (*Populus*) are typically under-represented because of lower production, dispersal and preservation potential, even when they dominate the local landscape (Leitner and Gajewski, 2004; Peros et al., 2017; however, see Carter et al. (2017) and Morris et al. (2019)).

Contrary to plant macroremains that can generally only be used as presence/absence data, pollen grains are much more abundant in sediments and can be used to derive relative proportions of pollen taxa, which enables more refined statistical inferences about past climates. The reliability of these percentages directly depends on the number of counted pollen grains being counted (Maher, 1981; Maher, 1972). The optimal number of grains depends on the research aim, but counting sufficient numbers of pollen grains is critical to account for the different production and dispersal strategies as the over-representation of some taxa can mask the presence of other climatically informative taxa with lower pollen production rates ('false absence', see larch on Fig. 4) and negatively impact the accuracy of the climate reconstruction. For climate reconstructions, it is usually recommended to identify a minimum of 300 pollen grains in individual fossil samples to estimate reliable percentages of the dominant taxa but counting 500 grains or more is usually necessary to estimate the proportions of rare taxa. However, counting high number of pollen grains may not be possible in certain environments or sediment types due to preservation or other taphonomic issues (e.g. Djamali and Cilleros, 2020; Keen et al., 2014; Lytle and Wahl, 2005).

### 3.2.2. Representativeness of a pollen sample

A pollen assemblage is not a direct measure of local vegetation composition, as many biological, climatic and environmental processes influence the production, dispersal and preservation of pollen grains (Fig. 4; Gajewski, 2006; Goring et al., 2013; Huntley, 2012; Prentice, 1986). Most pollen grains are mainly transported by wind (the dominant vector) that mixes grains in the air from various locations, and by streams and rivers that can carry grains over land depending on the local hydrology (Fig. 4). In some rare cases, pollen grains can also be transported to the depositional setting by animals, possibly introducing a signal that reflects foraging and mobility preferences (e.g. Chase et al., 2012; McAndrews and Turton, 2007). Pollen assemblages thus represent a mix of locally and remotely produced pollen grains whose chances of being observed decrease with increasing distance of the plant source to the sampling location (Jackson and Lyford, 1999; Prentice, 1985; Sugita, 1994). The relative proportion of local and distant taxa is related to the catchment characteristics of the studied archive.

The environmental setting of a pollen archive can strongly influence the scale of climatic inferences that can be derived from a pollen record (Jackson, 1994; Jacobson and Bradshaw, 1981). Records extracted from small lakes, mires, bogs or forest-hollows usually contain a stronger local signal (small pollen source area) compared to records extracted from large lakes (larger pollen source area) that act as regional integrators (Prentice, 1985). Ice cores (e.g. Bourgeois et al., 2000; Brugger et al., 2018) and marine cores (e.g. Dupont et al., 2019; Klotz et al., 2006) represent extreme endmembers in terms of local versus regional pollen ratios, as none of the collected pollen grains are locally produced. While fossil pollen are generally recovered from lakes, wetlands or marine sediments, they have also been recovered from a diverse range of more unusual archives, such as the middens of packrats (Davis and Anderson, 1987) and rock hyraxes (Chase et al., 2012; Scott, 1990), animal faeces and remains (e.g. Akeret et al., 1999; Gil-Romera et al., 2014; Kosintsev et al., 2010), ice cores (e.g. Bourgeois et al., 2000;

Koerner et al., 1988), and speleothems (e.g. Brook et al., 2010; Sniderman et al., 2016). When possible, characterising the source archive and the likely area represented by pollen assemblages is important to defining the type and scale of climate inferences that can be drawn from a fossil pollen record (e.g. local change vs. regional- or continental-scale variability). Irrespective of the type of archive, environments with low pollen productivity can also be highly influenced by remote pollen producers (potentially conflicting with assumption 1). For instance, observing wind-blown pine pollen grains is quite common in tundra and high-Arctic ecosystems because the large quantities of pollen grains produced by distant, wind-pollinated pines contributes to the local pollen signal (e.g. Gajewski, 2006; Wang et al., 2017). In such cases, separating local from extra-local taxa can reduce the noise of the pollen data and improve the local climate signal.

### 3.2.3. Variable taxonomic resolution

Pollen grains are rarely identifiable to the species level using light microscopy (e.g. Halbritter et al., 2018). Pollen grains from temperate and boreal environments are generally identified at the sub-genus – i.e. a group of species but not the entire genus – to sub-family level, while resolution is generally lower in tropical regions (genus to family level). The diversity of species comprising a given pollen type can complicate the estimation of reliable pollen-climate responses, even if its constituent species have a clear response to climate. For instance, Finsinger et al. (2007) have shown that *Ostrya*-type pollen has a bimodal distribution across Italy, as it is produced by both the European hornbeam (*Ostrya carpinifolia*), which is more abundant to the colder and wetter north, and the oriental hornbeam (*Carpinus orientalis*), which is more abundant to the warmer and seasonally drier south. Only the techniques that can account for these distinct climate preferences and estimate multimodal climate responses or the techniques that do not model the responses can deal well with such issues without breaking Assumption 5. The most diverse pollen types (e.g. the grass family (Poaceae) which comprises hundreds of species that are prevalent from the Arctic tundra to tropical savannas; Fig. 4) can be problematic for some reconstruction methods (e.g. the indicator species, see Section 4.1), especially in environments where 1) the relationship between the pollen and climate can be a mix of various individual responses (thus challenging Assumption 2), and where 2) the potential species turnover within the pollen type can be masked, thus possibly breaking Assumption 4 over time.

Pollen taxonomy is continuously being improved to differentiate these diverse pollen types into subgroups (e.g. Bush and Weng, 2007; Julier et al., 2018; May and Lacourse, 2012; Schüler and Hemp, 2016). Increasing the taxonomic resolution of pollen analyses is useful for palaeoecological studies, especially in hotspots of botanical diversity where tens to hundreds of plant species with different climatic niches produce indistinguishable pollen grains (e.g. in Amazonia (Burn et al., 2010) and South Africa (Valsecchi et al., 2013)), however it may not always be beneficial for climate reconstructions. While indicator species approaches benefit positively from a higher taxonomic resolution (cf. Section 4.1), most reconstruction techniques use large compilations of modern pollen samples to estimate the relationship between pollen and climate (Sections 4.2 to 4.4). Taxonomic improvements cannot be back-propagated to those existing compilations of modern samples, so that records with higher and/or novel taxonomic resolution have to be taxonomically simplified to match the resolution of the modern samples (taxonomical smoothing; Goring et al., 2013; Williams and Shuman, 2008).

### 3.2.4. Human impact

Most terrestrial environments are or have been influenced to some extent by humans. Recent evidence suggests that even the most remote locations of the tropical rainforests of Africa and South America have been modified by human activities during the Holocene with long-term effects still visible today (e.g. Marchant et al., 2018; McMichael et al.,

2017). Human activities (e.g. deforestation or agriculture) modify vegetation structure, composition and diversity (e.g. replacement of a forested area by a field on Fig. 4), which can obscure the observed relationships between pollen and climate (St Jacques et al., 2015). The issue has been recognised very early on in the development of the reconstruction techniques (e.g. Guiot, 1987; Howe and Webb, 1983; Pons, 1984), as the well-studied landscapes of Europe and the Mediterranean Basin are historically among the most human-impacted (e.g. Conedera et al., 2017; Gaillard et al., 2010; Giesecke et al., 2019; Kaplan et al., 2009). The arrival of Europeans to eastern North America, which is captured by the introduction and rise of ragweed (*Ambrosia*) pollen, also represents another striking example of human-driven landscape disturbance (e.g. McAndrews, 1988), with this distinctive environmental change being used as a precise dating point in some studies (e.g. Blois et al., 2011; Bunting et al., 1998). Anthropogenic land-uses (modification of vegetation compositions, loss or gain of species) can substantially influence pollen-climate relationships even if the actual impact of this influence on climate reconstructions is often difficult to quantify. Pollen-based climate reconstructions from environments characterised by strong human influence need to be evaluated carefully, as these could be biased. However, many reconstructions from such regions have been supported by non-botanical palaeoclimate proxies, indicating that modern climate reconstruction methods can – to some extent – cope with human impact (e.g. Cheddadi and Khater, 2016; Lotter et al., 2000; Mauri et al., 2015; Seppä et al., 2009).

### 3.3. Selection of climate indicators

There is no universally applicable rule that can define which pollen taxa should or should not be used for climate reconstructions and some expert judgment is often required. Some reconstruction techniques work well with the full suite of identified pollen taxa, while others perform better with a restricted set of taxa with strong sensitivities to the climate parameter being reconstructed. The optimal number of selected taxa will also vary between regions and taxonomic resolution of pollen records. Using too few taxa excludes part of the climatic signal recorded by the pollen, while long lists of taxa increase chances to include 1) taxa that are not sensitive to the studied climate variable across the study area or 2) rare taxa, for which the estimations of pollen percentages or pollen-climate relationships are insufficiently constrained. The total number of pollen taxa used in a reconstruction is most often a trade-off between using too few and too many taxa. Some simple rules can guide the selection of taxa, including:

- As terrestrial and aquatic taxa can be influenced by distinctly different sets of environmental variables, it is recommended to exclude all non-terrestrial pollen taxa and to rescale the sum of terrestrial plant taxa to 100% before reconstructing climate.
- All the taxa observed in the fossil record should be present in the modern dataset and in significant proportions (e.g. presence in at least 20–25 samples), as a large number of modern calibration samples of pollen or plant occurrences are required to establish a taxon's response to climate.
- Cosmopolitan taxa or pollen taxa identified at a low taxonomic resolution may be problematic because their signal-to-noise ratio is often low, but also because estimating a reliable pollen-climate relationship may be complex (e.g. grass species). With such taxa, quantifying reliable climate response curves can be difficult and, depending on the reconstruction method used, they may provide limited information or even introduce noise to the reconstruction.
- It may be useful to exclude observations with low percentages (e.g. < 0.5 or 1%) from the reconstruction. While low percentages can indicate presence in the environment, they can also result from lab or sampling contamination, or long-distance transport. Excluding taxa based on low percentages should always be associated with additional considerations of their ecology and

taphonomy. Different exclusion thresholds can also be defined for taxa with different pollen production rates.

### 3.4. Climate variable(s) that can be reconstructed from a fossil record

Vegetation compositions from different environmental settings are sensitive to different environmental variables. As such, the identification of which climate variable(s) can be realistically reconstructed from different vegetation compositions is critical. Plants are sensitive to a broad range of eco-physiological variables, such as maximum and minimum temperatures, the number of growing degree and frost days, the length of the dry period or the seasonality of rainfall and/or temperature (e.g. Harrison et al., 2010; Prentice et al., 1992; Woodward and Williams, 1987). All of these variables directly influence plant growth, survival and reproduction and ultimately define the presence/abundance or absence of plant species at a site (e.g. Guisan and Zimmermann, 2000). These bioclimatic variables are strongly correlated with monthly or seasonal mean values commonly used in climatology (e.g. winter temperature with number of frost days). Climate variables are thus commonly used as proxies for these more specific variables because they have the advantage of providing a common scale with other biotic and abiotic palaeoclimatic sensors (e.g. isotopes, biomarkers), hence improving comparability among palaeoclimate datasets, instrumental data, and climate model simulations.

The variable selection should be as independent as possible from the researcher's expectations. In other words, it is not because a researcher is primarily interested in studying temperature change that temperature can or should be reconstructed from all pollen records or for all time periods. One consequence of Assumptions 2 and 3 is also that reconstructing climate is not always possible when non-climatic factors leave a stronger impact on the pollen assemblages than climate itself (Juggins, 2013). This can include sea-level changes at coastal sites that modify the vegetation at the site (e.g. Quick et al., 2018), fire that gives different trajectories to climate-driven ecological changes (e.g. Cadd et al., 2019), or land use (e.g. Birks and Seppä, 2003; Finsinger et al., 2010; Howe and Webb, 1983). All of these can affect both modern and fossil records. Identifying where and when this influence is strong enough to supersede climate, for instance by adapting sample processing in the lab to specifically identify pollen of cultivated plants (e.g. Maezumi et al., 2018; Whitney et al., 2012), or by comparing the reconstruction with a vegetation-independent climate reconstruction (e.g. Finsinger et al., 2010; Peyron et al., 2005) can highlight phases if/when the reconstruction is likely to be most biased.

Unfortunately, no technique is equipped to automatically identify the variables that can be reconstructed (i.e. the variables that have driven the observed ecological changes), and they will always provide climate estimates of the prescribed variable irrespective of the exact drivers(s) of past vegetation change. Considering the ecology (and potential palaeoecology) and local environmental drivers at a site is critical to choose the variable(s) that can be reliably reconstructed based on pollen records. Examining performance indices and spread of uncertainties associated with the fitted reconstruction model (see Section 5.2 for details on how to calculate these) can identify whether a variable is a poor predictor of the pollen data. However, such analyses, alone are insufficient to affirm that a certain variable can be reconstructed, since good performance indices can also be obtained with potentially confounding, covarying variables (i.e. variables that correlate with the studied taxa in the modern settings but have no causal relationship with them). In such a situation, a confounding variable correlates with the studied taxa only because it is correlated with another variable, which ultimately drives the observed distributions of taxa. If the relationship between the driving and confounding variables changes over time (breaking Assumptions 1 and 3), reconstructions may become biased. For example, January and July temperature are usually positively correlated in spatial datasets today and, as such, both can appear as significant drivers of vegetation. However, in many regions

they are likely to have been negatively correlated during the Holocene due to orbital forcing. In this trivial example, the positive modern correlation between the two temperature variables in calibration datasets will potentially induce a positive correlation between reconstructed temperatures based on a pollen record, whereas in reality these may have been negatively correlated.

Atmospheric CO<sub>2</sub> imposes an additional, non-climatic control on vegetation dynamics and structure that must be accounted for when interpreting reconstructions (e.g. Cowling and Sykes, 1999). Atmospheric CO<sub>2</sub> concentration influences plant water-use efficiency, so that lower CO<sub>2</sub> concentrations, such as during the LGM (e.g. Monnin et al., 2001), induce plants to consume more water to maintain the same rate of photosynthesis (e.g. Harrison and Prentice, 2003; Jolly and Haxeltine, 1997; Ward et al., 2005). Low CO<sub>2</sub> concentrations are likely to create non-analogue conditions, i.e. conditions that have no modern equivalent, thus potentially breaking Assumption 5. The approaches that do not account for the CO<sub>2</sub> effect may be biased when reconstructing climate of periods with lower CO<sub>2</sub> concentrations and, for instance, ‘misread’ LGM vegetation as indicating conditions drier than they actually were. Low CO<sub>2</sub> also decreases available carbon for photosynthesis and may bias the temperature reconstruction towards lower values by attributing the effect of CO<sub>2</sub> limitation to temperature (Wu et al., 2007b; Wu et al., 2007a), although others have argued any temperature effect is limited (Williams et al., 2000).

#### 4. Approaches to quantification

Pollen-based climate reconstructions can be performed based on different approaches, each with unique advantages and weaknesses. Here, we present the most relevant that contributed to the development of the techniques in use today. Their relative strengths and weaknesses are discussed in Section 5.

### 5. Summary

#### 5.1. Indicator-species approach

Indicator-species approaches estimate independent climatic responses for each pollen taxon. The responses of the climate-sensitive taxa are then combined to estimate the climate parameter and the distribution of its uncertainties. These techniques use modern plant presence-only data (i.e. geolocalised occurrences) to infer the plant-pollen-climate relationships.

Indicator-species techniques are conceptually intuitive and were the first reconstruction techniques to be developed. Their first applications were Andersson (1903, 1909, 1910), who used the presence of macrofossils of hazel (*Corylus avellana*) as an indicator of warm season temperature in Sweden during the Holocene. This work was later extended by Iversen (1944), who combined independent bivariate estimations of the niche edges of three indicator species (*Ilex*, *Viscum* and *Hedera*) to produce the first multi-taxa reconstructions of Holocene winter and summer temperature minima in Scandinavia (Fig. 2). Despite the limited predictive power of these original techniques beyond their calibration area, they laid the foundations and the practical ecological background for the more sophisticated and computationally-intensive techniques used today.

##### 5.1.1. Mutual Climatic Range (MCR)

The MCR technique was developed by Grichuk (1969) using botanical remains, but the name ‘Mutual Climatic Range’ was proposed by Atkinson (1986) in a study based on fossil beetles. MCR pollen-climate relationships are defined from modern plant distribution data as the minimum and maximum climatic values where the taxon is observed (i.e. where it lives; Atkinson et al., 1987). Past climates are reconstructed from the intersection of these climate ranges to identify the climate interval where the climate ranges of all taxa in a sample

overlap, which results in a climate range and not a single ‘best value’. The precision of the reconstruction (i.e. the width of the reconstructed climate range) is defined by the affinities of all the observed taxa in a sample with the climate variable being reconstructed, even if, in practice, it is most often constrained by the presence of one or two stenotopic species (i.e. taxa with strict climate requirements). The absence of a central tendency measure in the MCR approach can make the reconstructions difficult to reconcile with independent records, because the range of values produced by MCR is, by construction, unweighted and no climate values are more likely than others within that range, even if the contrary has been claimed (Coope et al., 1998; Schenk et al., 2018).

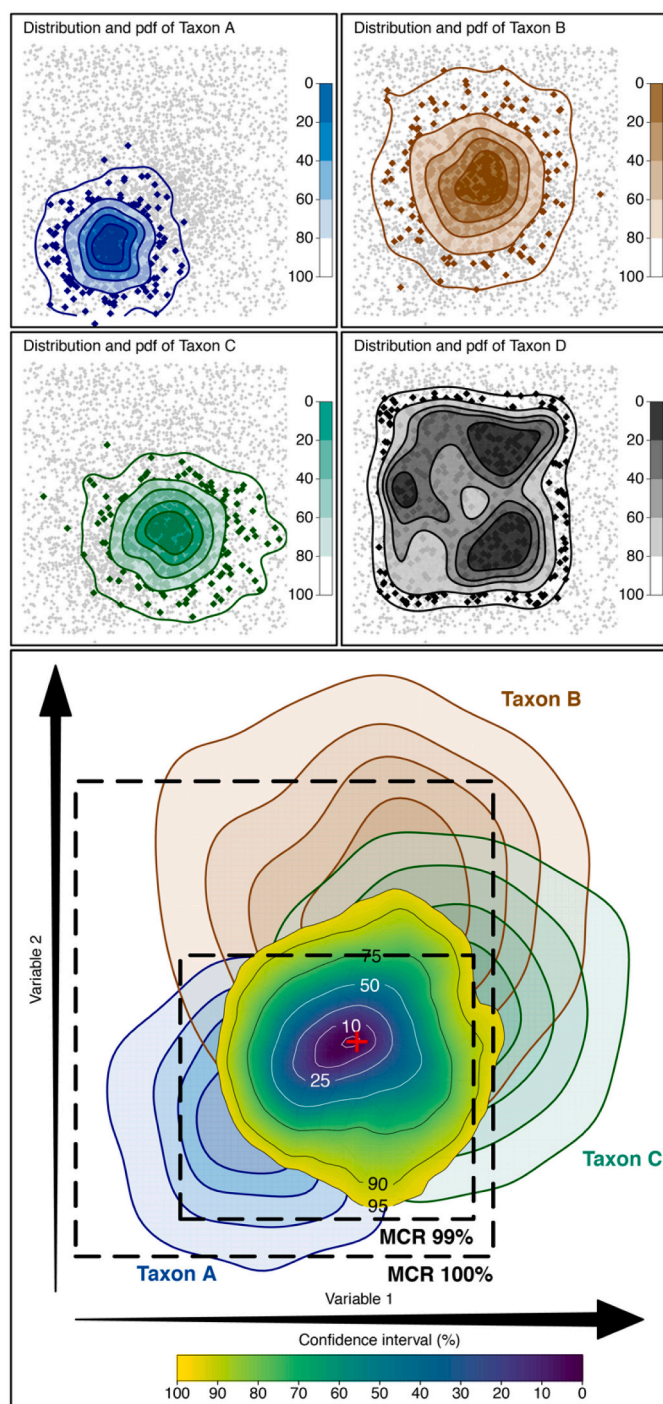
One fundamental limitation with MCR reconstructions concerns the sensitivity of the estimated climatic ranges, particularly in the presence of stenotopic species. The absolute minimum and maximum climate values where a taxon can live are difficult to estimate empirically and any error in the estimation of these ranges is directly transferred to the reconstructions. Different strategies have been proposed to reduce this sensitivity but none of these options has fully resolved the issue. For example, the total distribution range was truncated to the 10–90% interval (Sharpe, 2002), although this leads to biased inferences as true presences are truncated. Fauquette et al. (1998) proposed using varying tolerance ranges for each taxon in an assemblage and to associate high pollen percentages with narrower climatic ranges. More probabilistic approaches, such as fitting sinewaves to define the boundaries of the climate envelopes (Marra et al., 2004), have also been applied to account for all the occurrence data and diminish the importance of extreme observations. Similarly, Pross et al. (2000) narrowed the estimation of the ranges by modelling a most likely range within the observed full range. Ultimately, limitations with approaches like MCR are related to the statistical framework upon which they are based (and the lack of error estimates that guide and constrain interpretations), and this limits the measures that can be taken to resolve these issues. As a result, alternative techniques are now often preferred for late Quaternary pollen-based climate reconstructions.

##### 5.1.2. Probability Density Functions (pdfs)

The philosophy of the MCR approach was integrated into a probabilistic framework by Kühl et al. (2002) and Schölzel, 2005), who developed the first climate reconstruction statistical technique employing probability density functions (pdfs). In the pdf techniques, the pollen-climate relationships are described by conditional probability density functions that represent the dependence of the different taxa on climate (Kühl et al., 2002). A pdf describes the probability of observing any climate value considering the presence of one taxon, with the constraint that all probabilities (i.e. the area under the curve) sum to one. In contrast to MCR that focuses on the estimation of the minimum and maximum climate values only, the pdf approach uses all the information contained in the species' distributions to estimate entire climate responses and is, therefore, significantly less sensitive to the structure and range limit of the modern calibration dataset than MCR.

The pdf techniques combine modern climate and plant occurrence data (presence-only from their spatial distributions) to determine the specific climate envelope(s) of pollen taxa (Fig. 5). To account for the differential representation of the climate values across the entire study area (some climate values are more abundant than others), climate values are weighted by the inverse of their abundance in the modern climate space. Abundant climate values are down-weighted, while the rarest ones are relatively up-weighted (Bray et al., 2006; Kühl et al., 2002; Schölzel, 2005). While weighting climate values ensures that the pdfs do not reflect the modern distribution of climate, estimating ecologically reliable climate responses for all observed taxa is not always possible due to the taxonomic resolution of pollen records (e.g. grass pollen, see Fig. 4). When pollen taxa are composed of dozens or more species, their pdfs can be relatively flat and/or broad and such taxa are commonly excluded from the reconstruction. Similarly, taxa





**Fig. 5.** Schematic representation of the *pdf* approach. (top) Distribution of four taxa (coloured points) in a bivariate climate space (grey points). The blue, red and green taxa have a well-defined *pdf* in that space, while the black taxon is not directly sensitive to either of the two variables, as expressed by its width and multimodality. Colour scales represent the cumulative probability distributions. (bottom) Combination of the three sensitive taxa A, B and C to estimate the joint probabilistic distribution of the two climate variables. The best climate estimate is represented with a red cross. MCR reconstructions using the same data are represented with the dashed rectangles. The larger rectangle is based on all the input data (MCR 100%) whereas the smaller one only uses 99% of the values (exclusion of highest and lowest values, MCR 99%), highlighting the sensitivity of MCR to the quality of input data. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

expected to be primarily driven by unstudied climate variables or non-climate variables (e.g. Taxon D on Fig. 5) should also be excluded.

In the *pdf* methods, the *pdfs* of individual taxa are assumed to be independent so that climate can be reconstructed by multiplying the *pdfs* of the selected pollen taxa observed in a fossil sample. This *pdf* represents the probabilistic adaptation of the intersection used in MCR and describes the probabilities of the reconstructed climate variable conditioned by the co-existence of the taxa in the pollen sample (i.e. a weighted interval, Fig. 5). Because the height of each *pdf* is constrained by the range of climate values experienced by the taxon (i.e. all the probabilities must sum to one), the degree of influence of each taxon is inversely proportional to the width of its *pdf* (not represented on Fig. 5 where the *pdfs* are coloured by cumulated probabilities for simplification purposes) and generalist species have less influence than stenotopic taxa. For example, the spread of the distribution of Taxon A on Fig. 5 is more restricted than the spread of the distributions of Taxon B and Taxon C so that the co-existence interval resulting from the multiplication of their *pdfs* is shifted towards the optimum of Taxon A. The results are also substantially more precise than what would be obtained with MCR (dashed unweighted rectangles on Fig. 5).

Two *pdf* techniques have been developed for pollen data: the approach by Kühn et al. (2002) and the CREST (Climate REconstruction Software) technique (Chevalier et al., 2014). Based on similar fundamentals, the implementation of these two techniques differ in some Assumptions that reflect the characteristics of the environments for which they were developed (Europe for Kühn et al. (2002) and the southern African subtropics for Chevalier et al. (2014)). As Iversen (1944) pointed out that for some plants, higher winter temperatures can compensate for lower summer temperatures and *vice versa* (see also Jeffree and Jeffree, 1994), Kühn et al. (2002) argued that *pdfs* should preferably be bi- or tri-dimensional in order to account for the covariance of climate variables. However, fitting multidimensional *pdfs* for all taxa assumes that they are sensitive to all the variables to be reconstructed. Chevalier et al. (2014) found numerous taxa in southern Africa that were preferentially sensitive to either temperature- or moisture-related variables, and thus only implemented unidimensional *pdfs* in CREST. Both techniques allow the *pdfs* to be multimodal (i.e. the pollen-climate responses can have multiple peaks or optima; Chevalier et al., 2014; Gebhardt et al., 2008).

The use of presence-only distribution data to fit the *pdfs* (in contrast to presence-absence data for which absences have also been reliably assessed) has been criticised, with some arguing that presence-only data cannot be used to estimate reliable niches (Elith et al., 2011; Elith et al., 2006; Phillips et al., 2009; Sniderman et al., 2016). However, true presence/absence data are rarely available, while presence-only data are ubiquitous. Despite known limitations, presence-only data have been used with *pdf* techniques to reconstruct modern environmental parameters successfully (Chevalier et al., 2014) and to produce reconstructions of past conditions consistent with independent proxy records (e.g. Chevalier et al., 2020; Cordova et al., 2017; Litt et al., 2009; Neumann et al., 2007). As they have not been employed in many regions, more applications are necessary to fully assess their strengths and weaknesses. Currently, it is suggested that the definition of the *pdfs* is robust provided that the climate space of the plants' ranges is fully sampled (Chevalier et al., 2014). The recent release of a global calibration dataset enabling using *pdf* techniques from anywhere vegetation currently grows (Chevalier, 2019) could promote testing this hypothesis across various environmental gradients.

For more technical details about the two variants of *pdf* techniques, we refer the reader to Kühn et al. (2002), Gebhardt et al. (2008), Chevalier et al. (2014) and Chevalier et al. (2019). In addition to these two techniques, a third approach has also been recently proposed: CRACLE (Climate Reconstruction Analysis using Coexistence Likelihood Estimation; Harbert and Baryames, 2020; Harbert and Nixon, 2015).

CRACLE is based on similar principles as the other two *pdfs* techniques (*i.e.* individual responses are fitted before being multiplied to estimate climate parameters) but has not yet been applied to pollen data. Reconstructions using Pliocene fossil beetles (Fletcher et al., 2019), plant macrofossils (Fletcher et al., 2017; Harbert and Nixon, 2018) and modern vegetation plots (Harbert and Baryames, 2020) suggest good reconstruction performance and an application to pollen data appears promising.

## 5.2. The regression approach

This approach includes linear and non-linear regression techniques that model the relationships between one or more climate variable(s) and the pollen taxa from a collection of modern pollen assemblages (*i.e.* a ‘calibration’ dataset), before applying these relationships to fossil pollen assemblages.

### 5.2.1. Linear regression techniques

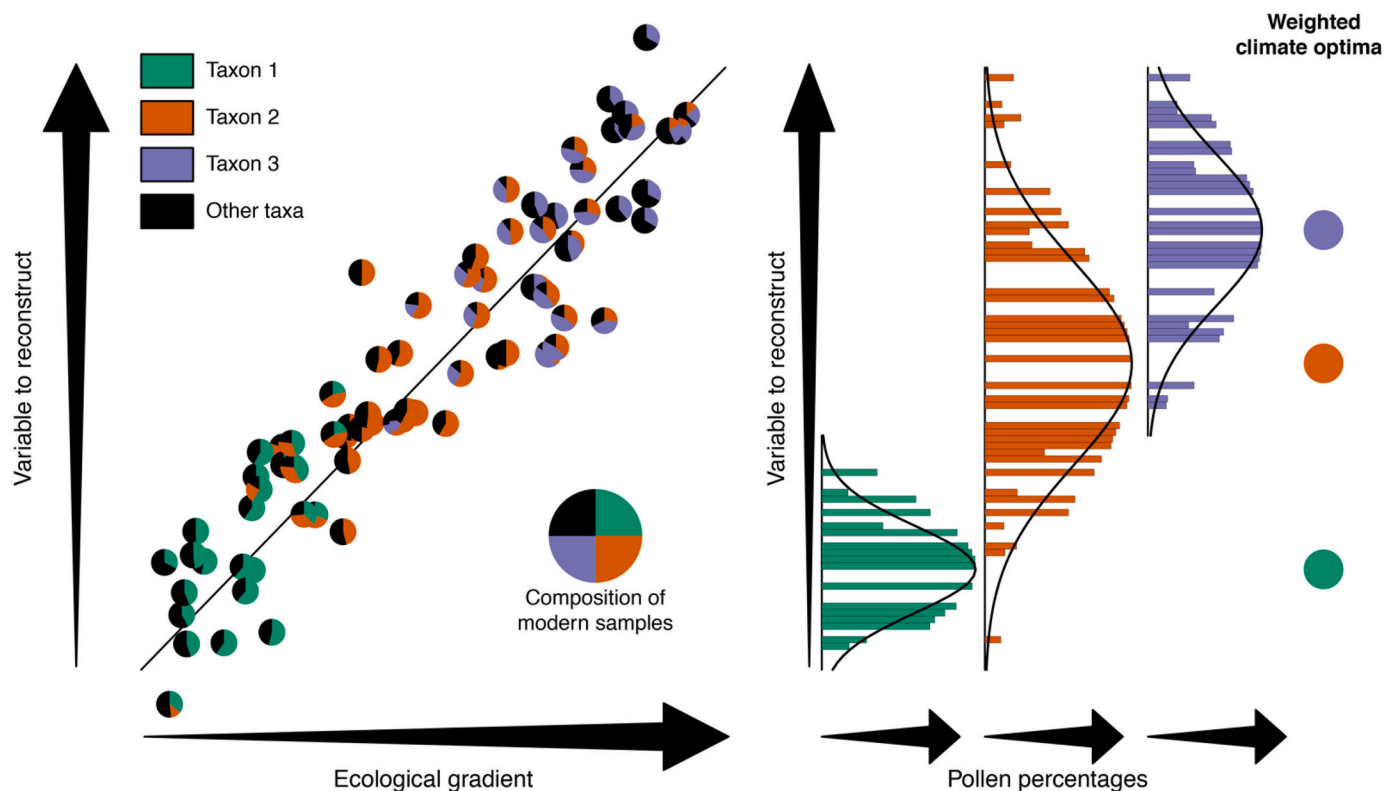
Linear regression techniques were the first regression techniques to be applied to pollen data (Webb and Bryson, 1972) and were extensively used during the 1970s and 1980s (*e.g.* Bartlein and Webb, 1985; Gajewski, 1988; Webb and Clark, 1977). Linear pollen-climate responses mean that any fixed change of the studied climate variable will be mirrored by a constant change in a taxon's percentage (*e.g.* for any increase of 1 °C, the pollen percentage of the taxon increases by 10%), although the amplitude of this change may differ between taxa. The assumption of linearity restricts the range of fossil samples for which an equation can be applied, as pollen-climate relationships are not linear. To employ a linear approach, the strategy is thus to select a small geographical region where the pollen-climate relationships can be linearised, and to develop multiple regression equations that can be applied to fossil pollen data. The trade-off between the number of

equations and the accuracy of these equations is a function of the pollen record, period of time, and past vegetation composition (Bartlein and Webb, 1985; Bartlein and Whitlock, 1993).

Although this technique is no longer used in favour of alternative techniques (described hereafter), some reconstructions have proven to be robust when used in suitable conditions (*e.g.* small calibration region and limited climate change within the studied time). For instance, the eastern North America reconstructions of the last 1000 years made by Gajewski (1988) are nearly identical to reconstructions developed using newer techniques (*e.g.* Wahl et al., 2012) and are still used in global reconstructions (Pages2k Consortium, 2017).

### 5.2.2. Partial Least Square (PLS) regression

Linear regression approaches are prone to violations of their underlying statistical assumptions, such as the linearity of the response of the taxa and the independence of the predictor variables (pollen taxon abundances show a high degree of multicollinearity in pollen datasets; Birks, 1995). Partial Least Squares regression (PLS) – a dimensionality reduction approach (Höskuldsson, 1988; Wold et al., 2001) – has been proposed as an alternative to the classical linear regression-based approaches to account for the covariance of the predictors (*i.e.* many pollen taxa respond similarly to climate variability; Birks, 1998; Birks, 1995). The covariance is removed by combining the predictors together to estimate composite components that are independent of each other (*i.e.* their correlation is zero) and maximise the covariance with the response variable to be reconstructed. The components are then linearly combined to model the relationship between vegetation change and climate. While adding many components improves the apparent statistical performance of the transfer function, it can also lead to an overfitted model that has poor predictive power on the fossil data, despite high performance statistics with the training set(s) (Birks, 1998; Næs, 2002). The optimal number of components can be determined by



**Fig. 6.** Conceptual representation of WA and how climate optima are inferred for the taxa. For simplification purposes, we only focus on three coloured taxa in this example and the ‘black’ areas represent all the taxa that are not considered. (left) Modern pollen assemblages (the ‘ecological gradient’) plotted against the studied climate gradient. (right) Climate values are extracted for each modern sample one taxon occurs in and weighted by the corresponding pollen percentage to derive the ‘climate optimum’ of the taxon (circle on the right).

cross validation (see Section 5.2). Usually, between one and three components are used in the transfer function. As with classical linear regression, the PLS technique is most suitable for calibration datasets covering short environmental gradients where linear responses can be expected.

### 5.2.3. Non-linear unimodal techniques (WA, WA-PLS)

Linear responses are acceptable when focusing on small climate gradients. However, more complex responses become more suitable with longer gradients. The most commonly used non-linear approaches for pollen-based climate reconstructions are weighted averaging (WA) and weighted-averaging partial-least squares (WA-PLS) regression. WA and WA-PLS were first used in palaeolimnology to reconstruct past water chemistry variables, such as pH, salinity, and water depth from fossil assemblages (Birks et al., 1990b; Birks et al., 1990a; Dixit et al., 1991; Oksanen et al., 1988; ter Braak and van Dame, 1989). These techniques account for the non-linear relationships of plant species with their environment. These models assume that in their respective climate niches plant species have the highest abundances and that niche breadth determines the environmental tolerance of the species (Roux, 1979; ter Braak, 1985; ter Braak et al., 1993; ter Braak and Looman, 1986).

As species tend to occur with maximum abundance near the climatic conditions they are most adapted to or competitive in, WA estimates such climate optima from a calibration dataset by calculating the average climatological conditions a species occurs in weighted by the abundance of this species (*i.e.* by calculating a weighted average of the climate values where the taxon is observed; Fig. 6). To reconstruct climate from fossil data, the optima of pollen types in a given fossil pollen assemblage are weighted by their percentage and averaged, assuming that the most abundant types were closest to their optimal climatic conditions at the time the pollen assemblage was deposited (Birks, 1998; Brewer et al., 2007a; ter Braak and Looman, 1986). An extension of WA weights the relevance of individual taxa based on the width of their distribution, with widely distributed taxa receiving lower weight than narrowly dispersed ones (ter Braak and van Dame, 1989), thus assuming that taxa with narrower distributions are more reliable indicators. This ‘tolerance down-weighted’ WA approach has not been much employed, as in practice, it has little effect on the reconstructed values (Juggins and Birks, 2012).

WA performs well in statistical performance tests, and often outperforms other comparable techniques (Birks, 1995; Juggins and Birks, 2012), but it has well-established weaknesses. Since modern samples in the training set are usually unevenly distributed in relation to the environmental gradient of interest, estimation of the optima may be biased (ter Braak and Looman, 1986). The default choice of Gaussian responses can also be problematic for some variables, such as precipitation or moisture-related variables, and near the edges of the environmental gradients where the responses are known to be skewed (Austin and Gaywood, 1994; Guisan and Zimmermann, 2000). As with most techniques, reconstructions using WA also suffer from the so-called ‘edge effect’, where the reconstructed values of the climate variable of interest are biased near the ends of the gradient in the training set (low values are overestimated, while high values are underestimated). To reduce this problem WA applies a correction of inferred values referred to as ‘deshrinking’ (Birks, 1995). A regression model is fit between the ‘reconstructed’ and the ‘target’ values from modern samples to estimate how much the range is reduced (‘shrunk’). The reconstructions are then rescaled with respect to the range of the target (Nolan et al., 2019). A linear model is usually used to estimate and correct the bias in the data, but monotonic smooth spline regressions may provide a better fit (Birks and Simpson, 2013).

WA-PLS was later developed to address some of the problems identified in WA (ter Braak et al., 1993; ter Braak and Juggins, 1993). WA-PLS combines some advantages of WA regression with the multicollinearity reduction approach of PLS. In practice, WA-PLS with one

single component is equivalent to WA with inverse deshinking. The additional components use the residual structure in the species data (*i.e.* the part of the data that are not explained by the first component) for improving the species optimum estimates (ter Braak et al., 1993). As such, they can be seen as additional WA fitted on the residuals to increase the extracted climate signal (Nolan et al., 2019). One benefit of PLS is that WA-PLS suffers less from the edge effect than WA. As in PLS regression, the selection of the number of WA-PLS components to include should be based on cross-validation (see Section 5.2), as selecting too many components will result in a model with poor predictive power. The final model should represent the minimal adequate model (*i.e.* having fewest components possible), a low root-mean-square error in cross-validation and low maximum bias. In practice, WA-PLS models based on two or three components perform best, as their inclusion still usually represents an improvement from a one-component WA-PLS model, while avoiding the risk of overfitting. Additionally, WA and WA-PLS provide robust solutions to account for many other intrinsic characteristics of palaeoecological datasets, such as the high number of taxa, many zero values (*i.e.* absences), data expressed as percentages and covariance among taxa (*e.g.* Birks, 1998; ter Braak and Juggins, 1993).

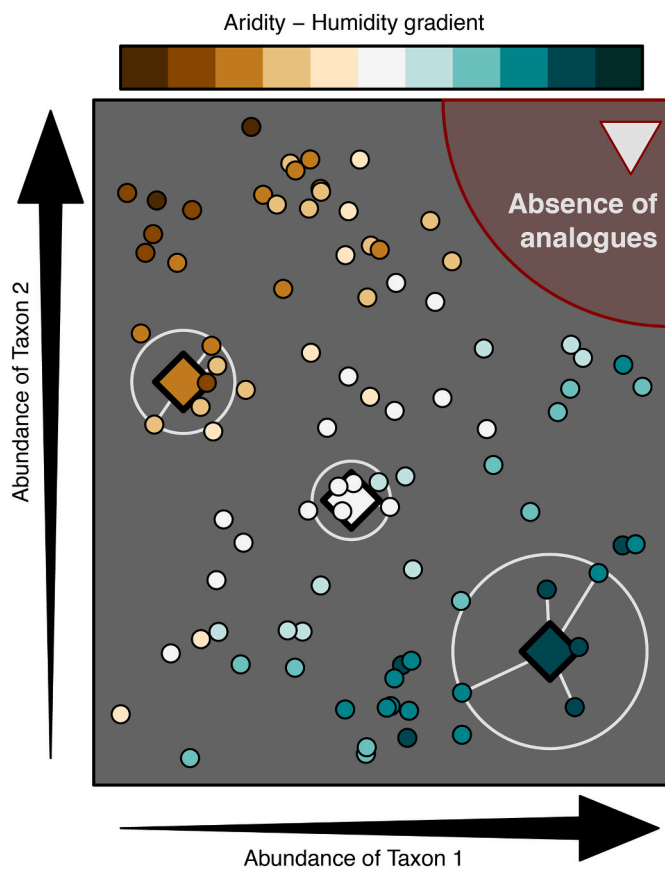
WA-PLS has been widely used with pollen to reconstruct late Pleistocene and Holocene precipitation in China (Chen et al., 2015; Li et al., 2017; Li et al., 2015; Lu et al., 2018), and temperature in Siberia (Klemm et al., 2013) and in Europe (Finsinger et al., 2010; Lotter et al., 2012; Lotter et al., 2000; Sakari Salonen et al., 2013; Seppä et al., 2009; Veski et al., 2015). Tests with simulated and real training sets have shown that WA-PLS performs well in various ecological and climatic conditions, especially with training sets with high species turnover and low noise (Juggins and Birks, 2012). Interestingly, Lotter et al. (2000) have shown that unimodal response models (WA and WA-PLS) performed at least as well as their linear counterpart (PLS) over a short temperature gradient and performed significantly better when including a much larger area. This is one of the reasons why unimodal models are commonly preferred over simpler linear responses, as they are less sensitive to the size of the studied climate gradient. However, WA and WA-PLS may not be optimal techniques in continental or global scale studies, as the responses of some pollen taxa to the variable of interest can become multimodal.

### 5.3. The classification approach

Classification techniques compare fossil pollen assemblages to collections of assemblages for which climate is known to identify which assemblages are most similar to the fossil ones. These techniques do not estimate the optimum or tolerance of the taxa and are called ‘non-parametric’. The most similar assemblages are called ‘analogues’ and are used to reconstruct climate.

#### 5.3.1. The Modern Analogue Technique (MAT)

MAT was first proposed in palaeoclimatological studies by Hutson (1978) for marine data and by Overpeck et al. (1985) and Guiot and Pons (1986) for pollen data. MAT is a specific version of a broader statistical technique called ‘k-nearest-neighbours’ (k-NN; Cover and Hart, 1967) adapted to palaeoecological datasets. The approach is based on the measure of the degree of similarity (‘degree of analogy’) between fossil pollen assemblages and elements of a collection of modern assemblages (Jackson and Williams, 2004). The climate values associated with the most similar pollen assemblage(s) (*i.e.* the closest modern analogue(s)) are averaged, possibly using a weight defined by the measure of similarity, to estimate the climate related to fossil assemblages (Fig. 7, Overpeck et al., 1985). While MAT does not explicitly model the modern response of the taxa to climate, some methodological decisions still have to be made to estimate a climate value, such as defining a metric to measure the similarity between modern and fossil assemblages (*i.e.* the dissimilarity metric), the



**Fig. 7.** Conceptual diagram describing MAT. Two taxa with different aridity/humidity preferences coexist. Taxon 1 prefers humid climates (green samples), while Taxon 2 prefers drier conditions (brown samples). In this example, three fossil pollen assemblages (diamonds with thick black borders) are compared to a set of modern assemblages, each one associated with an aridity/humidity value (coloured circles). The five closest analogues (defined by the white circles) are retained and averaged to estimate past conditions. The white triangle in the top right corner represents a fossil sample with high proportions of both Taxon 1 and Taxon 2, a type of assemblage that does not have close analogues in the modern environment, as indicated with the red area. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

number of closest analogues to retain, the threshold for non-analogue values (*i.e.* the dissimilarity value after which samples are considered too different to be reliable analogues) and the spatial and climatic coverage of the modern samples.

The similarity between pollen samples can be measured in different ways (Overpeck et al., 1985; Prentice, 1980). The squared chord distance (SCD) is routinely used with MAT, as it works well with most pollen assemblages. SCD is a particular distance metric that down-weights the most abundant taxa, while maintaining the relative ranking of the taxa (Overpeck et al., 1985). As a result, the signal of the taxa with low pollen production rates is not masked by the taxa with higher rates. The threshold that determines which modern pollen samples sufficiently resemble the fossil pollen assemblage to be considered reliable analogues depends on the parameters of the study, including the size and coverage of the modern samples set (*i.e.* how many modern samples are available and how they are distributed in the geographic and climate spaces), the taxonomic diversity of the samples and the scale of the study (Gavin et al., 2003; Jackson and Williams, 2004; Overpeck et al., 1992). The definition of a good ‘analogue’ has been discussed since the creation of the approach (Anderson et al., 1989; Sawada et al., 2004; Simpson, 2012) and is frequently presented as

threshold values. A ‘no-analogue’ situation occurs when none of the assemblages from the training set resemble the fossil assemblage with sufficient accuracy to be considered an acceptable analogue (Fig. 7) and can be caused by either a restricted collection of modern samples or by past conditions that have no counterparts in the modern settings.

Another critical choice relates to the number of analogues to use to reconstruct climate. While it may seem more intuitive to retain the closest analogue alone, as it is the one that is most similar to the fossil sample, selecting more than one is advantageous as it helps to account for the natural variability of pollen data and minimize sampling error. For example, modern samples taken from a particular site a few years/meters apart are likely to exhibit some differences, despite experiencing the same climate (see Section 3). Selecting too many analogues can, however, induce undesired bias, especially near the limits of the designed climate space (the ‘edge effect’ defined in Section 4.2.3). For example, to reconstruct the aridity/humidity for a site located in the wettest part of the climate space represented by the training set (bottom right corner on Fig. 7), selecting more and more analogues will progressively include drier samples from the rest of the climate space represented in the calibration data, thus inducing an undesired dry bias on the reconstruction (Gajewski, 2015; Viau et al., 2008). The selected number of analogues can be based on either 1) selecting all the samples with a dissimilarity measure lower than a defined threshold, 2) selecting a fixed number of analogues, or 3) a combination of these two factors. The choice should be based on an assessment of the model prediction error by cross validation (Section 5.2.3 and *e.g.* Simpson, 2012).

In a broad-scale cross validation study across North America comparing modern samples against themselves, Williams and Shuman (2008) concluded that a small number of closest analogues, from 3 to 7, outperforms the use of only the closest analogue. They also investigated the effect of different weighting types to average the climate values of the selected analogues: the ‘no weighting’ approach, in which all the analogues have the same weight (*i.e.* the same importance), and distance-based weighting (the inverse of the distance and the inverse of the squared distance), whereby closer analogues have higher weights. These different weighting types have little impact on continental-scale reconstructions, with the inverse-distance weighting being the most commonly used (Cao et al., 2017; Marsicek et al., 2018; Mauri et al., 2015). With distance-based weights, the choices of the number of analogues is less critical as poor analogues will be given little weight.

Errors of MAT reconstructions are commonly estimated from the climatic dispersion of the selected closest modern analogues around the fossil data (Guiot, 1990). This definition is problematic as it implies errors cannot be calculated for reconstructions based on one analogue only. An alternative option is to use the results of the modern validation of the dataset (Section 5.2.3) as a proxy for the mean error of reconstruction (*e.g.* the root-mean-square error of prediction; RMSEP). None of these options provide an accurate estimation of the true error associated with a climate reconstruction. In addition, these indices do not include many factors that are likely to inflate the uncertainties, in particular the uncertainties related to the analogue distance. For example, the selected closest analogues may be distant analogues of the fossil pollen samples (*i.e.* not so good analogues) but be close analogues to each other (*i.e.* experiencing similar climates). This would result in a small climate uncertainty, despite a poor selection of analogues. In addition, Telford and Birks (2005, 2009, 2011a) have shown that because MAT tends to select analogues that are geographically close to each other or to the target sample, the uncertainties of the reconstruction can be artificially reduced (Telford and Birks, 2009), as spatially close analogues are likely to experience very similar climates. This violates the assumption of the independence of the modern samples. Telford and Birks (2009) proposed a graphical test to assess if spatial autocorrelation is likely to have a strong impact on the reconstructions and recommend using an *h*-block cross validation to assess the predictive skill of the model (Section 5.2.3).

As MAT does not require an explicit model of the pollen-climate relationship, these relationships are not required to be either linear or unimodal over the study area. Therefore, the climatic and environmental span of the modern dataset can be much larger than for other techniques. The most recent MAT developments have thus been primarily focused on increasing the size of the modern datasets to include larger climatic and ecological gradients and increase the chance of finding closer analogues. Larger datasets facilitate the application of MAT on extensive networks of fossil pollen datasets (tens to hundreds) and thus they enable reconstructing continental-scale climate maps (e.g. Davis et al., 2003; Gajewski et al., 2000; Ladd et al., 2018; Mauri et al., 2015; Sawada et al., 2004) or time-series (e.g. Marsicek et al., 2018; Viau et al., 2008). However, including more analogues also increases the risk of false positive matches, especially when the calibration dataset encompasses wide spatial areas where the low taxonomic resolution of pollen data can lead to similar pollen compositions to be observed (low SCD values) despite experiencing very different climates. Williams and Shuman (2008) have proposed that several broadscale North American taxa (e.g. *Pinus*, *Alnus* or *Tsuga*) could be split into eastern and western pollen taxa to circumvent this issue. Their regional splitting allows capturing the specific climate response of each subgroup and reduces the risk of false positives. Such an approach is also useful to other techniques, such as WA/WA-PLS or the *pdf* techniques, to estimate more local climate responses (e.g. Donders et al., 2011).

5.3.2. The Revised Analogue Techniques (RAM)

The Revised Analogue Technique (RAM) is composed of two distinct steps: a pollen data modelling step to interpolate modern samples using a technique called Response Surface (Bartlein et al., 1986; Huntley, 1993; Prentice et al., 1991; Webb et al., 1993) and a MAT-based climate reconstruction step (Gonzales et al., 2009; Waelbroeck et al., 1998). With Response Surface, the climate response of individual pollen taxa is estimated for two or more variables simultaneously. Pollen percentage data are mapped in a two- or higher-dimension climate space of interest

(e.g. July temperature and annual precipitation for instance, such as in Anderson et al., 1991) and a data smoother is used to interpolate the values of the pollen percentages as a function of these variables. The choice of the smoother depends on the nature and quality of the modern pollen sample collection and the goal(s) of the study, but it has commonly been either global 2<sup>nd</sup> or 3<sup>rd</sup> order polynomial curves (Bartlein et al., 1986) or locally weighted smoothers (e.g. LOWESS; Huntley et al., 1989; Prentice et al., 1991). As for any regression-based approach, evaluation of model residuals is important to ensure a good fit. The series of response surfaces fitted for selected taxa are then stacked together to create novel pollen assemblages at every grid point of the climate space. These ‘composite’ modern pollen samples can then be used to fill in coverage gaps within the sampled climate space. The percentages of these composite samples should be rescaled to a 100% to be comparable with the modern and fossil pollen samples. Because the response surfaces are regression models, vegetation compositions from non-analogue compositions can also be inferred through extrapolation by assuming, for instance, symmetric pollen responses surfaces (Gonzales et al., 2009). However, extrapolating too far from the observed pollen samples creates unreliable pollen samples and palaeoclimatic inferences.

The climate reconstruction step is classically performed with MAT using all the modelled pollen assemblages, but any reconstruction technique could be employed at this stage. One advantage of the response surfaces approach compared to the classical MAT relies on increased ecological and climate coverage provided by the extended pool of modern pollen samples. It also filters out non-climatic noise in the pollen data and only retains primary pollen-climate signals. The higher coverage within the existing climate space increases the chances of finding closer analogues to the fossil pollen assemblages (Gonzales et al., 2009; Xu et al., 2010). One fundamental issue of the first RAM studies is that the composite pollen samples estimated from the response surfaces were most often only derived from a limited number of real pollen samples. This can bias the reconstructions and reduce the

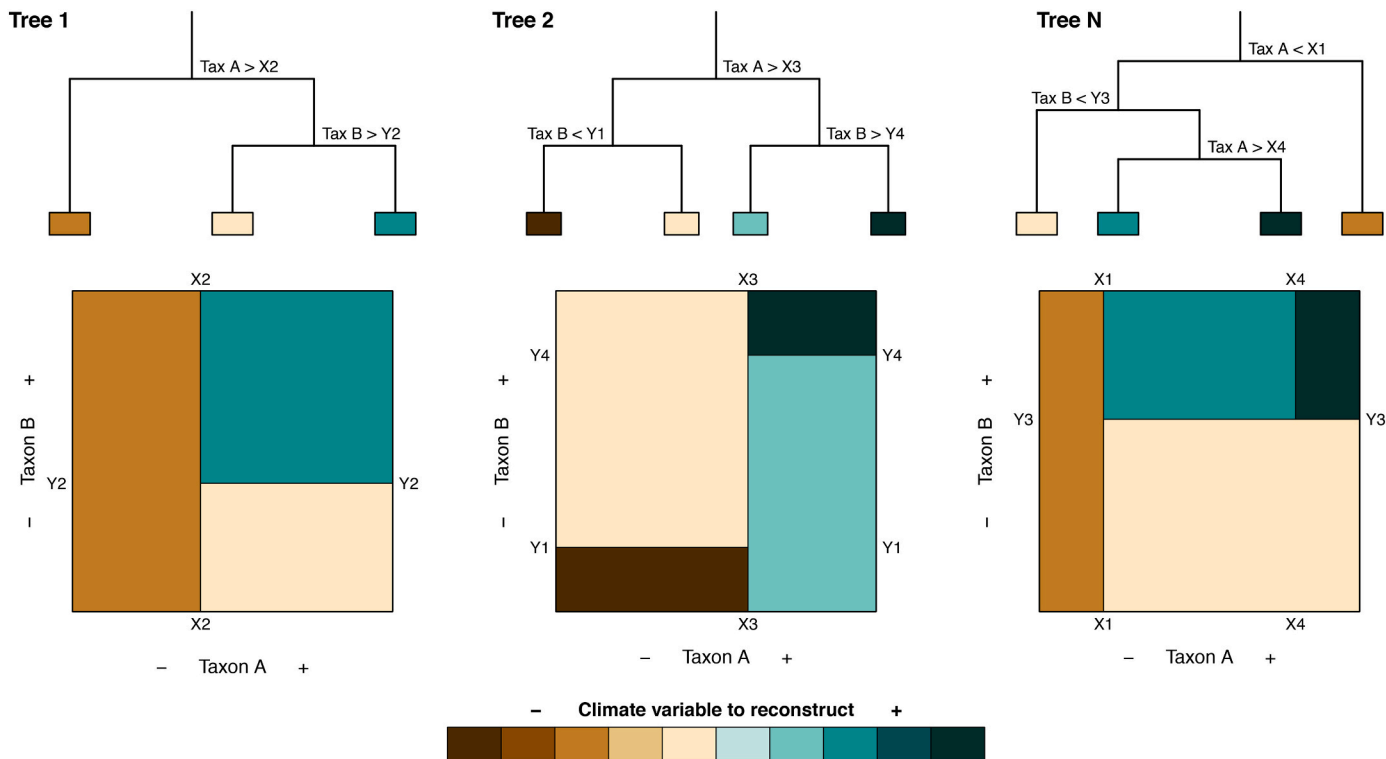


Fig. 8. Schematic representation of regression trees. In this example, N trees are fitted on a pollen dataset composed of two taxa. The different trees split the samples according to the different pollen percentages to estimate climate values. To reconstruct climate from an ensemble of trees fitted on modern data, the fossil pollen sample should be classified by each individual tree and the ensemble of estimated climate values averaged.

reconstruction errors, which makes RAM appear to outperform other techniques (Telford et al., 2004). RAM has not been much used because of these known issues but the development of progressively larger and more detailed compilations of modern pollen samples (e.g. Davis et al., 2020; Whitmore et al., 2005; Zheng et al., 2014) represents a possible opportunity to revisit this approach and explore its relative strengths and weaknesses.

### 5.3.3. Artificial Neural Networks (ANNs)

ANNs were introduced to continental palaeoclimatology by Guiot et al. (1996) and to palaeoceanography by Malmgren and Nordlund (1997) to better account for the non-linear and sometimes multimodal response of palaeoenvironmental proxies. An ANN is a statistical non-linear regression technique that consists of a number of simple, interconnected mathematical functions, called neurons, that process the signal coming from input variables (i.e. the pollen assemblages) to estimate the output environmental variables (e.g. Caudill and Butler, 1994). The interconnections between taxa and neurons are defined by coefficients that are iteratively tuned to best fit the output variables. ANNs allow very close data fitting, although it is difficult to analyse the causality among input and output variables (Gurney, 2014; Rumelhart et al., 1994). The number of input neurons corresponds to the number of taxa and the number of neurons in the output layer is equal to the number of climate variables to be reconstructed (Gurney, 2014; Rumelhart et al., 1994). In most pollen-based applications, the architecture of the ANN is composed of one hidden layer with a number of neurons defined by the complexity of the problem. At each neuron, the incoming signal is transformed by a non-linear function (called activation function) and sent to the next layer. An example of a possible activation function is the sigmoid, which translates the input in the [0,1] domain and attenuates extreme values by introducing non-linearity in the model. One of the most commonly used ANN techniques is the feedforward network trained with a back-propagation learning algorithm (e.g. Rumelhart et al., 1988).

The most recent applications of ANNs for assemblage data come from the marine realm and the study of foraminifera or dinocyst assemblages to estimate past sea-surface temperature (de Vernal et al., 2005; Kucera et al., 2005; Malmgren et al., 2001; Peyron and de Vernal, 2001). The main reason why ANNs are not more frequently used in palaeoclimatology is related to the complexity of finding the proper structure of the network. Because of its non-linear property, ANNs can fit the training data set very well, integrating the noise inherent to any

pollen dataset as climate signals. This overfitting of the data restricts the predictive power of the transfer function outside of the range of the training set. In addition, understanding why an ANN produces an output from a set of input variables is challenging (statistical black box), rendering analysis and causal interpretations of the results difficult and sometimes speculative. Overall, ANNs do not perform better than the simpler reconstruction techniques (Salonen et al., 2019) and, as such, have been little used to reconstruct climate from fossil pollen data.

### 5.3.4. Regression trees

Regression trees (or decision trees) are a specific family of machine-learning techniques that have been increasingly used with ecological data (De'ath, 2007; Elith et al., 2008; Prasad et al., 2006). Classification trees are used to partition the data by successively separating the pollen samples based on the relative pollen percentages (Fig. 8). Single regression trees have often limited predictive power. As such, more complex techniques based on the same principles have been developed and recently been applied to paleoecological data (Salonen et al., 2019). Random Forest (RF; Breiman, 2001; Prasad et al., 2006) is based on the estimation and combination of a large number of regression trees (i.e. thus creating a 'forest'), each tree being estimated from a randomised ensemble of modern pollen samples by bootstrap. If we call  $n$  the number of samples composing the modern dataset, the bootstrap approach consists of randomly sampling  $n$  samples with replacement to create a new dataset of the same size. In this case, some samples are selected multiple times, and some are never selected. Unselected samples become part of the test dataset. An alternative approach based on the combination of regression trees is called Boosted regression Trees (BRT; De'ath, 2007; Elith et al., 2008). BRT differs from RF in the definition of the random modern datasets. In RF, each sample gets the same probability of being selected, while in BRT the samples that were insufficiently described in the previous tree get a higher probability of being selected. This approach is called 'boosting' and increases the performance of the model over the elements that are least well predicted.

Regression trees have not been much used but their performance was similar to the classical MAT, WA and WA-PLS under different testing conditions, despite having higher uncertainties (Nolan et al., 2019). One potential negative of these approaches is that the model does not take advantage of ecological theory in any way, which can sometimes lead to unrealistic results. While more testing is necessary,

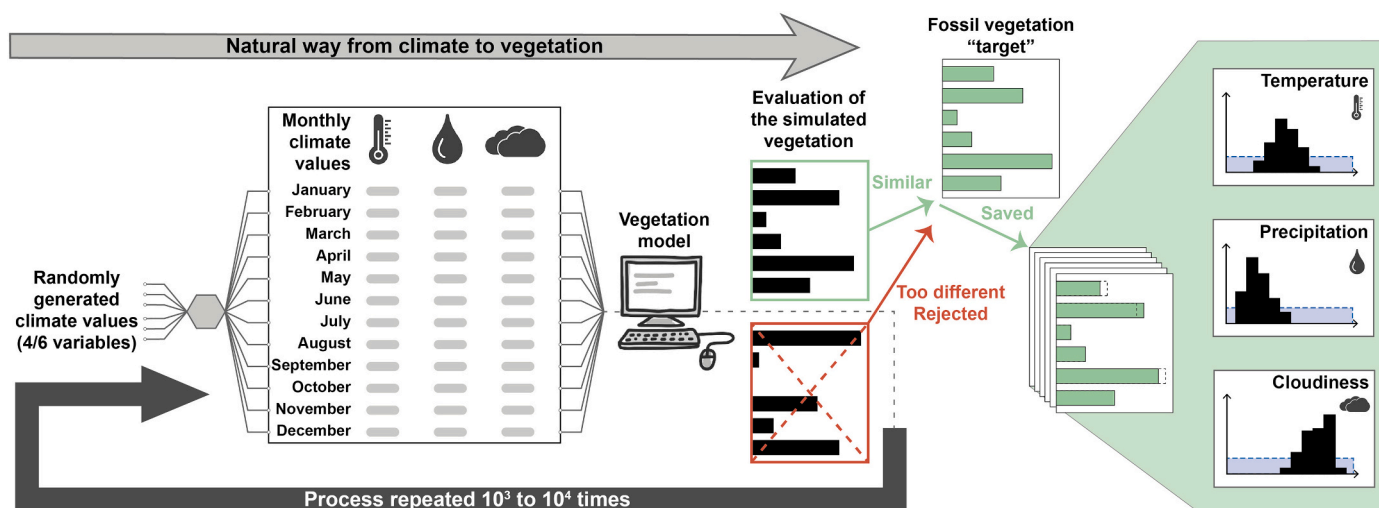


Fig. 9. Schematic representation of the inverse modelling approach. From left to right: Random states of four or six climate variables are generated and then transformed into 36 monthly values of temperature, precipitation and cloudiness using empirical relationships. The 36 variables are used to drive a vegetation model, and the outputs are compared to a 'target' fossil assemblage. Climate conditions associated with similar simulated assemblages are saved to generate probabilistic distributions of climate. The process is repeated a few thousand times to estimate distributions for climate variables (Guiot et al., 2000).

the work of Salonen et al. (2019) and Nolan et al. (2019) strongly suggests that machine-learning algorithms (RF, BRT but also ANNs) could become part of the pool of more widely used reconstruction techniques in the near future.

#### 5.4. The process-based approach

As opposed to the techniques that find local or global relationships between modern and fossil pollen data, the process-based approach explores different climate scenarios to identify which conditions are the most able to reproduce the fossil data using vegetation models.

##### 5.4.1. The inverse modelling technique

Guiot et al. (2000) introduced the inverse-vegetation modelling approach to address the low CO<sub>2</sub> concentration issue (see Section 3.4). In contrast to the standard statistical techniques that work from pollen assemblages towards climate (Fig. 3), Guiot et al. (2000) developed computational algorithms to maintain the causal relationship from climate towards vegetation. Climate values are randomly sampled from uniform distributions with prescribed ranges (the blue rectangular distributions in Fig. 9) to produce a large variety of climatic scenarios that are then fed to vegetation models that explicitly represent the mechanistic effects of atmospheric CO<sub>2</sub> concentrations on vegetation, such as BIOME4 (Kaplan, 2001) and BIOME5-beta (Izumi and Bartlein, 2016). These vegetation models are used to simulate the equilibrium vegetation productivity associated with the input climate scenarios that can then be compared with fossil pollen assemblages. Sometimes referred to as ‘Inverse Modelling through Iterative Forward Modelling’ (IMIFM), Inverse Modelling has been used to successfully reconstruct climate over multiple regions, such as Europe (Guiot et al., 2000), Africa (Lézine et al., 2019; Wu et al., 2007b) and North America (Izumi and Bartlein, 2016).

Since vegetation models require a large range of climate parameters (36 monthly values of temperature, precipitation and cloudiness), generating plausible random climate scenarios is a critical issue. The common strategy is to reduce the dimension of the input climate vector from 36 elements to four (January and July temperature and January and July precipitation; Guiot et al., 2000) or six (the same four variables complemented with April and October precipitation; Izumi and Bartlein, 2016) and to use empirical modern relationships to derive the other monthly variables. These climate values are then fed to a vegetation model. One limitation of Inverse Modelling relates to the outputs of vegetation models that can only simulate the net primary productivity of a reduced number of Plant Functional Types (PFTs). PFTs are broad plant classification units defined by stature, leaf-form, phenology and climatic adaptations that were originally developed to enable comparisons of pollen data and vegetation model outputs and promote global-scale studies of vegetation-climate interactions, and have been used to identify distributional changes of biomes in geographic space (e.g. Ni et al., 2010; Prentice et al., 1996; Trondman et al., 2015; Williams et al., 1998). The simulated PFT productivities predicted by the vegetation model for a particular randomly generated climate are then compared to the fossil pollen-derived PFT assemblage and the climates that produce reasonably coherent PFT assemblages are retained. The process is repeated until the model converges, which usually occurs when a few thousand climate scenarios have been retained. The distribution of the retained climate scenarios estimates the distribution of climate that can most likely simulate biomes matching the fossil data (Fig. 9). In addition to climate variables, this approach can also estimate other variables, including (1) the fractionation of  $\delta^{13}\text{C}$  in the plant, which can be compared with isotopic measurements in bulk sediment (Hatté and Guiot, 2005) and (2) the water absorbed by the plant versus runoff, which can then be directly compared to lake-level data when the data are obtained from lacustrine sediments (Guiot et al., 2009).

Inverse Modelling circumvents some issues associated with the most

standard approaches, including accounting for the physiological effect of the low CO<sub>2</sub> atmospheric concentration during recent glacial periods. Based on mechanistic vegetation modelling, Inverse Modelling is also fairly insensitive to non-analogue climates or vegetations. It is, however, highly dependent on the vegetation model used and despite its unique properties, the inverse modelling technique has several weaknesses that have limited its application. One of the most important criticisms relates to the nature of vegetation model outputs, which are not strictly comparable with pollen assemblages. PFT productivity at a location differs strongly from pollen percentages in both its meaning and botanical resolution. As a result, a correspondence metric has to be defined to match vegetation model outputs (PFT primary production) with pollen percentage data, which adds another layer of complexity and increases the uncertainties (Guiot and Kaniewski, 2015). Vegetation models are also based on numerous modelling assumptions that define how the different PFTs behave under different climate scenarios. These assumptions are often complex to evaluate with high accuracy, so that the simulations integrate many uncertainties (e.g. initial state, parameters, process representation) that will be propagated to the palaeoclimate reconstructions (Dietze et al., 2018). Further developments and testing are probably required for its establishment as a commonly employed tool for quantitative palaeoclimatology.

#### 5.5. Bayesian approaches

Bayesian statistical models differ from the aforementioned models in that they treat the model parameters and the unobserved data as random variables. This allows for the characterisation of uncertainty in both the data and processes and permits the use of prior information about the pollen-climate relationship when available. The Bayesian framework allows for the estimation of a posterior probability distribution of climate variables. The complexity of the Bayesian approach has limited its applicability in the past, but recent software developments have resulted in increased use of Bayesian models and inference.

##### 5.5.1. Bayesian pollen-based climate reconstruction techniques

The Bayesian approach is not a reconstruction technique in itself. Instead, it describes a way of specifying observed and unobserved data and variables and the relationships among them. While models specified using a more traditional approach are used to estimate the single most likely climate scenario given the data, Bayesian approaches estimate the posterior distribution of the climate variable given the data (this approach is similar to that used by the *pdf* techniques and Inverse Modelling). As is the case with non-Bayesian approaches as well, many climate reconstruction models can be derived from the same set of underlying Bayesian principles (Berliner, 1996). A Bayesian model specifies the relationships between the data (fossil pollen), the process (formalised conceptual knowledge about the unobserved natural process), and the *prior* assumptions (description of our beliefs about the values of the quantities to be estimated). The estimation of the *posterior distributions of model parameters* (i.e. in this case the climate variable of interest) is typically done using Markov Chain Monte Carlo (Gilks et al., 1996; Gilks et al., 1995), although other algorithms do exist.

In the case of Bayesian pollen-climate models, the data are typically modelled as counts of fossil pollen – as opposed to pollen percentages, where the total pollen count for a sample represents the counting effort or the concentration/preservation in the sediments – and the taxon-specific pollen count in a sample informs about relative abundance of that taxon given the total pollen count in that sample. The observation model is typically implemented as a multinomial distribution where the parameters of interest are the compositional probabilities that determine the likelihood of finding a particular taxon in a fossil sample of a given total count (i.e. the statistical law that describes the probability of observing the taxon *k* times in a sample of *n* pollen grains, knowing the probability *p* of observing the taxon). The underlying, unobserved relationship between taxonomic relative abundances and climate

variable(s) of interest is the process model. It can include the formalisation of any phenomena thought to determine the probabilities of counts, including unimodality, non-linear responses to climate or other effects. In addition, spatiotemporal autocorrelation (*i.e.* the measure of the connectedness between samples as a function of their proximity in space and/or time) in the species responses to the climate variable(s), or between the climate variable(s), can be accounted for in the process model. The *prior* model completes the specification and is used to calculate the posterior distribution. A particular benefit of a Bayesian approach is the principled, probability-based estimate of uncertainty, which is in strong contrast with some of the traditional approaches. While uncertainties of the different regression and classification techniques are errors around the best fit, Bayesian uncertainties are inherently different and summarise all the processes in action. Further, inference of the full probability distribution of past climate histories makes it possible to estimate characteristics of these climate histories, including the timing of specific events, and trends in both local and global climatic changes.

Several Bayesian approaches have been developed to reconstruct climate from fossil data (*e.g.* Holmström et al., 2015; Ilvonen et al., 2016; Parnell et al., 2016a; Tipton et al., 2019). The first examples of Bayesian modelling application in palaeoclimatic studies used chironomid records from lake sediments to reconstruct past temperatures (Korhola et al., 2002; Toivonen et al., 2001; Vasko et al., 2000). Toivonen et al. (2001) proposed the Bayesian Unimodal Model (BUM) that can be viewed as a Bayesian counterpart of weighted-averaging (WA, see Section 4.2.2) and Gaussian logit models (Birks et al., 1990b; ter Braak and van Dame, 1989; not discussed in this review). Vasko et al. (2000) introduced a modification of BUM, the Bayesian hierarchical Multinomial Regression model (BUMMER) that accounts for dependency among species. Adopting some elements of the BUMMER model, Haslett et al. (2006) were the first to use Bayesian hierarchical modelling in pollen-based climate reconstructions. Their model was further developed in Salter-Townshend and Haslett (2012), who complexified the response curves of the pollen counts to climate to account for multimodality, and Parnell et al. (2015), who included age uncertainties in the reconstruction. In parallel, the BUMMER model was also modified to better correspond to pollen data by Holmström et al. (2015). Ilvonen et al. (2016) and Salonen et al. (2012) applied modifications and extensions of BUMMER to reconstruct temperature from pollen records, and Li et al. (2015) used it to reconstruct past

precipitation. For an in-depth review of Bayesian approaches to pollen-based reconstructions, see Ohlwein and Wahl et al. (2012).

#### 5.5.2. Advantages, disadvantages and ongoing developments

The Bayesian approach provides a natural framework for hierarchical modelling because the model is defined using conditional probability distributions (*e.g.* Tingley et al., 2012). The relationships between various components can be explicitly described by specifying the conditional dependencies among variables, allowing for a scaffold model development approach, starting with a simple structure and increasing complexity as needed (*e.g.* Hobbs and Hooten, 2015). For example, it is possible to start with the single pollen record BUMMER model (Vasko et al., 2000) and then increase complexity by adding modules for multiple pollen records, dating uncertainty, or complicated response surfaces (Holmström et al., 2015; Ilvonen et al., 2016; Parnell et al., 2015; Tipton et al., 2019). Bayesian hierarchical models are sufficiently flexible to model any of the processes deemed to be relevant to the generation of the observations. For example, if it is hypothesized that the response of a pollen taxon varies across a landscape, that assumption can be included in the model by allowing for spatially varying coefficients. Or if spatial or temporal autocorrelation is present, it can be directly modelled from the data.

The Bayesian approach makes it also possible to incorporate disciplinary expertise or pre-existing information in the form of a *prior* knowledge in the pollen-climate model. Different types of *prior* information about past temperature history have been used. For example, Holmström et al. (2015) constrained the variability of their Holocene reconstructions by the variability simulated by the NCAR Climate System Model, while Guiot and Kaniewski (2015) used 20th century variability inflated by 1.5 as a *prior* to represent Holocene variability. More recently, Weitzel et al. (2019) used the spatial structure simulated by climate models to constrain the reconstructions and extrapolate them into surfaces of past climate anomalies. *Prior* information about past climate histories can also be borrowed from scientific expertise, as it is known that some periods of the late Pleistocene were colder/warmer than present and that abrupt, high-magnitude climate change events were more likely/unlikely to occur. In Bayesian modelling, it is possible to account for the possibility of the occurrence of abrupt changes by specifying a *prior* that describes the expected behaviour in a dynamic and volatile climate system (Parnell et al., 2016a). Conversely, in the absence of certainty in the prior information, *priors* can be

**Table 1**  
Main strengths and weaknesses of the climate reconstruction approaches presented in Section 4.

Method	Strengths	Weaknesses
<i>Pdf</i> techniques	<ul style="list-style-type: none"> <li>● Globally applicable</li> <li>● Good estimation of the uncertainties</li> <li>● Can deal to some extent with non-analogue situations if species-environment relationships are well constrained</li> </ul>	<ul style="list-style-type: none"> <li>● Based on the response of plants (instead of pollen) to climate</li> <li>● Use of presence-only data</li> </ul>
Regression techniques	<ul style="list-style-type: none"> <li>● Useful at the local to regional scale</li> <li>● Can deal to some extent with non-analogue situations if species-environment relationships are well constrained</li> </ul>	<ul style="list-style-type: none"> <li>● Often assume pre-defined shape of the pollen-climate relationships (<i>e.g.</i> linear or unimodal).</li> <li>● Struggle with complex (<i>e.g.</i> multimodal) species responses</li> </ul>
Calibration techniques	<ul style="list-style-type: none"> <li>● No assumptions about the shape of the pollen-climate relationships</li> <li>● Applicable at broad spatial scales</li> <li>● Limited performance in no-analogue situations</li> </ul>	<ul style="list-style-type: none"> <li>● Sensitive to spatial autocorrelation (artificial reduction of the uncertainties)</li> <li>● Poor estimation of the uncertainties</li> <li>● High sensitivity to the dissimilarity metric and dissimilarity threshold</li> </ul>
Machine learning (ANN, BRT, RF)	<ul style="list-style-type: none"> <li>● Good fitting of the calibration data</li> <li>● Potentially applicable to complex species-environment relationships</li> <li>● Stochastic approaches</li> </ul>	<ul style="list-style-type: none"> <li>● Statistical black boxes</li> <li>● Performance statistics often similar to simpler approaches</li> <li>● Can overfit the data in noisy datasets</li> </ul>
Bayesian approaches	<ul style="list-style-type: none"> <li>● Explicit and rigorous characterisation of uncertainties</li> <li>● Allow for the formalisation of processes that generate the data (<i>e.g.</i> use of counts instead of percentages)</li> <li>● Use of <i>a priori</i> knowledge</li> </ul>	<ul style="list-style-type: none"> <li>● Specialist expertise required to specify the model and estimate parameters</li> <li>● Specifying appropriate prior probability distributions may be challenging</li> </ul>
Inverse modelling	<ul style="list-style-type: none"> <li>● Relies on ecologically realistic forward modelling of how climate drives vegetation change</li> <li>● Can model physiological effects of CO<sub>2</sub> atmospheric concentration</li> <li>● No modern calibration dataset needed.</li> </ul>	<ul style="list-style-type: none"> <li>● Simulates PFT productivity and not pollen</li> <li>● Low taxonomic resolution</li> <li>● Strongly dependent on available vegetation model</li> </ul>



specified in a way that minimises their contribution to the inferred posterior result.

Finally, the modular structure of a hierarchical Bayesian model also allows for accounting of uncertainty in processes and observations in the individual model components. These uncertainties are propagated through the different stages of the model, making it possible to specify the total uncertainty in the reconstructed climate variable and attribute this uncertainty to different sources. Formally, the uncertainties are typically represented by parameters that describe the variance of a distribution and their propagation through the modelling steps takes place in a systematic, principled manner following the general laws of probability (e.g. Hobbs and Hooten, 2015). For example, many recent studies have stressed the importance of considering the chronological error in pollen-based climate reconstructions (Blaauw et al., 2018; Breitenbach et al., 2012; Brewer et al., 2008; Chevalier and Chase, 2015; Ilvonen et al., 2016; Parnell et al., 2015). As the chronologies used in the reconstructions are never error-free, the chronological error introduces uncertainties to the estimations that should be part of the statistical model used in the reconstruction. Parnell et al. (2015) introduced a fully Bayesian model for palaeoclimate inference with detailed and sophisticated handling of time uncertainty by adding the Bayesian *Bchron* algorithm as a chronology module to the hierarchical reconstruction model (Parnell et al., 2015; Parnell et al., 2008). This way, errors in the chronology become an integral part of the reconstruction process and their effect is reflected in uncertainty of the reconstructed climate (Breitenbach et al., 2012).

Bayesian modelling has many advantages, but it remains little used in the field of quantitative palaeoclimatology. In the past, the computational requirements of Bayesian methodology have often limited their practical usefulness, but recent advances in algorithm design and the availability of increasingly powerful hardware have largely reduced this issue. The need to write and run custom-made code that was only applicable for a particular model has also limited the use of Bayesian modelling. The development and availability of general-purpose software for pollen-based climate reconstruction is likely to make the Bayesian approach more attractive and more easily applicable in the near future (Parnell et al., 2015; Tipton et al., 2019).

## 6. Performing high-quality reconstructions

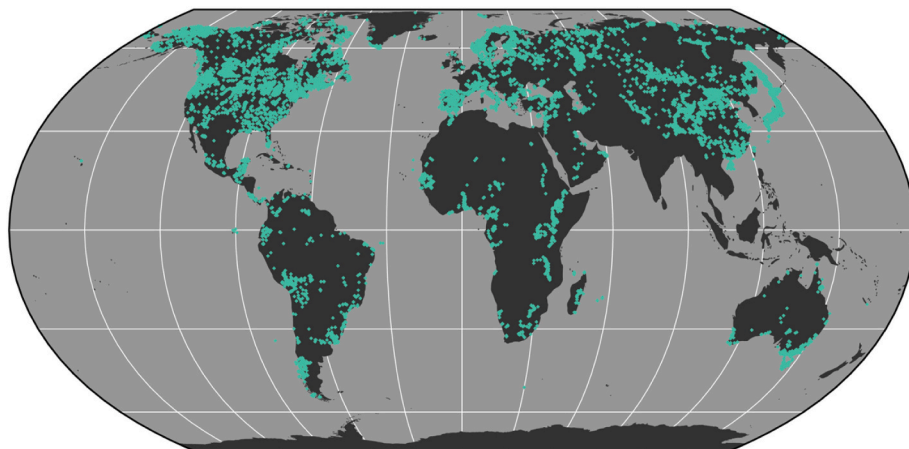
### 6.1. Selection of the appropriate technique(s)

All the reconstruction techniques have their own set of advantages and limitations (Table 1) and many criteria influence the selection of the most appropriate technique(s) to be used on fossil pollen records. In the following section, we present general guidelines that are suggestive of the strategies adopted by the community thus far, but specific applications and/or research questions may require using one specific

technique over another. These guidelines must be also combined with the appropriate calibration datasets (Section 5.2). When possible, we recommend using multiple techniques (e.g. Brewer et al., 2008; Lotter et al., 2000; Nolan et al., 2019; Salonen et al., 2019) to examine the effects of technique choice.

The first factor to consider is the location of the study site and availability of suitable calibration datasets. In regions where modern pollen samples are unavailable (Fig. 10), or may not provide adequate representations of the vegetation, it may nevertheless be possible to employ *pdf* techniques calibrated with the global GBIF data (Chevalier, 2018) or the inverse modelling approach using, for instance, one of the existing regional pollen taxa-PFT transformation schemes (Binney et al., 2017; Elenga et al., 2001; Jolly et al., 1998; Marchant et al., 2009; Peyron et al., 1998; Tarasov et al., 1998; Williams et al., 1998). The *pdf* techniques have a finer botanical resolution than inverse modelling (pollen types v. PFTs) and should allow for more precise palaeoclimatic reconstructions. However, inverse modelling can be more advantageous to explore boundary conditions that differ significantly from those of today (i.e. in no-analogue climate situations), such as periods with significantly different atmospheric CO<sub>2</sub> concentrations (e.g. during glacial periods) or different seasonal insolation distribution in the early Holocene. While the *pdf* techniques can deal with some of these effects – the *pdf* techniques estimate the climatic response of each taxon independently and not the response of the pollen composition as a whole – they cannot address the effects of atmospheric CO<sub>2</sub> concentrations on plants' responses to climate (Assumption 5). In contrast, the mechanistic response of plant PFTs to this forcing is integrated in the vegetation models used by inverse modelling, which makes this approach particularly adapted to study specific time periods with different boundary conditions even in areas where sufficiently large sets of modern pollen samples are available.

Using PFTs instead of pollen taxa has also been proposed as a way to limit no-analogue situations (e.g. due to the impact of past and modern human-influenced landscapes, or to different climate boundary conditions), irrespective of the chosen technique (Davis et al., 2003; Peyron et al., 1998). Because PFTs are defined by converging characteristics, the PFT-climate relationships are expected to be more robust than the pollen taxa-climate ones. This approach can reduce the number of no-vegetation analogue situations and produce more consistent climate reconstructions (Mauri et al., 2015). However, the resilience of the PFT approach is limited to replacements of species within the same PFT type (i.e. a tree species being replaced by another tree species with similar characteristics). If the changes occur at the ecosystem level (e.g. a forest replaced by grassland or shrubland vegetation, as has been widely observed in many late Holocene European pollen records), the PFT-based approach probably does not perform better than the pollen-taxa based reconstructions. Detailed testing of this hypothesis with different techniques is needed to better identify when and where it could be most



**Fig. 10.** Map of modern pollen samples available from numerous public databases and studies (Cao et al., 2014; Davis et al., 2020; Flantua et al., 2015; Fletcher and Thomas, 2007; Gajewski et al., 2002; Herbert and Harrison, 2016; Montade et al., 2019b; Montade et al., 2019b; Whitmore et al., 2005; Williams et al., 2018a). The map is based on the 'Equal Earth' map projection to better represent the relative sizes of the different continents (Šavrič et al., 2019).

appropriate.

Regression, classification and Bayesian techniques are powerful for analysing pollen records from data-rich regions (Fig. 10). As these techniques use modern pollen samples to link pollen and climate, they avoid some of the assumptions inherent to other techniques (e.g. linking plant distributions with pollen assemblages for the *pdf* techniques or relating pollen percentages to PFT productivity in inverse modelling). The choice between classical techniques such as MAT or WA/WA-PLS may depend on several criteria. The scale of the study is possibly the most critical, as WA/WA-PLS need a specifically designed training set to ensure unimodality of the responses of taxa to climate (see Section 4.2). This means that WA/WA-PLS is usually better suited for local- to regional-scale studies, as the low taxonomic resolution of some pollen types may break the unimodality assumption the supra-regional and continental-scale (Birks et al., 2010). MAT does not require mathematical formalising of pollen-climate relationships and can thus be used for studies that cover many vegetation and climate zones (Marsicek et al., 2018; Mauri et al., 2015; Viau et al., 2006). Similarly, machine-learning approaches can be used over large ecological gradients as they do not rely on any assumption about the pollen-climate relationships. However, considering their tendency to overfit the data (*i.e.* to find unrealistic relationships in the data), it is recommended to use these approaches primarily in areas where dense networks of modern pollen samples exist (see Fig. 10).

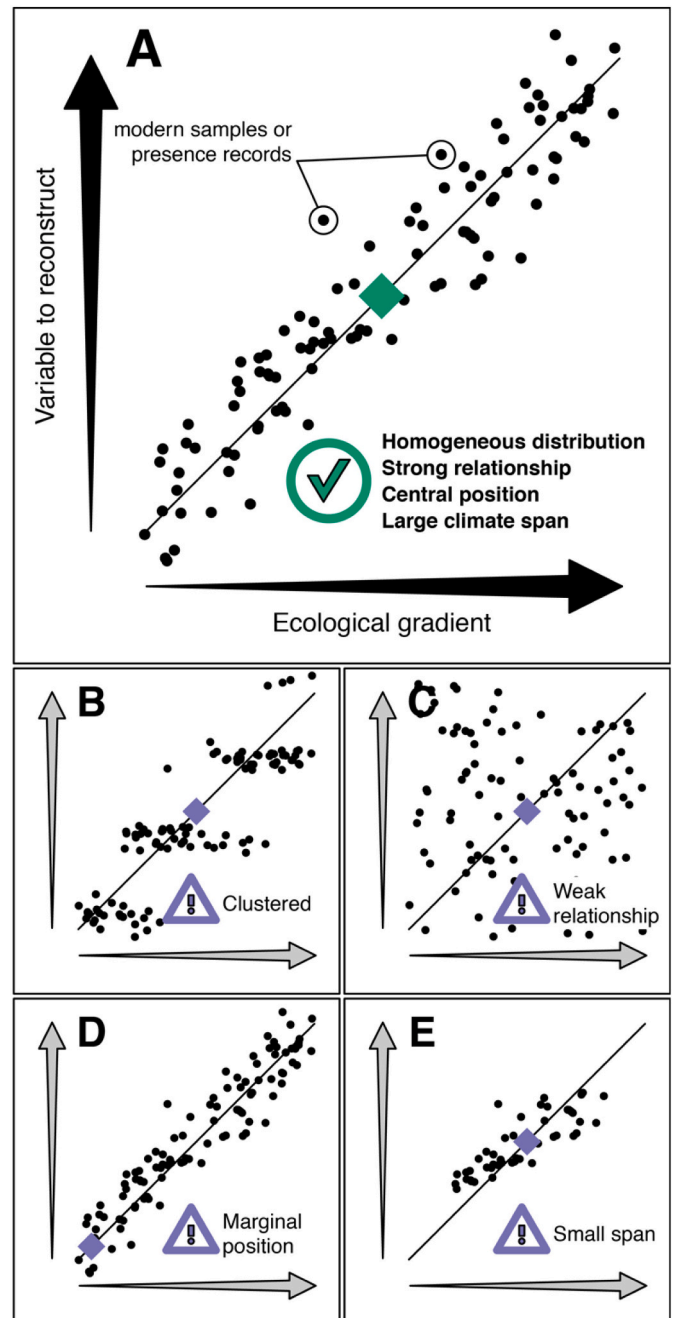
In addition, spatial autocorrelation should always be considered before choosing MAT, as it can bias the evaluation of the performance of the technique (Telford and Birks, 2005). Transfer function techniques that estimate the pollen-climate relationships from the whole modern dataset (*e.g.* regression, *pdfs* and Bayesian techniques) rather than from individual samples (*e.g.* MAT) are more robust against spatial structure in the training set (Telford and Birks, 2011b; Telford and Birks, 2005). The influence of spatial autocorrelation can be assessed using an h-block cross validation, even if the test also has limitations (see Section 5.2.3).

## 6.2. Design of a modern pollen dataset

For approaches relying on modern calibration data, designing a robust modern pollen dataset is critical part of the reconstruction process, as the reconstructed values depend strongly on the empirical relationships between the modern distribution and abundance of taxa and contemporary environmental measurements. Most techniques use modern pollen samples (*i.e.* modern pollen assemblages measured from soil, mires, moss pollsters or lake sediments, Sections 4.2, 4.3 and 4.5), while others require modern plant presence-only data (*i.e.* collections of geolocalised observations of plant species, Section 4.1). We use the generic term *modern datasets* when we refer to both. Despite these differences, designing an appropriate modern dataset follows similar rules, even if they can be difficult to quantify objectively. The overall quality of a modern dataset depends on various factors, such as 1) the nature of the samples composing the training set itself, 2) the spatiotemporal scale of the study, 3) the number and types of identified taxa, and 4) the location of the target site(s) in the sampled climate space (Fig. 11). The following sections provide general guidelines to follow in designing modern datasets and balancing among these different factors. In practice, following all these recommendations is rarely possible and the design of a modern dataset is often a practical compromise between the quality (*i.e.* the selected data follow all the guidelines) and the quantity (*i.e.* modern datasets have a finite size and pruning too many samples can lead to a loss of information) of modern data.

### 6.2.1. Depositional environments of the modern pollen samples

For all the techniques using modern pollen samples to establish the link between fossil pollen and climate, the depositional environment of



**Fig. 11.** (A) Conceptual illustration of an ideal relationship between the ecological gradient (the pollen on the x-axis) and the variable to reconstruct (y-axis) in modern datasets. The modern data are homogeneously distributed along the gradient of the climate variable to reconstruct. The variable describes a large fraction of the ecological variance (small spread around the trend line), the modern climate of the location where the target pollen record was retrieved is centred on the climate gradient, and the span of the modern dataset is large enough to reconstruct potentially large environmental changes in the past. Panels (B) to (E) illustrate suboptimal characteristics of modern datasets: (B) Clustered dataset that could create step-wise reconstructions (*i.e.* jumps from one cluster to the next), (C) weak relationship between the modern ecological gradient and the variable to reconstruct, (D) modern climate of the target site is at the extreme of the covered climate range, creating risks of biases due to past non-analogue climates (*i.e.* beyond the calibration range) and edge effects and (E) small dataset that only covers a fraction of the potential climate gradient with an effect similar to D (resulting limited range of possible reconstructed values and strong edge effect).

such modern samples has been viewed as an important factor, as it defines the spatiotemporal characteristics of the sample. In dryland regions for example, the quantity of preservational environments is limited and most modern pollen grains may be strongly impacted by post-depositional deterioration (e.g. Bryant Jr et al., 1994; Faegri et al., 1989). The samples are less likely to represent a stable multi-decadal vegetation mean and, considering the strong inter- and intra-annual climate variability in such regions, may only reflect periods of particularly recent pollen production, which may not be representative of mean climate states. In more humid regions where such collections are currently unavailable (Fig. 10), more samples could become available given sustained sampling and counting effort.

Seppä and Bennett (2003) argued that training-set samples should only be derived from the same type of sedimentary environment as the fossil material to minimise the influence of taphonomic processes on pollen-plant relationships (e.g. proportion of local vs. regional pollen taxa, differential preservation of pollen grains, see Section 3). This is supported by Goring et al. (2010) and Jantz et al. (2013), who showed that the use of a modern dataset using samples from different depositional environments generally increases model errors. As such, it is recommended to use samples from the same depositional environment when they are abundant. However, since fossil records are mostly derived from lacustrine sediments or, more infrequently, peat from wetlands (bogs and mires), collecting a large set of lake- or peat-derived modern samples is only achievable in study areas characterised by a high spatial density of lakes and wetlands, such as temperate and boreal regions (e.g. Downing et al., 2006). In areas with a low density of lakes and wetlands, moss and/or soil-samples (e.g. Gajewski et al., 2002; Ortega-Rosas et al., 2008), marine sediments (e.g. Montade et al., 2019b) or pollen-trap samples (de Nascimento et al., 2015; Jantz et al., 2013) can be used to complement the coverage of the lacustrine and peat samples, despite known differences in taphonomy and preservation (e.g. Hagemans et al., 2019). In particular, pollen trap data may represent an important source of modern samples. However, while samples from sediments, peat, mosses, and soil typically represent an average of the pollen deposited over several years, pollen-trap samples are generally collected seasonally or annually. Therefore, trap data should only be used to reconstruct climate when many years of trap data (e.g. > 10 years) are averaged to reduce the influence of inter-annual pollen deposition variability and improve comparability (Giesecke et al., 2010; Hicks, 2006).

In addition, information about the depositional environment is not always available, especially for legacy datasets (i.e. old datasets originally stored and shared on obsolete data formats and hardware infrastructures), which comprise multiple generations of collections, perhaps with differing standards and methodologies. For example, in the latest Eurasian Modern Pollen database (EMPD2), metadata regarding depositional environment for a quarter of all the modern pollen samples (~2000 samples) is unknown (Davis et al., 2020). To develop their modern dataset, users often have to make a trade-off between increasing the size of the modern dataset and including sites from different depositional environments. Optimising the number and quality of modern samples in the modern dataset is critical to increase the representation of each climate and vegetation type, and, by extension, extend the range of potential climate values that can be reconstructed from fossil data (Fig. 11E) but also to maintain a certain degree of homogeneity between the fossil and modern samples.

### 6.2.2. Span, coverage and size of the modern dataset

The accuracy and reliability of any reconstruction depends on the span (i.e. the width of the environmental gradient covered by the samples), the size (i.e. the number of data points), and the coverage of the environmental gradient (i.e. the distribution of the samples along that gradient) (Juggins and Birks, 2012). The span should range beyond the climatic and environmental values likely to have generated the fossil assemblages to avoid over/underestimation of the extreme values

(edge effect). This aspect is critical, as an unfavourable position (Fig. 11D) and/or span (Fig. 11E) of the modern dataset will bias the model and artificially restrict the potential range of the reconstructions. When a study site is located near the centre of the climatic range covered by the modern dataset and the distribution of samples along this climate gradient is homogeneous, the possibilities for reconstructing lower or higher climatic values are unrestricted (Fig. 11A). In contrast, when the site's modern climate is close to the extreme edge of the training set, then only changes in one direction will be accurately reconstructed, while changes in the other direction will be muted or wrongly attributed (Fig. 11D).

The size of the modern dataset (i.e. the number of individual samples) and the distribution of these samples along the studied gradient also have a significant impact on the pollen-climate relationships. While the minimum number of required samples depends on the context of the study (e.g. pollen diversity, time span), aiming for several hundred samples represent a good target for a single site reconstruction without extreme pollen turnover, even if this number will also depend on the methodological choices. For MAT for instance, larger modern datasets sampling large climatic and ecological gradients will improve the quality and quantity of information necessary to reconstruct climate, as this will increase the chances of finding good analogues. In contrast, methods (e.g. regression techniques) and studies focusing on regional pollen climate relationships will often be based on smaller calibration datasets to better capture the regional pollen-climate relationships. In all cases, a homogenous distribution of these samples along the climate gradient is also preferred (Telford and Birks, 2011b). When the distribution of samples is too clustered in the climate space (Fig. 11B), the pollen-climate relationships can incorporate such irregularities and the reconstructed values can 'jump' from one climate cluster to the next, especially classification techniques. The techniques that model the modern relationship between pollen and climate can be resilient to small gaps (e.g. regression and Bayesian techniques). However, all techniques will suffer from this when the variable to reconstruct varies more between clusters than within clusters (i.e. when the clusters are separated by large data gaps).

Records of plant species distributions are usually more continuous than distributions of modern pollen samples. Therefore, the techniques based on these data for the calibration step are less sensitive to the homogeneity of the sampling. However, in some areas, sharp transitions in sampling density exist (e.g. across country borders; Chevalier, 2019; Rutherford et al., 2003). Some preliminary results suggest a limited impact on the *pdfs* as long as the full range of climatic conditions where species live remains fully sampled (Chevalier et al., 2014). A sub-sampling of the densest areas can be applied to obtain a more balanced distribution of data and limit the bias of the estimated parameters towards the climate of more data-rich regions. More tests in various conditions are necessary to fully evaluate the potential impact of these effects on the reconstructions.

### 6.2.3. Testing and validation of reconstruction models and modern datasets

Before reconstructing past climate with a model, one should make sure that it has significant predictive power, which can be defined as the capacity of the model to estimate reliable climate values from unknown (i.e. fossil) pollen samples. In most applications, past climates are unknown, and the predictive power of the models is evaluated against modern climate, although in some situations comparisons with reconstructions based on other proxy or climate model data are possible. When comparing inferences with modern climate, data used to develop the climate reconstruction model and the data used to validate it must be different (i.e. independent) to avoid circularity. In the case of MAT for instance, the tested analogue would always be matched against itself. In this trivial example of circularity, the use of the same data to develop and validate the model biases performance indicators, leading to an improved accuracy and reduced errors. One common solution to circumvent this problem is to use cross-validation, which divides the

modern dataset in two parts: one subset to inform the model ('training dataset') and another subset to test it ('test dataset'). The cross-validation must be repeated multiple times to assess the ability of the selected model to characterise all the climate conditions represented in the modern dataset.

The assessment of a model's performance should always include looking at the relationship between inferred and observed values and at the distribution of the residuals (*i.e.* the difference between the reconstructed and the climate values). The relationship between inferred and observed values can be summarised by the correlation between the reconstructed and the target values. A high correlation is important but insufficient as a model could have a high correlation but be off target by a constant or increasing/decreasing value along the gradient of interest (*e.g.* the reconstructed values are always higher than the target value by 2 °C, or systematically and increasingly deviate from the target value towards one end of the examined climate gradient). Plotting the reconstructed against the target climate values enables easy assessment of how the samples are distributed relative to the  $y = x$  line. The mean position of the samples above or below the line informs about the accuracy or trueness of the model and indicates whether there is a bias in inferences, while the spread of inferred samples for similar target values informs about its precision. The mean offset between the target and the reconstructed values can be measured by the root-mean-square error of prediction (RMSEP), which summarises how far the reconstructed values are from their target on average (*i.e.* the mean of the squared differences between the reconstructed and target climate values). A model with high precision and accuracy will have an RMSEP close to 0.

Another important indicator of model performance is maximum bias, defined as the largest absolute error between the target and the prediction averaged for certain intervals of the climate gradient. A model can be good on average (low mean RMSEP) but fail in some parts of the climate space (high local bias). Calculating the maximum bias per section, for instance by splitting the studied gradient in 10 sections as in Birks (1998), can highlight climatic conditions that likely cannot be reconstructed in the past considering the modern calibration dataset

(Lotter et al., 1997; ter Braak et al., 1993). This usually happens near the edges of the sampled climate space (lowest and highest values are generally over- and under-estimated, respectively; see Fig. 11D) or if the modern climate space is heterogeneously sampled.

A common form of cross-validation is the  $k$ -fold cross-validation. In this case, the modern dataset is randomly divided into a certain number of groups  $k$  of the same size (commonly 5 to 10 groups, see Fig. 12). Each time, one group is used to test the model defined by all the other groups (*e.g.* Barrows and Juggins, 2005; Mauri et al., 2015). Another commonly used form of cross-validation is called leave-one-out (LOO) cross-validation, in which the test set is composed of only one sample at a time, all the other samples being used to inform the model. LOO is equivalent to  $k$ -fold with as many groups as there are samples in the modern dataset. With LOO, the tested model is the closest to the model that will be used to reconstruct climate from fossil samples (only one sample is removed from the modern dataset). The procedure can, however, require considerable computer resources, especially for models based on hundreds to thousands of modern pollen samples. Bootstrapping cross-validation is also commonly employed to evaluate a model (see definition in section 4.3.4). For each bootstrap dataset, some samples will be selected more than once, while the unselected samples will define the test group. The process must be repeated a large number of times to obtain robust statistics (Juggins and Birks, 2012).

Spatial autocorrelation and/or block structure in the modern dataset (Fig. 11B) can bias the performances indicators estimated from the LOO,  $k$ -fold and bootstrap cross-validations (Telford and Birks, 2005). An alternative approach to alleviate this issue is to use an  $h$ -block cross-validation, which is a more specific form of LOO cross-validation where all samples within  $h$  kilometres of the target site are excluded from the reconstruction step to ensure that spatial autocorrelation does not impact the values. However, estimating  $h$  is non-trivial (Trachsel and Telford, 2016), as a too small  $h$  will not remove all the spatial autocorrelation, while a too large  $h$  will exclude too much information to reconstruct climate. Trachsel and Telford (2016) have proposed three techniques to estimate  $h$ . Salonen et al. (2019) have explored

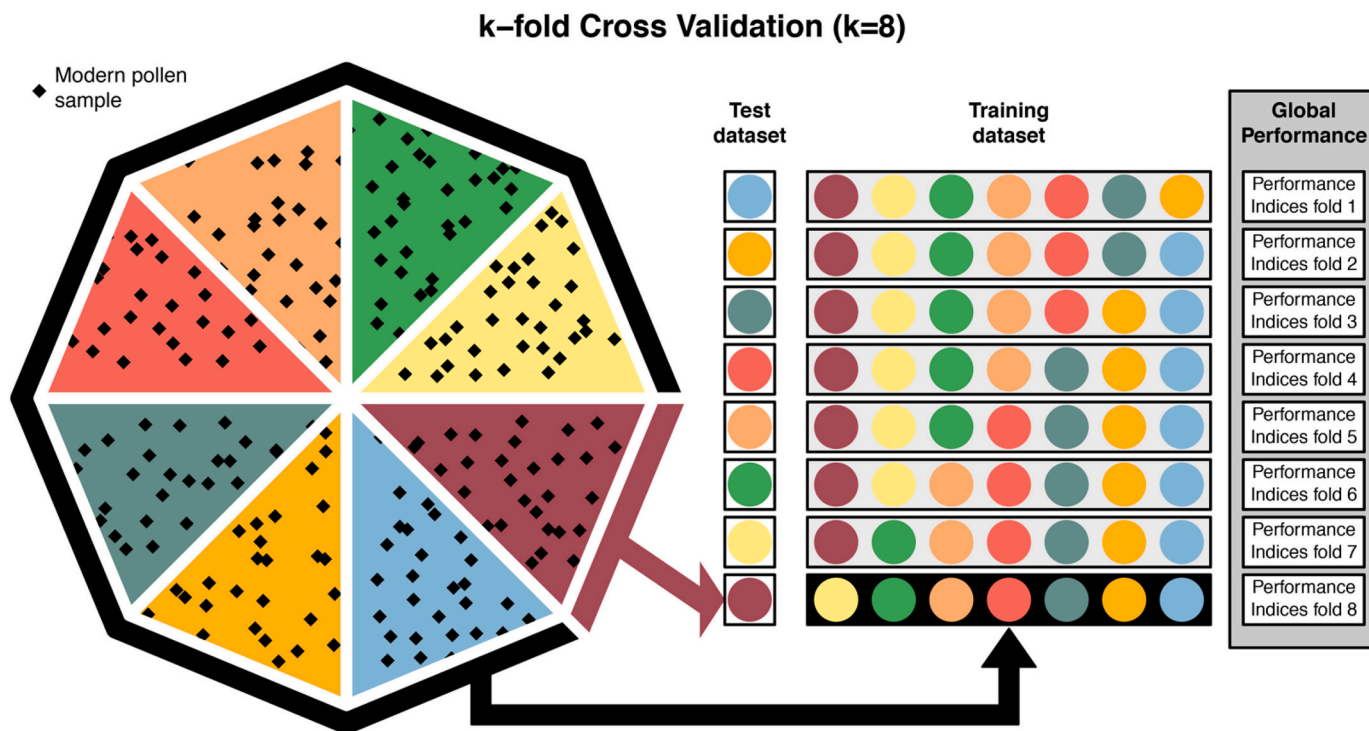


Fig. 12.  $k$ -Fold cross-validation illustrated for  $k = 8$ . The modern dataset (all the black squares) is divided into eight groups of the same size (coloured triangles). Each group is individually used as a test set while the remaining groups form the training set. Residuals are calculated for each fold and the combination of these fold-specific residuals allows estimating the performance of the entire modern dataset at reconstructing climate.

acceptable ranges of  $h$  for European and North American pollen records and argue that values of  $h$  between 300 and 800 km are acceptable thresholds. Such values need to be re-assessed for every study, but they provide a reasonable range to start screening for  $h$ . However,  $h$ -block cross-validation is only based on horizontal distances and does not include elevation changes. Climate gradients over vertical distances are much steeper than over horizontal distance, so that excluding samples purely based on horizontal distance can be a significant issue in mountainous areas, as it would exclude potential samples from lower/higher elevations that could experience very different climates.

For techniques that use modern plant distributions instead of modern surface samples to create the pollen-climate relationship, such as the *pdf* techniques, validating the model based on modern pollen samples is more straightforward as the calibration set (modern plant occurrences) is independent from the validation set (modern pollen samples). As such, cross-validating the modern dataset is not necessary. All the modern samples can be reconstructed at once and the global RMSEPs and maximum biases analysed as described previously.

### 6.3. Evaluation of palaeoclimate reconstructions

With the development of recent statistical tools and computing power, it has become easier to produce quantitative reconstructions, thus shifting the complexity of the work from the extraction of signal from multivariate pollen data to the development of sensible modern datasets and the evaluation of the reconstructions. Evaluating a reconstruction is a complex process that relies on both statistical indicators and expert knowledge. Commonly used indicators include: 1) the amplitude of change (e.g. Is the reconstructed amplitude of change coherent with expectations considering the location and studied time period?), 2) the consistency of the reconstruction across time (e.g. Is the sample-to-sample variability more important than the reconstructed signal itself?) or 3) the percentage of explained variance (e.g. Does the reconstruction explain an acceptable fraction of the total variance in the pollen data?). Passing one or more of these criteria does not guarantee the validity of the reconstruction, but it reduces the risk of incorrect assignment (rejecting valid reconstructions or accepting erroneous ones). Expert knowledge of the local ecology of the studied site should always be included and considerations of the variability of key pollen taxa with clear climate sensitivities to the studied parameters can help decide whether the reconstruction is robust or if there are confounding factors (Juggins et al., 2015).

For all the techniques that use modern pollen samples, assessing the distance between the closest modern analogues to a fossil sample can inform about the reliability of the reconstruction. Square chord distances of the closest analogue are commonly plotted against time to highlight potential no-analogue situations, which would indicate that the reconstructed value could be erroneous and/or that uncertainties may be underestimated. The higher the difference between a fossil sample and the closest modern sample(s), the lower the confidence in any reconstruction from that sample should probably be. No fixed threshold can be defined to determine if a fossil pollen sample has good analogues in the modern calibration data, as the usual metric distances used to measure the similarity of pollen assemblages varies as a function of many study-specific parameters. Sawada et al. (2004) demonstrated that the diversity of pollen assemblages influences the dissimilarity indices such as SCD. High diversity samples tend to have a higher SCD with modern samples than low diversity samples. Using the 5th percentile of the measured distance between modern dataset samples has become the standard threshold identifying fossil samples with good modern analogues (e.g. Jackson and Williams, 2004). While imperfect, this empirical approach can be employed consistently and is justifiable. Mapping the selected analogues in geographical space can help to better understand the drivers of the reconstruction, as it allows to reconnect with the ecology supporting the reconstructions. Looking at the spread of the uncertainties around the reconstructed value can also be a

good indicator of the quality of the reconstruction, especially with methods that estimate the distribution of climate histories (e.g. K uhl et al., 2007).

Another evaluation tool is the palaeoSig test developed by Telford and Birks (2011c), which assesses whether a reconstruction shows significant changes in the inferred variable compared to a null distribution of reconstructions derived from transfer functions trained on simulated climate variables. The palaeoSig test compares the percentage of variance explained by a reconstruction to an ensemble of reconstructions obtained from a randomised calibration dataset. If the real reconstruction explains more variance than most of the randomised reconstructions (e.g. 95% of them), the reconstruction is deemed statistically significant (Telford and Birks, 2011c). However, palaeoSig does not test if the reconstructed values make sense climatically; it only tests if there is a statistically significant trend in the data relative to an ensemble of random reconstructions. Climate reconstructions with low climate variability will unavoidably ‘fail’ the palaeoSig test. This does not necessarily mean the reconstruction is incorrect, it simply means that there is no statistically significant trend in the data (i.e. the reconstruction could be reliable, but climate may have been stable over the studied period). Alternatively, in the presence of a confounding factor, reconstructions can ‘pass’ the palaeoSig, while being erroneous (Juggins, 2013). As with any evaluation tool, palaeoSig should only be one element used to assess the reliability of a climate reconstruction.

Multi-site, multi-technique or multi-proxy reconstructions can also be used as an additional way of evaluating reconstructions. For multi-site comparisons (comparing a reconstruction with independent regional records of the same variable), expecting good correlation between records is reasonable, as climate is spatially autocorrelated. Multi-technique approaches (different reconstruction techniques applied to the same pollen dataset) usually produce similar trends but can be characterised by different amplitudes of change or intrinsic noise (Brewer et al., 2008; Lebamba et al., 2012; Peyron et al., 2013). While differences in amplitude can be hard to reconcile and can be caused by many factors, they have been interpreted as an estimate of the climate uncertainty associated with the reconstruction. Composite reconstructions can be obtained by combining the reconstructions obtained with the different techniques (Birks, 2014). Finally, highlighting similarities in trends and amplitude between multi-proxy reconstructions (reconstructing the same variable from one core or nearby locations using different proxies) can provide additional support (Birks and Ammann, 2000; Finsinger et al., 2010; Fortin and Gajewski, 2016; K uhl et al., 2010). Sometimes they do not converge due to different ecologies of the proxies and/or differential sensitivities to various climatic aspects (e.g. Lotter et al., 2012; Samartin et al., 2017), but in other cases, the reconstructions can closely resemble each other (Fortin and Gajewski, 2016).

While the agreement between two proxy records is unlikely to be exact, strong similarities are expected if the proxies reflect the same or similar aspects of the climate system. Considering the uncertainties of the records being compared is often key to reconcile the records, as the uncertainty ranges of the two records can overlap and be in agreement despite the reconstructed most likely climate change trajectories being different. Many factors can explain subtle differences (e.g. local meso-climate, low number of analogues, uncertain dating), but overall, climate is spatially correlated, and the reconstructions should reflect that. Any strong departure from this may reflect that either the modern dataset was inappropriate or that the reconstructed variable is not the main driver of the pollen changes in the record. It can also mean that the two records being compared do not reflect the same climatic aspects, which may require a re-assessment of the interpretation of the other record.

In conclusion, there is no single solution that exists to either validate or invalidate a pollen-based climate reconstruction. It is only through critical consideration of multiple, independent lines of evidence (climatic and ecological, multi-site, multi-proxy and multi-techniques

intercomparisons, palaeoSig, analogue distance, etc.) and by explaining them through consistent ecological and climatic processes that a reconstruction can be properly assessed.

## 7. Datasets and software

### 7.1. Modern and fossil pollen databases

Limited or difficult access to modern and fossil pollen data, metadata and modern plant distributions have often hindered the application of climate reconstruction techniques. Until recently, open-access pollen data have mainly been gathered into independent, individually structured regional databases (Table 2) that were not directly compatible with each other. Over the past decade, the Neotoma Paleocology

Database, a global database that includes pollen as well as other palaeoecological proxies, has been gradually incorporating pollen data from many of these regional databases to facilitate access and usability of the data (Williams et al., 2018a). Importantly, the taxon names of the various datasets have been standardised in Neotoma to follow new international standards and facilitate the use and combination of these datasets, even if the constituent databases retain their own identities and governance to ensure that the regional specificities are not overly simplified. All the datasets in Neotoma are also curated to become more compliant with the FAIR principles (FAIR: Findable, Accessible, Interoperable and Reusable; Wilkinson et al., 2019, Wilkinson et al., 2016). Neotoma is continually growing and presently contains over 4700 fossil and ~3600 modern pollen records all over the world (as of August 2020).

**Table 2**

List of open-access pollen datasets, modern plant distribution data, modern climatologies and numerical tools to run pollen-based climate reconstructions.

	Name	Reference(s)	Source(s)
Modern and Fossil pollen databases /datasets	Neotoma Paleocology Database	Williams et al. (2018a)	<a href="https://www.neotomadb.org">https://www.neotomadb.org</a>
	North American Pollen Database	Grimm (2000)	Included in Neotoma
	Eurasian Pollen Database	Binney et al. (2016, 2017)	<a href="https://doi.org/10.5285/6aeba247-52d1-4e84-949f-603742af40c1">https://doi.org/10.5285/6aeba247-52d1-4e84-949f-603742af40c1</a>
	European Pollen Database	Fyfe et al. (2009)	<a href="http://www.europeanpollendatabase.net/getdata">http://www.europeanpollendatabase.net/getdata</a> (Also in Neotoma)
	North American Modern Pollen database	Whitmore et al. (2005)	<a href="http://www.lpc.uottawa.ca/data/modern/index.html">http://www.lpc.uottawa.ca/data/modern/index.html</a> (Partly in Neotoma)
	Eurasian Modern Pollen database	Chevalier et al. (2019) Davis et al. (2020)	<a href="https://doi.pangaea.de/10.1594/PANGAEA.909130">https://doi.pangaea.de/10.1594/PANGAEA.909130</a>
	African Pollen Database (both modern and fossil data)	Gajewski et al. (2002) Vincens et al. (2007)	<a href="http://fpd.sedoo.fr/fpd">http://fpd.sedoo.fr/fpd</a> (Being transferred to Neotoma)
	Latin American Pollen Database (both modern and fossil data)	Flantua et al. (2015)	Constituent database in Neotoma
	Non-specific pollen data repositories	Marchant et al. (2002)	<a href="http://www.pangaea.de">www.pangaea.de</a> <a href="https://www.ncdc.noaa.gov/data-access/paleoclimatology-data">https://www.ncdc.noaa.gov/data-access/paleoclimatology-data</a>
Modern plant distribution databases	GBIF for CREST dataset	Chevalier (2018, 2019)	<a href="http://www.doi.org/10.6084/m9.figshare.6743207">www.doi.org/10.6084/m9.figshare.6743207</a> <a href="http://www.gbif.org">www.gbif.org</a>
	CRACLE calibration accessible via the R package	Harbert and Baryames (2020)	<a href="https://github.com/rsh249/cRacle">https://github.com/rsh249/cRacle</a>
	North American Biodiversity Information Serving Our Nation (BISON)		<a href="https://bison.usgs.gov/#home">https://bison.usgs.gov/#home</a>
	US Forest Inventory and Analysis EU-Forest	Mauri et al. (2017) Giovanni et al. (2016)	<a href="https://www.fia.fs.fed.us/">https://www.fia.fs.fed.us/</a> <a href="https://doi.org/10.6084/m9.figshare.c.3288407.v1">https://doi.org/10.6084/m9.figshare.c.3288407.v1</a>
Modern climatologies	Climate Research Unit (CRU)	Harris et al. (2014) New et al. (2002)	<a href="http://www.cru.uea.ac.uk/data">http://www.cru.uea.ac.uk/data</a>
	Worldclim	Fick and Hijmans (2017) Hijmans et al. (2005)	<a href="http://worldclim.org/version2">http://worldclim.org/version2</a>
	NCEP/NCAR reanalysis	Kalnay et al. (1996)	<a href="https://www.esrl.noaa.gov/psd/data/gridded/data.ncep.reanalysis.html">https://www.esrl.noaa.gov/psd/data/gridded/data.ncep.reanalysis.html</a>
Software & Packages	CHELSA climatology	Karger et al. (2016)	<a href="http://chelsa-climate.org/downloads">http://chelsa-climate.org/downloads</a>
	rioja R package	Juggins, 2015	<a href="https://cran.r-project.org/web/packages/rioja/index.html">https://cran.r-project.org/web/packages/rioja/index.html</a>
	analogue R package	Simpson and Oksanen (2013)	<a href="https://cran.r-project.org/web/packages/analogue/index.html">https://cran.r-project.org/web/packages/analogue/index.html</a>
	MATTOOLS R package	Sawada (2006)	<a href="https://cran.r-project.org/web/packages/MATTOOLS/index.html">https://cran.r-project.org/web/packages/MATTOOLS/index.html</a>
	BIOINDIC R package		<a href="https://www.eccorev.fr/spip.php?rubrique55">https://www.eccorev.fr/spip.php?rubrique55</a>
	palaeoSig R package	Telford and Trachsel (2015)	<a href="https://cran.r-project.org/web/packages/palaeoSig/index.html">https://cran.r-project.org/web/packages/palaeoSig/index.html</a>
	Bclim R package	Parnell et al. (2016b)	<a href="https://cran.r-project.org/web/packages/Bclim/index.html">https://cran.r-project.org/web/packages/Bclim/index.html</a>
	cRacle R package	Harbert and Baryames, 2020	<a href="https://github.com/rsh249/cRacle">https://github.com/rsh249/cRacle</a>
	BayesComposition R package	Nolan et al. (2019)	<a href="https://github.com/jtipton25/BayesComposition">https://github.com/jtipton25/BayesComposition</a>
	vegan R package	Oksanen et al. (2007)	<a href="https://cran.r-project.org/web/packages/vegan/index.html">https://cran.r-project.org/web/packages/vegan/index.html</a>
	Neotoma R package	(Goring et al., 2015)	<a href="https://github.com/ropensci/neotoma">https://github.com/ropensci/neotoma</a>
CREST R package		<a href="https://github.com/mchevalier2/crestr">https://github.com/mchevalier2/crestr</a>	
CREST software	Chevalier et al. (2014)	<a href="https://chevaliermanuel.wixsite.com/webpage/crest">https://chevaliermanuel.wixsite.com/webpage/crest</a>	
Past software	Hammer et al. (2001)	<a href="https://folk.uio.no/ohammer/past">https://folk.uio.no/ohammer/past</a>	
C2 software	Juggins (2007)	<a href="https://www.staff.ncl.ac.uk/stephen.juggins/software/C2Home.htm">https://www.staff.ncl.ac.uk/stephen.juggins/software/C2Home.htm</a> ;	
Polygon software	Nakagawa (2008)	<a href="http://polysystems.rits-palaeo.com">http://polysystems.rits-palaeo.com</a>	

Despite this effort, access to some regional databases continues to be problematic, as they are not available from well-identified public repositories. Many small-scale regional modern sample datasets also continue to be developed and are typically available outside of Neotoma (e.g. Julier et al., 2018; Montade et al., 2019b; Montade et al., 2019a; Urrego et al., 2015). Supporting further field campaigns, especially in quantification deserts, and making these data publicly accessible (e.g. by adding them to relevant regional Constituent Databases in Neotoma, or any other repository, such as Pangaea [www.pangaea.de](http://www.pangaea.de) or NOAA [www.ncdc.noaa.gov](http://www.ncdc.noaa.gov)) is critical to improve the global coverage of modern pollen assemblages and, by extension, the applicability of climate reconstruction methods.

To our knowledge, only the Global Biodiversity Information Facility (GBIF) provides a set of globally dense distributions of plants that can be directly used with the *pdf* techniques (Chevalier, 2019; Chevalier, 2018). Other high-resolution datasets of the distribution of most European trees (Mauri et al., 2017) and North American (Biodiversity Information Serving Our Nation (BISON) and US Forest Inventory and Analysis (FIA)) vegetation are also available (see Table 2). Additional collections can often be obtained from herbaria and national museums by direct request.

## 7.2. Modern climate data

Modern climate datasets are used by all the techniques to define the relationships between plant/pollen taxa with climate at present time, following the standard space-for-time assumption that spatial relationships among plant taxa and environmental variables can be used to draw inferences from fossil pollen time series about past environmental variations (Section 3). The most commonly used datasets are listed in Table 2. These datasets average climatic conditions over different time periods. For instance, WorldClim 2.0 reports average values of the period 1970–2000, while CHELSA represents the period 1979–2013. Generally, all modern samples are assigned climate values based on the same climate data. To provide a common background for climate reconstructions, many pollen calibration datasets are already associated with a specific climate dataset. For instance, in the latest version of the Eurasian Modern Pollen Database, the samples have been paired with 30 arcsecond WorldClim 2.0 data (Davis et al., 2020). However, the climate baseline associated with the earliest modern samples collected in 1970s and 1980s and the most recent ones has shifted, especially considering contemporary climate warming. This potential disequilibrium between vegetation and climate may challenge the first fundamental assumption in regions where climate has changed the most over the past decades (Section 3.1).

In addition, climate data are often available at different spatial resolutions ranging from 1° or 2° to much finer 30 arcsecond grids and there is no consensus regarding the choice of the optimal spatial resolution, which can impact the reconstruction (Fortin and Gajewski, 2012; Ladd et al., 2015). For instance, Ladd et al. (2015) do not recommend using high-resolution modern datasets, such as the WorldClim 30 arcsecond (~1 km) grid, for broad-scale climate reconstructions. They argue that the resolution should generally not be finer than the area represented by the data in the training set (i.e. source area for modern pollen samples as described in Section 3, or resolution of modern plant distribution data). Similarly, *pdf* studies have upscaled the climate data to match the lowest resolution of the botanical data employed (e.g. Truc et al., 2013). The selected resolution is particularly relevant for mountainous regions where climate varies strongly over short elevational distances (Furlanetto et al., 2019) but is less important in instances when climate fields are spatially smooth (e.g. temperature generally, particularly in low-relief landscapes; Ladd et al., 2015). More studies are needed to better understand how different climate data and their various spatiotemporal resolutions and extents influence climate reconstructions.

## 7.3. Software

Several software packages for carrying out modern calibrations and transfer function techniques have been developed (Table 2). Most of them have been written in the open source R statistical computing language (R Core Team, 2018) and are available online via GitHub / CRAN. These currently include rioja, analogue, vegan, MATTOOLS, BIOINDIC, Bclim and the palaeoSig package for testing the statistical significance of climate reconstructions. MAT-capable software with a graphical user interface include Polygon and Past. For Windows users, most techniques can be implemented in C2 without the need for programming. The *pdf*-techniques can be applied using the CREST python software (a R package is currently under development) and the CRACLE model can be applied using the cRacle R package. Many of the Bayesian techniques described in Section 4.4 can be applied from the Bayes-Composition R package on GitHub, although at this time, it is still under development and the documentation may be incomplete (J. Tipton, *pers. comm.*). The Bayesian model Bclim can be used using the eponym R package (Parnell et al., 2016b). All data stored in Neotoma can be accessed through a dedicated R package (Goring et al., 2015), thus facilitating the connection with the reconstruction methods.

## 8. Looking ahead: Opportunities and next steps

Reliable and accurate palaeoclimate reconstructions are necessary to address the issues posed by modern climate change. As a result of the significant increase of curated palaeodata through online, open-access databases and repositories during the last decade, three important avenues of development for pollen-based climate quantifications have emerged: 1) Developing new techniques using the wealth of available data to extract more reliable information from fossil pollen records, 2) Providing tools to facilitate access to and use of the reconstruction techniques and 3) Developing and using new datasets and techniques to produce more reconstructions in quantification deserts. Building on the global abundance of pollen records, the development of such new frameworks and their increased use will contribute to better understand past climate variability across the planet, from local to continental scales.

### 8.1. Harnessing open palaeodata to move from site-based to network-based reconstructions

The methodologies presented here all reconstruct climate from each sample/record independently from the others, that is without considering what may have happened before or after the time of deposition, or at nearby locations. The unprecedented density of easily accessible pollen records available from open-access, online databases (Table 2) can now support the transition from such sample-to-sample independent reconstructions to more holistic approaches in which samples would no longer be reconstructed in isolation. Some efforts have already been made in that direction and many studies have shown that combining multiple site-based reconstructions enabled extracting robust shared temporal trends (e.g. Chevalier and Chase, 2015; Davis et al., 2003; Gajewski, 2015; Marsicek et al., 2018; Seppä et al., 2009) and spatial patterns (e.g. Ladd et al., 2018; Mauri et al., 2015; Viau et al., 2006). In these studies, however, the multi-samples and/or multi-records connectedness was added after reconstructing the samples individually, the specificities of each site averaged out, and the inherent errors of each individual reconstructions were most often unaccounted for. In contrast, some Bayesian models integrate some of these principles (e.g. Tingley and Huybers, 2010a, 2010b; Weitzel et al., 2019; Gebhardt et al., 2008; Simonis et al., 2012), but the complexity of applying these models has limited their use. Connecting samples through time and space (i.e. sharing information between sites and records) and accounting for the sample-to-sample and record-to-record similarities will 1) produce less noisy, better constrained reconstructions, 2) extract

the most common extra-local climate signals and 3) increase the reliability of the reconstructions. Focusing on developing spatial reconstructions will thus promote a better, more in-depth understanding of past regional and global environmental variability.

Ideally, mutual influences between pollen samples and records should be probabilistic in nature to reflect the uncertainty of the pollen-climate relationships. Bayesian models provide a natural framework to incorporate some of these inter-sample and inter-site connections (e.g. Holmström et al., 2015; Weitzel et al., 2019). Their use was already highlighted as an important avenue for future work in a previous review of the field (Birks et al., 2010). While progress has been made in that direction (e.g. Holmström et al., 2015; Ilvonen et al., 2016; Parnell et al., 2016a), the use of Bayesian models remains limited in pollen-based climate reconstructions. Birks et al. (2010) argued that the lag in using such techniques was related to the limited computing power at the time. This is no longer the case, at least at the site-level basis (Holmström et al., 2015). Despite some unquestionable advantages (Ohlwein and Wahl, 2012), reconstructions using Bayesian models remain little used and the classical techniques remain dominant in the field. One possible explanation for the slow transition may be the expertise and time required to develop, apply, and interpret such models. It is also likely that the advantage of replacing existing models by Bayesian approached can appear, to date, only modest, as the classical models perform well for site-based reconstructions. However, models that estimate robust climate uncertainties are necessary to compare pollen and non-pollen climate reconstructions. One important objective of the coming years should, therefore, be to develop more general Bayesian models that can be employed by non-experts, and importantly to assess their relative strengths and weaknesses when compared to the classical models.

8.2. Improving reporting standards to increase global use and methodological transparency

The many decisions that must be taken before and after applying any of reconstruction technique (e.g. identifying taxa to be used, selecting the number of analogues, selecting the number of components of the model, validating a reconstruction and accessing and formatting appropriate modern datasets; Section 4), require expertise to produce reliable reconstructions (Juggins, 2013). Accessing and formatting appropriate modern datasets (Section 5) is another hurdle that complicates access to and likely limits the use of these tools. Increasing the level of details about the different steps of the reconstruction process reported in the publications, from the design of the modern dataset to the assessment of the reliability of a reconstruction, could help to overcome these constraints. The required information is usually distilled in the text but grouping it into one single and standardised table and/or graphical output would help to quickly assess how the model was run and help potential users to get a better understanding of the process. Table 3 summarises some parameters and data treatment decisions that are essential to report.

As most of the techniques are already available as R packages and/or software solutions (Section 6), implementing standardised log outputs to summarise the procedure, such as a publication-ready table or figure, directly in these numerical tools represent a key step to standardise reporting. Similar approaches have been adopted by other statistical models commonly used, such as the age-depth modelling technique Bacon (Blaauw et al., 2020; Blaauw and Christen, 2011) that automatically generates graphical plots summarizing the selected parameters. Similar solutions adapted to pollen-based climate reconstructions could be developed. Advantages of such an approach include that results can be easily evaluated and replicated and it allows for an accessible way to revise reconstructions if, for instance, some elements of the fossil or modern datasets get updated (e.g. the samples get assigned a better-resolved climatology, or errors in the pollen taxa names or calibration data are corrected) after the original publication.

Table 3 Suggested parameters and data treatments on pollen-based climate reconstructions to report in publications.

Technique	Modern dataset		Specific treatments	Name and version		The model	Fossil pollen dataset	Uncertainties
	Name and version	Name and version of the modern dataset used. If the data are not publicly available, upload them on a public repository to DOI them.		Report name and version of all the software, package or vegetation model used	Dimension and shape of the pdfs, Selection of climate sensitive taxa, Minimum number of points to define a pdf, Weighting of the pdfs			
Pdf techniques	Report name and version of the modern dataset used. If the data are not publicly available, upload them on a public repository to DOI them.	Definition of the study area (e.g. by countries and/or coordinates), Linkage of plant species to pollen taxa	Selection of samples (density across the gradient, pollen counts, pollen diversity, sampling context, etc.)	Report name and version of all the software, package or vegetation model used	Dimension and shape of the pdfs, Selection of climate sensitive taxa, Minimum number of points to define a pdf, Weighting of the pdfs	Exclusion of aquatics, Exclusion of rare taxa, Low percentages turned to 0, Taxonomic smoothing	Cross-validation statistical indices, Reconstruction errors, analogue-quality, residual length	
Regression techniques					Deshrinking strategy, Number of components, Numbers of hidden neurons, Number of trees (if appropriate)			
MAT					Dissimilarity metric, Number of analogues, dissimilarity threshold, Weighting of the analogues			
ANN, BRT, RF					Structure of the neural networks or the trees, number of trees			
Bayesian Inverse modelling	Name and version of the PFT scheme				Priors			
	Name and version of the vegetation model				Number of input variables (4 or 6), Number of iterations, Dissimilarity metric			



Ultimately, the ideal approach would be to comply with the FAIR principles by releasing any new data used, listing the provenance of existing publicly available data, and sharing the code used to format the modern and fossil pollen data and reconstruct the climate variable(s). Sharing scripts is the most transparent solution and the simplest way to describe all the data manipulations that were necessary to obtain climate reconstructions. As such, it is recommended to prefer programmatic interfaces over point-and-click programs to support reproducibility of the user's workflow. Many online platforms now exist for basic script sharing such as GitHub (<https://github.com>) or Bitbucket (<https://bitbucket.org>), but also for more user-friendly interface, such as RMarkdown annotated scripts (<https://rmarkdown.rstudio.com>). These open-access scripts could in turn be used as example protocols for beginners to practice and better understand the techniques. Adapting these scripts to new datasets would then be easier than writing scripts from the beginning and would, eventually, facilitate the use of pollen-based climate reconstruction techniques and contribute to broader-scale application in the many regions where more climate quantifications are needed (see Fig. 1).

### 8.3. Resolving quantification deserts

Fossil pollen records have been retrieved from all continents. In contrast, pollen-based climate reconstructions are highly concentrated in the Northern Hemisphere extratropics (Fig. 1). While this discrepancy can be partly explained by the absence of modern pollen datasets in many regions (Fig. 10) and the belief for a long-time that vegetation in the tropics was stable (Flenley, 1979), the limited access to the necessary resources and technical knowledge required to use climate reconstruction techniques has further contributed to the persistence of large quantification deserts. However, the recent trends towards open-access and online databases coupled with the continuous development of new reconstruction techniques based on alternative principles (e.g. Inverse Modelling and *pdf* techniques) provide an unprecedented opportunity to produce these urgently needed reconstructions. Extracting the wealth of palaeoclimatic information contained in pollen records from quantification deserts represents an important challenge to be addressed in the coming years. Reaching this objective will require engaging the entire community, from palynologists to data and computer scientists, as various elements will have to be assembled together, including:

1. Collecting more modern pollen samples and high-resolution fossil pollen records from quantification deserts to fill in data gaps,
2. Making these new data accessible to the community *via* existing open data online resources in a timeframe that is beneficial to all,
3. Developing more user-friendly, broad audience software to facilitate access to climate reconstructions techniques,
4. Providing training in the use of the techniques and software to the regional expert, for example at workshops, and, most importantly, by improving reporting standards and script sharing as discussed in Section 7.2.

Shrinking the spatiotemporal extent of quantification deserts will in turn enable a better characterisation of past climate dynamics across the tropics and the Southern Hemisphere continents and promote a more holistic understanding of the natural variability of some of the most important features of Earth climate. Producing extensive data coverage across these regions will also contribute to bridging the data gap that prevails between Antarctica and the Northern Hemisphere extratropics and help addressing problems of global significance. Both regions have been extensively studied but remain partly disconnected by the quantification deserts, which hinders our understanding of global climate teleconnections and variability (e.g. Kaufman et al., 2020). Ultimately, extended datasets will also be important to evaluate the most recent climate model simulations using truly 'out-of-the-box'

datasets (*i.e.* data that were not used in previous data-model comparisons) and to reinforce our confidence in climate forecasts.

## 9. Conclusion

This paper reviews over 75 years of the development and application of quantitative pollen-climate reconstruction methods since the pioneer reconstruction of Iversen (1944). The advance of increasingly sophisticated statistical and computational modelling approaches to pollen-climate analysis paralleled by rapid improvements in the size, quality and availability of modern datasets has provided a wealth of information about the past climates of many regions (e.g. Bartlein et al., 2011; COHMAP Members, 1988; Marsicek et al., 2018). These developments in methods and datasets have also been driven by demand for more spatially extensive climate reconstructions from the climate modelling community that requires palaeodata in climate model evaluation (e.g. Harrison et al., 2015). Despite the successful development of alternative climate proxies in recent years – including those based on isotopes and biomarkers – fossil pollen remains the most important source of quantitative climate data for terrestrial un-glaciated regions (e.g. Kaufman et al., 2020). Very few climate proxies can rival pollen for its spatial and temporal coverage for the late Quaternary period, and for its ability to reconstruct multiple paleoclimatic variables. In this paper, we aim to encourage not only greater use of pollen data for climate reconstructions, but also the better documentation of the underlying methodological choices, tools and datasets. To help in this process, we provide guidance on the appropriate application of the different techniques currently available, and the availability of the necessary tools and datasets required to apply these methodologies. Better transparency and documentation will help both palynologists interested in reconstructing climate and also experts from other fields interested in how pollen reconstructions compare to other paleoclimatic proxies and earth system models.

In the future, we hope to see greater use of the pollen-climate proxy outside of the traditional study regions of North America and Europe. In quantification deserts (Africa, Latin America, Southeast and Central Asia and Australasia, see Fig. 1), producing reliable reconstructions will require overcoming multiple challenges, not only scientific, such as the need for larger modern datasets, but also to help and encourage the scientific community to implement climate reconstruction techniques. As this and other data become more available, we can also expect to see key methodological developments in multi-sample and multi-site approaches that moves away from the traditional approach of treating each pollen sample as a single independent data point. This transition to multi-sample and multi-site analyses will help to better extract the background regional climate signal in pollen records that can sometimes be dominated by local noise from non-climatic influences, or, said differently, to use the traditional strength of pollen data in its sheer quantity to overcome some of its traditional limitations. The strength of using the abundance of fossil pollen records will facilitate the generation of spatially explicit reconstructions of late Quaternary climate change at both regional and global scale, which remains one of the greatest challenges in modelling the impact of our future climate.

### Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

### Acknowledgements

All authors have contributed to the design and writing of the manuscript. MC designed all the figures. The figures can be accessed in a vectorial format from <https://github.com/mchevalier2/Papers/tree/>

master/Chevalier\_etal\_Review\_Climate\_Quantif\_Pollen. The PCMP (Pollen-Climate Methods Intercomparison Project) workshop was funded by an International Exploratory Workshop Grant from the Swiss National Science Foundation (SNF/FNS) (#IZ32Z0\_173407). Participation of early career scientists at the workshop was assisted by travel grants from the International Union for Quaternary Research (INQUA) (Project no. 1705P) and the Past Global Changes organisation (PAGES). MC and PSS were supported by the SNF/FNS funded 'HORNET' Project (#200021\_169598), KG and TL by Natural Sciences and Engineering Research Council of Canada (NSERC) Discovery Grants, KI by CLAVICHORD Project (H2020-MSCA-IF-EF, 705895), LNP by the FNS grant P2LAP2\_187745, and MZ by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) Project number 2901391021 – SFB 1266. SYM was supported by the European Commission (Marie Curie) Fellowship 792197. The development of CREST was supported by the European Research Council (ERC) under the European Union's Seventh Framework Programme (FP7/2007-2013)/ERC Starting Grant 'HYRAX', grant agreement no. 258657.

## References

- Akeret, Ö., Haas, J.N., Leuzinger, U., Jacomet, S., 1999. Plant macrofossils and pollen in goat/sheep faeces from the Neolithic lake-shore settlement Arbon Bleiche 3, Switzerland. *Holocene* 9, 175–182. <https://doi.org/10.1191/095968399666631581>.
- Anderson, P.M., Bartlein, P.J., Brubaker, L.B., Gajewski, K., Ritchie, J.C., 1989. Modern analogues of Late-Quaternary pollen spectra from the western interior of North America. *J. Biogeogr.* 16, 573–596. <https://doi.org/10.2307/2845212>.
- Anderson, P.M., Bartlein, P.J., Brubaker, L.B., Gajewski, K., Ritchie, J.C., 1991. Vegetation-pollen-climate relationships for the Arcto-Boreal Region of North America and Greenland. *J. Biogeogr.* 18, 565–582.
- Andersson, G., 1903. Klimatet i Sverige efter istiden. *Nord. Tidskr.* H 1, 1–26.
- Andersson, G., 1909. The climate of Sweden in the Late-Quaternary period, facts and theories. *Sver. Geol. Undersöknings Årsbok Ser. C* 3, 1–87.
- Andersson, G., 1910. Die Veränderungen des Klimas seit dem Maximum der letzten Eiszeit. In: 11th International Geological Congress (Stockholm, 1910).
- Atkinson, T.C., 1986. Climatic calibration of coleopteran data. In: *Handbook of Holocene Paleocology and Paleohydrology*. Wiley.
- Atkinson, T.C., Briffa, K.R., Coope, G.R., 1987. Seasonal temperatures in Britain during the past 22,000 years, reconstructed using beetle remains. *Nature* 325, 587–592. <https://doi.org/10.1038/325587a0>.
- Austin, M.P., Gaywood, M.J., 1994. Current problems of environmental gradients and species response curves in relation to continuum theory. *J. Veg. Sci.* 5, 473–482.
- Barrows, T.T., Juggins, S., 2005. Sea-surface temperatures around the Australian margin and Indian Ocean during the Last Glacial Maximum. *Quat. Sci. Rev.* 24, 1017–1047. <https://doi.org/10.1016/j.quascirev.2004.07.020>.
- Bartlein, P.J., Webb, T.I., 1985. Mean July temperature at 6000 yr BP in eastern North America: regression equations for estimates from fossil-pollen data. In: Harrington, C.R. (Ed.), *Climatic Change in Canada 5: Critical Periods in the Quaternary Climatic History of Northern North America*, pp. 301–342.
- Bartlein, P.J., Whitlock, C., 1993. Elk Lake pollen record. *Geol. Soc. Am.* 275–293.
- Bartlein, P.J., Webb, T.I., Fleri, E., 1984. Holocene climatic change in the northern Midwest: Pollen-derived estimates. *Quat. Res.* 22, 361–374. [https://doi.org/10.1016/0033-5894\(84\)90029-2](https://doi.org/10.1016/0033-5894(84)90029-2).
- Bartlein, P.J., Prentice, I.C., Webb, T.I., 1986. Climatic Response Surfaces from Pollen Data for some Eastern North American Taxa. *J. Biogeogr.* 13, 35–57. <https://doi.org/10.2307/2844848>.
- Bartlein, P.J., Harrison, S.P., Brewer, S.C., Connor, S., Davis, B.A.S., Gajewski, K., Guiot, J., Harrison-Prentice, T.I., Henderson, A., Peyron, O., Prentice, I.C., Scholze, M., Seppä, H., Shuman, B.N., Sugita, S., Thompson, R.S., Viau, A.E., Williams, J.W., Wu, H., 2011. Pollen-based continental climate reconstructions at 6 and 21 ka: a global synthesis. *Clim. Dyn.* 37, 775–802. <https://doi.org/10.1007/s00382-010-0904-1>.
- Berliner, L.M., 1996. Hierarchical Bayesian time series models. In: *Maximum Entropy and Bayesian Methods*, pp. 1–8. [https://doi.org/10.1007/978-94-011-5430-7\\_3](https://doi.org/10.1007/978-94-011-5430-7_3).
- Binney, H.A., Lozhkin, A.V., Anderson, P., Andreev, A.A., Bezrukova, E., Blyakharchuk, T.A., Jankovska, V., Khazina, I., Krivonogov, S., Kremenetski, K.V., Novenko, E.Y., Ryabogina, N., S. V.Z., 2016. Eurasian Pollen Data From 21 Kiloannum to the Present. <https://doi.org/10.5285/6aeba247-52d1-4e84-949f-603742af40c1>.
- Binney, H.A., Edwards, M.E., Macias-Fauria, M., Lozhkin, A., Anderson, P., Kaplan, J.O., Andreev, A.A., Bezrukova, E., Blyakharchuk, T.A., Jankovska, V., Khazina, I., Krivonogov, S., Kremenetski, K.V., Niels, J., Novenko, E.Y., Ryabogina, N., Solovieva, N., Willis, K.J., Zernitskaya, V.P., Jankovská, V., 2017. Vegetation of Eurasia from the last glacial maximum to present: key biogeographic patterns. *Quat. Sci. Rev.* 157, 80–97. <https://doi.org/10.1016/j.quascirev.2016.11.022>.
- Birks, H.J.B., 1995. Quantitative palaeoenvironmental reconstructions. *Stat. Model. Quat. Sci. data. Tech. Guid.* 5, 161–254.
- Birks, H.J.B., 1998. Numerical tools in palaeolimnology-progress, potentialities, and problems. *J. Paleolimnol.* 20, 307–322.
- Birks, H.J.B., 2014. Quantitative palaeoenvironmental reconstructions from holocene biological data. *Glob. Chang. Holocene* 9780203785, 107–123. <https://doi.org/10.4324/9780203785027>.
- Birks, H.J.B., 2019. Contributions of Quaternary botany to modern ecology and biogeography. *Plant Ecol. Divers.* 12, 189–385. <https://doi.org/10.1080/17550874.2019.1646831>.
- Birks, H.H., Ammann, B., 2000. Two terrestrial records of rapid climatic change during the glacial-Holocene transition (14,000,000 calendar years B.P.) from Europe. *Proc. Natl. Acad. Sci. U. S. A.* 97, 1390–1394. <https://doi.org/10.1073/pnas.97.4.1390>.
- Birks, H.J.B., Seppä, H., 2003. Pollen-based reconstructions of Late-Quaternary climate in Europe – progress, problems, and pitfalls. *Acta Palaeobot.* 44, 317–334.
- Birks, H.J.B., Seppä, H., 2010. Late-Quaternary palaeoclimatic research in Fennoscandia – a historical review. *Boreas* 39, 655–673. <https://doi.org/10.1111/j.1502-3885.2010.00160.x>.
- Birks, H.J.B., Simpson, G.L., 2013. "Diatoms and pH reconstruction" (1990) revisited. *J. Paleolimnol.* 49, 363–371. <https://doi.org/10.1007/s10933-013-9697-7>.
- Birks, H.J.B., Juggins, S., Line, J.M., 1990a. Lake surface-water chemistry reconstructions from palaeolimnological data. In: *Surf. Waters Acidif. Program*. Cambridge Univ. Press, Cambridge, pp. 301–313.
- Birks, H.J.B., Line, J.M., Juggins, S., Stevenson, A.C., ter Braak, C.J.F., 1990b. Diatoms and pH reconstruction. *Philos. Trans. R. Soc. Lond. B* 327, 263–278. <https://doi.org/10.1098/rstb.1990.0062>.
- Birks, H.J.B., Heiri, O., Seppä, H., Bjune, A.E., 2010. Strengths and weaknesses of quantitative climate reconstructions based on Late-Quaternary biological proxies. *Open Ecol. J.* 3, 68–110.
- Blaauw, M., Christen, J.A., 2011. Flexible Paleoclimate Age-Depth Models using an Autoregressive Gamma Process. *Bayesian Anal.* 6, 457–474. <https://doi.org/10.1214/11-BA618>.
- Blaauw, M., Christen, J.A., Bennett, K.D., Reimer, P.J., 2018. Double the dates and go for Bayes – impacts of model choice, dating density and quality on chronologies. *Quat. Sci. Rev.* 188, 58–66. <https://doi.org/10.1016/j.quascirev.2018.03.032>.
- Blaauw, M., Christen, J.A., Aquino, L.M.A., 2020. rbacon: Age-Depth Modelling using Bayesian Statistics.
- Blois, J.L., Williams, J.W., Grimm, E.C., Jackson, S.T., Graham, R.W., 2011. A methodological framework for assessing and reducing temporal uncertainty in paleovegetation mapping from Late-Quaternary pollen records. *Quat. Sci. Rev.* 30, 1926–1939. <https://doi.org/10.1016/j.quascirev.2011.04.017>.
- Blois, J.L., Williams, J.W., Fitzpatrick, M.C., Jackson, S.T., Ferrier, S., 2013. Space can substitute for time in predicting climate-change effects on biodiversity. *Proc. Natl. Acad. Sci.* 110, 9374–9379. <https://doi.org/10.1073/pnas.1220228110>.
- Bourgeois, J.C., Koerner, R.M., Gajewski, K., Fisher, D.A., 2000. A Holocene ice-core pollen record from Ellesmere Island, Nunavut, Canada. *Quat. Res.* 54, 275–283. <https://doi.org/10.1006/qres.2000.2156>.
- Bray, P.J., Blockley, S.P.E., Coope, G.R., Dadswell, L.F., Elias, S.A., Lowe, J.J., Pollard, A.M., 2006. Refining mutual climatic range (MCR) quantitative estimates of palaeotemperature using ubiquity analysis. *Quat. Sci. Rev.* 25, 1865–1876. <https://doi.org/10.1016/j.quascirev.2006.01.023>.
- Breiman, L., 2001. Random forests ensemble classifier trained with data resampling strategy to improve cardiac arrhythmia diagnosis. *Mach. Learn.* 45, 5–32. <https://doi.org/10.1016/j.combiomed.2011.03.001>.
- Breitenbach, S.F.M., Rehfeld, K., Goswami, B., Baldini, J.U.L., Ridley, H.E., Kennett, D.J., Prufer, K.M., Aquino, V.V., Asmerom, Y., Polyak, V.J., Cheng, H., Kurths, J., Marwan, N., 2012. Constructing proxy records from age models (COPRA). *Clim. Past* 8, 1765–1779. <https://doi.org/10.5194/cp-8-1765-2012>.
- Brewer, S.C., Guiot, J., Barboni, D., 2007a. Pollen methods and studies| use of pollen as climate proxies. In: *Encyclopedia of Quaternary Science*. Elsevier, Oxford, pp. 2497–2508. <https://doi.org/10.1016/B978-0-08-088523-0.00338-5>.
- Brewer, S.C., Guiot, J., Torre, F., 2007b. Mid-Holocene climate change in Europe: a data-model comparison. *Clim. Past* 3, 499–512.
- Brewer, S.C., Guiot, J., Sánchez Goñi, M.F., Klotz, S., 2008. The climate in Europe during the Eemian: a multi-method approach using pollen data. *Quat. Sci. Rev.* 27, 2303–2315. <https://doi.org/10.1016/j.quascirev.2008.08.029>.
- Brook, G.Z., Scott, L., Railsback, L.B., Goddard, E.A., 2010. A 35ka pollen and isotope record of environmental change along the southern margin of the Kalahari from a stalagmite and animal dung deposits in Wonderwerk Cave, South Africa. *J. Arid Environ.* 74, 870–884. <https://doi.org/10.1016/j.jaridenv.2009.11.006>.
- Brugger, S.O., Gobet, E., Sigl, M., Osmond, D., Papina, T., Rudaya, N., Schwikowski, M., Tinner, W., 2018. Ice records provide new insights into climatic vulnerability of Central Asian forest and steppe communities. *Glob. Planet. Change* 169, 188–201. <https://doi.org/10.1016/j.gloplacha.2018.07.010>.
- Bryant Jr., V.M., Holloway, R.G., Jones, J.G., Carlson, D.L., 1994. Pollen preservation in alkaline soils of the American Southwest. *Sediment. Org. Part* 47–58.
- Bunting, M.J., Morgan, C.R., Van Bakel, M., Warner, B.G., 1998. Pre-European settlement conditions and human disturbance of a coniferous swamp in southern Ontario. *Can. J. Bot.* 76, 1770–1779. <https://doi.org/10.1139/b98-161>.
- Burn, M.J., Mayle, F.E., Killeen, T.J., 2010. Pollen-based differentiation of Amazonian rainforest communities and implications for lowland palaeoecology in tropical South America. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* 295, 1–18. <https://doi.org/10.1016/j.palaeo.2010.05.009>.
- Bush, M.B., Wang, C., 2007. Introducing a new (freeware) tool for palynology. *J. Biogeogr.* 34, 377–380. <https://doi.org/10.1111/j.1365-2699.2006.01645.x>.
- Bush, M.B., Silman, M.R., Urrego, D.H., 2004. 48,000 years of climate and forest change in a biodiversity hot spot. *Science* (80-) 303, 827–829. <https://doi.org/10.1126/science.1090795>.
- Cadd, H., Fletcher, M.-S., Mariani, M., Heijnis, H., Gadd, P.S., 2019. The influence of fine-scale topography on the impacts of Holocene fire in a Tasmanian montane landscape. *J. Quat. Sci.* 34, 491–498. <https://doi.org/10.1002/jqs.3114>.
- Campbell, I.D., McDonald, K., Flannigan, M.D., Kringayark, J., 1999. Long-distance















- von Post, E.J.L., 1918. Skogsträdpollen i sydsvenska torvmosslagerföljder. *Forh. ved* 16, 433–465.
- Waelbroeck, C., Labeyrie, L., Duplessy, J.-C., Guiot, J., Labracherie, M., Leclaire, H., Duprat, J., 1998. Improving past sea surface temperature estimates based on planktonic fossil faunas. *Paleoceanography* 13, 272–283. <https://doi.org/10.1029/98PA00071>.
- Wahl, E.R., Diaz, H.F., Ohlwein, C., 2012. A pollen-based reconstruction of summer temperature in central North America and implications for circulation patterns during medieval times. *Glob. Planet. Change* 84–85, 66–74. <https://doi.org/10.1016/j.gloplacha.2011.10.005>.
- Wang, Y., Heintzman, P.D., Newsom, L., Bigelow, N.H., Wooller, M.J., Shapiro, B., Williams, J.W., 2017. The southern coastal Beringian land bridge: cryptic refugium or pseudoregion for woody plants during the Last Glacial Maximum? *J. Biogeogr.* 44, 1559–1571. <https://doi.org/10.1111/jbi.13010>.
- Ward, J.K., Harris, J.M., Cerling, T.E., Wiedenhoef, A., Lott, M.J., Dearing, M.D., Coltrain, J.B., Ehleringer, J.R., 2005. Carbon starvation in glacial trees recovered from the La Brea tar pits, southern California. *Proc. Natl. Acad. Sci. U. S. A.* 102, 690–694. <https://doi.org/10.1073/pnas.0408315102>.
- Webb, T.I., 1986. Is vegetation in equilibrium with climate? How to interpret Late-Quaternary pollen data. *Vegetation* 67, 75–91. <https://doi.org/10.1007/BF00037359>.
- Webb, T.I., Bryson, R.A., 1972. Late- and Postglacial Climatic change in the Northern Midwest, USA: quantitative estimates derived from fossil pollen spectra by multivariate statistical analysis. *Quat. Res.* 2, 70–115. [https://doi.org/10.1016/0033-5894\(72\)90005-1](https://doi.org/10.1016/0033-5894(72)90005-1).
- Webb, T.I., Clark, D.R., 1977. Calibrating micropaleontological data in climatic terms: a critical review. *Ann. N. Y. Acad. Sci.* 288, 93–118. <https://doi.org/10.1111/j.1749-6632.1977.tb33605.x>.
- Webb, R.S., Anderson, K.H., Webb, T.I., 1993. Pollen response-surface estimates of Late-Quaternary changes in the moisture balance of the northeastern United States. *Quat. Res.* 40, 213–227. <https://doi.org/10.1006/qres.1993.1073>.
- Weitzel, N., Hense, A., Ohlwein, C., 2019. Combining a pollen and macrofossil synthesis with climate simulations for spatial reconstructions of European climate using Bayesian filtering. *Clim. Past* 15, 1275–1301. <https://doi.org/10.5194/cp-15-1275-2019>.
- Whitmore, J., Gajewski, K., Sawada, M., Williams, J.W., Shuman, B.N., Bartlein, P.J., Minckley, T., Viau, A.E., Webb, T.I., Shafer, S., Anderson, P., Brubaker, L., 2005. Modern pollen data from North America and Greenland for multi-scale paleoenvironmental applications. *Quat. Sci. Rev.* 24, 1828–1848. <https://doi.org/10.1016/j.quascirev.2005.03.005>.
- Whitney, B.S., Mayle, F.E., Punyasena, S.W., Fitzpatrick, K.A., Burn, M.J., Guillen, R., Chavez, E., Mann, D., Pennington, R.T., Metcalfe, S.E., 2011. A 45kyr palaeoclimate record from the lowland interior of tropical South America. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* 307, 177–192. <https://doi.org/10.1016/j.palaeo.2011.05.012>.
- Whitney, B.S., Rushton, E.A.C., Carson, J.F., Iriarte, J., Mayle, F.E., 2012. An improved methodology for the recovery of Zea mays and other large crop pollen, with implications for environmental archaeology in the Neotropics. *The Holocene* 22, 1087–1096. <https://doi.org/10.1177/0959683612441842>.
- Whittaker, R.H., Levin, S.A., Root, R.B., 1973. Niche, habitat, and ecotope. *Am. Nat.* 107, 321–338. <https://doi.org/10.1086/521238>.
- Wilkinson, M.D., Dumontier, M., Aalbersberg, I.J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L.B., Bourne, P.E., Bouwman, J., Brookes, A.J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C.T., Finkers, R., Gonzalez-Beltran, A., Gray, A.J.G., Groth, P., Goble, C., Grethe, J.S., Heringa, J., 't Hoen, P.A., Hoof, R., Kuhn, T., Kok, R., Kok, J., Lusher, S.J., Martone, M.E., Mons, A., Packer, A.L., Persson, B., Rocca-Serra, P., Roos, M., van Schaik, R., Sansone, S.-A., Schultes, E., Sengstag, T., Slater, T., Strawn, G., Swertz, M.A., Thompson, M., van der Lei, J., van Mulligen, E., Velterop, J., Waagmeester, A., Wittenburg, P., Wolstencroft, K., Zhao, J., Mons, B., 2016. The FAIR Guiding Principles for scientific data management and stewardship. *Sci. Data* 3. <https://doi.org/10.1038/sdata.2016.18>.
- Wilkinson, M.D., Dumontier, M., Sansone, S., Bonino da Silva Santos, L.O., Prieto, M., Batista, D., McQuilton, P., Kuhn, T., Rocca-Serra, P., Crosas, M., Schultes, E., 2019. Evaluating FAIR maturity through a scalable, automated, community-governed framework. *Sci. Data* 6, 174. <https://doi.org/10.1038/s41597-019-0184-5>.
- Williams, J.W., Shuman, B.N., 2008. Obtaining accurate and precise environmental reconstructions from the modern analog technique and North American surface pollen dataset. *Quat. Sci. Rev.* 27, 669–687. <https://doi.org/10.1016/j.quascirev.2008.01.004>.
- Williams, J.W., Summers, R.L.R., Webb, T.I., 1998. Applying plant functional types to construct biome maps from eastern North American pollen data: comparisons with model results. *Quat. Sci. Rev.* 17, 607–627.
- Williams, J.W., Webb, T.I., Shurman, B.N., Bartlein, P.J., 2000. Do low CO<sub>2</sub> concentrations affect pollen-based reconstructions of LGM climates? A response to “physiological significance of low atmospheric CO<sub>2</sub> for plant–climate interactions” by Cowling and Sykes. *Quat. Res.* 53, 402–404. <https://doi.org/10.1006/qres.2000.2131>.
- Williams, J.W., Grimm, E.C., Blois, J.L., Charles, D.F., Davis, E.B., Goring, S.J., Graham, R.W., Smith, A.J., Anderson, M., Arroyo-Cabrales, J., Ashworth, A.C., Betancourt, J.L., Bills, B.W., Booth, R.K., Buckland, P.I., Curry, B.B., Giesecke, T., Jackson, S.T., Latorre, C., Nichols, J., Purdum, T., Roth, R.E., Stryker, M., Takahara, H., 2018a. The Neotoma Paleoeology Database, a multiproxy, international, community-curated data resource. *Quat. Res.* 89, 156–177. <https://doi.org/10.1017/qua.2017.105>.
- Williams, J.W., Newton, A.J., Kaufman, D.S., von Gunten, L., 2018b. Building and harnessing open paleodata. *PAGES Mag.* 26. <https://doi.org/10.22498/pages.26.2>.
- Wodehouse, R.P., 1935. *Pollen Grains*. McGraw-Hill Book Company, Inc, New York; London.
- Wold, S., Sjöström, M., Eriksson, L., 2001. PLS-regression: a basic tool of chemometrics. *Chemom. Intell. Lab. Syst.* 58, 109–130.
- Woodward, F.I., Williams, B.G., 1987. Climate and plant distribution at global and local scales. *Vegetation* 69, 189–197. <https://doi.org/10.1007/BF00038700>.
- Wright, H.E., 1993. *Global Climates Since the Last Glacial Maximum*. U of Minnesota Press.
- Wu, H., Guiot, J., Brewer, S.C., Guo, Z., Peng, C., 2007a. Dominant factors controlling glacial and interglacial variations in the treeline elevation in tropical Africa. *Proc. Natl. Acad. Sci. U. S. A.* 104, 9720–9724. <https://doi.org/10.1073/pnas.0610109104>.
- Wu, H., Guiot, J., Brewer, S.C., Guo, Z.T., 2007b. Climatic changes in Eurasia and Africa at the last glacial maximum and mid-Holocene: reconstruction from pollen data using inverse vegetation modelling. *Clim. Dyn.* 29, 211–229. <https://doi.org/10.1007/s00382-007-0231-3>.
- Xu, Q., Xiao, J., Li, Y., Tian, F., Nakagawa, T., 2010. Pollen-based quantitative reconstruction of Holocene climate changes in the Daihai lake area, Inner Mongolia, China. *J. Clim.* 23, 2856–2868. <https://doi.org/10.1175/2009JCLI3155.1>.
- Zheng, Z., Wei, J., Huang, K., Xu, Q., Lu, H., Tarasov, P.E., Luo, C., Beaudouin, C., Deng, Y., Pan, A., Zheng, Y., Luo, Y., Nakagawa, T., Li, C., Yang, S., Peng, H., Cheddadi, R., 2014. East Asian pollen database: modern pollen distribution and its quantitative relationship with vegetation and climate. *J. Biogeogr.* 41, 1819–1832. <https://doi.org/10.1111/jbi.12361>.