Predicting the future of species diversity: macroecological theory, climate change, and direct tests of alternative forecasting methods

Adam C. Algar, Heather M. Kharouba, Eric R. Young and Jeremy T. Kerr

A. C. Algar, H. M. Kharouba, E. R. Young and J. T. Kerr (Jeremy.kerr@uottawa.ca), Dept of Biology, Univ. of Ottawa, 30 Marie Curie, Ottawa, ON K1N 6N5, Canada. (Present address of H. M. K.: Dept of Zoology, Univ. of British Columbia, #2370-6270 Univ. Blvd., Vancouver, BC V6T 1Z4, Canada.)

Accurate predictions of future shifts in species diversity in response to global change are critical if useful conservation strategies are to be developed. The most widely used prediction method is to model individual species niches from point observations and project these models forward using future climate scenarios. The resulting changes in individual ranges are then summed to predict diversity changes; multiple models can be combined to produce ensemble forecasts. Predictions based on environment-richness regressions are rarer. However, richness regression models, based on macroecological diversity theory, have a long track record of making reliable spatial predictions of diversity patterns. If these empirical theories capture true functional relationships between environment and diversity, then they should make consistent predictions through time as well as space and could complement individual species-based predictions. Here, we use climate change throughout the 20th century to directly test the ability of these different approaches to predict shifts of Canadian butterfly diversity. We found that all approaches performed reasonably well, but the most accurate predictions were made using the single best richness-environment regression model, after accounting for the effects of spatial autocorrelation. Spatially trained regression models based on macroecological theory accurately predict diversity shifts for large species assemblages. Global changes provide pseudo-experimental tests of those macroecological theories that can then generate robust predictions of future conditions.

How will species diversity change in the future? The pervasiveness of anthropogenic climate change impacts on the world's physical and biological systems (Rosenzweig et al. 2008) makes this a critical question for scientists, conservationists, and decision makers. Recent climate and land-use changes have been followed by pronounced biological responses, including phenological changes (Root et al. 2003), range shifts (Parmesan et al. 1999, Root et al. 2003), and increasing diversity of generalist species while more specialized species decline (Menendez et al. 2006, White and Kerr 2006, 2007). These responses are expected to continue, plausibly causing rapid acceleration of extinction rates (Thomas et al. 2004), as climate and land-uses continue to change.

A substantial effort has been made to predict the response of biological systems to global change, particularly with respect to species diversity (Peterson et al. 2004, Thuiller et al. 2005, McKenney et al. 2007). However, different modeling methods yield highly divergent predictions, even when spatial assessments of model accuracy appear excellent (Araújo et al. 2005b, Kharouba et al. in press). Predicting future is challenging: there is no way to determine which

projection is correct and which is not. Given our inability to distinguish reliably between divergent predictions of future states (except by waiting), datasets that combine both historical and more recent data on environmental conditions and species distributions are indispensable for two reasons. First, historically calibrated models, whose performance can be evaluated through time before projecting into the future, are more likely to provide accurate predictions of global change (White and Kerr 2006, Willis and Birks 2006, Kerr et al. 2007, Nogués-Bravo et al. 2008). Second, such datasets allow for the direct evaluation and comparison of alternative predictive methods (Araújo et al. 2005a).

Spatial models of species' niches can be used to predict where species will shift in response to changing environmental conditions (McKenney et al. 2007). These models estimate species' niches across geographical space by linking observations of species' presences (and sometimes absences) to environmental or biotic conditions. To translate individual species' niche models into predictions of how species will respond to anticipated environmental change, new environmental data are substituted into the niche models to predict how species' distribution will change. These

distributions are then summed to produce predicted species richness.

An alternative approach to predicting diversity shifts could build on established macroecological hypotheses predicting species richness. Macroecologists have devoted decades of study to spatial patterns of diversity (Fischer 1960, Currie 1991, Hawkins et al. 2003). This research has lead to the development of empirical theory that predicts the global distribution of species diversity by proposing a direct link between species richness and a small number of climatic (or related) variables; in some cases, heterogeneity in the form of land-cover variety (Kerr et al. 2001) or elevation range (Kerr and Packer 1997), has also proven important. Strong environment-richness relationships have been found for nearly all taxa in which they have been investigated, including vertebrates (endothermic and ectothermic), invertebrates, and multiple plant taxa (Hawkins et al. 2003). Macroecological models reported in such studies not only explain substantial amounts of variation in species richness (R^2 values typically range from 0.6 to 0.9) within their respective study areas, but they can predict patterns in independent geographical regions (Francis and Currie 2003), though in some cases such models may have regional limitations (Whittaker et al. 2007). By applying the coefficients from one of these multiple regression models to future climate scenarios, predictions of future diversity shifts may be possible (Dormann et al. 2008).

Global change provides pseudo-experimental opportunities to test macroecological hypotheses, which often rely on exclusively spatial data (Kerr et al. 2007). One example is the contemporary climate hypothesis for broad-scale richness gradients. However, a direct relationship between contemporary climate and species richness is only one hypothesis that has been proposed for richness gradients. Alternative hypotheses include evolutionary history (Ricklefs 2007), and historical climate effects (Svenning and Skov 2007, Araújo et al. 2008). Here, we focus on the contemporary climate hypotheses for two reasons. One, the alternative hypotheses, in their current forms, do not make explicit predictions of richness change over the time scale (<100 yr) we consider here. Two, the contemporary climate approach has already been used to predict future diversity shifts (Dormann et al. 2008). Thus, while our primary goal is to evaluate the capacity of climate-richness regressions to predict observed diversity shifts, our study also provides a temporal test of the contemporary-climate hypothesis. If the observed links between environmental characteristics and species richness are causal, then spatial models based on macroecological theory should accurately predict temporal species richness trends as environment changes (White and Kerr 2006). Hypotheses that fail to predict temporal changes should be rejected, while those making successful temporal predictions may prove valuable in improving decisions intended to ameliorate biological impacts of global change (Kerr et al. 2007).

Relying on a single, spatial model can produce incorrect predictions of future diversity shifts (Araújo et al. 2005b, Dormann et al. 2008). Predictions from such "single-best" models are more sensitive to the assumptions and particular characteristics of the modeling method selected and differences among methods can degrade the reliability of niche predictions through time (Thuiller 2003, Beaumont et al. 2007). To combat this, some authors have suggested that diversity predictions should be made on the basis of predictions of ensembles of different models (Araújo et al. 2005b, Araújo and New 2007, Dormann et al. 2008). Ensemble forecasting avoids the requirement to select a single, best model from among many approximately equivalent models (Dormann et al. 2008). It also avoids potential errors due to over-fitting and can capture different components of the true basis for species' presences, parts of which may be represented in different models (Araújo and New 2007). For example, ensemble forecasting better predicted individual species ranges of British birds than a single-best model approach (Araújo et al. 2005b). However, it is unknown whether ensemble forecasting produces more reliable predictions than those based on macroecological theory that produces spatio-temporally consistent predictions of species diversity.

This paper has two purposes. First, we consider the potential of empirically-derived, macroecological models to predict diversity shifts in response to changing climate and test their performance against more widely used niche models. For the latter, we assess both single-best model and ensemble forecasting approaches. Second, by using historical and recent data on Canadian butterfly richness and climate change from throughout the 20th century, we provide a rare, pseudo-experimental test of whether spatial macroecological models also predict temporal changes in species diversity, as they should if those spatial models are causal, not merely correlative.

Methods

Species distribution models

Species distributions were modeled using occurrence records taken from the Canadian National Collection of Butterflies, which contains ca 300 000 georeferenced, dated records for 297 Canadian butterfly species (Layberry et al. 1998, Kerr et al. 2001, White and Kerr 2006, Kharouba et al. in press). Observation points were divided into two time periods: 1900–1930 (historical) and 1960–1990 (recent). Species with fewer than 10 geographically distinct records in either time period were excluded based on modeling accuracy concerns (Hernandez et al. 2006), leaving 139 species for analysis. Data were corrected for increased collection intensity in the recent time period. For each species, a random subset of occurrence points was chosen from the recent time period on which to train and test our models. Thus, for each species, the number of presence points was invariant with respect to time period (for details on additional data corrections, and a list of included species, see Kharouba et al. in press).

We used seven environmental variables in our distribution models. These included four climate variables, averaged across each time period (historical or recent): mean growing season (April–October) temperature, maximum growing season temperature, minimum annual temperature, and mean annual precipitation (Mekis and Hogg 1999, McKenney et al. 2001, McKenney pers. comm.). We also included human population density, measured in the

1921 and 1981 censuses (White and Kerr 2006), elevation, and land-cover data delineating generalized physiographic differences in Canadian vegetation and the boundaries of the major agricultural regions in Canada (Beaubien et al. 2000). Climate and human population density data were available for both time periods and elevation, obviously, was constant between the two. Land-cover data (included as a categorical variable) were not available for the historical period so data for the current time period were used to identify major habitat divisions and ecotones within Canada, which have remained relatively consistent during the 20th century (e.g. transition between prairies and aspen woodlands or broadleaf to coniferous forests). Areas where current land-cover data misrepresent natural variability (e.g. southern Ontario) were converted to human land-uses long before the beginning of the earliest time period in this study (Ramankutty and Foley 1999, Kerr and Cihlar 2003). Change in climate and human population density between the two time periods are shown in Fig. 1 and rely on data from Mekis and Hogg (1999), McKenney et al. (2001) and McKenney (pers. comm.). Species distributions were modeled at a resolution of 6.67 × 6.67 km, which is the

effective resolution of the historical climate data available in Canada (McKenney pers. comm.) but also sufficiently coarse to ensure sub-pixel georeferencing accuracy among all or nearly all butterfly observations.

Maximum entropy (Maxent) was used to model species distributions (Phillips et al. 2006), using the default settings. Maxent was developed specifically for use with presence-only occurrence data and consistently performs well relative to other methods (Elith et al. 2006). To build each model, the occurrence records were randomly partitioned into a training (70%) and testing set (30%). This process was repeated to produce 10 model repetitions for each species. Historical and recent distributions were modeled independently (see Kharouba et al. in press for details).

Historical and recent species richness

Our measures of species richness were limited to the 139 (predominantly common, widespread) species for which we had sufficient data. To determine past and recent species
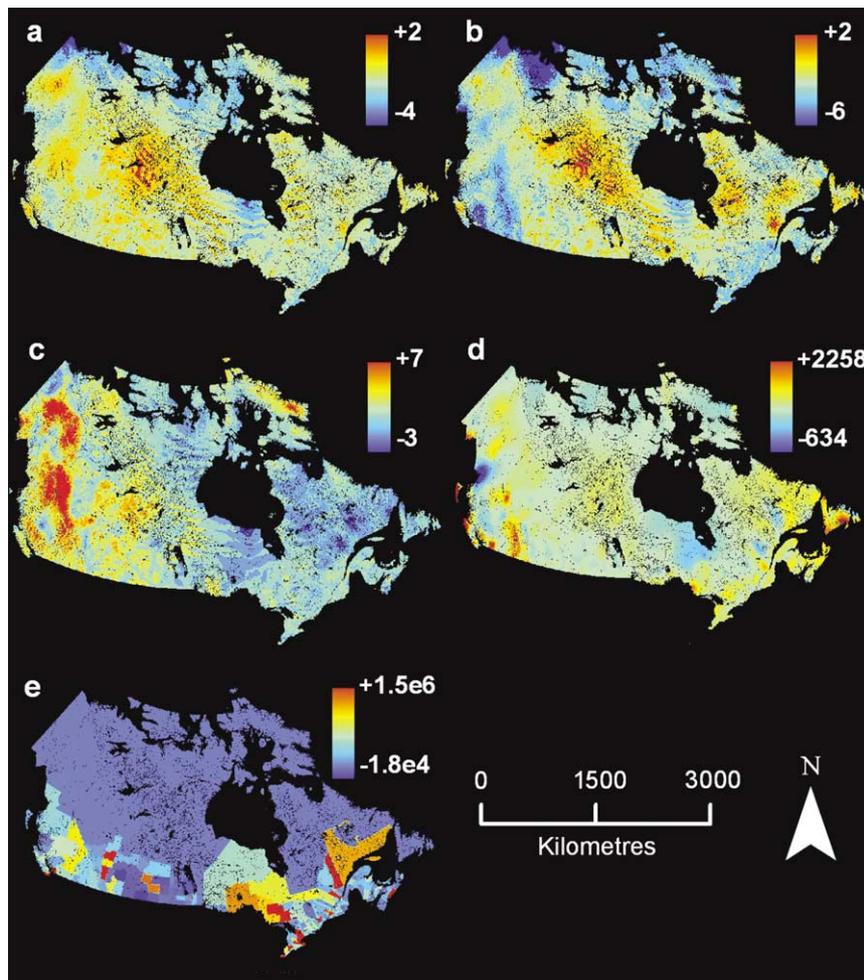

Figure 1. Climate and human population change in Canada from the historical (1900–1930) to recent (1960–1990) time periods. Difference maps were created by subtracting historical from recent conditions. Maps are for mean growing season temperature (a), maximum growing season temperature (b), minimum yearly temperature (c), mean annual precipitation (d), and human population density (e). Data are from a variety of sources: Mekis and Hogg (1999), McKenney et al. (2001), McKenney (pers. comm.), and White and Kerr (2006).

richness, we first converted each species' Maxent suitability map, which ranged from 0 indicating low suitability to 100 for high suitability, into a binary range. For each species, in each time period, a decision threshold was defined, above which the species was considered present and below which it was considered absent. The threshold was calculated as the lowest average suitability (across ten repetitions) of a true presence point (Liu et al. 2005, Pearson et al. 2007). Therefore, for each final model, the binary range included all pixels that were predicted to be at least as suitable as those where a species' had actually been observed (Pearson et al. 2007). This approach results in model failures being limited to commission errors (i.e. false positives). For each time period, species richness was measured by overlaying species' binary ranges on a grid system covering Canada composed of 100×100 km equal area quadrats. Following conventional macroecological techniques, species were considered present in a quadrat if their range overlapped it. While this can result in some false presences (when ranges barely overlap a quadrat), sampling at a 100×100 km reduces effects of single, or small groups of pixels with unrealistically high levels of climate change that may result from data reconstruction methods (e.g. Fig. 1). Quadrats with $<50\%$ land area were excluded to reduce potential area effects leaving 793 quadrats for analysis.

Predicting diversity shifts

We used two main approaches to predict how butterfly species richness would change from the historical to the recent time period. The individual niche model approach considered additive effects of individual species range shifts. The empirical diversity theory approach evaluated species richness as a whole by developing richness-climate regression models. Both methods, and their variants, are described in more detail below. A summary of prediction methods are given in Table 1 for quick reference.

The individual niche model approach
We made individual range-based predictions of species richness in the recent time period by projecting the niche model for each species forward through time by substituting the recent environmental data into the models created for the historical time period. The projected ranges were then overlaid on our quadrat system and summed to produce predicted species richness. For each species, as described above, ten repetitions of the historical models were run using unique test (30%) and training (70%) sets. These produced ten projected ranges for each species. Using these projected ranges, we implemented three alternate forecasting methods: single-best, committee average, and consensus.

We implemented the single-best approach by identifying, for each species, the model repetition with the greatest area under the curve (AUC) of the receiver operating characteristic. AUC is a commonly used measure of niche model accuracy (Fielding and Bell 1997, Elith et al. 2006). However, recent work has highlighted its drawbacks (Lobo et al. 2008). Any statistic used to select the best model has potential disadvantages, including R^2, AIC, BIC, and AUC. We recognize that single-best models could be chosen using alternate methods such as maximizing predictions of true presences (sensitivity). Nevertheless, AUC represents a well known method for distinguishing between models. Future research evaluating alternative ''single-best'' statistics by testing their selected models' ability to predict temporal changes is an area of interesting future research. Thresholds

Table 1. Summary of predictive approaches compared in this study.

	Approach	Description
Empirical diversity theory approaches	Best-OLS	Fits a series richness-environment regression to historical data; The best of these is the one with the highest R^2. These coefficients are then used to predict richness in recent time period.
	Best-CAR	Fits a richness-environment regression to historical data using the same environmental variables as the best-OLS. Conditional autoregression is used to estimate coefficients to account for the influence of spatial autocorrelation.
	Committee average	Fits a large number of richness-environment regressions to historical data. Identifies a candidate set of models with adjusted-R^2 values within 0.05 of the best model. The committee average is the mean prediction across all models in the candidate set.
	Consensus forecast	A principal components analysis is performed on the predictions of all models in the candidate set (defined above). The model with the highest loading of PC1 is identified and its predictions are considered to represent the consensus among all candidate models.
Individual niche model approaches	Committee average	For each species, ten repetitions of the historical Maxent niche model were projected forward. The mean of these projections was converted to a binary range for each species and the ranges summed to produce richness predictions.
	Single-best	For each species, the historical Maxent model repetition with the highest area under the curve (AUC) was projected forward and converted to a binary range. Ranges were then summed across all species.
	Consensus forecast	For each species, a principal components analysis was run on the 10 Maxent model projections and the repetition with the highest loading was converted to a binary range. Richness was calculated as described above.

for conversion from a continuous cumulative distribution map to a binary range were performed as above, but were based only on the single best repetition for each species. The best projections for all species were then summed, after binary conversion, to predict richness.

Committee averaging and consensus forecasting both incorporate information from all ten repetitions of the projected range for each species. The committee average prediction considered all candidate models equally, while the consensus forecast best captured the common trend among the entire set of predictions. The committee average prediction was calculated by taking the mean projected distribution across all ten repetitions for each species. Following Araújo et al. (2005b), to determine the consensus forecast, we performed a principal component analysis on each species' ten projected distributions. Treating species separately, we then identified the repetition that had the highest loading on the first principal component. As before, after identifying the appropriate projection, continuous maps were converted to binary ranges and summed to predict species richness.

The empirical diversity theory approach

To predict recent species richness using established macroecological diversity theory, we constructed richness-environment regression models using only the historical data. Regression models potentially included any or all of the environmental variables for which we had both historical and recent data. All of these variables (or extremely similar ones) have been examined as predictors of diversity in the macroecological literature. Mean climatic and human population density values were extracted for each quadrat. Land-cover and elevation were converted to measures of heterogeneity by sampling the number of land-cover types (Kerr et al. 2001) and the elevation range per quadrat (Kerr and Packer 1997). Additionally, we also considered quadratic terms and all first order interactions, resulting in 35 possible independent variables (see Dormann et al. 2008 for a similar approach).

To potentially be included in the candidate model set for ensemble forecasting, regression models had to meet several criteria. First, since macroecological theory indicates that relatively simple models can predict diversity (Francis and Currie 2003, Buckley and Jetz 2007), we limited the maximum number of variables in a model to five plus an intercept (see Dormann et al. 2008 for a similar decision). Second, due to high collinearity, mean growing season temperature and maximum growing season temperature

were forbidden from occurring in the same regression model (Table 2). Lastly, if models included quadratic or interaction terms, the matching linear terms had to be included.

One method of identifying the candidate model set is to consider all models that are indistinguishable with respect to their Bayesian information criteria (BIC; Hoeting et al. 1999). For our dataset, BIC did not identify a useful set of candidate models. The model with the lowest BIC always had a substantially (>10) lower BIC than all other models, limiting our set to one. However, these lower BIC values often corresponded to changes in adjusted R^2 of <0.01; a difference that we considered biologically insignificant. The large differences in BIC also meant that Bayesian model averaging (Hoeting et al. 1999) was not appropriate.

We determined our set of candidate models based on adjusted R^2. R^2 has long been the benchmark for model selection in macroecology. To identify the candidate set, we used the leaps function (Lumley and Miller 2007) in R 2.6.0 (R Development Core Team 2007) to find, at most, the best 10 000 regression models of each possible size. This process was repeated once with maximum growing season temperature included in the set of predictors, and once with mean growing season temperature included. The two sets of models were then combined. Redundant models, and those that did not meet the other criteria described above were then culled. All remaining models with adjusted R^2 values within 0.05 of the best model were included in the candidate set. While this criterion is arbitrary, it identifies a justifiable set of relatively simple models that explain nearly equivalent amounts of variation in geographical patterns of species richness.

The model with the highest adjusted R^2 was chosen as the single-best model using ordinary least squares regression (OLS). However, it is well known that spatial autocorrelation in model residuals can bias OLS regression (Lennon 2000, Kühn 2007). Thus, we also fitted a conditional autoregressive model (best-CAR) that included the same variables as the best-OLS model. For the CAR, the spatial weights matrix was an inverse function of distance between quadrat centres ($w_{ij} = d_{ij}^{-\alpha}$). Three weighting schemes were considered: $\alpha = 1.0$, $\alpha = 1.5$, and $\alpha = 2.0$. In all schemes, distances greater than the distance at which Moran's I was non-significant in OLS residuals were given weights of zero (Lichstein et al. 2002). The weighting scheme that resulted in the lowest level of autocorrelation in resulting model residuals was retained. The CAR was calculated using the

Table 2. Correlation matrix of variables used in regression analysis from the historical time period. Tmaxgr is maximum growing season temperature, Tavggr is mean growing season temperature, Tmin is minimum yearly temperature, Pcpyr is mean annual precipitation, Popn is human population densisty, Erange is elevation range and Vgtvar is the number of land-cover classes.

	Tmaxgr	Tavggr	Tmin	Pcpyr	Popn	Erange
Tmaxgr						
Tavggr	0.991					
Tmin	0.953	0.930				
Pcpyr	0.581	0.543	0.732			
Popn	0.570	0.570	0.556	0.361		
Erange	−0.190	−0.117	−0.105	0.131	−0.109	
Vgtvar	0.531	0.554	0.523	0.429	0.010	0.271

spdep package (Bivand et al. 2008) in R 2.6.0 (R Development Core Team 2007).

The committee average and consensus forecasts were determined analogously to the procedure used for the individual approach. The committee average was the mean predicted richness across all candidate models. The consensus model was identified by its loading on the first principal component of a PCA on all candidate predictions.

We considered one final predictive model, based on established empirical macroecological theory. This minimum macroecological model included only three variables (plus intercept): mean growing season temperature, mean annual precipitation and their interaction. Similar models have been shown to predict species richness globally (Francis and Currie 2003, Buckley and Jetz 2007), suggesting that these variables influence species richness and are not merely collinear with unidentified, true causes.

Evaluating predictive methods

We directly selected among approaches by comparing their predictions of richness in the recent time period to the actual pattern by regressing observed recent richness (dependent variable) against predicted recent richness (independent variable), following Piñeiro et al. (2008). We considered four criteria for evaluation. 1) Accuracy; evaluated by determining if the slope and intercept of observed-predicted regression were close to one and zero, respectively. 2) Precision; judged by the correlation (Pearson's r) between observed and predicted richness. 3) Linearity; determined by including a quadratic term in the regression of observed versus predicted richness values; the p-value of this term was used as a continuous index. 4) Prediction consistency; determined by evaluating the homoscedasticity of observed-predicted residuals by regressing the square of the residuals against the predicted values. Finally, predictive methods were ranked based on each of these criteria and the ranks summed across all criteria to produce an integrated measure of performance. Lower ranksums indicated stronger overall performance. We also compared ranks using a Kruskal-Wallis test.

Results

Species richness patterns

In both periods, species richness generally increased from north to south (Fig. 2), consistent with other studies of butterflies in Canada (Kerr 2001, Kerr et al. 2001, White and Kerr 2006). Generally, species richness increased throughout the 20th century in more northerly areas and decreased at lower latitudes (Fig. 2).

Niche and regression models

The mean AUC values across all species (based on the mean of 10 repetitions for each species) in the historical and recent time periods were 0.94 and 0.92 respectively. The mean of the maximum AUC values observed among model runs (i.e. that were used for single-best prediction) across all
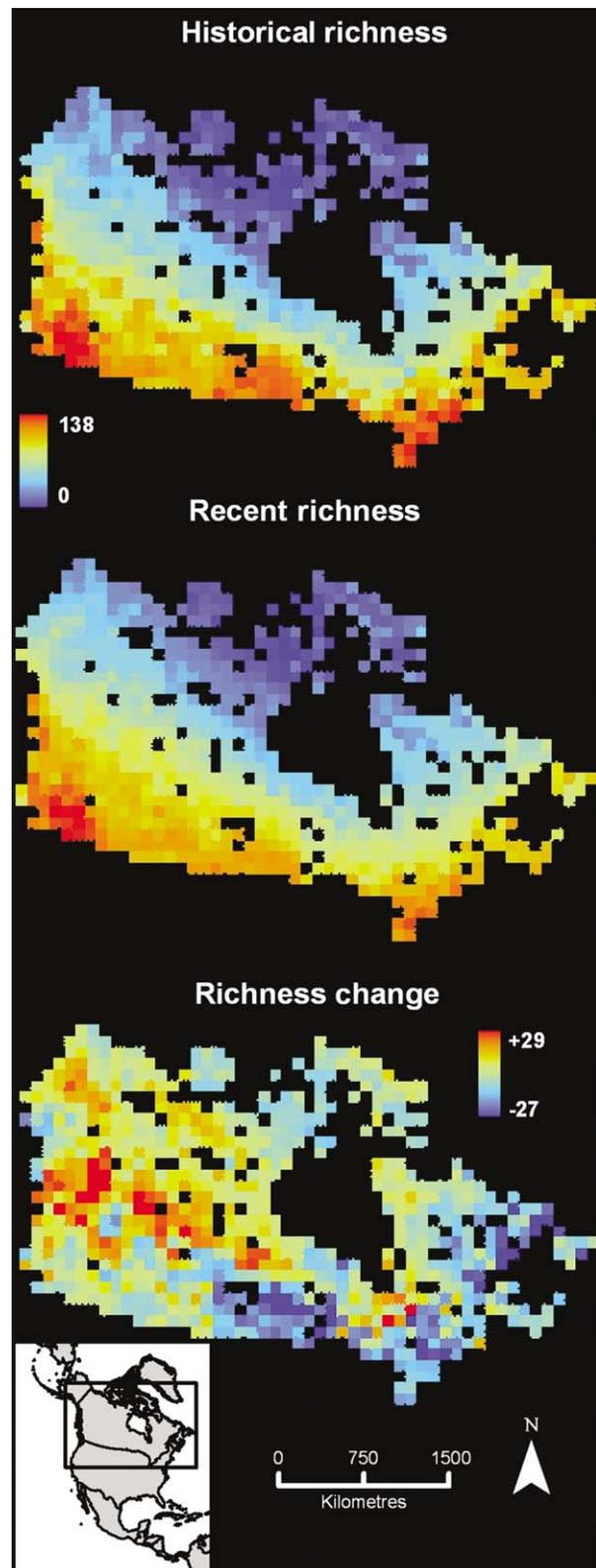

Figure 2. Patterns and change in butterfly species richness throughout Canada during the 20th century for the 139 species used in our analysis. In the historical (1900–1930) and recent time periods (1960–1990), reds represent high diversity and blues low diversity. For richness change, reds indicate increases in diversity, blues indicate decreases. Richness was measured using 100×100 km quadrats.

species from the historical time period was 0.97. Due to the recent identification of limitations of using AUC to evaluate model performance (Lobo et al. 2008), we have presented these values for descriptive purposes only.

The best regression model (highest adjusted R^2) was:

$$S = 1018.0 - 0.82(\text{Tmaxgr}) + 0.15(\text{Tmaxgr})^2$$
$$+ 4.33(\text{Tmin}) - 0.67(\text{Erange}) - 0.011(\text{Erange})^2$$

where S is species richness, Tmaxgr is maximum growing season temperature, Tmin is minimum yearly temperature, and Erange is elevation range. This model explained 91% of the geographical variation in historical species richness. The best-CAR was:

$$S = 722.0 - 0.61(\text{Tmaxgr}) + 0.11(\text{Tmaxgr})^2$$
$$+ 3.74(\text{Tmin}) - 0.49(\text{Erange}) - 0.0090(\text{Erange})^2$$

Our model selection procedure identified a set of 165 models with adjusted R^2 values within 0.05 of the best OLS model. The set of candidate models did not include the minimum macroecological model (adjusted $R^2 = 0.82$).

Comparing prediction methods

Accuracy and precision

All approaches predicted recent species richness relatively accurately; intercepts and slopes of the observed vs predicted richness regressions ranged from -0.9 to 4.3 and 0.9 to 1.0 respectively (Table 3). The single best and consensus niche model predictions had accurate predictions in areas with low richness (intercepts not significantly different from zero, $p > 0.05$), but displayed a tendency to over-predict richness in highly diverse areas (slopes significantly less than unity, $p < 0.001$, Fig. 3). With the exception of the best-CAR, the remaining approaches had a tendency to under-predict richness in species-poor areas (intercepts greater than zero) but their predictive accuracy improved as richness increased (Fig. 3). However, the best-CAR clearly outperformed all other approaches with slope and intercept values that were remarkably close to one and zero, respectively (intercept $\pm$ SE $= 0.22 \pm 0.64$, $H_0 = 0$, $p > 0.7$; slope $\pm$ SE $= 1.007 \pm 0.0087$, $H_0 = 1$, $p > 0.2$).

All approaches' predictions estimated recent species richness precisely (Table 3). Niche model predictions were more precise than regression predictions, though the improvement over the best-CAR approach was marginal. The minimum macroecological model had the lowest correlation, though it was still high ($r = 0.91$). This is not surprising since it was not included in the candidate set of models due to its lower R^2 value in the historical period.

Linearity and homoscedasticity

The best-CAR and minimum macroecological model were the only approaches without a statistically significant quadratic term in the observed vs predicted regression, though the single-best niche model prediction had a quadratic term at the edge of statistical significance ($p \approx 0.04$). Curvilinearity indicates that predictive accuracy is conditional on the diversity of the area in which the predictions are made, i.e. the slope of the observed vs predicted regression moves away from one in more (or less) diverse areas.

None of the observed vs predicted regressions had homoscedastic residuals (Table 3). The individual niche model approaches outperformed the empirical diversity theory approaches, including the best-CAR (though best-CAR performed relatively well with a residuals^2 vs predicted richness slope of 0.98). The niche model committee-averaged prediction had the slope closest to zero (0.18; Table 3). All approaches had larger residuals in highly diverse areas, indicating that predictions become more variable as richness increased.

Overall performance

The best-CAR approach performed well in every evaluation criterion (Table 3) and ranked best overall. The individual niche model approach with the best overall score was the single-best, though it scored poorly with respect to slope; the committee average had a slope much closer to unity. The approach with the worst overall performance was the minimum macroecological model. A Kruskal-Wallis test identified differences among approaches that were of borderline statistical significance (Kruskal-Wallis $\chi^2 = 13.8$, DF $= 7$, $p = 0.054$), despite very limited statistical power to detect differences.

Table 3. Comparison of multiple approaches of predicting diversity shifts. Slopes and intercepts are from observed-predicted richness regressions; Pearson's r is the correlation between observed and predicted richness, p-value quadratic is the p-value of including a predicted^2 term in observed-predicted regressions. Trend in residuals is the slope of the residuals^2-predicted regression, where residuals are from the linear observed-predicted richness linear regression. Bracketed terms are performance rankings relative to the other approaches.

	Approach	Slope[†]	Intercept[‡]	Pearson's r	p-value quadratic[‡]	Trend in residuals[‡]	Rank sums
Empirical diversity theory approaches	Best-OLS	0.96*** (4)	4.29*** (5)	0.96*** (5)	$<10^{-8}$ (6)	1.14*** (5)	25
	Best-CAR	1.01 (1)	0.23 (1)	0.97*** (4)	0.74 (1)	0.99*** (4)	11
	Committee average	0.97* (3)	3.22*** (4)	0.95*** (7)	0.08 (3)	1.39*** (6)	23
	Consensus forecast	0.95*** (5)	4.66*** (6)	0.95*** (6)	$<10^{-10}$ (7)	1.49*** (7)	31
	Min. macro. model	0.92*** (7)	5.34*** (8)	0.91*** (8)	0.12 (2)	1.60*** (8)	33
Individual niche model approaches	Committee average	0.98* (2)	4.95*** (7)	0.99*** (3)	$<10^{-29}$ (8)	0.19*** (1)	21
	Single-best	0.93*** (6)	–0.56 (2)	0.99*** (1)	0.04 (4)	0.44*** (3)	16
	Consensus forecast	0.92*** (8)	–0.86 (3)	0.99*** (2)	$<10^{-7}$ (5)	0.44*** (2)	20

[†]$H_0 = 1.0$, [‡]$H_0 = 0.0$
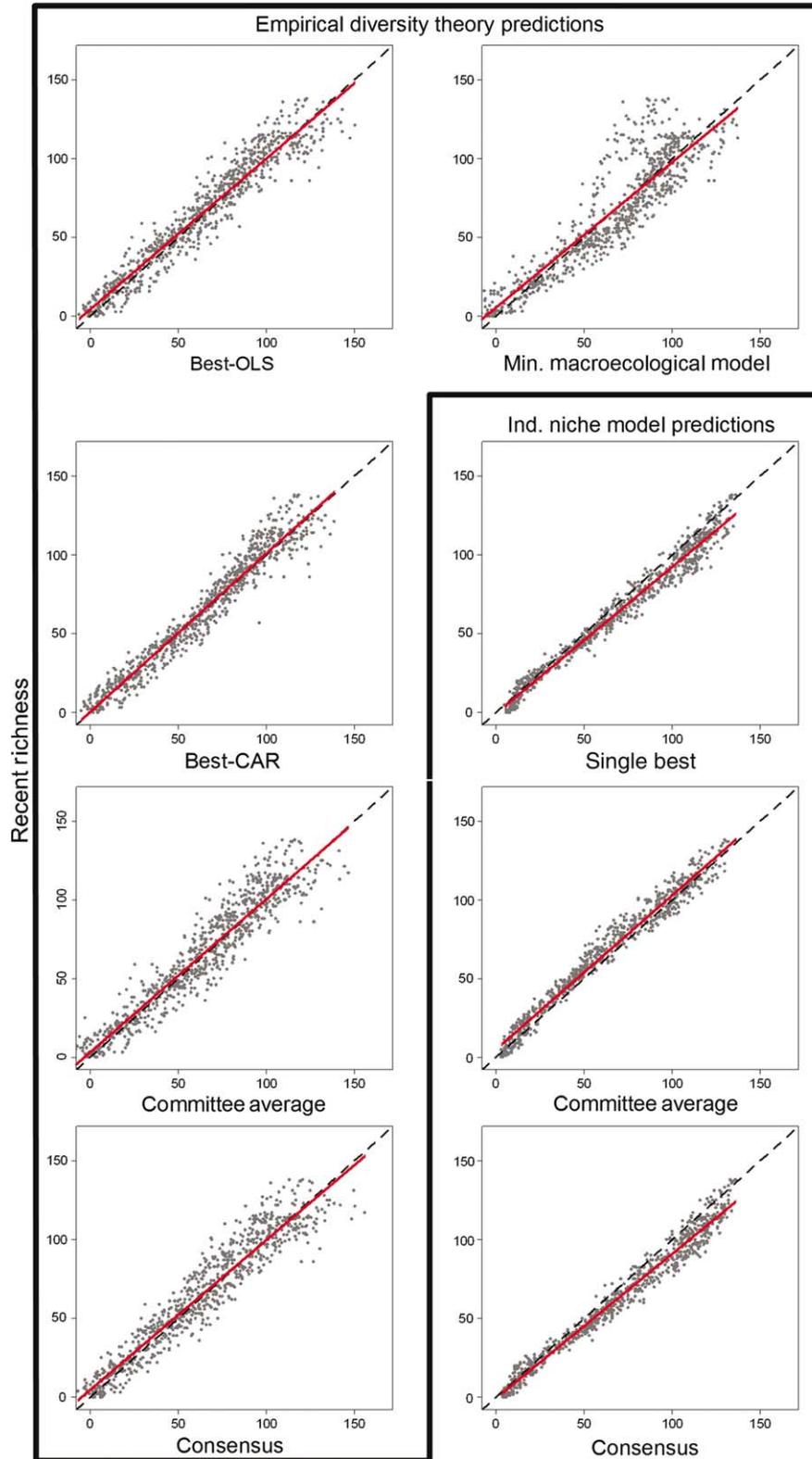*$p < 0.05$, **$p < 0.01$, ***$p < 0.001$

Figure 3. Observed vs predicted richness in the recent time period for various predictive approaches. Predictions are on the x-axis. See Table 1 for methods description. Models were trained on data on butterfly species richness in the early 20th century (1900–1930) and predictions of richness in the late 1900s (1960–1990) were generated based on observed environmental changes. Observed richness is the observed richness in the late 20th century. The red-line is the OLS regression line and the black line is the 1:1 line representing perfect agreement between expectations and observations. OLS is ordinary-least-squares and CAR is conditional autoregression.

Discussion

Species responses to recent global change have already been documented (Root et al. 2003, Kharouba et al. in press). With more climate changes predicted for the future (IPCC 2007), the need for reliable predictions of biological responses to global change is likely to become more acute. Different approaches to building such predictions are possible (Table 1). Here, we compared the ability of these approaches to predict shifts in species richness through time. We found that the most reliable predictions were made by a relatively simple regression model, after incorporating spatial autocorrelation. This approach, based on established macroecological diversity theory, outperformed the more common method of summing projected changes of individual species ranges.

If empirical macroecological theories of diversity represent actual functional relationships between environment and species richness, then they should successfully predict changes through time as well as space (H-Acevedo and Currie 2003, Kerr et al. 2007). Thus, global climate and land-use changes comprise a pseudo-experimental opportunity to test these hypotheses (Kerr et al. 2007). Macroecologists have long suggested that relatively simple environmental-richness relationships underlie broad-scale diversity gradients (Currie 1991, Hawkins et al. 2003). The temporal consistency of the best-CAR model's predictions is consistent with this explanation. However, as we have not explicitly tested temporal predictions of other macroecological hypotheses for richness gradients (e.g. historical climate; evolutionary history), we cannot reject these alternatives in favour of the contemporary climate hypothesis. Nevertheless, our results suggest that, at least for Canadian butterflies, diversity shifts can be reliably predicted using a relatively simple model based on links between climate, heterogeneity and species diversity. Further work using different taxa in different geographical domains will determine the generality of this result.

Our niche and regression models are both correlative approaches to modeling diversity shifts. An alternative approach, not examined here, is to generate a priori, process-based or mechanistic models predicting each species' distribution, and to aggregate these models to predict diversity shifts. Such mechanistic models are rare but have considerable promise to improve predictions of individual species' distributions and sometimes abundance (Buckley 2008). Their application to predicting range shifts for large assemblages will require further development of their generality and scope. All models, from the simplest correlation to most elaborate mechanistic model, must overcome the problem that future climates and land uses may lack current analogs, requiring extrapolation beyond current conditions (Williams et al. 2007). Nevertheless, despite changes in land-use and climate combinations throughout the 20th century, the empirical approach we have used successfully predicted diversity shifts within that period, implying that these models capture the critical relations governing where species can be detected. Whether this approach will remain reliable as climate changes continue and novel combinations become more extreme is uncertain. Contributions from any model that improves mechanistic links between shifting environmental conditions and species distributions are urgently needed.

The best-CAR regression model generated more reliable predictions than its OLS analog. The effect of spatial autocorrelation (the non-independence of geographically proximate data points) on regression analysis of geographical ecological data has received prolonged attention in recent years (Dormann 2007a). Recently, debate on this topic has focused on whether ignoring spatial autocorrelation biases coefficient estimates from OLS regression (Beale et al. 2007, Dormann 2007a, Hawkins et al. 2007, Kühn 2007). Our dataset provides a novel perspective on this debate. By incorporating a temporal component, we directly compared the predictive ability of coefficient estimates from OLS and conditional autoregression. The best-CAR's coefficients predicted recent species richness much better than OLS estimates, suggesting that accounting for spatial autocorrelation in regression analysis more closely captures true biological relationships between environment and species richness than standard OLS regression analysis.

It is important to note that although predictions based on individual niche model projections were inferior to those of the best-CAR, they performed well, and outperformed the OLS-based approaches. This result suggests that individual niche modeling approaches offer an improved mode of prediction over models based on OLS regression. Since species richness is, ultimately, the sum of individual species' presences and absences, why did the individual niche model approach not, at the very least, make predictions equivalent to the best-CAR's? One potential reason is that no niche model can be truly comprehensive, i.e. include all niche characteristics (Araújo and New 2007, Dormann 2007b). Similarly, no realistic regression model can be truly comprehensive (or there would be no residual variation). However, by including an influence of neighbouring quadrats (i.e. spatial effects) on species richness, the best-CAR provides an estimate of the relations between measured environmental variables and richness that are conditioned on the potential relations between richness and unmeasured, spatially structured variables. Although our best-CAR predictions of recent richness used only the coefficients of the measured environmental variables, these coefficients differed from the OLS estimates, due to the incorporation of spatial effects. Our niche models did not include analogous spatial effects. Therefore, the projections are based solely on the measured variables, without considering influences of unmeasured, but spatially structured variables. Potentially useful inclusion of spatial effects in distribution models has been demonstrated in North American birds (Bahn and McGill 2007, Allouche et al. 2008). Spatial effects could be modelled concomitantly with environmental variables in Maxent by including distance from the nearest presence point (Allouche et al. 2008) or by using spatial eigenvectors as predictors (De Marco et al. 2008). However, Maxent currently does not have the capacity to generate projections based solely on the subset of environmental relations as we have done for our best-CAR prediction. Such a procedure would allow for Maxent projections to be conditioned on unmeasured, spatially structured variables.

Given these differences in the best-CAR and aspatial niche models, the failure of the niche model approaches to out-predict the best-CAR approach suggests that our niche models lacked some relevant niche characteristic(s); possible exclusions could be either an abiotic variable, or the effects of biotic interactions (Araújo and Luoto 2007, Dormann 2007b). Empirical diversity theory approaches that used OLS regression performed less well than the individual niche modeling approaches, suggesting that unmeasured variables introduce more bias into these regression models than individual niche models. The ease with which spatial effects can be included in regression models, thereby correcting this bias (at least in our study system), is another advantage of this approach. Nevertheless, the high accuracy of the niche modeling approach's predictions suggest that they successfully identified the dominant factors determining species distributions and niche characteristics that we failed to include likely played only minor roles.

The niche modeling approach to predicting how species diversity will change has two potential advantages over regression modeling. First, the empirical diversity theory approach we have used is limited to predictions of species numbers, though Ferrier's generalized dissimilarity modeling (Ferrier 2002, Ferrier et al. 2004) and related approaches (Hortal and Lobo 2006) can incorporate measures of species turnover among regions or sites. The individual niche modeling approach can predict both species richness and composition, a major advantage if there are species (such as those listed as ''at risk'') whose distributions are of particular interest. Second, approaches using regression may be limited to predictions of widespread, generalist species richness, as we have used here. Geographical patterns of the diversity of widespread species are better predicted by climate than geographically restricted species (Jetz and Rahbek 2002). Therefore, the success of the empirical diversity theory approach will be determined primarily by the similarity of widespread species responses to climate change; these changes may not necessarily be representative of changes in geographically rare species' ranges. We could not explicitly evaluate empirical diversity theory or the niche modeling approach's ability to predict shifts in the diversity of geographically rare species as we only had sufficient data to model ranges of widespread, relatively common species. However, when sufficient data are available, niche models often perform better on species with small ranges (McPherson and Jetz 2007), though this could result, at least in part, from using AUC as a measure of model performance (Lobo et al. 2008). Nevertheless, niche modeling predictions may prove more effective than regression-based predictions for assemblages of geographically rare species or for complete assemblages where range-restricted species have sufficient geographic overlap to non-randomly affect the overall richness gradient, though the latter case may occur only rarely.

Committee averaging produced both accurate and precise predictions of richness, but overall it performed less well than the single-best approach. This result is expected if the single-best model accurately captures the functional links between environment and species richness.

When this is the case, the incorporation of other models will bias the final projection away from the best model's predictions. However, when a model with strong theoretical and empirical support is unavailable or a single best model cannot be easily identified (Dormann et al. 2008) committee averaging can be used to reliably predict diversity.

The minimum macroecological model was not as effective as other methods, but it performed reasonably well considering it was the only model whose variables we defined a priori. Instead its variables – mean growing season temperature, mean annual precipitation, and their interaction – were chosen based on an established hypothesis linking global diversity gradients to climate (Francis and Currie 2003). The model's predictions were likely hindered by two factors. First, it excluded effects of habitat heterogeneity, which is known to affect butterfly species richness in Canada (Kerr 2001, Kerr et al. 2001) and, second, precipitation effects are most prominent in sub-tropical and tropical areas (Hawkins et al. 2003) well south of Canada. Despite these limitations, the minimum macroecological model's predictions were still sufficiently accurate and precise to be useful in a conservation context if other, system-specific, regression models were unavailable.

Our results are unlikely to be biased by the necessary level of estimation involved in reconstructing historical and recent diversity patterns. Small errors in estimated ranges, especially at their margins, are unlikely to have substantial influence on our results due to our method of sampling richness at a relatively coarse grain (100×100 km). Although we constructed distribution models and generated predictions using the same environmental data, this approach is not tautological, since each species range was modeled independently. Thus, all species could potentially have responded to the relevant environmental variables in a manner that produced patterns of richness that were not highly correlated with climate. Thus, areas of high richness reflect underlying commonalities in species' niche requirements.

Empirical diversity theory, based on decades of macroecological research into spatial diversity gradients, produces reliable and remarkably accurate predictions of diversity shifts as environmental conditions change. Moreover, this approach, when incorporating spatial autocorrelation, outperformed the more common aspatial method of summing projected range shifts of individual species. This result highlights: 1) the usefulness of using global change to test ecological theory; 2) the importance of incorporating spatial information in lieu of complete information on environmental influences; and 3) demonstrates how ecologists can exploit past global change to produce the most reliable predictions possible of future biological responses to continually changing climate and land-use regimes.

References

Allouche, O. et al. 2008. Incorporating distance constraints into species distribution models. – J. Appl. Ecol. 45: 599–609.

Araújo, M. B. and Luoto, M. 2007. The importance of biotic interactions for modelling species distributions under climate change. – Global Ecol. Biogeogr. 16: 743–753.

Araújo, M. B. and New, M. 2007. Ensemble forecasting of species distributions. – Trends Ecol. Evol. 22: 42–47.

Araújo, M. B. et al. 2005a. Validation of species-climate impact models under climate change. – Global Change Biol. 11: 1504–1513.

Araújo, M. B. et al. 2005b. Reducing uncertainty in projections of extinction risk from climate change. – Global Ecol. Biogeogr. 14: 529–538.

Araújo, M. B. et al. 2008. Quaternary climate changes explain diversity among reptiles and amphibians. – Ecography 31: 8–15.

Bahn, V. and McGill, B. J. 2007. Can niche-based distribution models outperform spatial interpolation. – Global Ecol. Biogeogr. 16: 733–742.

Beale, C. M. et al. 2007. Red herrings remain in geographical ecology: a reply to Hawkins et al. (2007). – Ecography 30: 845–847.

Beaubien, J. et al. 2000. Land cover of Canada. – Natural Resources Canada, Ottawa.

Beaumont, L. J. et al. 2007. Where will species go? Incorporating new advances in climate modelling into projections of species distributions. – Global Change Biol. 13: 1368–1385.

Bivand, R. et al. 2008. spdep: spatial dependence, weighting schemes, statistics and models. – R package ver. 0.4–24.

Buckley, L. B. 2008. Linking traits to energetics and population dynamics to predict lizard ranges in changing environments. – Am. Nat. 171: E1–E19.

Buckley, L. B. and Jetz, W. 2007. Environmental and historical constraints on global patterns of amphibian richness. – Proc. R. Soc. B 274: 1167–1173.

Currie, D. J. 1991. Energy and large-scale patterns of animal-species and plant-species richness. – Am. Nat. 137: 27–49.

De Marco, P. et al. 2008. Spatial analysis improves species distribution modelling during range expansion. – Biol. Lett. 4: 577–580.

Dormann, C. F. 2007a. Effects of incorporating spatial autocorrelation into the analysis of species distribution data. – Global Ecol. Biogeogr. 16: 129–138.

Dormann, C. F. 2007b. Promising the future? Global change projections of species distributions. – Basic Appl. Ecol. 8: 609–628.

Dormann, C. F. et al. 2008. Prediction uncertainty of environmental change effects on temperate European biodiversity. – Ecol. Lett. 11: 235–244.

Elith, J. et al. 2006. Novel methods improve prediction of species' distributions from occurrence data. – Ecography 29: 129–151.

Ferrier, S. 2002. Mapping spatial pattern in biodiversity for regional conservation planning: where to from here? – Syst. Biol. 51: 331–363.

Ferrier, S. et al. 2004. Mapping more of terrestrial biodiversity for global conservation assessment. – BioScience 54: 1101–1109.

Fielding, A. H. and Bell, J. F. 1997. A review of methods for the assessment of prediction errors in conservation presence/absence models. – Environ. Conserv. 24: 38–49.

Fischer, A. G. 1960. Latitudinal variations in organic diversity. – Evolution 14: 64–81.

Francis, A. P. and Currie, D. J. 2003. A globally consistent richness-climate relationship for angiosperms. – Am. Nat. 161: 523–536.

H-Acevedo, D. and Currie, D. J. 2003. Does climate determine broad-scale patterns of species richness? A test of the causal link by natural experiment. – Global Ecol. Biogeogr. 12: 461–473.

Hawkins, B. A. et al. 2003. Energy, water, and broad-scale geographic patterns of species richness. – Ecology 84: 3105–3117.

Hawkins, B. A. et al. 2007. Red herrings revisited: spatial autocorrelation and parameter estimation in geographical ecology. – Ecography 30: 375–384.

Hernandez, P. A. et al. 2006. The effect of sample size and species characteristics on performance of different species distribution modeling methods. – Ecography 29: 773–785.

Hoeting, J. A. et al. 1999. Bayesian model averaging: a tutorial. – Stat. Sci. 14: 382–401.

Hortal, J. and Lobo, J. M. 2006. Towards a synecological framework for systematic conservation planning. – Biodiv. Inform. 3: 16–45.

IPCC 2007. Climate change 2007: the physical science basis. – Cambridge Univ. Press.

Jetz, W. and Rahbek, C. 2002. Geographic range size and determinants of avian species richness. – Science 297: 1548–1551.

Kerr, J. T. 2001. Butterfly species richness patterns in Canada: energy, heterogeneity, and the potential consequences of climate change. – Conserv. Ecol. 5: 1–17.

Kerr, J. T. and Packer, L. 1997. Habitat heterogeneity as a determinant of mammal species richness in high-energy regions. – Nature 385: 252–254.

Kerr, J. T. and Cihlar, J. 2003. Land use and land use intensity estimation in Canada from SPOT4/VEGETATION and ancillary data. – Global Ecol. Biogeogr. 12: 161–172.

Kerr, J. T. et al. 2001. Remotely sensed habitat diversity predicts butterfly species richness and community similarity in Canada. – Proc. Nat. Acad. Sci. USA 98: 11365–11370.

Kerr, J. T. et al. 2007. The macroecological contribution to global change solutions. – Science 316: 1581–1584.

Kharouba, H. M. et al. in press. Historically calibrated predictions of butterfly species' range shift using global change as a pseudo-experiment. – Ecology.

Kühn, I. 2007. Incorporating spatial autocorrelation may invert observed patterns. – Divers. Distrib. 13: 66–69.

Layberry, R. A. et al. 1998. The butterflies of Canada. – NRC Research Press.

Lennon, J. J. 2000. Red-shifts and red herrings in geographical ecology. – Ecography 23: 101–113.

Lichstein, J. W. et al. 2002. Spatial autocorrelation and auto-regressive models in ecology. – Ecol. Monogr. 72: 445–463.

Liu, C. R. et al. 2005. Selecting thresholds of occurrence in the prediction of species distributions. – Ecography 28: 385–393.

Lobo, J. M. et al. 2008. AUC: a misleading measure of the performance of predictive distribution models. – Global Ecol. Biogeogr. 17: 145–151.

Lumley, T. and Miller, A. 2007. leaps: regression subset selection. – R package ver. 2.7.

McKenney, D. W. et al. 2001. Canada's plant hardiness zones revisited using modern climate interpolation techniques. – Can. J. Plant Sci. 81: 129–143.

McKenney, D. W. et al. 2007. Potential impacts of climate change on the distribution of North American trees. – Bioscience 57: 939–948.

McPherson, J. M. and Jetz, W. 2007. Effects of species' ecology on the accuracy of distribution models. – Ecography 30: 135–151.

Mekis, E. and Hogg, W. D. 1999. Rehabilitation and analysis of Canadian daily precipitation time series. – Atmos. Ocean 37: 53–85.

Menendez, R. et al. 2006. Species richness changes lag behind climate change. – Proc. R. Soc. B 273: 1465–1470.

Nogués-Bravo, D. et al. 2008. Climate change, humans, and the extinction of the woolly mammoth. – PLoS Biol. 6: 0686–0691.

Parmesan, C. et al. 1999. Poleward shifts in geographical ranges of butterfly species associated with regional warming. – Nature 399: 579–583.

Pearson, R. G. et al. 2007. Predicting species distributions from small numbers of occurrence records: a test case using cryptic geckos in Madagascar. – J. Biogeogr. 34: 102–117.

Peterson, A. T. et al. 2004. Modeled climate change effects on distributions of Canadian butterfly species. – Can. J. Zool. – Rev. Can. Zool. 82: 851–858.

Phillips, S. J. et al. 2006. Maximum entropy modeling of species geographic distributions. – Ecol. Model. 190: 231–259.

Piñeiro, G. et al. 2008. How to evaluate models: observed vs predicted or predicted vs observed? – Ecol. Model. 216: 316–322.

R Development Core Team 2007. R: a language and environment for statistical computing. – R Foundation for Statistical Computing, Vienna.

Ramankutty, N. and Foley, J. A. 1999. Estimating historial changes in land cover: North American croplands from 1850–1992. – Global Ecol. Biogeogr. 8: 381–396.

Ricklefs, R. E. 2007. History and diversity: explorations at the intersection of ecology and evolution. – Am. Nat. 170: S56–S70.

Root, T. L. et al. 2003. Fingerprints of global warming on wild animals and plants. – Nature 421: 57–60.

Rosenzweig, C. et al. 2008. Attributing physical and biological impacts to anthropogenic climate change. – Nature 453: 353–357.

Svenning, J. C. and Skov, F. 2007. Could the tree diversity pattern in Europe be generated by postglacial dispersal limitation. – Ecol. Lett. 10: 453–460.

Thomas, J. A. et al. 2004. Comparative losses of British butterflies, birds, and plants and the global extinction crisis. – Science 303: 1879–1881.

Thuiller, W. 2003. BIOMOD – optimizing predictions of species distributions and projecting potential future shifts under global change. – Global Change Biol. 9: 1353–1362.

Thuiller, W. et al. 2005. Climate change threats to plant diversity in Europe. – Proc. Nat. Acad. Sci. USA 102: 8245–8250.

White, P. and Kerr, J. T. 2006. Contrasting spatial and temporal global change impacts on butterfly species richness during the 20th century. – Ecography 29: 908–918.

White, P. and Kerr, J. T. 2007. Human impacts on environment-diversity relationships: evidence for biotic homogenization from butterfly species richness patterns. – Global Ecol. Biogeogr. 16: 290–299.

Whittaker, R. J. et al. 2007. Geographical gradients in species richness: a test of the water-energy conjecture of Hawkins et al. (2003) using European data for five taxa. – Global Ecol. Biogeogr. 16: 76–89.

Williams, J. W. et al. 2007. Projected distributions of novel and disappearling climates by 2100 AD. – Proc. Nat. Acad. Sci. USA 104: 5738–5472.

Willis, K. J. and Birks, H. J. B. 2006. What is natural? The need for a long-term perspective inbiodiversity conservation. – Science 314: 1261–1265.