

ECOGRAPHY

Research article

Reliability of presence-only data for assessing plant community responses to climate warming

L. Camila Pacheco-Riaño¹✉, Sabine Rumpf², Tuija Maliniemi³,
Suzette G. A. Flantua⁴ and John-Arvid Grytnes¹

EDITOR'S
CHOICE

¹Department of Biological Sciences, University of Bergen, Bergen, Norway

²Department of Environmental Sciences, University of Basel, Basel, Switzerland

³Geography Research Unit, University of Oulu, Oulu, Finland

⁴Department of Biology, University of Bergen and Bjerknes Centre for Climate Research, Bergen, Norway

⁵Department of Biological and Environmental Sciences, University of Gothenburg, Gothenburg, Sweden

Correspondence: L. Camila Pacheco-Riaño (pachecoriano.c@gmail.com)

Ecography

2024: e07213

doi: 10.1111/ecog.07213

Subject Editor: Tim Newbold

Editor-in-Chief: Miguel Araújo

Accepted 5 March 2024



Climate warming has triggered shifts in plant distributions, resulting in changes within communities, characterized by an increase in warm-demanding species and a decrease in cold-adapted species – referred to as thermophilization. Researchers conventionally rely on co-occurrence data from vegetation assemblages to examine these community dynamics. Despite the increasing availability of presence-only data in recent decades, their potential has largely remained unexplored due to concerns about their reliability. Our study aimed to determine whether climate-induced changes in community dynamics, as inferred from presence-only data from the Global Biodiversity Information Facility (GBIF), corresponded with those derived from co-occurrence plot data. To assess the differences between these datasets, we computed a community temperature index (CTI) using a transfer function, weighted-averaging partial least squares regression (WA-PLS). We calibrated the transfer function model based on the species–temperature relationship using data before recent climate warming. Then we assessed the differences in CTI and examined the temporal trend in thermophilization. In a preliminary analysis, we assessed the performance of this calibration using three datasets: 1) Norwegian co-occurrence data, 2) presence-only data from a broader European region organized into pseudo-plots (potentially capturing a larger part of the species niches), and 3) a combined dataset merging 1) and 2). The transfer function including the combined dataset performed best. Subsequently, we compared the CTI for the co-occurrence plots paired up spatially and temporally with presence-only pseudo-plots. The results demonstrated that presence-only data can effectively evaluate species assemblage responses to climate warming, with consistent CTI and thermophilization values to what was found for the co-occurrence data. Employing presence-only data for evaluating community responses opens up better spatial and temporal resolution and much more detailed analyses of such responses. Our results therefore outline how a large amount of presence-only data can be used to enhance our understanding of community dynamics in a warmer world.



www.ecography.org

© 2024 The Authors. Ecography published by John Wiley & Sons Ltd on behalf of Nordic Society Oikos

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

Keywords: assemblages, biodiversity informatics, climate warming, GBIF, plant community, thermophilization

Introduction

Ongoing climate warming induces range shifts of species which track their optimal temperature conditions (Steinbauer et al. 2018, Lenoir et al. 2020, Mamantov et al. 2021, O'Sullivan et al. 2021). These shifts subsequently lead to local reorganizations of species communities and assemblages (Menéndez et al. 2006, Walther 2010), often resulting in a relative increase of warm-demanding species and/or a decreasing number of cold-demanding species, a pattern referred to as thermophilization (Gottfried et al. 2012). Thermophilization has been reported globally from a wide variety of habitats and areas, from mountain tops (Gottfried et al. 2012) to forests (Zellweger et al. 2020) and from temperate (Pacheco-Riaño et al. 2023) to tropical regions (Fadrique et al. 2018).

Studies on multi-decadal vegetation responses to climate change such as thermophilization commonly rely on permanent or semi-permanent plot-based datasets as the main source of information (Bertrand et al. 2011, Götzenberger et al. 2012, Chytrý et al. 2014, Kapfer et al. 2017, Fadrique et al. 2018, Freeman et al. 2021, Richard et al. 2021). Permanent plots are best suited to track vegetation dynamics, and multiple initiatives were set up at the beginning of the 21st century for continuous long-term monitoring (Gottfried et al. 2012, Chytrý et al. 2014, Haider et al. 2022). Although the number of permanent plots is continuously increasing, they are, however, geographically scattered, and most cover relatively short time spans. Historical co-occurrence plots (herein broadly defined as species records co-occurring in a specific site) were initially done to describe the structure and diversity of vegetation types (e.g. phytosociological plots, also called relevés), and are in this context often referred to as semi-permanent plots (Kapfer et al. 2017). Resurveys of these semi-permanent plots have also proven to be a valuable source of information to describe vegetation dynamics over decades (Kapfer et al. 2017), and to study range shifts (Lenoir et al. 2008, Felde et al. 2012, Rumpf et al. 2018) and thermophilization responses (Pacheco-Riaño et al. 2023).

To study vegetation dynamics over time, an alternative to co-occurrence plots is the aggregation of haphazardly collected species occurrence records (i.e. presence-only data from individual species observations) such as museum and herbaria collections or, more recently, observations from various structured or unstructured citizen science projects. Presence-only data have been collected extensively over the last century and their number has increased enormously over the past 20 years (Heberling et al. 2021). The world's largest biodiversity data network, mediated by The Global Biodiversity Information Facility (GBIF) (<http://gbif.org>), stands as the leading open-access data portal for geo-referenced species occurrence data collected from a myriad of different sources (König et al. 2019, Wüest et al. 2020). GBIF provides access to more than

1.5 billion species records from across the globe and the tree of life.

Presence-only data offer substantially more extensive temporal and spatial coverage compared to co-occurrence data due to the vast network of data collectors. The availability of presence-only data therefore opens up the opportunity to investigate community responses in various ways, including using presence-only data from regions that lack historical co-occurrence data or for regions with limited data availability in the contemporary context. It is also worth considering combining both data types, where the integration of presence-only data and co-occurrence data can help mitigate some of the biases inherent to each data type (Vellend et al. 2013). However, presence-only data are commonly considered unreliable for many community analyses and have so far only been exploited to a limited extent to assess community responses to global warming (Feeley 2012, Vellend et al. 2013, Bottin et al. 2020, Duchenne et al. 2021, Lajeunesse and Fourcade 2023). In addition to missing reliable information about species absence, presence-only records are prone to biases stemming from identification errors, geographically misplaced records, etc., due to the diversity of collectors and data sources (Vellend et al. 2013, Beck et al. 2014, Meyer et al. 2016, Hughes et al. 2021).

Assessing thermophilization relies on species co-occurrences in a specific area combined with species-specific thermal indicator values to calculate community temperature indices (CTI), i.e. the (weighted) average of the thermal indicator values for species assemblages. The CTI approach is an effective way to summarize thermophilization trends by comparing changes in CTI over time not only for plants but also for other organisms such as birds and butterflies (Devictor et al. 2008, 2012, Duque et al. 2015, Feeley et al. 2020, Richard et al. 2021). It can provide an unbiased estimation of thermophilization regardless of sampling differences, as long as there is no disproportionate collection of warm-demanding or cold-adapted species compared to their actual occurrences in the area (or vice versa). In other words, if the sampling efforts do not favour one type of species over the other in terms of their thermal preferences, the CTI can be assumed to accurately reflect the degree of thermophilization in each community. Presence-only data could thus hold a great potential to fill spatial and temporal gaps in studies of species dynamics and, therefore, allow for a more comprehensive understanding of climate-driven responses of species across their geographical ranges (König et al. 2019).

One approach to quantify the CTI is through calibrating a relationship between species occurrences and temperature by using transfer functions (Bertrand et al. 2011, Pacheco-Riaño et al. 2023). Transfer functions are statistical tools commonly used in paleoecology. In paleoecology, the calibration

phase of a transfer function establishes the relationship between species occurrences and climatic conditions from the present time to reconstruct past climates from sediment cores (Guiot and de Vernal 2007, Juggins and Birks 2012, Schäbitz et al. 2013). This approach builds on the assumption that species have approximately symmetrical, unimodal response curves with an ecological optimum for climate variables for each species (Hutchinson 1957). It also assumes that the relationship between species and climate remains constant through time (ecological uniformitarianism) (Rull 2010), so that the inferred species–climate relationship from the present time can be used to reconstruct past climates from community composition (Salonen et al. 2011). This approach has been used to reconstruct various environmental conditions in palaeoecology, from water chemistry (e.g. pH) using diatoms (ter Braak and Juggins 1993), to temperature (Chevalier et al. 2020) and precipitation (Lu et al. 2019) using fossil pollen. A corresponding approach has recently been utilized in modern ecology to estimate a CTI by inferring temperature from co-occurrence vegetation data (Bertrand et al. 2011, Pacheco-Riaño et al. 2023). In this case, historical co-occurrence plots sampled prior to major climatic changes were used to calibrate a transfer function (estimate species–climate relationship), which was subsequently used to project the CTI based on more recent vegetation plot data (Bertrand et al. 2011, Bhatta et al. 2018, Pacheco-Riaño et al. 2023). Based on this approach, thermophilization can be estimated as the difference between the floristically inferred temperature (i.e. CTI) and the observed temperature from the calibrating period (Pacheco-Riaño et al. 2023).

Exploiting the vast amount of presence-only data to analyse the responses of communities to climate warming requires, however, a rigorous evaluation of robustness and reliability (Bayraktarov et al. 2019). Using the CTI approach we make this evaluation by comparing the CTI from co-occurrence plots with presence-only data and determining the suitability of using presence-only data for this type of analysis. Therefore, our aim was to answer three key questions. First, are the CTI values derived from presence-only data obtained from GBIF different from the CTI values acquired from co-occurrence plot data for the same time and place? Second, do changes in community dynamics (i.e. thermophilization) inferred from presence-only data correspond with the results obtained from co-occurrence plot data? Last, can these two datasets be used interchangeably, either independently or in combination, to produce similar community responses during the model calibration and prediction phases? In our study, we integrated co-occurrence plot data from Norway with spatially and temporally aggregated presence-only data (referred to as pseudo-plots) from Europe. Within this framework, we assessed the differences in the CTI and the variation in thermophilization index, hypothesized minimal differences in CTI values between the two datasets, and anticipated a consistent thermophilization pattern independent of which dataset was used. If our hypotheses hold true, this will give some support for using the vast amount of presence-only data when assessing thermophilization patterns.

Material and methods

All analyses were conducted in R ver. 4.0.2 (www.r-project.org). The package ‘tidyverse’ ver.1.3.2 (Wickham et al. 2019) was used to clean and handle the data, ‘sf’ ver. 1.0-9 (Pebesma 2018) and ‘raster’ ver.3.6 (Hijmans 2023) for spatial data manipulation, and ‘ggplot2’ ver. 3.4.0 for visualization (Wickham et al. 2019).

Data

Co-occurrence dataset

We used the same non-standardized co-occurrence dataset used by Pacheco-Riaño et al. (2023), which consisted of 605 637 records of terrestrial vascular plants (3617 taxa including genera, species, subspecies, and varieties) from Norway. These records had their origins in a series of field campaigns, where each campaign resulted in a check-list from a specific location, conducted by various collectors across Norway since the early 1900s and until 2007. These records were digitized as field notes from The Agder naturmuseum og botaniske hage, University of Oslo, University of Trondheim, and the Norwegian University of Life Sciences, integrated into GBIF, and stored as occurrence data. To reconstruct the co-occurrence plots, we grouped the data by coordinates, year, elevation, and collector as described in Pacheco-Riaño et al. (2023), and only kept taxa at species level or merged lower taxonomical units to species level, using GBIF’s backbone taxonomy tool (GBIF.org 2020). This resulted in 41 993 co-occurrence plots with 2888 species covering the time span from 1905 to 2007 (Fig. 1).

Presence-only dataset

We downloaded from GBIF all global georeferenced records for those vascular plant taxa included in the non-standardized co-occurrence dataset described above (3617 taxa: genera, species, subspecies, and varieties; Supporting information). This included a total of 212 286 166 records from 6831 datasets (8 December 2022, see Data availability section; GBIF.org 2022).

We applied six automated cleaning procedures to eliminate known issues with presence-only data using the package ‘CoordinateCleaner’ (Zizka et al. 2019): 1) equal coordinates (records with identical longitude and latitude), 2) zero coordinates (plain zeros in the coordinates and a radius around), 3) capitals (radius around capital cities), 4) centroids (radius around country and province centroids), 5) sea coordinates (non-terrestrial records), and 6) biodiversity institutions (radius around biodiversity institutions). All flagged records, or records with missing coordinates or missing record years, were removed. As for the co-occurrence dataset, we harmonized the taxonomy using GBIF’s backbone taxonomy tool (GBIF.org 2020), and kept records identified at the species and intraspecific level, e.g. subspecies, but merged all records

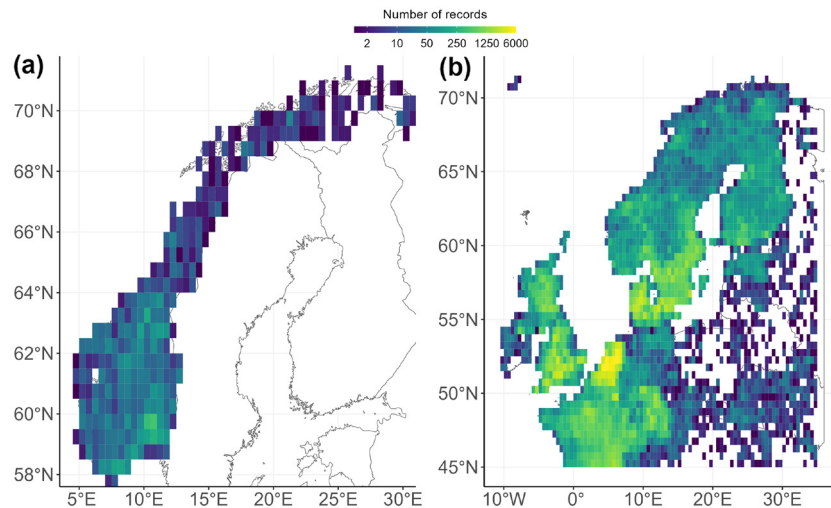


Figure 1. The density distribution of assemblages. (a) Co-occurrence plots covering Norway, and (b) presence-only pseudo-plots spanning across Europe. Both panels represent all vascular plant species recorded between 1905 and 2007/2016 per $0.5 \times 0.5^\circ$ degrees cell.

to species level. Subsequently, duplicated records with the same species names, coordinates, and year were removed, and only occurrences within Europe (longitude: $-12^\circ 0' 0''$ W to $35^\circ 0' 0''$ W, latitude: $45^\circ 0' 0''$ N to $72^\circ 0' 0''$ N) were kept. Last, to prevent any duplication with the co-occurrence dataset, we excluded records with the same species names, year, and geographical coordinates as contained in the co-occurrence records. This process resulted in a cleaned, presence-only dataset of 68 112 130 occurrence records of 2742 species (Fig. 1).

To build ‘communities’ with spatio-temporal information from the presence-only data, we created ‘pseudo-plots’ by aggregating species occurrences temporally and spatially by using the same grid cells as for the temperature dataset described below (i.e. 30 arc seconds resolution, approx. 463 m at a latitude of 60° N and longitude of 10° E). When there were multiple observations of a single species within a pseudo-plot, we randomly chose one record for each species, ensuring only one record per species in the pseudo-plots. Only pseudo-plots between 1905 and 2016 were kept for further analyses.

Temperature data

From here and henceforth, we will use the word ‘assemblage’ to refer to both co-occurrence plots and presence-only pseudo-plots. Temperature was extracted for each vegetation assemblage’s geographical location (co-occurrence and pseudo-plot) using CHELSAcruts. This climate dataset provides monthly data for every year from 1901 up to 2016 at a spatial resolution of 30 arc seconds (Karger et al. 2017, Karger and Zimmermann 2018). We calculated mean annual temperatures (MAT) by averaging maximum and minimum temperature values for all months of the respective year. The final MAT for each vegetation assemblage was calculated as the average of the 5 years preceding the respective field survey, including surveyed year (Steinbauer et al. 2018).

Community temperature index, CTI

To estimate CTI for each assemblage, a transfer function was first calibrated with the assemblages from the period before the recent major increase in temperature, the baseline period (1950–1979, Fig. 1; see Pacheco-Riaño et al. (2023) for a justification for selecting this period). Subsequently, this transfer function was used to project CTI for the assemblages before and after this baseline period (i.e. 1905–1949 and 1980–2007/16, respectively). The transfer function was calibrated using a weighted-averaging partial least squares regression (WA-PLS) using the ‘fxTWAPLS’ package (Liu et al. 2020). This is an efficient approach to creating unbiased CTI estimates (Bhatta et al. 2019, Liu et al. 2020).

The transfer function WA-PLS is based on estimating species optima along a gradient of MAT and assumes that the full temperature range of a species is captured in the calibration dataset. If this is not the case, the niches are truncated leading to an overestimation of the CTIs at the lower end and an underestimation at the higher end of the temperature range (Liu et al. 2020). Prior to the main analyses, we therefore evaluated the effect of using three different datasets in the transfer function calibration phase, and computed three different transfer functions based on either 1) the co-occurrence plots from Norway, 2) the pseudo-plots derived from presence-only data from Europe (assuming that the extent captured a wider thermal gradient than Norwegian plant species have and thus avoid niche truncation, at least at the warmer limit), and 3) the two datasets combined. To have sufficient species representation to calibrate the transfer function, we only retained species that occurred in more than ten assemblages in each dataset. In the calibration subset, this resulted in 717 species for the co-occurrence dataset (1287 assemblages), 2273 species for the presence-only dataset (8029 assemblages), and 2290 species for the combined dataset (9928 assemblages). In addition, we kept only those assemblages that contained a minimum of ten species for the projection phase. Through

the application of these criteria, we ensured that each assemblage had sufficient species representation to estimate CTI using a transfer function (Bhatta et al. 2018). Note that we did not utilize any abundance data for any of the datasets. Consequently, when constructing the CTI value, equal weight was assigned to all species in both the co-occurrence plots and pseudo-plots. The performance of each transfer function (i.e. WA-PLS) was evaluated through leave-one-out cross-validation, specifically focusing on the calibration assemblages located in Norway and utilizing only 75% of the calibrating data from these Norwegian assemblages. We also used these cross-validation analyses to identify the statistically reliable component of the WA-PLS, which was determined by the combination of the lowest root mean square error of prediction (RMSEP), the highest R^2 value, and the lowest average bias (Liu et al. 2020). For all three datasets the inclusion of the second component gave the RMSEP, the highest R^2 , and the lowest average bias within Norway. For this evaluation, only assemblages from the same period that was used for the calibration itself were included. The co-occurrence dataset had the highest values for RMSEP (3.54), R^2 (0.81), and average bias (0.117), followed by the presence-only dataset with 3.12, 0.68, and 0.043, respectively, and the combined dataset had the lowest values with 3.09, 0.67, and 0.032, respectively (Supporting information).

Plotting the predicted CTI versus the observed MAT for the assemblages in the baseline period demonstrated that all transfer functions overestimated the CTI at low temperatures and underestimated the CTI at high temperatures (Supporting information). To correct for this remaining over- and underestimation due to niche truncation in the ends, we used a local nonparametric regression (LOESS) with a span of 0.75 to fit the residuals of the WA-PLS model. This choice of a span of 0.75 indicates that approximately 75% of the data points contribute to each local regression, making it a relatively smooth estimate. The difference between the residuals of the WA-PLS model and the LOESS model was then used to correct the initially predicted CTI in the main analyses. We fitted the LOESS only for the vegetation assemblages within Norway due to their colder climate relative to the rest of Europe, preventing potential underestimation of CTI values (Supporting information). Finally, to reduce niche truncation effects near the edges during the projection phase (Liu et al. 2020), we further refined our dataset by excluding communities with extreme temperatures and focused our CTI projections on assemblages within a MAT range of -2.5°C to 8.5°C (initial MAT range -5.4° – 13°C), representing 95% of the calibration data (Pacheco-Riaño et al. 2023). Note that these corrections were expected to affect the two datasets (co-occurrence plots and pseudo-plots) equally, and when the main analyses were made, i.e. comparing the CTIs based on the two datasets, without these corrections it had no effect on the results. However, the calibration models performed better when employing the corrections (Supporting information) and were therefore kept in the following analyses. After these corrections, the co-occurrence dataset had an R^2 of 0.86, the presence-only dataset of 0.79, and the combined

dataset of 0.78 (Supporting information). Next, we carried out CTI projections for each vegetation assemblage, covering the period of 1905 to 1949 before the baseline and for the contemporary period spanning from 1980 to 2007/16. Because the evaluation of the different datasets indicated that the three different transfer functions had different advantages based on the RMSEP, R^2 , and average bias, we made projections with all three transfer functions to compare co-occurrence and presence-only in the subsequent analyses.

Comparison of CTI between co-occurrence and presence-only

Given that co-occurrence plots are commonly employed in CTI and thermophilization analyses (De Frenne et al. 2013, Fadrique et al. 2018, Richard et al. 2021), we established the CTI of co-occurrence plots as a benchmark. We then assessed the degree of similarity between the estimated CTI derived from pseudo-plots generated from presence-only data and the CTI estimated from co-occurrence plots for the three calibration models (co-occurrence, presence-only, and combined). To be able to make this comparison, we reproduced any spatiotemporal biases of the co-occurrence dataset in the presence-only dataset. For this we paired up the co-occurrence plots with the pseudo-plots by subsampling the latter, finding the closest available pseudo-plot that was sampled. We only paired up a pseudo-plot with a co-occurrence plot if the centre of a pseudo-plot was found within a radius of 1 km from the co-occurrence plot, if the difference in sampling time was less than 2 years, and if the MAT of the assemblages had a difference of less than 0.5°C . If a co-occurrence plot did not have a corresponding pseudo-plot given these criteria, it was discarded. As the co-occurrence dataset only extends until 2007, the paired period was constrained to the same timeframe, resulting in 3839 assemblage pairs for the period 1905–2007. For each assemblage pair we quantified the difference between the CTI estimated based on the pseudo-plot and on the co-occurrence plot (ΔCTI) as follows:

$$\Delta\text{CTI} = \text{CTI}_{\text{pseudo-plots}} - \text{CTI}_{\text{Co-occurrence plots}}$$

Positive ΔCTI indicates an overestimation of CTI for the pseudo-plot relative to the co-occurrence plot. A value of zero indicates equal CTI estimation while negative ΔCTI indicates an underestimation of CTI for the pseudo-plots. To statistically assess these differences, we conducted t-tests. Further, to evaluate the effectiveness of the pseudo-plots in estimating community responses, and to reveal any temporal deviance from the co-occurrence plots, we examined the temporal trend of thermophilization independently for the two already paired datasets. We estimated the thermophilization index as the difference between the CTI and the observed baseline temperature (mean MAT 1950–1979) for each vegetation assemblage in the paired dataset (Pacheco-Riaño et al. 2023). A positive value indicates an increase in the CTI (an increase in thermophilic species) compared to

the baseline period of 1950–1979, while a negative value suggests a decrease in the CTI (an increase in cryophilic species) during the same reference period. Note that this thermophilization trend was not bias-corrected as was done in Pacheco-Riaño et al. (2023), but the same potential spatiotemporal bias was kept for both datasets.

Results

Comparison of predicted values for co-occurrence and presence-only data

The average ΔCTI in the paired dataset ($\text{CTI}_{\text{pseudo-plots}} - \text{CTI}_{\text{co-occurrence plots}}$) for the study period (1905–2007) was low, independent of which calibration model was used, but it was statistically significantly different when calibrating the transfer function with the co-occurrence data only (Fig. 2a; paired t-test average difference = 0.062, SD error 0.016, $t = 3.7068$, $df = 3838$, $p\text{-value} < 0.001$). The model calibrated with presence-only data showed an average difference of 0.029 (Fig. 2b; $t = 1.35$, $df = 3838$, $p\text{-value} = 0.17$), and the model calibrated with the combined dataset showed an average difference of 0.014 (Fig. 2c; $t = 0.68$, $df = 3838$, $p\text{-value} = 0.49$). In all cases the presence-only CTI were slightly higher than those of the co-occurrence dataset.

Since the transfer function calibrated with the combined dataset showed the best performance measures (Supporting information) and showed the smallest difference in paired CTI, we relied on this calibration model as the basis for illustrating the temporal pattern of thermophilization for both datasets (Fig. 2d). This revealed that there was no statistically significant difference between thermophilization trends based on the co-occurrence and presence-only datasets, and that the thermophilization values increased steadily around the same time (Fig. 2d).

Ad hoc analyses of the CTI deviations

Across all transfer functions, there was a tendency for CTI overestimation in the presence-only data, particularly at lower temperatures (Fig. 2, Supporting information). This overestimation remained within the credibility interval when utilizing the transfer function based on the combined dataset, but deviations from the credibility interval were noted when using the transfer functions derived solely from either co-occurrence data or presence-only data (Fig. 2, Supporting information). Because the areas of colder temperatures tend to be placed in areas of higher topographical reliefs in Norway, we suspected that these deviations were due to elevational biases in the presence-only data towards more accessible and warmer valley bottoms compared to the

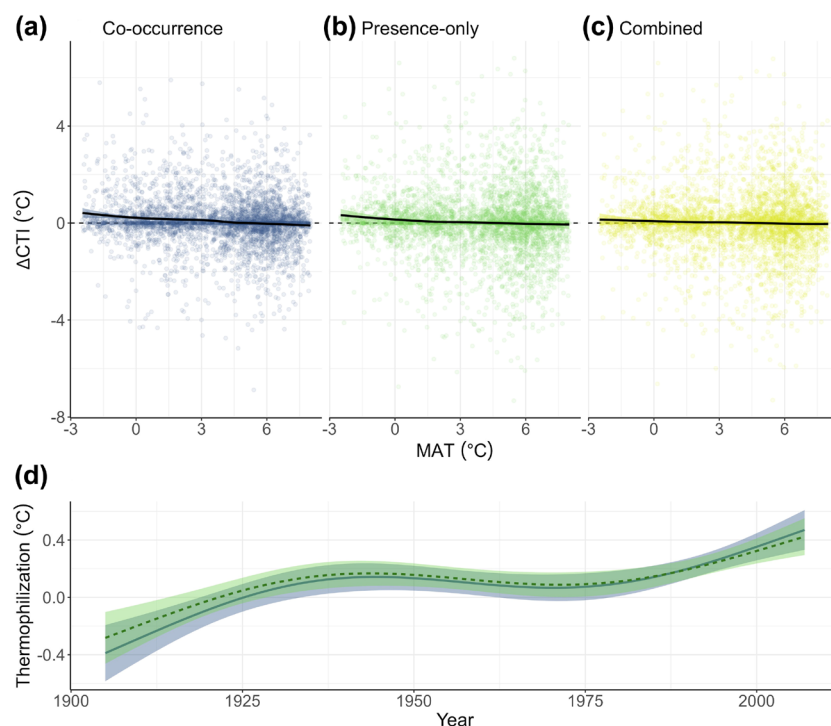


Figure 2. Comparison of the community temperature index (CTI) of the paired dataset (ΔCTI , 3839 pairs) for the different calibrated models (panels (a), (b) and (c)). In the paired dataset, co-occurrence plots were matched with presence-only by subsampling the latter and selecting the closest available pseudo-plot that had been sampled. x-axis: mean annual temperature (MAT) of the average of the 5 years preceding the respective assemblage survey of the co-occurrence dataset. y-axis: ΔCTI is the differences between the paired plots as follows: $\text{CTI}_{\text{pseudo-plots}} - \text{CTI}_{\text{co-occurrence plots}}$. A local nonparametric regression (LOESS) regression line is shown in each panel. (d) Temporal thermophilization trend. Thermophilization is estimated using the projection of the combined calibrated model using the paired dataset. Blue solid line: co-occurrence dataset; green dashed line: presence-only dataset.

co-occurrence plots. To evaluate this, we assessed elevation differences (Δ Elevation) between the co-occurrence plots and their correspondingly paired pseudo-plots. To obtain elevational data, we initially extracted elevation values from the digital elevation model (DEM) with a resolution of 25 m (Copernicus 2016). For the co-occurrence plots, we used the provided coordinates. However, for the pseudo-plots we first extracted the elevation from the DEM for each species' records within each pseudo-plot. Subsequently, we averaged them to get the elevation of the pseudo-plot. Delta elevation was then estimated by subtracting the average elevation for the presence-only pseudo-plots from the elevation of co-occurrence plots. Hence, negative delta elevation values indicate a lower elevation in the presence-only pseudo-plots compared to the co-occurrence plots, while positive values suggest a higher elevation in the presence-only pseudo-plots.

Our findings indicated a general tendency of the presence-only pseudo-plots being situated at lower elevations compared to the co-occurrence plots. Specifically, observations in the presence-only pseudo-plots were consistently found at lower elevations, showing a mean difference of -35.3 m with a SE of 1.41 (ANOVA F-value = 5.082 , p-value < 0.05 ; Supporting information). Additionally, we observed that larger differences in elevations corresponded to greater variations in CTI values. To quantify this trend, linear models were applied. For the co-occurrence calibration model, the slope was $-2.15 \times 10^{-4} \text{ } ^\circ\text{C m}^{-1}$ (SE = 2×10^{-4} , $p = 0.467$), while for the presence-only calibration model, the slope was $-1.18 \times 10^{-3} \text{ } ^\circ\text{C m}^{-1}$ (SE = 3.8×10^{-4} , $p = 0.003$). The combined model showed a slope of $-1.17 \times 10^{-3} \text{ } ^\circ\text{C m}^{-1}$ (SE = 3.7×10^{-4} , $p = 0.003$) (Fig. 3).

Discussion

Our assessment demonstrated the potential of employing presence-only data for evaluating community responses to

climate warming, specifically by quantifying the CTI and thermophilization. The aggregation of presence-only data in pseudo-plots yields CTI values remarkably consistent with those obtained by co-occurrence plots, which are conventionally perceived as more reliable (Bertrand et al. 2011, Pacheco-Riaño et al. 2023). Presence-only data typically originate from museum collections and citizen science projects. As such, they often suffer from different spatial and taxonomical biases, e.g. variable sampling density per area and low-quality control of species identification (Beck et al. 2014). Despite these challenges, especially prevalent in historical data, these data are crucial since they are sometimes the only way to understand communities in the past (Vellend et al. 2013). Although presence-only data inherently carry biases and identification errors, our study reveals a consistent estimate of the CTI for the paired plots and a consistent temporal trend in thermophilization from the two datasets. This consistency can be explained by the general well-distributed presence of species in the pseudo-plots within the thermal range of the co-occurrence plots (except in the cold end, discussion below). In simpler terms, there is no tendency to record more cold-adapted or warm-adapted species in presence-only data compared to co-occurrence data. Even if species in an area are under-sampled in the presence-only pseudo-plots, the likelihood of observing a species, regardless of its temperature indicator value, remains constant. Likewise for the misidentification of species, there is no relationship with consistently higher or lower thermal indicator values. In addition, even if some species were misidentified in the pseudo-plots, we expect that this would have a minor impact, given the substantial number of correct identifications that keep the overall consistency in CTI values between both datasets.

It is important to mention that both co-occurrence plots and presence-only data plots have spatial and temporal biases, and in the present study we replicate the spatial and temporal bias of the co-occurrence data in the presence-only data. For an accurate representation of the thermophilization of an area

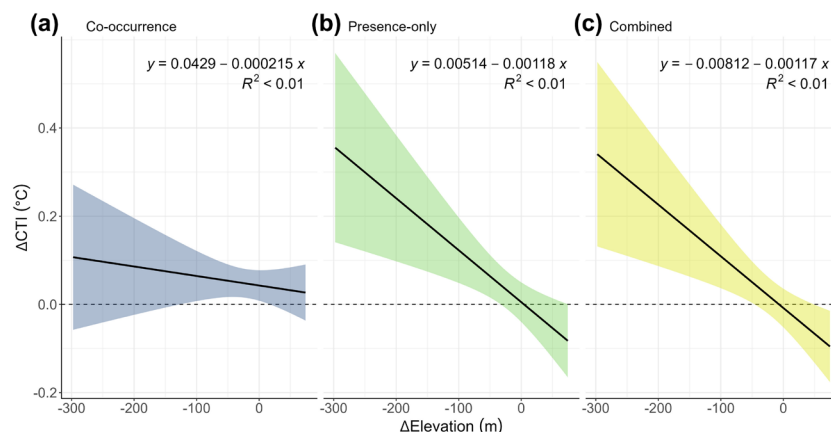


Figure 3. Relationship between community temperature index (CTI) of the paired dataset (Δ CTI) ($\text{CTI}_{\text{Pseudo-plots}} - \text{CTI}_{\text{co-occurrence plots}}$) and Δ Elevation (average elevation for specimens of presence-only pseudo-plot–elevation co-occurrence plots) in different calibrated models using the paired dataset. In the paired dataset, co-occurrence plots were matched with pseudo-plots by subsampling the latter and selecting the closest available pseudo-plot that had been sampled. Here we used the 95th percentile of the paired dataset.

it is imperative to thoroughly examine and account for these biases when studying community responses. Therefore, the thermophilization trend observed in this study (Fig. 2) does not accurately reflect the actual thermophilization for the area (for an unbiased accurate representation, see Pacheco-Riaño et al. 2023).

By including presence-only data in CTI analysis, we are able to cover larger geographic areas than analyses based on traditional co-occurrence plots alone (König et al. 2019). This can be useful towards global completeness and representativeness of species data and producing more realistic projections of the community's responses. Larger temporal and spatial coverage also provides improved opportunities to unravel the impacts of various climatic drivers and how climate interacts with other variables. This becomes particularly valuable in situations where co-occurrence surveys are restricted, either in terms of spatial or temporal coverage. The inclusion of presence-only data also provides a cost-effective way of monitoring ongoing dynamics in species communities, an alternative to more intensive co-occurrence surveys, which are often considered information of the highest quality but can be time-consuming and expensive (Dengler et al. 2011). This study was carried out in Norway where both good-quality co-occurrence plots and presence-only records have been sampled extensively over the last 100 years, making the comparison between the two data types possible. Although biodiversity data, including presence-only data, have spatial biases and some regions are undersampled (Vellend et al. 2013, Hughes et al. 2021), the demonstrated usefulness of presence-only data for assessing responses to climate changes opens up the possibility for conducting such studies in a much larger area than has been done until today. Nonetheless, certain regions, such as the tropics, or continents like Africa, will still be under-represented compared to temperate regions. Despite these limitations, presence-only data could be the only source of information for these areas at times (Feeley 2012). Although the CTI values were very similar, we observed that the differences in predicted CTI values between the two datasets (D CTI) were more pronounced towards the colder end of the thermal gradient. This deviation could potentially be attributed to the distinct methodologies employed in data collection. Co-occurrence data are often gathered through more or less organized expeditions and may as such cover a more representable distribution of the topographic relief in an area, including higher and more remote areas. The presence-only data, however, are often compiled from more random observations made in unplanned citizen science projects. In colder areas, which in most cases will entail mountain regions in our study area, the more accessible parts are in the valley bottoms (where roads are placed), resulting in a bias towards lower elevations in topographically heterogeneous areas. Our ad hoc analysis substantiated this expectation (Fig. 3), revealing that co-occurrence plots are predominantly situated at higher elevations in areas where CTI from pseudo-plots is overestimated compared to the co-occurrence plots; despite these differences in elevation, the CTI remains similar unless the differences in elevation are

substantial. Being aware of this, it would be possible to adjust for this potential bias by also incorporating elevation when aggregating species to pseudo-plots, or by adjusting for the bias in CTI by using the observed elevations of the presence-only data. This would be especially important when comparing pseudo-plots, or when combining co-occurrence plots and presence-only data in areas of different topographical relief.

Overall, we noticed that the variations among the CTIs are comparable when calibrating the transfer functions using all three types of datasets. However, the difference observed in the model relying only on the co-occurrence dataset can be attributed to the extent of the dataset employed for calibrating the model. We found that including plant communities from a larger area improved the transfer function, notably when including communities with thermophilic species from lower latitudes (e.g. from central Europe). This inclusion effectively addresses the issue of niche truncation at the warmer end of the thermal range. However, including presence-only data from outside our focus area did not improve the accuracy of the CTI values at the colder end; this was likely due to the lack of species in the dataset from higher latitudinal areas or elevations. This consideration is of particular significance given the ongoing global warming trend, as adjusting the overestimation of the cold end would be less necessary as there will be a higher representation of thermophilic species in the communities. However, giving special attention to species adapted to warmer conditions holds significant relevance in Norway, given its predominantly cold climate. In contrast, in locations with milder temperatures, especially those not situated at high latitudes or elevations, relying on presence-only data would help mitigate the truncation of species at both the cold and warm extremes. Moreover, we saw that the differences were larger for older assemblages than for more recent ones, which could be attributed to the improved accuracy of the newer information.

It is important to note that the accuracy of the CTI values produced by pseudo-plots will depend on the quality of the presence-only data being used. The exponential growth of presence-only data records in the last two decades has resulted in greater public access to these records (Jin and Yang 2020). However, before using these data for ecological analyses, cleaning and standardization procedures must be undertaken. This step is especially important due to the varying sources and properties of the data, as temporal, spatial, and taxonomic criteria all need to be considered (Vellend et al. 2013, Meyer et al. 2016).

Integrating presence-only data into species distribution modelling has been widely used (Beck et al. 2014, Pacifici et al. 2017, Smith et al. 2023). However, very few studies explore their use in understanding community responses, such as thermophilization, to environmental changes (Feeley 2012). Our study demonstrates that, by including presence-only data, we can better understand and learn from their advantages and drawbacks in biogeographical analyses. The explicit combination of this type of data can be a powerful approach for understanding the effects of multiple drivers

contributing to climate change. Although the outcomes produced by this method may have some flaws, they could be the first approximation for many regions and taxonomical groups, and provide a good starting point for further research. As demonstrated here, open data portals, such as GBIF, can be utilized to consolidate datasets that are used to analyse communities' responses to environmental change. To bridge existing data gaps, the digitization and mobilization of scientific biological collections and personal archives of researchers must be continued. Considering the valuable insights offered by presence-only data that we showed, and the urgent necessity to comprehend global biological changes, prioritizing the use of such data sources becomes crucial in research. The significance of both presence-only data and co-occurrence data emphasizes the importance of recording and digitalizing biological data, including herbarium specimens, human observations, and vegetation surveys. This will help to monitor species composition, abundance, and diversity changes and identify potential threats to ecosystem functioning, leading to a better understanding of the environmental impact of global climate change.

Our study compellingly indicates that presence-only data can be used to estimate community indicators (e.g. CTI) accurately. The data serve as an additional source of information to complement more traditional co-occurrence plot-based datasets. Our findings suggest that the data integration of presence-only data can be used to improve the calibration of transfer function models and our understanding of vegetation responses to climate change. Nevertheless, the overall patterns and trends of thermophilization remain largely consistent across the two datasets, suggesting that the thermophilization values are not significantly different when using pseudo-plots. Even more important is the fact that both datasets show a consistent trend in thermophilization, independent of the calibration dataset. We additionally presented an outline that can be used to study community responses for global change research. Our main findings, therefore, demonstrate that presence-only data can be used to quantify thermophilization. Though some careful attention is needed when integrating traditional co-occurrence plots with presence-only data (our ad hoc analyses, particularly, highlight the biases related to elevation for this purpose), there is a substantial potential to unlock new opportunities for rapid and cost-effective monitoring of communities in response to changes in climate.

Acknowledgements – We are immensely grateful to all data contributors for their invaluable efforts in collecting, maintaining, and sharing species occurrence data on GBIF. We would also like to extend our thanks to Sophie Weides for her invaluable comments on previous versions of the manuscript.

Funding – SGAF is grateful for the generous support from the Trond Mohn Stiftelse (TMS) and the University of Bergen through startup grant no. 'TMS2022STG03.' SGAF and J-AG would like to acknowledge the financial assistance provided by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program grant agreement

no. 741413, which is attributed to HJB Birks. LCP-R received partial support from the Olaf Grolle Olsen's Scholarship grant no. 20019/53/FOL. TM was supported by the University of Oulu and the Research Council of Finland within the framework of Profi6, grant no. 336449.

Author contributions

L. Camila Pacheco Riaño: Conceptualization (equal); Data curation (lead); Formal analysis (equal); Methodology (equal); Validation (equal); Visualization (lead); Writing – original draft (lead). **Sabine Rumpf:** Methodology (supporting), Writing – review and editing (equal). **Tuija Maliniemi:** Writing – review and editing (equal). **Suzette G. A. Flantua:** Supervision (equal); Writing – original draft (equal); Writing – review and editing (equal). **John-Arvid Grytnes:** Conceptualization (equal); Methodology (equal); Supervision (lead); Writing – original draft (equal); Writing – review and editing (equal).

Transparent peer review

The peer review history for this article is available at <https://publons.com/publon/10.1111/ecog.07213>.

Data availability statement

The data used in this study were obtained from GBIF at the following DOIs: presence-only dataset: <https://doi.org/10.15468/DL.VZVGK7>; University of Agder (2021) Vascular plant herbarium (KMN) UiA. Vers. 1.1804. Occurrence dataset, <https://doi.org/10.15468/2g6i0v>; University of Oslo <https://doi.org/10.15468/w8gru5>, <https://doi.org/10.15468/tvnjk7>; Norwegian University of Science and Technology <https://doi.org/10.15468/kkb2x0>; and Norwegian University of Life Sciences (NMBU) <https://doi.org/10.15468/mbhmt>. The climate data used are available from <http://chelsa-climate.org/chelsaocruts/>.

The R scripts required to reproduce the results are available from the Zenodo Digital Repository: <https://doi.org/10.5281/zenodo.10708517> (Pacheco-Riaño et al. 2024).

Supporting information

The Supporting information associated with this article is available with the online version.

References

- Bayraktarov, E., Ehmke, G., O'Connor, J., Burns, E. L., Nguyen, H. A., McRae, L., Possingham, H. P. and Lindenmayer, D. B. 2019. Do big unstructured biodiversity data mean more knowledge? – *Front. Ecol. Evol.* 6: 239.
- Beck, J., Böller, M., Erhardt, A. and Schwanghart, W. 2014. Spatial bias in the GBIF database and its effect on modeling species' geographic distributions. – *Ecol. Inform.* 19: 10–15.
- Bertrand, R., Lenoir, J., Piedallu, C., Dillon, G. R., De Ruffray, P., Vidal, C., Pierrat, J. C. and Gégout, J. C. 2011. Changes in

- plant community composition lag behind climate warming in lowland forests. – *Nature* 479: 517–520.
- Bhatta, K. P., Grytnes, J. A. and Vetaas, O. R. 2018. Downhill shift of alpine plant assemblages under contemporary climate and land-use changes. – *Ecosphere* 9: e02084.
- Bhatta, K. P., Birks, H. J. B., Grytnes, J.-A. and Vetaas, O. R. 2019. Corrigendum: weighted average regression and environmental calibration as a tool for quantifying climate-driven changes in vegetation. – *J. Plant Ecol.* 12: 787–789.
- Bottin, M., Peyre, G., Vargas, C., Raz, L., Richardson, J. E. and Sanchez, A. 2020. Phytosociological data and herbarium collections show congruent large-scale patterns but differ in their local descriptions of community composition. – *J. Veg. Sci.* 31: 208–219.
- Chevalier, M. et al. 2020. Pollen-based climate reconstruction techniques for late Quaternary studies. – *Earth Sci. Rev.* 210: 103384.
- Chytrý, M., Tichý, L., Hennekens, S. M. and Schaminée, J. H. J. 2014. Assessing vegetation change using vegetation-plot databases: a risky business. – *Appl. Veg. Sci.* 17: 32–41.
- Copernicus 2016. European digital elevation model EU-DEM v1.1. – <https://land.copernicus.eu/imagery-in-situ/eu-dem/eu-dem-v1.1?tab=metadata>
- De Frenne, P. et al. 2013. Microclimate moderates plant responses to macroclimate warming. – *Proc. Natl Acad. Sci. USA* 110: 18561–18565.
- Dengler, J., Jansen, F., Glöckler, F., Peet, R. K., de Cáceres, M., Chytrý, M., Ewald, J., Oldeland, J., Lopez-Gonzalez, G., Finckh, M., Mucina, L., Rodwell, J. S., Schaminée, J. H. J. and Spencer, N. 2011. The Global Index of Vegetation-Plot Databases (GIVD): a new resource for vegetation science. – *J. Veg. Sci.* 22: 582–597.
- Devictor, V., Julliard, R., Couvet, D. and Jiguet, F. 2008. Birds are tracking climate warming, but not fast enough. – *ProcRSoc. B* 275: 2743–2748.
- Devictor, V. et al. 2012. Differences in the climatic debts of birds and butterflies at a continental scale. – *Nat. Clim. Change* 2: 121–124.
- Duchenne, F., Martin, G. and Porcher, E. 2021. European plants lagging behind climate change pay a climatic debt in the north, but are favoured in the south. – *Ecol. Lett.* 24: 1178–1186.
- Duque, A., Stevenson, P. R. and Feeley, K. J. 2015. Thermophilization of adult and juvenile tree communities in the northern tropical Andes. – *Proc. Natl Acad. Sci. USA* 112: 10744–10749.
- Fadrique, B. et al. 2018. Widespread but heterogeneous responses of Andean forests to climate change. – *Nature* 564: 207–212.
- Feeley, K. J. 2012. Distributional migrations, expansions, and contractions of tropical plant species as revealed in dated herbarium records. – *Global Change Biol.* 18: 1335–1341.
- Feeley, K., Martinez-Villa, J., Perez, T., Silva Duque, A., Triviño Gonzalez, D. and Duque, A. 2020. The thermal tolerances, distributions, and performances of tropical montane tree species. – *Front. For. Global Change* 3: 25.
- Felde, V. A., Kapfer, J. and Grytnes, J. A. 2012. Upward shift in elevational plant species ranges in Sikkildalen, central Norway. – *Ecography* 35: 922–932.
- Freeman, B. G., Song, Y., Feeley, K. J. and Zhu, K. 2021. Montane species track rising temperatures better in the tropics than in the temperate zone. – *Ecol. Lett.* 24: 1697–1708.
- GBIF.org. 2020. Species matching. – <https://www.gbif.org/tools/species-lookup>.
- GBIF.org. 2022. Occurrence download. – The Global Biodiversity Information Facility, <https://doi.org/10.15468/DL.VZVGK7>.
- Gottfried, M. et al. 2012. Continent-wide response of mountain vegetation to climate change. – *Nat. Clim. Change* 2: 111–115.
- Götzenberger, L., de Bello, F., Bräthen, K. A., Davison, J., Dubuis, A., Guisan, A., Lepš, J., Lindborg, R., Moora, M., Pärtel, M., Pellissier, L., Pottier, J., Vittoz, P., Zobel, K. and Zobel, M. 2012. Ecological assembly rules in plant communities: approaches, patterns and prospects. – *Biol. Rev.* 87: 111–127.
- Guiot, J. and de Vernal, A. 2007. Transfer functions: methods for quantitative paleoceanography based on microfossils. Chapter 13. – In: Hillaire-Marcel, C. and Vernal, A. D. (eds), *Proxies in late cenozoic paleoceanography*, Vol. 1. Elsevier, pp. 523–563.
- Haider, S. et al. 2022. Think globally, measure locally: the MIREN standardized protocol for monitoring plant species distributions along elevation gradients. – *Ecol. Evol.* 12: e8590.
- Heberling, J. M., Miller, J. T., Noesgaard, D., Weingart, S. B. and Schigel, D. 2021. Data integration enables global biodiversity synthesis. – *Proc. Natl Acad. Sci. USA* 118: e2018093118.
- Hijmans, R. J. 2023. raster: geographic data analysis and modeling. – Rpackage ver. 3.6-13, <https://cran.r-project.org/package=raster>.
- Hughes, A. C., Orr, M. C., Ma, K., Costello, M. J., Waller, J., Provoost, P., Yang, Q., Zhu, C. and Qiao, H. 2021. Sampling biases shape our view of the natural world. – *Ecography* 44: 1259–1269.
- Hutchinson, G. E. 1957. Population studies: animal ecology and demography. – *Bull. Math. Biol.* 22: 415–457.
- Jin, J. and Yang, J. 2020. BDcleaner: a workflow for cleaning taxonomic and geographic errors in occurrence data archived in biodiversity databases. – *Global Ecol. Conserv.* 21: e00852.
- Juggins, S. and Birks, H. J. B. 2012. Quantitative environmental reconstructions from biological data. – In: Birks, H. J. B., Lotter, A. F., Juggins, S. and Smol, J. P. (eds), *Tracking environmental change using lake sediments. Developments in paleoenvironmental research*. Springer, pp. 431–494.
- Kapfer, J., Hédli, R., Jurasinski, G., Kopecký, M., Schei, F. H. and Grytnes, J. A. 2017. Resurveying historical vegetation data – opportunities and challenges. – *Appl. Veg. Sci.* 20: 164–171.
- Karger, D. N. and Zimmermann, N. E. 2018. CHELSAcruts - high resolution temperature and precipitation timeseries for the 20th century and beyond. – *EnviDat*, <https://www.doi.org/10.16904/enviDat.159>.
- Karger, D. N., Conrad, O., Böhner, J., Kawohl, T., Kreft, H., Soria-Auza, R. W., Zimmermann, N. E., Linder, H. P. and Kessler, M. 2017. Climatologies at high resolution for the earth's land surface areas. – *Sci. Data* 4: 170122.
- König, C., Weigelt, P., Schrader, J., Taylor, A., Kattge, J. and Kreft, H. 2019. Biodiversity data integration—the significance of data resolution and domain. – *PLoS Biol.* 17: e3000183.
- Lajeunesse, A. and Fourcade, Y. 2023. Temporal analysis of GBIF data reveals the restructuring of communities following climate change. – *J. Anim. Ecol.* 92: 391–402.
- Lenoir, J., Gégout, J. C., Marquet, P. A., de Ruffray, P. and Brisse, H. 2008. A significant upward shift in plant species optimum elevation during the 20th century. – *Science* 320: 1768–1771.
- Lenoir, J., Bertrand, R., Comte, L., Bourgeaud, L., Hattab, T., Muriene, J. and Grenouillet, G. 2020. Species better track climate warming in the oceans than on land. – *Nat. Ecol. Evol.* 4: 1044–1059.
- Liu, M., Prentice, I. C., Ter Braak, C. J. F. and Harrison, S. P. 2020. An improved statistical approach for reconstructing past climates from biotic assemblages: improved palaeoclimate recon-

- struction. – *Proc. R. Soc. A* 476: 47620200346.20200346, <http://doi.org/10.1098/rspa.2020.0346>.
- Lu, F., Ma, C., Zhu, C., Lu, H., Zhang, X., Huang, K., Guo, T., Li, K., Li, L., Li, B. and Zhang, W. 2019. Variability of East Asian summer monsoon precipitation during the Holocene and possible forcing mechanisms. – *Clim. Dyn.* 52: 969–989.
- Mamantov, M. A., Gibson-Reinemer, D. K., Linck, E. B. and Sheldon, K. S. 2021. Climate-driven range shifts of montane species vary with elevation. – *Global Ecol. Biogeogr.* 30: 784–794.
- Menéndez, R., Megías, A. G., Hill, J. K., Braschler, B., Willis, S. G., Collingham, Y., Fox, R., Roy, D. B. and Thomas, C. D. 2006. Species richness changes lag behind climate change. – *Proc. R. Soc. B* 273: 1465–1470.
- Meyer, C., Weigelt, P. and Kreft, H. 2016. Multidimensional biases, gaps and uncertainties in global plant occurrence information. – *Ecol. Lett.* 19: 992–1006.
- O’Sullivan, K. S. W., Ruiz-Benito, P., Chen, J. C. and Jump, A. S. 2021. Onward but not always upward: individualistic elevational shifts of tree species in subtropical montane forests. – *Ecography* 44: 112–123.
- Pacheco-Riaño, L. C., Høistad Schei, F., Flantua, S. G. A. and Grytnes, J.-A. 2023. Lags in the response of plant assemblages to global warming depends on temperature-change velocity. – *Global Ecol. Biogeogr.* 32: 719–733.
- Pacheco-Riaño, L. C., Rumpf, S., Maliniemi, T., Flantua, S. G. A. and Grytnes, J. A. 2024. Data from: Reliability of presence-only data for assessing plant community responses to climate warming. – Zenodo Digital Repository, <https://doi.org/10.5281/zenodo.10708517>.
- Pacifici, K., Reich, B. J., Miller, D. A. W., Gardner, B., Stauffer, G., Singh, S., McKerrow, A. and Collazo, J. A. 2017. Integrating multiple data sources in species distribution modeling: a framework for data fusion. – *Ecology* 98: 840–850.
- Pebesma, E. 2018. Simple features for R: standardized support for spatial vector data. – *R J.* 10: 439–446.
- Richard, B. et al. 2021. The climatic debt is growing in the understory of temperate forests: stand characteristics matter. – *Global Ecol. Biogeogr.* 30: 1474–1487.
- Rull, V. 2010. Ecology and palaeoecology: two approaches, one objective. – *Open Ecol. J.* 3: 1–5.
- Rumpf, S. B., Hülber, K., Klöner, G., Moser, D., Schütz, M., Wessely, J., Willner, W., Zimmermann, N. E. and Dullinger, S. 2018. Range dynamics of mountain plants decrease with elevation. – *Proc. Natl Acad. Sci. USA* 115: 1848–1853.
- Salonen, J. S., Ilvonen, L., Seppä, H., Holmström, L., Telford, R. J., Gaidamavičius, A., Stančikaitė, M. and Subetto, D. 2011. Comparing different calibration methods (WA/WA-PLS regression and Bayesian modelling) and different-sized calibration sets in pollen-based quantitative climate reconstruction. – *Holocene* 22: 413–424.
- Schäbitz, F., Wille, M., Francois, J.-P., Haberzettl, T., Quintana, F., Mayr, C., Lücke, A., Ohlendorf, C., Mancini, V., Paez, M. M., Prieto, A. R. and Zolitschka, B. 2013. Reconstruction of palaeoprecipitation based on pollen transfer functions – the record of the last 16 ka from Laguna Potrok Aike, southern Patagonia. – *Quat. Sci. Rev.* 71: 175–190.
- Smith, A. B., Murphy, S. J., Henderson, D. and Erickson, K. D. 2023. Including imprecisely georeferenced specimens improves accuracy of species distribution models and estimates of niche breadth. – *Global Ecol. Biogeogr.* 32: 342–355.
- Steinbauer, M. J. et al. 2018. Accelerated increase in plant species richness on mountain summits is linked to warming. – *Nature* 556: 231–234.
- ter Braak, C. J. F. and Juggins, S. 1993. Weighted averaging partial least squares regression (WA-PLS): an improved method for reconstructing environmental variables from species assemblages. – *Hydrobiologia* 269–270: 485–502.
- Vellend, M., Brown, C. D., Kharouba, H. M., McCune, J. L. and Myers-Smith, I. H. 2013. Historical ecology: using unconventional data sources to test for effects of global environmental change. – *Am. J. Bot.* 100: 1294–1305.
- Walther, G. R. 2010. Community and ecosystem responses to recent climate change. – *Phil. Trans. R. Soc. B* 365: 2019–2024.
- Wickham, H. et al. 2019. Welcome to the Tidyverse. – *J. Open Source Softw.* 4: 1686.
- Wüest, R. O., Zimmermann, N. E., Zurell, D., Alexander, J. M., Fritz, S. A., Hof, C., Kreft, H., Normand, S., Cabral, J. S., Szekely, E., Thuiller, W., Wikelski, M. and Karger, D. N. 2020. Macroecology in the age of Big Data – where to go from here? – *J. Biogeogr.* 47: 1–12.
- Zellweger, F. et al. 2020. Forest microclimate dynamics drive plant responses to warming. – *Science* 775: 772–775.
- Zizka, A., Silvestro, D., Andermann, T., Azevedo, J., Duarte Ritter, C., Edler, D., Farooq, H., Herdean, A., Ariza, M., Scharn, R., Svantesson, S., Wengström, N., Zizka, V. and Antonelli, A. 2019. CoordinateCleaner: standardized cleaning of occurrence records from biological collection databases. – *Methods Ecol. Evol.* 10: 744–751.