uOttawa

L'Université canadienne
Canada's university

FACULTÉ DES ÉTUDES SUPÉRIEURES ET POSTOCTORALES

uOttawa

L'Université canadienne
Canada's university

FACULTY OF GRADUATE AND POSDOCTORAL STUDIES

Julie Nadeau
AUTEUR DE LA THÈSE / AUTHOR OF THESIS

M.Sc. (Biology)
GRADE / DEGREE

Department of Biology
FACULTÉ, ÉCOLE, DÉPARTEMENT / FACULTY, SCHOOL, DEPARTMENT

Testing the Reliability of Canada-wide and Regional Species Distribution Models with Independent Field Surveys and Evaluating Their Use for Conservation
TITRE DE LA THÈSE / TITLE OF THESIS

J. Kerr
DIRECTEUR (DIRECTRICE) DE LA THÈSE / THESIS SUPERVISOR

D.J. Currie
CO-DIRECTEUR (CO-DIRECTRICE) DE LA THÈSE / THESIS CO-SUPERVISOR

A. Morin

R. Sargent

L. Fahrig

Gary W. Slater
Le Doyen de la Faculté des études supérieures et postdoctorales / Dean of the Faculty of Graduate and Postdoctoral Studies

Testing the reliability of Canada-wide and regional species distribution models with independent field surveys and evaluating their use for conservation

Julie L. Nadeau

Thesis submitted to the Faculty of Graduate and Postdoctoral Studies, University of Ottawa, in partial fulfillment of the requirements for the M.Sc. Degree in the Ottawa-Carleton Institute of Biology.

Thèse soumise à la Faculté des études supérieures et postdoctorales, Université d'Ottawa en vue de l'obtention de la maîtrise en sciences de l'Institut de biologie Ottawa-Carleton

Library and Archives
Canada

Published Heritage
Branch

395 Wellington Street
Ottawa ON K1A 0N4
Canada

Bibliothèque et
Archives Canada

Direction du
Patrimoine de l'édition

395, rue Wellington
Ottawa ON K1A 0N4
Canada

Your file Votre référence
ISBN: 978-0-494-69028-4
Our file Notre référence
ISBN: 978-0-494-69028-4

NOTICE:

The author has granted a non-exclusive license allowing Library and Archives Canada to reproduce, publish, archive, preserve, conserve, communicate to the public by telecommunication or on the Internet, loan, distribute and sell theses worldwide, for commercial or non-commercial purposes, in microform, paper, electronic and/or any other formats.

The author retains copyright ownership and moral rights in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

AVIS:

L'auteur a accordé une licence non exclusive permettant à la Bibliothèque et Archives Canada de reproduire, publier, archiver, sauvegarder, conserver, transmettre au public par télécommunication ou par l'Internet, prêter, distribuer et vendre des thèses partout dans le monde, à des fins commerciales ou autres, sur support microforme, papier, électronique et/ou autres formats.

L'auteur conserve la propriété du droit d'auteur et des droits moraux qui protège cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

In compliance with the Canadian Privacy Act some supporting forms may have been removed from this thesis.

While these forms may be included in the document page count, their removal does not represent any loss of content from the thesis.

Conformément à la loi canadienne sur la protection de la vie privée, quelques formulaires secondaires ont été enlevés de cette thèse.

Bien que ces formulaires aient inclus dans la pagination, il n'y aura aucun contenu manquant.

Canada

Acknowledgements

I wish to thank my thesis supervisors, Dr. Jeremy Kerr and Dr. David Currie for their guidance, knowledge and patience throughout this valuable learning experience, for their financial and academic support, and for opportunities to attend conferences. I thank my committee members Dr. Antoine Morin and Dr. Lenore Fahrig for their time, comments and helpful expertise during this project. I am thankful to Dr. Morin for all the statistical help he also provided me.

I am grateful to the Orville Erickson Memorial Scholarship Fund, the Pearl Weinberger Award and R.W. Tomlinson Limited for the support they provided me. I am thankful to Gatineau Park for a scientific permit, the City of Ottawa Community Forest, the Warwick Demonstration Forest, Queen's University Biological Field station, Mr. Ian Carlisle and R.W. Tomlinson Limited for granting me the right to carry out surveys on their land. Heartfelt thanks to the Ottawa field naturalist club for data access and reporting from local members.

I sincerely thank Ross Layberry, Peter W. Hall, and Don Lafontaine for invaluable advice, expertise and identification of butterflies and Mrs. Linda Jeays for insight on butterfly sampling. I thank all Kerr and Currie lab members for feedback on this thesis, statistical and GIS assistance, for interesting discussions and memorable times. In particular, thank you to Dr. Maxim Larrivée, Dr. Adam Algar, Jay Fitzsimmons and Alexander MacDonald for statistical help; to Heather Kharouba and Eric Young for their work on the national models; to Zuzu Gadallah for her expertise with remote sensing and the mysterious world of GIS; and to Nora Szabo for her friendship, her stellar work as a brilliant field assistant, and her help with statistics and GIS.

I also wish to thank my family, Nadeau, Williams and Pagé alike, and my friends for their encouragement and support, for always being there when I needed them, and understanding when I was just too busy. Papa et maman, je suis ici grâce à vous; je vous serai toujours reconnaissante pour vos encouragements et toutes les expériences que vous m'avez rendu possibles en croyant en moi. Finally, I thank my husband James Pagé for many reasons, but above all for his unwavering support along the way, his friendship, his love, and his attitude towards life that made this all possible.

Table of Content

List of Tables

List of Figures

Abstract

For purposes of conservation, we require methods to predict where species occur currently, and where they are likely to move in the future. Species distribution models can be useful tools, by predicting where species are in areas that cannot be surveyed but where management decisions must occur. However, they are rarely tested using independent field surveys to determine their accuracy. These models are also generally developed over broad spatial scales; in contrast, distributions must often be predicted over local to regional scales for conservation purposes. Using butterfly surveys along transects in eastern Ontario, we tested whether independent regional observations of butterflies could be predicted from distribution models developed on Canada-wide species ranges. We also developed a set of regional models using Maximum entropy software to compare with the independent field surveys. Our results show that regional field occurrences are not accurately predicted by the national models, which explain between 0 and 31% (median 4.2%) of the deviance in species presence/absence. Even when species distribution models developed using regional variables have high apparent accuracy (AUC values exceed 0.8), they do not appear to accurately predict occurrence at local spatial extents, explaining between 0 and 46% (median 3.4%) of the deviance in species presence/absence. Species distribution models' inaccuracy in predicting species' presence within their ranges indicates that field-based observations remain indispensable when making conservation decisions. Misusing models can be costly if they lead to errors in predicting the presence of species of concern or their habitat and can result in poor management decisions if used improperly.

Résumé

À des fins de conservation, nous devons tenter de prédire où les espèces se trouvent présentement, et où leurs répartitions se déplaceront vraisemblablement dans le futur. Les modèles de distribution des espèces peuvent être des outils de conservation utiles car ils donnent une prévision de la présence d'espèces dans des endroits qui ne peuvent être inventoriés mais où des décisions de gestion sont imminentes. Ces modèles sont utilisés mais rarement mis à l'épreuve à l'aide d'inventaires indépendants pour déterminer leur justesse. De plus, ils sont généralement développés à une grande échelle spatiale, tandis qu'ils doivent souvent être utilisés à l'échelle régionale à des fins de conservation. Nous avons évalué si des modèles de distribution d'espèces à l'échelle nationale peuvent prédire des inventaires de papillons indépendants à travers la région d'Ottawa. Nous avons développé une série de modèles régionaux en utilisant le logiciel Maxent, que nous avons aussi comparé avec les inventaires. Nos résultats démontrent que les observations d'inventaires ne sont pas prédites avec justesse par les modèles nationaux qui expliquent entre 0 et 31% (médiane 4.2%) des présences/absences, ni par ceux à l'échelle régionale, expliquant entre 0 et 46% (médiane 3.4%) des présences/absences. Même lorsque les modèles semblent démontrer une justesse élevée (AUC dépassant 0.8), et déterminer les limites de la répartition géographique des espèces, ils ne sont pas fiables pour prédire les endroits particuliers où les espèces se trouvent à l'échelle régionale. Ces résultats soulignent l'importance des observations sur le terrain pour permettre la prise de décision en conservation. L'utilisation erronée de modèles peut être coûteuse. Une mauvaise prédiction de la présence d'espèces en péril, ou de leur habitat, pourrait provoquer des décisions de gestion sans utilité ou même dangereuses pour la survie des espèces en question.

Introduction

Predicting the environmental conditions within which species occur is critical for purposes of conservation. Species distribution models (SDM) associate species occurrences to ecological and environmental characteristics in an attempt to forecast the suitable range of a species (Guisan and Thuiller 2005; Soberon and Peterson 2005). In terms of species' conservation, SDMs allow for science-based decision making in areas that have not necessarily been surveyed for a particular species, but where immediate conservation action is required. They have been used in the contexts of conservation biology and management (e.g. Peterson et al. 2000; Scott et al. 1993; Loiselle et al. 2003; Willis et al. 2007; Schadt et al. 2002). For example, Gap Analysis, used by the United States National Biological Service since the 1990s, uses species presence, land cover, elevation and other environmental data to identify priority areas for future reserves (Scott et al. 1993; Edwards et al.). SDM utility depends on their accuracy (Araújo et al. 2005; Marmion et al. 2009), which is often difficult to test rigorously due to limited data (but see Thuiller, 2004; Lawler et al., 2006).

Because global changes affect species distributions, effective models should also predict the effects of climate and land use changes on species' distributions (Pimm et al. 1995; Araújo and Rahbek 2006; Brook et al. 2008). However, predictions of species' responses to future climate change are uncertain and difficult to determine (Thuiller 2004; Araújo & New 2006; Lawler et al. 2006; Pearson et al. 2006). One view is that species' recent historical responses to climate change may be essential to calibrate models that predict future change and provide pseudo-experimental tests of model results (Kharouba et al. 2009). Kharouba et al. (2009) used historical (from early 20th century) species observations and environmental data to calibrate models of species distributions, which were then used to forecast contemporary species ranges and further compared to predictions of species distributions modelled using present climatic variables. They found that both historical and current models generally fit the data well when tested using random subsets of the original data. The predictive models estimated the actual modeled ranges well for most species. However, these results were not tested against independently gathered data, as is seldom done for models used to forecast species distributions (MacKey and Lindenmayer, 2001). Before predictive models can be used for conservation, particularly those models that include a predictive climate component, it is imperative to test the models' ability to predict species current distributions in data sets that were systematically surveyed, do not consist solely of presence-only data, and were not involved in model construction. This is where field testing should become a priority, and is necessary if models are to be used appropriately for a given conservation purpose.

Models are most often tested internally, using the area under the receiver operator characteristic curve (AUC), a standard estimate of model accuracy (Fielding and Bell, 1997). The available dataset is divided into two parts and a portion of the occurrence points is used to build the model (typically 70%) and the remainder of the data to test the model (Fielding and Bell, 1997). AUC is determined by calculating the area under the curve obtained when plotting the proportion of true positives (sensitivity) as a function of the proportion of false positives (respective 1-specificity values) for all possible presence/absence thresholds (Fielding and Bell, 1997). AUC values thus indicate how well the model discriminates between presences and absences and range between 0 and 1, where 0.5 indicates that the model performs no better than expected randomly (Swets 1988). Although AUC is a common goodness-of-fit measure for distribution models (Fielding and

Bell, 1997), recent studies have questioned the appropriateness of this measurement (Lobo et al. 2008; Peterson et al. 2008; Veloz 2009) namely because of the equal weight given to false positives (commission errors of the models), and false negatives (omission errors of the models) in its calculation, and because of the relationship between the training and testing sets used to calculate this value. For conservation purposes, false positives may be more significant because they could indicate falsely the presence of suitable habitat when it is in fact absent (Loiselle et al. 2003). This approach can however be viewed as more precautionary if large areas are being protected.

Models of species' distributions are most typically developed over broad spatial scales (national, continental or global extent, coarse grain). In contrast, conservation decisions frequently need predictions at the local to regional scales (Margules & Pressey 2000; Guisan and Thuiller 2005). Broad scale approaches are more common, we suspect, because of data availability and analytical issues: 1) coarse-grained species distribution and environmental data are more readily available than fine-grained data, and 2) species ranges and habitat variables have greater variance over broad spatial scales, which yields stronger statistical relationships. However, this broad approach is likely to capture variation in species distributions that is related to habitat characteristics that vary over broad spatial scales such as climate (Pearson and Dawson 2003; McPherson and Jetz 2007; Field et al. 2009). Within particular regions, factors such as climate vary much less; other factors are likely to affect where species may be found within local areas. Biotic interactions, microhabitat (e.g. host plant availability) and habitat disturbance, for example, are rarely included in broad scale models, but could strongly affect species distributions at regional scales (Uehara-Prado et al. 2009). Models built at broad scales, are often used for management at different scales due to their availability, even though the hypothesis that distribution models developed with broad-scale distribution data can be applied in specific regions to predict finer-scale variation in species presence has rarely been tested.

The objectives of this study are twofold. The first is to test whether regional variation in butterfly presence observed in systematic field surveys in Eastern Ontario can be predicted from broad scale (large extent, coarse grain) distribution models developed for Canada-wide species ranges. The second objective is to construct models using variables at a finer resolution over the Eastern Ontario region and determine if they provide significantly better predictions of regional species occurrences. This process allows a test of both broad scale and regional scale model performance relative to an independently constructed dataset based on systematically surveyed field observations of a large species assemblage.

Methods

Predictions of the variation in habitat suitability for each butterfly species were given by Maximum Entropy models (Maxent – Phillips et al. 2006) trained using two different data sets, one of national extent and using coarse grained environmental data, and one focusing on the Ottawa region related to small grained variables. The predictions of the Maxent models were then tested against a set of butterfly presence/absence observations from systematic transects field surveys in the Ottawa region (Figure 1).

Study taxa

Butterflies were used as a study taxon to carry out the tests of the predictive models. Butterflies are attractive to naturalists and straightforward to identify due to their well defined taxonomy. Consequently, a substantial database exists of observations of butterfly presences in Canada that were opportunistically gathered (following no systematic protocol) by naturalists through the 20th century (Layberry et al. 1998). Butterflies have been extensively used in distribution modelling thanks to the availability of data and because, as highly vagile poikilotherms, their activities (reproduction, migration, dispersal) and distributions are strongly related to climatic conditions (Hughes 2000, Hill et al. 2001; Peterson et al. 2004).

Maximum entropy species distribution models

We selected Maxent (Phillips et al. 2006) to model the distributions of suitable habitat of the butterfly species. Maxent relates observations of species presences to the environmental variables where the species was observed and determines the area of best fit (maximum entropy). The fitted model can then be used to calculate a suitability score between 0 and 1 for any given location, based on its environmental characteristics where 0 represents regions where the environmental data is not suitable for the species and 1 represents ideal conditions. Maps of habitat suitability for a species can thus be produced (Phillips et al. 2006). Maxent performs as well or better than most other modelling techniques (e.g., Bioclim, Domain, GARP; Elith et al. 2006), particularly for sample sizes <100 (Hernandez et al. 2006; Wisz et al. 2008). Maxent modelling accuracy remains more stable than other techniques when sample sizes vary between 5 and 100 observations (per species) (Hernandez et al. 2006) and is a suitable technique for presence-only data (Phillips et al. 2006). An AUC value is then calculated for the models to determine their accuracy.

Broad scale species distribution models

The first set of models used for this study was based on models built by Kharouba et al. (2009). Minor modifications were made to the Kharouba models (Jay Fitzsimmons and Eric Young, unpublished data) by removing a variable delimiting Canadian ecozones, and by using Maxent's logistic output, which allows for comparisons between species by standardising the output through scaling between 0 and 1. Records of butterfly occurrences were from the Canadian National Collection of Butterflies (Layberry et al. 1998), a collection composed of close to 300,000 butterfly records from 297 Canadian species, georeferenced to within 1km (for approximately 80% of the records). The majority of records are professionally identified and preserved butterfly specimens, kept in 40 collections throughout Canada. We used these constructed habitat models for 58 species found in the Ottawa region. Rarer species were not included because there were too few observations to construct habitat models.

Seven environmental variables drawn from standard published sources were used to build the models at the national scale. These variables included human population density (humans/km^2) from the 1981 census (White & Kerr, 2006), land cover and elevation (Beaubien et al. 2000), and climate data (Canadian Forest Service and the Atmospheric Environment Service; McKenney et al., 2001; Mekis & Hogg, 1999), including total annual precipitation, maximum monthly temperature averaged over the growing season (April to October), minimum annual temperature, and mean growing season temperature.

For a given species, 70% of the species occurrence data was used in Maxent to train a distribution model, and the remaining 30% was used as a testing set. The resulting model was then used to predict habitat suitability for the species across Canada in grid cells at a resolution of ~6.61 by 6.61 km. The size of the output cells for the broad scale models were limited by the coarseness of the climatic variables. This process was repeated ten times, randomly allocating observations to the training and the test data sets every time. The predicted suitability per grid cell was averaged over the 10 iterations to yield the final model used in this study.

Regional scale species distribution models

The second set of Maxent models was built for this study using finer-grained environmental and species occurrence data. Species occurrence records include both specimens and observations of butterflies by local naturalists in Eastern Ontario from 1967 to 2008 (Ross Layberry, pers. comm.; observations recorded in GBIF database). Some of these records are also present in the Butterflies of Canada database, but this regional dataset includes sightings of butterflies that were not collected and preserved in Canadian collections. Sightings ranged between 25 and 1258 observations per species, with only three species having fewer than 50 sightings. This falls well within the range specified as optimal by Hernandez et al. (2006) who indicated sample sizes of 100 yielded the highest AUC, but that AUC values were close to their maximum with 50 observations. Non-geographically unique data points were automatically deleted by Maxent.

Similar parameters to those of the broad scale models were measured using much more finely-grained environmental variables at the regional scale; the regional environmental observations used to construct Canada-wide models are too coarse to detect fine-scale variability within the study region. These regional variables included human population density in 2006 (LandScanTM Global Population Database; Oak Ridge, TN: Oak

Ridge National Laboratory; available at http://www.ornl.gov/landscan/), high resolution digital elevation data from the Shuttle Radar Topography Mission (SRTM, 90m resolution; Jarvis et al. 2008), incoming solar radiation (derived from the digital elevation data – Appendix 1) and land cover divided in 32 classes (GeoBase.ca). Incoming solar radiation was used as a surrogate for local temperature variability, because the macroclimatic variables (temperature and precipitation) vary little across the region (Appendix 1 for data processing prior to use in the models).

We constructed high resolution Maxent models for the Ottawa region, following the procedure outlined above. The models make suitability of presence predictions for cells having a resolution of ~180 meters (0.00177 decimal degrees). Cell size was chosen to match transect length (see following section) and corresponds to twice the resolution of the elevation data. For each species, the model was carried out 10 times, each time with a different random training and testing set, and the results were averaged over the 10 iterations.

Regional butterfly field surveys

We carried out field surveys of butterflies to test the accuracy on a regional scale of the aforementioned two sets of Maxent models (coarse-grained national extent model, and fine-grained regional extent models). Twenty-four sites in the Ottawa region were each visited between two and five times from May 18, 2007 to September 1st 2007, between 9:30am and 16:30pm on days with weather and wind conditions favourable to butterfly activity (Pollard and Yates, 1993). We carried out several visits throughout the summer to increase the probability of detection of butterfly species with different flight periods. Most sites (fourteen) were surveyed four times. Seven sites were surveyed five times, one site was surveyed three times and two were surveyed twice (because of the vagaries of weather).

Study sites were selected to include both high and low diversity sites to maximize variation of presence/absence in the region, and based on advice from the authors of the monograph "Butterflies of Canada" (Layberry et al. 1998) and local experts. The sites covered an array of habitats and could be classified in five general habitat types: field, forest path, rock barren, areas adjacent to marshes, and roadsides (Table 1, Figure 2). We aimed to maximize the diversity gradient among sites that could provide butterfly habitat, not to distinguish entirely unsuitable sites. Consequently, we did not sample sites where butterflies were not expected to be readily found (e.g. dense conifer forest interior or urban sites) and, as a result, samples included few sites for which the national models predicted very low suitability (close to 0). In general, we visited high butterfly diversity sites, which was also the case for specimens in the "Butterflies of Canada" (Layberry et al. 1998) and Canadian National Collection of Butterflies. Therefore, the variation between the survey sites was generally analogous to the variation in the broad scale models. Several species in this study are at the edges of their ranges in the Ottawa region, making this survey area particularly pertinent in the context of climate change.

Surveys were completed along standard transects (Pollard and Yates, 1993) 200m x 6m in size (Kerr et al. 2000; Grill and Cleary, 2003; Davros et al. 2006), by two observers, each surveying one side of the transect. Transects were walked at a pace of 10m/min to ensure constant sampling effort (Davros et al. 2006), stopping only to capture and identify butterflies, record information, photograph specimens and to georeference the observation locations. We also recorded all other butterfly species that we observed outside transects.

We carried out the analyses outlined below with both the "transect" species and all species at each site.

Butterflies were identified in the field whenever possible. We used a standard butterfly net when we needed to capture the butterflies for identification and then released the specimens. Specimens that could not be identified in the field were collected, preserved in ethyl acetate and identified later using field guides (Opler 1994; Layberry et al. 1998; Brock and Kauffman 2003). Butterflies that could neither be caught nor identified on the wing were recorded down to genus or as unknown. Voucher specimens were collected for all species that were difficult to identify (Grill and Cleary, 2003).

To determine how well we detected species present in the area, we constructed species rarefaction curves using the Ottawa 2007 field sightings, and the Estimate S software (Colwell 2005). We used Coleman's rarefaction index (Gotelli and Colwell 2001, Colwell and Coddington, 1994) to model the number of species observed as a function of sampling effort. We performed three analyses, one pooling all species for all sites in the region, another splitting the data according to the number of visits per site, and the last on an individual site basis. We also re-sampled all sites once during the summer of 2008 to verify the completeness of the data collected in the summer of 2007.

Analyses

For both the national scale and regional scale models, we used ArcGIS (Environmental Systems Research Institute 2006) to obtain the suitability predicted by the modelling results for each species at each field site surveyed in 2007. To do so, the sighting points from the butterfly field surveys were overlaid on a map of the predictive models' suitability outputs. A suitability value of each species' model at each sighting point was extracted, allowing a statistical comparison of species presences/absences from field surveys with the modelled habitat suitability. If a transect crossed multiple cells, suitabilities of the cells were averaged by taking the proportion of transect in a cell into account (Appendix 1 for details).

To determine how well the models could predict the presence of species in our field sites, we modelled species presence/absence at the field sites surveyed in 2007 (dependent variable) as a function of habitat suitability (estimated by the Maxent models) using logistic regressions (S-Plus statistical software: Insightful Corp. 2007). The regressions were executed for every species found in the area of study for which a model had been derived. Of the 67 species present at the field sites around Ottawa, there were sufficient data on 58 to build and test national habitat models, and 65 to test regional models.

We tested the predictions of the Canada-wide models and the Ottawa region models using both the butterfly species found on the transect and all species observed at each site (i.e. on or off the transect) as functions of the predicted habitat suitability. We conducted a sensitivity analysis for the Canada-wide models by repeating the two sets of regressions after eliminating six sites that fell within the same cell as another site. For the Canada-wide models, all sets of analyses led to the same conclusions; we therefore only present the results of the regressions using all the sites and all of the species here. The same applies for the Ottawa regional models' analysis.

Pseudo-R^2 values were calculated to assess goodness-of-fit of the logistic regressions using equation 1 (Guisan and Zimmermann 2000, Prugh et al. 2008).

$$\text{Pseudo-}R^2\ (D^2) = \frac{\text{Null deviance} - \text{Residual deviance}}{\text{Null deviance}} \qquad \text{(Eq. 1)}$$

Null deviance and residual deviance were obtained from the S-plus output of each logistic regression. The pseudo-R^2 (also referred to as D^2) indicates the amount of the deviance explained by the model. Its scale ranges from 0 to 1 with higher numbers indicating a better fit and thus resembles the ordinary least squares (OLS) R^2, although pseudo-R^2 values are typically smaller than OLS R^2 values (Hagle and Mitchell 1992).

The resulting pseudo-R^2 values from the models were further plotted against butterfly species commonness, the suitability range of the surveyed sites for each species, and the AUC value of the regional Maxent models to determine if the prediction accuracy of the logistic models varied as a function of these factors. The pseudo-R^2 values not being distributed normally, they were transformed using an arcsine square root transformation prior to this secondary analysis.

Results

Field surveys

During the summer of 2007, 4064 butterflies from 67 species were surveyed at the 24 eastern Ontario field sites (details in Appendix 2). Of these, 1739 were European skippers. Only 326 of the remaining 2325 specimens (8%) could not be identified to the species level. In 2007, 1531 specimens from 90 species were observed by naturalists in the Ottawa region (Ross Layberry, pers.comm.), but a number of these species were specialists or naturally rare and were excluded from our modelling efforts. Between 43 and 90 species per year were found by local naturalists in the Ottawa region between 1998 and 2007 (average: 68.2 ±16.3; Ross Layberry, pers. comm.).

Our sampling in 2007 appears to have been adequate to record most of the species that were present in the region. A Coleman's rarefaction curve indicates that most species of the region were found with the 24 sites surveyed (Figure 3a). The rarefaction curves of sites visited four and five times overlapped (Figure 3b) and the sites surveyed 5 times only show a small number of extra species, indicating a similar number of species was found at sites visited four and five times. Not enough sites were visited twice or three times to allow interpretation from the rarefaction curves from those sites. Furthermore, most species at each site seem to have been detected within three visits of a site when the cumulative number of species is plotted against the number of visits, arranged in chronological order (Figure 4a). When Coleman's rarefaction curves are carried out for individual sites, it appears that more visits would have been needed to detect all species at some of the sites, as not all the curves tend toward an asymptote (Figure 4b).

Additional visits to the field sites in 2008 yielded few new species in addition to those detected in the previous year. We observed 271 specimens representing 43 species that could be identified (20 specimens, representing 6.87%, were not identified) (Appendix 2). Individual sites had a median of 1 new species, beyond those present in 2007, thus showing an increase in new species of between 0 and 25% (Figure 5; Appendix 2 for details). Even the sites visited twice in 2007 only increased their species number by $\leq 7.7\%$, suggesting that we missed few species in our 2007 censuses even at the least visited sites. For these reasons, all sites were kept in the analysis comparing field survey observations to models predictions at both scales, and we did not correct for sampling effort in subsequent analyses.

Broad scale species distribution models: evaluation with field surveys

National Maxent models predicted field species occurrences poorly (Figure 6; Appendix 3 for pseudo-R^2 per species) despite the fact that they described Canada-wide distributions of most Canadian butterfly species well when assessed internally according to Kharouba (2007) (Average AUC = 0.979; SE = 0.0014). The pseudo-R^2 values ranged between 0 and 0.31 (explaining 0 to 31% of the deviance in species presence/absence) with a median value of 0.042, when the models were tested against independent field data. None of the relationships were significant at $p<0.05$.

The poor predictive ability of the Maxent models appears to be marginally related to low range of surveyed suitability within the region (Figure 7). Most of our sampling sites had high predicted suitability (mean suitability was 0.7). For 41 of the 58 species that were tested with models, the predicted suitability of at least one site was less than 0.5. For 18 of these 41 species, only one site surveyed had a suitability <0.5. The pseudo-R^2 depends upon the range in suitability among sites (F-statistic = 4.332, df =1 and 53, R^2 = 0.0756, p-value=0.0422), and therefore low pseudo-R^2 values could result in part from the little variation in range of suitability within the region. This is however only marginally significant, and the range in suitability explains very little of the variation in the pseudo-R^2 values.

Species were sometimes present at sites with low predicted suitability, and often absent at sites with high predicted suitability (Figure 8). Most of the observations (84%) were carried out at a predicted suitability above 0.5. The distributions of presences and absences follow a similar distribution with 87% of the presences (and 83% of the absences) occurring at sites having a predicted suitability above 0.5. The average suitability for the sites where species were present when all species are pooled was 0.72 ± 0.19, while the average suitability at the sites where they were absent was 0.68 ± 0.21. This difference is statistically significant (Wilcoxon rank-sum test: z = 2.6522, p-value = 0.008; Table 2), indicating that suitability is marginally higher at the sites where species were present (Figure 8), but is not conclusive considering the amount of overlap between the two groups.

Model performance does not depend upon how common a species is (Figure 9). The pseudo-R^2 of logistic regressions relating presence/absence to habitat suitability is unrelated to the number of sites occupied by a species (F-statistic = 0.2588, df =1 and 53, $R^2 = 0.0049$, p-value = 0.6130).

Regional scale species distribution models

Regional Maxent models described the habitat of butterflies well in the Ottawa region, with an average AUC of 0.88 (SD = 0.04) when tested internally for 71 species (Table 3 and Appendix 4 for the contribution of each factor in the model). This AUC falls within the 0.7 to 0.9 range, thus considered good on the Swets scale (Swets, 1988). For most of the models (62 out of 71), elevation was identified by Maxent as the most important of the environmental factors contributing to the model. Land cover was in all these cases the second most important factor, and the most important variable in the nine models for which elevation was not. Incoming solar radiation was not indicated as an important factor but is weakly correlated with elevation (r = -0.33) as it is calculated in part using elevation.

However, regional Maxent models do not predict independent regional species occurrences accurately. Sixty-five species models could be tested with data from the independent field surveys; the other 6 models represented species in the area that were not found during the field study. Observed presence/absence was significantly related to Maxent predictions for only one species (fewer than would be expected by chance). Regional Maxent models could statistically account for 0% to 46% of the deviance in species presence/absence, with a median of only 3.4%. Agreement is poor for the majority of the models (Figure 10; Appendix 4 for pseudo-R^2 for all species). There was also no relationship between the AUC of the Maxent models and the logistic regression pseudo-R^2 values (Spearman rank correlation rho = -0.1460, p-value = 0.2581), indicating that models with a high AUC do not predict presence/absence in the independent surveys better.

In contrast to the broad-scale models, those developed exclusively within the study region predict large differences in habitat suitability, spanning a wider range of suitabilities for each species (Figure 11). All species were surveyed at sites both above and below a suitability of 0.5. Generally, the range of suitabilities among sites for each species (suitability of site with highest suitability value – suitability of site with lowest surveyed suitability) was between 0.5 and 1.0. Pseudo-R^2 values vary slightly as a function of the range in suitability (Figure 12), indicating that the models perform somewhat better with a wider range of suitabilities (F= 9.898, df = 1 and 59, R^2=0.1437, p-value = 0.0026). The small R^2 indicates that once again the effect is small.

Species were often present at sites of low predicted suitability by the models and absent at sites with high predicted suitability (Figure 11). Roughly half of the observations (55%) were carried out at a predicted suitability above 0.5. Presences and absences follow a different distribution with 65% of the presences and 51% of the absences occurring at sites having a predicted suitability above 0.5. The remaining 35% of presences were at sites with a predicted suitability below 0.5. The average suitability for the sites where species were present when all species are pooled was 0.54 ± 0.18, while the average suitability at the sites where they were absent was 0.48 ± 0.21; these values are significantly different (Wilcoxon rank-sum test: $z = -4.8315$, p-value < 0.0001; Table 2), indicating suitability is marginally higher at the sites where species were present (Figure 11), but once again the overlap is too important for this result to be considered biologically conclusive.

The predictive ability of regional scale models was marginally related to the commonness of the species (Figure 13). Species with the highest pseudo-R^2 tended to be rare ($F= 6.242$, df = 1 and 59, $R^2=0.0957$, p-value is 0.01528) indicating that only a very small number of presences (one or two sightings) can drive the relationship to have a higher pseudo-R^2.

Discussion

Maxent species distribution models failed to predict species presence/absence in an independent field survey in the Ottawa region, even though their apparent accuracies are high. This is true for both models calibrated at the national and the regional scales. The models appear to be useful on the basis of high AUC and because they seem to provide one of the few means of generating predictions of species' distributions in areas that are not surveyed (Kharouba et al. 2009). They can be a convenient complement to surveys, which are often too expensive and time consuming to be carried out in a timely fashion. Although AUC values remain one of the principal means for accuracy assessment, they should be interpreted very conservatively (Lobo 2008; Veloz 2009) due, at least in part, to the spatial autocorrelation of the training and testing sets used during model construction and evaluation (Veloz 2009). Even if broad scale models are one of the few tools available to predict species distributions regionally, their accuracy among individual field sites is questionable and their practical utility may be limited to broader extents and coarser spatial grains.

Our results are consistent with those few studies that have tested species distribution model accuracy with independent field observations. Models' predictions typically show only moderate agreement with independent field observations (Greaves et al. 2006; Inglis et al. 2006) even when the predictive models' accuracy appears excellent (i.e. AUC values in constructing the models approach their maximum value). In this study, models developed using Maximum Entropy for Canadian butterflies (Kharouba et al. 2009) were unrelated to field observations. Similar models that we developed using much higher resolution data specific to the study region did not improve the agreement between models' predictions of species distributions relative to field observations of true species presences/absences.

The mismatch between model predictions of species habitat suitability and species presence in the field is unlikely to be a result of sampling intensity. We recorded 4064 butterfly observations (2325 when excluding European skippers) for 67 different species at 24 sites in 2007. The rarefaction curve for the entire region indicates we have found most of the species in the area, and most species at each site seem to have been detected after three visits of the site. Surveys in 2008 increased the numbers of species at each site only marginally, indicating both that field observations were sufficiently complete to detect nearly all species and that inter-annual variation in butterfly species' presences at individual field sites were small and unlikely to alter our conclusions. Sampling intensity during this study was higher than would be expected in most conservation situations.

It is possible that Maxent model predictions failed to predict regional presence/absence because our accuracy assessment relies on data from a single year. It is conceivable that meta-population dynamics cause local extinctions and recolonizations. A temporary absence in the year we sampled would appear as a lack of agreement between predicted high suitability and observed absence. A species may be consistently present within a region but only intermittently present at a field site. Beyond the potential effects of sampling (see above), we believe this possibility is unlikely for two reasons. First, model predictions do not match field observations for ubiquitous species (e.g. the introduced Cabbage White butterfly) better than for the rarest species for which localized extinction could more strongly limit field-based detection. To the extent that there is any trend, it is that rarer species are marginally better predicted by regionally-calibrated models than common species. Second, if meta-population dynamics were the main cause of the lack of fit, we would have expected species that were absent from particular sites in 2007 to have recolonized those sites more frequently in 2008. The small number of new species observed in any site in 2008 is inconsistent with this expectation.

Failure of the species distribution models to accurately predict at the regional level could have arisen due to simple scale mismatch. This certainly accounts, at least in part, for the poor predictive ability at the regional level of models tuned at the national level. Different factors are believed to limit species' distributions at different sampling scales, with climatic factors predominating at the broadest spatial extents, while habitat characteristics and biotic interactions are more important at small extents (Pearson and Dawson 2003). Although SDMs are capable of detecting boundaries at broad scales (e.g. Elith et al. 2006), such as those imposed by climate, they may do less well in predicting local absence within a species' range.

Poor model fit to field observations may also occur if butterflies select habitats on the basis of factors that we did not measure. For example, well known interactions between butterflies and host plants can affect SDM accuracy. Including other variables such as biotic interactions and dispersal (Pearson and Dawson 2003) in the modelling process might improve the predictive ability of the models when assessed with independent field surveys. However, our regional models used high resolution, spatially detailed data, and had AUC values above 0.74, suggesting that the models should make accurate predictions; they still failed to predict regional presence/absence within the very region they were developed.

It is important to distinguish between model predictions of suitable territory for a species (the role of habitat models) versus direct field observation of species' presence and absence. The models predict habitat suitable for a particular species. The actual occupation of all that habitat by the species is likely to be different. Even if SDMs, such as those using Maximum Entropy, correctly identify an area as suitable, species may be absent from the area for many reasons. These may include historical contingency, such as a disturbance from which the species has not yet recovered sufficiently to permit local recolonization, difficulty dispersing beyond a barrier of unsuitable area or other factors such as competitive exclusion.

Despite the many visits at each site, we must note that it is impossible to confirm a failure to detect a species as an absence (Anderson et al. 2003). Butterflies have to be both present and observed in a cell to be considered present, which might also account for the high number of absences at sites presenting high predicted suitability. If the surveys were not as complete as the data seem to indicate, not finding species that were actually present at the sites surveyed could also have led to a lack of fit of the models. However, the rarefaction curves indicate that no new species were observed after a certain number of site visits suggesting most species were observed at the site. If a species was falsely thought to be absent, one would expect new species to be observed at subsequent site visits. Few species were found at sites where national models' suitability were below 0.5 (as also found by Greaves et al. 2006), for a total of 13% of all presences. For regional model analysis, 35% of the presences were found at a predicted suitability below 0.5. It must be noted that this trend might happen because for the national models, we surveyed the highest part of the suitability distribution produced by the models. For the regional models, because they were trained and tested at a similar scale, the entire range of suitability distribution is surveyed, thus allowing more potential for presences to be found at low suitability sites.

Presences at low suitability are potentially more problematic than absences at high suitability, because the former are clearly errors of the models. Absences at high predicted suitability might just indicate failure to detect a species even if it is present at the site, or that the site is within the species fundamental niche, but not the realised one (Soberon and Peterson 2005). A final problem is that butterflies disperse relatively readily. Some species may have been observed at field sites predicted to be unsuitable while travelling between patches, seeking to avoid predators, or simply have strayed into unsuitable territory during flight.

Maximum Entropy also generates "pseudo-absences" from a random sample of locations, some of which may be occupied. Maxent fails, we propose, because the distribution of environmental conditions of pseudo-absences (when only presences have been recorded) differs significantly from the environmental conditions of real, observed absences (from a systematic field survey). To reduce sampling bias and bias introduced by sampling effort, further work should consider using Phillips et al. (2009) target-group background technique, which compares a species' occurrences only to surveyed locations for the taxon, instead of selecting pseudo-absences at random from the entire study region.

In most modelling processes, the training and the testing datasets are not independent; they are random subsets of a single dataset, which could also be problematic. This can potentially inflate the AUC value of the model (Veloz, 2009). Absences are important to take into account, as they give an indication of how much the models overpredict suitability. For example, if we only considered presences from the surveys, the national models would be thought to perform well, with 87% of all presences found at a predicted suitability > 0.5 and only 13% < 0.5. The regional models would appear to have a lower predictive ability with 65% of presences found at a predicted suitability > 0.5 and 35% < 0.5. This assessment would miss the high number of absences at high suitability, and the fact that a different suitability range is surveyed for the national and the regional models.

These findings have important implications for the use of SDMs for conservation and management decisions as well as for designating habitat for species at risk. If a species' survival hinges on protecting its critical habitat (as is the case in Canada and Ontario), and if we are unable to identify this habitat accurately, its protection can be compromised. When AUC values are compared to independent field observations, models overestimate the number of areas where species are found (commission errors – species not found in the field where predicted by the model). Over-prediction errors, although they can appear to yield a conservative distribution in the context of conservation, can be detrimental for predicting where species will be present; they might lead to protecting areas where the species of interest is not, or to failures to detect species that had declined subsequently in their distributions (Loiselle et al. 2003). The analysis shows a very high number of absences at sites with high predicted suitability for both sets of models. These results underscore the importance of testing models for model accuracy using independent field observations, and the need to be cautious in translating even region-specific models into conservation recommendations.

Conclusion

Our tests of Maxent models (presenting high AUC values) accuracy with independent field surveys indicate that national models with coarse resolution and regional models of fine resolution of butterfly species distribution cannot be used to predict where species will be found in the Ottawa area. Species distribution models are useful to forecast the general distribution of species over very broad areas that include the species' geographic limits. However, our results suggest that their application to localized conservation decisions, for regional scale predictions of exact species presence / absence, is limited. The types of distribution models used in this study should be considered as indicating potential habitat for the species, and can therefore be used to narrow down search areas for species of interest, not to indicate specific areas where the species is present.

While intensive biological surveys are instrumental for reserve selection (Balmford and Gaston 1998), their cost and high resources requirement often prevent their feasibility and hinder their completion. This study points to the need for intensive fieldwork to evaluate species distribution models regionally. According to our results, it is crucial to test SDMs with independent presence/absence field data to help build them more effectively and calibrate them before they are used for conservation.

Tables

Table 1: Field sites of butterfly transect surveys during the summers of 2007 and 2008. Field sites are identified numerically with corresponding geographic coordinates, general survey area and vegetation type. See Figure 2 for site location on the map of the Ottawa region.

Site Number	Geographic coordinates	Area	Vegetation type
1	N 45.38831, W 76.07194	Carp	Rock Barren
2	N 45.49690, W 76.07914	Constance Bay	Forested path
3	N 45.49693, W 76.07946	Constance Bay	Roadside
4	N 45.53252, W 75.99522	Gatineau Park; Luskville Falls	Field
5	N 45.25382, W 76.03848	Manion Corners (Howie Road, site 1)	Adjacent to marsh
6	N 45.25677, W 76.04151	Manion Corners (Howie Road, site 2)	Forested path
7	N 45.25046, W 76.04619	Manion Corners (Howie Road, site 3)	Field
8	N 45.26165, W 76.04810	Manion Corners (Howie Road, site 4)	Roadside
9	N 45.39206, W 75.95617	Mud Pond Area (Marchhurst)	Adjacent to marsh
10	N 45.63697, W 76.01350	Gatineau Park; Lac Philippe Region	Field
11	N 45.62452, W 76.00679	Gatineau Park; Lac Philippe Region	Adjacent to marsh
12	N 45.07148, W 75.81922	Marlborough Forest	Forested path
13	N 45.04595, W 75.86580	Marlborough Forest	Field
14	N 44.54810, W 76.36423	Chaffey's Locks	Roadside
15	N 44.54455, W76.37270	Chaffey's Locks	Forested path
16	N 44.57111, W 76.33164	Chaffey's Locks	Field
17	N 45.59963, W 76.09676	Gatineau Park; Eardley Masham Area	Field
18	N 45.42966, W 76.17934	Carp	Field
19	N 45.43206, W 76.17516	Carp	Adjacent to marsh
20	N 45.29126, W 75.55011	South Gloucester	Field
21	N 45.29353, W 75.55119	South Gloucester	Forested path
22	N 45.19886, W 75.05257	Warwick Forest, Finch Trail (Finch)	Forested path
23	N 45.58456, W 75.46329	Cumberland, East	Roadside
24	N 45.08590, W 75.48626	Winchester	Adjacent to marsh

Table 2: Average suitability of sites at which each species was present and average suitability of sites where each species was absent, pooled for all species, for both the Canada-wide and the regional models. Wilcoxon rank-sum test results comparing the average suitability of sites at which the butterflies were found to the average suitability of sites where the species were not found are included along with the sample size (n) and the significance (p) of the test (α = 0.05).

Model	Average Suitability Species Present ±SD	n present	Average Suitability Species Absent ±SD	n absent	z value	Significance (p value)
Broad scale	0.72 ± 0.19	387	0.68 ± 0.21	981	-4.0364	0.0001
Regional scale	0.54 ± 0.18	436	0.48 ± 0.21	1125	-4.8315	< 0.0001

Table 3: Average Area under the curve (AUC ± SD) for Ottawa regional scale models built using Maximum Entropy. Average is presented for n=71 species models. The average percent contribution of each environmental variable included in the models is also presented. See Appendix 4 for detailed results per species.

	Percent Contribution (%)			
AUC ± SD	**Elevation**	**Solar Radiation**	**Land Cover**	**Population Density**
0.8848 ± 0.0353	61.55	2.54	30.97	4.93

Figures

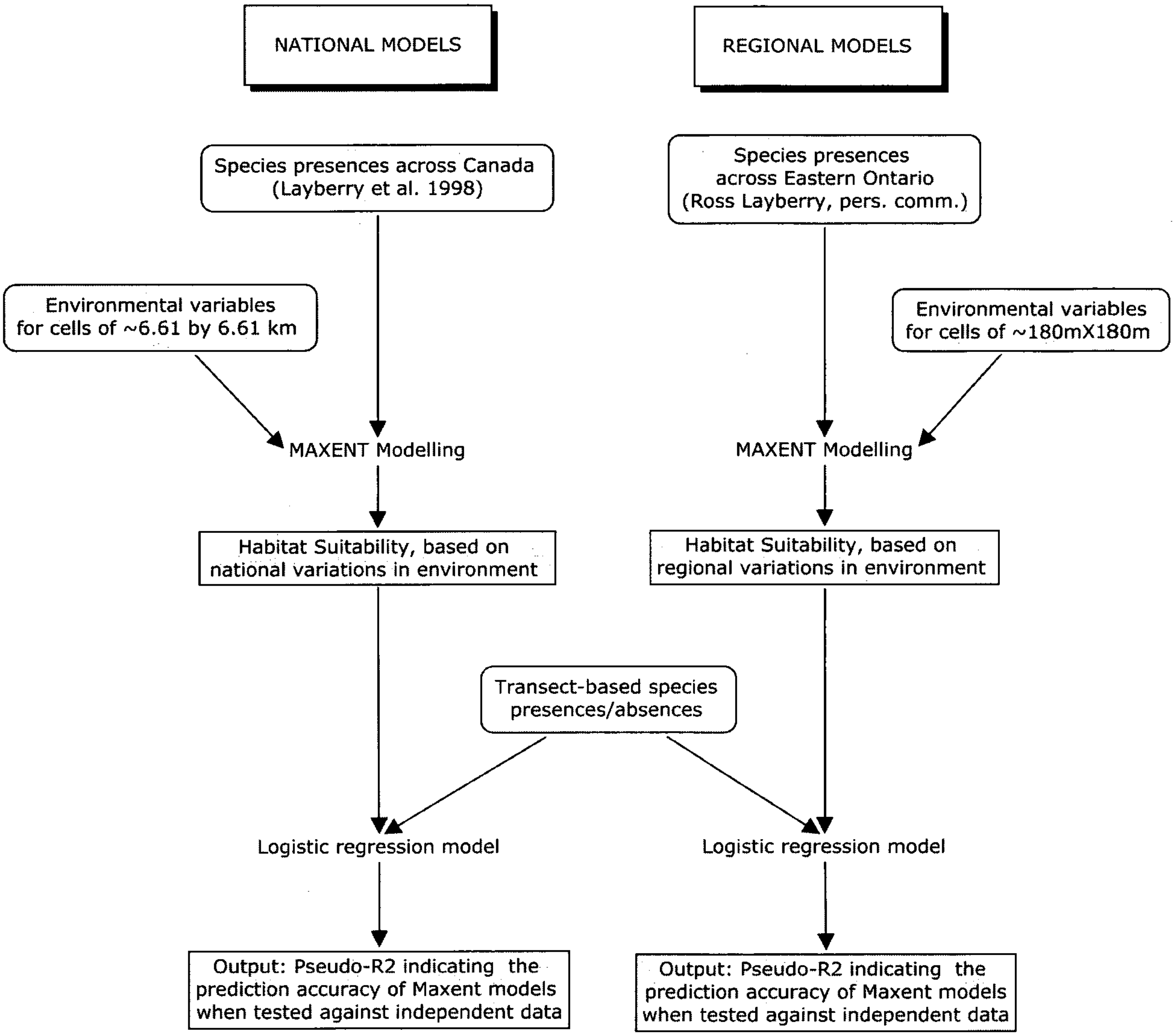

Figure 1: Experimental design and methodology for testing the predictive accuracy of Maxent models built at two different scales against independent systematically surveyed presence/absence data in the region of Ottawa, Canada with the use of logistic regressions. This figure was built with IHMC CmapTools (http://cmap.ihmc.us/conceptmap.html).

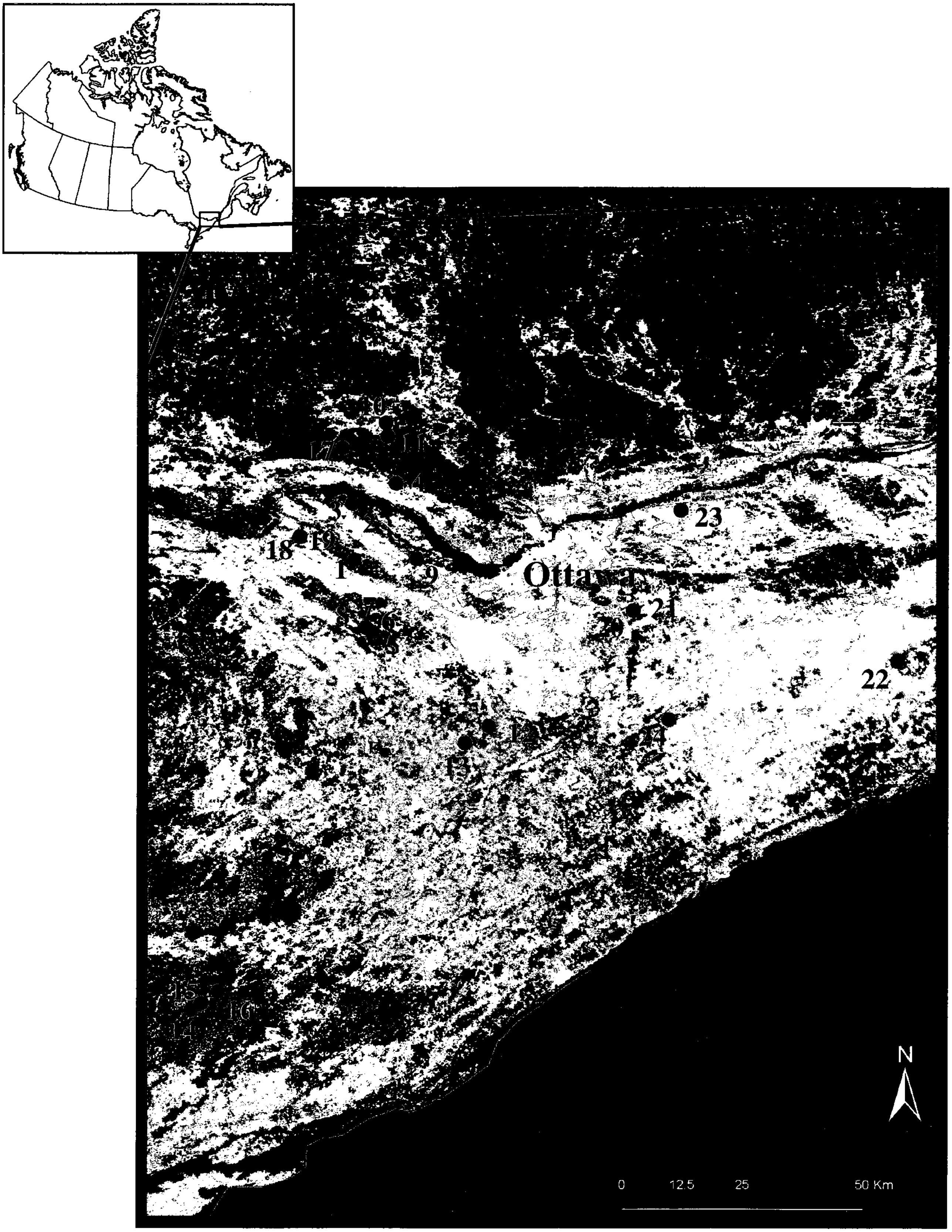

Figure 2: Map of field site locations for the butterfly transect surveys during the summers of 2007 and 2008. Refer to Table 1 for field site numbers, description and vegetation type. The background is the landcover used for the regional modelling

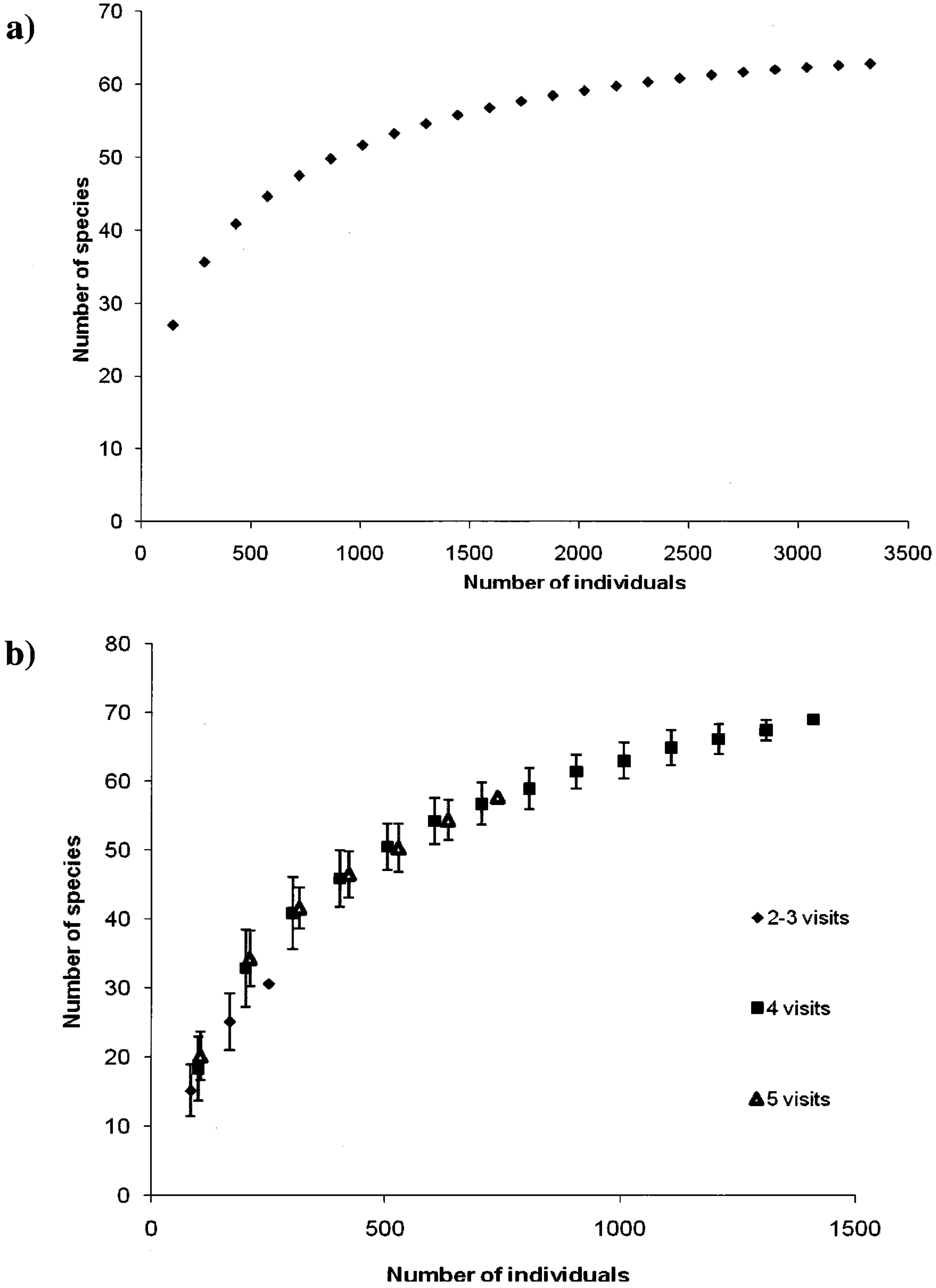

Figure 3: Coleman's rarefaction curves representing a) all the species found from all individuals at all sites, and b) the overlap of curves for sites visited 4 and 5 times.

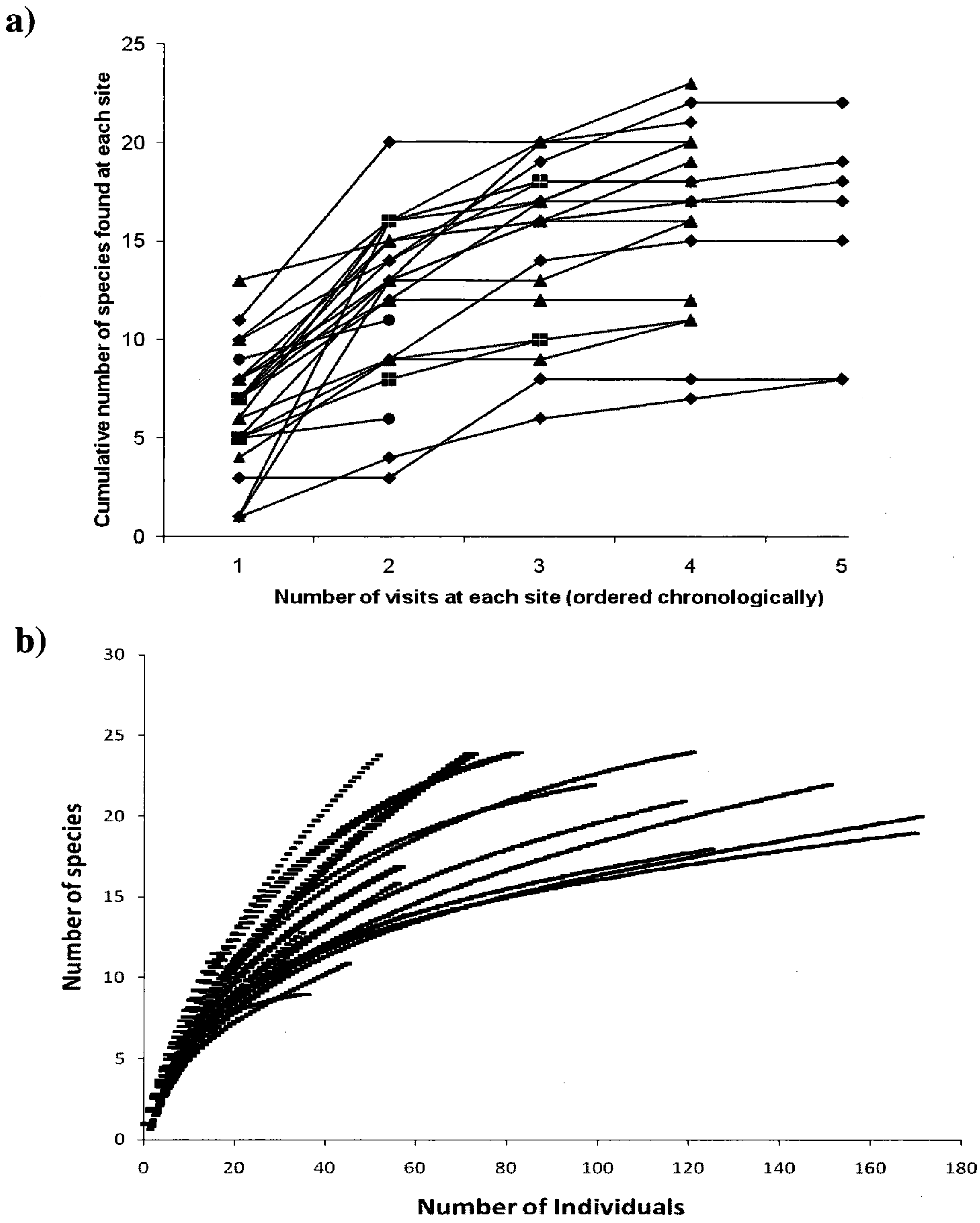

Figure 4: a) Cumulative number of species observed at each site as a function of the number of visits per site, arranged in chronological order. Series for sites visited twice are represented by black circles; series for sites visited three times are represented by a white cross in a black square; series for sites visited four times, by a triangle; series for sites visited 5 times, by a diamond. b) Coleman's rarefaction curve representing the species found as a function of each new individual at each site. Each line represents a survey site.

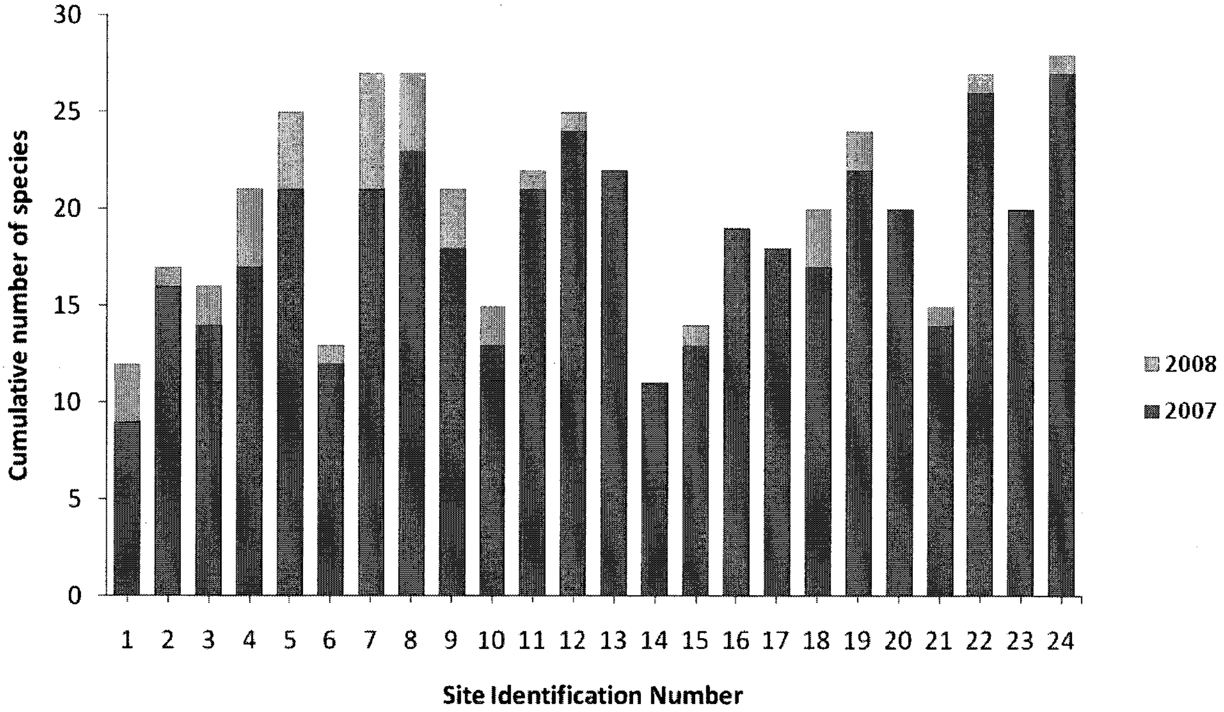

Figure 5: Total number of species per site observed in 2007 (black bars) and additional number of species per site observed in 2008, but not in 2007 (grey), for each site surveyed.

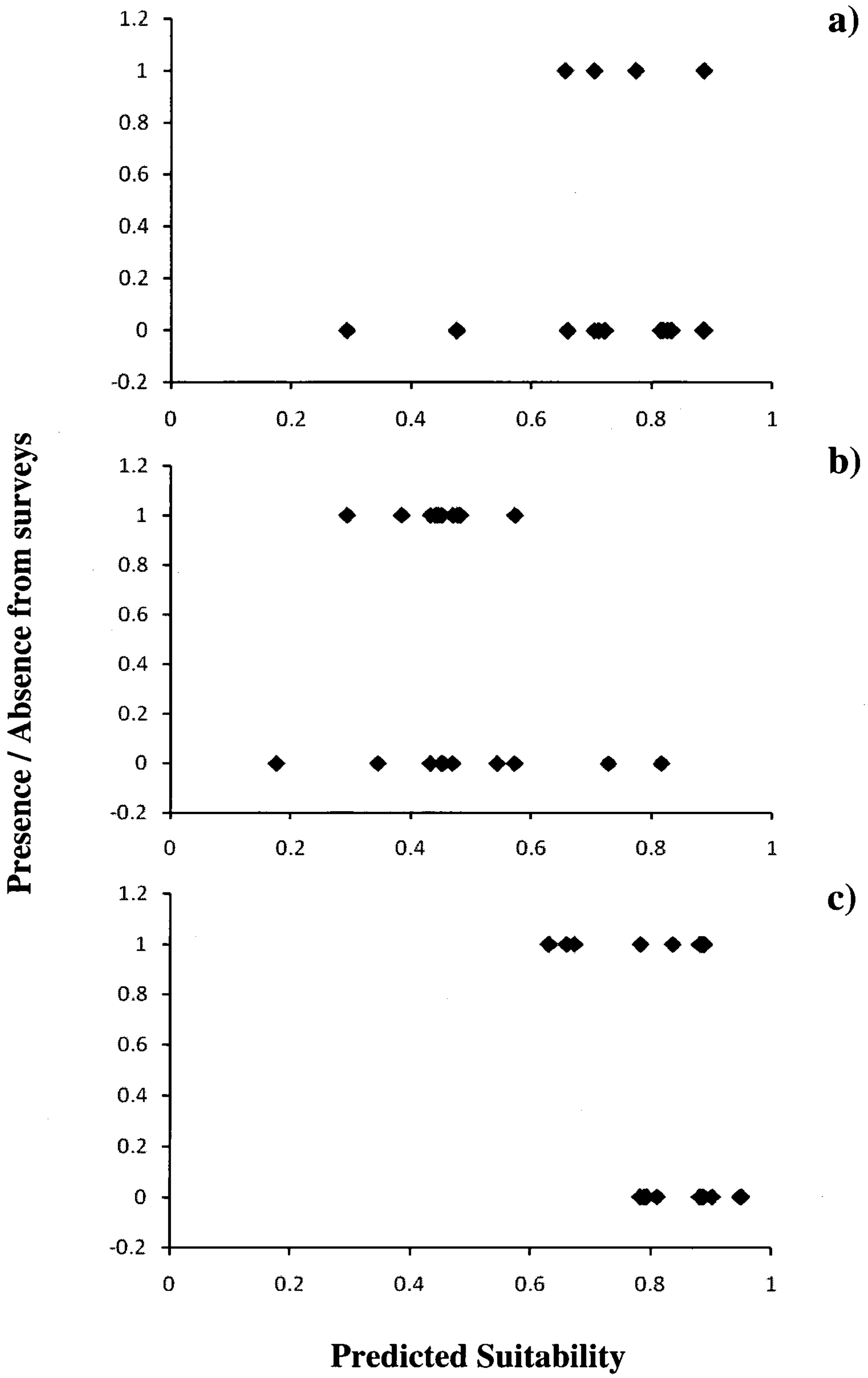

Figure 6: Species presence/absence on the surveys as a function of the suitability predicted by the national Maxent models for a) the American Lady, b) the Little Wood Satyr c) the White Admiral. Each point represents the presence or absence of the species at one survey site.

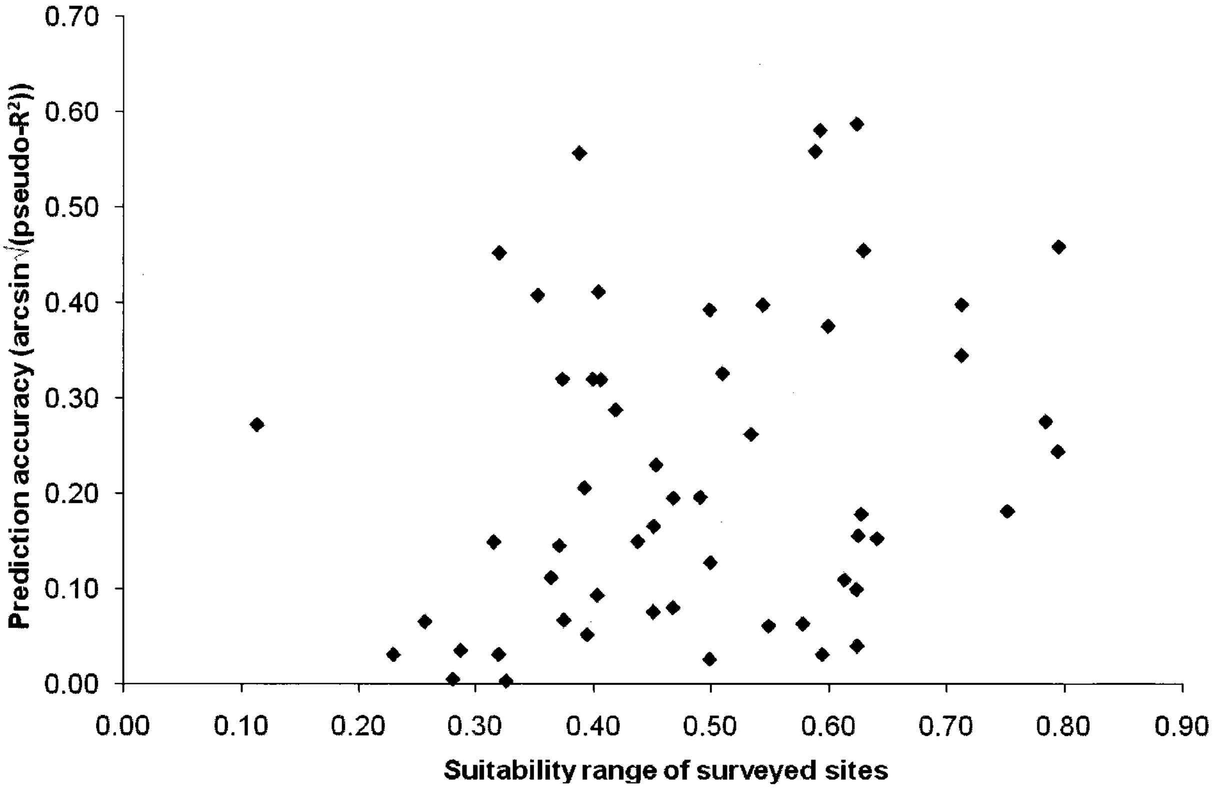

Figure 7: Prediction accuracy of the national models (measured as the arcsin√(pseudo-R^2) of the logistic regression) as a function of the suitability range of the national models at sites surveyed for each species (suitability range = site of highest predicted suitability – site of lowest predicted suitability). Each point represents a species with n=55 species. All species for which a pseudo-R^2 could be calculated were included in the plot.

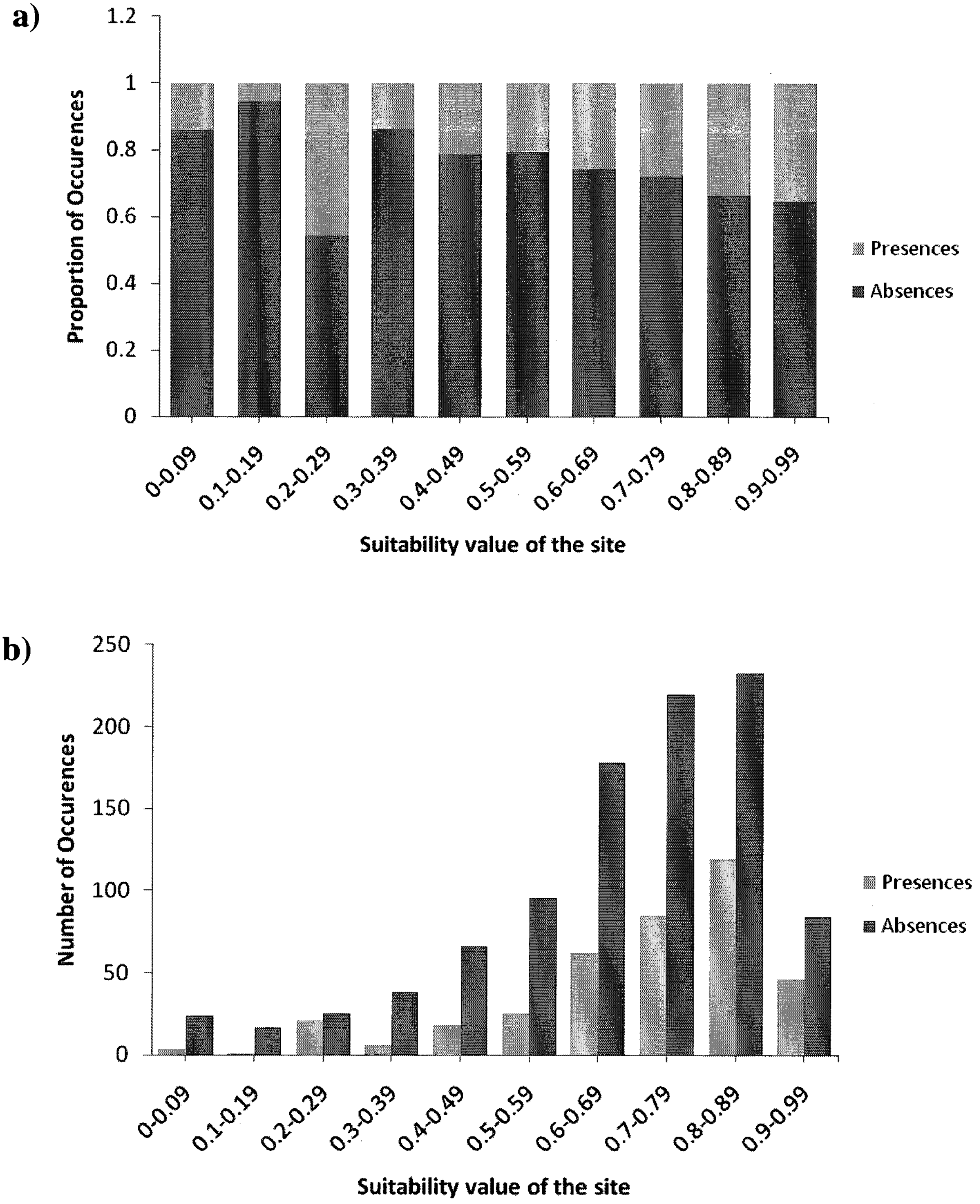

Figure 8: a) Proportion of occurences and b) number of presences and absences as a function of the suitability predicted by the national models at each site in the Ottawa region. Total observations n = 1368; presences n = 387; absences n = 981.

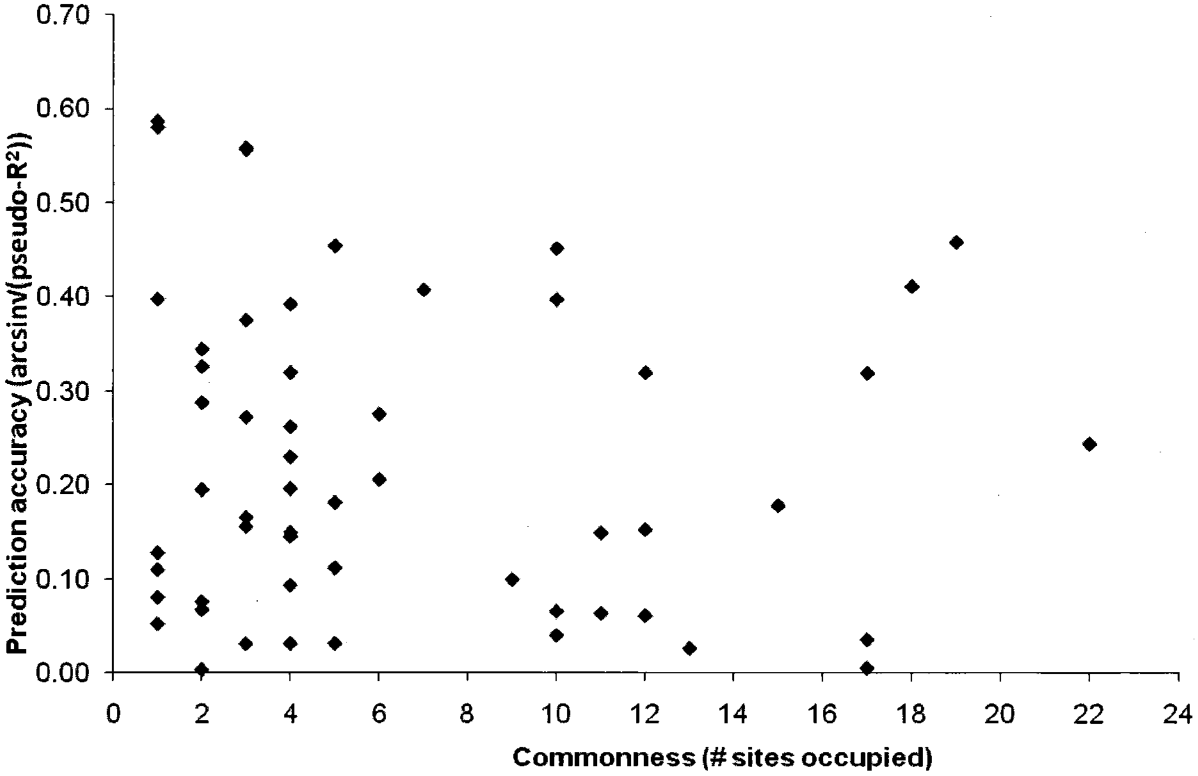

Figure 9: Prediction accuracy of the national models (measured as the arcsin√(pseudo-R^2) of the logistic regression) as a function of commonness (measured as the number of field sites where a species was found during the field surveys). Each point represents a species.

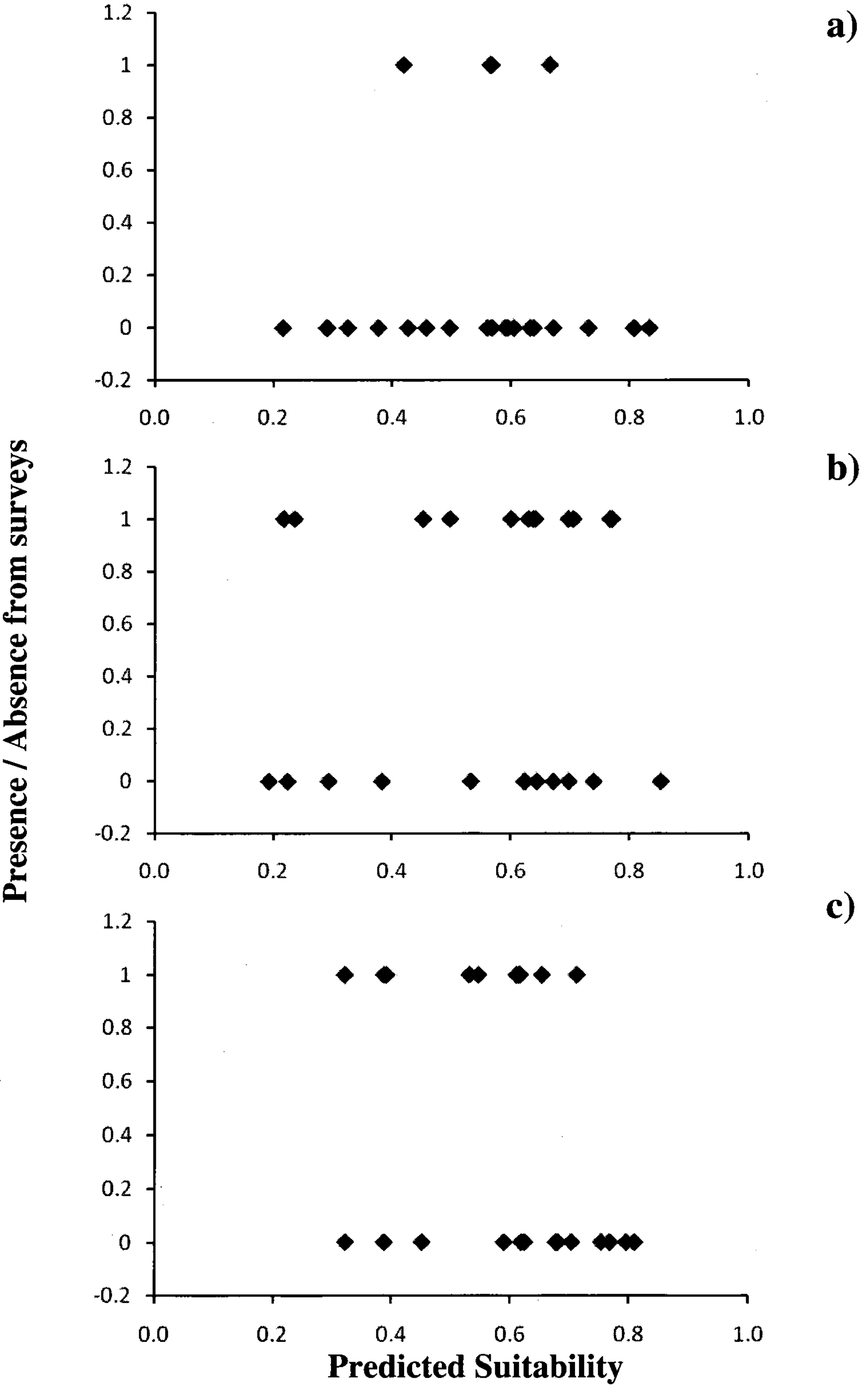

Figure 10: Species presence/absence on the surveys as a function of the suitability predicted by the regional models for a) the American Lady, b) the Little Wood Satyr c) the White Admiral. Each point represents the presence or absence of the species at one survey site.

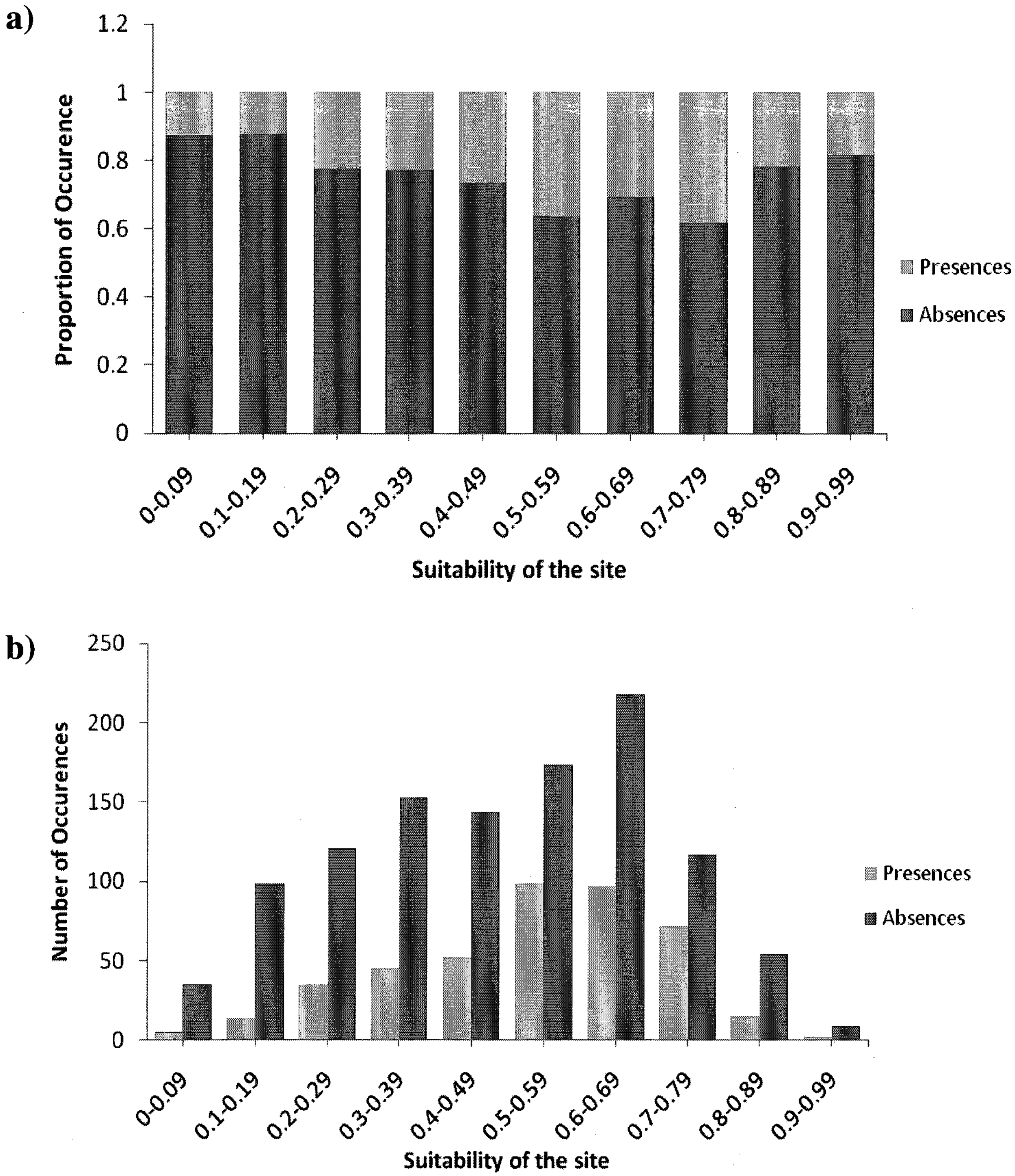

Figure 11: a) Proportion of occurences and b) number of presences and absences as a function of the suitability predicted by the regional models at each site in the Ottawa region. Total observations n = 1560; presences n = 436; absences n = 1124.

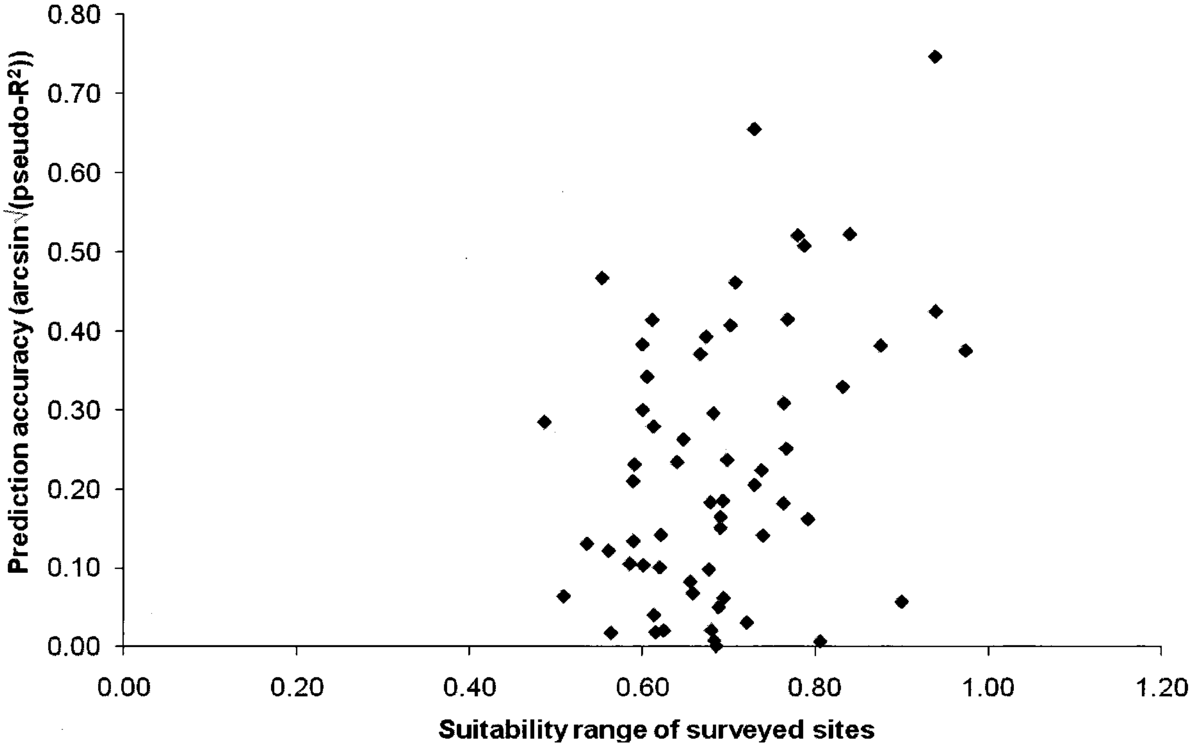

Figure 12: Prediction accuracy of the regional models (measured as the arcsin√(pseudo-R^2) of the logistic regression) as a function of the suitability range of the site where each species was found (suitability range = site of highest predicted suitability – site of lowest predicted suitability). Each point corresponds to a species.

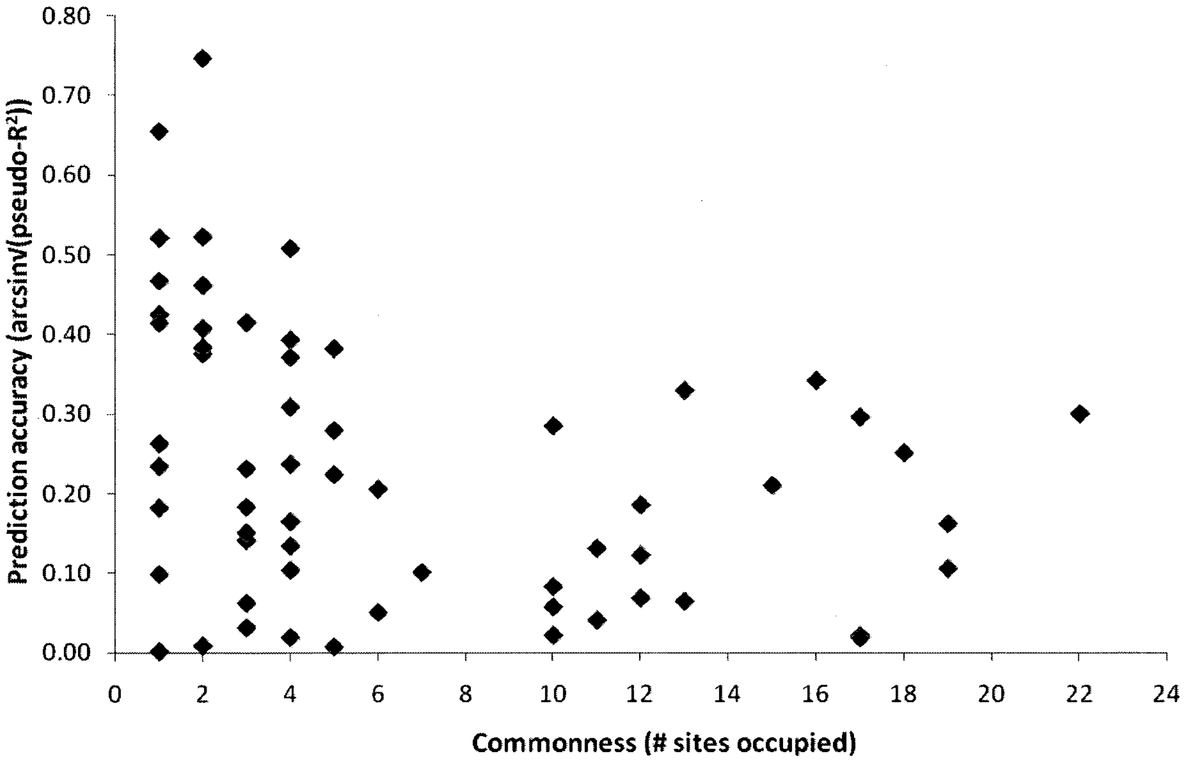

Figure 13: Prediction accuracy of the regional models (measured as the arcsin√(pseudo-R^2) of the logistic regression) as a function of commonness (measured as the number of field sites where a species was found during the field surveys). Each point corresponds to a species.

References

Anderson, R. P., D. Lew, and A. T. Peterson. 2003. Evaluating predictive models of species' distributions: Criteria for selecting optimal models. Ecological Modelling 162: 211-232.

Araújo M. B. and M. New. 2006. Ensemble forecasting of species distributions TREE 22: 42-47.

Araújo M.B & C. Rahbek. 2006. How Does Climate Change Affect Biodiversity? Science 313: 1396 – 1397.

Araújo M.B., Whittaker, R.J., Ladle, R., Erhard, M. 2005. Reducing uncertainty in projections of extinction risk from climate change. Global Ecology & Biogeography 14: 529–538

Balmford A, and K.J.Gaston. 1998. Why diversity surveys are good value. Nature 398: 204-205.

Beaubien, J., J. Cihlar, and R. Latifovic. 2000. Land cover of Canada. Natural Resources Canada, Ottawa, Canada. (Beaubien, J., Cihlar, J., & R. Latifovic. (2002). Land cover classification of LANDSAT ETM+ scene 02 August 2000 P52/R10. Satellite information for Land cover of Canada (SILC) Project, Natural Resources Canada, Canada Centre for Remote Sensing, Ottawa, Ontario.)

Brock J.P., and K.Kaufman. 2003. Butterflies of North America (Kaufman Focus Guides). Houghton Mifflin Harcourt, New York. 384 p.

Brook B.W., Sodhi N.S., Bradshaw C.J.A. 2008. Synergies among extinction drivers under global change. Trends in Ecology and Evolution 23: 453–460.

Colwell, R. K. 2005. EstimateS: Statistical estimation of species richness and shared species from samples. Version 8.0. User's Guide and application published at: http://purl.oclc.org/estimates.

Colwell and Coddington, 1994. Estimating terrestrial biodiversity through extrapolation. Philosophical Transactions of the Royal Society (Series B) 354: 101-118.

Davros NM, Debinski DM, Reeder KF, Hohman WL. 2006. Butterflies and the Continuous Conservation Reserve program filter strips: landscape considerations. Wildlife Society Bulletin. 34(4) 936-943.

Edwards Jr, T.C., Deshler, E.T., Foster, D., and G.G. Moisen. 1996. Adequacy of Wildlife Habitat Relation Models for Estimating Spatial Distributions of Terrestrial Vertebrates. Conservation Biology 10: 263-270.

Elith J., Graham, C. H., Anderson, R. P., Dudik, M., Ferrier, S., Guisan, A., Hijmans, R. J., Huettmann, F., Leathwick, J. R., Lehmann, A., Li, J., Lohmann, L. G., Loiselle, B. A., Manion, G., Moritz, C., Nakamura, M., Nakazawa, Y. , Overton, J. M., Peterson, A. T., Phillips, S. J., Richardson, K., Scachetti-Pereira, R., Schapire, R. E., Soberón, J. , Williams, S., Wisz, M. S. , and N. E. Zimmermann. 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 Ltd., Redlands, California.

Field, R., Hawkins, B.A., Cornell, H.V., Currie, D.J., Diniz-Filho, J.A.F., Guégan, J.-F., Kaufman, D.M., Kerr, J.T., Mittelbach, G.G., Oberdor, T., O'Brien, E. & Turner, J.R.G. 2009. Spatial species-richness gradients across scales: a meta-analysis. Journal of Biogeography 36: 132-147.

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.

Gotelli, N., & R. K. Colwell. 2001. Quantifying biodiversity: Procedures and pitfalls in the measurement and comparison of species richness. *Ecology Letters* 4 , 379-391.

Greaves, G.J., Mathieu, R., and P.J. Seddon. 2006. Predictive modelling and ground validation of the spatial distribution of the New Zealand long-tailed bat (*Chalinolobus tuberculatus*). *Biological Conservation* 132, 211-221.

Grill A, Cleary DFR. 2003. Diversity patterns in butterfly communities of the Greek nature reserve Dadia. *Biological Conservation* 114 (3): 427-436.

Guisan, A., and N. E. Zimmermann, 2000. Predictive habitat distribution models in ecology Ecological Modelling 135: 147-186.

Guisan A., and W. Thuiller. 2005. Predicting species distribution: offering more than simple habitat models. Ecology Letters 8(9): 993-1009.

Hagle, T.M., and G.E. Mitchell. 1992. Goodness-of-fit measures for probit and logit. American journal of political science 36: 762-784.

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 modelling methods. Ecography **29**:773-785.

Hill J.K., Collingham, Y., Thomas, C., Blakeley, D., Fox, R., Moss, D., and B. Huntlry. 2001. Impacts of landscape structure on butterfly range expansion. Ecology Letters 4: 313-321.

Hughes, L. 2000. Biological consequences of global warming: Is the signal already apparent? Trends in Ecology and Evolution 15:56-61.

Inglis, G.J., Hurren, H., Oldman, J., and R. Haskew. 2006. Habitat suitability index and particle dispersion models for early detection of marine invaders. Ecological Applications 16: 1377-1390.

Insightful Corporation. S-PLUS 8.0 for Windows.

Jarvis A., H.I. Reuter, A. Nelson, E. Guevara, 2008, Hole-filled seamless SRTM data V4, International Centre for Tropical Agriculture (CIAT), available from http://srtm.csi.cgiar.org.

Kerr JT, Sugar A, Packer L. 2000. Indicator taxa, rapid biodiversity assessment, and nestedness in an endangered ecosystem. *Conservation Biology* 14 (6): 1726-1734.

Kharouba, H.M. 2007. Predicting and mitigating the impacts of global change on species' distributions. 79 p. University of Ottawa, Ottawa, Canada.

Kharouba H.M., Algar A., and J.T. Kerr 2009. Historically calibrated predictions of butterfly species' range shift using global change as a pseudo-experiment. Ecology. Ecology 90: 2213-2222.

LandScan™ Global Population Database. Oak Ridge, TN: Oak Ridge National Laboratory. Available at http://www.ornl.gov/landscan/.

Lawler J.J., White D., Neilson R. P., 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. 280p.

Lobo, J.M., Jiménez-Valverde, A., and R. Real. 2008. AUC: a misleading measure of the performance of predictive distribution models. Global Ecology and Biogeography 17, 145-151.

Loiselle, B.A., Howell, C.A., Graham, C.H., Goerck, J.M., Brooks, T., Smith, K.G., and P.H. Williams. 2003. Avoiding pitfalls of using species distribution models in conservation planning. *Conservation Biology* 17, 1591–1600.

Mackey BG, Lindenmayer DB. 2001. Towards a hierarchical framework for modeling the spatial distribution of animals. Journal of Biogeography 28:1147–1166.

Marmion M., Parviainen M., Luoto M., Heikkinen R.K. & W. Thuiller. 2009. Evaluation of consensus methods in predictive species distribution modeling. Diversity and Distribution 15: 59-69

Margules C.R., and R.L. Pressey. 2000. Systematic conservation planning. Nature 405: 243-253.

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.

Mekis E., and W.D. Hogg. 1999. Rehabilitation and analysis of Canadian daily precipitation time series, *Atmosphere-Ocean* **37** (1) (1999), pp. 53–85

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

Opler P.A.1994. A Field Guide to Eastern Butterflies. Houghton Mifflin Harcourt. Boston. 486p.

Pearson, R. G., and Dawson, T. P. 2003. Predicting the impacts of climate change on the distribution of species: are bioclimatic models useful? Global Ecology and Biogeography 12: 361-371.

Pearson RG, Thuiller W, Araujo MB, Martinez-Meyer E, Brotons L, et al. 2006. Model-based uncertainty in species range prediction. Journal of Biogeography 33: 1704–1711.

Peterson A.T., Egbert, S.L., Sánchez-Cordero V., Price, K. 2000. Geographic analysis of conservation priority : endemic birds and mammals in Veracruz, Mexico. Biological Conservation 93: 85-94.

Peterson AT, Papeş M, Soberón J. 2008. Rethinking receiver operating characteristic analysis applications in ecological niche modelling. Ecological Modelling 213: 63–72.

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.

Phillips S. J., R. P. Anderson, and R. E. Schapire. 2006. Maximum entropy modeling of species geographic distributions. Ecological Modelling **190**:231-259.

Phillips SJ, Dudik M, Elith J, Graham CH, Lehmann A, et al. 2009. Sample selection bias and presence-only distribution models: implications for background and pseudo-absence data. Ecol Appl 19: 181–197.

Pimm SL, Russel GJ, Gittleman JL, Brooks TM (1995) The future of biodiversity. Science 269: 347–350.

Pollard E and Yates TJ. 1993. Monitoring Butterflies for Ecology and Conservation. Chapman and Hall, London. 274p.

Prugh L.R., Hodges K.E., Sinclair, A.R.E., & J.S. Brashares. 2008. Effect of habitat area and isolation on fragmented animal populations. PNAS 105: 20770-20775.

Schadt , S., Knauer, F., Kaczensky, P., Revilla, E., Wiegand, T., and L Trepl. 2002 Rule-based assessment of suitable habitat and patch connectivity for the Eurasian lynx Ecological Applications 12: 1469-1483.

Scott, J.M., Davis F., Csuti B., Noss R, Butterfield B., Groves C., Anderson H., Caicco S., Derchia F, Edwards T.C., Ulliman J., and R.G. Wright. 1993. Gap Analysis: A geographic Approach to Protection of Biological Diversity. Wildlife Monographs 123: 1-41.

Soberon J., and AT. Peterson. 2005. Interpretation of models of fundamental ecological niches and species' distributional areas. Biodiversity informatics. 2: 1-10.

Swets, J. A. 1988. Measuring the accuracy of diagnostic systems. Science 240:1285-1293.

Thuiller W. 2004. Patterns and uncertainties of species' range shifts under climate change. Global Change Biology 10: 2020–2027.

Uehara-Prado, M., J. d. O. Fernandes, et al. (2009). Selecting terrestrial arthropods as indicators of small-scale disturbance: A first approach in the Brazilian Atlantic Forest. Biological Conservation 142: 1220-1228.

Veloz, S.D. 2009. Spatially autocorrelated sampling falsely inflates measures od accuracy for presence only niche models. Journal of Biogeography 36: 2290-2299.

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

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.

Wisz, M.S., Hijmans, J.L., Peterson, A.T., Graham, C.H., Gisan, T., and NCEAS Predicting Species Distributions Working Group. 2008. Effect of sample size on the performance of species distribution models. Diversity and Ditribution 14: 763-773.

Appendix 1: Detailed methods

Data processing for regional scale models

The regional models used high resolution datasets that varied within the region and that were comparable in their potential biological effects on butterflies with the coarse-grained variables used in national models. These datasets included finer scale land cover and incoming solar radiation built with high resolution elevation, in place of the climate variables that do not vary greatly at the scale of the field study. The geographical extent of the models covers the area where the transect surveys were carried out, and includes a small buffer around the field sampling region.

Incoming solar radiation (for the summer solstice) was calculated for the study area from a 90m resolution digital elevation map (Jarvis et al. 2008) using the Solar Radiation Extension of the Spatial Analyst in ArcGIS (Environmental Systems Research Institute 2006). Land cover was derived from classified Landsat 5 and Landsat 7 ortho-images (Land Cover, circa 2000 – vector; GeoBase.ca). These data were available as vector files and I transformed them into raster grids with cell sizes of approximately 200m (of 0.00167 degrees) and mosaicked the tiles together using ArcGIS (Environmental Systems Research Institute 2006) to form a landcover with 32 classes. The 200m cell size was chosen to reflect the size of the transect. It would have been possible to increase the resolution of the analysis, but anything below 200m would have been biologically difficult to interpret due to the motility of the butterflies. The environmental variables used in the models cover the time period of the butterfly sightings to ensure butterfly sightings corresponded as faithfully as possible to the years land cover, elevation, incoming solar radiation and population density were derived from.

Suitability calculation for transects covering several cells

For each field site, we calculated the proportional distance of the transect that was crossing each cell and multiplied this proportion by the suitability of the cell. These partial suitabilities were then added to yield the total suitability at each site.

Appendix 2: Species and sites data

Number of identified specimens and species per site for the 2007 and 2008 field surveys. Also includes the number of species found in 2008 that had not been seen at the site in 2007, along with the total number of species per site and the percent increase in number of species per site in 2008

Site Number	# identified specimens in 2007*	# identified specimens in 2008	# identified species in 2007	# identified species in 2008	# 2008 species not observed in 2007	Total # of species	Percent increase in species (%)
1	36	6	9	4	3	12	25.00
2	67	3	19	3	1	20	5.00
3	69	2	15	2	2	17	11.76
4	284	31	19	10	4	23	17.39
5	127	16	26	11	3	29	10.34
6	17	4	12	4	1	13	7.69
7	119	12	21	10	6	27	22.22
8	52	14	23	9	4	27	14.81
9	56	42	18	7	3	21	14.29
10	98	21	13	4	2	15	13.33
11	311	23	21	10	1	22	4.55
12	88	8	24	4	1	25	4.00
13	177	5	21	2	0	21	0.00
14	45	4	11	1	0	11	0.00
15	35	2	13	2	1	14	7.14
16	171	12	19	3	0	19	0.00
17	56	4	17	2	0	17	0.00
18	58	15	17	8	3	20	15.00
19	109	14	23	7	2	25	8.00
20	67	3	20	2	0	20	0.00
21	44	3	14	2	1	15	6.67
22	74	6	25	5	1	26	3.85
23	127	1	20	1	0	20	0.00
24	156	4	25	2	1	26	3.85

*if more than 10 European skippers were observed during a visit, only the first 10 are included in the count

Appendix 3: Logistic regression pseudo-R^2 value per species

Pseudo-R^2 value of logistic regressions of presence/absence of species for the independent field surveys as a function of the suitability predicted from the national and regional models.

Species	National Model Pseudo-R^2	Regional Model Pseudo-R^2
Acadian Hairstreak	0.0242	0.0039
American Lady	0.0010	0.0004
Aphrodite Fritillary	0.0010	0.0494
Arctic Skipper	0.0521	0.0108
Atlantis Fritillary	0.0726	0.0333
Baltimore Checkerspot	0.0992	0.1469
Black Swallowtail	0.1930	0.0001
Brown Elfin	0.0163	N/A
Cabbage White	0.0000	0.0005
Canadian Tiger Swallowtail	0.0038	0.0150
Chryxus Arctic	0.3068	N/A
Clouded Sulfur	N/A	N/A
Columbine Duskywing	0.0327	0.1391
Compton Tortoiseshell	0.0027	0.0540
Common Ringlet	0.0013	0.0852
Common Roadside Skipper	0.0377	0.1348
Common Wood Nymph	0.0989	0.0003
Delaware Skipper	N/A	0.1628
Dreamy Duskywing	0.0088	0.2367
Dun Skipper	0.1961	0.0112
Eastern Comma	0.1029	0.1401
Eastern Pine Elfin	0.0065	0.0097
Eastern Tailed Blue	N/A	0.1047
European Skipper	N/A	0.1126
Eyed Brown	0.0126	0.0761
Green Comma	0.0808	0.1986
Grey Comma	0.0046	0.1573
Great Spangled Fritillary	0.0222	0.0171
Harison Checkerspot	0.0273	N/A
Hickory Hairstreak	N/A	0.0201
Hoary Elfin	0.1502	0.1702
Hobomok Skipper	0.0041	0.0017
Indian Skipper	N/A	0.3712
Juvenal's Duskywing	0.0223	0.1318
Least Skipper	0.0381	0.0925
Leonard Skipper	0.0121	0.2031
Little Wood Satyr	0.0232	0.0047
Long Dash Skipper	0.0991	0.0340
Meadow Fritillary	0.0419	0.0026
Milbert's Tortoiseshell	0.2809	0.0526
Monarch	0.0585	0.0876
Mourning Cloak	0.0007	0.0042
Mustard White	0.0010	0.0199
Northern Broken Dash	N/A	0.0261
Northern Crescent	0.1601	0.0619
Northern Cloudywing	0.0673	0.0270
Northern Pearly Eye	0.0099	N/A
Olympia Marble	N/A	0.0676
Orange sulfur	0.1465	0.0552
Pearl Crescent	0.0016	0.0033
Peck's Skipper	0.0210	0.0180
Question Mark	N/A	0.0000
Red Admiral	0.0316	0.0436
Silver-bordered Fritillary	0.2792	0.0010
Silvery Blue	0.0044	0.0069
Silvery Checkerspot	0.1143	0.2491
Silver spotted Skipper	0.3010	0.1621
Spring Azure	0.1574	0.0102
Striped hairstreak	0.0058	0.0001
Summer Azure	0.1501	0.0005
Tawny-edged Skipper	0.1347	0.0227
Tawny Crescent	0.0000	0.4614
Viceroy	0.0742	0.0417
White Admiral	0.1909	0.0790
Mean	0.0725	0.0752
Median	0.0327	0.0417

Appendix 4: Regional model AUC accuracy and contribution of environmental variables

Area under the curve (averaged over the 10 iteration of the models) for each butterfly species for the regional modelling, including the standard deviation. The table also reports the percent contribution of each variable, as well as the overall average of AUC, and percent contribution of each environmental variable included in the model.

		Percent Contribution (%)			
Species	AUC ± SD	Elevation	Solar Radiation	Land Cover	Population Density
Acadian Hairstreak	0.9056 ± 0.0172	68.26	3.40	25.63	2.71
American Lady	0.8507 ± 0.0258	68.42	3.42	26.10	2.06
Aphrodite Fritillary	0.8735 ± 0.0238	55.65	3.72	31.37	9.26
Appalachian Brown	0.9017 ± 0.0243	64.37	1.93	30.86	2.84
Arctic Skipper	0.8954 ± 0.0187	66.05	3.51	29.16	1.28
Atlantis Fritillary	0.8605 ± 0.0296	48.59	2.22	47.92	1.27
Baltimore Checkerspot	0.9271 ± 0.0148	71.20	1.66	22.98	4.16
Black Swallowtail	0.9203 ± 0.0124	83.86	1.33	12.63	2.19
Bronze Copper **	0.9016 ± 0.0245	62.24	0.79	28.26	8.70
Brown Elfin	0.8418 ± 0.0710	43.80	0.74	55.02	0.45
Cabbage White	0.9005 ± 0.0082	79.98	2.50	17.08	0.44
Canadian Tiger Swallowtail	0.8509 ± 0.0161	65.24	5.82	28.23	0.71
Chryxus Arctic	0.7408 ± 0.1080	7.31	4.06	87.97	0.66
Clouded Sulfur	0.9020 ± 0.0092	76.07	1.36	15.60	6.96
Columbine Duskywing	0.8734 ± 0.0290	56.87	2.24	38.05	2.84
Compton Tortoiseshell	0.8846 ± 0.0208	65.89	5.79	26.96	1.37
Common Ringlet	0.8973 ± 0.0100	76.20	3.04	20.44	0.33
Common Roadside Skipper	0.9177 ± 0.0380	23.91	0.44	51.17	24.48
Common Wood Nymph	0.8942 ± 0.0101	78.69	1.24	19.64	0.43
Coral Hairstreak **	0.8453 ± 0.0427	53.07	1.98	43.02	1.93
Crossline Skipper	0.8616 ± 0.0385	50.33	0.60	30.81	18.26
Delaware Skipper	0.9256 ± 0.0255	45.88	0.77	28.72	24.63
Dion Skipper **	0.9216 ± 0.0345	35.79	0.07	40.80	23.34
Dreamy Duskywing	0.8725 ± 0.0247	59.11	4.40	34.74	1.74
Dun Skipper	0.8891 ± 0.0121	70.37	5.36	23.68	0.59
Eastern Comma	0.8790 ± 0.0213	71.02	2.74	25.34	0.91
Eastern Pine Elfin	0.8880 ± 0.0305	54.33	0.74	33.77	11.16
Eastern Tiger Swallowtail **	0.9280 ± 0.0232	68.90	1.63	25.45	4.02
Eastern Tailed Blue	0.9302 ± 0.0149	85.13	0.63	11.37	2.87
European Skipper	0.8606 ± 0.0116	74.20	3.41	21.92	0.47
Eyed Brown	0.8884 ± 0.0134	74.22	1.89	23.21	0.68

				Percent Contribution (%)			
Species	AUC ± SD			Elevation	Solar Radiation	Land Cover	Population Density
Gorgone Checkerspot **	0.9688	±	0.0147	28.14	0.00	66.53	5.33
Green Comma	0.7888	±	0.0655	31.61	1.24	59.99	7.16
Grey Comma	0.8583	±	0.0272	65.91	2.55	30.13	1.42
Great Spangled Fritillary	0.8818	±	0.0127	73.94	2.66	23.10	0.30
Harris's Checkerspot	0.9070	±	0.0246	62.88	3.69	30.60	2.83
Hickory Hairstreak	0.8478	±	0.0487	74.00	0.43	25.56	0.00
Hoary Elfin	0.8925	±	0.0657	12.74	0.58	54.83	31.85
Hobomok Skipper	0.8747	±	0.0144	66.69	4.89	27.14	1.29
Indian Skipper	0.8883	±	0.0303	40.52	1.13	36.03	22.33
Juvenal's Duskywing	0.8568	±	0.0216	57.06	4.91	37.15	0.88
Least Skipper	0.9171	±	0.0180	77.77	1.33	20.47	0.43
Leonard Skipper	0.8500	±	0.0362	50.12	3.14	41.85	4.89
Little Wood Satyr	0.8862	±	0.0126	68.38	2.53	28.30	0.79
Long Dash Skipper	0.8838	±	0.0138	71.23	3.08	23.76	1.93
Meadow Fritillary	0.9048	±	0.0172	69.44	4.64	24.98	0.94
Milbert's Tortoiseshell	0.8570	±	0.0244	61.45	5.36	32.00	1.19
Monarch	0.8732	±	0.0099	79.18	2.40	18.12	0.30
Mourning Cloak	0.8879	±	0.0132	75.66	2.42	21.43	0.49
Mustard White	0.9027	±	0.0132	72.27	1.69	21.00	5.04
Northern Broken Dash	0.9282	±	0.0176	59.63	0.21	38.80	1.37
Northern Crescent	0.9104	±	0.0095	74.84	1.36	20.57	3.23
Northern Cloudywing	0.8711	±	0.0198	65.35	5.71	27.86	1.08
Olympia Marble	0.8545	±	0.0558	41.43	14.50	37.04	7.03
Orange sulfur	0.8682	±	0.0200	73.57	3.63	19.47	3.32
Painted Lady **	0.8833	±	0.0241	74.47	1.10	22.25	2.18
Pearl Crescent	0.9246	±	0.0151	62.24	0.37	31.14	6.25
Peck's Skipper	0.8470	±	0.0321	73.42	3.31	22.01	1.26
Question Mark	0.8956	±	0.0234	74.69	1.36	23.46	0.49
Red Admiral	0.8875	±	0.0151	76.84	2.75	19.45	0.96
Silver-bordered Fritillary	0.9045	±	0.0159	70.77	4.29	22.89	2.05
Silvery Blue	0.8826	±	0.0139	70.13	1.27	27.89	0.71
Silvery Checkerspot	0.7976	±	0.0770	14.88	1.69	64.17	19.26
Silver spotted Skipper	0.8956	±	0.0249	61.92	0.67	35.70	1.71
Spring Azure	0.8834	±	0.0168	66.17	5.33	27.71	0.79
Striped hairstreak	0.9093	±	0.0226	63.13	4.62	29.69	2.57
Summer Azure	0.8990	±	0.0122	77.18	0.94	21.21	0.66
Tawny-edged Skipper	0.8953	±	0.0152	67.88	1.86	29.22	1.03
Tawny Crescent	0.9481	±	0.0311	12.23	0.14	48.39	39.25
Viceroy	0.9190	±	0.0092	77.80	1.25	19.78	1.17
White Admiral	0.8626	±	0.0140	67.65	2.06	29.41	0.88

** Species for which models could be build but not tested with the independent field surveys