

Historically calibrated predictions of butterfly species' range shift using global change as a pseudo-experiment

HEATHER M. KHAROUBA,¹ ADAM C. ALGAR, AND JEREMY T. KERR²

Department of Biology, University of Ottawa, 30 Marie Curie, P.O. Box 450 Station A, Ottawa, Ontario K1N 6N5 Canada

Abstract. Global changes have the potential to cause a mass extinction. Predicting how species will respond to anticipated changes is a necessary prerequisite to effectively conserving them and reducing extinction rates. Species niche models are widely used for such predictions, but their reliability over long time periods is known to vary. However, climate and land use changes in northern countries provide a pseudo-experiment to test model reliability for predicting future conditions, provided historical data on both species distributions and environmental conditions are available. Using maximum entropy, a prominent modeling technique, we constructed historical models of butterfly species' ranges across Canada and then ran the models forward to present-day to test how well they predicted the current ranges of species. For the majority of species, projections of how we predicted species would respond to known climate changes corresponded with species' observed responses (mean autoregressive $R^2 = 0.70$). This correspondence declined for northerly and very widely distributed species. Our results demonstrate that at least some species are tracking shifting climatic conditions across very large geographic areas and that these shifts can be predicted accurately using niche models. We also found, however, that models for some species fail when projected through time despite high spatial model accuracies during model training, highlighting the need to base management decisions on species assemblages, not individual species.

Key words: *butterflies; Canada; climate change; land use change; macroecology; maximum entropy; range shifts; space-for-time substitution; species assemblages; species niche models.*

INTRODUCTION

Accurate predictions of how species will respond to climate change are critical. Climate change in the past century is thought to have already caused many species to shift northward, to higher altitudes (e.g., Grabherr et al. 1994, Parmesan 1996, Parmesan and Yohe 2003, Hickling et al. 2006), and to exhibit altered phenologies (Root et al. 2003). Although forecasts of future extinction rates are highly uncertain, it is clear that climate change is highly likely to accelerate them dramatically (Thomas et al. 2004, see replies: Thuiller et al. 2004, Buckley and Roughgarden 2004, Harte et al. 2004). A particularly intense focus on estimation of range shifts among species has led to rapid progress in the use of niche modeling to predict where species are likely to move (e.g., Peterson et al. 2001, 2005, Midgley et al. 2002, Araújo and Luoto 2007) given anticipated climate and land uses changes from well-established global and regional models (e.g., Sala et al. 2000, Chapin et al. 2001, Meehl et al. 2007). Historical data are urgently needed to calibrate predictions of future global

change impacts on biodiversity (Willis et al. 2007) and provide context for natural vs. anthropogenic changes (Willis and Birks 2007).

To estimate future climatic changes on species, many studies project future species distributions under different climate scenarios, based on models constructed with data from the present day (e.g., Thomas et al. 2004, Peterson et al. 2004, Pearson et al. 2006). These models estimate species' niches across geographical space by relating presence records of the species to environmental predictors within a particular time period (see Botkin et al. 2007). By substituting new variables that reflect anticipated environmental changes into these spatial models, the distribution of suitable habitat for a particular species can be predicted; a process that necessitates the assumption that the observed spatial relationships used to construct the initial model will remain largely consistent through a period of global change (Kerr et al. 2007). Niche models intended to predict climate change impacts on species' ranges are most commonly constructed using contemporary data. The space-for-time assumption can be unreliable (Lawler et al. 2006, Pearson et al. 2006), with different modeling methods projecting massive differences in range sizes for the same species and region. Moreover, slight differences between the predicted current distributions of species derived from different models can be magnified when projecting distributions through time

Manuscript received 10 July 2008; revised 7 October 2008; accepted 20 October 2008; final version received 26 November 2008. Corresponding Editor: R. A. Raguso.

¹ Present address: Department of Zoology, University of British Columbia, 2370-6270 University Boulevard, Vancouver, British Columbia V6T 1Z4 Canada.

² Corresponding author. E-mail: jkerr@uottawa.ca

under anticipated climates (Thuiller 2003, Beaumont et al. 2007), a result that seems unsurprising given reported differences in the reliability of different modeling methods (Elith et al. 2006), each of which employs a distinctive algorithm and requires different assumptions. New work, however, demonstrates that differences in model accuracy between species greatly outstrip differences in model accuracy due to the use of different modeling techniques (Guisan et al. 2008).

Here we present the first attempt of which we are aware to test whether accurate predictions of species range shifts over long time periods (30–60 years) and very large geographical areas are likely. We conducted this analysis across Canada, covering a geographical extent of nearly 10 million km². Canada, as a northern country, has already experienced large climate changes, but differs from comparably high-latitude areas in Western Europe in having extensive areas with relatively little permanent human land uses. For well-studied taxa, this improves the likelihood of discriminating between climate and land use change impacts on species distributions. Secondarily, we consider the important issue of whether and to what degree species' traits influence the predictability of range shifts (if any) during the 20th century (McPherson and Jetz 2007). These models provide a critical baseline observation of rates of species range shift as a function of observed climate changes and will allow calibrated predictions of future range shifts given large anticipated climate changes in northern environments.

METHODS

Niche modeling

Maximum entropy (Maxent) was chosen to model species distributions given that it was developed specifically for use with presence-only occurrence data (see Phillips et al. 2006), and it performs well compared to other methods (Elith et al. 2006). Maxent estimates a target probability distribution for each species by finding the probability distribution of maximum entropy (i.e., closest to uniform), subject to a set of constraints (environmental variables) that represents the incomplete information about the target distribution (see Phillips et al. 2006). For each model (which predicts where a species is found across geographical space derived from its occurrence records relative to environmental predictors), Maxent produces a cumulative map of predicted suitability where 0 means the environmental conditions are predicted to be unsuitable for the species, and 100 is predicted to be perfectly suitable for the species given the environmental variables used in the model. A complete list of species included can be found in Appendix A.

Species data

Species niches were modeled using occurrence records taken from the Canadian National Collection of Butterflies, which contains ~300 000 georeferenced, dated records for 297 Canadian butterfly species

(Layberry et al. 1998). This database includes only those records that could be georeferenced precisely (e.g., to a particular location, >80% of records to within 1 km), and butterfly specimens stored at one of many museums across Canada. Butterfly collecting increased in intensity throughout Canada during the later part of the 20th century. There was a mean of 34 geographically unique records per species used in this analysis in the 1900–1930 period, and 200 in the 1960–1990 period. For each species, the same number of records was chosen at random from the records for 1960–1990 as were available from the 1900–1930 time period to reduce sampling biases associated with increased sampling intensity through time. Therefore, the numbers of records used in all three models (historical, projected, and current) were identical. Only geographically unique records from 1900–1930 and 1960–1990 were used (i.e., a location was only included once even if it had been sampled repeatedly over time). Species with fewer than 10 geographically distinct records in either time period were excluded based on modeling accuracy concerns (Hernandez et al. 2006), although most species were widely collected from within their geographical ranges in Canada. There were 139 species that met these preselection requirements (see Appendix A).

Six predictor variables were selected for niche models: mean growing season temperature, maximum growing season temperature, ecozones, land cover, total annual precipitation, and human population density. To measure climate, monthly precipitation and temperature data were obtained from climate normals for 1901–1930 and for 1961–1990 from the Canadian Forest Service (McKenney et al. 2001; D. W. McKenney, *personal communication*). These data were aggregated to produce the growing season (April–October) temperature, maximum growing season temperature, and annual precipitation data sets for all of Canada. Growing season and maximum growing season temperature were selected because they were the most common limiting temperature variables in preliminary analyses. Growing season temperature was also previously found to be a consistent predictor of butterfly species richness across Canada (White and Kerr 2006). Physical land cover data, which describe the broad ecosystems and major agricultural regions of Canada (Beaubien et al. 2000), as well as Canada's 15 major ecological regions, were included to measure the influence of vegetation, which is known to strongly affect butterfly species richness in Canada (Kerr 2001, Kerr et al. 2001). As an approximation of human land use pressure, human population density based on the census of 1921 and 1981 was used. Briefly, population density was measured as the mean number of people per square kilometer in each of the 238 census divisions across the country (White and Kerr 2006). Agricultural land uses in Canada have tracked the rapid expansion of human populations in Canada, a relationship known to be sufficiently strong to allow historical land use extent to be inferred directly from observed

trends in human population density (Ramankutty and Foley 1999). The landcover, ecozone, and human population density data were resampled to a resolution of ~ 6.6 km, which is the minimum interpretable resolution for the climate data sets (D. W. McKenney, *personal communication*) and considerably greater than the georeferencing precision in our data set (which was < 1 km).

Temporal experiment

Our approach involved four steps (see Fig. 1): (a) modeling species distributions with environmental data and occurrence records from 1900–1930 (this step will be referred to as the “historical model”); (b) projecting these niche models into the future by substituting environmental data from 1960 to 1990 into niche models from 1900 to 1930 (i.e., running the models forward through time to nearly the present day, henceforth referred to as the “projected model”); (c) modeling species distributions independently using environmental data and occurrence records from 1960 to 1990 (this step will be referred to as the “current model”); and (d) testing the ability of the projected models to predict the current models.

To build each model, the occurrence records were randomly divided into a training (70%) and testing set (30%). For each step (a–c), randomizations were repeated 10 times to produce 10 models for each species and then the mean predicted suitability for each pixel in Canada based on the 10 model outputs was calculated to produce the final model (Phillips et al. 2006). Generally, there were no large differences in accuracy of the different models but probability of suitability changed slightly between models.

Evaluating the projected models

Before assessing how well the projected models predicted the current models, the accuracy of each model was calculated as the area under the curve (AUC) of the receiver operating characteristic. AUC has been used extensively in the species’ distribution modeling literature and measures the ability of a model to discriminate between sites where a species is present vs. those where it is absent (Fielding and Bell 1997, Elith et al. 2006, see also Lobo et al. 2007). AUC ranges from 0 to 1, where a score of 0.5 indicates that the model performs no better than random, and a score of 1 indicates perfect discrimination (Fielding and Bell 1997). For each species and for the historical and current model, the mean AUC based on the 10 individual models was calculated (listed fully in Appendix A). Values between 0.7 and 0.9 are considered useful, and values exceeding 0.9 are considered excellent (Swets 1988).

Next, the ability of the projected models to predict the current models using spatial correlation was assessed at two scales (for pixels and for ecodistricts; this evaluation method will now be referred to as “predictive ability”).

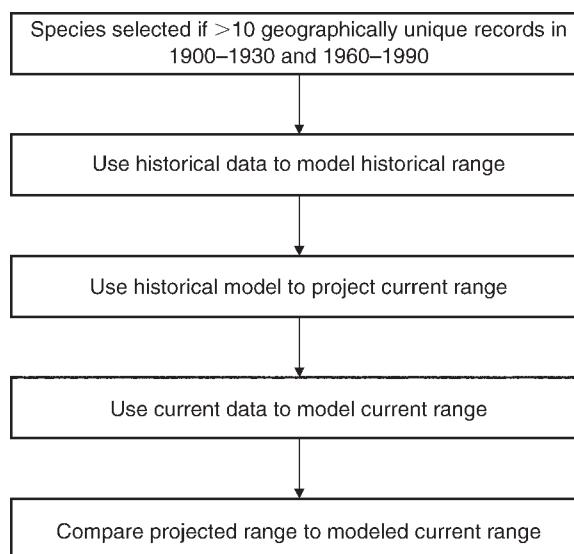


FIG. 1. The methodology used to assess the degree to which species distribution models based on historical data can predict species’ niche shifts through time.

For each species, the probability of suitability generated by the projected model was regressed against the probability of suitability from the current model using linear regression across all pixels (area of each pixel = $6.61 \text{ km} \times 6.61 \text{ km} = 43.7 \text{ km}^2$, $n = 2.2 \times 10^5$). Since conservation decisions are often made at a regional scale, the predictive ability was also assessed across ecodistricts (mean area = $12\,000 \text{ km}^2$, $n = 1062$), which are the smallest ecological units used for policy decisions in Canada. Ecodistricts have distinctive species assemblages, as well as internal consistency in environmental conditions including topography, geology, soil, and dominant vegetation structure (more information *available online*).³ For each ecodistrict, the mean predicted suitability of constituent pixels was calculated and then both a conditional autoregressive and ordinary least-squares (OLS) regression models were constructed to compare projected to current distribution models for each species. This analysis identifies the degree to which models built using data from one time period and then run through time to the present-day agree with current species niches. Positive spatial autocorrelation can influence statistical tests by increasing the chances of committing Type I errors (incorrectly rejecting the null hypothesis; Legendre and Legendre 1998). Conditional autoregressive models provide more robust probability tests and estimates of variable coefficients in the presence of spatial autocorrelation (Lichstein et al. 2002, Diniz-Filho et al. 2003). Conditional autoregressive models cannot be calculated using higher resolution, pixel-based data because of computational limitations.

³ (<http://sis.agr.gc.ca/cansis/nsdb/ecostrat/>)

TABLE 1. Pairwise correlations among all variables expected to influence predictive ability (R^2 values from regression models of projected vs. current model for species) and model accuracy (measured as percentage of observations correctly predicted as suitable); $n = 139$ species.

Variable	Threshold	Sample size	AUC	Wing size	Suitable area	Latitude	Temperature range
Threshold	1	-0.71	0.12	0.083	-0.69	-0.092	0.30
Sample size	-0.71	1	-0.11	-0.025	0.70	0.091	-0.37
AUC	0.12	-0.11	1	0.094	-0.33	-0.51	-0.17
Wing size	0.083	-0.025	0.094	1	-0.050	-0.11	-0.16
Suitable area	-0.69	0.70	-0.33	-0.050	1	0.42	-0.38
Latitude	-0.09	0.091	-0.51	-0.11	0.42	1	0.23
Temperature range	0.30	-0.37	-0.17	-0.16	-0.38	0.23	1

Notes: Presented are Pearson r values. Threshold is the threshold used to classify the continuous output from Maxent (probability of suitability) into a binary presence-absence map for the historical model. Sample size is the number of observation records used in models. AUC is the mean area under the curve of the historical model based on 10 model outputs. Wing size was measured as wing span when fully extended. Species' suitable area is as predicted by the historical model. Latitude is the center point of a species' suitable area based on the historical model. Temperature range is the range of growing-season temperature across a species' suitable area based on the historical model.

For each species, we tested how well models projected through time matched independently constructed, current models. Predictive ability was assessed differently for ecodistricts vs. pixels: among ecodistricts, for which it was possible to construct conditional autoregressive models, we calculated predictive ability as the coefficient of determination of the full regression model (which includes a spatial autocorrelative component). Because our pixel-based data are too numerous to calculate a conditional autoregressive model, we based our assessments for pixel-based predictive ability on OLS regressions. We provide these data for comparative purposes, but caution that OLS results include a spatial component that may not represent a true biological effect.

The accuracy of the projected models was also calculated based on counts of observation points. For each species, the percentage of occurrence points for 1960–1990 that were correctly classified as suitable by the projected model was determined (which is 1 – omission error, where omission error is the proportion of observations that are found outside the predicted niche for the species). For each model, the model output (a map of probability of suitability) was converted into a binary map of areas predicted to be suitable and unsuitable. A decision threshold was defined for each model, above which species were considered to be present and below which species were considered to be absent. Many different approaches have been employed for setting thresholds (Liu et al. 2005). A simple, but effective, approach was used (Liu et al. 2005, Pearson et al. 2007). When the occurrence records were initially divided up into training and testing groups to build each model, each record was ranked in suitability compared to the other occurrence records and assigned a predicted suitability value. To set a threshold for each final model (i.e., to classify the continuous output from Maxent into a binary presence-absence map), the observation with the lowest predicted suitability in the training set of each model was used and the mean suitability of that record across the 10 models was taken. Therefore, for each final

model, all pixels that were predicted to be at least as suitable as those where a species had actually been observed were identified (Pearson et al. 2007). These binary maps were used to depict estimated species richness gradients in Canada and to provide a secondary estimate of model accuracy based on inclusion of observation points within modeled niches.

Lastly, we created richness maps based on all three models (historical, projected, and current). We summed the binary range maps across all species for each model to give a spatial representation of how the models compared when the entire assemblage of species was considered across Canada. This approach permitted consideration of whether projected models systematically underestimated or overestimated actual richness in regions of Canada.

Evaluating the influence of niche characteristics

Species' characteristics are known to affect the spatial accuracy of niche models (e.g., Stockwell and Peterson 2002, Segurado and Araújo 2004, McPherson and Jetz 2007, Pöyry et al. 2008). Forecasts of species' geographical responses to global change through time would be more useful if a priori estimates of the temporal accuracy of niche models were possible. We regressed the predictive ability of distribution models (i.e., conditional autoregressive [CAR] R^2 values) against niche characteristics from the historical time period (see Table 1 for complete listing), which allowed for tests of whether there are species or model characteristics that predict how well a species tracks its geographical niche through time. We inspected bivariate plots of all variables before constructing pairwise correlations with potential variables to identify potential problems with multicollinearity among potential predictor variables (Table 1). Correlations between individual predictor variables and predictive ability, and between predictors and model accuracy, were also examined to gauge the individual importance of predictors (Table 2). We constructed regression models testing whether the

TABLE 2. Correlations (Pearson r ; $n = 139$ species) between species' niche characteristics and predictive ability (measured as r^2 values from regression models of projected vs. current model for species) across pixels and ecodistricts, and model accuracy (measured as percentage of observations correctly predicted as suitable) of projected and current model.

Variable	Predictive ability									
	Across ecodistricts with CAR		Across ecodistricts with OLS		Across pixels		Projected model accuracy		Current model accuracy	
	r	P	r	P	r	P	r	P	r	P
Threshold	N/A		N/A		N/A		-0.56	<0.001	-0.64	<0.001
Historical suitable area	-0.22	0.0081	-0.25	0.003	-0.30	0.0004	0.44	<0.001	0.36	<0.0001
Sample size	-0.051	0.55	-0.059	0.49	-0.086	0.32	0.43	<0.001	0.48	<0.001
Historical latitude	-0.33	<0.001	-0.33	<0.001	-0.39	<0.001	-0.038	0.66	0.037	0.67
Historical temperature range	-0.17	0.049	-0.18	0.032	-0.18	0.030	-0.48	<0.001	-0.037	0.66
Wing size	0.17	0.047	0.18	0.031	0.18	0.036	0.15	0.078	0.035	0.68
Historical AUC	0.32	<0.001	0.38	<0.001	0.41	<0.001	0.23	0.0057	0.030	0.73

Notes: Predictive ability across ecodistricts was measured from conditional autoregressive models (CAR) and ordinary least-squares regression models (OLS). Variables were selected based on these correlations and correlations from Table 1 for the regression model presented in Table 3. Threshold is the threshold used to classify the continuous output from Maxent (probability of suitability) into a binary presence-absence map for the historical and then current model; N/A indicates not applicable. Species' suitable area is as predicted by the historical model. Sample size is the number of observation records used in models. Latitude is the center point of a species' suitable area based on the historical model. Historical temperature range is the range of growing-season temperature across a species' suitable area based on the historical model. AUC is the mean area under the curve of the historical model based on 10 model outputs. Correlations were considered significant at $P < 0.05$.

coefficients of determination for OLS and CAR models could be predicted from an array of factors that might affect predictive capacity, including accuracy of the historical model, the wing size of each butterfly species, the size of each species' geographical range in Canada, the number of observations available to build models of each species' range, and environmental conditions within their ranges (including mean and maximum growing season temperatures). We included wing size in this list, as it is a quantifiable index of dispersal potential for butterflies and predicted that large-winged butterfly species would track their niches through time more effectively.

All statistical analyses were conducted using R 2.6.1 (R Foundation for Statistical Computing 2007), while all geographic data were manipulated using Arc/Info Grid 9.1 (ESRI 2005).

RESULTS

Evaluating the projected models

When evaluated internally, all models constructed using purely spatial data were very accurate. The mean AUC value for the historical model across all 139 species that passed the preselection phase was 0.920 (SE = 0.0063), and for the current model the mean was 0.940 (SE = 0.0055). Mean growing season temperature, ecozone, and population density frequently influenced model output the most, based on Maxent estimations of variables' relative importance.

Historical models projected through time showed significant correspondence with independent, current models. The mean predictive ability (OLS R^2) was 0.46 (SE = 0.016) based on pixel-level analysis, and the average slope was less than one (slope = 0.78, SE = 0.022). At the resolution of ecodistricts, for which

conditional autoregressive (CAR) models could be calculated, the mean R^2 of the complete models was 0.70 (SE = 0.17) and the average slope was 0.79 (SE = 0.32). There is significant variation in predictive ability (Fig. 2A), but CAR results indicate that projected models predicted current niche models for most species (i.e., models built on historical data and then run through time provide a reasonably strong match for models built using exclusively current data).

Model accuracy was also assessed using observation points only. On average, the current model was 93% (SE = 6.0%) accurate (i.e., percentage of observation records correctly predicted as suitable by current model). Projected models, on average, correctly predicted a mean of 58% (SE = 29%) of the current occurrence records. There was considerable variability in accuracy across species (Fig. 2B).

In general, when historical models were projected into the present day, they suggested habitat suitabilities that tend to be slightly lower than suitabilities that are actually observed using recent data only (based on probability of suitability outputs, the average slope from species-by-species regressions of projected vs. current model outputs was <1 when measured across pixels and ecodistricts for 91 of 139 species included). The mean area predicted as suitable for each species (based on binary maps indicating species' presence/absence) by the projected model was somewhat smaller than the current predicted range ($P = 0.0035$).

Across Canada, there are large differences in the degree to which species (or species' niches) apparently respond to climate change. The ecozone where projected models underestimated current habitat suitabilities the most is the Boreal Plains, where richness increased the most between study periods (Fig. 3). Human popula-

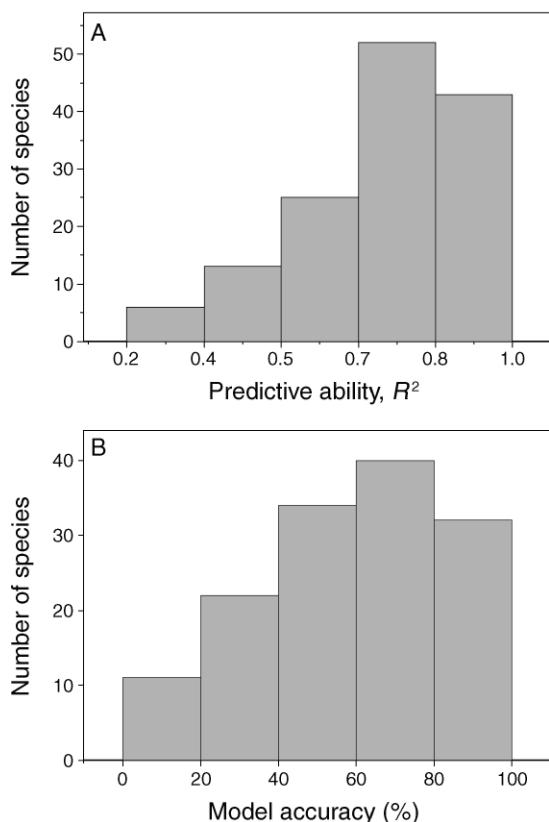


FIG. 2. The variability in predictive ability and model accuracy across species. (A) The distribution of r^2 values from the conditional regression models linking projected and current species' niches at the ecodistrict scale. (B) The distribution of the variation in the percentage of occurrence records from 1960–1990 correctly predicted as suitable by the projected model across all species ($n = 139$ species models).

tion density has increased rapidly there, and land use changes have been extensive (Kerr and Cihlar 2003). Projected models overestimated current habitat suitabilities the most, and richness declined to the greatest extent, in the Pacific Maritime ecozone (Fig. 3).

Variation in predictive ability and model accuracy

Predictive ability was not strongly correlated with any of the factors individually (i.e., $\langle \text{Pearson } r < 0.5$ in most cases; Table 2). At both scales, predictive ability is best explained by the mean AUC of the historical model (i.e., predictive ability through time improves for species that were modeled in the first time period with the greatest accuracy) and may be affected positively by wing size (but we caution that effects are statistically nonsignificant; Table 3). However, these variables explain very little total variation in how successfully a historical species niche model can be projected through time across Canada (Table 3).

Projected model accuracy (percentage of observation records from 1960 to 1990 that were correctly predicted as suitable by projected model) was most correlated with

the threshold used to convert the continuous distribution into a binary niche map (lower threshold value, greater model accuracy), and the number of observation records included in the model. This is to be expected as the chances of correctly predicting a presence increases with the number of occurrence records being tested. Species with lower set thresholds also have larger predicted suitable areas and thus greater model accuracy (i.e., fewer observations are omitted) according to this statistic. Although we report these accuracy statistics for completeness, we rely more on model outputs prior to thresholding (i.e., conversion of continuous data to presence-absence models).

DISCUSSION

Accurate predictions of species' geographical responses to global change through substantial time periods and over broad areas are possible. Species ranges estimated using niche models, as in this study, resemble range maps estimated using older techniques but have more readily quantifiable accuracies. However, projection accuracy varies with scale (i.e., grain size) and improves when measured at regional, rather than pixel-by-pixel, scales. Species' biological characteristics or environmental associations can sometimes help determine whether species' niches can be modeled successfully using spatial data (McPherson and Jetz 2007, Guisan et al. 2008), but our results suggest that when spatial models are projected through time, their behavior sometimes becomes erratic even when spatial models appear accurate. Some species' responses to climate change seem unpredictable even when using our proposed, historically calibrated models, much as spatial models fail for some species (Guisan et al. 2008). Projecting human impacts on species over the coming decades will be less accurate, even using the pseudo-experimental, historically calibrated approaches developed here, if

TABLE 3. Results from forward and backward stepwise regression linking predictive ability (measured as r^2 values from regression models of projected vs. current model for species) to the spatial accuracy of the historical model and the width of a species' environmental niche ($n = 139$ species).

Scale and predictor	Standardized coefficient	Probability	Model R^2
Ecodistricts ($n = 1062$) with CAR			
AUC	0.694	0.0002	0.125
Wing size	0.0016	0.0855	
Ecodistricts ($n = 1062$) with OLS			
AUC	1.11	<0.0001	0.165
Wing size	0.0023	0.0609	
Pixels ($n = 2.2 \times 10^5$)			
AUC	1.12	<0.0001	0.219
Wing size	0.0018	0.0733	

Notes: Predictive ability across ecodistricts was measured from conditional autoregressive models (CAR) and ordinary least-squares (OLS) models. AUC is the mean area under the curve of the projected model based on 10 model outputs.

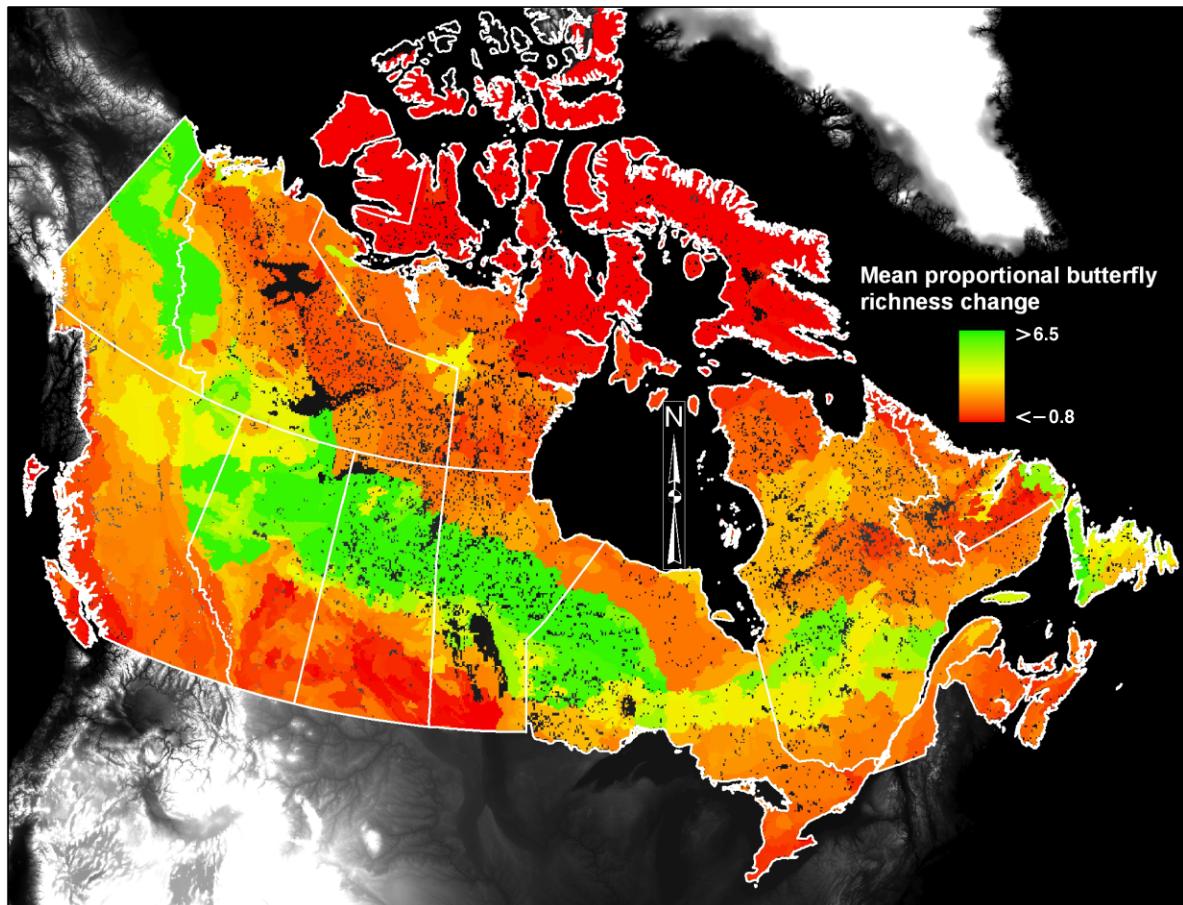


FIG. 3. Proportional differences in model predictions ($[(\text{current richness} - \text{projected richness})/\text{historical richness}] \times 100\%$) for butterfly niches (139 species models included) over the last century in Canada, averaged over each ecodistrict. The scale extends from red (indicating that projected richness was slightly higher than observed richness) to green (areas where projected richness was lower than actual richness). The scale is contrast-enhanced to maximize differences, but these actually range from approximately -1% to $+6.9\%$. Values for individual pixels are less accurate and more variable. Data are overlaid on a digital elevation model for North America, and political boundaries in Canada are indicated by white lines.

anticipated conditions lack historical precedent (see Thuiller 2003, Araújo et al. 2005, Kerr et al. 2007).

Inclusion of additional niche characteristics (e.g., data for host plant presence, if available) can improve spatial niche models for specialist butterfly species (Araújo and Luoto 2007). This study excludes such specialists. Topographic relief also makes environmental, including even satellite-based, measurements more complex and error prone (Kerr and Ostrovsky 2003), so estimates of niche parameters in such areas are less certain. A theoretically unsatisfying, but practical-minded, macroecological solution to this problem is to base projections on large species assemblages in the reasonable expectation that general trends will be robust to idiosyncratic results for individual species (Martinez-Meyer et al. 2004, Kerr et al. 2007).

Projected models may deviate from current models somewhat because of “clamping.” Clamping occurs when novel environmental conditions arise that are not

part of the historical environment from which species were modeled. Overall, clamping is not widespread or common: to date, in Canada, environmental changes within species’ modeled range have been small relative to the total amount of environmental variation historically present within their range. For instance, summer temperatures may have risen by, say, 2°C at a location where one species is found, but elsewhere within that range, it may already have been that warm naturally. Environmental regimes that most clearly lack historical precedents occurred outside the actual geographical range of most species, such as in the extreme northern environments of Canada, where there are nearly no butterflies. Clamping is unlikely to have resulted in substantial deviations of the projected model from the current model, although regional deviations are possible.

Projections of species’ niches through time are more likely to match observed shifts among species with more accurate current species niche models and for southern

species with narrower environmental niches. To some degree, greater spatial model accuracy leads to greater temporal accuracy, suggesting that the space-for-time substitution assumption necessary to project species' ranges into the future (e.g., Lawler et al. 2006, Pearson et al. 2006) does not preclude accurate model projections. Butterfly responses to global changes considered here may also have been influenced to a small extent by their dispersal characteristics. We expected large-winged butterfly species to disperse readily, potentially tracking shifting climatic conditions from year to year more effectively than less-capable dispersers with smaller wings (Dennis et al. 2000, Pearson and Dawson 2003). For moths, wing span has been shown to positively influence migration rate (Nieminen et al. 1999). We retained wing size in regression models despite marginal statistical nonsignificance because this variable varies relatively little among butterfly species, so it might have somewhat greater biological significance than regression models suggest. It is a weak predictor, however, and we caution against overinterpretation. Nonbiological contributions to the small decline in predictability of northern species may arise from reduced sampling intensity in the north. Models' predictive capacity does not simply reflect the degree of overlap between historical and current ranges (a species' historically suitable range divided by their current range) is statistically unrelated to predictive ability of models for those species (i.e., how closely models projected from the historical to current period match the current model; Pearson $r = 0.044$, $P = 0.60$).

Geographical ranges of butterflies in Canada have shifted in response to climatic changes in Canada (White and Kerr 2006) and are expected to continue to do so in the future (Peterson et al. 2004): In neither case, at least among the generalist species included here, has an effect of host plant limitation been detected or projected. Generalist species in Canada appear to be tracking climate changes with some success, to the extent that projections of their historical ranges into the present day reflect their current ranges, albeit imperfectly. We do not have evidence in this study, as among intensively monitored European butterfly communities (Menendez et al. 2006), regarding specialist species' responses to climate or land use changes, although there is considerable cause for concern in this area given suspicions of biotic homogenization among Canadian butterflies (White and Kerr 2007).

Although we equalized sampling intensity between time periods and sampling effects were not statistically detectable in any of our analyses, we cannot rule out hidden effects of sampling variation on our results. Butterflies have been collected in Canada for more than a century, but have not been sampled with the remarkably intensive, systematic efforts applied to the British flora and fauna (e.g., Hill et al. 2002, Menendez et al. 2006). We equalized sampling intensity by randomly selecting the same number of species obser-

vations in both study periods to reduce effects of geographic and temporal sampling biases. After equalizing sampling intensity between both time periods covered by this study, some residual geographical bias may yet have remained, as the records randomly chosen from the second time period could still have included more northerly regions than the records from the first time period simply because there was a larger pool of records to sample in the second time period. Our statistical analysis rules out a systematic, large impact of such bias: If such sampling bias actually did explain the trend of range expansion we detected, then differences in predicted suitable area between the current and projected species' range would relate to differences in numbers of records between the two time periods. In fact, predicted differences in species' range sizes between the two time periods are *unrelated* to differences in number of records ($r^2 = 0.0042$, $P = 0.45$) between time periods. Sampling effects of this kind are likely small and do not vary systematically with per-species sampling intensity. Expanded, systematic collecting will improve our ability to make strong predictions about species responses to global changes in the future (Kerr et al. 2007).

For many butterfly species, species niche models derived from purely spatial data predict how those species' niches have changed during the 20th century, a period during which climate and land use changes have been significant in Canada. The next step is to project species niche models into the future using only those species that have responded predictably in the past (given these model parameters), then with all species together, to assess the magnitude of uncertainty arising from the inclusion of species with niches that do not behave predictably through time. Such projections will be most reliable when limited to species that have responded predictably to recent global changes. Policy-relevant recommendations regarding where management activities should be directed are urgently required. Even when calibrated on past observations of niche shift, it is still possible that inaccurate predictions for some species are simply a cost that must be borne when making global change predictions. Historical calibration of such models provides one way to reduce that cost. Predicting the future, however, is likely to remain an uncertain enterprise. It would be wise to quantify and allow for this uncertainty when developing management strategies to mitigate the effects of accelerating climate change on biological diversity.

ACKNOWLEDGMENTS

The authors thank the Canadian Biodiversity Information Facility, particularly Larry Speers, Peter Hall, Don Lafontaine, and Ross Layberry, and the many institutions that contributed to butterfly collection data. The research was funded by NSERC, the Canadian Foundation for Innovation, and the Ontario Ministry of Research and Innovation. Robert Guralnick and an anonymous reviewer provided helpful comments that improved the manuscript.

LITERATURE CITED

- Araújo, M. B., and M. Luoto. 2007. The importance of biotic interactions for modelling species distributions under climate change. *Global Ecology and Biogeography* 16:743–753.
- Araújo, M. B., R. G. Pearson, W. Thuiller, and M. Erhard. 2005. Validation of species-climate impact models under climate change. *Global Change Biology* 11:1504–1513.
- Beaubien, J., J. Cihlar, and R. Latifovic. 2000. Land cover of Canada. Natural Resources Canada, Ottawa, Ontario, Canada.
- Beaumont, L. J., A. J. Pitman, M. Poulsen, and L. Hughes. 2007. Where will species go? Incorporating new advances in climate modelling into projections of species distributions. *Global Change Biology* 13:1368–1385.
- Botkin, D., et al. 2007. Forecasting effects of global warming on biodiversity. *Bioscience* 57:227–236.
- Buckley, L. B., and J. Roughgarden. 2004. Biodiversity conservation: effects of changes in climate and land use. *Nature* 430:33–35.
- Chapin, F. S., III, O. E. Sala, and E. Huber-Sannwald, editors. 2001. *Global biodiversity in a changing environment: scenarios for the 21st century*. Springer-Verlag, Berlin, Germany.
- Dennis, R. L. H., B. Donato, T. H. Sparks, and E. Pollard. 2000. Ecological correlates of island incidence and geographical range among British butterflies. *Biodiversity and Conservation* 9:343–359.
- Diniz-Filho, J. A. F., L. M. Bini, and B. A. Hawkins. 2003. Spatial autocorrelation and red herrings in geographical ecology. *Global Ecology and Biogeography* 12:53–64.
- Elith, J., et al. 2006. Novel methods improve prediction of species' distributions from occurrence data. *Ecography* 29: 129–151.
- ESRI [Environmental Systems Research Institute]. 2005. Arc/Info and ArcMap for Windows NT. Version 9.1. ESRI, Redlands, California, USA.
- Fielding, A. H., and J. F. Bell. 1997. A review of methods for the assessment of prediction errors in conservation presence/absence models. *Environmental Conservation* 24:38–49.
- Grabherr, G., M. Gottfried, and H. Pauli. 1994. Climate effects on mountain plants. *Nature* 369:448.
- Guisan, A., N. E. Zimmermann, J. Elith, C. H. Graham, S. Phillips, and A. T. Peterson. 2008. What matters for predicting the occurrences of trees: techniques, data, or species' characteristics? *Ecological Monographs* 77:615–630.
- Harte, J., A. Ostling, J. L. Green, and A. Kinzig. 2004. Biodiversity conservation: climate change and extinction risk. *Nature* 430:145–148.
- Hernandez, P. A., C. H. Graham, L. L. Master, and D. L. Albert. 2006. The effect of sample size and species characteristics on performance of different species distribution modeling methods. *Ecography* 29:773–785.
- Hickling, R., D. B. Roy, J. K. Hill, R. Fox, and C. D. Thomas. 2006. The distribution of a wide range of taxonomic groups are expanding polewards. *Global Change Biology* 12:450–455.
- Hill, J. K., C. D. Thomas, R. Fox, M. Telfer, S. Willis, J. Asher, and B. Huntley. 2002. Responses of butterflies to twentieth century climate warming: implications for future ranges. *Proceedings of the Royal Society of London B* 269:2163–2171.
- Kerr, J. T. 2001. Butterfly species richness patterns in Canada: energy, heterogeneity, and the potential consequences of climate change. *Conservation Ecology* 5:10. (<http://www.consecol.org/vol5/iss1/art10/>)
- Kerr, J. T., and J. Cihlar. 2003. Land use and land use intensity estimation in Canada from SPOT4/VEGETATION and ancillary data. *Global Ecology and Biogeography* 12:161–172.
- Kerr, J. T., H. M. Kharouba, and D. J. Currie. 2007. The macroecological contribution to global change solutions. *Science* 316:1581–1584.
- Kerr, J. T., and M. Ostrovsky. 2003. From space to species: ecological applications for remote sensing. *Trends in Ecology and Evolution* 18:299–305.
- Kerr, J. T., T. R. E. Southwood, and J. Cihlar. 2001. Remotely sensed habitat diversity predicts butterfly species richness and community similarity in Canada. *Proceedings of the National Academy of Sciences* 98:11365–11370.
- Lawler, J. J., D. White, R. P. Neilson, and A. R. Blaustein. 2006. Predicting climate-induced range shifts: model differences and model reliability. *Global Change Biology* 12: 1568–1584.
- Layberry, R. A., P. W. Hall, and J. D. Lafontaine. 1998. *The butterflies of Canada*. NRC Research Press, Canada Institute for Scientific and Technical Information, University of Toronto Press, Toronto, Ontario, Canada.
- Legendre, P., and L. Legendre. 1998. *Numerical ecology*. Second edition. Elsevier Science, Amsterdam, The Netherlands.
- Lichstein, J. W., T. R. Simons, S. A. Shriver, and K. E. Franzreb. 2002. Spatial autocorrelation and autoregressive models in ecology. *Ecological Monographs* 72:445–463.
- Liu, C., P. M. Berry, T. P. Dawson, and R. G. Pearson. 2005. Selecting thresholds of occurrence in the prediction of species distributions. *Ecography* 28:385–393.
- Lobo, J. M., A. Jiménez-Valverde, and R. Real. 2007. AUC: a misleading measure of the performance of predictive distribution models. *Global Ecology and Biogeography* 17: 145–151.
- Martinez-Meyer, E., A. T. Peterson, and W. W. Hargrove. 2004. Ecological niches as stable distributional constraints on mammal species, with implications for Pleistocene extinctions and climate change projections for biodiversity. *Global Ecology and Biogeography* 13:305–314.
- McKenney, D. W., M. F. Hutchison, J. L. Kesteven, and L. A. Venier. 2001. Canada's plant hardiness zones revisited using modern climate interpolation techniques. *Canadian Journal of Plant Science* 81:129–143.
- McPherson, J. M., and W. Jetz. 2007. Effects of species' ecology on the accuracy of distribution models. *Ecography* 30: 135–151.
- Meehl, G. A., et al. 2007. Global climate projections. Pages 747–845 in S. Solomon, D. Qin, M. Manning, M. Marquis, K. Averyt, M. Tignor, H. L. Miller, and Z. Chen, editors. *Climate change 2007: The physical science basis*. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press, Cambridge, UK.
- Menendez, R., A. Gonzalez Megias, J. K. Hill, B. Braschler, S. G. Willis, Y. Collingham, R. Fox, D. B. Roy, and C. D. Thomas. 2006. Species richness changes lag behind climate change. *Proceedings of the Royal Society B* 273:1465–1470.
- Midgley, G. F., L. Hannah, D. Millar, M. C. Rutherford, and L. W. Powrie. 2002. Assessing the vulnerability of species richness to anthropogenic climate change in a biodiversity hotspot. *Global Ecology and Biogeography* 11:445–451.
- Nieminen, M., H. Rita, and P. Uuvana. 1999. Body size and migration rate in moths. *Ecography* 22:697–707.
- Parmesan, C. 1996. Climate and species' range. *Nature* 382: 765–766.
- Parmesan, C., and G. Yohe. 2003. A globally coherent fingerprint of climate change impacts across natural systems. *Nature* 421:37–43.
- Pearson, R. G., and T. P. Dawson. 2003. Predicting the impacts of climate change on the distribution of species: Are bioclimate envelope models useful? *Global Ecology and Biogeography* 12:361–371.
- Pearson, R. G., C. J. Raxworthy, M. Nakamura, and A. Townsend Peterson. 2007. Predicting species distributions from small numbers of occurrence records: a test case using cryptic geckos in Madagascar. *Journal of Biogeography* 34: 102–117.

- Pearson, R. G., W. Thuiller, M. B. Araújo, E. Martinez-Meyer, L. Brotons, C. McClean, L. Miles, P. Segurado, T. P. Dawson, and D. C. Lees. 2006. Model-based uncertainty in species range prediction. *Journal of Biogeography* 33: 1704–1711.
- Peterson, A. T., E. Martinez-Meyer, C. Gonzalez-Salazar, and P. W. Hall. 2004. Modeled climate change effects on distributions of Canadian butterfly species. *Canadian Journal of Zoology* 82:851–858.
- Peterson, A. T., V. Sanchez-Cordero, J. Soberon, J. Bartley, R. H. Buddemeier, and A. G. Navarro-Siguenza. 2001. Effects of global climate change on geographic distributions of Mexican Cracidae. *Ecological Modelling* 144:21–30.
- Peterson, A. T., H. Tian, E. Martinez-Meyer, J. Soberón, V. Sánchez-Cordero, and B. Huntley. 2005. Modeling distributional shifts of individual species and biomes. Pages 211–228 in T. E. Lovejoy and L. Hannah, editors. *Climate change and biodiversity*. Yale University Press, New Haven, Connecticut, USA.
- Phillips, S. J., R. P. Anderson, and R. E. Schapire. 2006. Maximum entropy modeling of species geographic distributions. *Ecological Modelling* 190:231–259.
- Pöyry, J., M. Luoto, R. K. Heikkinen, and K. Saarinen. 2008. Species traits are associated with the quality of bioclimatic models. *Global Ecology and Biogeography* 17:403–414.
- R Foundation for Statistical Computing. 2007. R version 2.6.1. (<http://www.R-project.org/>)
- Ramankutty, N., and J. A. Foley. 1999. Estimating historical changes in global land cover: croplands from 1700–1992. *Global Biogeochemical Cycles* 13:997–1027.
- Root, T. L., J. T. Price, K. R. Hall, S. H. Schneider, C. Rosenzweig, and J. A. Pounds. 2003. Fingerprints of global warming on wild animals and plants. *Nature* 421:57–60.
- Sala, O. E., et al. 2000. Global biodiversity scenarios for the year 2100. *Science* 287:1770–1774.
- Segurado, P., and M. B. Araújo. 2004. An evaluation of methods for modeling species distributions. *Journal of Biogeography* 31:1555–1568.
- Stockwell, D. R., and A. T. Peterson. 2002. Effects of sample size on accuracy of species distribution models. *Ecology Modelling* 148:1–13.
- Swets, J. A. 1988. Measuring the accuracy of diagnostic systems. *Science* 240:1285–1293.
- Thomas, C. D., et al. 2004. Extinction risk from climate change. *Nature* 427:145–148.
- Thuiller, W. 2003. BIOMOD. Optimizing predictions of species distributions and projecting potential future shifts under global change. *Global Change Biology* 9:1353–1362.
- Thuiller, W., M. B. Araújo, R. G. Pearson, R. J. Whittaker, L. Brotons, and S. Lavorel. 2004. Biodiversity conservation: uncertainty in predictions of extinction risk. *Nature* 430. [doi:10.1038/nature02716]
- White, P. J., and J. T. Kerr. 2006. Contrasting spatial and temporal global change impacts on butterfly species richness during the 20th century. *Ecography* 29:908–918.
- White, P. J., and J. T. Kerr. 2007. Human impacts on environment-diversity relationships: evidence for biotic homogenization from butterfly species richness patterns. *Global Ecology and Biogeography* 16:290–299.
- Willis, K. J., M. B. Araujo, K. D. Bennett, B. Figueroa-Rangel, C. A. Froyd, and N. Myers. 2007. How can a knowledge of the past help to conserve the future? Biodiversity conservation and the relevance of long-term ecological studies. *Philosophical Transactions of the Royal Society B* 362:175–186.
- Willis, K. J., and H. J. B. Birks. 2007. What is natural? The need for a long-term perspective in biodiversity conservation. *Science* 314:1261–1265.

APPENDIX A

Table showing numbers of records and model accuracy by species name (*Ecological Archives* E090-154-A1).

APPENDIX B

Map showing change in mean growing season temperature between 1900–1930 and 1960–1990 (*Ecological Archives* E090-154-A2).