Occupancy models reveal regional differences in detectability and improve relative abundance estimations in fossil pollen assemblages

A. Michelle Lawing [a, *], Jessica L. Blois [b], Kaitlin C. Maguire [c], Simon J. Goring [d], Yue Wang [e], Jenny L. McGuire [e, f, **]

[a] Department of Ecology and Conservation Biology, Texas A&M University, College Station, TX, USA
[b] Department of Life and Environmental Sciences, University of California, Merced, CA, USA
[c] Orma J. Smith Museum of Natural History, The College of Idaho, Caldwell, ID, USA
[d] Department of Geography, University of Wisconsin, Madison, Madison, WI, USA
[e] Schools of Biological Sciences and Earth and Atmospheric Sciences, Georgia Institute of Technology, Atlanta, GA, USA
[f] Interdisciplinary Graduate Program in Quantitative Biosciences, Georgia Institute of Technology, Atlanta, GA, 30332, USA

ARTICLE INFO

ABSTRACT

The late Quaternary fossil record provides crucial data that demonstrate how organisms respond to climate change. These records have been used to great effect, demonstrating that no-analog communities frequently occur during periods of no-analog climate, and that taxa demonstrate individualistic responses to change. However, our efforts to reconstruct biotic responses to environmental change are frequently hampered by inconsistent sampling and differential preservation of fossil taxa. We leveraged occupancy modeling methods and the fossil pollen record across eastern North America to identify circumstances under which occupancy modeling improves our ability to estimate relative abundance in four pollen taxa (*Cornus*, *Fagus*, *Picea*, and *Pinus*) through time (15 kya to present) and to identify localities where data are unreliable reflections of the local community. We found that integrating observed pollen abundance and detectability improves model performance. Low genus richness and large basin area were consistently important determinants of low detection. Our occupancy models were most informative for taxa with high enough variation in observed pollen abundance for models to be adequately calibrated. We combined occupancy model estimates of pollen abundance and detectability with a Getis-Ord statistical approach to identify spatial clusters of high or low detectability, identifying regions where a taxon's pollen is more (or less) reliable. This work will advance the integration of ecological and paleontological sciences by allowing us to better identify whether a pollen taxon is truly absent from a fossil site or if it has simply gone undetected, allowing us to produce more robust paleoecological models. This approach will bolster our ability to identify the responses of plant communities to past climatic and anthropogenic change so that we can improve our predictions of future responses.

1. Introduction

Increasing concerns about the future of species under anthropogenic climate change is driving ecologists to look to the past for evidence of how species dealt with previous shifts in climate (Willis et al., 2010; McGuire and Davis, 2014; Dietl et al., 2015; Barnosky et al., 2017). By tracing species and community dynamics through time, we can identify how ecosystems change in response to environmental changes and highlight potential drivers of those changes (Hadly Barnosky, 1994; Williams and Jackson 2007; Terry 2009; Blois et al., 2010). We use this understanding to improve ecological models and predictions of ecosystem responses to environmental change (Williams et al., 2007; Nogués-Bravo, 2009; Varela et al., 2011; Blois et al., 2013; McGuire and Davis, 2013; Maguire et al., 2015).

The palynological (fossil pollen) record is a particularly

 * Corresponding author.
 ** Corresponding author. Schools of Biological Sciences and Earth and Atmospheric Sciences, Georgia Institute of Technology, Atlanta, GA, USA.
    E-mail addresses: alawing@tamu.edu (A.M. Lawing), jmcguire@biology.gatech.edu (J.L. McGuire).

promising system to reveal insights for translating past biological dynamics into future predictions and has been used to reconstruct plant responses to past climate change (Williams et al., 2004; Blois et al., 2013; Dawson et al., 2016; Maguire et al., 2016; Wang et al., 2020). The record is composed of high-resolution sediment cores that document changes in pollen composition over hundreds to thousands of years (Fig. 1a). Palynological records show that no-analog plant communities, composed of taxa that are not typically associated today, were common during past periods of no-analog climate (Williams and Jackson, 2007). This leads us to anticipate dynamic, individualistic responses to modern climate change. As glaciers retreated following the latest glaciation, the pollen record indicates that plant taxa filled in the newly exposed landscape at unexpectedly rapid rates, emphasizing the importance of micro-refugia for survival (Gavin et al., 2014). Paleoecologists use pollen-vegetation models to translate pollen relative abundances, counted from sediment cores, into plant community compositions, providing metrics that demonstrate changes in plant identities, affiliations, and diversity through time (Dawson et al., 2016). In this paper, we propose an approach that will estimate relative pollen abundance while accounting for imperfect detection. The results of our modeling approach can be used to estimate vegetation composition more accurately by reducing errors in the detection of pollen due to the processes of pollen transport and sediment deposition.

The fossil pollen record is comprised of millions of data records from fossil localities worldwide (Brewer et al., 2012; Maguire et al., 2015), but the fossil record suffers from incomplete detection (Weng et al., 2006; Liow, 2013), which prevents us from confidently using this remarkable resource to its maximum potential. The field of taphonomy documents and develops methods to account for many aspects of incomplete detection in the fossil pollen record, pinpointing processes that contribute to preservation bias. Kujawa et al. (2016) demonstrated that the accuracy of pollen-vegetation reconstruction declines over time, likely because taphonomic processes differentially degrade pollen granules of certain taxa the longer those processes act. In addition to taphonomic bias, detection ability incorporates collection and identification biases and other biases resulting from the nuances associated with a particular site or study. Although many paleontologists consider taphonomy and study bias when performing individual site reconstructions (Behrensmeyer et al., 2000; Jackson, 2012; Kidwell, 2013), associated metadata are often lost in the amalgamation of data reported from primary literature when it is translated into records within large databases (Markwick and Lupia, 2002). Thus, it would be useful to apply an occupancy modeling approach, a method well-developed in the ecological literature to account for imperfect detection, to paleoecological data (Liow 2013). Although many ecological fields recognize the need for assessing detection (Borchers et al., 2002; Royle et al., 2013), it is still not consistently
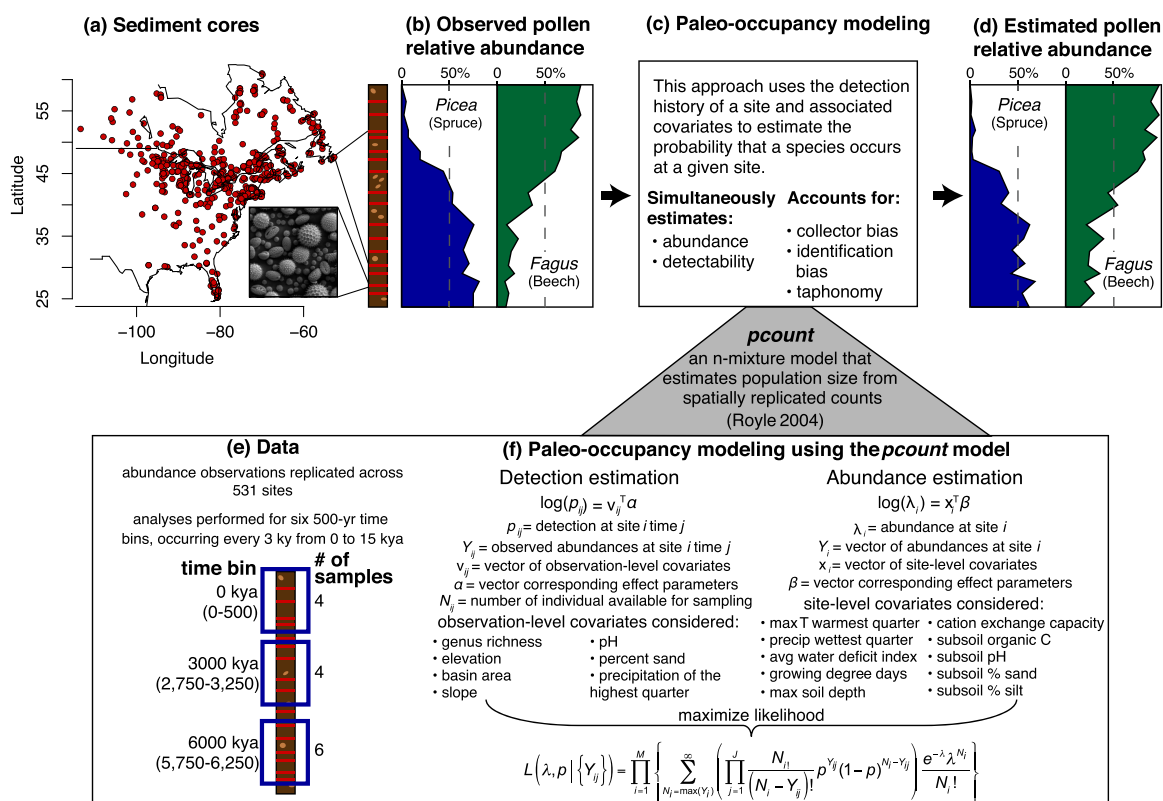

Fig. 1. Occupancy modeling framework for fossil pollen. (a) Sediment cores from eastern North America (red points) used in this study. One sediment core is expanded to show its vertical profile and stratigraphic context for pollen sampling (red lines in core). Time is represented vertically, so older samples are deeper in the core. One pollen sample is expanded to show a scanning electron microscopy photo of a theoretical pollen sample. Pollen samples are taken throughout a sediment core and each sample is counted to obtain relative abundances of pollen taxa. (b) Observed relative abundances change through time and are observed measures that do not account for taphonomy. (c) Occupancy models simultaneously estimate abundance and detectability of taxa using a statistical, maximum likelihood estimation. (d) One output of a occupancy model is a taphonomy-corrected estimate of pollen relative abundance that can be used in a pollen-vegetation model to reconstruct past vegetation communities. (e) Data used in our occupancy modeling framework are replicate samples within designated time bins. (f) Detection estimates are calculated using consistency of observations in repeated samples and variability in covariates. Abundance estimation leverages variability in detection and site covariates to refine measures of pollen relative abundance. Detection and abundance are simultaneously estimated in a maximum likelihood framework. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

incorporated even in contemporary ecological studies. A literature review of ecological articles found that only 23% account for imperfect detection (Kellner and Swihart, 2014).

Occupancy modeling is a method to statistically account for the issue of imperfect detection while simultaneously estimating either occupancy probability or abundance (MacKenzie et al., 2004). Occupancy models account for sampling effort and environmental covariates that determine the detectability of a species (MacKenzie et al., 2004; MacKenzie, 2006). Failure to consider imperfect detection can lead to strikingly different conclusions (Williams et al., 2002). In the occupancy modeling framework, occupancy is defined as the probability of a taxon's presence in a defined region, whether or not it was actually observed in that region (MacKenzie, 2006). Here, we use *pcount* as a strategy for refining fossil pollen relative abundance estimates. The *pcount* model estimates population sizes by examining replicated counts of abundance observations (Fig. 1; Royle 2004). The model simultaneously estimates the detectability and abundance of a pollen taxon at a site (Fig. 1c). Detectability is estimated by examining the consistency of observations for repeated samples, along with covariates that are relevant to those samples/observations (Fig. 1 c & f). We adapt a strategy developed by Liow (2013) that takes advantage of time-averaging and age uncertainties to use repeated samples from within a single time horizon as replicates for the purpose of estimating the detectability of a pollen taxon (Fig. 1e). Abundance estimations leverage the detectability estimate along with site-level covariates to refine estimations of pollen relative abundance (Fig. 1f). Detectability and abundance estimations are simultaneously calculated in a maximum likelihood framework (Fig. 1f).

Using this occupancy modeling framework for fossil pollen, we address three sets of questions. 1) We ask whether adding a detection component to a model of observed pollen abundance improves estimates of the relative abundance of pollen taxa through time and which covariates best explain variation in detectability of pollen taxa across a landscape. Because of study bias and biased preservation in the fossil record, we hypothesize that adding a detection component will significantly improve model estimates of relative abundance. 2) We ask whether the spatio-temporal variation in estimated relative abundance and detection probabilities vary within and across taxa. We expect detection probabilities to vary across taxa, because some taxa are high pollen producers, while other taxa have limited pollen productivity, and detection probability is positively related to the frequency of occurrence of pollen in a sediment core (Weng et al., 2006). We also hypothesize that detectability will be lower for older pollen samples (Kujawa et al., 2016). 3) We ask how reliable different geographic regions are for interpreting relative abundances from the fossil pollen record. To do so, we identify the spatial structure of detectability and estimated pollen abundance in eastern North America. This strategy allows us to identify instances where low detectability and low estimated abundances converge to create conditions where taxon occupancy is too uncertain to be reliable.

2. Methods

We used occupancy modeling to simultaneously estimate abundance and detection in fossil pollen taxa across eastern North America. Below, we discuss the fossil pollen data, the covariates we use to calibrate abundance and detection, and our approach to occupancy modeling. All modeling was executed in the R Statistical Programing environment (R Core Development Team, 2018) and relies on the package *unmarked* (Fiske and Chandler, 2011).

2.1. Pollen data

We used a dataset of 531 eastern North American sites from Maguire et al. (2016), based on sites originally examined by Blois et al. (2013). These data are archived in the Neotoma Paleoecology Database (Williams et al., 2018). Each site in this dataset represents one sediment core and each sediment core has sediment sampled at multiple depths (Fig. 1a). Pollen was counted from each sample until a standard limit was reached (typically, 300 or 500 pollen grains), thus pollen counts are relative abundances (Fig. 1b) (Gavin et al., 2003; Whitmore et al., 2005; Williams and Shuman, 2008). Taxonomic resolution of original pollen identifications varied, so only pollen taxa identified to the genus level were retained and pollen taxa identified to the species level were aggregated to the related genus level. Please see Maguire et al. (2016), Blois et al. (2013), and Williams et al. (2018) for further details on the nature of these pollen data.

We relied on the chronology inferred by Blois et al. (2013) and Blois et al. (2011) for samples within sites, which relate sediment age to depth using chronological controls such as radiocarbon dates, tephra layers, or biostratigraphic markers (e.g., *Ambrosia* rise associated with Euro-American Settlement). Pollen sample age accuracy based on cross-validation using this method is roughly 500 years (Blois et al., 2011), which led to our selection of 500-year time bins for occupancy analyses (Fig. 1e). We aggregate samples within 500-year time bins centered at six time periods in the past (0, 3, 6, 9, 12, and 15 calibrated kya).

We selected four pollen taxa from the dataset that represent different levels of pollen productivity, from low to high (Fig. 2): *Cornus* (dogwood), *Fagus* (beech), *Picea* (spruce), and *Pinus* (pine). All four of these pollen taxa are present in at least 10 sites in each of the six time periods considered, and all are important components of North American vegetation assemblages throughout the Holocene.

The occupancy model used herein is unable to handle extremely sparse matrices, i.e. matrices that contain many NA values. However, the models are more robust when there are more replicate samples. In the context of a sediment core, each replicate sample represents a sediment sample from a particular depth from which pollen was counted (Fig. 1e), and the number of replicate samples within any particular 500-year time window varies among cores depending on how densely the original researchers sampled the sediment core for pollen (number of red lines within blue bins, Fig. 1e). We attempted to find a balance between maximizing the number of replicate samples from each site to include in the analysis while minimizing NAs, so we limited the number of replicate samples and sites in a dataset for a particular time period. Sites with fewer than the number of replicates listed in Table 1 were coded as having missing data and replicates were truncated from sites with more than the established number of replicates. There are only a few sites that have a very large number of replicates, so this choice to avoid sparse matrices should not result in major differences in our interpretations.

2.2. Covariates of abundance and detectability

We relied on a contemporary understanding of pollen deposition (Appendix S1) to assign covariates for abundance and detectability. We assessed the strength of an effect of each covariate on 1) whether a plant taxon could successfully live in a region (abundance/site covariate) or 2) whether the pollen from that taxon would make it into the fossil record and be sampled and identified by a collector (detectability/observation covariate). The abundance/site covariates typically vary at the site level and represent the ecological niche of the plant taxa in determining their ability to
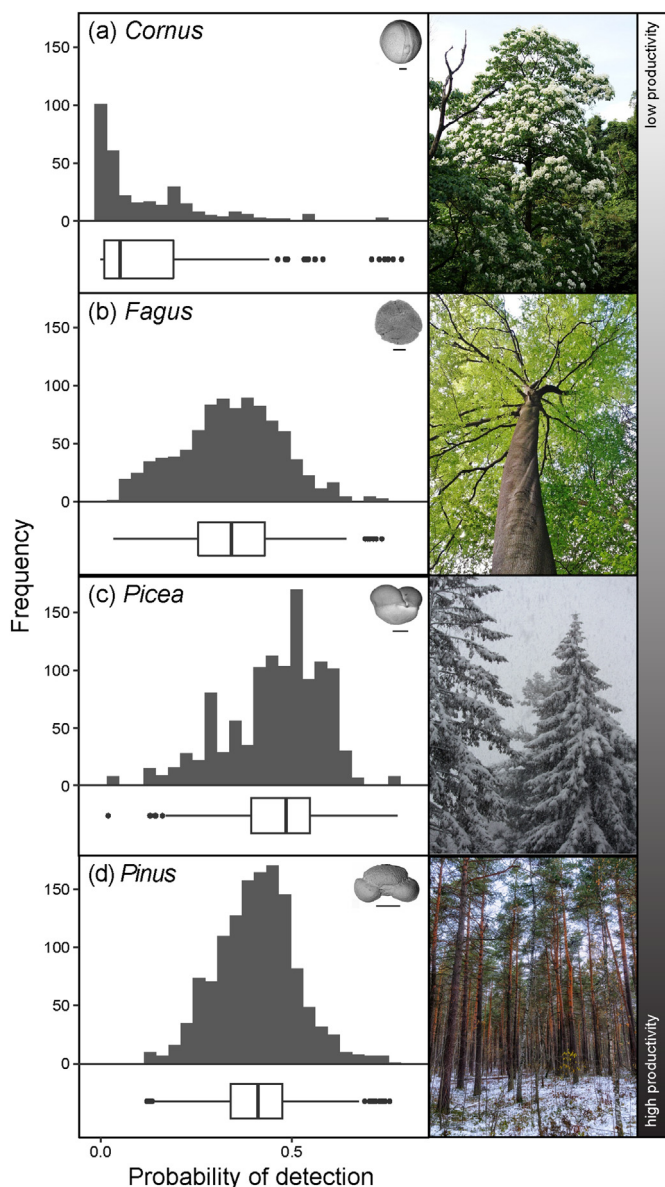
Fig. 2. Probability of detection at the most recent time period. Histograms are organized by pollen productivity. The boxplots show the median, quartiles, and outliers of detection for each taxon. Scanning electron microscope images of pollen grains are modified from Halbritter and Hesse (2004) for (a) *Cornus*, Bouchal et al. (2018) for (b) *Fagus*, Runions et al. (1999) for (c) *Picea*, and Schwendemann et al. (2007) for (d) *Pinus*. Bars representing 10 μm are under each pollen grain.

Table 1
Number of sites and replicate samples for four taxa at six time periods.

	Cornus		*Fagus*		*Picea*		*Pinus*	
kya	Sites	Replicates	Sites	Replicates	Sites	Replicates	Sites	Replicates
0	122	3	233	6	269	7	297	8
3	63	2	146	3	179	3	189	3
6	67	2	116	3	147	3	151	3
9	50	2	78	3	96	4	99	4
12	43	4	65	5	73	5	74	5
15	11	2	16	2	20	2	20	2

persist at varying relative abundances. The detectability/observation covariates typically vary at the pollen sample level and represent the taphonomic considerations in the inclusion of that

pollen type within the sample. Although we initially ran a subset of occupancy models systematically sampling all combinations of variables across both abundance and detection components for each time bin and for each taxon, running the full set of combinations was too computationally intensive. In the end, covariates were selected based on consideration of the process of pollen deposition, as summarized from the literature (Fig. 1f and discussed below). We expect to find, as many past studies have shown, that climate strongly influences the estimates of relative pollen abundance (Jackson, 1990), but that soil characteristics will also exhibit some influence (Davidson et al., 1999). We hypothesize that physiography (e.g., elevation, slope, basin size) will have a moderate influence on detection (Prentice, 1985; Sugita, 1993) and some aspects of soil and climate will have a small, but detectable influence on detection. Physiographic variables were included as detection covariates rather than abundance covariates because they have been strongly demonstrated to affect pollen taphonomy (Prentice, 1985; Sugita, 1993) and covary strongly with many of the climate covariates included as abundance covariates (Maguire et al., 2016).

Climate is a major component of the environmental niche of a species and thus determines where on the landscape a species may occur. It is a fundamental control on both taxonomic distribution and abundance, and thus is included as an abundance covariate. We considered four climate covariates that are minimally correlated ($r < 0.65$) at our sites: maximum temperature of the warmest quarter (°C), precipitation of the wettest quarter (mm), average yearly water deficit index, and growing degree days (days). We assume these four climate covariates to be relatively strong drivers of taxon abundance (Appendix S1). In the context of fossil pollen, various climate covariates also influence deposition and detectability. In particular, wind speed has been shown to influence pollen relative abundance (Jackson and Lyford, 1999). Higher precipitation limits wind dispersal, preserving more local pollen, and thus affecting local detectability. We used precipitation of the wettest quarter as a taphonomic covariate related to detectability, and we expect that it has a significant, but relatively weak, influence on detectability (Appendix S1). All climate covariates used in our analysis were described and derived in Maguire et al. (2016).

Soil characteristics influence which plant species are present in a given geographic region, thus affecting both taxonomic distribution and abundance. We selected six soil covariates that are minimally correlated ($r < 0.65$) at our sites as environmental covariates to estimate relative abundances: maximum soil depth, subsoil cation exchange capacity, subsoil organic carbon, subsoil pH, subsoil sand fraction, and subsoil silt fraction (Liu et al., 2014). While we do consider the difference of soil characteristics among different cores, we simply assumed that the soil characteristics are constant through time within a sediment core. Although this assumption is unrealistic (Willis et al., 1997), we do not have comprehensive soil maps at the continental scale through the time periods of interest, thus this is a limitation in this study. We expect soils to influence abundance, but less so than climate covariates (Appendix S1).

Some soil characteristics, including the pH and density/porosity of the sediments, can affect the detectability of taxa. The ideal environment for pollen preservation is in slightly acidic non-oxidizing soil. Soils with high pH cannot break down pollen granules that are stuck together, so those granules are difficult to identify and the resulting detectability is expected to be low (Dimbleby, 1957). Chemical oxidation in sandy soils thins the walls of pollen grains and decreases pollen detectability, as well (Havinga, 1971). We used subsoil pH and subsoil sand fraction as taphonomic covariates to calculate detectability, and we expect them to influence detectability but relatively less so than physiographic covariates (Appendix S1). Although we assumed that the

soil characteristics do not change through time, we think the influence of these variables will change throughout our six time periods because corrosion and grinding of pollen granules should compound over time. Note that we focus on soil properties because they may influence pollen preservation during transport to lake sediments, but all sediment cores were recovered from lakes, bogs, or marshes (Blois et al., 2013).

Physiographic variables influence the likelihood that pollen will make its way into a lake and thus be preserved in an appropriate depositional environment. These variables are slowly changing through time and influence detection. We used three physiographic variables measured at 1 km^2 (30 arcsecond) resolution as taphonomic covariates to calculate detectability: mean elevation (m) and mean slope (m) within 10 km of each fossil site, and basin area (km^2) (Danielson and Gesch, 2011). Basin area and slope values are log transformed. The physiographic variables are not highly correlated at our sites (in all cases, r < 0.65). We expect these covariates to be strongly associated with detection (Appendix S1).

Samples with high taxon richness have lower relative abundances on average. Thus, the probability of a given taxon being detected will be lower at higher richness. Conversely, if more taxa are present in a sample, it might indicate a depositional setting with high fidelity to the vegetation on the landscape, which would be positively associated with detection. We use genus richness as a taphonomic covariate to calculate detectability (Appendix S1).

2.3. Occupancy modeling

In occupancy modeling, a single site is sampled repeatedly throughout a season. Each time a site is sampled, a species is observed at some abundance, which might be zero. Occupancy modeling takes advantage of these repeated observations to simultaneously estimate the probability of detection and the abundance of that species given a history of observations (MacKenzie et al., 2002). It simultaneously considers covariate information that influences either abundance or detectability estimations.

We applied the *pcount* occupancy model, which fits an N-mixture model of Royle (2004), to fossil pollen count data across all sites with repeat samples for each of our time bins (Fig. 1c and f). Here, a "site" is a series of pollen samples taken from a single location within a 500-year time bin (Fig. 1a). To estimate abundance, we used repeated samples of observed relative abundances, often referred to as naive estimates in the occupancy modeling literature, from within each time bin (Fig. 1b and e). Taxa at the repeated samples for a given time period represent a single season in the occupancy modeling literature (and here represents a 500-year time window). These samples can vary in observed relative abundance and may be more or less detectable given other taphonomic factors (i.e., observation covariates; Fig. 1f). Thus, detectability varies at the observation level, represented here by each repeated sample.

We constructed four models for each taxon at each time interval to compare model strengths with and without our two sets of covariates, the detection covariates and the abundance covariates. Detection covariates correspond with repeated pollen samples, and these affect taphonomic processes that may lead to deposition and degradation of pollen granules. Abundance covariates correspond with sites, and these consist of environmental information that would affect a plant's ability to live in that location (similar to the covariates used in species distribution models; Fig. 1f). For consistency and comparability, we used the same covariates for all taxa at all time periods. Constructing these four models allows us to determine whether or not it is important to account for detection when modeling the abundance of pollen taxa across a landscape.

We used AICc for model comparison and to account for small sample sizes (Burnham and Anderson, 2002).

Detection probabilities and relative abundance estimations were calculated for all sites using the best occupancy model for each taxon-by-time comparison. Occupancy model outputs are sufficient for interpretation, but in addition, we chose to evaluate the occupancy model outputs with a Getis-Ord statistical approach in order to gain an understanding of the spatial patterns associated with the occupancy results. The Getis-Ord statistical approach, also called hot-spot analysis, is a spatially explicit approach that relies on z-scores to calculate a Getis-Ord Gi statistic that identifies spatial clusters of high or low values. This statistic uses values of neighboring sites to calculate a local average and compares that to a global average to estimate whether a deviation is significantly greater than would be expected by random chance. In our occupancy-model context, the Getis-Ord Gi statistic indicates whether, when, and where there are spatial clusters of relatively high or relatively low detectability and relative abundance estimations (Getis and Ord, 1992; Ord and Getis, 1995). For the Getis-Ord analyses, we choose a neighborhood size of 60 km, which applies across all sites. We used this analysis to evaluate the spatial structure of detection and regional reliability in observed estimates of relative abundance. We hypothesize that regional deposition rates will be influenced by regional differences in paleoclimate and in physiography (e.g., larger basin size will increase the regional signature, while dampening the local signature of vegetation) (Prentice, 1985; Sugita, 1994).

Finally, we performed a model validation by projecting models that were fitted to the 0−500 years ago time bin onto the 3 kya time bin. This is intended to ensure that the relationship between covariates and pollen abundance as estimated with our occupancy models are consistent through time. We compared projected pollen abundances with observed abundances from the 3 kya time bin using a log-log linear regression model to calculate the amount of explained variance in the projected estimates from the observed abundances. We also compared differences between estimated and observed pollen abundances within each time bin to infer how accounting for detectability changes abundance estimations. We anticipate that by accounting for detection, this approach will help improve estimates of relative abundance of pollen taxa in the palynological record.

3. Results

3.1. Does adding a detection component and covariates improve estimations of abundance?

For all taxa except *Cornus*, the best model nearly always accounted for both abundance and detection covariates (16 of 18 model comparisons; Appendix S2). For *Cornus*, detectability contributed to the best model in the four most recent time periods, but abundance only contributed in two models (6 kya and 12 kya). *Cornus* had low occupancy across sites, and where found, occurred in low abundances.

The ability of covariates to account for variation in abundance and detectability estimates was largely consistent with the process-based model of pollen deposition (Appendix S1, S2). As predicted, we found that the climate covariates had relatively higher coefficient estimates than soil covariates and both sets were generally informative for abundance estimation (Tables 2 and 3; Appendix S3). Contrary to the general pattern, *Cornus* had few influential covariates, and *Fagus* at 15 kya as well as *Pinus* at 0, 3, 6, and 9 kya had mixed strengths in the influences of climate and soil covariates on abundance (Table 2).

We predicted that there would be a strong influence on

Table 2

Coefficient estimates for abundance covariates. Double dashes (–) mark estimates not significantly different from zero. Best models are indicated with symbols for 1) no detection and no abundance covariates (□△), 2) no detection and full abundance covariates (□▲), 3) full detection and no abundance covariates (■△), and 4) full detection and full abundance covariates (■▲). Only significant coefficient estimates are reported. Standard errors, z scores, and p-values are available in Appendix S3.

Taxon	kya	best model	intercept	maximum soil depth	subsoil cation exchange	subsoil organic carbon	subsoil pH	subsoil silt fraction	subsoil sand fraction	max temp warmest quarter	precipitation wettest quarter	yearly water deficit	growing degree days
Cornus	0	■△	–	–	–	–	–	–	–	–	–	–	–
	3	□△, ■△	7.63	–	–	–	–	–	–	–	–	–	–
	6	■▲	–	–	–	–	–	7.16	8.77	9.56	–	3.4	–
	9	□△, ■△	–	–	–	–	–	–	–	–	–	–	–
	12	□▲	–	–	–	–	–	–	–	13.42	–	–	–
	15	□△	–	–	–	–	–	–	–	–	–	–	–
Fagus	0	■▲	3.05	–	–	0.11	−0.13	0.34	0.41	–	−0.46	1.16	1.05
	3	■▲	3.28	–	0.09	0.24	−0.16		0.23	–	−1.24	1.39	1.52
	6	■▲	2.3	–	0.2	−0.12	0.97	0.4	0.3	1.11	−0.46	2.44	2.12
	9	■▲	1.33	–	−0.27	−0.17	−0.49	1.53	1.72	–	−0.86	–	2.44
	12	■▲	−0.66	1.51	−1.97	−1.11	1.3	–	−2.67	−3.28	4.27	−1.84	1.46
	15	□▲	−325.5	100.25	−640	–	510.84	−436	−348	–	46.15	301.05	–
Picea	0	■▲	3.38	0.05	0.41	0.15	−0.36	–	−0.07	0.17	–	0.2	−1.21
	3	■▲	3.2	−0.47	0.51	0.15	−0.24	−0.23	−0.1	−0.57	0.52	−0.53	−1.17
	6	□▲	3.13	−0.19	0.55	0.2	−0.31	0.21	0.32	−0.61	0.5	−0.36	−1.13
	9	■▲	3.38	−0.47	−0.27	–	−0.37	0.57	0.52	0.18	–	−0.44	−1.74
	12	■▲	5.3	0.11	−0.15	–	0.21	0.33	0.37	−2.02	0.05	−0.33	1.34
	15	■▲	6.25	−0.1	–	0.08	0.14	−0.21	0.65	−0.65	−0.8	1.52	1.15
Pinus	0	■▲	5.26	0.18	0.1	−0.14	−0.45	0.05	0.37	0.13	–	−0.09	−0.08
	3	■▲	5.24	0.13	0.16	−0.13	−0.88	0.25	0.43	0.21	−0.19	−0.25	−0.27
	6	■▲	5.46	0.15	0.05	−0.04	–	0.09	0.48	0.97	−0.08	0.51	−0.71
	9	■▲	5.92	−0.09	0.24	–	−0.39	0.47	0.7	0.04	0.24	−0.24	−0.22
	12	■▲	5.04	−0.07	−0.27	−0.1	0.34	−0.06	0.24	1.46	−0.44	0.84	−0.87
	15	■▲	5.22	0.43	−1.11	−0.08	–	−0.75	−1.01	2.7	0.4	0.6	−2.12

Table 3

Coefficient estimates for detection covariates. Double dashes (–) mark estimates not significantly different from zero. Best models are indicated with symbols for 1) no detection and no abundance covariates (□△), 2) no detection and full abundance covariates (□▲), 3) full detection and no abundance covariates (■△), and 4) full detection and full abundance covariates (■▲). Only significant coefficient estimates are reported. Standard errors, z scores, and p-values are available in Appendix S3.

Taxon	kya	best model	intercept	genus richness	mean elevation	basin area (km²)	mean slope	subsoil pH	subsoil sand fraction	precipitation of the wettest quarter
Cornus	0	■△	−2.57	0.87	1.25	−0.68	–	–	–	–
	3	□△, ■△	−10.96	–	–	–	1.79	–	−1.28	–
	6	■▲	−13.37	3.01	4.77	–	−13.63	–	–	–
	9	□△, ■△	–	–	–	–	–	–	–	–
	12	□▲	–	–	–	–	–	–	–	–
	15	□△	–	–	–	–	–	–	–	–
Fagus	0	■▲	−0.71	0.49	0.4	0.09	–	−0.27	–	−0.28
	3	■▲	−0.74	1.01	−0.23	−0.38	0.51	0.99	−0.13	0.85
	6	■▲	−1.44	0.54	–	−0.57	0.84	−0.28	–	−2.35
	9	■▲	−1.23	0.74	−0.63	−1.75	0.71	–	–	–
	12	■▲	–	–	1.24	−4.58	−1.77	3.4	0.46	1.83
	15	□▲	–	–	–	–	–	–	–	–
Picea	0	■▲	−0.21	–	0.21	0.11	−0.06	0.31	−0.05	0.49
	3	■▲	0.39	0.11	−0.23	−0.29	0.24	0.84	0.25	–
	6	□▲	−0.27	–	–	–	–	–	–	–
	9	■▲	−0.9	0.12	0.22	–	−1.07	–	0.4	−0.3
	12	■▲	−0.49	0.04	–	0.56	0.2	−0.54	−0.22	−0.34
	15	■▲	−1.69	–	1.97	−0.85	−2.12	0.58	−0.77	0.38
Pinus	0	■▲	−0.39	0.39	−0.1	0.13	−0.31	–	0.07	−0.25
	3	■▲	0.07	0.56	0.13	−0.39	−0.34	0.73	–	0.32
	6	■▲	−0.06	0.38	−0.37	0.27	−0.22	−1.11	−0.24	−1.29
	9	■▲	−0.51	0.28	−0.28	−0.82	−0.06	0.21	–	−0.7
	12	■▲	−0.47	0.38	−0.1	−0.56	−0.1	0.16	−0.18	–
	15	■▲	−1.73	–	0.42	−0.69	–	–	–	−1.37

detectability from physiographic and biological covariates and a detectable, but not strong, influence from soil and climate covariates. We found that adding climate, soil, physiographic, and biological covariates improved the fit of the models, but that there was no real pattern in the strength of covariates influence on detectability (Table 3).

3.2. How do occupancy models modify estimations and predictions about relative abundance?

Estimated relative abundance was higher than the observed relative abundance at nearly all sites for all taxa and time periods (Fig. 3). There was no relationship between estimated and observed relative abundance in *Pinus*, except at 15 kya. Regardless of
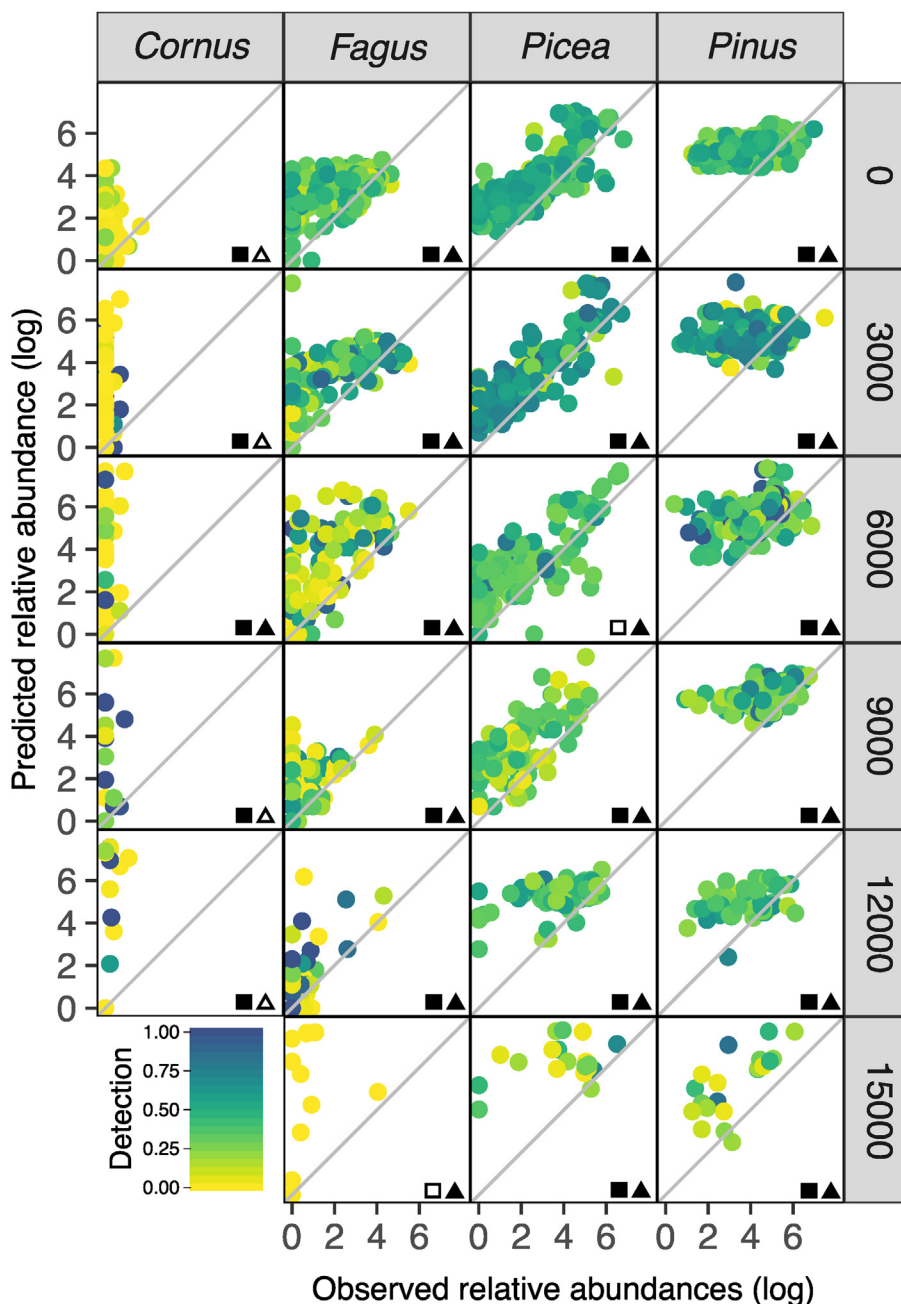
Fig. 3. Observed and estimated relative abundance for each taxon (column) by time (row) group. The blue points represent high detection probabilities and the yellow points represent low detection probabilities. Best models depicted are identified with symbols (Table 2). The line of equality is shown in gray. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

observed abundances, in this highly productive, wind-dispersed taxon, estimated abundance is always relatively high. At 15 kya, there was a relationship between estimated and observed abundance, and the y intercept was greater than zero, which indicated that estimated abundances are consistently higher than observed. *Picea* demonstrated the strongest relationship between estimated and observed abundance. In both *Picea* and *Fagus*, this relationship was stronger in recent times and deteriorates back in time. *Cornus* had consistently low observed abundances that were frequently estimated to be higher when covariates were taken into consideration.

When we test the transference of our models across time, we identified clear relationships between observed and projected

relative abundances in pollen for *Fagus* ($R^2 = 0.41$; $p < 0.001$), *Picea* ($R^2 = 0.66$; $p < 0.001$), and *Pinus* ($R^2 = 0.16$; $p < 0.001$). This indicates that the models are informative for estimating relative pollen abundance (Fig. 4). There was no relationship between the observed and projected relative abundances for *Cornus* pollen ($p = 0.57$), which indicated that the model for this pollen taxon is uninformative for estimating relative abundances (Fig. 4, Appendix S4).

3.3. What factors explain detectability of taxa across a landscape?

All detectability covariates were important explanatory variables for at least half of the taxon-by-time combinations, and in
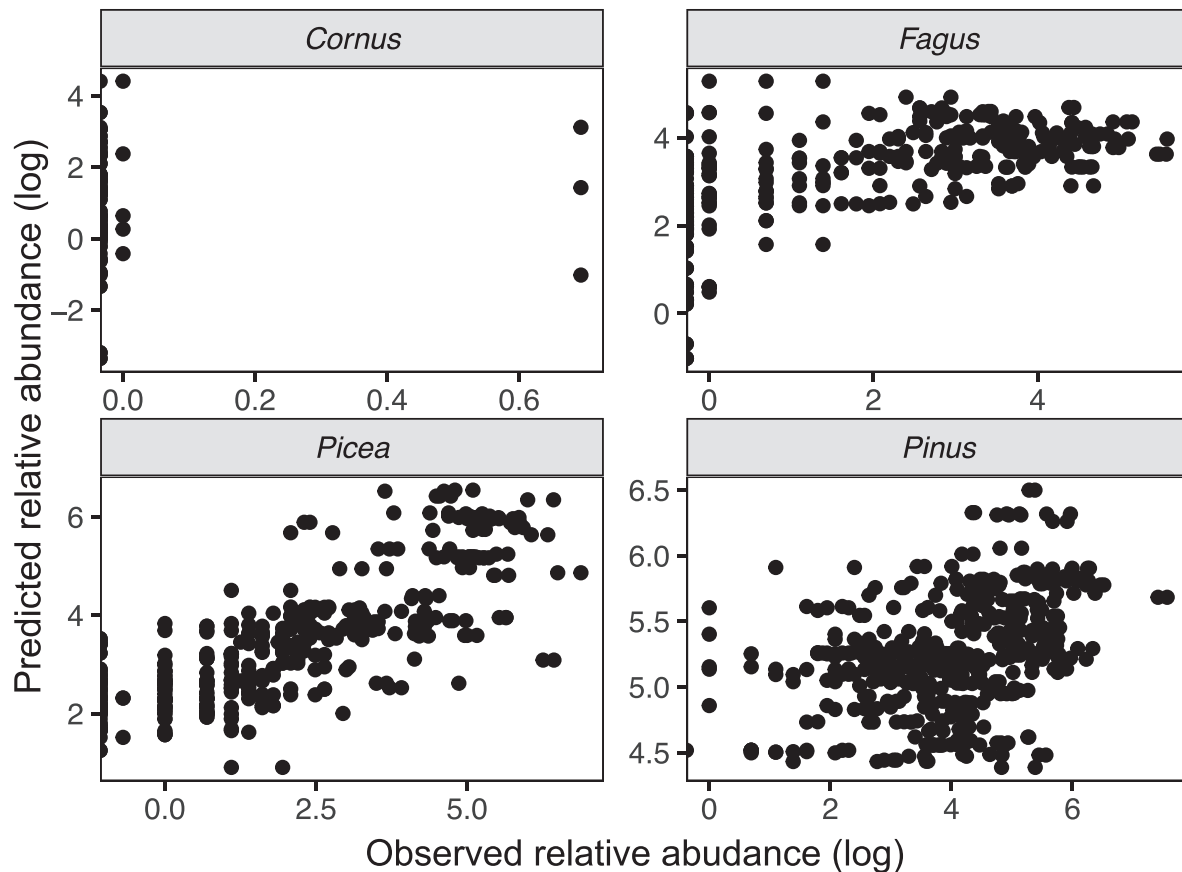
Fig. 4. Scatterplot of observed relative abundances at 3 kya and predicted relative abundances from a model calibrated at the 0−500 years ago time period and projected onto the covariates at 3 kya. There is a significant relationship between observed and predicted abundances for *Fagus*, *Picea*, and *Pinus*, but not for *Cornus* (see Appendix S4).

fact, many were relevant for most of the model comparisons (Table 2). The sign (±) of the coefficient estimates for genus richness and basin area were consistent. The coefficients on genus richness were always positive, indicating that higher genus richness is associated with higher detectability. The coefficients for basin area were mostly negative, indicating that, in general, smaller basin sizes contribute to higher detectability. Elevation, slope, pH, sand, and precipitation are also important explanatory factors of detectability for most taxa and time periods; however, the sign on the coefficient estimates fluctuate, indicating that the direction of the relationship depends on the specifics of the taxon-by-time occurrences (Table 2).

3.4. Do detection probabilities vary between taxa and what is the spatio-temporal variation in abundance and detection estimates in each taxon?

The probability of detection ranged between zero and one and varied within species across sites (Fig. 5). Distributions of detection probabilities differed. For *Fagus*, *Picea*, and *Pinus*, the detection probabilities were approximately normally distributed with the central tendency away from a zero-detection probability. *Cornus* detectability was greatest near zero and truncated at zero. *Picea* had the highest detection probability, followed by *Pinus*, *Fagus*, and *Cornus*, respectively. This order suggested that pollen productivity had a large influence on detectability. In general, detectability decreased with time, except for in *Picea* and *Pinus* at 3 kya, where it moderately increased (Fig. 5).

Detection probabilities and estimated abundance varied across

taxa and through time. Plants with higher pollen productivity, *Pinus* and *Picea*, had higher estimated abundance (Fig. 5). Regions of low detectability but high abundance estimates are identified within several maps, notably for *Cornus* and *Fagus*, and are increasingly frequent as sites become older. The high detectability scores for *Picea* and *Fagus* appeared to be the result of some regions of highly detectable low abundance estimates. However, there were few regions of low abundance estimated for *Pinus*, indicating that high detectability scores for this taxon were commonly paired with high abundance estimates.

3.5. How reliable are regions for interpreting abundance?

The analysis of regional reliability of relative abundance and detectability showed patterns of spatial clustering for both high and low values of relative abundance and high and low values of detectability. The Getis-Ord Gi statistic calculated for this analysis detected significant regional differences in estimates of relative abundance and detectability across taxa and through time (Fig. 6). Although we combine the results from these two analyses within Fig. 6 to demonstrate the relationship between abundance and detectability, the Getis-Ord Gi statistics were calculated separately for relative abundance and detectability estimates. The important aspect here was detectability, because without a reasonable amount of detectability, abundance measures will be low. Regions of high relative abundance or high detectability values that are significantly greater than expected based on the global distribution are equivalent to hot spots. Regions of low relative abundance or low detectability values that are significantly less than the expected
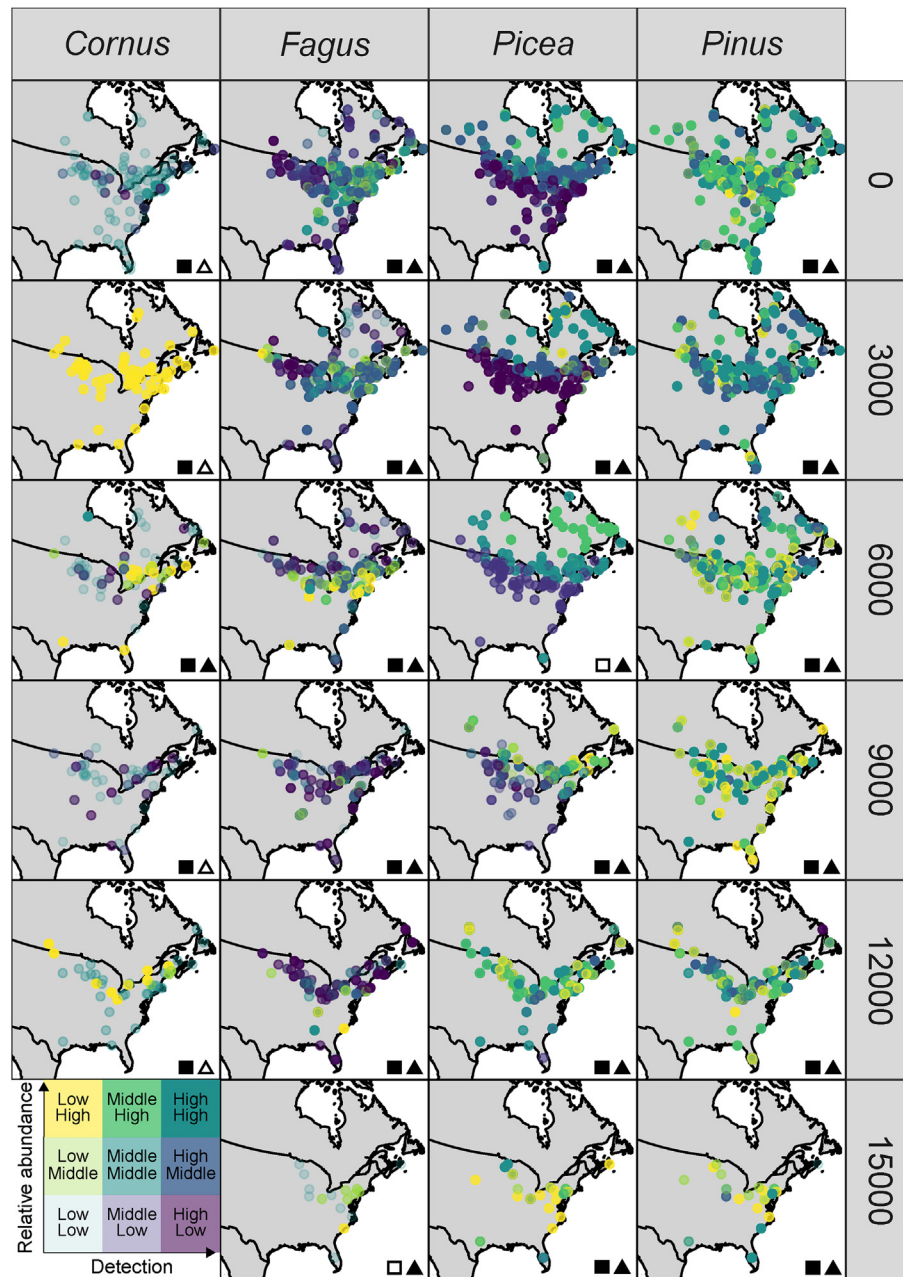
Fig. 5. Relative abundance and detection probabilities within each taxon (columns) by time (rows) group for the best model. Best models depicted are identified with symbols (Tables 2 and 3). Probabilities of abundance and detection are represented by a two-dimensional color gradient. End members of the color gradient are light blue for low relative abundance and low detectability, yellow for high relative abundance and low detectability, purple for low relative abundance and high detectability, and teal for high relative abundance and high detectability. Color categories were grouped by splitting each axis of probabilities into three quantiles. The 2D color legend replaces mapped estimates for *Cornus* at 15 kya, because the best model was the null model. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

based on the global distribution are equivalent to cold spots. We recommend performing any in-depth study on one of these taxa in regions of high detection confidence. *Picea* and *Pinus* had northeastern regional clusters of high detectability and low relative abundance estimates (purple; Fig. 6). *Picea* at 0 kya and *Pinus* at 0 and 3 kya had regional clusters in northern Quebec with low detectability and high relative abundance estimates (yellow; Fig. 6). There were few distinct regional clusters at 15 kya and generally low detection.

4. Discussion

4.1. Abundance and detection estimation: the occupancy modeling approach

Imperfect detection of species is a general problem in ecological studies, but is particularly acute in paleoecology (Weng et al., 2006; Liow, 2013; Kujawa et al., 2016). Here, we show that occupancy models can improve interpretation of the fossil pollen record by integrating estimates of detectability from replicated samples (i.e. detection probability), in addition to detection and abundance
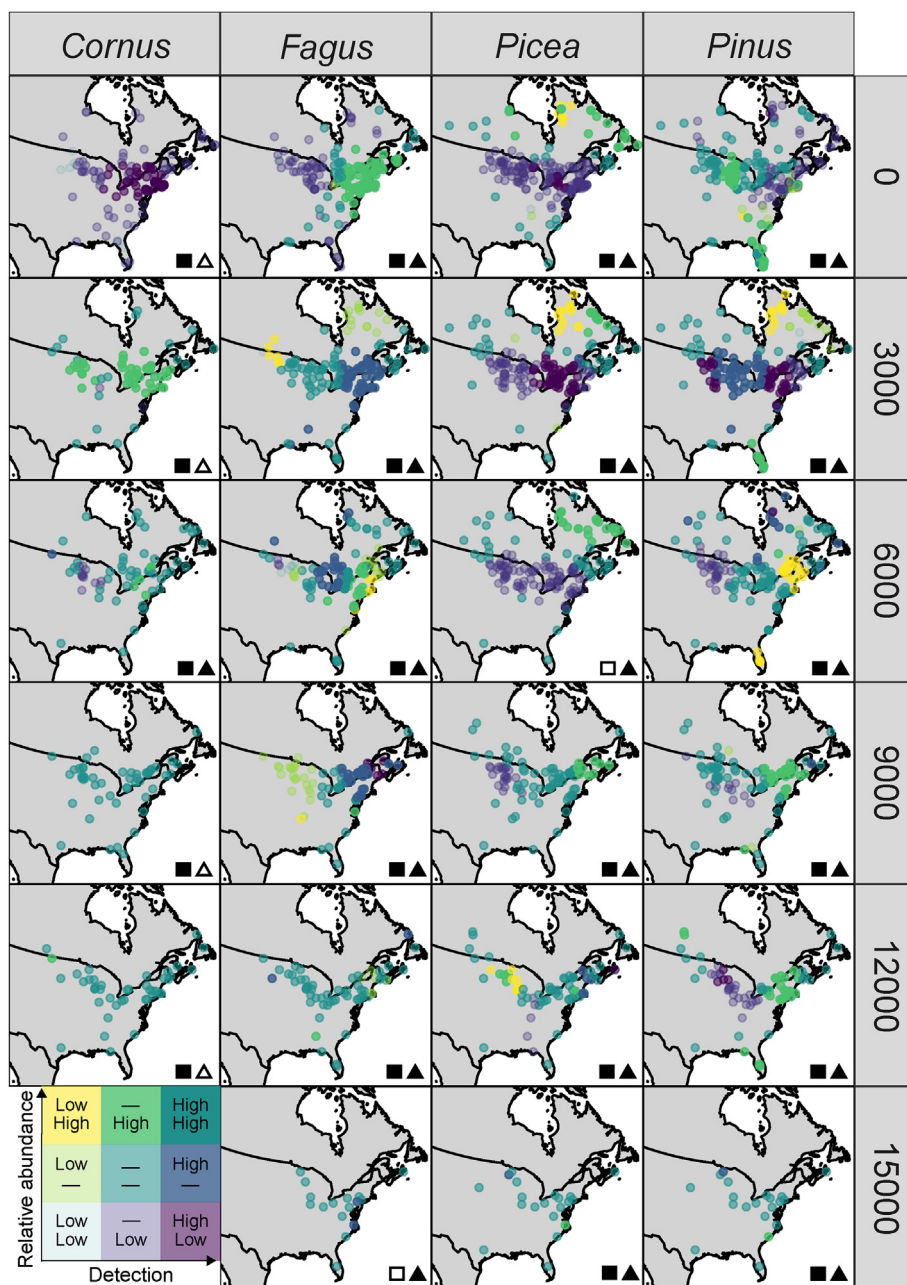
Fig. 6. Hot spot analysis calculated with Getis-Ord Gi statistic for each taxon (column) by time (row) group. Probabilities of relative abundance and detection from the best models are compared with other sites nearby to calculate Gi and determine if a region has particularly high or low probabilities. Best models depicted are identified with symbols (Tables 2 and 3). The two dimensional color gradient represents significantly low (low), not significant (−), or significantly high (high) Gi for relative abundance and detection. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

covariates, into relative abundance reconstructions (Tables 2 and 3; Figs. 2, 3 and 5). A strength of this approach is that detection probabilities and abundance estimations are decoupled (Fig. 5). As a result, we integrate these outcomes using the Getis-Ord approach to identify regions where a given taxon is highly detectable whether or not it is likely to be abundant. This can demonstrate geographic regions where we could confidently focus taxon-specific studies (Fig. 6). A region that has low detection but high abundance estimates could be interpreted as one where over-dispersion of a motile pollen is likely to have occurred. Importantly, we can identify regions likely to have true absences (purple in Fig. 5) as opposed to low detection (pale blue in Figs. 5 and 6). For example, using this method we can have confidence that *Picea*

pollen was truly not present in the regions to the east of the Great Lakes over the last 3000 years, and that the genus likely has a hard-to-detect distribution in northern Quebec (yellow in Fig. 5).

Both the abundance and detection covariates that we identified as likely relating to the relative abundances of pollen granules significantly improve most models (Tables 2 and 3). Importantly, we find that detectability covariates are important explanatory variables for at least half of the models, although relationships vary across taxa and through time (Table 2). However, the strength of the relationships of abundance covariates from the process-based model do generally hold. For example, the climate covariates more strongly influence abundance estimates than soil covariates, but both are important more often than not.

Plant productivity and dispersal affect both detectability and relative abundance estimations. *Cornus*, which has very low pollen productivity and biologically-facilitated dispersal, has the lowest observed relative abundances and the most extreme increases in estimated relative abundances (Fig. 3). However, due to the nature of these data (i.e., extremely low sample sizes), there is little power to estimate detectability. Thus, occupancy modeling is not expected to improve estimates of relative abundance for this group. Occupancy modeling is probably most informative for taxa with similar characteristics to *Fagus*, a taxon with a combination of relatively low pollen productivity and high dispersal ability via wind (Fig. 3). Likewise, *Picea*, which has high pollen productivity and is also anemophilous, has abundance estimates that seem to be improved by occupancy modeling. *Pinus*, however, is estimated to be highly abundant everywhere when occupancy models are applied, likely due to its incredibly high pollen motility and omnipresence in the observed pollen record.

Occupancy models are informative to estimate pollen relative abundance for taxa with moderate to high variation in relative abundance across a landscape (e.g., probably not taxa like *Cornus* or *Pinus*). These models produce updated relative abundance estimations for pollen taxa that differ from observed pollen relative abundances (Fig. 3). However, there is negligible explained variance in the log-log linear regression model for testing the association between estimated and observed relative abundance of *Cornus* at 3 kya (Fig. 4). Although the predicted-observed association of *Pinus* at 3 kya shows a regression slope significantly different from zero, the amount of explained variation is still low at 16%, so it is unclear whether it is justifiable to apply a complicated modeling approach to adjust *Pinus* relative abundances prior to use in downstream modeling efforts. *Fagus* and *Picea* have the highest explained variance in the predicted-observed relationship (41% and 66%, respectively), and we recommend applying this occupancy modeling framework to adjust those relative abundance estimates prior to use in downstream modeling efforts.

4.2. Assumptions and considerations

Our approach assumes processes influencing deposition, preservation, and vegetation on the landscape, reflected in abundance and detection covariates, are relatively stable within the 500-year time bins. Support for this assumption differs depending on the underlying ecological dynamics within a time period. Additionally, these 500-year time bins reflect the confidence interval of the age estimates of the sediment cores (Blois et al., 2011). Overall, grouping samples across periods of large environmental change runs the risk of classifying absences as false absences instead of true absences. In our study, we examine six time intervals (0, 3, 6, 9, 12, and 15 kya). At least two of these intervals correspond to times of rapid change (12 kya is during the Younger Dryas period and 15 kya encompasses the start of the Bolling-Allerod period). Our choice of 500 years is largely pragmatic, striving for a narrow enough time window to satisfy the assumptions but also capture multiple samples of a taxon. Thus, the decrease in detection probabilities at older time periods (Figs. 5 and 6) may be due not just to sample age but to the rapidity of climate change during the 12 and 15 kya time periods, which resulted in real differences in fossil pollen abundance estimates across samples.

In a traditional occupancy-modeling framework, repeat visits to a site result in multiple observations of abundance in a relatively short window of time. More visits to a site increase sample sizes to estimate abundance and detectability, thus increasing the power of statistical analysis. However, because of the inability of occupancy models to handle sparse matrices, we truncated our replicate samples to even out the number of samples per site (see Methods).

Thus, because of uneven temporal sampling across sites, we sacrificed the power that comes from having an increased number of replicates in favor of having more sites. Despite this sacrifice, including detection covariates in our modeling framework generally improved model estimations of relative abundance.

4.3. Relationship between fossil pollen and vegetation

Occupancy models improve most estimations of relative abundances of fossil pollen. However, fossil pollen does not directly indicate the relative abundances of plants on the landscape. To estimate plant assemblage composition, many factors must be considered, including plant pollen productivity and pollen dispersal distances, none of which are included in our models. Models have been developed to convert pollen abundances to landscape vegetation (called PVMs or pollen-vegetation models), including STEPPS (Dawson et al., 2016) and LOVE or REVEALS (Sugita, 2007a, b). We suggest that the outputs of occupancy models can serve as inputs for these conversion models, which should reduce uncertainty and improve estimations of plant communities and their evolution over time.

In addition, many studies also directly use pollen abundance estimates to examine changes in plant taxon relative abundances in response to specific drivers (Blois et al., 2013; Maguire et al., 2016). The next step for implementing an occupancy-modeling approach to model vegetation changes on a landscape through time would be to determine if using corrected relative abundances improves model fit within PVMs or distribution models relative to the original abundances. Caution should be taken when replacing estimated relative abundance for observed relative abundance by examining the fit of the model to other proxies such as plant macrofossils, which may more accurately indicate local presence.

5. Conclusions

Here, we are able to demonstrate that occupancy models are useful for correcting relative pollen abundance estimates in a way that accounts for taphonomic covariates (Fig. 3). Accurate relative abundance estimates for fossil pollen are important because they are used in subsequent models estimating vegetation (Williams and Shuman, 2008; Dawson et al., 2016), reconstructing paleoclimate (Bartlein et al., 2011; Marsicek et al., 2018), or estimating past distributions (Williams et al., 2004). Yet, fossil pollen abundance estimates are subject to a variety of potential biases, many of them related to our ability to accurately detect fossil pollen. Further, the increasing use of databases has the potential to amplify existing bias across many datasets (e.g. Inman et al. (2018)). There are many approaches aimed at quantifying and correcting for these biases (reviewed recently by Dawson et al. (2016)), with different data requirements and complexity. We suggest the Getis-Ord approach to further assess the relative confidence that we have in the detectability of pollen taxa for the appropriate selection of regional analyses.

We are borrowing information through time in the form of repeated samples, and explicitly accounting for the potential influence of physiographic and edaphic factors in our models of relative abundance. In doing so, occupancy modeling may be able to account for some of the taphonomic biases present in fossil pollen data and maximize the potential of large datasets to address ecological questions. Further, occupancy modeling provides a flexible framework that applies across scales. For example, we have applied this approach to estimate abundance and detection probabilities across many sites, but it can also be applied at a local scale to one or a few sites, where more detailed knowledge of the local physical context may alter the choice of abundance and detection

covariates to include in the process model. The next step is to add adjusted relative abundances into vegetation reconstruction models to capture a more realistic picture of the vegetation landscape through time.

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

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.quascirev.2020.106747.

Author contributions

JM initially conceived the project; ML led the analyses; JM and ML led the writing; JM, KM, JB and ML designed methodology; JB and SG developed conceptual framework for covariates; KM compiled pollen data and traits; KM and JM compiled covariates data; YW, ML, JM designed model validation methodology. All authors contributed to writing, revising, and approve the final version of the manuscript.

Data Accessibility

Derived data and R code are archived at Mendeley Data and available at https://doi.org/10.17632/rnbrkxxm9m.1.

References

Barnosky, A.D., Hadly, E.A., Gonzalez, P., Head, J., Polly, P.D., Lawing, A.M., Eronen, J.T., Ackerly, D.D., Alex, K., Biber, E., Blois, J., Brashares, J., Ceballos, G., Davis, E., Dietl, G.P., Dirzo, R., Doremus, H., Fortelius, M., Greene, H.W., Hellmann, J., Hickler, T., Jackson, S.T., Kemp, M., Koch, P.L., Kremen, C., Lindsey, E.L., Looy, C., Marshall, C.R., Mendenhall, C., Mulch, A., Mychajliw, A.M., Nowak, C., Ramakrishnan, U., Schnitzler, J., Das Shrestha, K., Solari, K., Stegner, L., Stegner, M.A., Stenseth, N.C., Wake, M.H., Zhang, Z., 2017. Merging paleobiology with conservation biology to guide the future of terrestrial ecosystems. Science 355, eaah4787.

Bartlein, P.J., Harrison, S.P., Brewer, S., Connor, S., Davis, B.A.S., Gajewski, K., Guiot, J., Harrison-Prentice, T.I., Henderson, A., Peyron, O., Prentice, I.C., 2011. Pollen-based continental climate reconstructions at 6 and 21 ka: a global synthesis. Clim. Dynam. 37, 775–802.

Behrensmeyer, A.K., Kidwell, S.M., Gastaldo, R.A., 2000. Taphonomy and paleobiology. Paleobiology 26, 103–147.

Blois, J.L., McGuire, J.L., Hadly, E.A., 2010. Small mammal diversity loss in response to late-Pleistocene climatic change. Nature 465, 771.

Blois, J.L., Williams, J.W., Fitzpatrick, M.C., Ferrier, S., Veloz, S.D., He, F., Liu, Z., Manion, G., Otto-Bliesner, B., 2013. Modeling the climatic drivers of spatial patterns in vegetation composition since the Last Glacial Maximum. Ecography 36, 460–473.

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.

Borchers, D.L., Buckland, S.T., Stephens, W.E., Zucchini, W., 2002. Estimating animal abundance: closed populations, 13. Springer Science & Business Media.

Bouchal, J.M., Güner, T.H., Denk, T., 2018. Middle Miocene climate of southwestern Anatolia from multiple botanical proxies. Clim. Past 14, 1427–1440.

Burnham, K., Anderson, D., 2002. Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach, second ed. Springer, New York, New York, USA.

Danielson, J.J., Gesch, D.B., 2011. Global multi-resolution terrain elevation data 2010 (GMTED2010). Open-File Report.

Davidson, D.A., Carter, S., Boag, B., Long, D., Tipping, R., Tyler, A., 1999. Analysis of pollen in soils: processes of incorporation and redistribution of pollen in five soil profile types. Soil Biol. Biochem. 31, 643–653.

Dawson, A., Paciorek, C.J., McLachlan, J.S., Goring, S., Williams, J.W., Jackson, S.T., 2016. Quantifying pollen-vegetation relationships to reconstruct ancient forests using 19th-century forest composition and pollen data. Quat. Sci. Rev. 137, 156–175.

Dietl, G.P., Kidwell, S.M., Brenner, M., Burney, D.A., Flessa, K.W., Jackson, S.T., Koch, P.L., 2015. Conservation paleobiology: leveraging knowledge of the past to inform conservation and restoration. Annu. Rev. Earth Planet Sci. 43, 79–103.

Dimbleby, G.W., 1957. Pollen analysis of terrestrial soils. New Phytol. 56, 12–28.

Fiske, I.J., Chandler, R.B., 2011. An R package for fitting hierarchical models of wildlife occurrence and abundance. J. Stat. Software 43, 1–23.

Gavin, D.G., Fitzpatrick, M.C., Gugger, P.F., Heath, K.D., Rodríguez-Sánchez, F., Dobrowski, S.Z., Hampe, A., Hu, F.S., Ashcroft, M.B., Bartlein, P.J., 2014. Climate refugia: joint inference from fossil records, species distribution models and phylogeography. New Phytol. 204, 37–54.

Gavin, D.G., Oswald, W.W., Wahl, E.R., Williams, J.W., 2003. A statistical approach to evaluating distance metrics and analog assignments for pollen records. Quat. Res. 60, 356–367.

Getis, A., Ord, J.K., 1992. The analysis of spatial association by use of distance statistics. Geogr. Anal. 24, 189–206.

Hadly Barnosky, E., 1994. Ecosystem dynamics through the past 2000 years as revealed by fossil mammals from lamar cave in yellowstone national park, USA. Hist. Biol. 8, 71–90.

Halbritter, H., Hesse, M., 2004. Principal modes of infoldings in tricolp (or) ate Angiosperm pollen. Grana 43, 1–14.

Havinga, A.J., 1971. An Experimental Investigation into the Decay of Pollen and Spores in Various Soil Types. Sporopollenin. Elsevier, pp. 446–479.

Inman, R., Franklin, J., Esque, T., Nussear, K., 2018. Spatial sampling bias in the Neotoma paleoecological archives affects species paleo-distribution models. Quat. Sci. Rev. 198, 115–125.

Jackson, S.T., 1990. Pollen source area and representation in small lakes of the northeastern United States. Rev. Palaeobot. Palynol. 63, 53–76.

Jackson, S.T., 2012. Representation of flora and vegetation in Quaternary fossil assemblages: known and unknown knowns and unknowns. Quat. Sci. Rev. 49, 1–15.

Jackson, S.T., Lyford, M.E., 1999. Pollen dispersal models in Quaternary plant ecology: assumptions, parameters, and prescriptions. Bot. Rev. 65, 39–75.

Kellner, K.F., Swihart, R.K., 2014. Accounting for imperfect detection in ecology: a quantitative review. PLoS One 9 (10), e111436.

Kidwell, S.M., 2013. Time-averaging and fidelity of modern death assemblages: building a taphonomic foundation for conservation palaeobiology. Palaeontology 56, 487–522.

Kujawa, E.R., Goring, S., Dawson, A., Calcote, R., Grimm, E.C., Hotchkiss, S.C., Jackson, S.T., Lynch, E.A., McLachlan, J., St-Jacques, J.M., Umbanhowar Jr., C., 2016. The effects of anthropogenic land cover change on pollen-vegetation relationships in the American Midwest. Anthropocene, pp. 60–71.

Liow, L.H., 2013. Simultaneous estimation of occupancy and detection probabilities: an illustration using Cincinnatian brachiopods. Paleobiology 39, 193–213.

Liu, S., Wei, Y., Post, W.M., Cook, R.B., Schaefer, K., Thornton, M.M., 2014. NACP MsTMIP: unified North American soil map. In: Center, O.D.A.A. (Ed.).

MacKenzie, D.I., 2006. Modeling the probability of resource use: the effect of, and dealing with, detecting a species imperfectly. J. Wildl. Manag. 70, 367–374.

MacKenzie, D.I., Bailey, L.L., Nichols, J.D., 2004. Investigating species co-occurrence patterns when species are detected imperfectly. J. Anim. Ecol. 73, 546–555.

MacKenzie, D.I., Nichols, J.D., Lachman, G.B., Droege, S., Andrew Royle, J., Langtimm, C.A., 2002. Estimating site occupancy rates when detection probabilities are less than one. Ecology 83, 2248–2255.

Maguire, K.C., Nieto-Lugilde, D., Blois, J.L., Fitzpatrick, M.C., Williams, J.W., Ferrier, S., Lorenz, D.J., 2016. Controlled comparison of species-and community-level models across novel climates and communities. Proc. Biol. Sci. 283, 20152817.

Maguire, K.C., Nieto-Lugilde, D., Fitzpatrick, M.C., Williams, J.W., Blois, J.L., 2015. Modeling species and community responses to past, present, and future

episodes of climatic and ecological change. Annu. Rev. Ecol. Evol. Syst. 46, 343–368.

Markwick, P.J., Lupia, R., 2002. Palaeontological databases for palaeobiogeography, palaeoecology and biodiversity: a question of scale, 194. Geological Society, London, Special Publications, pp. 169–178.

Marsicek, J., Shuman, B.N., Bartlein, P.J., Shafer, S.L., Brewer, S., 2018. Reconciling divergent trends and millennial variations in Holocene temperatures. Nature 554, 92.

McGuire, J.L., Davis, E.B., 2013. Using the palaeontological record of Microtus to test species distribution models and reveal responses to climate change. J. Biogeogr. 40, 1490–1500.

McGuire, J.L., Davis, E.B., 2014. Conservation paleobiogeography: the past, present and future of species distributions. Ecography 37, 1092–1094.

Nogués-Bravo, D., 2009. Predicting the past distribution of species climatic niches. Global Ecol. Biogeogr. 18, 521–531.

Ord, J.K., Getis, A., 1995. Local spatial autocorrelation statistics: distributional issues and an application. Geogr. Anal. 27, 286–306.

Prentice, I.C., 1985. Pollen representation, source area, and basin size: toward a unified theory of pollen analysis. Quat. Res. 23, 76–86.

R Core Development Team, 2018. R: A Language and Environment for Statistical Computing.

Royle, J.A., 2004. N-mixture models for estimating population size from spatially replicated counts. Biometrics 60, 108–115.

Royle, J.A., Chandler, R.B., Sollmann, R., Gardner, B., 2013. Spatial capture-recapture. Academic Press.

Runions, C.J., Rensing, K.H., Takaso, T., Owens, J.N., 1999. Pollination of Picea orientalis (Pinaceae): saccus morphology governs pollen buoyancy. Am. J. Bot. 86, 190–197.

Schwendemann, A.B., Wang, G., Mertz, M.L., McWilliams, R.T., Thatcher, S.L., Osborn, J.M., 2007. Aerodynamics of saccate pollen and its implications for wind pollination. Am. J. Bot. 94, 1371–1381.

Sugita, S., 1993. A model of pollen source area for an entire lake surface. Quat. Res. 39, 239–244.

Sugita, S., 1994. Pollen representation of vegetation in quaternary sediments: theory and method in patchy vegetation. J. Ecol. 82, 881–897.

Sugita, S., 2007a. Theory of quantitative reconstruction of vegetation I: pollen from large sites REVEALS regional vegetation composition. Holocene 17, 229–241.

Sugita, S., 2007b. Theory of quantitative reconstruction of vegetation II: all you need is LOVE. Holocene 17, 243–257.

Terry, R.C., 2009. The dead do not lie: using skeletal remains for rapid assessment of historical small-mammal community baselines. Proc. Biol. Sci. 277, 1193–1201.

Varela, S., Lobo, J.M., Hortal, J., 2011. Using species distribution models in paleobiogeography: a matter of data, predictors and concepts. Palaeogeogr. Palaeoclimatol. Palaeoecol. 310, 451–463.

Wang, Y., Shipley, B.R., Lauer, D.A., Pineau, R., McGuire, J.L., 2020. Plant biomes demonstrate that landscape resilience today is the lowest it has been since end-Pleistocene megafaunal extinctions. Global Change Biol. 26, 5914–5927.

Weng, C., Hooghiemstra, H., Duivenvoorden, J.F., 2006. Challenges in estimating past plant diversity from fossil pollen data: statistical assessment, problems, and possible solutions. Divers. Distrib. 12, 310–318.

Whitmore, J., Gajewski, K., Sawada, M., Williams, J., Shuman, B., Bartlein, P., Minckley, T., Viau, A., Webb Iii, T., Shafer, S., 2005. Modern pollen data from North America and Greenland for multi-scale paleoenvironmental applications. Quat. Sci. Rev. 24, 1828–1848.

Williams, B.K., Nichols, J.D., Conroy, M.J., 2002. Analysis and Management of Animal Populations. Academic Press.

Williams, J., Shuman, B., 2008. Obtaining accurate and precise environmental reconstructions from the modern analog technique and North American surface pollen dataset. Quat. Sci. Rev. 27, 669–687.

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., 2018. The Neotoma Paleoecology Database, a multiproxy, international, community-curated data resource. Quat. Res. 89, 156–177.

Williams, J.W., Jackson, S.T., 2007. Novel climates, no-analog communities, and ecological surprises. Front. Ecol. Environ. 5, 475–482.

Williams, J.W., Jackson, S.T., Kutzbach, J.E., 2007. Projected distributions of novel and disappearing climates by 2100 AD. Proc. Natl. Acad. Sci. Unit. States Am. 104, 5738–5742.

Williams, J.W., Shuman, B.N., Webb III, T., Bartlein, P.J., Leduc, P.L., 2004. Late-Quaternary vegetation dynamics in North America: scaling from taxa to biomes. Ecol. Monogr. 74, 309–334.

Willis, K., Bailey, R., Bhagwat, S., Birks, H., 2010. Biodiversity baselines, thresholds and resilience: testing predictions and assumptions using palaeoecological data. Trends Ecol. Evol. 25, 583–591.

Willis, K.J., Braun, M., Sümegi, P., Tóth, A., 1997. Does soil change cause vegetation change or vice versa? A temporal perspective from Hungary. Ecology 78, 740–750.