Ecological Niche Modelling: *Salix* in Ontario

Philippe Stankowski

A Graduate Thesis
Submitted in partial fulfillment of the requirements
For the degree of
Master of Science in Forestry

Faculty of Natural Resources Management
Lakehead University

July 2010

ABSTRACT

Stankowski, P.A. 2010. Ecological niche modelling: *Salix* in Ontario. Master of Science in Forestry, Lakehead University. Advisor, Dr. W.H. Parker.

Keywords: *Salix* (willow), climate change, continuous/discontinuous data, BIOCLIM, logistic regression, model validation, species/climate equilibrium, uncertainty.

It has long been recognized that organisms exist in environments peculiar to taxa. Climate is thought to be of primary importance in determining the natural geographic distribution of species. Growing concern over climate change and its potential consequences to biodiversity has prompted the rapid development of numerous analytical techniques to correlate *quantifiable* environmental characteristics with the *known* location of species. Ecological niche modelling focuses primarily on determining the dimensions of the niche space of species as a means of predicting their geographic distributions. Recently, many alarming predictions have emerged ranging from the mass extinction of taxa over the next century, to, at the very least, their partial redistribution. Regardless of consequence, however, these predictions indicate a deep and gnawing uncertainty.

To explore some of the fundamental sources of uncertainty with respect to niche modelling two studies were undertaken. In the first, simple correlative models were developed for *Salix* (willow) species occurring in Ontario, Canada, to determine the algorithmic sensitivity of logistic regression to extreme cases of distributional and environmental data sets. Provincial distribution models were developed for 30 willow species to examine (*i*) the predictive ability of logistic regression analysis, and (*ii*) the effects of using different distributional and environmental data sets. Two original measures of model accuracy and over-prediction were employed and evaluated using independent data. Models based on unique combinations of monthly climate data, rather than a fixed selection of annual and seasonal variables, predicted distributions most accurately for all species. Models based on a fixed set of variables, while generating the highest average probabilities of occurrence for certain species with limited ranges, resulted in the greatest under- and overestimates of willow distributions. Regardless of distributional and environmental data input, no algorithm maximized model performance for all species. Individual species models require individual approaches; *i.e.*, the variable selection technique, the set of environmental factors used as predictors, and the nature of species distributional data must be carefully matched to the intended application.

A central assumption underpinning research into the potential future habitat of terrestrial biota is that species are *presently* in equilibrium with their environments and that quantitative climate models *adequately* represent the distribution of species. The second study examines the effect of the assumption of species/climate equilibrium upon projected distributions using different historic and future data sets. Distributional models were developed for 24 *Salix* (willow) species occurring in the province of Ontario, Canada, using three historical climate data

sets. Although historical data very accurately represented the distributions of willows, the inherent variability within the models of species based on different periods greatly influenced the direction and magnitude of projected distributional change. Even large-scale models of the climatic niche dimensions of species are temporally variable. These findings imply that many of the recent predictions of the potential consequences of climate change to terrestrial biota may be unrealistic.

TABLE OF CONTENTS

TABLES

FIGURES

Chapter One

Ecological niche modelling: Concepts, trials, and directions

Concepts

It has long been recognized that organisms exist in environments peculiar to taxa (Darwin, 1859). This view forms the conceptual framework within which ecological niche theory has developed. Although formally articulated by Joseph Grinnell in 1917 as the set of conditions in which a species can exist, no formal representation of the niche concept has been achieved that is universally accepted (Pennington, 2006).

Growing concern over climate change and its potential consequences to biodiversity has prompted the rapid development of numerous analytical techniques to correlate *quantifiable* environmental characteristics with the *known* location of species. Recently, many alarming predictions have emerged ranging from the mass extinction of taxa over the next century, to, at the very least, their partial redistribution. Regardless of consequence, however, these predictions indicate a deep and gnawing uncertainty. Should species have to relocate to survive? And, if so, where might they go?

Climate has long been held to be of primary importance in regulating the natural geographic distribution of species. Evidence from the fossil record (Hunter Jr. *et al.*, 1988; Jackson, 2000; Davis and Shaw, 2001) and from biogeographical studies of more recent times (Parmesan and Yohe, 2003; Araújo *et al.*, 2005) demonstrates the profound influence of climate on the abundance and range of species. There exists strong evidence opposing the idea that rapid environmental change overwhelms evolutionary processes (Davis *et al.*, 2005) and paleoecological examinations of fossilized plants and pollen demonstrate that postglacial redistributions of species were highly individualistic (Jackson, 2000). It has been theorised that

species can respond to climate change in three ways: migration, adaptation, or extinction (Holt, 1990).

A branch of ecological research deeply rooted in the tenets of biogeography has recently gained immensely in public attention becoming influential even in matters of global policy (Tokumine, 2002; Biggs *et al.*, 2008; Joyner *et al.*, 2010). Ecological niche modelling focuses primarily on determining the dimensions of the niche space of species as a means of predicting their geographic distributions. Among the most popular approaches to simulate the impacts of climate change on the distribution of species are statistical bioclimate envelopes (Pearson and Dawson, 2003; Heikkinen *et al.*, 2006). These models generally focus on the Grinellian niche concept in describing a range of environmental variables over which species can occur. Niche modelling approaches can differ characteristically in three ways: (i) the type of species' distributional data used, (ii) the selection of environmental variables, and (iii) the statistical technique used to fit a model. Studies undertaken to evaluate the response of species to climate change under past, present, and future scenarios have employed distributional data in the form of either *known* presences only (Beaumont *et al.*, 2007), or presence and absence data (Thomas *et al.*, 2004; Peterson *et al.*, 2005), using either *a priori* selections of environmental variables (Pearson *et al.*, 2002; Calef *et al.*, 2005; Flantua *et al.*, 2007; McKenney *et al.*, 2007) or some statistical method to determine a *best* set (Iverson and Prasad, 1998; Bakkenes *et al.*, 2002; Hamann and Wang, 2006). Ultimately, the selection of variables is subjective depending largely upon the scope of the research interest, the availability of data, and the skill and knowledge of the modeller (Austin, 2002; Austin *et al.*, 2006).

Over the past decade, tremendous effort has been given to reducing the predictive errors of ecological niche models. Numerous methods to correlate environmental data with the

distribution of species have been developed and tested (Elith *et al.*, 2006; Heikkinen *et al.*, 2006; Pearson *et al.*, 2006). The effect of scale and grain size on niche models has been explored (Guisan *et al.*, 2007; Trivedi *et al.*, 2008). The sensitivity of several methods to the geographic completeness, sampling bias, and georeferencing errors in species' distributional data has been examined (Peterson and Cohoon, 1999; Graham *et al.*, 2008; Loiselle *et al.*, 2008; Feeley and Silman, 2010). The effect of species' range size on algorithmic performance has been assessed (Manel *et al.*, 2001; Stockwell and Peterson, 2002) and the need to account for the prevalence of species has been discussed (McPherson *et al.*, 2004).

Above and beyond the sources of uncertainty already cited, several issues common to all niche models remain unclear. How are the geographic locations of species known? To which environmental factors might they correlate? How few data are required to make accurate predictions? How can they be generalised? And, what method may be most suitable to determine a relationship? Beyond the general consensus that a model should be as simple as possible but never more so, there is little agreement over how best to represent the dimensions of a species' niche. Indeed, many studies have been successful at predicting the distribution of species despite using altogether different approaches (Iverson and Prasad, 1998; Bakkenes *et al.*, 2002; Pearson *et al.*, 2002; Burns *et al.*, 2003; Calef *et al.*, 2005; Thuiller *et al.*, 2005; Hanann and Wang, 2006; Rehfeldt *et al.*, 2006; McKenney *et al.*, 2007; Peterson *et al.*, 2008; Beever *et al.*, 2010; Joyner *et al.*, 2010; Stankowski and Parker, 2010).

Trials

To explore some of the fundamental sources of uncertainty with respect to niche modelling two studies were undertaken. In the first, simple correlative models were developed for *Salix* (willow) species occurring in Ontario, Canada, to determine the algorithmic sensitivity of logistic

regression to extreme cases of distributional and environmental data sets. The second study examines the effect of the assumption of species/climate equilibrium upon projected distributions using different historic and future data sets. Briefly summarized, the findings imply that many of the recent predictions of the potential consequences of climate change to terrestrial biota may be unrealistic.

In the first analysis, the results indicate that the distributions of individual species are most accurately described using individual approaches; *i.e.*, different combinations of predictors resulting from different variable selection techniques. One size does not fit all (Stankowski and Parker, 2010). A combination of available distributional data types is recommended to determine the potential niche dimensions of species; *i.e.*, models should intuitively be fitted at larger scales using continuous inferred distributions and tested at smaller scales using known collection points. However, sampling bias and prevalence in the distributional data sets of species must be accounted for. With respect to the selection of ecological predictors, much remains uncertain. At what spatial and temporal scales are data relevant? In the first analysis, elevation, a highly significant predictor for many species, was found to have made negligible contributions to the final regression equations. In the second analysis however, the exclusion of this variable effectively reduced the probability levels for the models of several species.

The results from the second analysis indicate the climatic dimensions of the ecological niche of willow species in Ontario vary qualitatively and quantitatively through time. Models developed in different historical periods were generally dissimilar in both the direction and magnitude of projected distributional change. This was true even for the models of species based on overlapping time periods sharing similar selections of predictor variables. A stitch in time is therefore not enough. Should the assumption of species/climate equilibrium form the basis for

projected estimates of distributional change? And if so, over which historical period should this climatic equilibrium apply? Any attempt to determine climatic niche dimensions should intuitively include information concerning the longevity and lifecycles of species *i.e.*, climate data should be sampled over equal intervals and along gradients relative to the lifespan of species.

Directions

Better predictions can be made when there is a foreknowledge of what is relevant and what is not. Given the increasing availability of information, the number of permutations in a given explanatory data set can define an extremely large search space. Although access to sophisticated processing arrays has permitted the analysis of larger and larger data sets, statistics are ultimately blind to the nature of the system under consideration. The strong correlation observed between historical climate and the distributions of willows suggests the potential use of these models as tools to develop an understanding of the niches of individual species. However, much more work is needed.

Understanding the results of ecological niche modelling exercises hinges greatly upon knowledge of the evolutionary responses of individual species (Davis *et al.*, 2005). An ignorance of the interaction between species and environment makes simple cause and effect assumptions inapplicable to the problem of predicting distributions (Betts and Shugart, 2005). Further research with willow distributional models should consider polyploidy, hybridization, and introgression, highlighting key ecological processes such as regeneration, phenology, and reproduction (Chuine and Beaubien, 2001), as well as the differences in response to the environment of male and female plants (Kevan, 1990; Jones *et al.*, 1999). Establishing a

relational basis in both *time* and *space* for ecological niche models may simultaneously improve

the accuracy of our prediction and develop our concept of niche theory.

References

Austin, M.P., 2002. Spatial prediction of species distribution: An interface between ecological theory and statistical modelling. Ecol. Model. 157, 101–118.

Austin, M.P., Belbin, L., Meyers, J.A., Doherty, M.D., Luoto, M., 2006. Evaluation of statistical models used for predicting plant species distributions: role of artificial data and theory. Ecol. Model. 199, 197–216.

Araújo, M.B., Pearson, R.G., Thullier, W., Erhard, M., 2005. Validation of species-climate impact models under climate change. Glob. Change Biol. 11, 1504–1513.

Bakkenes, M., Alkemade, J.R.M., Ihle, F., Leemans, R., Latour, J.B., 2002. Assessing effects of forecasted climate change on the diversity and distribution of European higher plants for 2050. Glob. Change Biol. 8, 390–407.

Beaumont, L.J., Pitman, A.J., Poulsen, M., Hughes, L., 2007. Where will species go? Incorporating new advances in climate modelling into projections of species distributions. Glob. Change Biol. 13, 1368–1385.

Beever, E.A., Ray, C., Mote, P.W., Wilkening, J.L., 2010 Testing alternative models of climate–mediated extirpations. Ecol. Appl. 20, 164–178.

Betts, R.A., Shugart, H.H., 2005. Dynamic ecosystem and earth system models *in* Climate Change and Biodiversity. Yale University Press, New Haven, pp. 232–251.

Biggs, R., Simons, H., Bakkenes, M., Scholes, R.J., Eickhout, B., Vuuren, van D., Alkemande, R., 2008. Scenarios of biodiversity loss in southern Africa in the 21st century. Global Environ. Chang. 18, 296-309.

Burns, C.E., Johnston, K.M., Schmitz, O.J., 2003. Global climate change and mammalian species diversity in U.S. national parks. Proc. Natl Acad. Sci. USA 100, 11474–11477.

Calef, M.P., McGuire, A.D., Epstein, H.E., Rupp, T.S., Shugart, H.H., 2005. Analysis of vegetation distribution in Interior Alaska and sensitivity to climate change using a logistic regression approach. J. Biogeogr. 32, 863–878.

Chuine, I., Beaubien, E.G., 2001. Phenology is a major determinant of tree species range. Ecology Letters, 4, 500–510.

Darwin, C.R., 1859. On the Origin of Species by Means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life (John Murray, London).

Davis, M.B., Shaw, R.G., 2001. Range shifts and adaptive responses to Quaternary climate change. Science, 292, 673–679.

Davis, M.B., Shaw, R.G., Etterson, J.R., 2005. Evolutionary responses to changing climate. Ecology 86, 1704–1714.

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.McC., Peterson, A.T., Phillips, S.J., Richardson, K., Scachetti–Pereira, R., Schapire, R.E., Soberón, J., Williams, S., Wisz, M.S., Zimmermann, N.E., 2006. Novel methods improve prediction of species' distribution from occurrence data. Ecography 29, 129–151.

Feeley, K.J., Silman, M.R., 2010. Modelling the responses of Andean and Amazonian plant species to climate change: the effects of georeferencing errors and the importance of data filtering. J. Biogeogr. 37, 733–740.

Flantua, S.G.A., van Boxel, J.H., Hooghiemstra, H., van Smaalen, J., 2007. Application of GIS and logistic regression to fossil pollen data in modelling present and past spatial distribution of the Columbian savanna. Clim Dyn. 29, 697–712.

Graham, C.H., Elith, J., Hijmans, R.J., Guisan, A., Peterson, A.T., Loiselle, B.A., and the NCEAS Predicting Species Distributions Working Group, 2008. The influence of spatial errors in species occurrence data using in distribution models. J. Appl. Ecol. 45, 239–247.

Grinnell, J., 1917. The niche–relationships of the California thrasher. Auk 34, 427–433.

Guisan, A., Graham, C.H., Elith, J., Huettmann, F., and the NCEAS Species Distribution Modelling Group, 2007. Sensitivity of predictive species distribution models to change in grain size. Divers. Distrib. 13, 332–340.

Hamann, A., Wang, T., 2006. Potential effects of climate change on ecosystem and tree species distribution in British Columbia. Ecology 87, 2773–2786.

Heikkinen, R.K., Luoto, M., Araújo, M.B., Virkkala, R., Thuiller, W., Sykes, M.T., 2006. Methods and uncertainties in bioclimatic envelope modelling under climate change. Prog. Phys. Geogr. 30, 751–777.

Holt, R.D., 1990. The microevolutionary consequences of climate change. Trends Ecol. Evol. 5, 311–315.

Hunter Jr., M.L., Jacobson Jr., G.L., Webb III, T., 1988. Paleoecology and the coarse-filter approach to maintaining biological diversity. Conserv. Biol. 2, 375–385.

Iverson, L.R., Prasad, A.M., 1998. Predicting abundance for 80 tree species following climate change in the Eastern United States. Ecol. Monogr. 68, 465–485.

Jackson, S.T., 2000. Out of the garden and into the cooler? A Quaternary perspective on deep-time paleoecology. Paleontological Society Papers 6, 287–308.

Jones, M.H., Macdonald, S.E., Henry, G.H.R., 1999. Sex- and habitat-specific responses of a high arctic willow, Salix arctica, to experimental climate change. Oikos, 87, 129–138.

Joyner, T.A., Lukhnova, L., Pazilov, Y., Temiralyeva, G., Hugh–Jones, M.E., Aikimbayev, A., Blackburn, J.K., 2010. Modeling the potential distribution of Bacillus anthracis under multiple climate change scenarios for Kazakhstan. PLoS ONE 5, e9596. doi:10.1371/journal.pone.0009596.

Kevan, P.G., 1990. Sexual differences in temperature of blossoms on a dioecious plant, Salix arctica: significance for life in the arctic. Arctic and Alpine Research, 22, 283–289.

Loiselle, B.A., Jørgensen, P.M., Consiglio, T., Jiménez, I., Blake, J.G., Lohmann, L.G., Montiel, O.M., 2008. Predicting species distributions from herbarium collections: does climate bias in collection sampling influence model outcomes? J. Biogeogr. 35, 105–116.

Manel, S., Williams, H.C., Ormerod, S.J., 2001. Evaluating presence-absence models in ecology: the need to account for prevalence. J. Appl. Ecol. 38, 921–931.

McKenney, D.W., Pedlar, J.H., Lawrence, K., Campbell, K., Hutchinson, M.F., 2007. Potential impacts of climate change on the distribution of North American trees. Bioscience 57, 939–948.

McPherson, J.M., Jetz, W., Rogers, D.J., 2004. The effects of species' range size on the accuracy of distribution models: ecological phenomenon or statistical artifact? J. Appl. Ecol. 41, 811–823.

Parmesan, C., Yohe, G., 2003. A globally coherent fingerprint of climate change across natural systems. Nature, 421, 37–42.

Pearson, R.G., Thuiller, W., Araújo, M.B., Martinez–Meyer, E., Brotons, L., McClean, C., Miles, L., Seguardo, P., Dawson, T.P., Lees, D.C., 2006. Model–based uncertainty in species range prediction. J. Biogeogr. 33, 1704–1711.

Pearson, R.G., Dawson, T.P., 2003. Predicting the impacts of climate change on the distribution of species: are bioclimate envelope models useful? Global Ecol. Biogeogr. 12, 361-371.

Pearson, R.G., Dawson, T.P., Berry, P.M., Harrison, P.A., 2002. SPECIES: a spatial evaluation of climate impact on the envelope of species. Ecol. Model. 154, 289–300.

Pennington, D., 2006. Representing the dimensions of an ecological niche. Proceedings 5th International Semantic Web, November 6, 2006, Athens, Georgia.

Peterson, A.T., Cohoon, K.P., 1999. Sensitivity of distributional prediction algorithms to geographic completeness. Ecol. Model. 117, 159–164.

Peterson, A.T., Tian, H., Martínez-Meyer, E., Soberón, J., Sánchez-Cordero, V., Huntley B., 2005. Modeling distributional shifts of individual species and biomes *in* Climate Change and Biodiversity. Yale University Press, New Haven, pp. 211–228.

Peterson, A.T., Ortega–Huerta, M.A., Bartley, J., Sánchez–Cordero, V., Soberón, J., Buddemeier, R.H., Stockwell, D.R.B., 2002. Future projections for Mexican faunas under global climate change scenarios. Nature 416, 626–629.

Peterson, A.T., Stewart, A., Mohamed, K.I., Araújo, M.B., 2008. Shifting global invasive potential of European plants with climate change. PLoS ONE 3, e2441. doi:10.1371/journal.pone.0002441.

Rehfeldt, G.E., Crookston, N.L., Warwell, M.V., Evans, J.S., 2006. Empirical analyses of plant-climate relationships for the western United States. Int. J. Plant Sci. 167, 1123–1150.

Stankowski, P.A., Parker, W.H., 2010. Species distribution modelling: does one size fit all? A phytogeographic analysis of Salix in Ontario. Ecol. Model. 221, 1655–1664.

Stockwell, D.R.B., Peterson, A.T., 2002. Effects of sample size on accuracy of species distribution models. Ecol. Model. 148, 1–13.

Thomas, C.D., Cameron, A., Green, R.E., Bakkenes, M., Beaumont, L.J., Collingham, Y.C., Erasmus, B.F.N., de Siqueria, M.F., Grainger, A., Hannah, L., Hughes, L., Huntley, B., van Jaarsveld, A.S., Midgley, G.F., Miles, L., Ortega–Huerta, M.A., Peterson, A.T., Phillips, O.L., Williams, S.E., 2004. Extinction risk from climate change. Nature 427, 145–148.

Thuiller, W., Lavorel, S., Araújo, M.B., Sykes, M.T., Prentice, C., 2005. Climate change threats to plant diversity in Europe. Proc. Natl Acad. Sci. USA 102, 8245–8250.

Tokumine, S., 2002. Bioclimatic modelling of sub-Saharan African plants. Progress report to Conservation International, September 2002. Centre for Ecology Law and Policy, Environmental Department, University of York, UK.

Trivedi, M.R., Berry, P.M., Morecroft, M.D., Dawson, T.P., 2008. Spatial scale affects bioclimate model projections of climate change impacts on mountain plants. Glob. Change Biol. 14, 1089–1103.

Chapter Two

Species distribution modelling: Does one size fit all?

Abstract

Empirical models for predicting the distribution of organisms from environmental data have often focused on tenets of ecological niche theory. However, even at large scales, there is little agreement over how to represent the dimensions of a species' niche. The performance of such models is greatly affected by the nature of species' distributional and environmental data. Provincial distribution models were developed for 30 willow species in Ontario to examine (*i*) the predictive ability of logistic regression analysis, and (*ii*) the effects of using different distributional and environmental data sets. Two original measures of model accuracy and over-prediction were employed and evaluated using independent data. Models based on unique combinations of monthly climate data, rather than a fixed selection of annual and seasonal variables, predicted distributions most accurately for all species. Models based on a fixed set of variables, while generating the highest average probabilities of occurrence for certain species with limited ranges, resulted in the greatest under- and overestimates of willow distributions. Comparisons of models demonstrated climatic patterns among willows of differing habit and habitat. The distribution of dwarf willow species, present only in the Ontario arctic, followed gradients of summer maximum temperatures. The distribution of the tree species in the southerly portions of the province followed gradients of fall and winter minimum temperatures. Regardless of distributional and environmental data input, no algorithm maximized model performance for all species. Individual species' models require individual approaches; *i.e.*, the variable selection technique, the set of environmental factors used as predictors, and the nature of species distributional data must be carefully matched to the intended application. Unless

sampling bias and species' prevalence can be accounted for, models based on collection point data are best used to guide field surveys. While inferred range data may be better suited to determine potential ecological niches, overestimation of species' prevalence and environmental tolerance must be recognized. A combination of available distributional data types is recommended to best determine species niches, an important step in developing conservation strategies.

Keywords: BIOCLIM, continuous data, discontinuous data, logistic regression, model validation, occurrence probability, over-prediction index, *Salix*, willow

Introduction

Although niche theory has long been a primary focus in ecological research, no formal representation of the niche concept has been achieved that is accepted by all ecologists (Pennington 2006). Habitat modelling and studies of environmental variability in space have often sought to express the *functional* or *fundamental* niche defined by Hutchinson (1957) as the hyperdimensional space composed of the collection of environmental conditions where a species can exist. At large scales, climate is thought to exert a primary influence on the natural distribution of plants. Consequently, many studies have modelled the *known* distributions of species in terms of climate by using a chosen subset of factors representing the natural forces thought to limit plant growth and survival (for review see: Heikkinen *et al.* 2006). While there are many other non-climatic factors influencing the distribution of plants, these are rarely considered at the species level owing to problems in obtaining data.

Distribution models are influenced by scale and resolution (Pearson and Dawson 2003; Trivedi *et al.* 2006) and by the nature of the available distributional data. While *known* distributions are usually detailed as continuous geographic ranges, individual plant species are

unable to grow in all locations within these areas. Conversely, distributional data that reflect only known collection locations for species may underestimate their actual environmental limits (Stockwell and Peters 1999; Townsend Peterson and Cohoon 1999). Comparisons of the predictive abilities of different approaches have demonstrated that the best models include both presence and absence data for analysis (Pearson *et al.* 2006). For this reason, statistical models have generally outperformed classification tree approaches and genetic algorithms (Meynard and Quinn 2007).

Beyond recognition that all plant species have a need for heat sums and moisture to grow, there is little agreement as to *which* and *how many* climate variables should be included in a distribution model. While extremes of all climate variables will limit a species' distribution, in any one location, there may be only a few, truly limiting factors, and interactions *among* certain variables may be most important. Although he was working with physiognomic vegetation characteristics rather than individual species, Box (1981) suggested that larger sets of climate data might constrain potential ranges with superfluous requirements (model over-fitting).

Plant species occupy different niches, and there is no agreement as to whether unique sets or the same variables should be used for different species distribution models. Many studies employing distribution models to help estimate the impacts of climate change on plants are based on a single 'parsimonious' set of predictor variables utilized for all species (e.g. Bakkenes *et al.* 2002; Pearson and Dawson 2002; Thuiller *et al.* 2005; McKenney *et al.* 2007; Townsend Peterson *et al.* 2008). These variables are frequently generated by BIOCLIM, a *spatial bioclimatic prediction system* that derives surrogate variables from monthly or weekly climate estimates when coupled with a digital elevation model (Nix 1986).

This study addresses three questions. (*i*) To what extent is model performance affected by the nature of species' distributional data? (*ii*) How do individual species' distribution models based on a fixed set of predictors compare with those based on unique combinations of variables? (*iii*) Can model performance be maximized using a given variable selection technique for all species? To provide answers, comparative logistic regression analyses were conducted using different environmental data sets and two distribution data sets for 30 *Salix* L. (willow) species growing in the province of Ontario. Two original measures of model performance were determined and evaluated using independent environmental data for all of Canada.

Methods

Study species

Willow species in Ontario are morphologically diverse and well-distributed (Table 1). They inhabit every region of the province and range in stature from dwarf species less than 3 cm tall in the arctic to trees 25 m tall in the southern deciduous forests. As pioneer species, willows are well-adapted to disturbance, and are able to persist in a variety of habitats from very wet lowlands to well-drained uplands. Ecologically, willows provide food and shelter for a variety of organisms and are important stabilizers of riparian shores, sand dunes, glacial outwash, and anthropogenic disturbed areas (Argus 2006). Of the 30 willow species reviewed in this study, four are considered rare and vulnerable in Ontario (Oldham and Brinker 2009). They are *S. arctica* var. *kophophylla, S. myricoides, S. maccalliana, S. pseudomonticola*.

Table 1. Habit, location, cell count (#), and prevalence (%), for *Salix* species in Ontario

SPECIES data			Collection point		Inferred range	
Salix spp.	Height (m)	*Location	#	%	#	%
reticulata L.	0.03 - 0.15	N	120	0.7	603	3.3
arctophila Cock. ex Heller	0.03 - 0.15	N	147	0.8	689	3.8
arctica Pall. var. *kophophylla* (Schneid.) Polunin	0.3 - 0.5	N	33	0.2	380	2.1
myrtillifolia Anderss.	0.1 - 0.9	N, C	327	1.8	12563	69.1
brachycarpa Nutt.	0.3 - 0.9	N	130	0.7	833	4.6
pedicellaris Pursh	0.2 - 1.5	O	944	5.2		
vestita Pursh	0.2 - 1.5	N	117	0.6	827	4.5
lanata L. ssp. *calcicola* (Fern. &Wieg) Hult.	1.0 - 2.0	N	53	0.3	479	2.6
glauca L. ssp. *callicarpaea* (Trautv.) Böcher	0.2 - 2.5	N	133	0.7	1307	7.2
candida Fluegge	0.3 - 2.5	O	796	4.4		
cordata Michx.	1.0 - 3.0	S, N	339	1.9	2286	12.6
humilis Marsh.	0.3 - 3.0	S, C	881	4.8	11826	65.0
planifolia Pursh	1.0 - 3.0	C, N	527	2.9	14386	79.1
maccalliana Rowlee	0.9 - 3.5	C	89	0.5	6714	36.9
pyrifolia Andress.	0.4 - 4.0	C	723	4.0	14979	82.3
myricoides Muhl.	0.25 - 5.0	S, C	219	1.2	5213	28.7
serissima (Bailey) Fern.	1.0 - 5.0	O	803	4.4	16074	88.4
purpurea L.	1.5 - 5.0	S	20	0.1		
eriocephala Michx.	0.2 - 6.0	S, C	520	2.9	7780	42.8
pellita Andress.	0.5 - 6.0	C, N	416	2.3	14476	79.6
pseudomonticola Ball	1.0 - 6.0	C	130	0.7	6500	35.7
petiolaris J.E.Sm.	1.0 - 6.0	S, C	1264	6.9	8533	46.9
lucida Muhl.	4.0 - 6.0	S, C	1039	5.7	12889	70.9
discolor Muhl.	2.0 - 4.0 (8.0)	S, C	1139	6.3	10283	56.5
exigua Rowlee	4.0 - 9.0	S, C	586	3.2	5911	32.5
bebbiana Sarg.	0.5 - 10.0	O	2423	13.3		
fragilis L.	3.0 - 15.0 (20.0)	S	317	1.7		
amygdaloides Anderss.	3.0 - 20.0	S	316	1.7	2244	12.3
nigra Marsh.	> 20.0	S	51	0.3	1589	8.7
alba L.	10.0 - 25.0	S	196	1.1		

Bold typeface indicates introduced species; * designates rare species; * (N, C, S, O) = northern, central, southern, omnipresent

Model scale and resolution

Regional scale distribution models for 30 willow species were developed for the province of

Ontario, Canada, which extends from 41°51' - 56°49' N and 74°21' - 95°21' W, and covers a

land area of approximately 1 076 395 km^2. Provincial data for willow distributions, as well as

historical climate and elevation, were obtained and converted to grids with cell dimensions of 5'
x 5' using the Grid Package of Arc/GIS (ESRI 2008). Grids were composed of 18 190 cells,
each representing a geographic area of approximately 60 km^2, depending on latitude and
longitude.

Distributional data

Distributions for willow species in Ontario were available as discontinuous and continuous
data sets. Detailed *collection point* maps for 30 willow species (Soper and Heimburger 1982)
provided the discontinuous data set. *Inferred range* maps for North America, available for 27 of
these species (Argus 2007), provided the continuous data set. Owing to the unavailability of
continuous range maps for introduced species, models for *S. alba*, *S. fragilis*, and *S. purpurea*
were based solely on discontinuous data. The continuous distributions for *S. bebbiana*, *S.
candida*, and *S. pedicellaris* were excluded from the analysis because there was no variation in
the response variable; *i.e.*, the distributions of these species covered the entire study area. Fifty-
four willow distributions were scanned and converted into *presence* (1) and *absence* (0) grids.
Species distributions based on discontinuous data occupied an average of 2.7% (0.1-13.3%) of
the study area, in contrast with 36.5% (2.1-88.4%) for those based on continuous data.

Environmental data

Climate data for Ontario were available in the form of 30-year monthly averages for the
period of 1961-1990 from D. McKenney of the Canadian Forest Service. Data included average
minimum and maximum monthly temperatures (°C) and average monthly precipitation measures
(mm). Provincial elevation data (m) were obtained from a digital elevation model for Canada
downloaded from the Natural Resources Canada website (http://geogratis.ca/geogratis/en/option/
select.do?id=8880).

Two environmental data sets were used to model the distribution of willow species in

Ontario. The first contained 37 variables: average monthly temperature minima and maxima,

mean monthly precipitation, and elevation. The second consisted of seven, six of which were

climate variables adopted from BIOCLIM, considered to be critical to plant physiological

function and survival. These variables, derived from the monthly climate data, included mean

annual temperature and precipitation, the average temperature of the warmest and coolest months

(July and January respectively), the average seasonal precipitation of the warmest quarter (June,

July, and August), and the coolest quarter (December, January, and February) over the 30-year

period. Elevation was the seventh variable in this data set.

Logistic regression models

Data files were compiled for each species consisting of presence and absence scores and the

41 environmental variable values for each cell in the Ontario grid. Predictive models were

defined based on a randomly selected 70% of the cells from each species grid. Logistic

regression analysis (SAS 2008) was used to determine *which* and *how many* of the environmental

factors were statistically significant ($p < 0.05$) with respect to the observed species distributions.

The default settings were used (*i.e.*, the threshold value of $p > 0.5$ for scored *presences* was used

for comparisons of cases, and the procedure was allowed a maximum of 25 iterations to

determine a 'best-fit' model).

Three models for the distribution of each species and environmental data set combination

were generated. Models were based on (*i*) logistic regression analysis without variable selection,

(*ii*) forward stepwise logistic regression, and (*iii*) an inflection point selection of the chosen

stepwise variables. For analyses of the environmental data set comprising 37 variables, the

stepwise procedure was programmed to select a maximum of 12 significant predictors. Lastly,

an inflection point selection of the chosen stepwise environmental variables was determined for each species distribution.

The maximum likelihood estimates of each logistic regression analysis were used to create probability distribution maps using ArcMap (ESRI 2006) based on the odds ratio formula:

$$\hat{Y}_i = \frac{e^u}{1 + e^u} \qquad [1]$$

where $\hat{Y}_i$ is the estimated probability that the ith case (each grid cell) is in one of the categories (presence or absence) and u is the usual linear regression equation:

$$u = A + B_1X_1 + B_2X_2 + \cdots + B_kX_k \qquad [2]$$

with constant A, coefficients B_j, and predictors, X_j for k predictors ($j = 1, 2, \ldots, k$).

Model evaluation

In addition to concordance levels, two measures of model performance were calculated from the odds ratio estimates. (*i*) The average probability of willow occurrence was determined across the scored cell presences for each species distribution. (*ii*) An index of over-prediction was calculated. This index was determined by dividing the number of scored cell *absences*, with probabilities equal or greater than the average probability across the scored cell *presences*, by the number of scored cell *presences* for each species distribution.

Model validation

The distribution models determined by logistic regression for willow species in Ontario were used to make range predictions for all of Canada. Each model was evaluated using an independent historical climate data set for Canada averaged over the 1950-2000 period (http://www.worldclim.org/current). The resulting probability distributions were visually compared with recently published North American *Salix* range maps (Argus 2007).

Results

Model performance

i. Concordance

The highest concordance levels were achieved for models of species with limited

distributions (Table 2). Models based on unique combinations of monthly climate variables,

especially those based on the full set of 37 predictors produced the highest average levels of

concordance. However, these models failed to converge for several arctic species (3 using

discontinuous and 5 using continuous data). Models for *S. purpurea* and *S. nigra* based on the

discontinuous data achieved maximum concordance levels with the BIOCLIM and stepwise

variable selections respectively. With the exception of *S. purpurea*, models based on BIOCLIM

variables produced the lowest average concordance levels.

The results of the stepwise and inflection point analyses of the BIOCLIM variables are not

reported for several reasons. Models based on the discontinuous data had the lowest levels of

concordance and produced the least representative maps of species input distributions. The

inflection point models based on the continuous data were similarly inadequate. Furthermore,

nearly all the stepwise models based on the continuous data retained the full complement of

seven variables; *i.e.*, stepwise models were identical to those produced by logistic regression

without variable selection.

Table 2. Logistic regression model concordance levels for *Salix* species in Ontario

| SPECIES data | *Collection point* | | | | *Inferred range* | | | |
CLIMATE data	full set	stepwise	inflection	BIOCLIM	full set	stepwise	inflection	BIOCLIM
reticulata	n/a	**99.2**	98.4	98.7	n/a	**100.0**	99.7	99.9
arctophila	**98.8**	98.5	98.1	98.2	n/a	**99.5**	99.5	99.5
arctica	n/a	**99.7**	98.7	99.7	n/a	**100.0**	99.7	99.8
myrtillifolia	90.3	88.5	87.2	83.5	**99.9**	99.2	97.7	96.8
brachycarpa	**96.6**	96.2	95.5	92.4	n/a	**99.9**	99.0	98.1
pedicellaris	**74.4**	70.8	66.6	60.8				
vestita	**98.1**	97.5	91.8	93.9	**99.8**	98.3	97.3	97.8
lanata	n/a	**99.4**	98.6	99.0	n/a	**99.9**	99.7	99.8
glauca	**95.5**	95.1	94.5	94.7	**99.3**	98.8	96.3	95.5
candida	**80.3**	76.9	73.0	70.3				
cordata	**89.2**	87.7	81.5	79.4	**99.2**	98.9	96.5	92.3
humilis	**82.6**	81.2	78.1	78.6	**98.6**	98.2	94.6	94.6
planifolia	**84.2**	81.0	70.9	68.8	**99.8**	99.5	97.5	97.7
maccalliana	**91.1**	84.4	65.2	70.9	**99.6**	98.7	92.5	89.9
pyrifolia	**84.0**	81.7	75.1	73.1	**99.3**	97.9	96.0	96.1
myricoides	**92.6**	91.6	82.1	71.0	**98.7**	95.7	81.5	76.5
serissima	**87.2**	83.2	80.0	76.2	**98.7**	97.6	95.1	92.9
purpurea	92.6	91.8	90.4	**97.4**				
eriocephala	**90.3**	89.3	87.9	76.9	**99.4**	99.1	98.0	97.8
pellita	**81.9**	76.7	71.2	66.5	**99.7**	99.4	97.4	97.2
pseudomonticola	**96.0**	91.1	81.6	77.7	**99.5**	98.8	96.1	89.7
petiolaris	**93.6**	92.8	90.9	90.5	**99.7**	99.5	98.8	97.7
lucida	**87.3**	86.2	83.8	83.0	**99.3**	99.0	98.0	98.0
discolor	**89.9**	87.8	86.4	85.2	**99.6**	99.3	98.6	98.6
exigua	**87.6**	86.1	83.0	77.1	**94.5**	93.7	87.9	84.6
bebbiana	**91.7**	90.4	90.0	88.0				
fragilis	**98.0**	97.5	95.3	96.2				
amygdaloides	**97.0**	96.6	95.7	95.6	**99.9**	99.9	98.8	99.3
nigra	96.1	**97.9**	96.8	96.6	**100.0**	100.0	99.8	99.7
alba	**97.4**	93.7	95.0	94.9				
AVERAGE	**90.5**	89.7	86.1	84.5	**99.2**	98.8	96.5	95.4

Bold typeface indicates the highest concordance levels; n/a indicates a failure of convergence

ii. Probability

No variable selection method consistently produced the highest probabilities for species models based on continuous data. The highest probabilities of willow presence over the observed input distributions were achieved using the BIOCLIM variables for models of species

with the lowest and highest prevalence (Table 3). The BIOCLIM models also produced the

highest probabilities for all arctic species. In contrast, models for these species based on unique

combinations of monthly climate variables produced the lowest probabilities.

Table 3. Average probabilities of grid cells scored present for *Salix* species in Ontario

SPECIES data	Collection point				Inferred range			
CLIMATE data	full set	stepwise	inflection	BIOCLIM	full set	stepwise	inflection	BIOCLIM
reticulata	n/a	0.395	0.334	**0.835**	n/a	0.083	0.900	**0.995**
arctophila	0.217	0.314	0.294	**0.560**	n/a	0.053	0.405	**0.982**
arctica	n/a	0.267	0.073	**0.960**	n/a	0.949	0.796	**0.997**
myrtillifolia	**0.284**	0.163	0.147	0.134	**0.953**	0.910	0.875	0.910
brachycarpa	0.004	0.000	0.087	**0.221**	n/a	0.201	0.261	**0.893**
pedicellaris	**0.251**	0.083	0.074	0.073				
vestita	**0.302**	0.246	0.117	0.011	0.259	0.269	0.654	**0.960**
lanata	n/a	0.088	0.129	**0.919**	n/a	0.937	0.490	**0.992**
glauca	0.001	0.020	0.215	**0.431**	0.399	0.077	0.410	**0.787**
candida	**0.332**	0.159	0.130	0.101				
cordata	**0.181**	0.035	0.071	0.076	0.577	0.649	**0.734**	0.652
humilis	**0.166**	0.103	0.119	0.068	0.798	0.850	0.839	**0.885**
planifolia	0.108	**0.287**	0.056	0.112	0.805	0.915	**0.956**	0.937
maccalliana	**0.444**	0.148	0.015	0.015	**0.766**	0.741	0.654	0.482
pyrifolia	**0.318**	0.243	0.043	0.048	0.958	**0.985**	0.929	0.916
myricoides	**0.257**	0.184	0.057	0.048	**0.902**	0.865	0.461	0.386
serissima	**0.357**	0.293	0.120	0.103	0.949	**0.969**	0.958	0.950
purpurea	**0.769**	0.225	0.058	0.118				
eriocephala	**0.423**	0.257	0.170	0.188	**0.967**	0.913	0.882	0.831
pellita	**0.184**	0.147	0.122	0.055	0.911	**0.959**	0.938	0.957
pseudomonticola	**0.521**	0.458	0.040	0.027	0.606	**0.783**	0.776	0.698
petiolaris	0.455	**0.504**	0.223	0.187	0.953	**0.961**	0.696	0.617
lucida	**0.418**	0.183	0.147	0.157	0.904	**0.965**	0.935	0.918
discolor	**0.555**	0.199	0.177	0.137	0.908	0.924	**0.927**	0.780
exigua	**0.356**	0.220	0.149	0.195	**0.658**	0.584	0.543	0.537
bebbiana	0.587	0.547	**0.611**	0.539				
fragilis	0.140	**0.351**	0.199	0.160				
amygdaloides	**0.408**	0.283	0.165	0.098	**0.929**	0.927	0.634	0.763
nigra	0.420	**0.446**	0.079	0.086	**0.946**	0.932	0.834	0.901
alba	**0.470**	0.233	0.105	0.155				
AVERAGE	**0.331**	0.236	0.144	0.227	0.797	0.725	0.729	**0.822**

Bold typeface indicates the highest probabilities

Even though the models based on discontinuous data for *S. arctophila*, *S. brachycarpa* and *S. glauca* did not suffer failures of convergence, the exceedingly low probabilities are indicative of model over-fitting. Using the discontinuous data, the highest average probabilities were achieved for models based on the full set of 37 predictors. Higher probabilities were reached for only four models based on stepwise selections of variables (*S. planifolia*, *S. petiolaris*, *S. fragilis*, and *S. nigra*) and one inflection point model (*S. bebbiana*).

iii. Over-prediction

The lowest over-predictions of willow distributions were achieved for models based on the continuous data using unique combinations of monthly climate variables (Table 4). Using the discontinuous data, the lowest over-prediction ratios were achieved with models based on stepwise selections of monthly climate variables. Using the continuous data, the lowest over-prediction ratios were achieved with models based on inflection point selections of monthly climate variables. The highest over-predictions of willow distributions were for models of species with the lowest prevalence in the data sets, especially for those based on the discontinuous data. The BIOCLIM models generally produced the highest over-predictions of willow distributions, especially for the arctic species, many of which have extremely limited ranges within Ontario and in North America.

Table 4. Over-prediction ratios for distribution models of *Salix* species in Ontario

SPECIES data	Collection point				Inferred range			
CLIMATE data	full set	stepwise	inflection	BIOCLIM	full set	stepwise	inflection	BIOCLIM
reticulata	n/a	**1.164**	1.766	5.133	n/a	**0.013**	0.051	1.218
arctophila	**0.993**	2.358	1.399	2.399	n/a	n/a	**0.509**	1.828
arctica	n/a	**4.758**	5.000	10.273	n/a	**0.381**	0.609	1.775
myrtillifolia	6.893	**2.669**	4.524	4.907	0.131	0.053	**0.013**	0.017
brachycarpa	**1.362**	n/a	3.138	8.838	n/a	**0.010**	0.275	2.863
pedicellaris	5.090	**2.727**	3.074	2.955				
vestita	5.270	7.852	**1.270**	3.157	0.438	0.342	**0.142**	1.248
lanata	n/a	**1.226**	9.528	17.038	n/a	**0.151**	0.459	3.085
glauca	n/a	**0.723**	1.723	3.560	0.382	**0.272**	0.466	1.546
candida	6.106	3.334	**1.173**	2.135				
cordata	3.541	**1.772**	2.264	2.147	0.642	0.725	**0.130**	0.219
humilis	3.002	**2.275**	2.564	3.029	0.152	**0.009**	0.021	0.027
planifolia	**3.992**	7.234	4.390	7.504	**0.005**	0.010	0.007	0.009
maccalliana	32.297	**15.022**	26.143	35.780	0.196	0.151	**0.086**	0.099
pyrifolia	3.589	**3.282**	4.312	6.427	0.144	0.031	0.008	**0.006**
myricoides	13.680	6.345	**2.123**	4.695	1.085	0.341	**0.302**	0.367
serissima	4.361	3.326	1.297	**1.278**	0.073	0.052	**0.007**	0.016
purpurea	46.150	**5.350**	22.300	17.950				
eriocephala	2.463	**0.924**	1.558	1.231	0.384	0.029	**0.013**	0.016
pellita	5.486	**2.732**	4.007	8.047	0.014	0.009	**0.005**	0.005
pseudomonticola	19.772	**7.497**	12.414	22.055	**0.094**	0.097	0.200	0.241
petiolaris	1.336	1.632	**0.573**	0.628	0.215	0.032	**0.008**	0.081
lucida	2.354	2.129	1.293	**1.113**	0.194	0.102	0.008	**0.006**
discolor	1.944	**0.742**	0.813	0.850	0.173	**0.007**	0.009	0.037
exigua	2.373	**0.658**	0.671	1.091	0.322	0.197	**0.077**	0.127
bebbiana	0.286	**0.121**	0.159	0.138				
fragilis	2.613	2.443	**1.232**	1.440				
amygdaloides	4.632	3.256	2.068	**1.347**	0.347	0.209	**0.015**	0.051
nigra	20.729	8.966	9.169	**7.119**	0.233	0.179	**0.011**	0.031
alba	9.174	1.969	**1.456**	1.569				
AVERAGE	8.057	**3.603**	4.447	6.194	0.275	0.148	**0.143**	0.622

Bold typeface indicates the lowest over-prediction ratios

Significant predictor variables

Logistic regression of species distributions against the 37 variable set produced models with an average of 12 (1-27) significant predictors based on the discontinuous data, and 21 (3-33) based on the continuous data. Stepwise models, based on discontinuous data, selected fewer

than 12 (4-9) significant variables for several willow species with low prevalence: *S. arctica, S. arctophila, S. lanata, S. glauca, S. purpurea, S. nigra, S. alba*, and *S. maccalliana*. Similarly, the stepwise models based on continuous data selected fewer than 12 (6-10) significant variables for four arctic species: *S. reticulata, S. arctica, S. arctophila,* and *S. lanata*. Inflection points were reached with the inclusion of only 2 variables for 14 species models using the discontinuous data and 17 using continuous data inputs. For most of the arctic species models under either distributional input, an inflection point was identified following the inclusion of 2 variables. Most of the models for provincially well-distributed species required 3-4 variables.

The species with the lowest prevalence in the data sets required the fewest predictors to model accurately, and there was considerable agreement in the selections of variables within species models based on different distributional data. In contrast, well-distributed species were generally modelled the least accurately (lower concordance and probabilities of occurrence), and there was seldom any agreement in the variable selections between species models using different distributional data.

June maximum temperature was the strongest predictor in nearly all the stepwise models of the dwarf willows in the arctic. August maximum temperature and elevation frequently appeared within the first 4 variable selections of these species models. July minimum temperature was the strongest predictor for stepwise models of *S. pseudomonticola*. Fall and winter minimum temperatures most strongly predicted the distributions of the tree species in the southern portions of Ontario. September minimum temperature was the strongest predictor in the stepwise models of *S. amygdaloides*, *S. fragilis*, and *S. alba*. January minimum temperature was the strongest predictor in the stepwise models for *S. nigra* and *S. purpurea*.

Model implementation in GIS

In general, the maximum average probabilities and lowest over-prediction levels, for models based on the either distributional data input, were achieved using stepwise and inflection point selections of monthly climate variables. To illustrate, three willow species with contrasting provincial distributions were chosen. The stepwise and BIOCLIM models based on continuous data for *S. arctica*, are presented (Fig. 1). While model concordance levels (100% and 99.8% respectively) and average probabilities across the input occurrence cells (0.949 and 0.997) did not differ appreciably, the over-prediction ratio for the stepwise model was approximately 4.7 times lower than that of the BIOCLIM model (0.381 and 1.775). The stepwise model closely followed the input distribution (Fig. 1-B) while the BIOCLIM model over-estimated this species' range to the point of suggesting areas of suitable habitat in the southernmost portions of the province (Fig. 1-C).
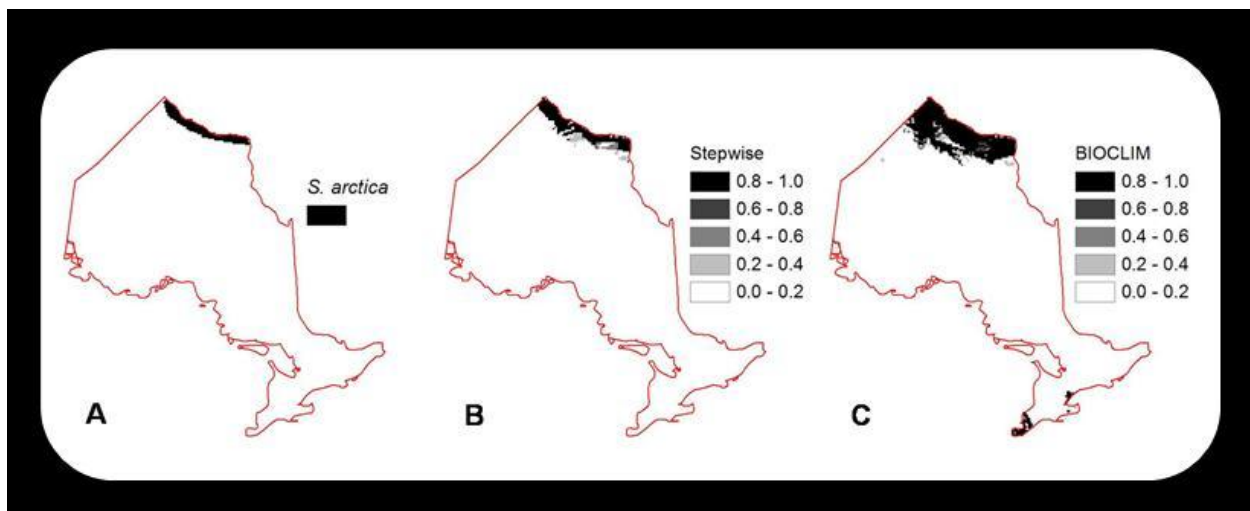
Fig. 1. Ontario distributions for *S. arctica* showing (A) continuous data input (Argus 2007); and probabilities of occurrence based on (B) a stepwise selection of monthly climate variables (June maximum temperature and September minimum temperature were the strongest predictors); and (C) BIOCLIM variables

The inflection point and BIOCLIM models based on continuous data for *S. discolor,* are

presented (Fig. 2). Although concordance levels were identical using both approaches, the

inflection point model generated the highest average probability across the input occurrence cells

of 0.927 versus 0.780 based on BIOCLIM, as well as a lower over-prediction ratio of 0.009

compared to 0.037. The inflection point model, based on the effects of May and August

precipitation and December minimum temperature, accurately depicted the input distribution,

correctly delimiting the northern tolerance of this species (Fig. 2-B). The BIOCLIM model,

however, was far less representative, selecting areas of unsuitable habitat in the western and

southern portions, as well as over-estimating the northern limit of this species' provincial

distribution (Fig. 2-C).
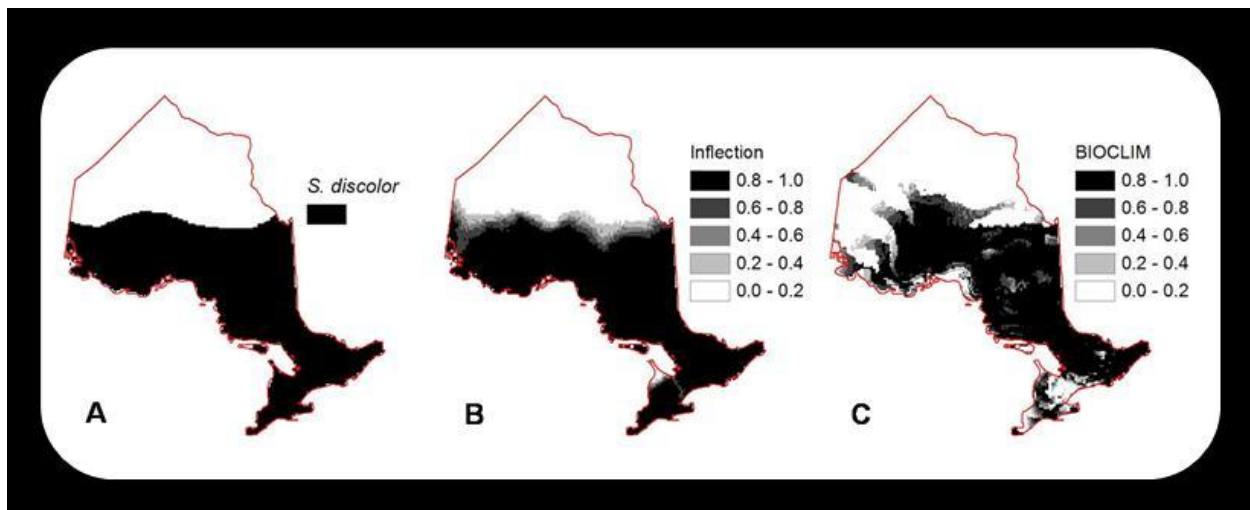
Fig. 2. Ontario distributions for *S. discolor* showing (A) continuous data input (Argus 2007); and probabilities of occurrence based on (B) an inflection point selection of monthly climate variables (May and August precipitation and December minimum temperature); and (C) BIOCLIM variables

The stepwise and BIOCLIM models based on continuous data for *S. amygdaloides* are

presented (Fig. 3). Although both models reached similar levels of concordance (99.9% and

99.3% respectively), the average probabilities across the input occurrence cells differed

(stepwise 0.927 and BIOCLIM 0.763) as did the over-prediction ratios (stepwise 0.209 and

BIOCLIM 0.051). The stepwise model most faithfully represented the input distribution and

identified areas of similar climate falling outside the range (Fig. 3-B). Despite its lower over-

prediction ratio, the model based on BIOCLIM variables failed to adequately represent areas in

the western and southeastern portions of this species' provincial distribution (Fig. 3-C).
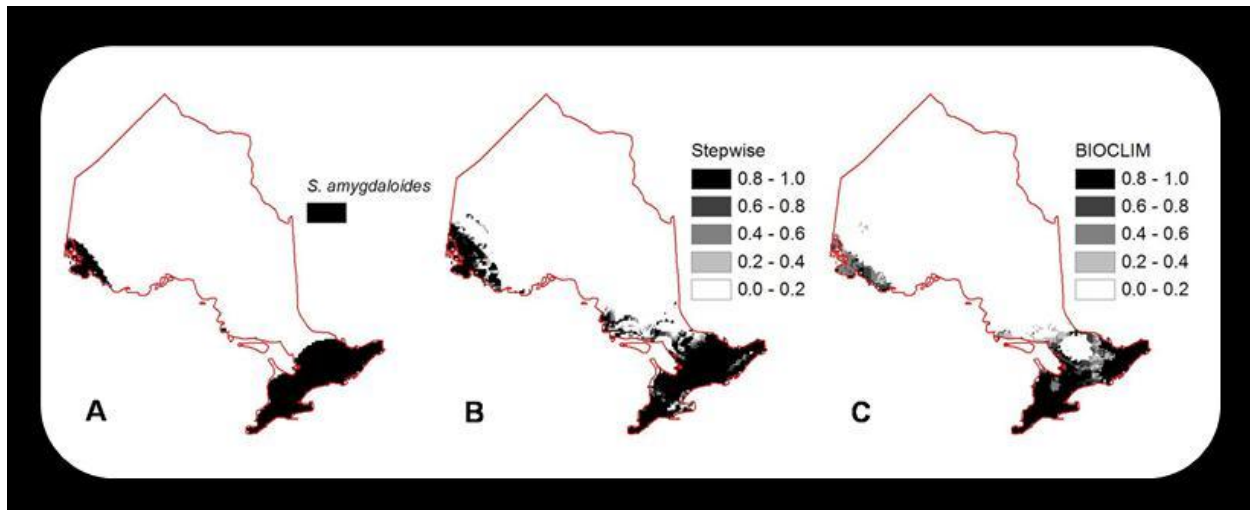
Fig. 3. Ontario distributions for *S. amygdaloides* showing (A) continuous data input (Argus
2007); and probabilities of occurrence based on (B) a stepwise selection of monthly climate
variables (September minimum temperature and August maximum temperature were the
strongest predictors); and (C) BIOCLIM variables

Model validation

Generally, models based on the *discontinuous* species data, regardless of variable selection

method, were unrepresentative of willow species' Canadian ranges. The exceptions were models

based on unique combinations monthly climate variables for species with distributions

concentrated in the southern portions of the province. In contrast, many of the models based on

the *continuous* species data, especially those based upon a stepwise or inflection point selection

of monthly climate variables, showed considerable agreement. Models based on the full set of

37 variables expressed fair representations of the Canadian ranges for 15 species. Regardless of

distributional input, the validation attempts for the arctic species models showed only partial agreement with published Canadian ranges, and frequently failed to represent the original Ontario distributions.

Models based on BIOCLIM variables, were generally unrepresentative of willow species ranges within Ontario and across Canada. The most surprising exception was the BIOCLIM model based on discontinuous data for *S. bebbiana*, which failed to adequately reproduce the observed Ontario input distribution using the training data but produced a fairly accurate representation of the Canadian range for this species. The only other exceptions were BIOCLIM models based on discontinuous data for *S. discolor* and *S. humilis*, and those based on continuous data for *S. lucida* and *S. pyrifolia*. While producing fair representations of these species' distributions in Ontario and in the immediately adjacent provinces, the BIOCLIM models greatly overestimated the ranges in other parts of the country.

One example validation series is presented for the models based on the continuous data of *S. maccalliana* (Fig. 4). The predicted distribution using all 37 environmental variables closely follows the entire Canadian distribution (Fig. 4-A). Stepwise (12 variables) and inflection point (2 variables) methods also closely approximate the east to west limits of the species (Figs. 4-B and C). Although the inflection method shows deviation to the north in western Canada, striking patterns of February minimum and March maximum temperatures are evident. The model based on BIOCLIM variables predicts species presence to the north and west with almost no overlap of the actual distribution, *not even in Ontario* (Fig. 4-D).
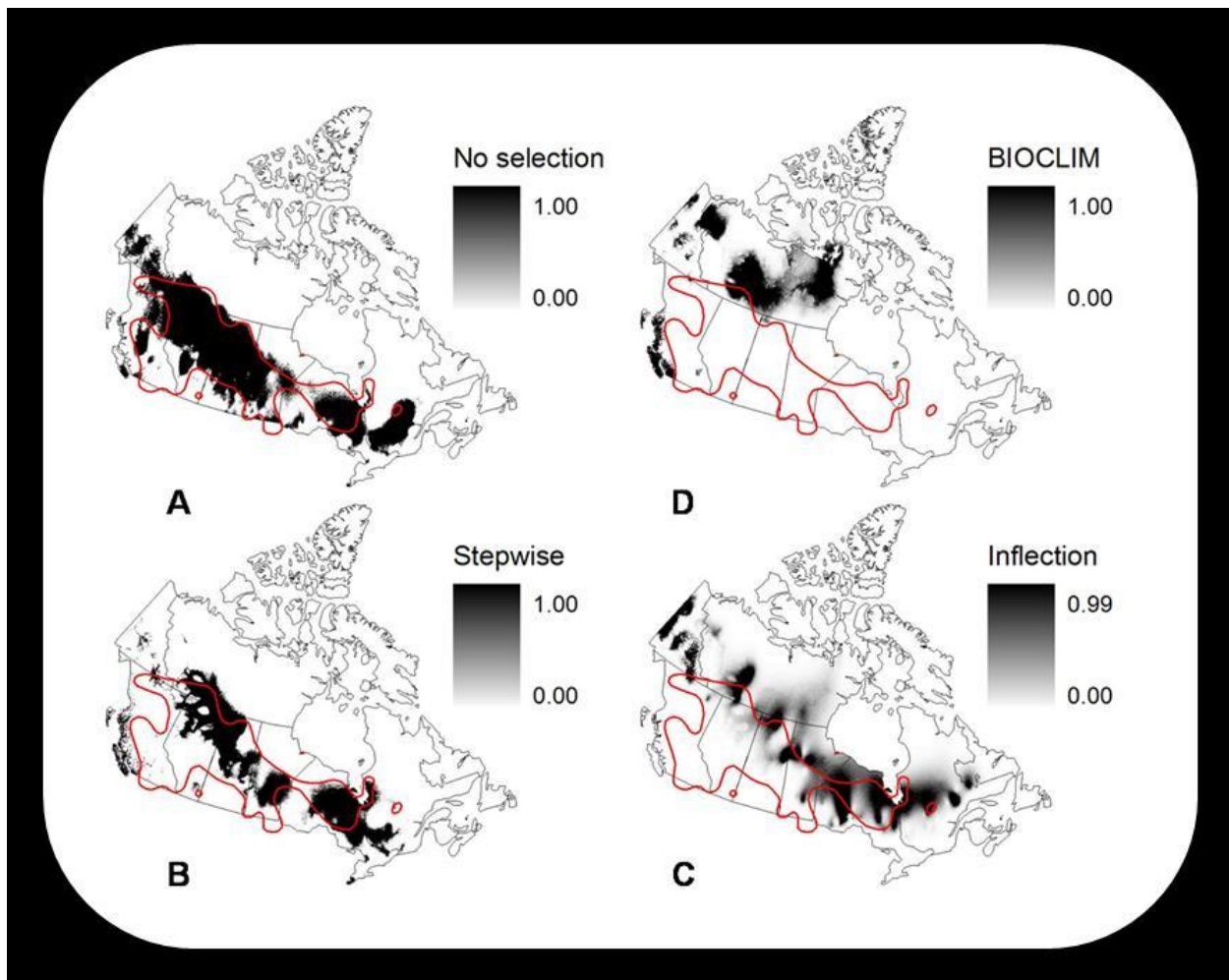
Fig. 4. Canadian distributions for *S. maccalliana* showing inferred range (solid line, Argus 2007); and probabilities of occurrence using Ontario data based on (A) logistic regression without variable selection; (B) a stepwise selection of variables; (C) an inflection point selection of variables; and (D) BIOCLIM variables

Discussion

This study assessed the predictive accuracy of logistic regression by modelling the distributions of willow species in Ontario using different distributional and environmental data sets. Six hundred and forty-eight probability maps were generated for this analysis. Models based on unique combinations of monthly climate data, rather than a fixed set of annual and seasonal averages, predicted the distribution of *all* species most accurately. Comparisons of models based on different distributional inputs revealed similarities in the selection of predictor

variables mainly for species with limited provincial ranges. No variable selection method optimized predictions for all species.

The nature of the distributional data

i. Collection point data

Using the discontinuous data, logistic regression models based on the full set of monthly climate data (37 variables) produced the most accurate predictions for the majority of willow species distributions. While these distributions consisted of *known* collection points, nothing was known about the absences comprising an average of 97.3% (86.7-99.9%) of the sampled data sets. The models for provincially well-distributed species required more environmental data to form accurate predictions. The distributions of these species were represented by relatively few occurrences dispersed over broad areas; thus, the increased variation in environmental tolerances required more predictor variables to be retained in the final models. Models based on all 37 predictors for arctic species, whose distributions occupied less than 3% of the area in Ontario, generally suffered from over-fitting (Table 2). Logistic regression is sensitive to the case (species presence) to variable (environmental predictor) ratio, and is thought to bias its results towards the more prevalent category of presence or absence (Fielding and Bell 1997).

Despite extremely low prevalence in the distributional datasets, the models of *S. purpurea* and *S. nigra* did not suffer the same failures of convergence as did those of the arctic species when using all 37 variables. These species, owing to their limited southerly distributions, were represented in respective datasets by only 14 and 36 *presences* across the 12 733 analysed cells. Logistic regression can achieve very high levels of accuracy with fewer points than previously reported (Stockwell and Peterson 2002). These results indicate that range size for willow does not affect model performance, and a small sample size may adequately represent dominant

ecological patterns, as were detected in these southern cases. This finding is at odds with the results of other studies which have suggested that models tend to over- or underestimate the probability of species' occurrence if sample prevalence is atypically high or low (Cumming 2000; Pearce and Ferrier 2000). Although the effects of species range size on model accuracy appear to be largely artefactual, McPherson *et al.* (2004) have recommended that intermediate levels of sampling prevalence should be used to achieve optimal model performance.

The concentration and pattern of known collection points, rather than their quantity, greatly influenced species model accuracy. For certain common and widespread species, a bias in the distribution of collection points towards areas of human development was evident. Despite high concordance levels, the modelled distributions for these species reflected the dominant concentration of collection points. As an example, the models for *S. bebbiana* reached concordance levels ranging from 88.0-91.7%, with above-average probabilities of occurrence (0.539-0.611), and the lowest over-prediction ratios (0.121-0.286). However, the predicted distributions, while accurately reflecting the southerly bias in the collected specimen data, failed to capture the panprovincial extent of this common and widespread species.

ii. Inferred range data

Using the continuous data, the stepwise and inflection point models based on monthly climate data represented willow distributions most accurately. The greater prevalence in the continuous data sets, in which an average of 36.5% (2.1-88.4%) of the samples contained presences, had the effect of increasing the confidence in the test, and, consequently, fewer variables were required to predict species distributions accurately. However, while eliminating the sampling bias of known collection points, models based on the continuous data likely provided an overly

optimistic estimate of species range tolerances, as areas of known unsuitable habitat such as lakes and rivers were not excluded from the analysis.

The range maps used for these analyses were only rough estimates based on known collection points, and therefore provided no measure of species prevalence across the study area. To illustrate how this may be problematic, *S. bebbiana* and *S. candida* occupy identical provincial area according to the available range maps, but while the former is very commonly observed in the province, the latter is much less so. Both species therefore cannot be assumed to have identical probabilities of occurrence, but this difference would not be reflected in the results of a test which ignores prevalence.

The nature of the environmental data

Models built using a fixed set of BICOLIM variables (annual means and quarterly sums) produced the least accurate predictions of willow distributions. Mean annual temperature is dependent on the maximum and minimum temperatures of the warmest and coolest months, just as mean annual precipitation is dependent on the precipitation of the warmest and coolest quarters. The highly correlated nature of these variables led to redundancies in the analyses which may have caused the erroneous predictions. These undesirable properties of the BIOCLIM variables may have negative implications for many recent predictions based on fixed, and perhaps overly parsimonious, climate data sets concerning the potential future habitat of species in the face of climate change (Iverson and Prasad 1998; Araújo *et al.* 2005; Thuiller *et al.* 2005; Hijmans and Graham 2006; Prasad *et al.* 2006; Rehfeldt *et al.* 2006; McKenney *et al.* 2007; Townsend Peterson *et al.* 2008). As a further consequence of using a small, fixed set of intercorrelated variables to predict distributions, information about the ecology of individual species is lost.

Flantua *et al.* (2007) suggested that the use of additional non-climate variables may improve model performance. Pearson and Dawson (2003) have recommended that distribution models at the regional scale should include topographic data. Elevation was a significant predictor in many models and it was highly significant ($p < 0.0001$) in nearly all the models of the arctic species. Despite its frequent selection in many species models, the contribution of the elevation variable to the final regression equations was negligible. Elevation coefficient values for the models of arctic species ranged from 0.0004 using the discontinuous data sets to 0.001 using the continuous data. Elevation made similarly small contributions to the models for well-distributed species (0.0002-0.0006). Elevation is therefore not considered relevant at the scale and resolution of the distribution models developed in this study.

Model validation

With few exceptions (Calef *et al.* 2005; Araújo *et al.* 2006; Peterson *et al.* 2008), attempts to validate distribution models have relied primarily on test statistics to gauge performance and accuracy (Manel *et al.* 2001; Anderson *et al.* 2003; McPherson *et al.* 2004; Pearson *et al.* 2006). While it has been suggested that the confidence in using logistic regression to model distributions is highest for interpolation, rather than extrapolation of the results (Calef *et al.* 2005), many of the validation attempts in this study were successful in predicting species ranges well beyond the provincial boundaries of the test. In all cases the most representative validation attempts were generated using unique combinations of predictors from monthly climate variables.

However, models for which parameters are estimated in one period may not directly apply to a different period (Ibáñez *et al.* 2005). In this study, the distributions of species with limited provincial ranges were modelled the most accurately using the test data. Nonetheless, the

predictions of Canadian ranges were successful for the models of willow species found in the southern portions of the province. The models for the arctic species, however, showed only partial agreement with the published Canadian ranges, and frequently failed to adequately represent the observed Ontario distributions. A second validation attempt for the arctic species, using 1971-2000 climate data (McKenney *et al.* 2006), resulted in distributions that were slightly more representative although a distinct shift eastward in many ranges was observed. Many of these arctic species have extremely limited provincial distributions and among them, *S. arctica* is considered rare and vulnerable. Consequently, there is some question about the usefulness of the distribution models calculated in this study for temporal extrapolation purposes. These results raise concerns about the validity of basing distribution models upon a single time series of climate data, especially for rare species.

Ecological significance of the predictors

This study has allowed a first formal evaluation of the potential environmental influences on the distribution of willow species in Ontario. However, more work is needed to better understand this interaction. Comparisons of models within well-distributed species did not show much agreement in the selections of predictor variables. Since different combinations of climate variables produced similar spatial patterns for these species, the monthly climate averages are correlated with one another in some way.

In spite of great variation in the selections of climate predictor variables within species models, strong climatic patterns were detected in the comparisons of models for several species of similar habit and habitat. The validation attempts for models of the tree species present in the southern portions of the province demonstrated a consistency in fall and winter temperatures across different time periods. September, October, and January minimum temperatures remained

stable over the distributions of these species for three periods (1950-2000, 1961-1990, 1971-2000), with averages of 9.2°C (4-13°C), 3.5°C (1-6°C), and -11.2°C (-16 - -8°C) respectively.

The models of the dwarf species present in the arctic identified strong patterns in summer temperatures over the 1961-1990 period. June and August maximum temperatures were the strongest predictors in these distribution models with average values of 14.1°C (11.7-16.3°C) and 16.5°C (16-18.5°C) respectively. Despite high levels of model performance with relatively few significant predictors, the mapping validation attempts for these species models were not entirely successful. Many of these rare arctic species have distributions extending much further north into the polar circle. In addition to their consistent proximity to sea level, the arctic species occupy higher latitude ranges where conditions for persistence are harsher; *i.e.* the physical factors regulating their distributions are more restrictive.

Kevan (1990), while discussing the importance of temperatures within the catkins of *S. arctica* with respect to reproductive success, presented evidence that the timing of blossoms correlated with the month of June when solar radiation is at its highest. Over the period of their development through insolation, willow catkins received additional heat units 25-50% beyond those indicated by the normal ambient air temperatures (7-16°C). Although his study was conducted much further north (81°49'N, 71°18'W), his findings provide a possible explanation of the failures of mapping validation attempts for the arctic species. The average 1961-1990 June maximum temperature in Ontario over the distribution for *S. arctica* was 12.6°C (12-13°C), a tolerance threshold so narrow as to make extrapolation of this species' provincial range impossible with these data. Multiple time series of climate data, as well as other environmental factors, such as measures of insolation, may be better suited to model the distributions of arctic willow species.

Understanding the results of single-species modelling exercises hinges greatly upon a knowledge of evolutionary responses (Davis *et al.* 2005). Further research with *Salix* distributional models should also consider the additional evolutionary processes of hybridization and introgression, as well as the differences in response to the environment of male and female plants (Kevan 1990; Jones *et al.* 1999). Furthermore, Betts and Shugart (2005) caution that simple cause and effect assumptions are inapplicable to the problem of predicting distributions unless the influence of species upon the environment is considered.

Conclusion

All species have different climatic tolerances and respond differently in their environment. The response of individuals at distributional extremes will likely differ from those occupying the central parts of their ranges. A central assumption underpinning many of the distribution modelling studies to date is that species are in equilibrium with their environment. This study determined Ontario distribution models based on 1961-1990 monthly climate averages for willow specimen data collected mostly before 1982. Model validation attempts revealed great spatial variability in the climate of northern latitudes using the data interpolated over different periods (1950-2000 and 1971-2000). Whether this finding accurately reflects historical trends, or merely a deficiency in the northern data owing to fewer weather stations, is difficult to ascertain. Perhaps willow species for which the 1961-1990 averages predicted accurately across the different periods are in equilibrium with their environmental conditions.

While finding a balance between overly parsimonious and over-fitted models is critical to reducing uncertainty in model prediction (Heikkinen *et al.* 2006), the results of this study indicate that the models of individual species, even on large scales and at relatively coarse resolution, require unique combinations of predictors. More information concerning the

prevalence of species, and corresponding with multiple time series of environmental data, would likely enhance the predictive ability of distribution models.

The nature of the distributional data of species must be carefully matched to the intended application of any modelling exercise. Unless sampling bias and prevalence can be accounted for, the discontinuous data are best used to guide field surveys. Exclusive use of known collection points should not be used to determine niche dimensions, especially if the purpose is to forecast future potential habitat under predicted climate change. While continuous data are better suited to this purpose, overestimation of prevalence and environmental tolerances of individual species should be recognized.

Acknowledgements

References

Anderson, R.P., Lew, D., Townsend Peterson, A., 2003. Evaluating predictive models of species'distributions: criteria for selecting optimal models. Ecol. Model. 162, 211-232.

Araújo, M.B., Pearson, R.G., Thullier, W., Erhard, M., 2005. Validation of species-climate impact models under climate change. Global Change Biol. 11, 1504-1513.

Argus, G.W., 2007. Salix (Salicaceae) Distribution maps and a synopsis of their classification in North America, north of Mexico. Harvard Papers in Botany. 12, 335-368.

Argus, G.W., 2006. Guide to Salix (Willow) in the Canadian Maritime provinces (New Brunswick, Nova Scotia, and Prince Edward Island). http://aknhp.uaa.alaska.edu/willow/ pdfs/GuideMaritimeSalix28JUN06.pdf (downloaded April 2008).

Betts, R.A., Shugart, H.H., 2005. Dynamic ecosystem and earth system models. Climate Change and Biodiversity. (eds. T.E. Lovejoy & L. Hannah) pp. 232-251. Yale University Press, New Haven, CT.

Box, E.O., 1981. Predicting physiognomic vegetation types with climate variables. Vegetatio. 45, 127-139.

Bakkenes, M., Alkemade, J.R.M., Ihle, F., Leemans, R., Latour, J.B., 2002. Assessing effects of forecasted climate change on the diversity and distribution of European higher plants for 2050. Global Change Biol. 8, 390-407.

Calef, M.P., McGuire, A.D., Epstein, H.E., Rupp. T.S., Shugart, H.H., 2005. Analysis of vegetation distribution in interior Alaska and sensitivity to climate change using a logistic regression approach. J. Biogeogr. 32, 863-878.

Cumming, G.S., 2000. Using between-model comparisons to fine-tune linear models of species ranges. J. Biogeogr. 27, 441-455.

Davis, M.B., Shaw, R.G., Etterson, J.R., 2005. Evolutionary responses to changing climate. Ecology. 86, 1704-1714.

ESRI., 2006. ArcGIS Desktop 9.2. Environmental Systems Research Institute Inc., Redlands, CA.

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

Flantua, S.G.A., van Boxel, J.H., Hooghiemstra, H., van Smaalen, J., 2007. Application of GIS and logistic regression to fossil pollen data in modelling present and past spatial distribution of the Columbian savanna. Clim. Dynam. 29, 697-712.

Heikkinen, R.K., Luoto, M., Araújo, M.B., Virkkala, R., Thuiller, W., Sykes, M.T., 2006. Methods and uncertainties in bioclimatic envelope modelling under climate change. Prog. Phys. Geogr. 30, 751-777.

Hijmans, R.J., Graham, C.H., 2006. The ability of climate envelope models to predict the effect of climate change on species distributions. Global Change Biol. 12, 2272-2281.

Hutchinson, G.E., 1957. Concluding remarks. Cold Springs Harb. Symp. Quant. Biol. 22, 415-427.

Ibáñez, I., Clark J.S., Dietze, M.C., Feeley, K., Hersh, M., LaDeau, S., McBride, A., Welch, N.E., Wolosin, M.S., 2005. Predicting biodiversity change: outside the climate envelope, beyond the species-area curve. Ecology. 87, 1896-1906.

Iverson, L.R., Prasad, A.M., 1998. Predicting abundance for 80 tree species following climate change in the eastern United States. Ecol. Monogr. 68, 465-485.

Jones, M.H., Macdonald, S.E., Henry, G.H.R., 1999. Sex- and habitat-specific responses of a high arctic willow, Salix arctica, to experimental climate change. Oikos. 87, 129-138.

Kevan, P.G., 1990. Sexual differences in temperature of blossoms on a dioecious plant, Salix arctica: significance for life in the arctic. Arctic Alpine Res. 22, 283-289.

Manel, S., Williams, H.C., Ormerod, S.J., 2001. Evaluating presence-absence models in ecology: the need to account for prevalence. J. Appl. Ecol. 38, 921-931.

McPherson, J.M., Jetz, W., Rogers, D.J., 2004. The effects of species' range size on the accuracy of distribution models: ecological phenomenon or statistical artifact? J. Appl. Ecol. 41, 811-823.

McKenney D., Papadopol, P., Campbell, K., Lawrence, K., Hutchinson, M., 2006. Spatial models of Canada- and North America-wide 1971-2000 minimum and maximum temperature, total precipitation and derived bioclimatic variables. Canadian Forest Service, Great Lakes Forestry Centre, Sault Ste. Marie, Ont. Tech Note 106.

McKenney, D.W., Pedlar, J.H., Lawrence, K., Campbell, K., Hutchinson. M.F., 2007. Potential impacts of climate change on the distribution of North American trees. Bioscience. 57, 939-948.

Meynard, C. N., Quinn, J.F., 2007. Predicting species distributions: a critical comparison of the most common statistical models using artificial species. J. Biogeogr. 34, 1455-1469.

Nix, H.A., 1986. A biogeographic analysis of Australian Elapid Snakes. Atlas of Elapid Snakes of Australia. Australian Flora and Fauna Series Number 7 (ed. R. Longmore), pp. 4-15. Australian Government Publishing Service, Canberra.

Oldham, M.J., Brinker, S.R., 2009. Rare Vascular Plants of Ontario, 4th edn. Natural Heritage Information Centre, Ontario Ministry of Natural Resources. Peterborough, Ontario.

Pearce, J., Ferrier, S., 2000. Evaluating the predictive performance of habitat models developed using logistic regression. Ecol. Model. 133, 225-245.

Pearson, R.G., Dawson, T.P., 2003. Predicting the impacts of climate change on the distribution of species: are bioclimate envelope models useful? Global Ecol. Biogeogr. 12, 361-371.

Pearson, R.G., Dawson, T.P., Berry, P.M., Harrison, P.A., 2002. SPECIES: a spatial evaluation of climate impact on the envelope of species. Ecol. Model. 154, 289-300.

Pearson, R.G., Thuiller, W., Araújo, M.B., Martinez-Meyer, E., Brotons, L., McClean, C., Miles, L., Seguardo, P., Dawson, T.P., Lees, D.C., 2006. Model-based uncertainty in species range prediction. J. Biogeogr. 33, 1704-1711.

Pennington, D., 2006. Representing the dimensions of an ecological niche. Proceedings 5th International Semantic Web Conference Workshop: Terra Cognita - Directions to the Geospatial Semantic Web, November 6, 2006, Athens, Georgia.

Peterson, A.T., Stewart, A., Mohamed, K.I., Araújo, M.B., 2008. Shifting global invasive potential of European plants with climate change. PLoS ONE. 3, e2441. doi:10.1371/journal.pone.0002441

Prasad, A.M., Iverson, L.R., Liaw, A., 2006. Newer classification and regression tree techniques: bagging and random forests for ecological prediction. Ecosystems. 9, 181-199.

Rehfeldt, G.E., Crookston, N.L., Warwell, M.V., Evans, J.S., 2006. Empirical analyses of plant-climate relationships for the western United States. Int. J. Plant Sci. 167, 1123-1150.

SAS Institute Inc., 2008. SAS 9.2 system for Unix. SAS Institute Inc., Cary, NC, USA.

Soper, J.H., Heimburger, M.L., 1982. Shrubs of Ontario. pp. 14-85. The Royal Ontario Museum, Toronto.

Stockwell, D., Peters, D., 1999. The GARP modelling system: problems and solutions to automated spatial prediction. Int. J. Geogr. Inf. Sci. 13, 143-158.

Stockwell, D.R.B., Townsend Peterson, A., 2002. Effects of size on accuracy of species distribution models. Ecol. Model. 148, 1-13.

Thuiller, W., Lavorel, S., Araújo, M.B., Sykes, M.T., Prentice, C.I., 2005. Climate change threats to plant biodiversity in Europe. Proc. Natl. Acad. Sci. 102, 8245-8250. doi:10.1073/pnas.0409902102

Townsend Peterson, A., Cohoon, K.P., 1999. Sensitivity of distributional prediction algorithms to geographic data completeness. Ecol. Model. 117, 159-164.

Trivedi, M.R., Berry, P.M., Morecroft, M.D., Dawson, T.P., 2008. Spatial scale affects bioclimate model projections of climate change impacts on mountain plants. Glob. Change Biol. 14, 1089-1103.

Chapter Three

Future distribution modelling: A stitch in time is not enough

Abstract

The last two decades have seen an increasing number of studies assessing the impact of climate change upon biodiversity. A central assumption underpinning research into the potential future habitat of terrestrial biota is that species are *presently* in equilibrium with their environments and that quantitative climate models *adequately* represent the distribution of species. Recently, many alarming predictions have emerged concerning the extinction and redistribution of species. Here, we show that even large-scale models of the climatic niche dimensions of species are temporally variable. Distributional models were developed for *Salix* (willow) species occurring in the province of Ontario, Canada, using three historical climate data sets. Although historical data very accurately represented the distributions of willows, the inherent variability within the models of species based on different periods greatly influenced the direction and magnitude of projected distributional change. We expose a fundamental uncertainty with respect to predicting the responses of species to climate change.

Introduction

It has long been recognized that organisms exist in environments peculiar to taxa (Darwin, 1859). This relationship forms the conceptual framework within which ecological niche theory has developed. Although formally articulated as the set of conditions in which a species can sustain itself *without immigration* (Grinnell, 1917), no formal representation of the niche concept has been achieved that is unanimously accepted (Pennington, 2006). This lack of consensus reflects the complexity underlying the systems of organization that exist in nature.

One branch of ecological research deeply rooted in the tenets of biogeography has recently gained public attention, becoming influential even in matters of global policy (Joyner *et al.*, 2010; Biggs *et al.*, 2008; Tokumine, 2002; Hales *et al.*, 2002). Ecological niche modelling can be used as a means of predicting the distribution of species from environmental data (Pearson and Dawson, 2003). A central premise of niche modelling is that, on large scales, climate has a primary influence over the terrestrial distribution of biota. Many studies using climate data to determine the niche dimensions of species have indeed been successful at predicting those species' distributions (Beever *et al.*, 2010; Joyner *et al.*, 2010; Stankowski and Parker, 2010; Peterson *et al.*, 2008; McKenney *et al.*, 2007; Hanann and Wang, 2006; Rehfeldt *et al.*, 2006; Calef *et al.*, 2005; Thuiller *et al.*, 2005; Burns *et al.*, 2003; Bakkenes *et al.*, 2002; Pearson *et al.*, 2002).

Growing concern over climate change and its potential consequences to biodiversity has prompted the rapid development of numerous analytical techniques to correlate *quantifiable* climate characteristics with the *known* location of species (Beever *et al.*, 2010; Feeley and Silman, 2010; Graham *et al.*, 2008; Loiselle *et al.*, 2008; Guisan *et al.*, 2007; Elith *et al.*, 2006; Heikkinen *et al.*, 2006). Scientific and technical advances of the twentieth century have permitted the creation of complex mathematical general circulation models (GCMs) that simulate global climate (Raper and Giorgi, 2005). Numerous GCMs have been developed worldwide and, coupled with different greenhouse gas emissions scenarios, used to predict potential future climate. The climatic niche dimensions of species can therefore be projected under different scenarios of climate change to identify potential future distributional areas.

Recently, many alarming predictions have been made ranging from the mass extinction of taxa over the next century (Malcolm *et al.*, 2006; Thuiller *et al.*, 2005; Thomas *et al.*, 2004;

Bakkenes *et al.*, 2002; Peterson *et al.*, 2002), to, at the very least, their partial redistribution (Peterson *et al.*, 2008; McKenney *et al.*, 2007; Hanann and Wang, 2006; Rehfeldt *et al.*, 2006; Calef *et al.*, 2005; Burns *et al.*, 2003; Pearson *et al.*, 2002; Iverson and Prasad, 1998). Regardless of consequence these predictions indicate a deep and gnawing uncertainty (Beaumont *et al.*, 2007; Hijmans and Graham, 2006; Pearson *et al.*, 2006). Should species have to relocate to survive? And if so, where might they go?

Even on global and continental scales there is little agreement over how best to represent the climatic dimensions of the niche of species. Pivotal questions remain unanswered: How few data are required to make accurate predictions? How can they be generalized? By what method will their utility be determined? The validity of choosing one analytical method over another remains problematic (Guisan *et al.*, 2007; Araújo and Guisan, 2006; Elith *et al.*, 2006; Heikkinen *et al.*, 2006; Pearson *et al.*, 2006; Anderson *et al.*, 2003; Loiselle *et al.*, 2003; Peterson and Cohoon, 1999), and largely depends upon the skill and knowledge of the modeller (Austin *et al.*, 2006; Austin, 2002). Here, we present qualitative and quantitative estimates of the climatic dimensions of the niches of willow species in Ontario using different historical data sets to demonstrate that an assumption of equilibrium in a given period greatly affects the direction and magnitude of projected distributional change.

Model Details

We developed distributional models for 24 provincially occurring *Salix* (willow) species. Data for willow distributions were available in the form of range maps (Argus, 2007). While the use of continuous ranges may provide an overly optimistic estimate of the occurrence of willows, models based solely on collection points were generally found to be inadequate (Stankowski and Parker, 2010). To examine the historical variability of the climatic niche dimensions of willows,

we used three data sets that have formed the basis for projected estimates of the impacts of climate change on the distribution of species: 1931–1960 (Bakkenes *et al.*, 2002), 1961–1990 (Peterson *et al.*, 2008; Hamann and Wang, 2006; Rehfeldt *et al.*, 2006; Thuiller *et al.*, 2005; Thomas *et al.*, 2004; Pearson *et al.*, 2002), and 1971–2000 (McKenney *et al.*, 2007). Each data set was comprised of 36 provincial grids with cell dimensions of 0.083° (0.05'), representing monthly maximum and minimum temperature (°C) and average precipitation (mm).

Logistic regression was used to determine the relation between the geographic distribution of the 24 willow species and the 36 climate variables using a procedure described in detail elsewhere (Stankowski and Parker, 2010). Although potentially not as robust as other novel approaches (Guisan *et al.*, 2007; Elith *et al.*, 2006; Heikkinen *et al.*, 2006), this method has been widely reviewed for distributional modelling studies (Calef *et al.*, 2005; Bakkenes *et al.*, 2002; Stockwell and Peterson, 2002) and many of its limitations are well understood (Meynard and Quinn, 2007; McPherson *et al.*, 2004; Manel *et al.*, 2001; Cumming, 2000; Pearce and Ferrier, 2000). We use it here as a generic example to discuss conceptual difficulties that are independent of analytical method. Models for each willow species in each climate period were determined (SAS, 2008) and used to create maps (ESRI, 2006) of probable provincial occurrence. For the purposes of this study, forward stepwise selections of 6 monthly climate variables were sufficient to model willow distributions.

Historical Variability

The models developed for each historical period accurately represented the known geographic distribution of willow species (A, B, and C of Figs. 1 – 6). However, there was little similarity in the selections of climate predictors within the distributional models of species (Table 1). Models based on 1931–1960 data were most dissimilar when compared to those based

on the two more recent periods. Four of the 24 models shared only one predictor with the models

corresponding to the more recent two time periods. A further four models shared none. Even

between the models based on 1961–1990 and 1971–2000 data sets, there was surprisingly little

overlap. Four models shared four predictors, and an additional four shared three.

Of the variables common among the time periods, few shared rank; *i.e.,* had the same relative

strength of a predictor within a given model. There was only a 15% similarity (21 variable

matches of 144 possibilities) between the models of species in 1961–1990 and 1971–2000 space;

5% between 1931–1960 and 1961–1990; and 4% between 1931–1960 and 1971–2000. The

variation between models was not uniformly distributed among species. Of the 24 species

considered in this analysis, only four shared the first predictor in their distribution models across

all three periods, and only one of these shared the first two (see supplementary information).

Table 1. Similarity within models for 24 *Salix* species in three periods

Climate period comparisons	1931-1960 1971-2000	1931-1960 1961-1990	1961-1990 1971-2000	All three periods
Shared predictors	21 (15%)	23 (16%)	42 (29%)	11 (8%)
Predictors sharing rank	10 (7%)	11 (8%)	25 (17%)	5 (3%)
Primary sequential predictors*	6 (4%)	7 (5%)	21 (15%)	5 (3%)

*See supplementary information

Despite differences in selections of predictors, the distributional models in each of the three

periods reflected strong historical climatic patterns. The distributions of several arctic species

were predicted by increasingly earlier monthly maximum temperatures, shifting from July during

the period 1931–1960, to June in 1961–1990, and May in 1971–2000. June maximum

temperature was the strongest predictor in 11 of 14 models for these high-latitude dwelling dwarf

shrubs based on the two more recent periods (1961–1990 and 1971–2000). Distributions for two

common and widespread boreal species revealed patterns of increasingly late winter minimum temperatures, shifting from November in 1931–1960, to December in 1961–1990, January in 1971–2000. September and January minimum temperatures respectively predicted the distributions of the tree species, which occur mainly along the southern provincial boundaries, in the 1931–1960 and 1971–2000 periods. However, different random samples from the original distributional data for *S. amygdaloides* resulted in alternate selections of these predictors in models for the interim period of 1961–1990 suggesting multicollinearity.

Projected Equilibrium

To test the assumption of historic species/climate equilibrium, each model was projected under conditions of *expected* future climate change using the output from four general circulation models (GCMs) for three periods (2011–2040, 2041–2070, and 2071–2100) under three scenarios (A2, B1, and A1B). The three scenarios follow the storylines laid out by the Intergovernmental Panel on Climate Change's (IPCC) fourth Assessment Report (AR4) and can be roughly summarized as follows. The A2 scenario assumes a heterogeneous world with a continuously increasing global population and a regionally oriented and somewhat fragmented economic growth. It is expected in this scenario that carbon dioxide (CO_2) emission levels will be maximal. In the B1 scenario, the global population is expected to peak sometime mid-century declining thereafter. Rapid changes in economic structures towards a service and information economy with reductions in materials intensity coupled with the introduction of clean and resource efficient technologies are expected to stabilize atmospheric CO_2 at 550 ppm. The A1B scenario is used as a central case, assuming little change in the growth rates of future emissions.

High-resolution interpolations of the IPCC AR4 GCMs for Canada were provided by the Canadian Forest Service (Price *et al.*, 2010 in review). These GCM outputs originated from the

Program for Climate Model Diagnosis and Intercomparison (PCMDI) Third Climate Model Intercomparison Project (CMIP3). The four GCMs were selected on the basis of data available in 2008 and because they were well recognized within the global GCM community. They were: the Canadian Centre for Climate Modelling and Analysis (CCCma) Third Generation Coupled Global Climate Model Version 3.1 Medium Resolution (CGCM31MR); the Australian Commonwealth Scientific and Industrial Research Organisation Mark 3.5 Climate System Model (CSIROMK35); the Japanese Centre for Climate System Research (CCSR) Model for Interdisciplinary Research on Climate Version 3.2 Medium Resolution (MIROC32MR); and the U.S. National Center for Atmospheric Research (NCAR) Community Climate Model version 3.0 (NCARCCSM3).

Over 1000 probability maps were generated in this analysis. For each species, 108 *expected* future distributions were examined. Regardless of consequence, there was little agreement among the CGM projections concerning the magnitude and direction of expected future change. Generally, those models using the NCAR data contrasted with one another among the three time periods (compare A4, B4 and C4 in each of Figs. 1 – 6) and with those of the other GCMs (compare A4 to A1, A2 and A3, etc.). However, these differences were expressed *even among models based on overlapping periods sharing the strongest climatic predictors*.

Interestingly, within a given GCM and scenario combination, very little change was observed in the projected distributions of species across future time periods (i.e., the projections of a given species' historical model for 2011-2040 resembled those for 2041-2070 and 2071-2100). For illustrative purposes, we present only the partial results (12 potential futures out of 108 generated possibilities) using only the GCM output for 2041–2070 (A1B scenario) for six willow species with contrasting distributions (Table 2). Two of these are widespread and common: *S. discolor*

and *S. amygdaloides*, while the other four are considered provincially rare and vulnerable: *S. arctica*, *S. maccalliana*, *S. myricoides* and *S. pseudomonticola* (Oldham and Brinker, 2009).

Table 2. Partial predictive results for the models of 6 *Salix* species in three periods

Salix species	Prevalence	Climate Period		
		1931-1960	1961-1990	1971-2000
*S. arctica**	2.1%	June Prec.	June Max. T.	June Max. T.
Mean	*n* = 267	55.7 mm	13.1 °C	12.8 °C
Std.dev.		1.7 mm	0.4 °C	0.5 °C
Avg. *P*		0.81	0.85	0.86
*S. maccalliana**	36.9%	Jan. Min. T.	Feb. Min. T.	July Prec.
Mean	*n* = 4706	-27.0 °C	-26.4 °C	97.8 mm
Std.dev.		1.3 °C	1.5 °C	4.5 mm
Avg. *P*		0.86	0.84	0.84
*S. myricoides**	28.7%	May Prec.	Nov. Prec.	Aug. Prec.
Mean	*n* = 3617	67.8 mm	62.4 mm	82.2 mm
Std.dev.		7.0 mm	18.3 mm	6.7 mm
Avg. *P*		0.79	0.59	0.66
*S. pseudomonticola**	35.7%	July Min. T.	July Min. T.	July Min. T.
Mean	*n* = 4503	9.8 °C	8.9 °C	9.1 °C
Std.dev.		0.9 °C	1.5 °C	1.4 °C
Avg. *P*		0.83	0.84	0.87
S. discolor	56.5%	May Prec.	May Prec.	May Prec.
Mean	*n* = 7116	68.3 mm	66.7 mm	68.1 mm
Std.dev.		5.9 mm	9.9 mm	10.3 mm
Avg. *P*		0.94	0.95	0.94
S. amygdaloides	12.3%	Sept. Min. T.	Sept. Min. T.	Jan. Min. T.
Mean	*n* = 1542	8.4 °C	8.5 °C	-14.8 °C
Std.dev.		1.8 °C	1.8 °C	4.6 °C
Avg. *P*		0.88	0.91	0.88

Average prevalence (%), number of cells present (*n*), strongest distributional predictor means and standard deviations, and the average probability of occurrence across Ontario in the final regression models; *denotes provincially rare and vulnerable species

In addition to their contrasting distributions, these species contrasting characteristically in terms of habit, habitat, genetic variability, and value. The A1B scenario for the period 2041-2070 was chosen to illustrate a central case assuming little change in the *status quo*. This sample of the results, albeit small, is nonetheless representative of the general patterns observed for *all* species under the different GCM/future period/scenario combinations.

June precipitation was the strongest predictor in the model of *S. arctica* based on the 1931–1960 data (Fig. 1. A). The projections for this model show a strong similarity in their northwestern patterns of redistribution (Figs. 1. A1 – A3) with the exception of the NCAR output (Fig. 1. A4). June maximum temperature was the strongest predictor in the models based on 1961–1990 and 1971–2000 data (Figs. 1. B, C). However, the projections of these models for the two most recent time periods were completely dissimilar. Projections of no suitable provincial climate (Figs. 1. C1, C3) contrast with easterly and westerly redistributions (Figs. 1. B1, B3) and unlikely southerly pockets (Figs. 1. B4, C4).
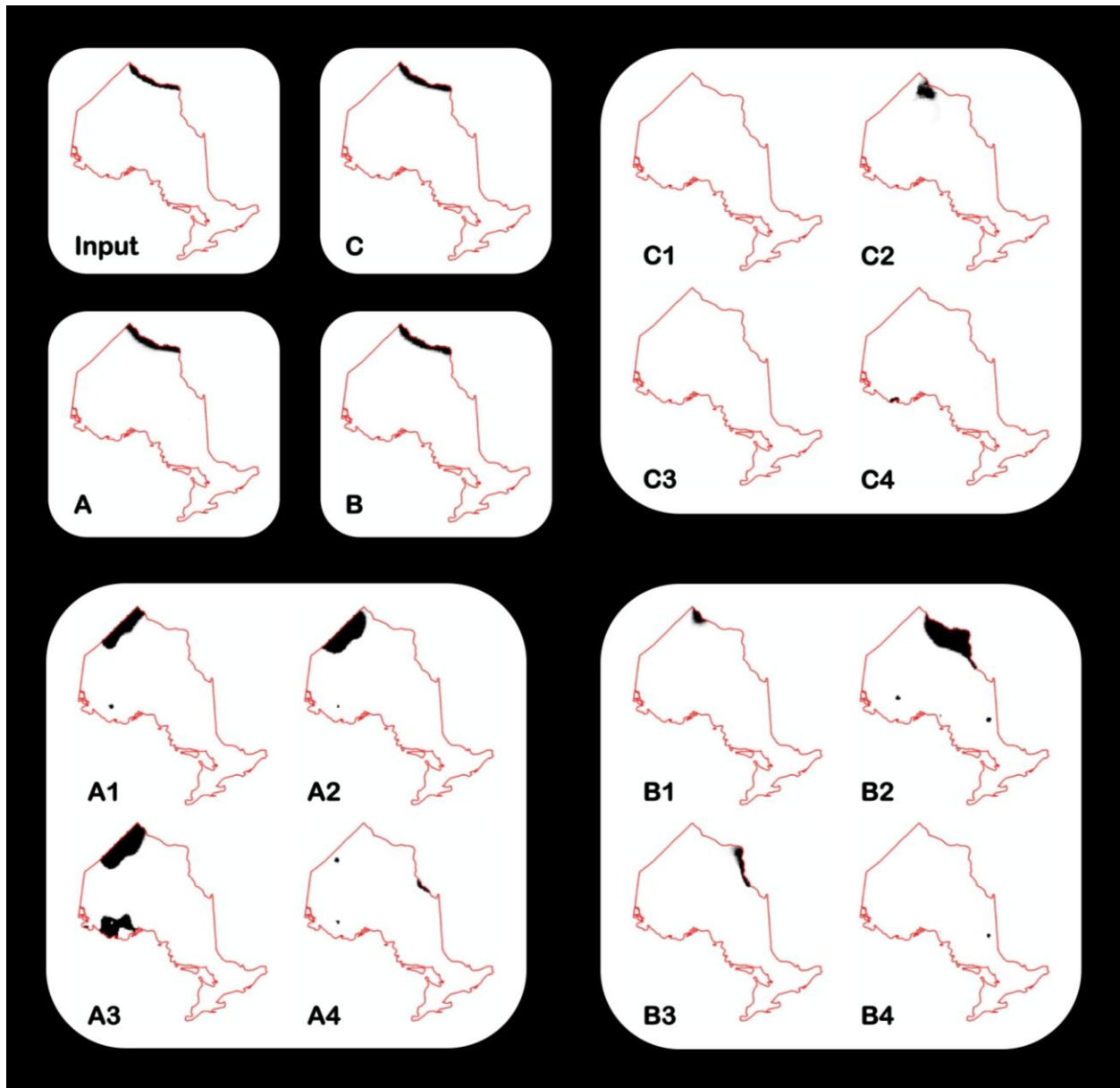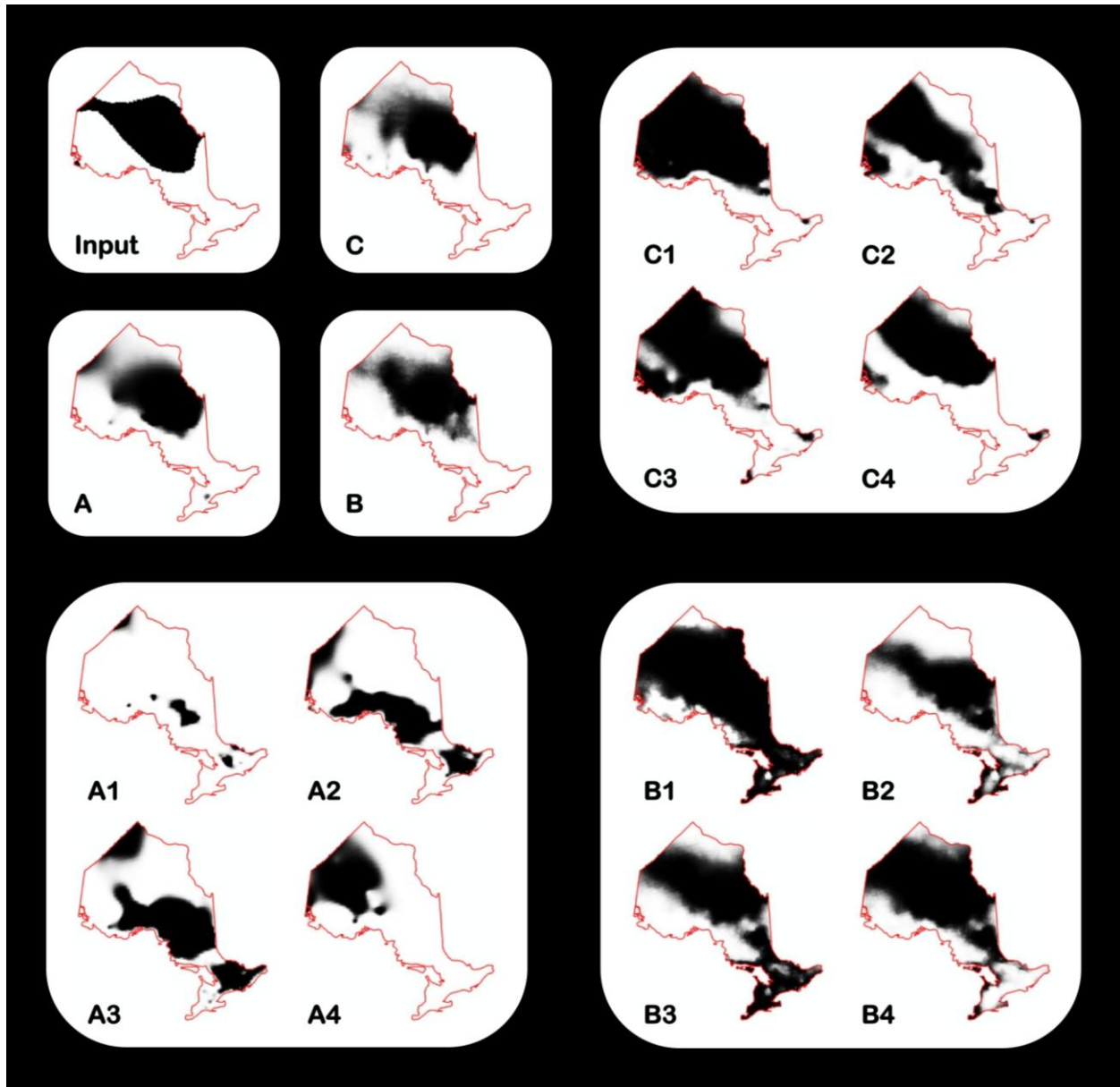
Fig. 1. *S. arctica* Ontario distribution (Input) modelled in three periods: (A) 1931-1960, (B) 1961-1990, and (C) 1971-2000; and projections for the period 2041-2070 (A1B scenario) using four GCM outputs: (1) CGCM31, (2) CSIROmk35, (3) MIROC32mr, and (4) NCARccsm

The models of *S. maccalliana* showed little agreement among selected variables, with winter minimum temperatures selected for the earlier periods (Figs. 2. A, B) in contrast to summer precipitation selected for the most recent period (Fig. 2. C). The 1931–1960 models of *S. maccalliana* project major range contractions (Fig. 2. A1) and multi-directional redistributions

(Figs. 2. A2 – A4). Despite the fact that the models were based on completely different sets of predictors, some consistency between the projections based on the two overlapping periods was observed (compare Figs. 2. B1 – B4, and C1 – C4).

Fig. 2. *S. maccalliana* Ontario distribution (Input) modelled in three periods: (A) 1931-1960, (B) 1961-1990, and (C) 1971-2000; and projections for the period 2041-2070 (A1B scenario) using four GCM outputs: (1) CGCM31, (2) CSIROmk35, (3) MIROC32mr, and (4) NCARccsm

The models of *S. myricoides*, which were among the weakest in terms of probability levels

(0.59 – 0.79), nonetheless reflected strong patterns of spring and summer precipitation variables

(Figs. 3. A – C). However, the future distributions of *S. myricoides* are extremely variable. The

1931–1960 model based on May precipitation projects major redistributions at odds with one

another (Figs. 3. A1 – A4). The projections of the 1961–1990 and 1971–2000 models, based on

November and August precipitation respectively, with the exception of those based on the

CSIRO output (Figs. 3. B2, C2), similarly contrast (Figs. 3. B1, B3, B4 and C1, C3, C4).
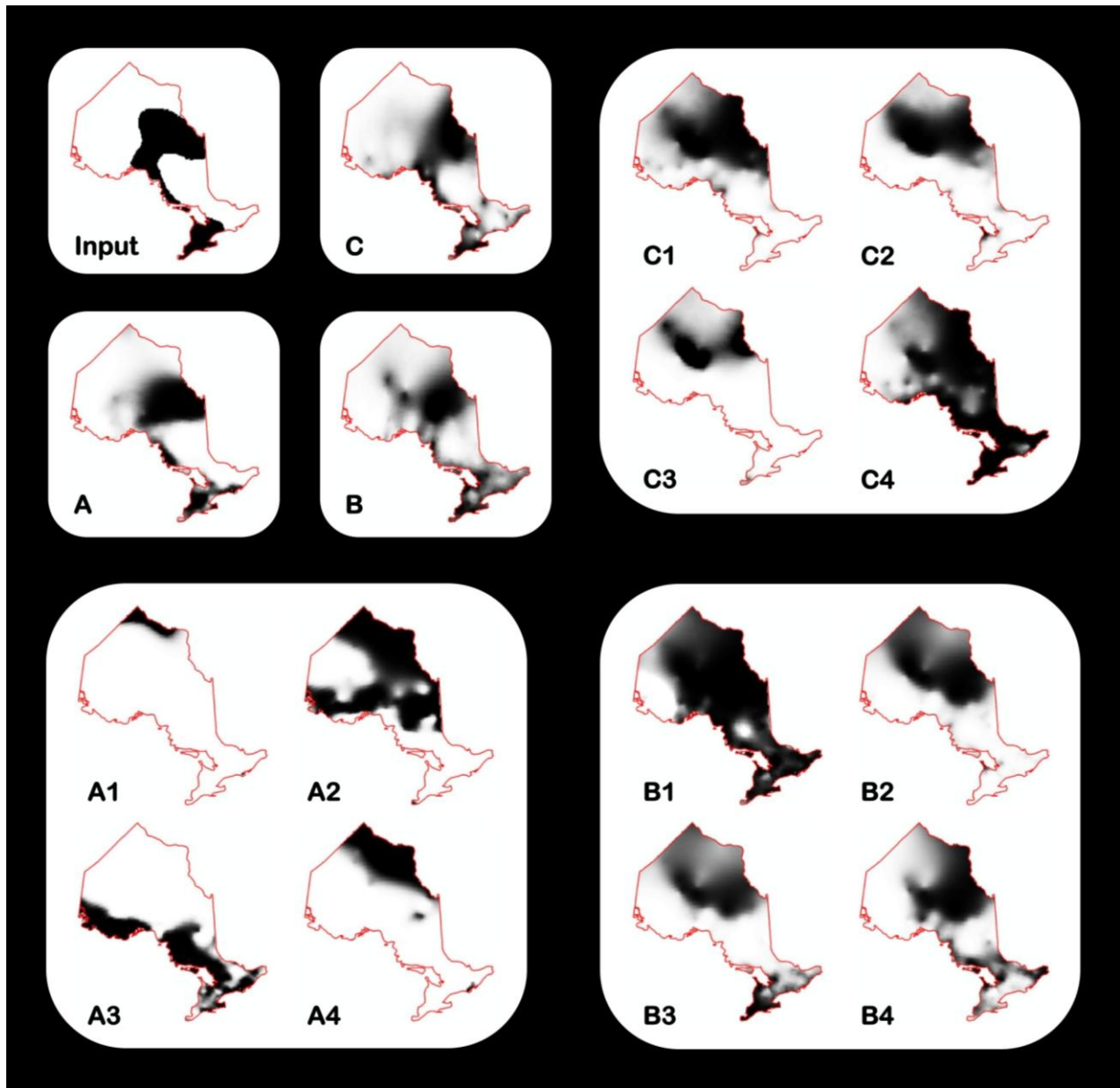
Fig. 3. *S. myricoides* Ontario distribution (Input) modelled in three periods: (A) 1931-1960, (B) 1961-1990, and (C) 1971-2000; and projections for the period 2041-2070 (A1B scenario) using four GCM outputs: (1) CGCM31, (2) CSIROmk35, (3) MIROC32mr, and (4) NCARccsm

July minimum temperature was the strongest predictor in all three periods for models of *S. pseudomonticola* (Fig. 4. A – C). Projections of the 1931–1960 model for *S. pseudomonticola* indicate somewhat similar range contractions (Figs. 4. A1 – A3), with the exception of the NCAR output showing range retention in the northeast (Fig. 4. A4). However, even for the

1961–1990 and 1971–2000 models, that shared the first four predictors in sequence making them the most similar of all species considered in this study, the projected distributions greatly contrast. The 1961–1990 projections using the CGCM and MIROC output (Figs. 4. B1, B3) show little change, while those using the CSIRO and NCAR output indicate major range contractions towards northern and southern coastal areas (Figs. 4. B2, B4). The 1971–2000 projections, using the CGCM and MIROC output, show substantial northwestern and southwestern range expansions (Figs. 4. C1, C3). The projection based on the CSIRO output indicates a dramatic reduction towards the provincial interior (Fig. 4. C2), while that using the NCAR output shows little future change (Fig. 4. C4).
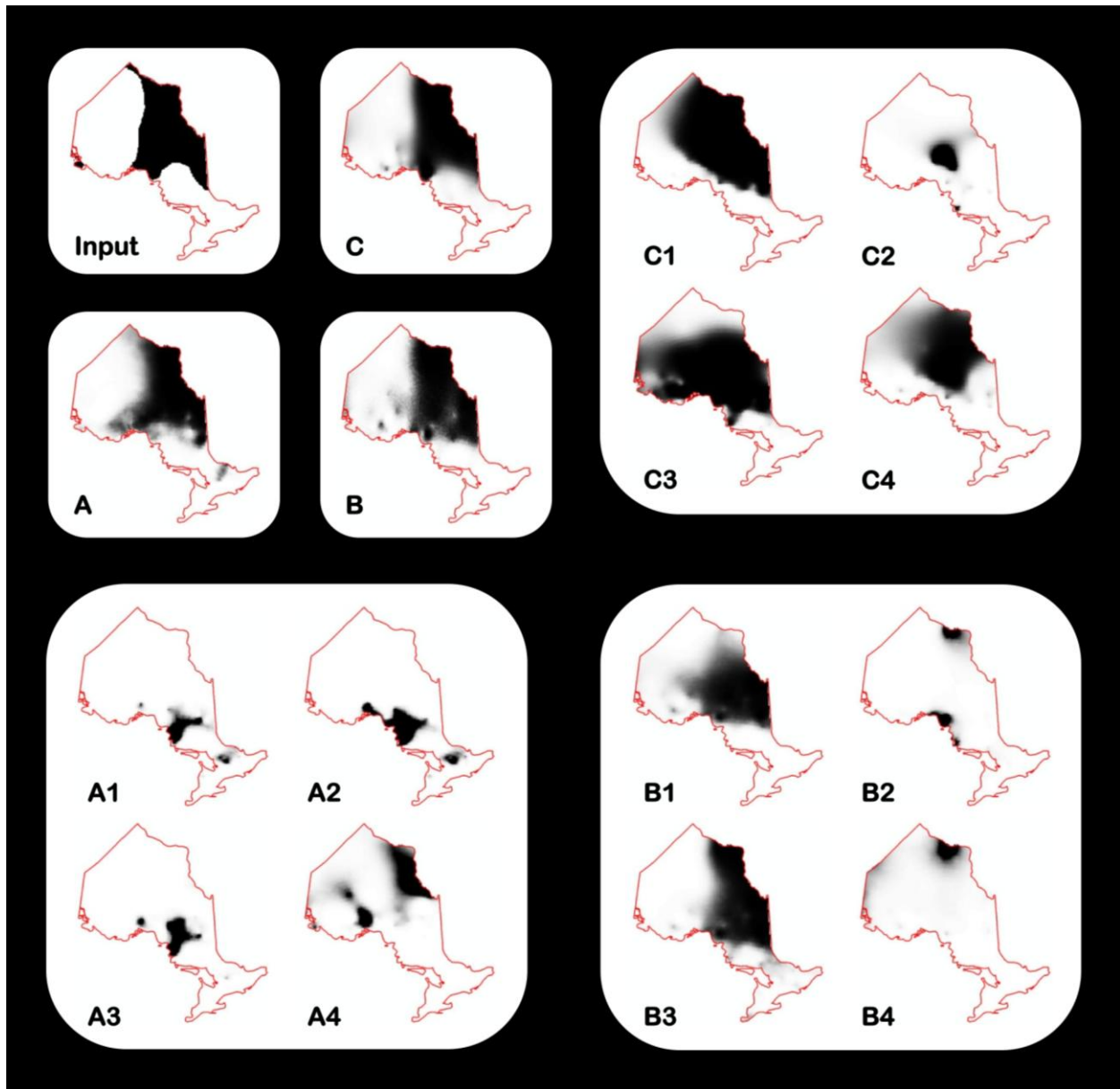
Fig. 4. *S. pseudomonticola* Ontario distribution (Input) modelled in three periods: (A) 1931-1960, (B) 1961-1990, and (C) 1971-2000; and projections for the period 2041-2070 (A1B scenario) using four GCM outputs: (1) CGCM31, (2) CSIROmk35, (3) MIROC32mr, and (4) NCARccsm

May precipitation was the strongest predictor in all three periods for models of *S. discolor* (Figs. 5. A – C). In contrast with the results for other willow species, the projected distributions of *S. discolor* indicate nearly no change in future climatic niche space. However, projections for

the models based on 1961–1990 and 1971–2000 data using the NCAR output indicate range

contractions (Figs. 5. B4, C4).
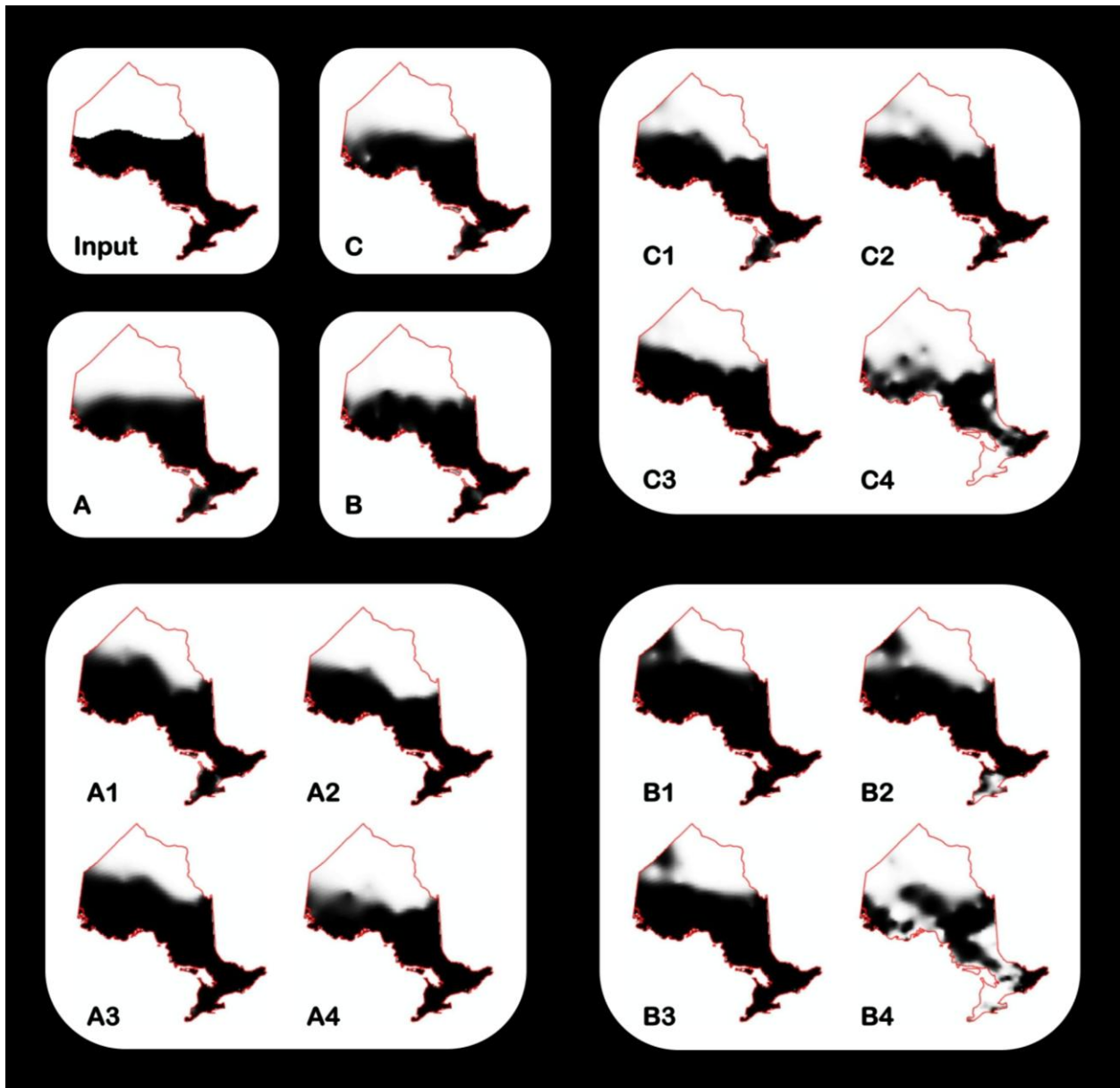
Fig. 5. *S. discolor* Ontario distribution (Input) modelled in three periods: (A) 1931-1960, (B) 1961-1990, and (C) 1971-2000; and projections for the period 2041-2070 (A1B scenario) using four GCM outputs: (1) CGCM31, (2) CSIROmk35, (3) MIROC32mr, and (4) NCARccsm

September minimum temperature was the strongest predictor in the models based on 1931–

1960 and 1961–1990 data for *S. amygdaloides* (Figs. 6. A, B) with a shift to January minimum

temperature in the more recent period (Fig. 6. C). The projections of these models were very similar in that three indicate northward range shifts, while one (NCAR) shows major shrinkage (Figs. 6. A1 – A4 and B1 – B4). In contrast, projections of the 1971–2000 models indicate minor southerly range contractions (Figs. 6. C1 – C4).
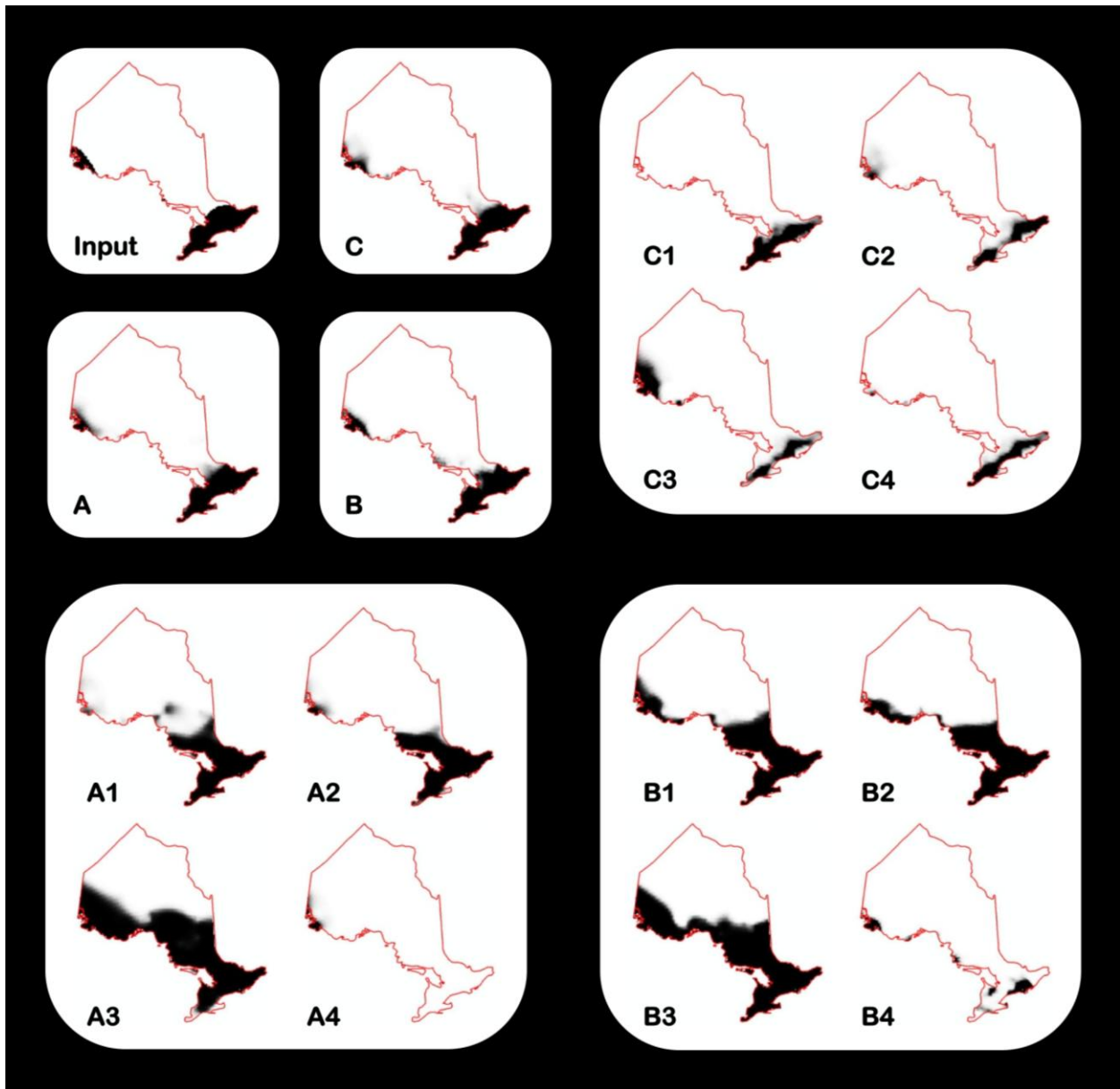
Fig. 6. *S. amygdaloides* Ontario distribution (Input) modelled in three periods: (A) 1931-1960, (B) 1961-1990, and (C) 1971-2000; and projections for the period 2041-2070 (A1B scenario) using four GCM outputs: (1) CGCM31, (2) CSIROmk35, (3) MIROC32mr, and (4) NCARccsm

Discussion

Prior to this study, issues relating to the use of different historical periods of climate data for future distributional modelling had not been considered. Our results show that the climatic dimensions of the ecological niche of willow species in Ontario are variable qualitatively and quantitatively through time. Climatic tolerances determined for the distribution of species in one period do not necessarily apply to other, *even overlapping*, periods.

Many studies have shown that estimates of the potential future distributions of species are highly variable depending on the GCM employed (Joyner *et al.*, 2010; Beaumont *et al.*, 2007; Araújo and Guisan, 2006; Hijmans and Graham, 2006; Pearson *et al.*, 2006; Peterson *et al.*, 2002). Our results compound this uncertainty, implying that many of the recent predictions of the potential consequences of climate change to species (Peterson *et al.*, 2008; McKenney *et al.*, 2007; Hanann and Wang, 2006; Malcolm *et al.*, 2006; Rehfeldt *et al.*, 2006; Calef *et al.*, 2005; Thuiller *et al.*, 2005; Thomas *et al.*, 2004; Burns *et al.*, 2003; Bakkenes *et al.*, 2002; Pearson *et al.*, 2002; Peterson *et al.*, 2002; Iverson and Prasad, 1998) are far too simplistic. In our analysis, the projected distributions of willows were dissimilar even for models based on similar selections of predictor variables in overlapping climate periods. Notable exceptions were the models for *S. discolor* whose projections indicate nearly no change across the future periods. Even so, these results for Ontario contrast sharply with the findings of Hamann and Wang (2006) who predicted a 64% habitat loss over the same period of this common and widespread species in the province of British Columbia, Canada.

If an estimate of climatic equilibrium is based on the analysis of a single period, the direction and magnitude of prediction can be misleading. The dissimilarity evident in the projections of different historical models within species complicates the development of any successful

conservation strategy. The questions remain: Should the assumption of species/climate equilibrium form the basis for projected estimates of distributional change? And if so, over which historical period should this climatic equilibrium apply?

The strong correlation we observed between historical climate and the distributions of willows suggests the potential use of these models as tools to develop our understanding of the niches of individual species. Beyond the uncertainty arising from the use of different GCM projections, issues surrounding the species/climate relationship need attention. Previously, we have shown that even on large scales, the distributions of species are most accurately predicted in climatic terms using data of the highest available resolution both spatially and temporally (Stankowski and Parker, 2010). However, of what spatial and temporal relevance to a given species are the data used to model its distribution? Any attempt to determine climatic niche dimensions should include information concerning the longevity and lifecycles of species; *i.e.*, intuitively, climate data should be sampled over equal intervals and along gradients relative to the lifespan of species. Establishing this relational basis in both *time* and *space* may simultaneously improve the accuracy of our predictions and develop our concept of niche theory.

Supplementary Information is appended.

Acknowledgements

References

Anderson, R.P., Lew, D., Townsend Peterson, A., 2003. Evaluating predictive models of species' distributions: criteria for selecting optimal models. Ecol. Model. 162, 211–232.

Araújo, M.B., Guisan, A., 2006. Five (or so) challenges for species distribution modelling. J. Biogeogr. 33, 1677–1688.

Argus, G.W., 2007. Salix (Salicaceae) Distribution maps and a synopsis of their classification in North America, north of Mexico. Harvard Papers in Botany 12, 335–368.

Austin, M.P., 2002. Spatial prediction of species distribution: An interface between ecological theory and statistical modelling. Ecol. Model. 157, 101–118.

Austin, M.P., Belbin, L., Meyers, J.A., Doherty, M.D., Luoto, M., 2006. Evaluation of statistical models used for predicting plant species distributions: role of artificial data and theory. Ecol. Model. 199, 197–216.

Bakkenes, M., Alkemade, J.R.M., Ihle, F., Leemans, R., Latour, J.B., 2002. Assessing effects of forecasted climate change on the diversity and distribution of European higher plants for 2050. Glob. Change Biol. 8, 390–407.

Beaumont, L.J., Pitman, A.J., Poulsen, M., Hughes, L., 2007. Where will species go? Incorporating new advances in climate modelling into projections of species distributions. Glob. Change Biol. 13, 1368–1385.

Beever, E.A., Ray, C., Mote, P.W., Wilkening, J.L., 2010 Testing alternative models of climate–mediated extirpations. Ecol. Appl. 20, 164–178.

Biggs, R., Simons, H., Bakkenes, M., Scholes, R.J., Eickhout, B., Vuuren, van D., Alkemande, R., 2008. Scenarios of biodiversity loss in southern Africa in the 21st century. Global Environ. Chang. 18, 296-309.

Burns, C.E., Johnston, K.M., Schmitz, O.J., 2003. Global climate change and mammalian species diversity in U.S. national parks. Proc. Natl Acad. Sci. USA 100, 11474–11477.

Calef, M.P., McGuire, A.D., Epstein, H.E., Rupp, T.S., Shugart, H.H., 2005. Analysis of vegetation distribution in Interior Alaska and sensitivity to climate change using a logistic regression approach. J. Biogeogr. 32, 863–878.

Cumming, G.S., 2000. Using between-model comparisons to fine–tune linear models of species ranges. J. Biogeogr. 27, 441–455.

Darwin, C.R., 1859. On the Origin of Species by Means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life (John Murray, London).

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.McC., Peterson, A.T., Phillips, S.J., Richardson, K., Scachetti–Pereira, R., Schapire, R.E., Soberón, J., Williams, S., Wisz, M.S., Zimmermann, N.E., 2006. Novel methods improve prediction of species' distribution from occurrence data. Ecography 29, 129–151.

ESRI., 2006. ArcGIS Desktop 9.2. Environmental Systems Research Institute Inc., Redlands, CA.

Feeley, K.J., Silman, M.R., 2010. Modelling the responses of Andean and Amazonian plant species to climate change: the effects of georeferencing errors and the importance of data filtering. J. Biogeogr. 37, 733–740.

Graham, C.H., Elith, J., Hijmans, R.J., Guisan, A., Peterson, A.T., Loiselle, B.A., and the NCEAS Predicting Species Distributions Working Group, 2008. The influence of spatial errors in species occurrence data using in distribution models. J. Appl. Ecol. 45, 239–247.

Grinnell, J., 1917. The niche–relationships of the California thrasher. Auk 34, 427–433.

Guisan, A., Graham, C.H., Elith, J., Huettmann, F., and the NCEAS Species Distribution Modelling Group, 2007. Sensitivity of predictive species distribution models to change in grain size. Divers. Distrib. 13, 332–340.

Hales, S., de Wet, N., Maindonald, J., Woodward, A., 2002. Potential effect of population and climate changes on global distribution of dengue fever: an empirical model. Lancet http://image.thelancet.com/extras/01art11175web.pdf.

Hamann, A., Wang, T., 2006. Potential effects of climate change on ecosystem and tree species distribution in British Columbia. Ecology 87, 2773–2786.

Heikkinen, R.K., Luoto, M., Araújo, M.B., Virkkala, R., Thuiller, W., Sykes, M.T., 2006. Methods and uncertainties in bioclimatic envelope modelling under climate change. Prog. Phys. Geogr. 30, 751–777.

Hijmans, R.J., Graham, C.H., 2006. The ability of climate envelope models to predict the effect of climate change on species distributions. Glob. Change Biol. 12, 2272–2281.

Iverson, L.R., Prasad, A.M., 1998. Predicting abundance for 80 tree species following climate change in the Eastern United States. Ecol. Monogr. 68, 465–485.

Joyner, T.A., Lukhnova, L., Pazilov, Y., Temiralyeva, G., Hugh–Jones, M.E., Aikimbayev, A., Blackburn, J.K., 2010. Modeling the potential distribution of Bacillus anthracis under multiple climate change scenarios for Kazakhstan. PLoS ONE 5, e9596. doi:10.1371/journal.pone.0009596.

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

Loiselle, B.A., Jørgensen, P.M., Consiglio, T., Jiménez, I., Blake, J.G., Lohmann, L.G., Montiel, O.M., 2008. Predicting species distributions from herbarium collections: does climate bias in collection sampling influence model outcomes? J. Biogeogr. 35, 105–116.

Malcolm, J.R., Liu, C., Neilson, R.P., Hansen, L., Hannah, L., 2006. Global warming and extinctions of endemic species from biodiversity hotspots. Conserv. Biol. 20, 538–548.

Manel, S., Williams, H.C., Ormerod, S.J., 2001. Evaluating presence-absence models in ecology: the need to account for prevalence. J. Appl. Ecol. 38, 921–931.

McKenney, D.W., Pedlar, J.H., Lawrence, K., Campbell, K., Hutchinson, M.F., 2007. Potential impacts of climate change on the distribution of North American trees. Bioscience 57, 939–948.

McPherson, J.M., Jetz, W., Rogers, D.J., 2004. The effects of species' range size on the accuracy of distribution models: ecological phenomenon or statistical artifact? J. Appl. Ecol. 41, 811–823.

Meynard, C. N., Quinn, J.F., 2007. Predicting species distributions: a critical comparison of the most common statistical models using artificial species. J. Biogeogr. 34, 1455–1469.

Oldham, M.J., Brinker, S.R., 2009. Rare vascular plants of Ontario, 4th ed. (Natural Heritage Information Centre, Ontario Ministry of Natural Resources, Peterborough, Ontario).

Pearce, J., Ferrier, S., 2000. Evaluating the predictive performance of habitat models developed using logistic regression. Ecol. Model. 133, 225–245.

Pearson, R.G., Thuiller, W., Araújo, M.B., Martinez–Meyer, E., Brotons, L., McClean, C., Miles, L., Seguardo, P., Dawson, T.P., Lees, D.C., 2006. Model–based uncertainty in species range prediction. J. Biogeogr. 33, 1704–1711.

Pearson, R.G., Dawson, T.P., 2003. Predicting the impacts of climate change on the distribution of species: are bioclimate envelope models useful? Global Ecol. Biogeogr. 12, 361-371.

Pearson, R.G., Dawson, T.P., Berry, P.M., Harrison, P.A., 2002. SPECIES: a spatial evaluation of climate impact on the envelope of species. Ecol. Model. 154, 289–300.

Pennington, D., 2006. Representing the dimensions of an ecological niche. Proceedings 5th International Semantic Web, November 6, 2006, Athens, Georgia.

Peterson, A.T., Cohoon, K.P., 1999. Sensitivity of distributional prediction algorithms to geographic completeness. Ecol. Model. 117, 159–164.

Peterson, A.T., Ortega–Huerta, M.A., Bartley, J., Sánchez–Cordero, V., Soberón, J., Buddemeier, R.H., Stockwell, D.R.B., 2002. Future projections for Mexican faunas under global climate change scenarios. Nature 416, 626–629.

Peterson, A.T., Stewart, A., Mohamed, K.I., Araújo, M.B., 2008. Shifting global invasive potential of European plants with climate change. PLoS ONE 3, e2441. doi:10.1371/journal.pone.0002441.

Price, D., McKenney, D., Joyce, L., Siltanen, M., Papadopol, P., Lawrence, K., 2010. High resolution interpolation of IPCC AR4 GCMs for Canada. Canadian Forest Service Report, Edmonton, in review.

Raper, S.C.B., Giorgi, F., 2005. Climate change projections and models *in* Climate Change and Biodiversity. Yale University Press, New Haven. pp. 199-210.

Rehfeldt, G.E., Crookston, N.L., Warwell, M.V., Evans, J.S., 2006. Empirical analyses of plant-climate relationships for the western United States. Int. J. Plant Sci. 167, 1123–1150.

SAS Institute Inc., 2008. SAS 9.2 system for Unix. SAS Institute Inc., Cary, NC, USA.

Stankowski, P.A., Parker, W.H., 2010. Species distribution modelling: does one size fit all? A phytogeographic analysis of Salix in Ontario. Ecol. Model. 221, 1655–1664.

Stockwell, D.R.B., Townsend Peterson, A., 2002. Effects of sample size on accuracy of species distribution models. Ecol. Model. 148, 1–13.

Thomas, C.D., Cameron, A., Green, R.E., Bakkenes, M., Beaumont, L.J., Collingham, Y.C., Erasmus, B.F.N., de Siqueria, M.F., Grainger, A., Hannah, L., Hughes, L., Huntley, B., van Jaarsveld, A.S., Midgley, G.F., Miles, L., Ortega–Huerta, M.A., Peterson, A.T., Phillips, O.L., Williams, S.E., 2004. Extinction risk from climate change. Nature 427, 145–148.

Thuiller, W., Lavorel, S., Araújo, M.B., Sykes, M.T., Prentice, C., 2005. Climate change threats to plant diversity in Europe. Proc. Natl Acad. Sci. USA 102, 8245–8250.

Tokumine, S., 2002. Bioclimatic modelling of sub-Saharan African plants. Progress report to Conservation International, September 2002. Centre for Ecology Law and Policy, Environmental Department, University of York, UK.

Supplementary Information

Table S1. Variable comparisons within models for 24 *Salix* species in three periods

Salix spp.		1931-1960	1961-1990	1971-2000	*Salix* spp.		1931-1960	1961-1990	1971-2000	*Salix* spp.		1931-1960	1961-1990	1971-2000
reticulata	1	junprc	junmxt	junmxt	*cordata*	1	junprc	novmit	novmit	*pellita*	1	novmit	decmit	janmit
	2	octprc	mayprc	mayprc		2	novmit	aprmit	maymxt		2	mayprc	octprc	sepprc
	3	febprc	marprc	febmit		3	maymxt	augmxt	sepmxt		3	novmxt	febmxt	febmit
	4	marprc	febprc	maymit		4	sepmit	janmxt	maymit		4	febmxt	aprprc	octmxt
	5	decmit	novmit			5	sepmit	febprc	aprmit		5	augmit	octmxt	augprc
	6		janmxt			6	decprc	octprc	aprmxt		6	marprc	julmit	julprc
arctophila	1	julmxt	junmxt	junmxt	*humilis*	1	maymxt	maymxt	maymxt	*pseudomonticola*	1	julmit	julmit	julmit
	2	augmxt	augmxt	novprc		2	novprc	augprc	augprc		2	decprc	junmit	junmit
	3	sepmxt	maymit	febmit		3	marmxt	sepprc	sepprc		3	janmit	febmit	febmit
	4	julprc	decmit	janprc		4	octmxt	aprprc	junmxt		4	novmit	octmit	octmit
	5	marprc	aprmit	aprmxt		5	febmit	marprc	aprmxt		5	junprc	decmit	aprmit
	6	febprc	marprc	augprc		6	octprc	junprc	maymit		6	sepprc	janmxt	maymit
arctica	1	junprc	junmxt	junmxt	*planifolia*	1	decmit	decmit	janmit	*petiolaris*	1	febmxt	mayprc	mayprc
	2	sepprc	sepmit	janprc		2	junprc	febmxt	marmxt		2	octmit	augprc	sepmit
	3	decprc	aprprc	marprc		3	julprc	aprprc	julprc		3	marmxt	novmit	marmxt
	4	octprc	febprc	aprmxt		4	mayprc	sepprc	sepmxt		4	sepmxt	octmxt	julmit
	5	novprc	maymit	augprc		5	octmit	octprc	octprc		5	junmxt	decmxt	octmit
	6	janmit		maymit		6	julmit	maymxt	sepprc		6	febmit	aprprc	augmxt
myrtillifolia	1	novmxt	novmit	marprc	*maccalliana*	1	janmit	febmit	julprc	*lucida*	1	junmxt	maymxt	maymxt
	2	decmxt	decmit	novmit		2	mayprc	marmxt	augmxt		2	sepmit	sepmit	octmit
	3	augprc	mayprc	febprc		3	sepprc	octmit	octmxt		3	maymit	aprprc	julprc
	4	junprc	octmit	marmit		4	decprc	mayprc	julmxt		4	sepmxt	marmit	mayprc
	5	janmxt	febprc	octmxt		5	junmxt	marmit	marmit		5	janprc	maymit	octmxt
	6	febmxt	sepmit	aprprc		6	marmit	marprc	junmit		6	novmxt	octmit	novmxt
brachycarpa	1	julmxt	junmxt	maymxt	*pyrifolia*	1	junprc	junmxt	junmxt	*discolor*	1	mayprc	mayprc	mayprc
	2	julmit	augmxt	augmxt		2	junprc	sepmit	julprc		2	maymxt	augprc	janmit
	3	augmit	maymit	febprc		3	sepprc	sepprc	marprc		3	janmit	decmit	sepmxt
	4	febmit	octprc	octprc		4	maymxt	octprc	sepmit		4	sepprc	julprc	febprc
	5	mayprc	febprc	maymit		5	julprc	aprmit	maymit		5	aprmit	junmxt	augprc
	6	julprc	decprc	julmxt		6	junmit	decprc	junprc		6	marmit	janprc	julmxt
vestita	1	junprc	junmxt	junmxt	*myricoides*	1	mayprc	novprc	augprc	*exigua*	1	decprc	sepmxt	novmxt
	2	julprc	junprc	julprc		2	maymxt	augprc	sepprc		2	sepprc	maymxt	janmxt
	3	sepmit	sepmit	mayprc		3	augmxt	julprc	mayprc		3	mayprc	octmxt	julprc
	4	maymit	mayprc	maymxt		4	sepprc	mayprc	decmxt		4	janprc	novprc	decmit
	5	augmxt	julprc	aprmit		5	junprc	decmxt	julprc		5	octmxt	febprc	junmit
	6	marprc	octmit	julmit		6	augprc	novmit	octmxt		6	sepmxt	julmxt	febmit
lanata	1	julmxt	junmxt	junmxt	*serissima*	1	junprc	sepprc	junprc	*nigra*	1	sepmit	janmit	janmit
	2	sepmit	augmxt	marprc		2	octprc	octprc	sepprc		2	marmit	maymxt	maymxt
	3	maymit	novmit	augprc		3	sepprc	sepmxt	augmit		3	febmxt	sepmit	octmit
	4	janmit	aprprc	aprprc		4	augmit	sepmit	junmit		4	novmxt	marmxt	marmit
	5		febprc			5	julmit	aprprc	julmxt		5	augprc	sepprc	marmxt
	6		maymxt			6	augprc	febprc	julprc		6	octmxt	febprc	decmit
glauca	1	julmxt	maymxt	maymxt	*eriocephala*	1	febprc	febprc	febprc	*amygdaloides*	1	sepmit	sepmit	janmit
	2	sepmxt	augmxt	augmxt		2	augprc	augprc	augprc		2	sepmxt	augmxt	marmit
	3	maymxt	octprc	febprc		3	augmit	julprc	julprc		3	sepprc	aprmit	octmit
	4	decmit	febprc	sepprc		4	aprmit	novprc	junprc		4	aprmit	junprc	junmit
	5	aprmxt	augprc	octprc		5	mayprc	decmit	aprprc		5	novmxt	julprc	janprc
	6	maymit	janprc	mayprc		6	maymit	octprc	janprc		6	janprc	sepmxt	octprc

Shading denotes similarity

Table S2. Stepwise model details for 6 *Salix* species in three periods

S. arctica	Prevalence	268	12721		266	12601		267	12699	
	1931-1960	*P*	%	1961-1990	*P*	%	1971-2000	*P*	%	
Model step (#)	1	junprc	0.582	99.5	junmxt	0.562	99.3	junmxt	0.475	99.1
	2	sepprc	0.606	99.6	sepmit	0.706	99.7	janprc	0.752	99.8
	3	decprc	0.664	99.7	aprprc	0.779	99.8	marprc	0.825	99.9
	4	octprc	0.795	99.9	febprc	0.843	99.9	aprmxt	0.846	99.9
	5	novprc	0.809	99.9	maymit	0.847	99.9	augprc	0.850	99.9
	6	janmit	0.809	99.9				maymit	0.855	99.9
S. maccalliana	Prevalence	4681	12635		4700	12727		4738	12751	
	1931-1960	*P*	%	1961-1990	*P*	%	1971-2000	*P*	%	
Model step (#)	1	janmit	0.469	70.1	febmit	0.494	74.6	julprc	0.547	81.0
	2	mayprc	0.648	89.6	marmxt	0.706	92.5	augprc	0.679	90.6
	3	sepprc	0.744	94.2	octmit	0.740	93.8	octmxt	0.737	93.6
	4	decprc	0.770	95.1	mayprc	0.781	95.8	julmxt	0.770	95.3
	5	junmxt	0.806	96.7	marmit	0.826	97.0	marmit	0.782	95.8
	6	marmit	0.856	98.0	marprc	0.839	97.1	junmit	0.838	97.4
S. myricoides	Prevalence	3570	12673		3649	12594		3631	12690	
	1931-1960	*P*	%	1961-1990	*P*	%	1971-2000	*P*	%	
Model step (#)	1	mayprc	0.401	75.6	novprc	0.331	67.9	augprc	0.358	69.4
	2	maymxt	0.561	87.8	augprc	0.451	81.5	sepprc	0.465	81.4
	3	augmxt	0.744	96.0	julprc	0.492	83.2	mayprc	0.502	83.4
	4	sepprc	0.757	96.5	mayprc	0.516	84.6	decmxt	0.578	88.8
	5	junprc	0.765	96.7	decmxt	0.557	86.3	julprc	0.641	92.0
	6	augprc	0.787	97.5	novmit	0.593	88.6	octmxt	0.664	93.2
S. pseudomonticola	Prevalence	4494	12692		4550	12705		4466	12645	
	1931-1960	*P*	%	1961-1990	*P*	%	1971-2000	*P*	%	
Model step (#)	1	julmit	0.559	83.5	julmit	0.622	87.2	julmit	0.627	88.5
	2	decprc	0.600	86.6	junmit	0.678	90.6	junmit	0.701	92.4
	3	janmit	0.722	93.5	febmit	0.747	94.1	febmit	0.793	95.9
	4	novmit	0.781	95.9	octmit	0.810	96.1	octmit	0.855	97.4
	5	junprc	0.804	96.4	decmit	0.821	96.7	aprmit	0.863	97.5
	6	sepprc	0.834	97.5	janmxt	0.841	97.1	maymit	0.871	97.7
S. discolor	Prevalence	7069	12744		7198	12728		7080	12650	
	1931-1960	*P*	%	1961-1990	*P*	%	1971-2000	*P*	%	
Model step (#)	1	mayprc	0.845	94.9	mayprc	0.884	96.8	mayprc	0.888	96.9
	2	maymxt	0.917	97.4	augprc	0.932	98.1	janmit	0.902	97.8
	3	janmit	0.930	98.2	decmit	0.940	98.6	sepmxt	0.930	98.2
	4	sepprc	0.933	98.5	julprc	0.946	98.7	febprc	0.937	98.5
	5	aprmit	0.937	98.4	junmxt	0.950	98.8	augprc	0.938	98.5
	6	marmit	0.938	98.6	janprc	0.954	99.0	julmxt	0.941	98.8
S. amygdaloides	Prevalence	1522	12621		1571	12712		1532	12634	
	1931-1960	*P*	%	1961-1990	*P*	%	1971-2000	*P*	%	
Model step (#)	1	sepmit	0.601	96.1	sepmit	0.646	97.1	janmit	0.647	96.8
	2	sepmxt	0.773	98.4	augmxt	0.783	98.8	marmit	0.807	99.2
	3	sepprc	0.838	99.2	aprmit	0.833	99.3	octmit	0.849	99.4
	4	aprmit	0.857	99.5	junprc	0.877	99.6	junmit	0.861	99.5
	5	novmxt	0.865	99.2	julprc	0.899	99.7	janprc	0.870	99.4
	6	janprc	0.881	99.3	sepmxt	0.912	99.8	octprc	0.882	99.5

Average prevalence (number of cells present vs. number of cells analysed), stepwise selections of monthly climate variables, as well as the average probability (*P*) across Ontario input distributions and logistic model concordance values (%) for each step