Quaternary Science Reviews 198 (2018) 115–125

Contents lists available at ScienceDirect

Quaternary Science Reviews

journal homepage: www.elsevier.com/locate/quascirev

Spatial sampling bias in the *Neotoma* paleoecological archives affects species paleo-distribution models

Richard Inman [a], [b], [*], Janet Franklin [c], Todd Esque [a], Kenneth Nussear [d]

[a] *US Geological Survey, Western Ecological Research Center, Las Vegas Field Station, 160 N. Stephanie Street, Henderson, NV, 89074, USA*
[b] *School of Geographical Sciences and Urban Planning, Arizona State University, PO Box 875302, Tempe, AZ, 85287, USA*
[c] *Department of Botany and Plant Sciences, University of California – Riverside, 900 University Ave, Riverside, CA, 92521, USA*
[d] *Department of Geography, University of Nevada – Reno, 1664 N. Virginia Street, Reno, NV, 89557, USA*

ARTICLE INFO

Article history:
Received 14 May 2018
Received in revised form
15 August 2018
Accepted 16 August 2018

Keywords:
Holocene
Paleogeography
North America
Data treatment
Data analysis
Neotoma
Spatial sampling bias
Species distribution modeling
Niche breadth
Macrofossils

ABSTRACT

The ability to infer paleo-distributions with limited knowledge of absence makes species distribution modeling (SDM) a useful tool for exploring paleobiogeographic questions. Spatial sampling bias is a known issue when modeling extant species. Here we quantify the spatial sampling bias in a North American packrat midden archive and explore its impact on estimating paleo-distributions. We test whether (1) spatial sampling bias inherent in this macrofossil record can influence estimates of paleo-distributions, (2) this bias can alter the ability to measure shifts in distributions and climatic niche breadth from the Northgrippian subdivision of the Holocene (8.3 ka – 4.2 ka) to present day (1950–2000 yr), and (3) bias correction methods can improve estimates of paleo-distributions and analyses of range shifts and niche breadth. We estimate spatial sampling bias for the mid-Holocene period with a three-stage statistical model, each representing a hypothesized source of bias: fossil site availability, preservation and accessibility. This approach enables the use of SDM to evaluate three separate paleo-distributions calibrated on the packrat midden archive: those without bias correction (σ-naïve), those created with a standard method (σ-standard), and those created with a novel alternative (σ-modeled) incorporating the three-stage model of bias. We find that paleo-distributions modeled for the mid-Holocene without bias correction (σ-naïve) provided poor estimates of hindcast paleo-distributions, and that the σ-modeled correction method improved paleo-distributions for our six species with, on average, 50% higher overlap to hindcast distributions than σ-naïve paleo-distributions (σ-standard results fell between σ-naïve and σ-modeled).

Published by Elsevier Ltd.

1. Introduction

A core focus of biogeography rests in understanding the determinants of species distributions and the processes by which they change. Towards that goal, rapid development of paleoecological archives and analytical tools over the past 20 years has enabled investigations of broad macro-ecological, evolutionary and conservation questions about the mechanisms and forces altering patterns of biodiversity throughout the history of our planet (Brewer et al., 2012; Swetnam and Allen, 1999). Species distribution modeling (SDM) has become a widely used tool in paleobiogeographic studies because in a presence-background modeling framework, SDM does not rely on explicit knowledge of absence localities (Franklin, 2010), which are often difficult to determine (Brotons et al., 2004). The ability to use presence-only data, such as fossil records, allows SDM to infer paleo-distributions with limited knowledge of absence and address paleobiogeographic questions of niche stability (e.g. Stigall, 2012), range dynamics (e.g. Nogués-Bravo et al., 2008; Veloz et al., 2012), speciation (e.g. Peterson and Nyári, 2008), and extinction (e.g. Lorenzen et al., 2011), among many others.

When used to investigate paleo-distributions, SDM offers two analytical opportunities: 1) validation of models calibrated on extant species that have been 'hindcast' to past environmental conditions, and 2) direct calibration with past environmental conditions (paleo-SDM). The former has been used most often, driven by the wealth of spatially explicit ecological archives available for extant species (Moreno-Amat et al., 2017). In this approach, models

---

[*] Corresponding author. US Geological Survey, Western Ecological Research Center, Las Vegas Field Station, 160 N. Stephanie Street, Henderson, NV, 89074, USA.
E-mail address: rdinman@usgs.gov (R. Inman).

https://doi.org/10.1016/j.quascirev.2018.08.015
0277-3791/Published by Elsevier Ltd.

are calibrated using present-day conditions and are applied to environmental data representing paleoclimatic conditions to create spatial predictions for the historic period of interest. Paleoecological archives are then used to quantify the accuracy of these hindcast projections (e.g. Franklin et al., 2015; Nogués-Bravo, 2009) or to qualitatively evaluate the congruence between hindcast projections and fossil localities (e.g. Carnaval and Moritz, 2008). We refer to these predictions of paleo-distributions as 'hindcast' distributions. These hindcast projections describe areas where a species could have occurred given its present-day realized niche; but rarely do these hindcast distributions consider exogenous factors such as fire or megafaunal disturbances that may have constrained paleo-distributions (Gill et al., 2009; Nogués-Bravo, 2009). In contrast, paleo-SDM uses paleoecological archives for calibrating models under past environmental conditions. Paleo-SDM is used less often, likely because paleo-distribution data are usually sparse in time and space and can be poorly resolved chronologically and taxonomically (Moreno-Amat et al., 2017). However, paleo-SDM allows a key assumption of hindcasting to be relaxed, namely, the requisite of niche conservation through time. Under niche conservation, species-environment parameters are maintained through time, even with environmental change (Wiens and Graham, 2005).

By directly calibrating on fossil archives, paleo-SDM can estimate new species-environment parameters for different chronological periods. Each unique set of parameters may provide evidence of a changed niche, and in conjunction with spatial predictions of habitat potential, can be compared across different time slices to evaluate potential change in niche and habitat (e.g. Saupe et al., 2014). Often, studies using paleo-SDM rely on the spatial predictions of habitat potential because many clear metrics exist for evaluating overlap among time slices, such as the Sorenson's similarity index (Sørensen, 1948), Schoener's D (Schoener, 1968) or Godsoe's ESP (Godsoe, 2013). These metrics, among others, allow for simple comparisons of spatial predictions rather than complex assessments of niche complexity, dimensionality, or breadth. When coupled with hindcasting, direct tests of niche conservation can be made between distributions derived from hindcasting and those from fossil calibrated models (e.g. Veloz et al., 2012; Walls and Stigall, 2011). These coupled approaches have been used often, contributing evidence that niches evolve slowly (e.g. Martinez-Meyer et al., 2004; Peterson, 2011; Stigall, 2012). However, paleoecological archives used for paleo-SDM often have small sample sizes, which can be biased in space and time (Varela et al., 2011; Vilhena and Smith, 2013).

Spatial bias is problematic for paleo-SDM because these methods assume niches are sampled over the full range of environmental conditions in which they occur (Phillips et al., 2009) and that sampling mirrors the background distribution of environmental covariates (Araújo and Guisan, 2006). These assumptions are not often met with paleoecological archives due to spatial variation of taphonomic conditions in different deposits (Allison and Bottjer, 2010), which results in spatially biased fossils where more fossils are found in certain areas due not to a greater prevalence of an organism, but instead due to a lack of fossils in other areas. Spatial bias in biodiversity data for extant species is often addressed with one or more bias correction methods that have been developed for presence-background frameworks. These methods stem from careful filtering of observation data (e.g. Boria et al., 2014) or from estimating the biased sampling distribution (σ) and manipulating background selection weights to result in proportional background samples (e.g. Phillips et al., 2009). In order be effective, these bias correction methods require large sample sizes. Paleoecological archives, however, rarely offer large samples (Moreno-Amat et al., 2017; Varela et al., 2011), and the ability to estimate σ from them is not often tested. An alternative method to estimate σ that does not rely on large sample sizes may improve estimates of paleo-distributions, thereby allowing analyses of range shifts and niche characteristics through time.

Here we explore potential effects of sampling bias on paleo-SDM and investigate an alternative for estimating σ with a focal group of extant plant species. Specifically, we test whether (1) spatial sampling bias inherent in a commonly used paleoecological archive, the USGS/NOAA North American Packrat Midden database (here after NAPM database; Strickland et al., 2013), can influence estimates of paleo-distributions, (2) this bias can alter our ability to measure shifts in distributions and niche breadth from the Northgrippian subdivision of the Holocene (8.3 ka – 4.2 ka) (here after mid-Holocene; Cohen et al., 2013) to present day (1950–2000 yr), and (3) bias correction methods can improve paleo-distributions and analyses of range shifts and niche breadth. While there is evidence that the niches of many North American plant species have shifted since the last glacial period of the late Quaternary, changes since the mid-Holocene period have been minimal (Ordonez, 2013; Veloz et al., 2012). We therefore use paleo-distributions created by hindcasting present-day models to the mid-Holocene as our reference distributions and compare them to three separate paleo-distributions calibrated on the fossil record: those without bias correction (σ-naïve), those created with a standard method (σ-standard), and those created with a novel alternative (σ-modeled).

Using hindcast distributions for paleo-SDM evaluation assumes that hindcast distributions are representative of mid-Holocene distributions, and so to independently verify the paleo-SDM distributions we also used pollen records from lake sediment cores from the western USA not used to calibrate paleo-distributions. We aim to identify an effective method for reducing sampling bias in paleo-SDM and to highlight how bias may affect analyses of range shifts and niche breadth.

2. Materials and methods

2.1. Study area and environmental gradients

Our study region covered 3,171,335 km^2 of the western USA encompassing the locations of packrat middens represented in the NAPM database (Strickland et al., 2013). We assembled uncorrelated raster data describing 11 climatic and physiographic environmental conditions to characterize present-day (1950–2000 yr) and mid-Holocene time periods, and generalized each at a spatial scale of 1 km (Appendix A). Interpolated climate variables for current conditions were obtained from WorldClim (Hijmans et al., 2005) as average monthly values and climate variables representing the mid-Holocene were obtained as downscaled paleoclimate simulations from the National Center for Atmospheric Research (NCAR) from the Community Climate System Model (CCSM4; Gent et al., 2011), also available through WorldClim. We assumed that most changes between the two periods were limited to climatic variables and that any changes in surface physiographic conditions due to Holocene erosion or deposition were minimal and within 2x the maximum vertical error of the terrain elevation data; 10 m (Danielson and Gesch, 2011).

2.2. Species distribution modeling

We developed present-day and paleo-distributions for six extant species well represented in the NAPM database that occur in a range of habitat types, including two small perennial shrubs inhabiting cold desert habitats (*Artemisia tridentata*, *Coleogyne ramosissima*), a small deciduous tree inhabiting foothills and low mountain elevations (*Quercus gambelli*), a small conifer tree confined to northern latitudes (*Juniperus communis*), and two large