



Forecasting the future of biodiversity: a test of single- and multi-species models for ants in North America

Matthew C. Fitzpatrick, Nathan J. Sanders, Simon Ferrier, John T. Longino, Michael D. Weiser and Rob Dunn

M. C. Fitzpatrick (mfitzpatrick@umces.edu), Univ. of Maryland Center for Environmental Science, Appalachian Lab, Frostburg, MD 21532, USA. – N. J. Sanders, Dept of Ecology and Evolutionary Biology, 569 Dabney Hall, Univ. of Tennessee, Knoxville, TN 37918, USA, and Center for Macroecology, Evolution and Climate, Dept of Biology, Univ. of Copenhagen, DK-2100 Copenhagen, Denmark. – S. Ferrier, CSIRO Ecosystem Sciences, Black Mountain Laboratories, Clunies Ross Street, Black Mountain ACT 2601, GPO Box 1700, Canberra ACT 2601, Australia. – J. T. Longino, The Evergreen State College, Olympia, WA 98505, USA. – M. D. Weiser, Dept of Biology, North Carolina State Univ., Raleigh, NC 27695, USA. – R. Dunn, Dept of Biology and Keck Center for Behavioral Biology, North Carolina State Univ., Raleigh, NC 27695, USA.

The geographic distributions of many taxonomic groups remain mostly unknown, hindering attempts to investigate the response of the majority of species on Earth to climate change using species distributions models (SDMs). Multi-species models can incorporate data for rare or poorly-sampled species, but their application to forecasting climate change impacts on biodiversity has been limited. Here we compare forecasts of changes in patterns of biodiversity in North America derived from ensembles of single-species models to those from a multi-species modeling approach, Generalized Dissimilarity Modeling (GDM). We found that both single- and multi-species models forecasted large changes in ant community composition in relatively warm environments. GDM predicted higher turnover than SDMs and across a larger contiguous area, including the southern third of North America and notably Central America, where the proportion of ants with relatively small ranges is high and where data limitations are most likely to impede the application of SDMs. Differences between approaches were also influenced by assumptions regarding dispersal, with forecasts being more similar if no-dispersal was assumed. When full-dispersal was assumed, SDMs predicted higher turnover in southern Canada than did GDM. Taken together, our results suggest that 1) warm rather than cold regions potentially could experience the greatest changes in ant fauna under climate change and that 2) multi-species models may represent an important complement to SDMs, particularly in analyses involving large numbers of rare or poorly-sampled species. Comparisons of the ability of single- and multi-species models to predict observed changes in community composition are needed in order to draw definitive conclusions regarding their application to investigating climate change impacts on biodiversity.

Concern over global change and the potential for associated species loss have increased the emphasis on understanding the factors that influence patterns of biodiversity and on forecasting how these patterns may change in the future (Thomas et al. 2004, Araujo and Rahbek 2006, Fitzpatrick et al. 2008). However, many regions and taxonomic groups remain poorly sampled, hindering attempts to formulate conservation strategies for all but a handful of species, mostly vertebrates and temperate plants. Species-rich taxonomic groups such as insects and other invertebrates pose a particular challenge for conservation managers, and as a result such groups remain poorly considered in conservation planning even though they may constitute the majority of eukaryotic species on Earth (Erwin 1982, Stork 1997, Dunn 2005, Hamilton et al. 2010).

In the absence of sufficient biological data, statistical modeling is often used to relate occurrence data to

environmental factors that might limit individual species distributions or explain overall patterns of biodiversity. These approaches allow species distributions and/or emergent patterns of biodiversity (e.g. species richness or community types) to be mapped across regions of interest (Guisan and Zimmermann 2000, Ferrier et al. 2002, Ferrier and Guisan 2006). Once developed, the models can also be used to assess threats to species persistence under scenarios of environmental and/or climatic change using forecasts from global circulation or land cover change models (Yates et al. 2010).

An increasingly common approach used to forecast changes in the distributions of individual species and to assess extinction risks under climate change is the application of species distribution models (SDMs, Guisan and Thuiller 2005). The popularity of SDMs can be attributed to the fact that, unlike mechanistic niche models (Kearney and Porter

2009), which can be developed for those few taxa for which we have sufficient physiological understanding, they require only data on species occurrence and associated environmental conditions and thus can be applied across multiple taxa, regions, times, and spatial scales (Guisan and Thuiller 2005). Despite their practical strengths, the availability of occurrence and environmental datasets often limit the application of SDMs to relatively common or well-sampled species – those species whose geographic ranges are large relative to the density of field sampling. These species are often also those of least conservation concern, both because they are widespread and almost necessarily have broad environmental tolerances. In cases where groups are poorly known, or for groups that exhibit rapid spatial turnover relative to the grain of sampling, many species will not be sampled at all or will be represented by too few records to allow implementation of SDMs (Ferrier and Guisan 2006). Second, in some instances conservation managers may not be interested in the response of or threat to any single species. Potentially of greater importance is identifying where large numbers of species are at risk (Myers et al. 2000) or determining how species composition of a locale may change in the future. Such inferences can be achieved by combining SDMs for individual taxa to assess changes in species richness (Fitzpatrick et al. 2008), composition (Leathwick et al. 1996), or functional types (Thuiller et al. 2006). But as the number of species being modeled grows, it becomes increasingly time consuming to analyze and interpret models for each species. Additionally, it is unclear whether the assembled SDMs reliably characterize emergent patterns of biodiversity as SDMs rarely incorporate the influences of species interactions, dispersal, or evolution (Guisan and Thuiller 2005).

An alternative but relatively unexplored approach to forecast potential impacts of climate change on biodiversity involves modeling the emergent patterns of biodiversity directly (Ferrier 2002) rather than combining individual models for many species. In this approach, emphasis shifts from discrete entities such as species ranges to collective properties of biodiversity such as alpha diversity (the number of species in a locale) or beta diversity (change in species composition in space or time). Unlike species-level modeling, for which species with too little data must be excluded from further analysis for statistical reasons, community-level modeling strategies can make use of all available data across all species, regardless of the number of records per species (Ferrier and Guisan 2006). Thus, modeling the collective properties of biodiversity rather than individual species ranges may confer benefits for analyses involving very large numbers of species, particularly when a majority of these species are rarely recorded, as is the case for most taxa on Earth.

Beta diversity is a collective measure of biodiversity that has received renewed attention (Tuomisto 2010a, b) as it has a clear conceptual link to the reorganization of communities under environmental change. For example, general turnover metrics, such as percent turnover, are often computed to quantify climate change impacts using stacks of single-species SDMs (Thuiller et al. 2005, Broennimann et al. 2006, Lawler et al. 2009). Ferrier et al. (2002, 2007) recently developed Generalized Dissimilarity Modeling (GDM) as a means to characterize species turnover as a function of

environmental and geographic separation. Rather than modeling individual species distributions, GDM models compositional dissimilarity (i.e. spatial turnover of species composition) between all possible pairs of locations as a function of environmental differences between these locations. Potential strengths of the GDM approach, relative to SDMs, in assessing spatial patterns of biodiversity and for quantifying or forecasting biodiversity change (Ferrier and Guisan 2006, Ferrier et al. 2007), include an ability to 1) rapidly analyze datasets containing very large numbers of species; 2) make use of data for all species in these datasets, regardless of the number of records per species; and 3) extrapolate patterns in compositional turnover beyond sampled communities. An additional strength of the GDM approach relative to other community-level modeling approaches is the ability to accommodate both the curvilinear relationship between environmental (and/or geographic) separation and compositional dissimilarity between sites, and the variation (non-stationarity) in the rate of compositional turnover at different positions along environmental gradients. Importantly, and unlike some community-based modeling approaches, using GDM to forecast impacts of climate change on biodiversity does not assume that species will move together as fixed community types. Rather, GDM assumes that emergent rates of spatial turnover along environmental gradients under current climatic conditions can act as a reliable surrogate for temporal turnover given environmental change in time (i.e. climatic change). GDM may represent an important supplement to existing methods used to assess the impacts of climate change on biodiversity, yet its use in this context remains largely unexplored (Ferrier and Guisan 2006).

Here we use GDM to forecast impacts of climate change on ants in North America, and we compare these forecasts to those derived from ensemble forecasts derived from single-species SDMs (Araujo and New 2007). Ants represent an ideal taxon for comparisons of SDMs and GDM as ants, including those in North America, are represented by both widespread and narrowly endemic species and well-known and poorly-known species. As an example of the former, the ant species *Tapinoma sessile* occurs in habitats ranging from New England bogs to moist parts of southwestern deserts (Menke et al. 2010). As an example of the latter, the species *Amblyopone trigonignatha* is known from one specimen and one set of photographs of two individuals. As such, only a subset of species can currently be modeled with SDMs, whereas the data for all species can be considered in the fitting of models using GDM. Ants are arguably one of the better-studied insect taxa with regard to both basic biology and systematics (Hölldobler and Wilson 1990) and North America is one of the better-studied regions. In addition, ants can represent the dominant fraction of animal biomass in some habitats (Hölldobler and Wilson 1990, Folgarait 1998), and mediate many ecological processes and therefore the loss or addition of ant taxa can potentially have large ecological consequences. In short, as is the case for most groups of organisms on the planet, there are data suitable for the development of SDMs, but only for a subset of taxa. We ask 1) where do SDMs and GDM forecast the greatest changes in ant community composition? 2) Do these projected changes differ in a meaningful way between SDMs and GDM; and

3) if so, where and why? When considering community composition, we consider both changes in the genera present in a region and, for SDMs, changes in the functional groups present (e.g. specialist mite feeders, general omnivores, etc.) with the idea that the loss or arrival of functional groups may have greater impacts than changes in genera within functional groups, to the extent there is redundancy (e.g. many omnivorous ant genera share similar diets).

Methods

Biological data

We focus on native ant genera occurring within continental North America north of the Colombia–Panama border and excluding Greenland using a database of North American ant species derived from specimens in museums, published systematic revisions of taxa, specimens in AntWeb (<www.antweb.org>) and other records. We augmented these data with an additional 23 618 records from the personal collection of J. Longino. We obtained occurrence data (presence-only) for 104 genera and 1696 species, removing any species or genera not native to North America. Because the spatial accuracy of the occurrence records was variable, we aggregated the data to a cell size of 10 arc-minutes (ca 20 × 20 km), resulting in 6733 locations with ant occurrence data (Fig. 1, data to be made available as online supplement). Given that ant species are continuing to be named in North America and the distribution of named ant species is sometimes only provisionally known (e.g. as in the case of *A. trigonignatha*), we focused our analyses at the level of the genus. We assigned genera to functional groups following Weiser and Kaspari (2006), supplementing when necessary based on recent literature studies (Davidson et al. 2003) and studies from the primary literature (Supplementary material Appendix 1).

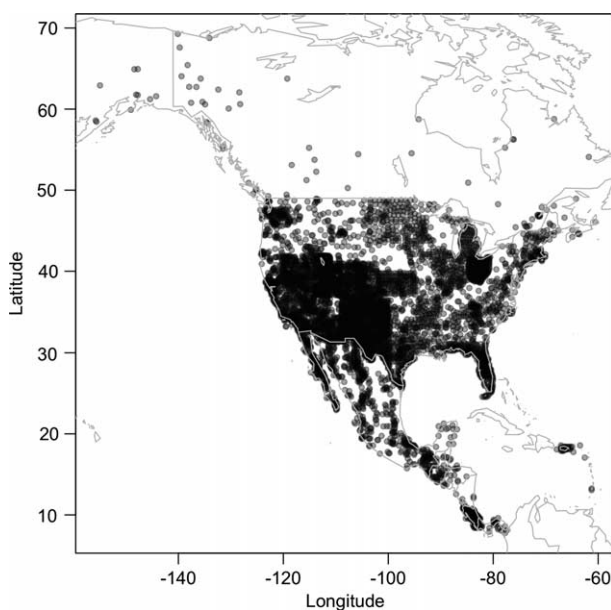


Figure 1. Locations with occurrence (presence-only) data for ants.

Environmental predictors

Numerous studies suggest both patterns of ant species richness and geographic ranges of individual species are related to climate and in particular temperature, precipitation, and humidity (Davidson 1977, Bestelmeyer 2000, Kaspari et al. 2000, 2004, Retana and Cerdá 2000, Sanders et al. 2003, 2007, Dunn et al. 2009). Precisely which aspects of climate (minimums, maximums, averages, or seasonality) influence individual distributions of ants rather than richness patterns per se have been less well explored. We thus initially considered a comprehensive set of 24 bioclimatic variables representing average conditions for the years 1950–2000 and used variable selection techniques (described below) to reduce this set to a few meaningful predictors. This full set was comprised of 19 predictors from the WorldClim database (<www.worldclim.org>, Hijmans et al. 2005) describing minimum, maximum, and seasonality in temperature and precipitation and four additional variables that we developed and that capture interactions between precipitation and temperature, including absolute minimum temperature, annual water balance, summer water balance, and water balance seasonality (see Svenning and Skov 2004 for details). Although remotely-sensed measures of available energy, such as net primary productivity and normalized difference vegetation index, may also be important determinants of biogeographical patterns of ant distributions, such variables are not available for the future. Thus, we instead used actual evapotranspiration (AET) as a measure of available energy computed using Turc's formula (cited in Kluge et al. 2006). We developed the same suite of 24 variables representing potential future climatic conditions in year 2050 as may develop under the assumptions of the HadCM3 general circulation model and the SRES A2a emission scenario, also from WorldClim. The HadCM3 general circulation model under the SRES A2a emission scenario forecasts at least 2°C warming across much of North America by 2050 with greater forecasted warming in the Arctic (+5°C) and in western and central North America (+3.5°C). Arid regions in southwestern North America are forecasted to receive 50% more annual precipitation, whereas regions south of and including the Yucatan Peninsula are forecasted to experience a decrease in annual precipitation of nearly 50%. We chose this scenario of future climate as it predicts relatively rapid increases in atmospheric carbon dioxide that are in line with current global trends (Canadell et al. 2007) and recognize that other models yield predictions that differ in their specifics of forecasted climate change. However, it is beyond the scope of the study to consider multiple scenarios of future climate. All climate variables were manipulated in ArcGIS 9.3 (ESRI 2009) such that they were spatially congruent and had a common resolution of 10 arc-minutes (ca 20 × 20 km), which matched the occurrence data.

Generalized dissimilarity modeling

GDM is a nonlinear, multivariate extension of Mantel correlation analysis (Manly 1998) for analyzing and predicting patterns of compositional dissimilarity (quantified with a presence-absence version of version of the Bray–Curtis

dissimilarity index) in relation to environmental gradients. It links sampled biological data with study-area wide environmental data using non-linear functions that can be used to rapidly predict spatial patterns in the composition of biodiversity across large regions and to estimate changes in these patterns under environmental change. Specifically, GDM uses GLMs (Generalized Linear Models) to model observed Bray–Curtis dissimilarity, d_{ij} between pairs of locations i and j as a function of n environmental variables, x_1 to x_n , using an exponential link function of the form

$$-\ln(1 - d_{ij}) = a_0 + \sum_{p=1}^n |f_p(x_{pi}) - f_p(x_{pj})| \quad (1)$$

where

$$d_{ij} = 1 - \frac{2A}{2A + B + C} \quad (2)$$

and where A is the number of genera common to both sites i and j ; B is the number of genera present only at site i ; and C is the number of genera present only at site j . The functions f_p transform the environmental variables to obtain the best fit of the observed dissimilarities, which in turn can be used to predict differences in community composition in space and time. To implement GDM, we used software available from www.biomaps.net.au/gdm (ver. 1.1) in R 2.10.0 (R Development Core Team 2009). See Ferrier et al. (2007) for further details and applications.

We converted the presence-only ant distribution data into a binary (0/1) sites-by-genera matrix, in which each ‘site’ consists of a ca 20×20 km grid cell where at least one ant genus had been collected. This matrix was used to derive the Bray–Curtis measure (equivalent to 1–Sørensen’s index, Sørensen 1948) of inter-site dissimilarity as the response variable for the GDM. Our site-by-genera matrix assumes that if a genus had not been collected in a grid cell it could be considered absent at that location. This is not true in many instances. Therefore we took three measures to account for potential biases introduced by the use of presence-only data and by differences in collection effort between locations. First, we used the ‘standard’ weighting function within GDM. Standard weighting weights sites proportionally to the number of species or genera observed at that site, such that sites with few genera carried less weight, and therefore less influence, in model fitting than sites with larger numbers of genera. This approach is likely the safest option to use if sampling effort is known to have varied substantially between sites, but it has the disadvantage of downweighting marginal (e.g. very cold) but potentially widespread conditions. We also ran models without weighting and found results were highly similar to those obtained using standard weighting. Second, to account for the inflation of observed dissimilarities due to the use of presence-only data, we excluded the intercept term fitted in the GDM model when making forecasts, which ensures that two sites that do not differ in environmental variables (i.e. environmental distance between sites is zero) will have a predicted dissimilarity of zero and that dissimilarities between all remaining pairs of sites will be scaled accordingly (Allnutt et al. 2008). Third, sites with low generic richness because of incomplete sampling can appear to be completely dissimilar to one another.

Therefore, to account for the potential influence of incomplete sampling, we excluded sites where fewer than five genera had been recorded. For the predictor variable, we constructed a corresponding sites-by-environment matrix using the variables described below. Using these two matrices we developed two GDM models, one using only genera with enough data for SDMs (66, see below) and another using all 104 genera in the database. Comparisons with the SDMs are made using the GDM with 66 genera.

Rather than blindly use all 24 available environmental variables to develop models, we used two procedures to perform variable reduction. First, we reduced the full set by selecting eight climate variables that minimized collinearity ($\rho < 0.7$), keeping those variables of correlated pairs that were in our opinion most biologically meaningful. These eight included annual mean temperature, mean diurnal range, isothermality, maximum temperature of the warmest month, annual precipitation, precipitation seasonality, precipitation of the driest quarter, and summer water balance. These eight variables were also used to model the distributions of ant genera using SDMs. Second, for the GDM models, we further reduced this set of eight using a custom backward-elimination variable selection routine written in the statistical program R v2.10.0 (R Development Core Team 2009). At each step in the routine, a GDM model was fit using the available n predictors (where $n = 8$ at the beginning of the routine). The predictor associated with the smallest amount of compositional change, as determined by summing the coefficients of the I-splines (Ferrier et al. 2007) when all other variables were held constant, was removed and a second GDM model was then fit using this reduced set of $n - 1$ predictors. The significance of removing the variable from the model was evaluated by calculating the difference in deviance between two models with and without the variable. The observed difference in deviance was then compared to a null distribution of differences in deviance obtained by fitting the two models using 500 random permutations of the order of the sites in the response (compositional dissimilarity) matrix. If no significant difference was found in the deviance between the two models, then the variable that was removed was not a significant predictor of compositional dissimilarity and could be eliminated from the model. Variable elimination continued until the difference in deviance between the two models became significant, thus indicating that no more variables could be eliminated without reducing model quality. This procedure resulted in a final model containing five variables: annual mean temperature, mean diurnal range, isothermality, maximum temperature of the warmest month, and summer water balance.

Our use of GDM to assess potential climate change impacts proceeded in two steps. First, we used GDM to relate compositional dissimilarity between all pairs of sites (i.e. all grid cells where ant genera were recorded) to their current environmental separation. This step provided functions that describe how ant community composition changes as a function of environmental separation and in this case, environmental separation occurs because sites are separated in space. Next, we used this model to predict how ant community composition would change in time given the amount of environmental separation between current

climate and climate in year 2050 at each location. In this case, environmental separation occurs not in space but in time because the climate at each location changes from t_1 to t_2 . The implicit assumption when GDM is used in this context is that the amount of compositional change modeled between two locations separated in current environmental space can be used to approximate how much a single location will differ in composition given an amount of environmental change in time.

Ant distribution models

We modeled individual distributions of each ant genus using the BIOMOD framework (Thuiller et al. 2009). BIOMOD contains nine algorithms for modeling species distributions and allows for the combination of multiple models (different algorithms, initial conditions, etc.) into a single ensemble, thereby reducing the influence of algorithm- and data-specific uncertainty and often producing a more robust model than can be achieved by any single method (Araújo et al. 2005, Araujo and New 2007, Marmion et al. 2009). For this study we used all nine algorithms within BIOMOD, which include artificial neural networks, classification trees, generalized additive models, generalized boosted regression tree models, generalized linear models, multivariate adaptive regression splines, mixture discriminate analysis, random forests, and surface range envelopes. See Thuiller et al. 2009 for details.

BIOMOD requires information on both presence and absence of modeled entities. Such data are unavailable for most species and regions (Graham et al. 2004), including those in our study. To circumvent this problem, we created artificial absence data (often called background or pseudo-absence data) using three approaches. These included 1) assuming, as we did for GDM, that a genus was absent at any location in the occurrence dataset where it had not been collected, 2) using a method proposed by Phillips et al. (2009) in which all collection locations in the occurrence dataset are used as absences, even locations where a given genus has been collected, and 3) selecting 500, 1000, or 10 000 random points from the entire study area, with an equal weight of presences vs background data. Of these three approaches, models fitted with 10 000 random points selected from the entire study area and with an equal weight of presence vs background data produced models with the best predictive accuracy, as measured statistically and by visual inspection. To avoid potential problems relating to small sample sizes, we developed models only for genera that had at least 20 spatially unique distribution records. Thirty-eight genera did not meet this criterion, leaving 66 genera.

Within BIOMOD, distribution data for each genus were partitioned randomly 10 times into calibration (70%) and evaluation (30%) datasets, and models were run on each of the 10 resulting datasets. Model accuracy was calculated using the True Skill Statistic (TSS, Allouche et al. 2006). The TSS, an established approach for assessing the accuracy of weather forecasts, compares the number of correct forecasts, minus those attributable to random guessing, to that of a hypothetical set of perfect forecasts. For a 2×2 confusion matrix, TSS is defined as $TSS = \text{sensitivity} +$

$\text{specificity} - 1$. It takes into account both omission and commission errors and success as a result of random guessing, and ranges from -1 to $+1$, where $+1$ indicates perfect agreement and values of zero or less indicate a performance no better than random. The TSS has recently been shown to be superior to kappa; notably, by not being affected by species prevalence (Allouche et al. 2006). The probability of occurrence was converted into binary presence/absence using the threshold maximizing TSS. The multiple models for each ant genus resulting from different algorithms and random splits of the occurrence data into training and test data were combined into a single ensemble using weighted averaging. The contribution of a model to the ensemble was based on its TSS score and the relative weight of the contribution was calculated using a decay of 1.6 (the default weighting in BIOMOD, Thuiller et al. 2009).

To calculate the Bray–Curtis dissimilarity between current and future distributions of ant genera from the SDMs, the single binary ensemble range prediction for each genus under current climate were stacked to produce a sites-by-genera matrix representing predicted ant community composition. This process was repeated for projected future distributions resulting in a sites-by-genera matrix for ant community composition under future climate. These two matrices were then used to calculate dissimilarity between present and future using the same formulation of the presence-absence version of the Bray–Curtis dissimilarity index used by GDM (eq. 2), but in this case A equaled the number of genera present at a site at both times i and j ; B equaled the number of genera present only at time i ; and C equaled the number of genera present only at time j . We performed these calculations assuming two extremes of dispersal; either individuals are able to disperse to all locations that become suitable (full-dispersal) or individuals cannot disperse at all and only lose range as climate changes (no-dispersal). Note that GDM assumes that forecasted changes in composition will occur regardless of the distance species would have to migrate in order to realize these forecasts, which is, in essence, to assume full dispersal.

To calculate changes in functional groups, we determined for each pixel in the study area whether it was forecasted to gain a functional type that was not present under current climate and whether it was forecasted to lose a functional type in the future such that it would no longer be represented by any genera in that location. Although some pixels may be forecasted to experience changes in composition of genera, such changes may not necessarily result in changes in the functional types present since multiple genera were assigned to the same functional type. We did not determine whether changes in functional groups were the result of shifts in the distribution of multiple genera or a single genus despite the fact that changes in multiple genera are likely to have greater consequences. We only report results for those functional groups that exhibited losses or gains.

Results

The predictive accuracies of the genus-level SDMs (True Skill Statistic values from the BIOMOD ensembles) were

consistently high and ranged from 0.990 to 0.999 (mean for all genera = 0.997, Supplementary material Appendix 1). The GDM model accounted for 40.9% of the deviance in observed turnover of ant genera. The largest amount of turnover was observed along the gradient of mean annual

temperature, followed by summer water balance (maximum height of curves in Fig. 2). In general, turnover was most rapid (slope of curves in Fig. 2) in hot environments, whereas colder environments exhibited relatively less compositional change per unit environmental distance.

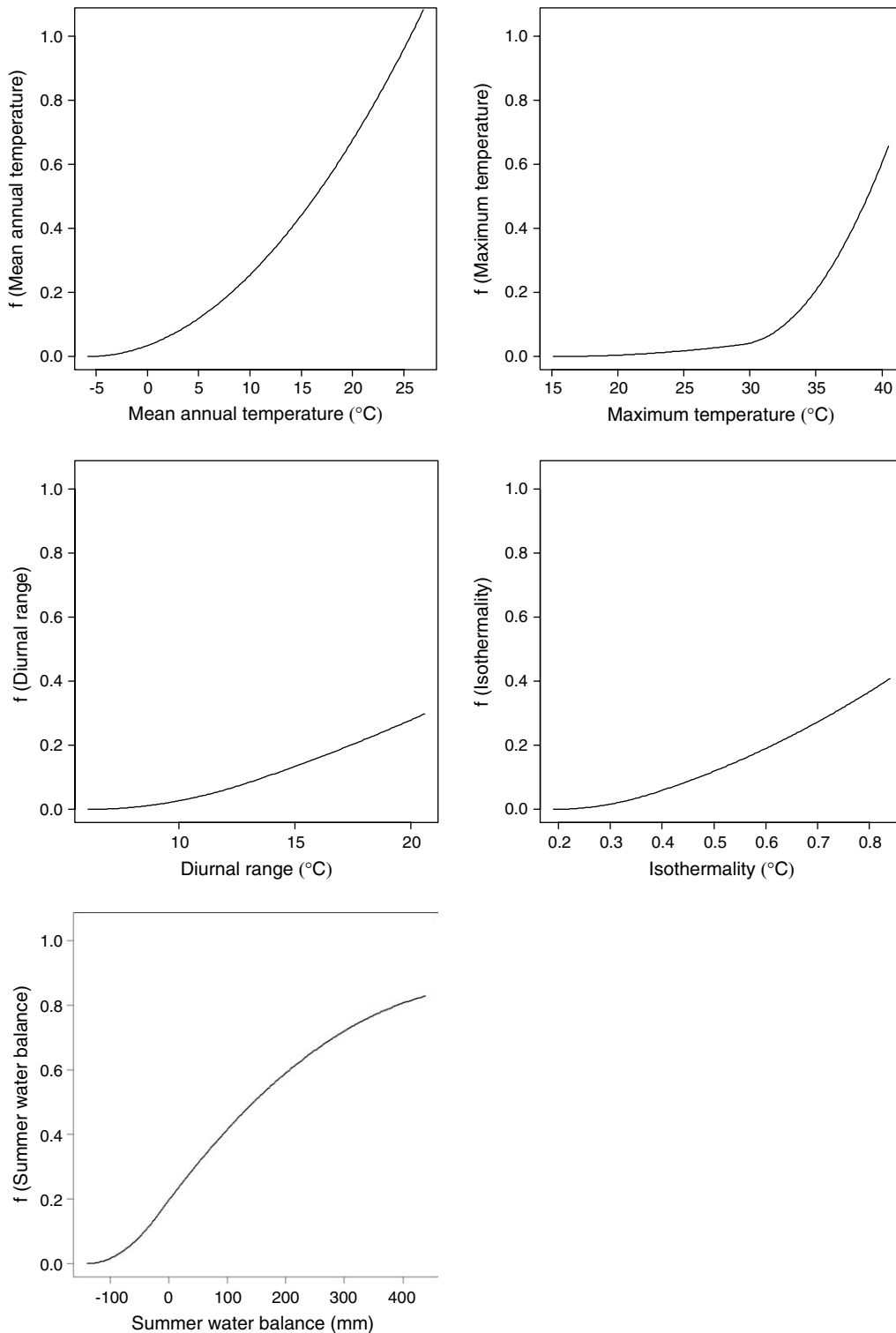


Figure 2. Fitted functions of observed turnover in composition of ant genera for a Generalized Dissimilarity Model using five environmental variables. The maximum height reached by each function provides an indication of the total amount of compositional turnover associated with that variable, holding all other variables constant. The slope of each function provides an indication of the rate of compositional turnover and how this rate varies along the gradient.

The r^2 value between GDM and the no-dispersal scenario from BIOMOD was more than three times that between the full-dispersal scenario and GDM (0.54 vs 0.17 respectively, Fig. 3). Compositional change forecasted using BIOMOD and assuming no-dispersal more closely matched the GDM forecasts than did those assuming full-dispersal (Fig. 3, 4), mainly because the assumption of no-dispersal prevented northward range shifts in many genera across large regions of northern North America, where GDM predicted little change in composition (Fig. 4). GDM tended to forecast greater amounts of compositional change and over a more contiguous area than did BIOMOD, including the south central United States (US), the Sierra Nevada of the western US, and most of Central America (Fig. 4). In Central America, ant generic richness is highest in the study region (Fig. 5a) and ranges

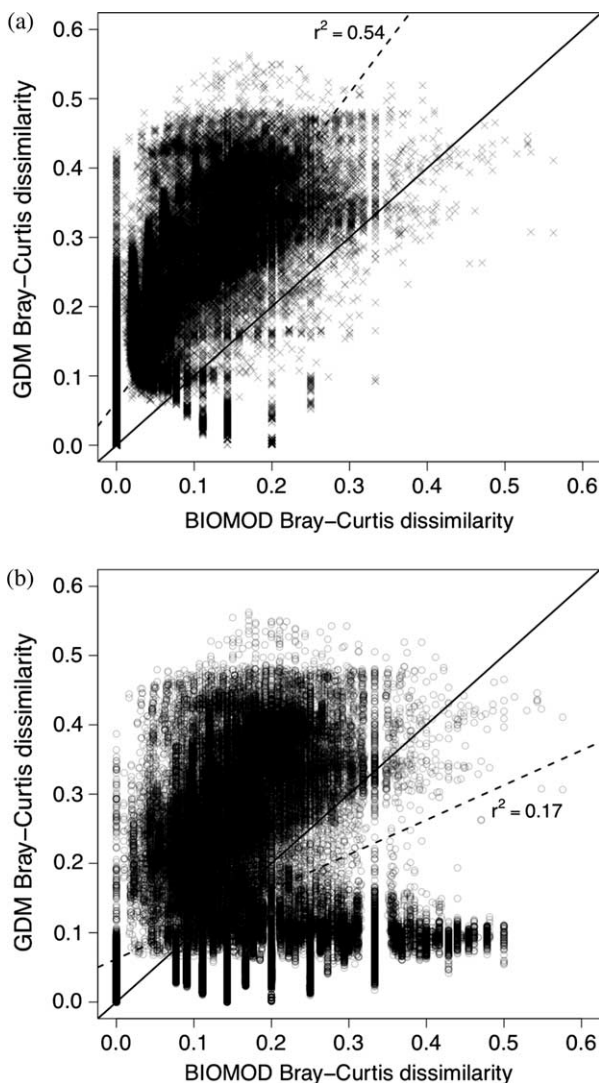


Figure 3. The relationship between forecasted compositional dissimilarity calculated from stacking 66 individual ant genus models from BIOMOD assuming either (a) no-dispersal (\times 's) or (b) full-dispersal (hollow circles) versus forecasted compositional dissimilarity from a GDM. Each point represents a grid cell in North America. The dashed lines are best-fit regressions. The solid black lines are unity.

are small (at least within the domain considered, some 'small-ranged' genera in our study domain actually have large ranges in the Neotropics) relative to grain size of the analysis. In contrast, when full dispersal was assumed, BIOMOD predicted greater amounts of compositional change than GDM in a narrow band across southern Canada (Fig. 4a), where SDMS projected northward range shifts of multiple genera and where current richness is low (Fig. 5b). Although BIOMOD and GDM tended to differ in the magnitude of forecasted change, both models highlighted northern Mexico and the Central Valley of California as regions having the greatest potential impacts and the intermountain West and the northeastern US as having the lowest forecasted changes. When all 104 genera were used in the GDM model (Fig. 4d), the results were similar to those of the GDM incorporating only those 66 genera with enough data for SDMs, but with generally higher forecasted turnover and particularly so in the Yucatan Peninsula and in Central America.

In terms of losses and gains of functional types determined from genus-level SDMs, southeastern Canada was forecasted to experience the most gains in functional types (Fig. 6). In contrast, loss of functional types was limited mainly to the southern half of the domain, though generalities were more difficult to draw for losses as compared to gains. In general, a greater portion of North America was predicted to lose rather than gain functional types, and gains would be realized only when assuming full-dispersal.

Discussion

What do two independent modeling approaches suggest regarding changes in patterns of ant biodiversity in North America under climate change? When considered qualitatively and at the broadest spatial extents, several notable similarities and differences between the forecasts of the two models emerge. Both SDMs and GDM suggest that turnover of ant genera may be high in relatively warm and dry environments such as those found in the south central US, California, and northern Mexico and relatively low in cold and wet regions elsewhere. Notable differences are that GDM also forecasts relatively large changes in ant community composition for the tropical portion of the study region and SDMs forecast greatest turnover in southeastern Canada, but only when full-dispersal is assumed. When the models agree, they suggest that warm rather than cold regions potentially could experience the greatest changes in ant fauna under climate change with corresponding alterations to ecosystem function to the extent that ants mediate such processes.

Although the models reach broadly similar conclusion regarding compositional change for some regions, the types of conclusions drawn from each approach differ. GDM says something about the drivers of compositional change, whereas SDMs say something about the nature, and thus the consequences, of those changes. For example, the region with the highest agreement between the models is relatively warm. The GDM model suggests that hot environments exhibit a large amount of turnover in ant genera per unit environmental distance (Fig. 2). Thus, relatively small

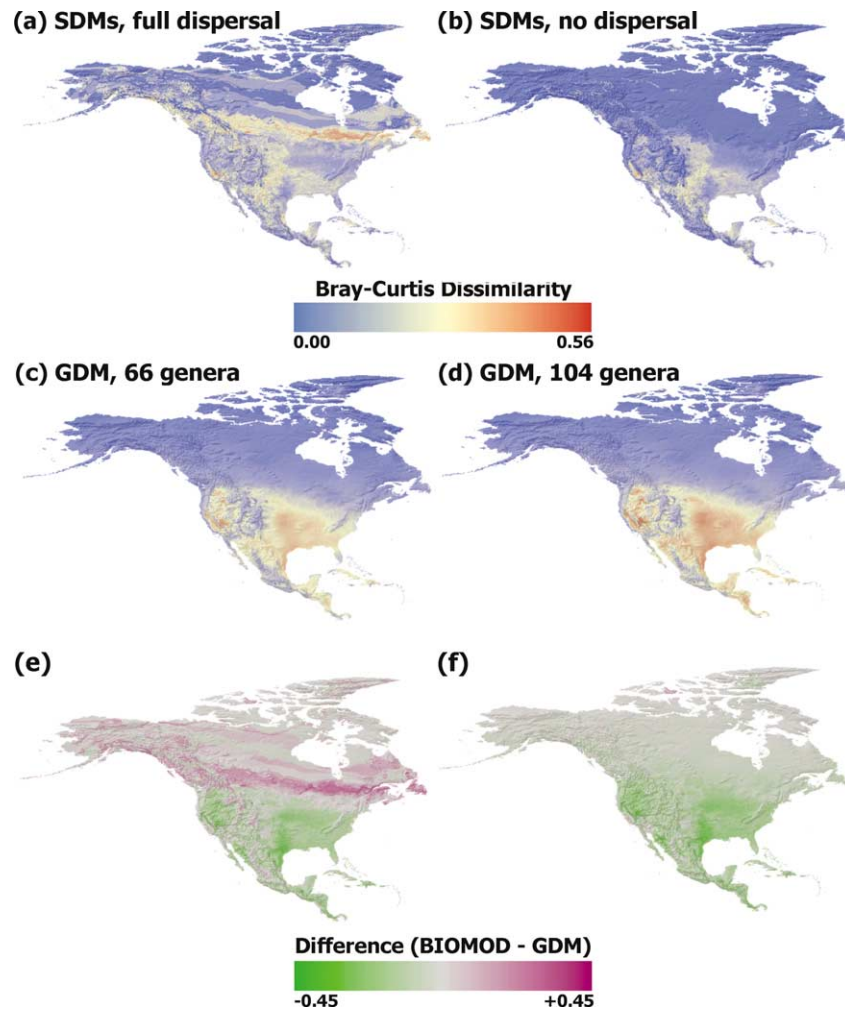


Figure 4. Compositional dissimilarity forecasted between present and 2050 for 66 individual ant genus models using BIOMOD and assuming either (a) full-dispersal or (b) no-dispersal, (c) GDM with the same 66 genera, and (d) GDM with all 104 genera for which distribution data were available. Panels (e) and (f) show the difference (BIOMOD – GDM) between the forecasts from GDM (66 genera model) and BIOMOD assuming full and no-dispersal respectively. Greens/purples indicate where forecasted dissimilarity from BIOMOD was less/greater than that from GDM.

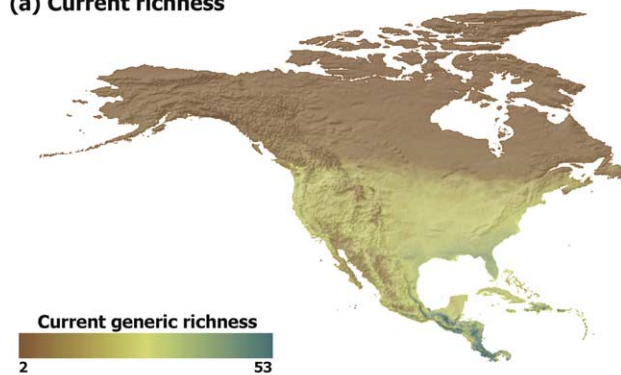
temperature increases in this region are expected to result in relatively large faunal changes. In contrast, SDMs can readily highlight which genera (and their associated functional roles) are expected to move and where. In our case, SDMs suggest that the southern half of North America may tend to experience a net loss in ant genera (Fig. 5), but some areas may gain functional roles (Fig. 6). In other words, the two approaches can be complimentary in terms of what they tell us. It is also important to consider the nature of the Bray–Curtis measure, which can reflect changes in the ant community due to gains or losses of genera as well as composition change with no net change generic richness. For SDMs, it is possible to decipher which of these types of changes the Bray–Curtis measure reflects and for the most part dissimilarity appears to be a result of gains and losses of genera rather than genus-for-genus compositional change (Fig. 4 a, b, Fig. 5). GDM does not readily provide similar insight because it models compositional change directly rather than genus-specific changes.

Although the models agree in some instances about where changes in ant composition may be greatest, there are also notable exceptions to this pattern that highlight

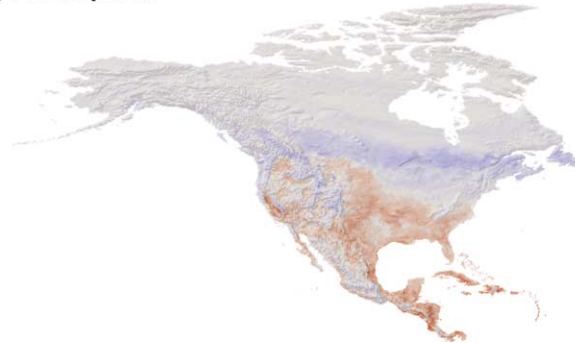
the differences between the two modeling approaches. In the southern half of North America, GDM tended to forecast higher turnover than did SDMs, especially when GDM incorporated all 104 genera rather than only the 66 for which there were enough data for SDMs (Fig. 4). Elsewhere in the study region, notably southern Canada, SDMs tended to predict greater turnover than GDM. Here, generic richness is relatively low (Fig. 5a), and genera are relatively widespread. Thus, overall patterns of biodiversity are driven by wide-ranging genera that can be modeled using SDMs. In the southern half of North America and in Central America in particular, the converse is true; most genera have small ranges, at least within the domain we considered, and as a consequence turnover in space is high relative to the grain of the analysis – the conditions under which the strengths of GDM are likely to be most apparent.

A biogeographical perspective provides further insight as some of the compositional changes forecasted by the models appear to coincide not simply with species range boundaries but also more generally with two of the major biogeographical transitions in the Americas. For example, we see high turnover in northern Mexico that appears to coincide with

(a) Current richness



(b) Full-dispersal



(c) No-dispersal

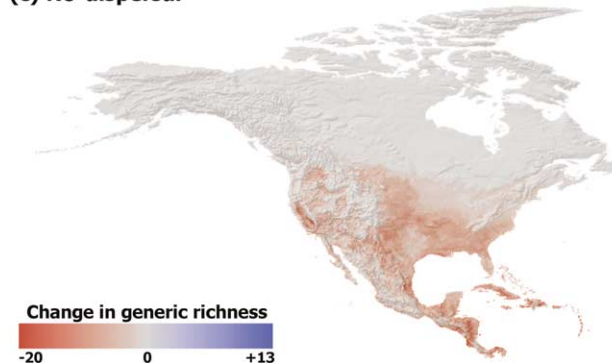


Figure 5. (a) Ant generic richness obtained by stacking individual BIOMOD predictions for each of 66 ant genera under current climate and forecasted change in richness obtained by stacking individual BIOMOD forecasts for year 2050 and assuming either (b) full- or (c) no-dispersal.

the movement of neotropical species northward into temperate regions (and with this shift, consequent increases in the area covered by ‘tropical’ functional groups, such as leaf-cutters) as well as the loss of some temperate species. The climate associated with this transition is relatively warm and dry and therefore the forecasted changes in community composition in these relatively warm areas by GDM partially may be an artifact of ant biogeography rather than ant ecology per se. By the same token, large differences exist between the Nearctic ant fauna and what might be called the Nearctic boreal fauna. It is along the biogeographic transition between these faunas that SDMs forecast high compositional dissimilarity in the north. The existing biogeographic history that underlies these two transitions is,

in other words, rearing up in the context of future predictions.

Our results suggest that GDM possesses an advantage over SDMs particularly at the edge of domains or in areas where knowledge of distributions is limited beyond some geographic point. In our case, SDMs can model both losses and gains of genera only in North America north of Mexico, as it is in this region where compositional change will be manifested by those genera occurring within the domain (excluding human-mediated invasions from distant regions). In contrast, Central America is at the southern edge of the study region. Because the ranges of most genera are predicted to shift northward, in Central America SDMs can essentially only model losses of genera (e.g. Fig. 5b, c) and not the arrival of species from South America. So almost inevitably SDMs forecast less compositional turnover in this region than might otherwise be expected. GDM does not rely on species distributions directly, but rather on patterns of compositional turnover in space. Thus GDM is less limited to sampled environments and therefore can extrapolate beyond the domain under consideration. This is not to say that GDM does such extrapolation perfectly, especially in cases when the relationship between diversity and some environmental variable differs from that which occur in the existing domain. But that it does it at all may prove an advantage not only in instances where domain boundaries are arbitrary rather than ecological, but also when non-analog environments predominate forecasted environmental changes and where SDMs may be limited (Williams and Jackson 2007, Fitzpatrick and Hargrove 2009).

Some of the differences between the SDMs and GDM arose from assumptions regarding dispersal, and these differences tended to be geographically structured. When full-dispersal was assumed for the SDMs, many species were forecasted to shift their ranges northward, resulting in an increase in species richness in Canada (Fig. 5b) and greater predicted turnover in this region than that forecasted by GDM (Fig. 4). When no-dispersal was assumed these gains in species richness disappeared and the forecasted change between SDMs and GDM became more similar (Fig. 4f). This result is surprising given that GDM forecasts assume something akin to full-dispersal. That the SDM and GDM forecasts differ most under full- rather than no-dispersal may be an artifact and may reflect a combination of sparse sampling and low generic richness in Canada (Fig. 1, 5a) which result in correspondingly less influence in a GDM. The use of presence-only data and the impacts of sampling bias generally in the development of GDMs warrants further investigation.

In addition to modeling the response of genera to climate change we also coded genera by their functional roles and in doing so inferred an aspect of the function of ants and how it may change. Modeling the consequences of changes in animal assemblages is complicated by several factors. Perhaps the most interesting of these complications is the two-headed nature of ecological consequences. Many ant genera have relatively unique diets or life histories that predispose them to having disproportionately large impacts when they arrive at new sites or disappear from currently occupied sites. Army ants, for example, prey on ant brood and their absence is likely to have large and ramifying effects

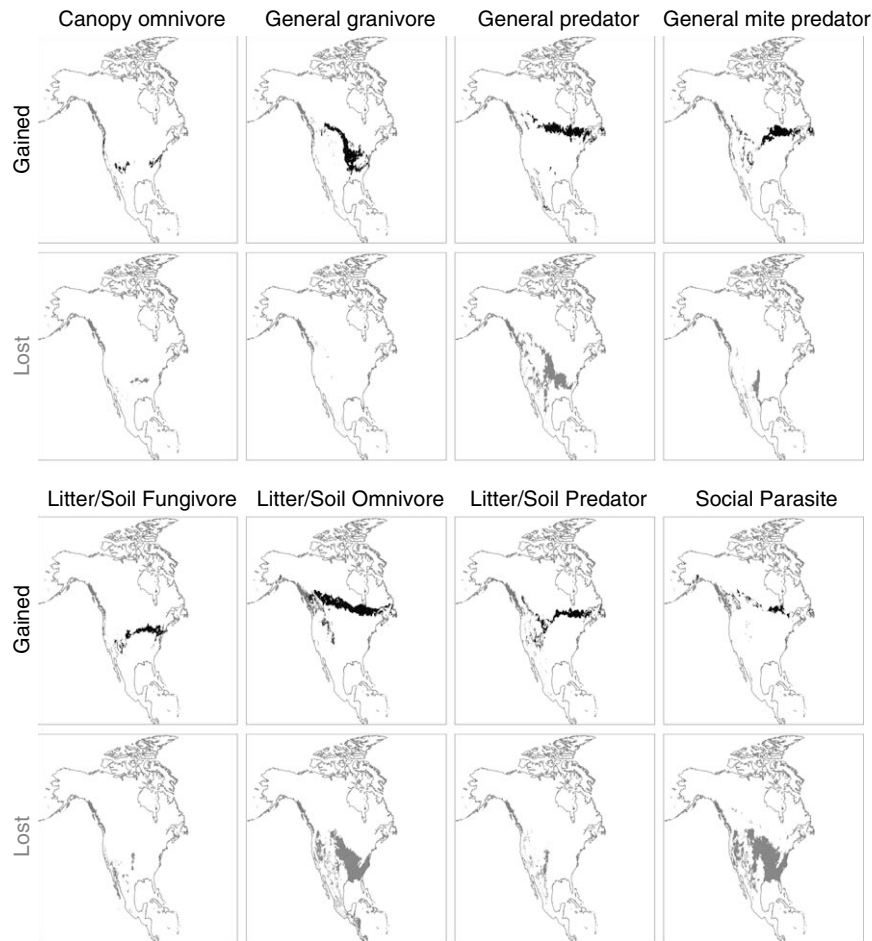


Figure 6. Losses and gains in ant functional groups. Note that these assume full-dispersal and that there would be no gains in functional groups if no-dispersal was assumed as genera can only lose range under this scenario. Only those functional groups that exhibited losses or gains are shown.

on the populations of co-occurring ants, but also other taxa such as social wasps (Franks 1982, Kaspari and O’Donnell 2003). Thus, the addition or loss of ant genera has the potential to have large and cascading consequences (Holway et al. 2002). At the same time, however, the specialized nature of some life histories of ants may mean that some species and genera may not track climate per se, but instead will track climate only to the extent that climate influences the resources upon which they depend. To the extent that this is true, responses of particular ant taxa may have significant time lags. Nonetheless, the movement of several prominent ant functional groups is noteworthy. For example, granivores, such as those species found in the genus *Pogonomyrmex*, are predicted to move northward in eastern North America, with potential consequences for seed dispersal.

Taken together, our results beg the question: which method is the better approach to forecasting likely levels of compositional change in response to climate change – GDM or SDMs? Clearly the answer to this question depends on context. In analyses involving very large numbers of species, where individual responses of species cannot be modeled due to rarity or insufficient sampling or are otherwise not of interest, GDM has clear benefits

to SDMs. We also found evidence that the ability of GDM to extrapolate beyond observed communities can lead to different conclusions than those based on SDMs. However, GDM is largely mute on the potential consequences of forecasted changes. Here, SDMs might provide more insight, but only for relatively common and widespread species. Finally, it bears mentioning the computation benefits in using GDM. Once the data were properly formatted, which requires similar amounts of time for GDM and SDMs, GDM took on the order of seconds to produce results. In contrast, our SDMs, which were admittedly rather computationally intense as compared to some other methods, took on the order of days.

Nonetheless, it remains unclear as to whether GDM or SDMs will provide more reliable results about future changes in biodiversity. Most likely, neither is quite right. Of the few other studies that have compared single-species to community-level models, community-level approaches exhibited model performances that were better (Elith et al. 2006), worse (Baselga and Araujo 2009) or similar to (Leathwick et al. 2006) that of single-species models. Baselga and Araujo (2009) found that individual-based Generalized Linear Models (GLMs) tended to more accurately predict observed distributions (as measured using

the kappa statistic) than did canonical quadratic ordination (CQO), a multi-species model, but the magnitude of these differences was not reported. Although Leathwick et al. (2006) found little difference between Generalized Additive Models and Multivariate Adaptive Regression Splines (MARS), they noted strong performance of MARS for species of low prevalence, reinforcing the benefits of multi-species models found in this study. In all cases, model performance under current climate is not an indication of how well the models will fare under extrapolation to future climates. To compare the ability of single- and multi-species models to extrapolate would require evaluations against observed changes in community composition through time, such as those available from long-term observational databases or from fossils, pollen or otherwise. An exploration of the ability of different SDMs and community-level models to make projections into different regions and times, as well as the incorporation of population and community dynamics into models, will likely lead to the most useful improvements to current methods (Botkin et al. 2007). Nevertheless, our results do suggest that current climatic change has the potential to influence the distribution of ant biodiversity in North America, and possibly any ecological processes many of those species mediate. Models agree that changes in ant community composition may be particularly significant in relatively warm environments, where large transitions in the genera and functional types present may occur.

Acknowledgements – MCF was supported by funding from UMCES. RRD and NJS were supported by a DOE-NICCR grant, DOE-PER grant DE-FG02-08ER64510 and NASA award NNX09AK22G. JL was supported by National Science Foundation grant DEB-0640015 (Project LLAMA). Comments from J. Elith and two anonymous reviewers vastly improved the manuscript. This is UMCES Appalachian Laboratory Scientific Contribution No. 4486.

References

- Allnutt, T. et al. 2008. A method for quantifying biodiversity loss and its application to a 50-year record of deforestation across Madagascar. – *Conserv. Lett.* 1: 173–181.
- Allouche, O. et al. 2006. Assessing the accuracy of species distribution models: prevalence, kappa and the true skill statistic (TSS). – *J. Appl. Ecol.* 43: 1223–1232.
- Araujo, M. B. and Rahbek, C. 2006. How does climate change affect biodiversity? – *Science* 313: 1396–1397.
- Araujo, M. B. and New, M. 2007. Ensemble forecasting of species distributions. – *Trends Ecol. Evol.* 22: 42–47.
- Araujo, M. B. et al. 2005. Reducing uncertainty in projections of extinction risk from climate change. – *Global Ecol. Biogeogr.* 14: 529–538.
- Baselga, A. and Araújo, M. B. 2009. Individualistic vs community modelling of species distributions under climate change. – *Ecography* 32: 55–65.
- Bestelmeyer, B. T. 2000. The trade-off between thermal tolerance and behavioural dominance in a subtropical South American ant community. – *J. Anim. Ecol.* 69: 998–1009.
- Botkin, D. B. et al. 2007. Forecasting the effects of global warming on biodiversity. – *BioScience* 57: 227–236.
- Broennimann, O. et al. 2006. Do geographic distribution, niche property and life form explain plants' vulnerability to global change? – *Global Change Biol.* 12: 1079–1093.
- Canadell, J. G. et al. 2007. Contributions to accelerating atmospheric CO₂ growth from economic activity, carbon intensity, and efficiency of natural sinks. – *Proc. Natl Acad. Sci. USA* 104: 18866–18870.
- Davidson, D. W. 1977. Species diversity and community organization in desert seed-eating ants. – *Ecology* 58: 712–724.
- Davidson, D. W. et al. 2003. Explaining the abundance of ants in lowland tropical rainforest canopies. – *Science* 300: 969–972.
- Dunn, R. R. 2005. Modern insect extinctions, the neglected majority. – *Conserv. Biol.* 19: 1030–1036.
- Dunn, R. R. et al. 2009. Climatic drivers of hemispheric asymmetry in global patterns of ant species richness. – *Ecol. Lett.* 12: 324–333.
- Elith, J. et al. 2006. Novel methods improve prediction of species' distributions from occurrence data. – *Ecography* 29: 129–151.
- Erwin, T. L. 1982. Tropical forests: their richness in Coleoptera and other arthropod species. – *Coleoptera Bull.* 36: 74–75.
- ESRI 2009. ArcMap version 9.3. – Environmental Research Inst.
- Ferrier, S. 2002. Mapping spatial pattern in biodiversity for regional conservation planning: where to from here? – *Syst. Biol.* 51: 331–363.
- Ferrier, S. and Guisan, A. 2006. Spatial modelling of biodiversity at the community level. – *J. Appl. Ecol.* 43: 393–404.
- Ferrier, S. et al. 2002. Extended statistical approaches to modelling spatial pattern in biodiversity in northeast New South Wales. II. Community-level modelling. – *Biodivers. Conserv.* 11: 2309–2338.
- Ferrier, S. et al. 2007. Using generalized dissimilarity modelling to analyse and predict patterns of beta diversity in regional biodiversity assessment. – *Divers. Distrib.* 13: 252–264.
- Fitzpatrick, M. C. and Hargrove, W. W. 2009. The projection of species distribution models and the problem of non-analog climate. – *Biodivers. Conserv.* 18: 2255–2261.
- Fitzpatrick, M. C. et al. 2008. Climate change, plant migration, and range collapse in a global biodiversity hotspot: the *Banksia* (Proteaceae) of Western Australia. – *Global Change Biol.* 14: 1337–1352.
- Folgarait, P. J. 1998. Ant biodiversity and its relationship to ecosystem functioning: a review. – *Biodivers. Conserv.* 7: 1221–1244.
- Franks, N. R. 1982. Social insects in the aftermath of swarm raids of the army ant *Eciton burchelli*. – In: Breed, M. D. et al. (eds), *The biology of social insects*. Westview Press, pp. 275–279.
- Graham, C. H. et al. 2004. New developments in museum-based informatics and applications in biodiversity analysis. – *Trends Ecol. Evol.* 19: 497–503.
- Guisan, A. and Zimmermann, N. E. 2000. Predictive habitat distribution models in ecology. – *Ecol. Model.* 135: 147–186.
- Guisan, A. and Thuiller, W. 2005. Predicting species distribution: offering more than simple habitat models. – *Ecol. Lett.* 8: 993–1009.
- Hamilton, A. J. et al. 2010. Quantifying uncertainty in estimation of tropical arthropod species richness. – *Am. Nat.* 176: 90–95.
- Hijmans, R. J. et al. 2005. Very high resolution interpolated climate surfaces for global land areas. – *Int. J. Climatol.* 25: 1965–1978.
- Hölldobler, B. and Wilson, E. O. 1990. *The ants*. – Harvard Univ. Press.
- Holway, D. A. et al. 2002. The causes and consequences of ant invasions. – *Annu. Rev. Ecol. Syst.* 33: 181–233.
- Kaspary, M. and O' Donnell, S. 2003. High rates of army ant raids in the Neotropics and implications for ant colony and community structure. – *Evol. Ecol. Res.* 5: 933–939.

- Kaspari, M. et al. 2000. Three energy variables predict ant abundance at a geographical scale. – *Proc. R. Soc. B* 267: 485–489.
- Kaspari, M. et al. 2004. Energy gradients and the geographic distribution of local ant diversity. – *Oecologia* 140: 407–413.
- Kearney, M. and Porter, W. 2009. Mechanistic niche modelling: combining physiological and spatial data to predict species' ranges. – *Ecol. Lett.* 12: 334–350.
- Kluge, J. et al. 2006. What drives elevational patterns of diversity? A test of geometric constraints, climate and species pool effects for Pteridophytes on an elevational gradient in Costa Rica. – *Global Ecol. Biogeogr.* 15: 358–371.
- Lawler, J. J. et al. 2009. Projected climate-induced faunal change in the Western Hemisphere. – *Ecology* 90: 588–597.
- Leathwick, J. R. et al. 1996. Predicting changes in the composition of New Zealand's indigenous forests in response to global warming: a modelling approach. – *Environ. Software* 11: 81–90.
- Leathwick, J. R. et al. 2006. Comparative performance of generalized additive models and multivariate adaptive regression splines for statistical modeling of species distributions. – *Ecol. Model.* 199: 188–196.
- Manly, B. F. 1998. Randomization, bootstrap and Monte Carlo methods in biology. – Chapman and Hall.
- Marmion, M. et al. 2009. Evaluation of consensus methods in predictive species distribution modelling. – *Divers. Distrib.* 15: 59–69.
- Menke, S. B. et al. 2010. Is it easy to be urban? Convergent success in urban habitats among lineages of a widespread native ant. – *PLoS One* 5: e9194.
- Myers, N. et al. 2000. Biodiversity hotspots for conservation priorities. – *Nature* 403: 853–858.
- Phillips, S. J. et al. 2009. Sample selection bias and presence-only distribution models: implications for background and pseudo-absence data. – *Ecol. Appl.* 19: 181–197.
- R Development Core Team 2009. R: a language and environment for statistical computing. – R Foundation for Statistical Computing.
- Retana, J. and Cerdá, X. 2000. Patterns of diversity and composition of Mediterranean ground ant communities tracking spatial and temporal variability in the thermal environment. – *Oecologia* 123: 436–444.
- Sanders, N. J. et al. 2003. Patterns of ant species richness along elevational gradients in an arid ecosystem. – *Global Ecol. Biogeogr.* 12: 93–102.
- Sanders, N. J. et al. 2007. Temperature, but not productivity or geometry, predicts elevational diversity gradients in ants across spatial grains. – *Global Ecol. Biogeogr.* 16: 640–649.
- Sørensen, T. 1948. A method of establishing groups of equal amplitude in plant sociology based on similarity of species and its application to analyses of the vegetation on Danish commons. – *K. Dan. Vidensk. Selsk. Biol. Skr.* 5: 1–34.
- Stork, N. 1997. Canopy arthropods. – Chapman and Hall.
- Svenning, J. C. and Skov, F. 2004. Limited filling of the potential range in European tree species. – *Ecol. Lett.* 7: 565–573.
- Thomas, C. D. et al. 2004. Extinction risk from climate change. – *Nature* 427: 145–148.
- Thuiller, W. et al. 2005. Climate change threats to plant diversity in Europe. – *Proc. Natl Acad. Sci. USA* 102: 8245–8250.
- Thuiller, W. et al. 2006. Using niche-based modelling to assess the impact of climate change on tree functional diversity in Europe. – *Divers. Distrib.* 12: 49–60.
- Thuiller, W. et al. 2009. BIOMOD – a platform for ensemble forecasting of species distributions. – *Ecography* 32: 369–373.
- Tuomisto, H. 2010a. A diversity of beta diversities: straightening up a concept gone awry. Part 1. Defining beta diversity as a function of alpha and gamma diversity. – *Ecography* 33: 2–22.
- Tuomisto, H. 2010b. A diversity of beta diversities: straightening up a concept gone awry. Part 2. Quantifying beta diversity and related phenomena. – *Ecography* 33: 23–45.
- Weiser, M. D. and Kaspari, M. 2006. Ecological morphospace of New World ants. – *Ecol. Entomol.* 31: 131–142.
- Williams, J. W. and Jackson, S. T. 2007. Novel climates, no-analog communities, and ecological surprises. – *Front. Ecol. Environ.* 5: 475–482.
- Yates, C. J. et al. 2010. Assessing the impacts of climate change and land transformation on *Banksia* in the South West Australian Floristic Region. – *Divers. Distrib.* 16: 187–201.

Supplementary material (Appendix E6653 at <www.oikos.ekol.lu.se/appendix>). Appendix 1.