Using Genetic Algorithms to Fit Species and Habitat Parameters for Modelling the Effect of Climate Change on Species Distributions with Stochastic Patch Occupancy Models

Gary Polhill  
*The James Hutton Institute, gary.polhill@hutton.ac.uk*

Alessandro Gimona  
*The James Hutton Institute, alessandro.gimona@hutton.ac.uk*

Follow this and additional works at: [https://scholarsarchive.byu.edu/iemssconference](https://scholarsarchive.byu.edu/iemssconference)

Part of the Civil Engineering Commons, Data Storage Systems Commons, Environmental Engineering Commons, Hydraulic Engineering Commons, and the Other Civil and Environmental Engineering Commons


[https://scholarsarchive.byu.edu/iemssconference/2014/Stream-G/5](https://scholarsarchive.byu.edu/iemssconference/2014/Stream-G/5)

This Event is brought to you for free and open access by the Civil and Environmental Engineering at BYU ScholarsArchive. It has been accepted for inclusion in International Congress on Environmental Modelling and Software by an authorized administrator of BYU ScholarsArchive. For more information, please contact scholarsarchive@byu.edu, ellen_amatangelo@byu.edu.
Using Genetic Algorithms to Fit Species and Habitat Parameters for Modelling the Effect of Climate Change on Species Distributions with Stochastic Patch Occupancy Models

Gary Polhill, Alessandro Gimona
The James Hutton Institute, Craigiebuckler, Aberdeen, UK
(gary.polhill, alessandro.gimona)@hutton.ac.uk

Abstract: Standard approaches to modelling the effect of climate change on species distributions model a direct link between climatic and other biophysical variables and species occupancy. Though these provide a reasonable estimate for the effects of climate change on species distributions in the future, there are a number of issues with these approaches that fail to account for dynamic landscape interactions. For example, the mass occupancy effect means that species may be observed in unsuitable habitat patches surrounding a well-populated area of highly suitable patches. Conversely, a highly suitable area may be too disconnected from other suitable patches to allow long-term species occupancy. The degree of isolation, however, is not fixed but depends on landscape dynamics. The dynamics of patch occupancy can be modelled using tools such as stochastic patch occupancy models, which use a habitat variable to represent the suitability of each patch for species occupancy, and include the computation of landscape habitat connectivity in determining whether a patch is occupied. Fitting biophysical and climatic variables to habitat suitability and using stochastic patch occupancy models to model the distributions offers a means to account for issues such as the mass occupancy effect, which can be partly responsible for autocorrelated error in purely statistical approaches, and also allows us to account for the effect of the rate and variability with which climatic variables change when modelling future species distributions. However, the method poses a more significant computing challenge to finding the fitting parameters. We report and reflect on preliminary work using genetic algorithms to search for these parameters.

Keywords: Stochastic Patch Occupancy Models, Genetic Algorithms, Climate Change Adaptation, Species Distributions.

1 INTRODUCTION

Standard approaches to computing the effects of climate change on species distributions use climate envelope models (CEM), which construct a model of the observed occupancy \((Q, t = 0)\) of \(p\) patches, or grid squares implicitly containing them, using the distribution of \(N\) biophysical variables \((V)\) thus:

\[
M = f(V_1(0), ..., V_N(0); \beta_1, ..., \beta_N)
\]

where \(f(.)\) is a function (e.g. linear, Gaussian or GAM) and \(\beta_i\) are fitting parameters for \(f(.)\) selected to minimise \(||M - Q||\). (Note that \(Q\) and \(V_i\) are vectors of length \(p\).)

This model is then used to predict occupancy based on a projected distribution of the biophysical variables at some point in the future \((t = T)\) derived from a climate change scenario:

\[
Q(T) = M(V_1(T), ..., V_N(T))
\]

The method is fast, and has proved effective in providing indicative effects of projected climatic changes on species distributions as a basis for conservation policy (e.g. Person and Dawson, 2003). However, it has a number of problems (e.g. Peterson et al., 1999; Beale et al., 2008; Duncan et al., 2009) that suggest exploring alternative approaches merits attention.
A basic problem with CEM is that species are assumed to attain instant equilibrium with climate, which, for species with short dispersal distance, is unrealistic. Two related consequences of such an assumption are that there is no consideration of landscape connectivity (i.e. occupancy of patches is assumed to be independent of that on other patches); and no allowance is made for the possible differences in effects of speed and variability of change on the expected distribution of the species at time $T$. It is increasingly clear that disequilibrium will dominate species distribution changes under climatic shifts, especially for most plant species which have a relatively short dispersal distances, of the order of (hundreds of) metres per year (e.g. Svenning and Sandel, 2014), with landscape fragmentation likely to slow down climate tracking (e.g. Bertrand et al., 2011).

Integrating CEM with a metapopulation approach (see e.g. Hanski and Gilpin, 1991) seems like a logical step to deal with the limitations above. Moilanen’s (1999; 2004) stochastic patch occupancy model (SPOM) is a metapopulation model providing a framework in which the dynamics of the spatial distribution of a single species based on the local (patch) availability of habitat can be simulated. It works by first computing the connectivity $S$ thus:

$$S_i(t) = H_i(t) \sum_{j \neq i} q_i(t-1) \exp(-\alpha d_{ij}) H_j(t)$$

where $i$ and $j$ iterate over the patches, $q_i(t)$ is the modelled occupancy of the species on patch $i$ at time $t$, $\alpha$ is a species dispersal parameter (lower values meaning higher dispersal) and $d_{ij}$ is the distance between patches $i$ and $j$.

If a patch $i$ is unoccupied, a colonisation probability is computed and used to update occupancy:

$$P_i(\text{colonisation}, t) = \frac{S_i^2(t)}{S_i^2(t) + 1}$$

Else, a local extinction probability is computed and used to update occupancy:

$$P_i(\text{extinction}, t) = \frac{\mu}{H_i^2(t)}$$

where $\mu$ is a species mortality parameter and $x$ is a species habitat mortality model parameter.

In the approach described here, we use a stylised landscape of biophysical variables, and construct a model $L$ similarly to the traditional approach that is focused on predicting the availability of habitat as a function of the biophysical variables rather than the occupancy directly.

$$L = f(V_1(0), ..., V_N(0); \beta_1, ..., \beta_N)$$

We then use the SPOM to model the spatial distribution of the species and its dynamic, i.e. the predicted distribution arising from changes in the biophysical variables. The model thus has to find the habitat fitting parameters $\beta$ and the species parameters $\alpha$, $\mu$ and $x$. Since the modelled occupancy to fit to $Q$ requires a SPOM simulation, conventional optimisation algorithms (typically making use of gradients on the fit metric / parameter surface) cannot be used. Here, we explore the use of algorithms applying an evolutionary metaphor.

2 METHOD

2.1 Setup

To create a scenario in which to explore the considerations in section 1, we developed an artificial landscape of two biophysical variables thus:

$$v_1 = kx + a \sin \phi x$$
$$v_2 = ky + a \sin \phi y$$
where $k$, $a$, and $\phi$ are parameters.

We created an artificial observed distribution of the species using the SPOM with given settings of the species parameters $a$, $\mu$ and $x$ and a function to compute the 'real' habitat $G$ based on a Gaussian model:

$$G = \prod_{i=1,2} \exp \left( - \frac{\mu_i - \lambda_i}{\sigma_i} \right)^2$$

where $\lambda_i$ and $\sigma_i$ are parameters.

We ran the SPOM for 100 time steps starting with full initial occupancy, to allow for the landscape scale occupancy to reach equilibrium, and took this state as the occupancy for $t = 0, Q$. Note that in SPOM simulations, equilibria in the strict sense are not reached, in that there is variability in occupancy from step to step. The word 'equilibrium' is used in the context of the SPOM to imply a relatively constant mean occupancy.

The species parameters were chosen to model a species with a moderate rate of dispersal and patch mortality. We used $a = 1$, $k = 2.27596$, $\phi = 4.5$, $\lambda_1 = \lambda_2 = 2.26442$ and $\sigma_1 = \sigma_2 = 1.51522$. These parameters were chosen to create a landscape in which there is good habitat for the species in a clearly defined region (or subregions) of the space. This creates a habitat map and species distribution shown in figure 1.

**Figure 1.** Spatial distribution of $G$ (habitat quality, left) and $Q$ (probability of occupancy after 100 time steps) when $a = 1.4, \mu = 0.02$ and $x = 1.3$ (right: grey level corresponds to mean occupancy over 20 samples, size of blue circles to variance in occupancy – larger meaning greater variance.)

To simulate the effect of climate change on these two variables, we created two time series:

$$c_1(t) = \frac{tm}{T} + \epsilon_1$$
$$c_2(t) = \frac{tm}{T} + \epsilon_2$$

where $m$ is a parameter, and $\epsilon_1$ and $\epsilon_2$ are variability parameters sampled stochastically from a normal distribution with mean 0 and standard deviations $s_1$ and $s_2$ (also parameters) respectively. $G(t)$, the habitat derived from changes in the biophysical variables is then computed as:

$$G(t) = \prod_{i=1,2} \exp \left( - \frac{\mu_i + c_i(t) - \lambda_i}{\sigma_i} \right)^2$$
This is used to provide an ‘actual’ effect of climate change on the species distribution to compare with those found by model fitting. To simulate a fast or slow change, we used values of $T = 10, 50$ respectively with $m = 2.45697$; to simulate low or high variance, we used values of $s_1 = s_2 = 0.40472$, $0.18100$ respectively. The $m$ parameter was chosen to ensure that the largest region of good habitat (top right in Figure 1) moved (to the bottom left as shown in Figure 2), requiring the species to migrate across the space during the period of simulated climate change. The variance parameters $s_1$ and $s_2$ were selected to balance creating different climate variability profiles with changes in the more extreme cases (fast change or high variance) that did not typically cause extinction.

2.2 Evolutionary search

The evolutionary search algorithm represents a candidate for the species and habitat parameters as a ‘genome’ (simply a vector $w$). A population $W$ of candidates is generated, their ‘fitnesses’ computed, and those members of $W$ with higher fitness are given a better chance of propagating the parameter settings they represent into the next generation. The algorithm uses a Gaussian function to model the habitat from the $\beta$ parameters, which each have a $\lambda$ and $\sigma$ component:

$$H_j = \prod_{i=1,2} \exp - \left( \frac{v_i - \beta_i^j}{\beta_i^j} \right)^2$$

A multicriteria fitness function is used, with components representing:

1. the number of time steps the SPOM ran before there was total extinction;
2. the degree to which the simulation had reached equilibrium occupancy (measured by the $p$-value of a test of difference of means in occupancy over time steps in the third and fourth quarters of the simulation);
3. the patch-by-patch similarity in mean occupancy between the observed and simulated values over the time steps in the latter half of the run (measured as the sum of squared difference between observed and simulated mean occupancy);
4. the similarity in total occupancy (measured as the number of patches minus the absolute difference between observed total occupancy (i.e. number of occupied patches) and the total simulated mean occupancy);
5. and the number of patches with mean occupancy close to 1 or 0.

Component 1 ensures that the parameter settings provide a viable population. Component 2 is more debatable as the implicit assumption is that the number of time steps for which the simulation is run for each searched parameter vector is sufficient. However it is included as a gauge of whether the other metrics applied to a stable population. Component 3 aims to make each individual patch have an occupancy probability close to the observed occupancy, whilst component 4 checks that overall levels of occupancy are correct. Component 5 was included because components 3 and 4 use mean occupancy rates over the latter half of the simulation, which can create local minima in which occupancy probabilities are close to 0.5.

The fitnesses are represented as a vector, avoiding the need to give weightings to each component. As a consequence, members of the population are partially ordered in fitness, with those for which no other member has better or equal fitness in all dimensions form a Pareto front in fitness space. Selection of breeding partners is based on a standard lottery approach in which the highest fitness member gets $\#W$ tickets, through to the lowest, which gets 1. Since the population is partially ordered for fitness, the sorting algorithm assigns members of the population on the Pareto front the highest numbers of tickets, and then proceeds through the lattice of partially-ordered fitnesses one layer at a time assigning lower and lower numbers of tickets.

We used ter Braak’s (2004) differential evolution Markov chain (DEMC) algorithm, with a generalisation to multicriteria fitness functions in which the Metropolis ratio $r$ (used to compute the probability of substituting one member of the population $w_1$ with another $w_2$) is:
where $f_i(.)$ is the $i$th fitness function, and $w_2$ replaces $w_1$ with probability $\min(1, r)$. We used a perturbation parameter $b$ of 0.01 (the DEMC algorithm adjusts bases on the genome randomly by sampling uniformly from $\pm b$, DEMC parameter $\gamma = 0.50742$, and a crossover probability CR of 0.1. The initial population is sampled from a prior distribution for each variable ($[0.5, 1.5]$ for $x$ and $\alpha$; $[0, 0.2]$ for $\mu$ and $[-5, 5]$ for all $\beta^k$ and $\beta^m$). In this early test of the work, we ran the GA for 200 generations.

For each candidate parameter setting, the SPOM is run for 100 time steps.

3 RESULTS AND DISCUSSION

3.1 Effect of speed and variability of climate change

![Example species distributions](image)

Figure 2. Example species distributions (white circles) overlaid on final habitat (colours as per figure 1) after various combinations of speed and variability of change. Top row (a-c): fast change with high variability, showing path dependence in equilibrium outcome. Bottom row (d-e): fast change with low variability, still showing potential differences in outcome; (f): slow change.

We ran the SPOM twenty times for each of fast or slow change with high or low variance, using the same climate time series for each replication in each combination of speed and variability of change. Figure 2 shows a selection of the results. With fast change (2(a-e)) the results were highly path dependent in that the final observed distribution is a function of particular species distribution patterns and dispersal and mortality dynamics during the course of the run. As should be expected, this is especially true with high variability (2(a-c)), where half the results involved total extinction of the species, with the rest of the results showing the species ending up on one or more of the new ‘islands’ of habitat created by climate change after a 100 time step run-out period to check the equilibrium distribution. With lower variability (2(d-e)), the species survived the fast change scenario in all runs, but the outcome was still path-dependent, as not all islands of habitat are populated. Less difference between the high and low variability runs were observed in the slow change scenario (2(f)), and the equilibrium outcome was much more consistent.
3.2 Comparison of evolutionary search with climate envelopes

We used Thuiller et al.’s (2013) Biomod2 package for R to compute climate envelopes. As should be expected, they show a good fit to the species distribution, of which the mean observations are shown in figure 1. The predictions of the CEM are also a reasonable fit (by inspection) to the slow change scenario in figure 2(f), particularly at higher expected observation levels, but they do not capture the potential variability in distribution shown by the fast change scenarios, particularly with high variability.

![Fit to equilibrium species distribution](image1)

![Predicted species distribution](image2)

**Figure 3.** CEM model and prediction for the artificial landscape. Red cells indicate a higher expectation of observing the species.

![Fits of two samples found by the evolutionary search algorithm](image3)

**Figure 4.** Fits of two samples found by the evolutionary search algorithm, with scales showing mean occupancy over the last quarter of the run.

The evolutionary search algorithm saved the 96 best samples during the course of the run. Although it had only been run for 200 generations, it had found some parameters providing reasonably good fits, in that the higher probabilities of occupancy in the simulations occur in occupied patches in the observed data (figure 4), albeit not as good as the CEM fit. Although we have not yet explored the forecasts of these fits, the work in section 3.1 demonstrates that, given a set of species and biophysical parameters learned through evolutionary methods, the SPOM could be used to explore the potential variability in outcomes these parameters entail.

5 CONCLUSION AND FUTURE WORK

We have shown that there can be considerable path dependence in outcome of species distributions, particularly for faster and more variable changes in climate. This path dependence can mean that species do not end up on the most likely areas. Developing techniques that can explore these scenarios in a way that captures this uncertainty will be more useful.
One of the matters not explored here is the issue of equilibrium. In the fast change runs, it took a number of time steps for the species to ‘catch up’ with the change in location of the best habitat. Figure 5 shows the species distribution corresponding figure 2(d) immediately after the change has taken place (rather than, as in figure 2(d), after a 100 time step burn-out period). If the factors influencing availability of habitat, whether through climate change, interaction with other species, or human activity, are the subject of continual change, knowing the equilibrium distribution of species after habitat change may not be particularly useful. For similar reasons, it may also not be reasonable to assume that an observed species distribution used to fit a model is at equilibrium.

![Figure 5. Distribution of species immediately prior to 100 step run-out in bottom left distribution in figure 2.](image)

A particular issue with the evolutionary approach is the large number of fitness function components, which creates the potential for several members of the Pareto front to exist. The graph in figure 6 shows how the size of the Pareto front changes over the course of the 200 generation run of the algorithm. Efforts to find effective fitness functions, preferably minimising the number of different ones used, would be beneficial in reducing this problem.

![Figure 6. Size of Pareto front during the course of the evolutionary algorithm.](image)

Future work could look at multiple species, using an extension of the SPOM to look at metacommunities, in which species can influence the occupancy of others (see, e.g. Polhill and Gimona, 2013). However, since much of the habitat provision in western Europe is a function of the decisions made by land managers (as opposed to non-human-induced change), integrating this work with the exploration of climate change adaptation methods of land managers is also a priority. Another focus of future work, given the greater availability of biophysical and climate data over species occupancy, is exploring whether the evolutionary dynamic modelling approach described here could be adapted for use with observations not assumed to be at equilibrium.
ACKNOWLEDGEMENTS

This work was funded by the Scottish Government Rural Affairs and the Environment Portfolio Strategic Research Theme 3 (Land Use).

REFERENCES