

Understanding range shift model error: The influence of generation time and rate of adaptation on species distribution model predictions.

Working group proposal

Short title: Range shifts and rates of adaptation

Submitted to NCEAS on August 31, 2012

Principle Investigator

Edmund M. Hart

Beaty Biodiversity Center

Dept. of zoology

University of British Columbia

4200-6270 University Blvd

Vancouver, B.C. V6T 1Z4

ehart@zoology.ubc.ca

Project Summary: Species range shifts is one of the most well documented responses of species to climate change and have been modeled using correlative niche models (species distribution models, SDMs) for more than a decade. These models are statistical correlations between a species realized niche and abiotic variables, because they ignore ecological theory there is often a difference between predictions and a species' actual distribution (error). One potential source of error is that these models ignore differing rates of adaptation to novel climates. Our group will seek to understand the source of this error using meta-analysis, and integrate adaptation into range shift theory. However much of the data from these models is locked in the form of published figures and maps. Therefore a second outcome of our working group will be a web-based data extraction tool. This will allow anyone to upload a figure and extract data from it and store it in DataONE. Beyond producing manuscripts, we will contribute tools for meta-analysis beyond the life of our working group.

Start date December 2012

End date December 2013

Data release December 2013

Resubmission? No

Problem Statement

Species' range shifts was one of the earliest documented ecological responses to climate change (Parmesan, 1996; Parmesan *et al.*, 1999; Parmesan & Yohe, 2003). Understanding range shifts is a pressing issue because as warming increases, species are exhibiting rapid distributional changes (?). Since the late 1990's ecologists have been using species' distribution models (SDM's) to try and predict how those ranges will shift over the next century (Davis *et al.*, 1998; Iverson & Prasad, 1998; Guisan & Zimmermann, 2000; Peterson, 2001). These models assume that the realized niche of a species is primarily determined by climate variables (Austin, 2002; Dormann, 2007). Early models ignored physiology, biotic interactions and rapid local adaptation, relying solely on correlations between current distribution and climate variables (Davis *et al.*, 1998; Pearson & Dawson, 2003; Guisan & Thuiller, 2005; Helmuth *et al.*, 2005). Aside from variance due to lack of ecological theory (Elith & Leathwick, 2009), SDM's can show great variance in their predictive abilities (Elith *et al.*, 2006; Kearney & Porter, 2009; Elith *et al.*, 2010). More recent SDM's have begun to incorporate mechanisms such as physiology (Crozier & Dwyer, 2006; Buckley *et al.*, 2010, 2011) and life history traits (Midgley *et al.*, 2006; Kearney & Porter, 2009; Pöyry *et al.*, 2009; Angert *et al.*, 2011) to improve fit. Including traits and physiology offers a significant improvement in model predictions (Angert *et al.*, 2011; Buckley *et al.*, 2011) but still only explain small percentage of variance. Other factors may be important in explaining the error in species actual ranges and their predicted distribution such as biotic interaction and adaptation.

Adaptation is important to understanding how species will respond to climate change (Visser, 2008; Lavergne *et al.*, 2010; Hoffmann & Sgrò, 2011), but is difficult to accurately measure in the field (Hansen *et al.*, 2012). Despite adaptation playing an important role in species range shifts to both current (Thomas *et al.*, 2001; Bridle & Vines, 2007) and historical climate change (Davis & Shaw, 2001) it is conspicuously absent from SDMs (?) (but see Kearney *et al.* (2009)). Shorter generation times allow for a more rapid adaptive response to strong selective pressures (Berteaux *et al.*, 2004; Somero, 2010; Hoffmann & Sgrò, 2011; Reed *et al.*, 2011; Shaw & Etterson, 2012; Walters *et al.*, 2012). The first reason is that species with shorter generation times can make use of standing additive genetic variation (Figure 1). Models of population persistence have demonstrated that shorter generation times can allow populations to persist by novel adaptation (?Hoffmann & Sgrò, 2011). For instance fur seals are predicted to be unable to adapt to rapid climate change because of long generation times, but this is not the case for other antarctic species with short generation times (Forcada *et al.*, 2008). One consequence of strong directional selection is the loss of additive

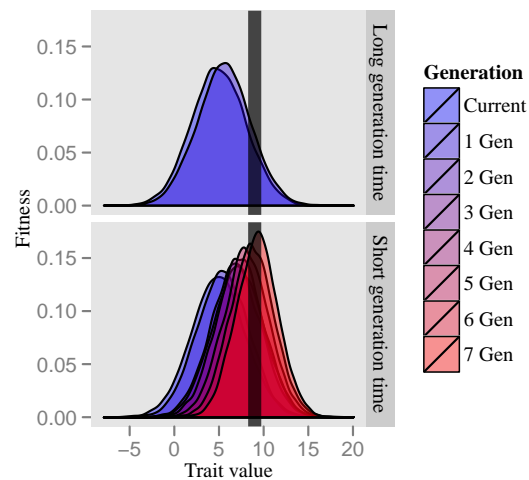


Figure 1: A species with a long generation time can only complete 1 generation in a fixed time, but a short generation time can complete 7 in the same time period, rapidly advancing towards the new fitness optimum (black line)

genetic variance necessary for adaptation (Lande & Shannon, 1996; Hoffmann & Sgrò, 2011). However, species with shorter generation times also have higher rates of molecular evolution due to increased mutation rates (Thomas *et al.*, 2010) adding to the total additive genetic variance. The expectation is species with short generation times will track their climate on the expanding front of the range, but not necessarily go extinct at the trailing edge, instead adapting to novel climate conditions. Therefore one of the fundamental assumptions of SDMs, that species will track their current climate conditions as they shift latitudinally, can be violated by species with short generation times.

Our goal is to (1) quantify SDM error across a range of taxa, (2) investigate how rates of adaptation introduce error model prediction, and (3) build tools to automatically extract data from figures in the published literature.

Proposed Activities

Despite more than a decade of publications on SDM's, adaptation is still absent from most models (Kearney & Porter, 2009). We want to analyze the large number of existing SDMs and attempt to integrate adaptation into range shift theory. By calculating a standard metric of error it is possible to construct models of that unexplained deviance from actual distributions. The challenge is that much of the data for these models is locked in the form of published figures. Methods already exist for data extraction such as the *digitize* package (Poisot, 2011) for R (Figure 2).

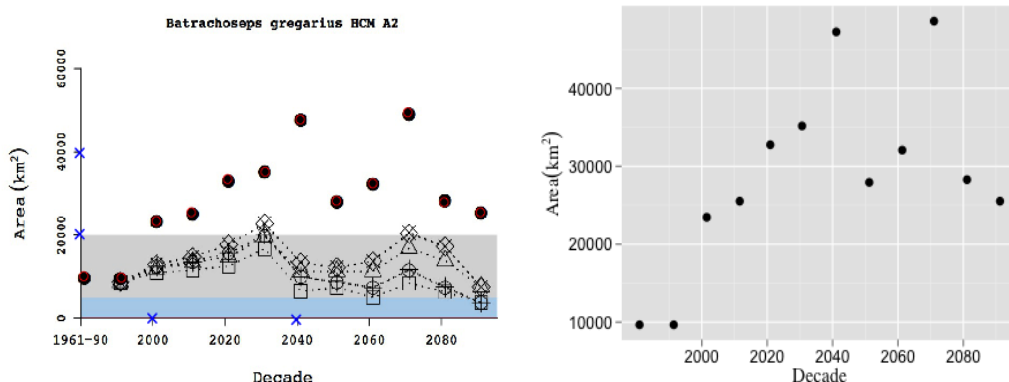


Figure 2: Left panel: Range shift predictions from Early & Sax (2011) of *Batrachoseps gregarius* with calibration points marked Right panel: Extracted data points using the *digitize* package for R (Poisot, 2011)

Using a combination of JavaScript and Python, we want to implement a web based interface for the extraction of data from a variety of figure types in existing publications. Furthermore, once we collect data on error rates in SDM prediction we can test other hypotheses such as the influence of biotic interactions or the breakdown of mutualistic networks.

1. Quantify the amount of error in published SDMs by comparing model predictions to original data sources.
 - By returning to the original data sources, we can measure the error rate of predicted

distributions and actual distributions. We will store this in a publicly accessible database on DataONE.

2. *Investigate the sources of error in SDMs, testing hypotheses about how rates of adaptation contribute to error.*

- We hypothesize that SDMs for species with shorter generation times should have greater error rates because they can rapidly adapt to new invaders and novel climate scenarios at the trailing edge of their distribution (Figure 3). Therefore they are less likely to track their current shifting climate. We will test this hypothesis by constructing mixed-effects models of SDM error.

3. *Develop a web-based interface for data extraction from digitized figures.*

- We will develop an open source tool that allows anyone to extract data from digital figures including: scatter plots, bar charts and georeferencable maps. The interface will store the data at DataONE which can then be used for any future meta-analysis.

Error analysis

Error can be calculated in three ways: difference in area of occupancy, difference in expanding front, and difference in trailing edge. The first two are the most common and we will have the largest sample from these. Because the trailing edge is of the most interest, it can be inferred from total area of occupancy. Species with less shift in the trailing edge will have a larger total area of occupancy. All these can be theoretically be compared by converting them to Z-scores and calculating a standard error statistic such as root-mean squared error. Once we have quantified error, we can construct mixed-effects models with error as the response variable. Using this framework we can add other covariates into our models to control for modeling artifacts, for instance it's known that different SDM construction methods have differences in performance (Elith *et al.*, 2006, 2010). Furthermore we will have assembled a database of residual model error. Our primary goal will be to analyze error in terms of generation time, but this does not exclude considering biotic interactions. Our group has three experts in networks: mutualistic networks (Chamberlain) and food webs (Poisot, Hart). Therefore we will also consider mutualistic pairs of species and examine if range shifts are limited by lack of co-expanding mutualists (Hellmann2008, Pelini2010) or enhanced by release from competitive interactions.

Data extraction from figures

Data from SDMs is often presented in the form of distribution maps or scatter plots (Figure 2). Rather than using existing tools to extract SDM data from figures, we have a more ambitious

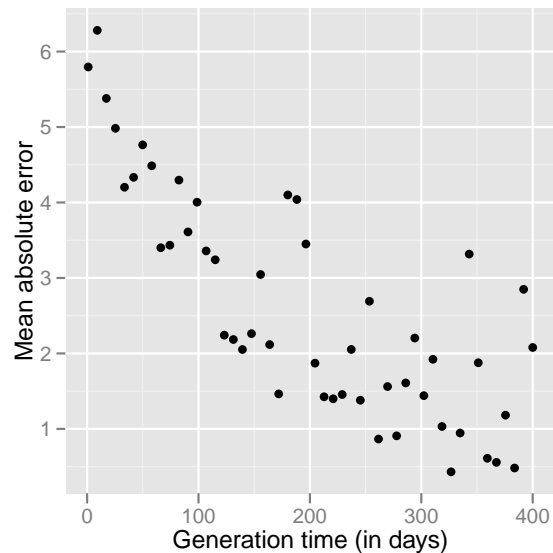


Figure 3: *Hypothesis*

goal of creating a new tool that we can use within the working group to extract data, and will be a resource for all scientists post-working group. While tools exist for data extraction from figures, they are all run natively on personal computers. Examples include GraphClick and ImageJ. We believe a data extraction tool that lives in the web will be much more powerful and benefit science. Living in the web, our tool will be cross-platform (Windows, Linux, OSX, etc.), which is extremely important to get wide adoption. In addition, our tool will allow for automatic data retrieval to a database on the backend, presentation of figures called via APIs from various publishers or user uploaded figures, automatic updating of the user interface (UI), authentication of users in order to track user statistics, and other features as needed. We will build this data extraction tool using a combination of the JavaScript and Python programming languages. Data will be captured in one of two general ways.

First, for some figures we will be able to automate data extraction (for an example, see Figure 2). Second, users will either upload their own figure or will be presented with a figure, at which point they define the axes via clicks on the page, then click on the points of interest on the figure to collect the data. Instead of the data going to a spreadsheet on a hard drive of a personal computer, the data will all go to a central database stored on servers in the cloud. Data in the database will be exposed via a RESTful API, which will allow anyone to search and retrieve data. After or during the NCEAS working group, depending on when the tool is stable, we will beta test the tool with the public to work out bugs. This tool will eventually allow for crowdsourcing of data extraction from figures at large scales, similar to Galaxy Zoo, Citizen Sky, and FoldIt. The massive scale at which this tool will operate will transform quantitative meta-analysis studies. Currently, researchers extract data once from a figure, and someone else may proof their data. However, with our tool, we can potentially get 10 different users to click on each and every point in a single figure, providing for greater accuracy and a measure of precision. To make this happen we need a tool that lives in the web and can take advantage of the massive scale at which science can be done.

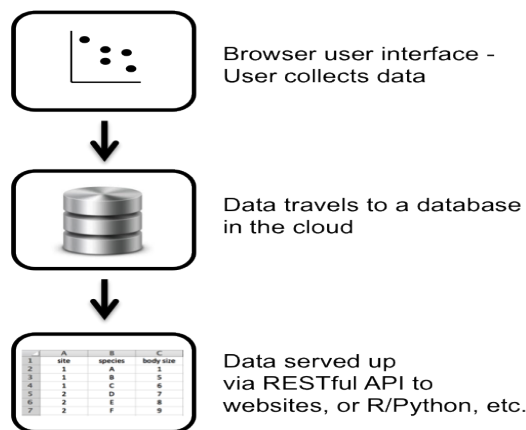


Figure 4: *Figure Extraction*

Table 1: *Participant list. I have organized a group of ecologists, evolutionary biologists and programmers representing a diversity career stages, institutions and NCEAS experience (* represents new NCEAS visitors.). In particular all of the programmers are biologists that have experience building computational tools for ecological questions. All listed participants are confirmed.*

Participant	Affiliation	Expertise / Notes
Jessica Hellmann	University of Notre Dame	Climate change, range shifts, adaptation
Lauren Buckley	University of North Carolina Chapel Hill	Climate change, range shifts, SDM development.
Amy Angert	University of British Columbia	Climate change, range shifts, SDM development
Jessica Blois	University California Merced	Climate change, range shifts SDM development
Scott Chamberlain*	Simon Fraser University	Evolution, software development, <i>NCEAS technical liaison</i>
Karthik Ram	Univeristy of California Berkeley	Climate change, GIS, software development
Tim Poisot	Universit du Qubec Rimouski	Network theory, evolution, software development
Rich FitzJohn*	The University of New South Wales	Evolution, software development
Jens Stevens*	The University of California Davis	Climate change, range shifts, <i>Graduate student</i>
Mark Hahnel*	Digital Science UK, Figshare	Software development, data management
Edmund Hart*	University of British Columbia	Climate change, software development, <i>In charge of data policy requirements</i>

Table 2: *Proposed timeline. We plan to have three meetings over the course of our working group, each time simultaneously developing our software product and writing manuscripts.*

Meeting	Objectives
Prior to meeting	<ul style="list-style-type: none"> • Develop web backend for data extraction from figures, have beta version working. • Create complete list of all relevant SDM papers to extract data from. • Download data for all species in SDM papers and assemble base range sizes and boundaries.
I. December 2012	<ul style="list-style-type: none"> • Beta test figure extraction software, develop web front end and work with NCEAS to set-up databases to store extracted data. Begin working on web front-end • Begin extracting data from figures and assembling figure database • Develop models of SDM error. • Create outline for potential manuscripts.
II. Summer 2013	<ul style="list-style-type: none"> • Complete web front end, and beta user interface for figure extraction. Work with NCEAS to finalize data storage • Finalize initial manuscripts for submission. • Develop ideas for other uses of the error data extracted from SDMs.
III. Winter 2013	<ul style="list-style-type: none"> • Publicly release website after any final tweaks. • Revise submitted manuscripts. • Outline and/or make revisions on further manuscripts using our existing data.

References

- Angert, A.L., Crozier, L.G., Rissler, L.J., Gilman, S.E., Tewksbury, J.J. & Chunco, A.J. (2011). Do species' traits predict recent shifts at expanding range edges? *Ecology letters*, 14, 677–89.
- Austin, M. (2002). Spatial prediction of species distribution: an interface between ecological theory and statistical modelling. *Ecological Modelling*, 157, 101–118.

- Berteaux, D., Réale, D., McAdam, A.G. & Boutin, S. (2004). Keeping pace with fast climate change: can arctic life count on evolution? *Integrative and comparative biology*, 44, 140–151.
- Bridle, J.R. & Vines, T.H. (2007). Limits to evolution at range margins: when and why does adaptation fail? *Trends in ecology & evolution*, 22, 140–7.
- Buckley, J., Butlin, R.K. & Bridle, J.R. (2011). Evidence for evolutionary change associated with the recent range expansion of the British butterfly, *Aricia agestis*, in response to climate change. *Molecular Ecology*, 21, no–no.
- Buckley, L.B., Urban, M.C., Angilletta, M.J., Crozier, L.G., Rissler, L.J. & Sears, M.W. (2010). Can mechanism inform species' distribution models? *Ecology letters*, 13, 1041–54.
- Crozier, L. & Dwyer, G. (2006). Combining Population-Dynamic and Ecophysiological Models to Predict Climate-Induced Insect Range Shifts. *The American naturalist*, 168.
- Davis, A.J., Jenkinson, L.S. & Lawton, J.H. (1998). Making mistakes when predicting shifts in species range in response to global warming. *Nature*, 391, 783–786.
- Davis, M.B. & Shaw, R.G. (2001). Range shifts and adaptive responses to Quaternary climate change. *Science (New York, N.Y.)*, 292, 673–9.
- Dormann, C.F. (2007). Promising the future? Global change projections of species distributions. *Basic and Applied Ecology*, 8, 387–397.
- Early, R. & Sax, D.F. (2011). Analysis of climate paths reveals potential limitations on species range shifts. *Ecology Letters*, pp. no–no.
- Elith, J., Anderson, R.P., Ferrier, S., Guisan, A., Graham, C.H., Dudik, M., Hijmans, R.J., Huettmann, F., Leathwick, J.R., Lehmann, A., Li, J., Lohmann, L.G., Loiselle, B.A., Manion, G., Moritz, C., Nakamura, M., Nakazawa, Y., Overton, J.M., Peterson, A.T., Phillips, S.J., Richardson, K., Scachetti-pereira, R., Schapire, R.E., Williams, S., Wisz, M.S. & Zimmermann, N.E. (2006). Novel methods improve prediction of species' distributions from occurrence data. *Ecography*, 2, 129–151.
- Elith, J., Kearney, M. & Phillips, S. (2010). The art of modelling range-shifting species. *Methods in Ecology and Evolution*, 1, 330–342.
- Elith, J. & Leathwick, J.R. (2009). Species Distribution Models: Ecological Explanation and Prediction Across Space and Time. *Annual Review of Ecology, Evolution, and Systematics*, 40, 677–697.
- Forcada, J., Trathan, P.N. & Murphy, E.J. (2008). Life history buffering in Antarctic mammals and birds against changing patterns of climate and environmental variation. *Global Change Biology*, pp. 2473–2488.
- Guisan, A. & Thuiller, W. (2005). Predicting species distribution: offering more than simple habitat models. *Ecology Letters*, 8, 993–1009.
- Guisan, A. & Zimmermann, N.E. (2000). Predictive habitat distribution models in ecology. *Ecological Modelling*, 135, 147–186.
- Hansen, M.M., Olivieri, I., Waller, D.M. & Nielsen, E.E. (2012). Monitoring adaptive genetic responses to environmental change. *Molecular Ecology*, 21, 1311–1329.

- Helmuth, B., Kingsolver, J.G. & Carrington, E. (2005). Biophysics, physiological ecology, and climate change: does mechanism matter? *Annual review of physiology*, 67, 177–201.
- Hoffmann, A. & Sgrò, C. (2011). Climate change and evolutionary adaptation. *Nature*.
- Iverson, L. & Prasad, A. (1998). Predicting abundance of 80 tree species following climate change in the eastern United States. *Ecological Monographs*.
- Kearney, M. & Porter, W. (2009). Mechanistic niche modelling: combining physiological and spatial data to predict species' ranges. *Ecology letters*, 12, 334–50.
- Kearney, M., Porter, W.P., Williams, C., Ritchie, S. & Hoffmann, A.A. (2009). Integrating biophysical models and evolutionary theory to predict climatic impacts on species' ranges: the dengue mosquito *Aedes aegypti* in Australia. *Functional Ecology*, 23, 528–538 ST – Integrating biophysical models and e.
- Lande, R. & Shannon, S. (1996). The role of genetic variation in adaptation and population persistence in a changing environment. *Evolution*.
- Lavergne, S., Moquet, N., Ronce, O. & Thuiller, W. (2010). Biodiversity and climate change: Integrating evolutionary and ecological responses of species and communities. *Annual Review of Ecology, Evolution, and Systematics*, 41.
- Midgley, G.F., Hughes, G.O., Thuiller, W. & Rebelo, a.G. (2006). Migration rate limitations on climate change-induced range shifts in Cape Proteaceae. *Diversity and Distributions*, 12, 555–562.
- Parmesan, C. (1996). Climate and species' range. *Nature*, 382, 765–766.
- Parmesan, C., Ryrholm, N., Stefanescu, C., Hill, J.K., Thomas, C.D., Descimon, H., Huntley, B., Kaila, L., Kullberg, J., Tammara, T. & Others (1999). Poleward shifts in geographical ranges of butterfly species associated with regional warming. *Nature*, 399, 579–583.
- Parmesan, C. & Yohe, G. (2003). A globally coherent fingerprint of climate change impacts across natural systems. *Nature*, 421, 37–42.
- Pearson, R.G. & Dawson, T.P. (2003). Predicting the impacts of climate change on the distribution of species: are bioclimate envelope models useful? *Global Ecology and Biogeography*, 12, 361–371 ST – Predicting the impacts of climate ch.
- Peterson, A.T. (2001). Predicting species' geographic distributions based on ecological niche modeling. *The Condor*, 103, 599–605.
- Poisot, T. (2011). The digitize Package : Extracting Numerical Data from Scatterplots. *The R Journal*, 3, 25–26.
- Pöyry, J., Luoto, M., Heikkinen, R.K., Kuussaari, M. & Saarinen, K. (2009). Species traits explain recent range shifts of Finnish butterflies. *Global Change Biology*, 15, 732–743.
- Reed, T., Schindler, D. & Waples, R. (2011). Interacting effects of phenotypic plasticity and evolution on population persistence in a changing climate. *Conservation Biology*.
- Shaw, R.G. & Etterson, J.R. (2012). Rapid climate change and the rate of adaptation: insight from experimental quantitative genetics. *The New phytologist*.

- Somero, G.N. (2010). The physiology of climate change: how potentials for acclimatization and genetic adaptation will determine 'winners' and 'losers'. *The Journal of experimental biology*, 213, 912–20.
- Thomas, C.D., Bodsworth, E.J., Wilson, R.J., Simmons, a.D., Davies, Z.G., Musche, M. & Conradt, L. (2001). Ecological and evolutionary processes at expanding range margins. *Nature*, 411, 577–81.
- Thomas, J.a., Welch, J.J., Lanfear, R. & Bromham, L. (2010). A generation time effect on the rate of molecular evolution in invertebrates. *Molecular biology and evolution*, 27, 1173–80.
- Visser, M.E. (2008). Keeping up with a warming world; assessing the rate of adaptation to climate change. *Proceedings. Biological sciences / The Royal Society*, 275, 649–59.
- Walters, R.J., Blanckenhorn, W.U. & Berger, D. (2012). Forecasting extinction risk of ectotherms under climate warming: an evolutionary perspective. *Functional Ecology*.