Controlled comparison of species- and community-level models across novel climates and communities

Kaitlin C. Maguire1,†, Diego Nieto-Lugilde2,†, Jessica L. Blois1, Matthew C. Fitzpatrick2, John W. Williams3, Simon Ferrier4 and David J. Lorenz3

1School of Natural Sciences, University of California—Merced, Merced, CA, USA
2Appalachian Laboratory, University of Maryland Center for Environmental Science, Frostburg, MD 21532, USA
3Center for Climatic Research, University of Wisconsin—Madison, Madison, WI, USA
4CSIRO Land and Water Flagship, Canberra, ACT, Australia

†These authors contributed equally to this study.

Received: 23 November 2015
Accepted: 16 February 2016

1. Introduction

Species distribution models (SDMs) remain among the most widely used methods for forecasting regional- to global-scale changes in species distributions, species assemblages and patterns of biodiversity in response to climate change [1]. When modelling entities or attributes above the species level using SDMs, a ‘predict first, assemble later’ approach [2] is typically employed, wherein individual models are fitted and projected for each species, the mapped predictions of which are then aggregated or ‘stacked’ to infer potential changes in community-level patterns such as species richness (e.g. [3]). An alternative, but relatively rarely used, approach involves combining data from multiple species to simultaneously analyse and map patterns of biodiversity at the community level. This ‘assemble and predict together’ strategy [2] underlies community-level models (CLMs) that fit a species co-occurrence matrix to environmental variables to predict both community structure and the distributions of individual species [2]. CLMs implicitly capture any process driving co-occurrence patterns, including shared climatic requirements of species, responses to unmeasured environmental variables and possibly biotic interactions [4]. SDMs implicitly also capture these processes (by setting the boundaries of a species’ realized niche), but do so in a manner that is isolated from the broader community context in which species occur (a Gleasonian approach).

By simultaneously modelling all observed species within a region of interest and incorporating information on co-occurrence, CLMs may have the potential to predict species distributions and changes in community composition better than SDMs [2], especially for large climatic shifts and novel climate regimes where individual taxon–climate relationships may break down [5], but broader...
biodiversity–climate relationships may remain comparatively stable [6]. The potentially greater transferability of CLMs to novel climates may be beneficial given the predicted emergence of no-analogue climates in the near future [7,8]. However, these ideas remain untested, and the relative ability of SDMs and CLMs to simulate the past emergence of no-analogue communities is unknown. Ferrier & Guisan [2] suggested that CLMs may effectively balance the assumptions underlying predictions of community types (that species move together as fixed community types in response to climate change—a Clementsian approach) versus the assumptions underlying predictions of individual species distributions (that the environmental variables shaping distributions may differ markedly between species) by modelling multiple species in relation to a shared set of environmental variables and by implicitly accounting for other drivers of co-occurrence [9]. The palaeo-record offers evidence that neither of these assumptions is fully valid [10,11]. CLMs may implicitly account for any factors that drive patterns of co-occurrence (including biotic interactions) and, therefore, could be advantageous given the well-documented role of biotic interactions in mediating responses of species and communities to climate change [12] and associated reshuffling of communities during current and past environmental changes [13,14]. Another major advantage of CLMs is an ability to model very large numbers of species, including rare (or poorly sampled) species [15,16]. By detecting shared environmental responses across species, CLMs can ‘borrow strength’ from more common (or better sampled) species and potentially produce better predictions than possible from SDMs [2,17–19]. Thus, for predicting species distributions and assemblages, CLMs allow fuller consideration of species that might otherwise be excluded from SDM-based analyses owing to an insufficiency of observations.

The extent to which the potential benefits of CLMs translate into improved predictions remains unknown. Most CLMs are multiresponse (i.e. multispecies) extensions of commonly used SDMs (e.g. constrained quadratic ordination (CQO) is a multiresponse version of generalized linear models (GLM)), while other CLMs have no direct SDM counterpart (e.g. generalized dissimilarity modelling). Previous studies that compared the performance of SDMs and CLMs when predicting species distributions, community composition and/or species richness have had mixed results, with some studies finding that CLMs tend to outperform SDMs [20–22], and others finding the opposite [4,15,23,24]. One only of these analyses directly compared CLM and SDM predictions against observed changes in species assemblages [15], and this analysis only compared one CLM with one SDM. The analysis presented here is the first to compare model predictions with observed species assemblages using head-to-head evaluations of multiple SDM algorithms and their direct CLM counterparts. In addition, previous analyses have not tested the transferability of CLMs to predict species distributions and community composition across longer time periods (more than 25 years), across periods of climate change similar in magnitude to that expected for this century, and in no-analogue climates.

Here, we used fossil pollen from sediment cores in eastern North America spanning the past 21 000 years (figure 1) to perform the first comprehensive SDM and CLM model comparison across the large and rapid climate changes of the Late Quaternary in order to tease apart the differences in SDMs and CLMs and to evaluate how these models may perform in predicting species distributions and assemblages under climate change. The palaeontological record, in general, and the fossil-pollen record of the Late Quaternary, in particular, provides a unique opportunity to test and compare ecological forecasting models across periods of climatic and biotic change [8,27]. We tested the ability of five SDM algorithms and their direct CLM counterparts to predict observed changes in fossil-pollen distributions and assemblages through time and test whether simultaneously modelling co-occurring taxa in a region increased model performance compared with SDMs. By comparing SDMs with their direct CLM counterparts, we controlled for differences in predictive performance among different classes of algorithms. We assessed how model performance changes as a function of the climatic and compositional novelty between time periods, in order to estimate the limits of model predictability and thereby provide guidelines for future projections. Lastly, we tested the hypothesis that CLMs have a higher predictive skill for rare taxa.

Figure 1. Temperature change and model comparison scheme. Solid line shows oxygen isotope data from 21 000 years ago (kyr BP) to the present from the North Greenland Ice Core Project (GICC05) [25] with the 2005 Greenland Ice Core Chronology age model [26]. Dashed line is mean maximum temperature from climate simulations used in this analysis. The Bølling–Allerød (B-A) and Younger Dryas (YD) Chronozones are highlighted in grey. The time periods used to fit SDM and CLM models are indicated with shaded boxes/rectangles, while projected time periods are indicated with open ellipses. This is illustrated for the 10 kyr BP time period (shaded box with thickened outline and grey arrows indicating projected time periods). Data were pooled for 15 to 14 kyr BP, 17.5 to 15.5 kyr BP and 21 to 15.5 kyr BP when building models (shaded rectangles) and for 17.5 to 15.5 kyr BP, 21 to 15.5 and 21 to 18 kyr BP when projecting models (open rounded rectangles). The number of localities used to fit models is indicated above the grey box representing the model fitting time period.
2. Material and methods

(a) Setting

Since the Last Glacial Maximum (LGM—21 kyr BP (thousands of years before present)), global mean temperature increased from 3°C to 5°C, with periods of rapid (decadal-to-centennial-scale) warming and cooling (e.g. the Bølling–Allerød and Younger Dryas onset and terminations; figure 1). Plant distributions [13] and realized niches [11,28] shifted in response to these climate changes, resulting in a reshuffling of taxon associations [13–14,29,30]. These climatic and ecological changes in the past were characterized by periods of climatic novelty and no-analogue communities as compared to contemporary climates and communities. The last deglaciation (21 kyr BP to present) is not an exact analogue for periods of climatic novelty and no-analogue communities as resulting in a reshuffling of taxon associations 

(b) Taxonomic occurrences

Taxon occurrence data were from the eastern North America fossil-pollen relative abundance dataset from Blois et al. [30], with the following modifications: (i) we removed sites that had marine/estuarine or unknown depositional environments, (ii) pollen abundances were expressed as the pollen sum for a particular taxon divided by the total sum for all genus-level taxa, rather than divided by the total upland sum for the site (which includes both genera as well as families and other higher level taxa), and (iii) pollen relative abundances were interpolated to 500-year time slices from 21,000 to present, rather than 1000-year time slices. Pollen data quality was calculated as in Blois et al. [30], and for each time period, only sites with a weighted quality value above 0.75 were included. If multiple sites fell within the same grid cell, their pollen abundances were averaged. Taxa co-occurrences were within a grid cell, with the majority of grid cells containing co-occurrences from within a single lake sediment core. Following Nieto-Lugilde et al. [31], the relative abundance pollen matrix was converted to a presence/absence matrix after applying a threshold scaled to 5% of the maximum abundance (see electronic supplementary material, table S1 and figure S1). Because the majority of fossil-pollen types considered here can be consistently identified, absences are considered true absences. We chose the 19 most abundant-through-time taxa at the generic level (see electronic supplementary material), the lowest taxonomic level in which most fossil pollen can be identified—except for Ostrya and Carpinus (Ostrya/Carpinus), which are palynologically indistinguishable, and Ambrosia-type, which can include Issa. Even though interspecific differences in climatic and ecological tolerances may differ from genus-level tolerances, the same problem exists, although to a lesser extent, for populations within species (e.g. [32]). In addition, pollen records best match forest inventory data at the genus level [31]; thus, analyses at the genus level are as appropriate as analyses at the species level, especially for pollen data, but may introduce another source of uncertainty. See Blois et al. [30] for a complete list of pollen types assigned to each genus.

(c) Environmental variables

We used palaeoclimate simulations from the Community Climate System Model Version 3 (CCSM3) SynTrace transient simulation with seasonally averaged model outputs saved at a decadal time step from 21 kyr BP to the present [33]. Climate variables were de-biased and downscaled to a 0.5° × 0.5° grid (approx. 50 × 50 km) and processed to create yearly, quarterly and monthly variables for every 500 years since 21 kyr BP. All climate variables are means across 200 years, centered on 500 year time slices (e.g. the 1.5 kyr BP climate variables are averaged across the years 1.6–1.4 kyr BP). From the original 27 variables (nine yearly, quarterly and monthly), we chose six uncorrelated variables (maximum temperature of the warmest quarter, minimum precipitation of the driest quarter, maximum precipitation of the wettest quarter, mean yearly potential evapotranspiration, mean yearly actual evapotranspiration and mean yearly water deficit index) using a Pearson correlation coefficient cut-off of 0.75 (see electronic supplementary material, figure S2).

(d) Model implementation

To enable direct comparison between the SDM and CLM versions of each model-class, we compared five SDMs and their direct CLM counterparts, covering a wide range of model-class types: (i) GLM [34] and CQO [35]; (ii) generalized additive models (GAM [34]) and constrained additive ordination (CAO [36]); (iii) single and multiresponse multivariate adaptive regression splines (MARS and MMARS, respectively, [37,38]); (iv) single classification and regression trees and multivariate regression trees (CARTs and MRTs, respectively, [2,19,39–41]) and (v) single and multiresponse artificial neural networks (ANN and MANN, respectively [42]). A description of each model-class type is provided in the electronic supplementary material, table S2. Models were fitted for 20 selected time periods and then projected to 12 time periods representing distinct climates through the Late Quaternary (figure 1). For fitting models, we pooled taxon occurrence and climate data for 15–14, 17.5–15.5 and 20–15.5 kyr BP to overcome low sample sizes (number of localities) while covering as many time periods as possible during the end of the last glacial period. Similarly, for model projections, we pooled 17.5–15.5, 21–18 and 21–15.5 kyr BP. Data were split into 10 random partitions of training (70%) and testing (30%) sets and models were fit and evaluated using each random set. All models were fitted with the same six climatic variables (i.e. we did not perform variable selection) with the exception of MARS/MMARS and CARTs/MRTs because these algorithms build variable selection into their model fitting. SDM and CLM pairs were tuned so that parameter complexity was equal between the SDM and its CLM counterpart (see electronic supplementary material, table S3). For example, with artificial neural networks, the number of nodes in the hidden layer was set to 20 for each taxon in the SDM (ANN) as well as for the CLM (MANN). To tune parameter values, models were fit with the 0 kyr BP occurrence and climate data and then the chosen parameters of the best model were propagated to the other time periods, thereby controlling model complexity across time. We also created step functions for variable selection and training functions to tune models for the 0 kyr BP time period and then projected these models to the selected 12 time periods, in order to examine if variable selection and parameter tuning influenced SDM and CLM performance (see the electronic supplementary material). All models and analyses were run in R v. 3.1.1 [43]. All R code, including the full list of R packages used and citations, are available on GitHub (https://github.com/fitzLab-AL/SDM-CLMcomp).

(e) Model evaluation

We evaluated the ability of models to predict taxa distributions in terms of both discrimination and reliability (also known as ‘calibration’) [44], which are useful metrics when comparing continuous model outputs to binary observations of the presences/absences. Reliability assesses the overall agreement between observations and predictions, and in particular, whether the model correctly predicts the probability that sites will be occupied or unoccupied. Reliability was measured using the
Brier score [45]. The Brier score is equivalent to mean squared error but is applicable when comparing continuous probabilities to mutually exclusive discrete outcomes, in this case the presence or absence. The Brier score ranges from 0 to 1, in which 0 indicates complete agreement between observed and predicted and 1 indicates complete disagreement. Discrimination quantifies the ability of the model to correctly distinguish among occupied and unoccupied sites (i.e. the extent to which predicted probabilities for occupied sites are higher than those for unoccupied sites), regardless of the correctness of the predicted probability of occupation. We measured discrimination using the area under the receiver operating characteristic curve (AUC). AUC ranges from 0 to 1, with a value of 1 indicating perfect agreement and 0.5 indicating model performance no better than random. We also evaluated the ability of the models to predict community composition by comparing predicted and observed community composition at each locality and examining their similarity using 1 – Bray–Curtis dissimilarity index. Bray–Curtis dissimilarity was calculated between the predicted probabilities of occurrence and observed presence/absence (in other words, observed absence is represented by a probability of 0 and observed presence is represented by a probability of 1). We assessed the statistical significance of differences in model performance between each SDM/CLM comparison and overall SDM and CLM performance using Wilcoxon tests on all 10 iterations for the 19 taxa for each metric.

(f) Model performance versus climatic and compositional novelty

Model performance was evaluated against climatic and compositional (biotic) novelty to examine the extent to which models can be reliably projected to new climatic regimes and no-analogue communities. Climatic novelty for each combination of projected and fitted time periods was calculated by averaging the minimum Euclidean distances between an occupied grid cell for the time period in which the model was fitted and all occupied grid cells in the time period to which the model was projected, based on the first and second principal component axes of a PCA analysis on the six climate variables from all occupied grid cells from all time periods, scaled to have unit variance. Similarly, compositional novelty for each combination of fitted and projected time periods was calculated as the mean of the minimum dissimilarities in relative abundance between each site in the fitted time period compared to all sites in the projected-to time period using Bray–Curtis dissimilarity [14]. To determine whether SDM and CLM performance was influenced by climatic novelty, compositional novelty or by both factors, we built linear models between model performance and each of the factors. We then evaluated the linear models with the Akaike information criteria (AIC) and calculated the AIC weights using the aaike.weights function in the mpcR package.

(g) Model performance and number of occurrences

One of the proposed strengths of CLMs is that by pooling data from all taxa they may better detect shared environmental responses across taxa with fewer occurrences, leading to improved predictions of their distributions. The number of occurrences (or the presence records) available for modelling a given taxon is a function of the number of available localities (sample size) and the prevalence of the taxon (the proportion of these sites at which the taxon of interest is present). The number of localities in the fitted time periods varied between 72 and 276 (including pooled time periods), potentially contributing to differences in model performance between SDMs and CLMs. To examine the influence of this on differences in performance between the SDMs and CLMs, we subsampled the number of localities from the 1 kyr BP (276 localities) and 9 kyr BP (197 localities) time periods at 50, 100, 150 and 200 (1 kyr BP only) localities and reran the analyses. We also examined the effect of prevalence (taxon rarity) on model performance to determine if CLMs model rare taxa better than SDMs.

3. Results

When models were projected to climatically similar time periods, both SDMs and CLMs had good ability to discriminate between occupied and unoccupied sites for individual species (AUC scores more than 0.7), with SDMs tending to outperform CLMs under these circumstances (figure 2a and figure 3, top row). However, the opposite was true when models were projected to climatically dissimilar time periods: both SDMs and CLMs performed poorly and their relative performance switched, with CLMs tending to outperform SDMs (figure 2a), especially when forecasting from the data-sparse LGM to more recent time periods (figure 3, top row). This pattern is especially evident in the GLM–CQO and GAM–CAO comparisons (electronic supplementary material, figure S4a), in which GLM and GAM performed particularly poorly relative to their CLM counterpart when forecasting. MRTs and MMARS also had higher AUC scores than their SDM counterparts, while ANN and MANN performed similarly through time (electronic supplementary material, figure S4).

In terms of model reliability (Brier score), SDMs and CLMs predicted species’ probability of occurrence similarly when projected to climatically similar time periods (figures 2b and 3, middle row); however, when models were hindcasted or forecasted to climatically novel periods, CLMs outperformed SDMs. This pattern was observed in three of the five paired model-class comparisons, except for GLM–CQO and GAM–CAO (electronic supplementary material, figure S4b). For these two model-classes, SDMs outperformed CLMs when models were fitted in, and projected to, climatically similar and temporally close time periods; however, they were similar to the other model-classes in that the CLM consistently outperformed the SDM in climatically novel and temporally distant time periods.

As with AUC and Brier scores, CLMs and stacked SDMs predicted community composition better when models were fitted in, and projected to, climatically similar time periods, with stacked SDMs tending to outperform CLMs (figures 2c, and 3, bottom row). When models were fitted and projected to climatically novel time periods, the models exhibited equal performance, with SDMs hindcasting slightly better and CLMs forecasting slightly better (figure 3, bottom row). When individual model-classes were examined, there were no significant differences between the SDM and CLM (electronic supplementary material, figure S4).

The performance of SDMs and CLMs tracked one another as compositional novelty increased between the fitted and projected time periods (figure 2d–f), with subtle differences: CLMs slightly outperformed SDMs at higher levels of compositional novelty according to the Brier Score, while SDMs outperformed CLMs according to Bray–Curtis similarity (1 – Bray–Curtis dissimilarity) at high and low values of compositional novelty. The inclusion of both climatic and compositional novelty as predictors of model performance is statistically supported over either variable alone (electronic
supplementary material, table S4), suggesting that climatic novelty and compositional novelty each have partially independent effects on model performance. When each variable was examined individually, compositional novelty was statistically more supported than climatic novelty according to all three metrics, the only exception to this being that climatic novelty is more supported for SDM discrimination (AUC).

The number of localities included in the model affected the performance (AUC) of SDMs more than CLMs (figure 4). As the number of localities decreased in the fitted time period (1 kyr BP), the performance of CLMs projected to other time periods did not change substantially according to all three metrics examined, whereas performance of SDMs tended to decrease (figure 4). We also observed this pattern in sample-size sensitivity tests using the 9 kyr BP time period (electronic supplementary material, figure S5). We found no evidence that the predictive skill of CLMs was consistently superior to SDMs for species with low prevalence (electronic supplementary material, figure S6). While CLMs performed better than SDMs at low prevalence for some taxa, other taxa showed the opposite or no trend, indicating that CLMs did not model rarer taxa better than SDMs overall.

4. Discussion

A major challenge in predicting future patterns of biodiversity is accommodating the complex and shifting relationships between the environment, species distributions and communities as the environment changes and communities reshuffle in response to those changes [8]. For example, realized niches can change as climates change [11], owing to changing biotic interactions [46,47] and emergence of novel climates [10]. Here, we found that overall model transferability was low (AUC < 0.65; 1 – Brier score < 0.75; Bray–Curtis similarity < 0.35) in the periods of greatest climatic and compositional novelty, irrespective of whether SDMs or CLMs were used. By simultaneously modelling all co-occurring taxa in a region, CLMs outperformed SDMs for the most novel climates and when forecasting from the data-sparse periods of the Pleistocene. This is most likely because CLMs (i) identify the primary climatic gradients driving biodiversity patterns and thereby buffer against idiosyncratic changes in individual species–environment relationships that can reduce the performance of SDMs and (ii) are more robust to small sample sizes (fewer localities or total number of presences and absences, but not necessarily taxon prevalence).
Figure 3. Mean SDM and CLM performance. Mean AUC scores (top row), 1 – Brier scores (middle row) and Bray–Curtis similarity (bottom row) of SDMs (left column) and CLMs (middle column) for each combination of fit in and projected to periods. Each square represents the average of all five modelling algorithms for all genera. Green shading indicates high performance and orange shading indicates low performance. The right column shows the differences between SDM and CLM scores across all five modelling algorithms and all genera. Blue shading indicates superior SDM performance while red shading indicates better CLM performance. Cells with black outlines indicate statistically significant differences between SDMs and CLMs based on a Wilcoxon test on the mean values for each metric. The thick solid black lines indicate models that were fit and projected to the same time period and thus divide hindcasted models (above line) from forecasted models (below line).
(a) Model performance and transferability

Overall differences between the combined averages (i.e. across all iterations, taxa and model types) of SDMs and CLMs were small (AUC < 0.1; Brier score < 0.15, Bray–Curtis similarity < 0.15). Small differences were to be expected, given that the algorithm for each SDM/CLM pair is essentially the same and differed only in whether the response was univariate or multivariate. Additionally, all models were fitted with the same climatic variables and each SDM/CLM pair was fitted with the same parameter complexity. However, average differences between SDMs and CLMs remained small even when models were permitted to choose the environmental variables and complexity of parameters (AUC < 0.1; Brier score < 0.15, Bray–Curtis similarity < 0.15). Neural networks (ANN and MANN) were a notable exception to this pattern as ANN outperformed MANN more frequently when variable selection and parameter complexity were chosen using a step function (electronic supplementary material, figure S3). The more revealing aspect of this analysis is the inability of models to predict patterns of biodiversity in novel climates with different community compositions. The transferability of a model between time periods is affected by the extent to which biological patterns are controlled by climate, the stability of these relationships between the two time periods of interest and the adequacy of the data. Modelled individual taxon–climate relationships (SDMs) and composition–climate relationships (CLMs) will break down as climates become increasingly novel relative to our reference window and new taxon associations emerge in response to these climates. Thus, model performance is expected to decline as climates become increasingly distant from modern analogues, for which taxon–climate and community–climate relationships are poorly understood [7,10,30]. Our results clearly demonstrate this decline (figure 2a–c). SDMs exhibited greater declines in their ability to predict taxon occurrences than CLMs under increasingly novel climates (figure 2a,b), suggesting that CLMs may be able to extrapolate taxon occurrences beyond known climatic regimes better than SDMs, though perhaps only marginally so. The higher performance of CLMs in predicting taxon occurrences in the most novel communities (figure 2c) is somewhat surprising given that SDMs allow individualistic responses while CLMs are somewhat more constrained [2,15]. Our analysis suggests that enough information was maintained about taxa associations along climate gradients that CLMs could borrow strength and make more informed predictions using low sample sizes. This result did not hold up for discrimination (figure 2d); however, discrimination capacity in SDMs depends on the representativeness of the environmental domain and therefore is a context-dependent property [48] and may not be a good measure of model transferability in this case. SDMs consistently over-fit the training data in this study (electronic supplementary material, figure S7) and others [4] (but see [15]), whereas CLMs showed less over-fitting, perhaps because they were constrained by multi-taxa inputs. This may explain why SDMs tended to have higher predictive skill for temporally adjacent time periods, whereas CLMs predicted better in novel climatic regimes and no-analogue communities than their SDM counterparts.

The relative discrimination capacity of SDMs and their CLM counterparts varied, with some SDMs generally outperforming their CLM counterpart (i.e. GLM, similar to Baselga & Araújo [4] and Bonthoux et al. [15], and GAM), while some CLMs generally outperformed their SDM counterpart (i.e. MMARS, similar to Leathwick et al. [38], and MRTs). ANN and MANN performed similarly, possibly given their use of a machine-learning algorithm designed for both univariate and multivariate analyses [42]. Model reliability (Brier score), on the other hand, clearly shows the superiority of all CLM algorithms to accurately predict the presence/absence, as discussed above.

Although others found CLMs better predicted community composition than SDMs [21,49], our analysis shows the opposite (figure 2c, and 3, bottom row; also see [23]). The one exception was forecasting from the LGM, where CLMs predicted community composition better than SDMs, likely because there were fewer localities during the LGM compared with more recent time periods, and CLMs were better than SDMs at predicting taxon occurrences with few sampling localities (figure 4). Neither stacked SDMs nor CLMs predicted community composition into novel climates particularly well (figure 2c). As compositional novelty increased, both SDMs and CLMs predicted community composition poorly, but on average, SDMs better predicted community composition (figure 2d). This mean result was largely driven by the superiority of GLM and GAM to predict community composition (electronic supplementary material, figure S4c). The SDM and CLM versions of the other three model-classes predicted

Figure 4. Number of localities and performance. Performance of models based on AUC scores (top row), 1 – Brier scores (middle row) and Bray–Curtis similarity (bottom row) when the number of localities used to fit the model in 1 kyr BP is the original number (n = 276) and subsampled to 100, 150 and 200 localities and then projected to chosen time periods. Solid line represents the mean value of each metric across all SDMs and the dashed line represents the mean across all CLMs.
community composition equally. Overall, results suggest that CLMs may not be able to capture emergent community properties in extreme instances of climatic and compositional novelty any better than stacked SDMs owing to individualistic taxon responses [15].

The differences in model performance and transferability between SDMs and CLMs is not owing just to the emergence of novel climates or the emergence of novel communities; rather, both are necessary to fully explain variation in model performance (electronic supplementary material, table S4). Climatic and compositional novelties were positively correlated until compositional novelty reached a plateau (figure 2g). This pattern was recovered for the 19 taxa in this analysis as well as all 109 taxa available in the original dataset and has been observed both temporally (this analysis, but see [14]) and spatially [50]. At the point at which compositional novelty plateaued in this analysis (compositional novelty approx. 0.3 and climatic novelty approx. 0.5) there was a shift in model performance across all metrics, after which model performance stabilized (figure 2a–f). After this threshold, model performance was mostly influenced by climatic novelty, when CLMs predicted distributions better than SDMs, further suggesting the superiority of CLMs when projecting to novel climates.

(b) Model performance and number of occurrences
The number of occurrences available to model a taxon’s distribution, which itself results from the interplay between the number of available localities and taxon prevalence, drove differences between model performance. When fitted with few localities, CLMs may have outperformed SDMs because of their multiresponse nature in which they can ‘borrow strength’ across taxa. On the other hand, when more localities were available, parameter estimation from SDMs was as good as or better than CLMs. Interestingly in our study, when sample size effect was removed, taxon prevalence was inconsistently correlated with model performance (electronic supplementary material), nor was there any consistent difference in the effect of prevalence on the predictive skill of SDMs and CLMs. It has been hypothesized [2,51] and demonstrated [17–19] that CLMs improve predicted distributions of rare taxa by incorporating information on co-occurrences with other taxa that have similar environmental tolerances (but see [15,19]). Whether or not CLMs ‘borrow strength’ to support rare taxa is complex—Bonthoux et al. [15] found that CLMs improved the understanding of rare species patterns but did not predict their distributions better than SDMs. Here we suggest that the superior predictive ability of CLMs may not be just the result of their ability to model rare taxa but also because they are more robust to low numbers of localities.

Overall, the difference in predictive performance of SDMs versus CLMs was modest even when variable selection and parameter complexity were introduced [19]. Similar to this study, several previous analyses examining the model performance of CLMs against SDMs reported slight differences or mixed results [4,15,22,24], indicating that model performance may be case-dependent and related to individual taxon–environment relationships [4]. Model performance might also be influenced by inaccurate palaeoclimate simulations as well as uncertainties in the pollen data and age models, or missing important covariates such as CO2 concentrations [6]. However, SDMs and CLMs should be equally affected by these uncertainties and therefore are unlikely to affect the main conclusions and comparative analyses presented here.

5. Conclusion
Both SDMs and CLMs performed poorly when projected into novel climates and assemblages. This finding calls into question the dependability and utility of using empirical models to project future distributions and communities. Overall, however, our results both reveal a path towards better predictions of future ecological assemblages given the large magnitude of climatic change expected for the future and highlight the need for better approaches given the expected increase in compositional novelty [7,52]. We confirm that CLMs are a similar or modestly superior approach to SDMs for modelling the responses of species diversity and distributions to climate change. Although CLMs are no better at predicting community composition than stacked SDMs [23] (but see [21]), they generally have higher predictive skill (with respect to discrimination and reliability) under conditions of novel climates and low taxon occurrences. For these reasons, CLMs deserve greater attention, application and examination than they have received to date.

Data accessibility. The presence/absence matrices of pollen taxa at each site per time slice and a list of final sites included in the analysis are available at Dryad data repository (http://dx.doi.org/10.5061/dryad.hk400).

Authors’ contributions. J.L.B., M.C.F., D.N.L., K.C.M., J.W.W. designed the study; J.L.B. compiled the pollen data; D.N.L. down-scaled and prepared the climate data; D.N.L. wrote the R code and scripts; D.N.L. and K.C.M. performed the analyses; J.L.B., M.C.F., S.F., D.N.L., K.C.M., J.W.W. analysed the output data; K.C.M. wrote the manuscript; all the authors contributed substantially to revisions, and gave final approval for publication.

Competing interests. The authors declare that they have no competing interests.

Funding. This work was supported by NSF (DEB-1257033 to J.L.B., DEB-1257164 to M.C.F. and DEB-1257508 to J.W.W.)

Acknowledgements. The authors thank Bette Otto-Bliesner, Zhengyu Liu and Feng He for the original climate simulations. We thank David Harris for helpful discussions and Janet Franklin and three anonymous reviewers for comments that greatly improved the manuscript. This is UMCS- Appalachian Laboratory Scientific contribution no. 5172.

References


5. Elith J, Leathwick J. 2009 Species distribution models: ecological explanation and prediction across


Correction to ‘Controlled comparison of species- and community-level models across novel climates and communities’

Kaitlin C. Maguire, Diego Nieto-Lugilde, Jessica L. Blois, Matthew C. Fitzpatrick, John W. Williams, Simon Ferrier and David J. Lorenz


One of the six climate variables used to fit the models was listed incorrectly in the Environmental variables section under Material and methods [1]. Mean yearly evapotranspiration ratio was used, not mean yearly potential evapotranspiration.

Reference