Jun 16th, 9:00 AM - 10:20 AM

Defining the spatiotemporal surveillance space for alien species' invasions using approximate Bayesian computation

Grant Hamilton
Queensland University of Technology, g.hamilton@qut.edu.au

Rune Rasmussen
Queensland University of Technology

Jana Mullerova
Academy of Sciences of the Czech Republic

Jan Pergl
Academy of Sciences of the Czech Republic

Petr Pysek
Academy of Sciences of the Czech Republic

Follow this and additional works at: 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

Defining the spatiotemporal surveillance space for alien species' invasions using approximate Bayesian computation

Grant Hamilton PhD[1,2] , **Rune Rasmussen PhD[1,2], Jana Müllerová PhD[3]**
Jan Pergl PhD[3], Petr Pyšek PhD[3,4]

[1]*Plant Biosecurity Research Centre, Australia*
[2]*Queensland University of Technology, School of Earth, Environmental and Biological Sciences, GPO Box 2434, Brisbane 001*
[3]*Institute of Botany, Academy of Sciences of the Czech Republic, CZ-252 43 Průhonice, Czech Republic*
[4]*Department of Ecology, Faculty of Science, Charles University in Prague, Viničná 7, CZ-128 44 Prague, Czech Republic*

g.hamilton@qut.edu.au

Abstract: The spatiotemporal dynamics of an alien species invasion across a real landscape are typically complex. While surveillance is an essential part of a management response, planning surveillance in space and time present a difficult challenge due to this complexity. We show here a method for determining the highest probability sites for occupancy across a landscape at an arbitrary point in the future, based on occupancy data from a single slice in time. We apply to the method to the invasion of Giant Hogweed, a serious weed in the Czech republic and throughout Europe.

Keywords: surveillance; range expansion; alien invasive species; spatial modeling; approximate Bayesian computation; simulation.

1   BACKGROUND

Predicting the areas into which an invasive species is likely to spread, and the timing of that spread, are critical elements of managing an invasion. Mathematical models for predicting spread in homogeneous space have existed for some time (Skellam 1951). Predicting spatial dynamics in the real world is much more difficult however, due to the underlying heterogeneity in the landscape across which invasions occur. Factors such as topography, the spatial distribution of hosts and other resources elements of habitat suitability all contribute to complexity that can be challenging to model.

Although simulation using cellular automata (CA) allows for the incorporation of complexity and thus presents a way forward, the data needed to parameterise a CA can be difficult to acquire, particularly since the dynamics of the invasive species may be different in its new range. Rasmussen and Hamilton (2012) presented a method to estimate the parameters of a range expansion using a snapshot of data using Approximate Bayesian Computation (ABC).

In this paper we demonstrate a probabilistic spatial model that can be used to predict the highest probability areas for occurrence of an invasive species across a heterogeneous landscape. The model is applied to the spread of Hogweed (*Heracleum mantegazzianum*) in the Czech republic, a serious weed of major concern (Pyšek et al. 2012).

2   MODELLING APPROACH

The range expansion model used in this study is similar to the model presented in Rasmussen and Hamilton (2012). The expansion model here has been implemented in concert with ABC to approximate the parameters of the expansion model, given point-in-time abundance data of a target species (see Rasmussen and Hamilton, 2012). In this model the number within each cells of the CA was used to

characterise the abundance of Hogweed within a geographical region represented by the cell. The CA operated over a grid of equally sized square cells, where each cell (x,y) involved the following components:

1. P(x,y) evaluated an abundance of Hogweed on the region of the cell at (*x,y*) at time *t.*
2. K(x,y) evaluated the carrying capacity as the maximum abundance of Hogweed that could occupy the region of the cell (*x,y*)
3. H(x,y) evaluated habitat suitability as the probability of establishment in the region of a cell (*x,y*).

The model involves two traversals over the set of cells at each time step *t*. The first traversal generates dispersal patterns of Hogweed, whereas the second traversal computes the growths and decays of Hogweed populations. In each first traversal, the model applies a dispersal rule on populated cells to generate dispersals to target neighbouring cells, where a random chance of establishment is determined a probability (H(*x,y*)). This dispersal rule involves a combination of short and long-distance dispersals. The model then applies a logistic growth update on all of the populated cells. After a predefined number of time steps, the range expansion model returns a dispersal pattern as a set of abundance values distributed over a grid. A deeper discussion about this model can be found in Rasmussen and Hamilton (2012). The parameters of interest in our analysis are given in Table 1.

Table 1. Parameters used in the forward simulation model. These parameters can be estimated with ABC using the method presented by Rasmussen and Hamilton (2012).

2.1

Element	Description
τ	The number of time steps (years) since invasion.
x_0	The x-coordinate of the initial population.
y_0	The y-coordinate of the initial population.
$P_0(x_0, y_0)$	The initial population size.
$\hat{r}$	The intrinsic rate of population growth.
p_{mig}	The probability of migration.
p_{LDD}	Probability of long distance dispersal.
σ_r	The standard deviation of the long-distance dispersal kernel.

Approximate Bayesian computation

Approximate Bayesian Computation (ABC) belongs to a family of likelihood-free Bayesian inference programs that attempt to approximate the posterior densities for problems where the measures of likelihood are *a priori* unknown (Marjoram *et al.*, 2003). ABC can be a sufficient solution to complex problems where a full-likelihood method is prohibitive (Beaumont *et al.*, 2002).

For our problem concerning a range expansion model, we consider a model $D_s \sim \phi(\theta)$, where D_s is the simulated data. An ABC program may be applied to estimate the posterior density $p(\theta \mid D_o)$ for a parameter θ and some observed data D_o. The Bayesian formula for the problem is:

$$p(\theta \mid D_o) = \frac{p(D_o \mid \theta)\pi(\theta)}{p(D_o)}$$

(1)

Where $p(D_o \mid \theta)$ is the *likelihood*, $\pi(\theta)$ is the *prior* and $p(D_o)$ is the evidence of the parameter θ. Since $p(D_o) = \int p(D_o \mid \theta)\pi(\theta) \, d\theta$ sums to a scaling constant, Equation (1) can be reduced to a proportionality:

$$p(\theta \mid D_o) \propto p(D_o \mid \theta)\pi(\theta)$$

(2)

ABC programs involve an ideal constraint $d(D_s, D_o) = 0$ to filter some rare parameters θ_i from a sufficiently large random sample, on the condition that simulated data from $\phi(\theta_i)$ satisfies the constraint. Here the distance metric $d(\cdot, \cdot)$ is Euclidean over the dimensions of the observed and simulated data, and $p(d(D_s, D_o) = 0 \mid \theta)$ approximates the likelihood $p(D_o \mid \theta)$. The output of an ABC program are the θ_i, because the $\theta_i \sim p(\cdot \mid D_o)$ and can be used to approximate the posterior.

2.2 Data and habitat suitability model

Hogweed data for a 2.5 x 2.5 km region (Figure 1) were collected in the Czech republic (see Mullerova et al 2011 for details). We used the 2006 time slice for experiments. The dataset involved categorical evaluations of hogweed within the cells of an 800x800 grid spanning the study region. We configured this a 200x200 grid. Grid cells were 10x10m, with abundance estimated by proportion of the grid cell covered by hogweed.

The landscape was classified using a landscape suitability model (1=most suitable, 2-least suitable, 3-uninhabitable), and converted to a probability of establishment as:

$$p_e(x) = A\left[1 - e^{(3-x)}\right] \qquad (3)$$

where A was set to $\frac{1}{1-e^2}$ so that $p_e = 1$ when $x = 1$ and $p_e = 0$ when $x = 3$. Given this assumption, the establishment probability for each 10×10-meter cell was computed as the average of the establishment probabilities found for the enclosed 2.5×2.5-meter cells and their suitability categories, using this formula.

Figure 1. The study region around Prameny in the Czech Republic. The point inside the red circle shows the point of invasion of Hogweed in the region estimated using the method presented by Rasmussen and Hamilton (2012)(Google maps, retrieved March 2014, URL: https://Maps.google.com)

2.3 Experimental setup

The forward expansion model is configured to simulated from an existing spatial population distribution with known abundances, forward in time for a given number of time steps (τ) across a heterogeneous landscape. Forward simulations of the expansion model require prior knowledge of the model's parameters for a given species and habitat, along with a habitat suitability model and an initial dispersal pattern. We used ABC to arrive at a large parameter sample that specify the most likely model parameters for hogweed in the study region. The time since invasion was estimated but is not presented here since it was not used as input for the forward simulation. We specified a number of time steps (Fτ) and drew parameters at random from the parameter sample to inform a large number of forward

simulations. For each $F\tau$ considered, the forward simulations generate a large set of dispersal patterns that each represents a possible future pattern. The average of these dispersal patterns shows the most likely dispersal pattern expected for the region.

The forward simulation experiment was set up to sample parameters from the posterior distributions of the ABC results. The 2006 dispersal pattern was used as the starting pattern and expansion fronts were predicted using forward simulations to 2015. These expansion fronts were predicted based on the average map derived from 10,000 simulated dispersal patterns for each prediction year. Prediction maps are shown in Figure 2.

3   SIMULATION RESULTS

3.1 ABC priors and results

Table 2. Prior bounds and results for the ABC experiment (SD-standard deviation of the posterior distribution) used as input for the forward simulation experiment

Parameters	Prior Bounds	Posterior Mean	SD
x_0	0-199 cells	39	23.4
y_0	0-199 cells	66	20.9
$\hat{r}$	0-0.5	0.204	0.045
p_{miq}	0-1	0.2	0.105
p_{LDD}	0-1	0.21	0.117
σ_r	1-1000m	447 m	84.3

3.2   Forward simulations

The expected occupancy of Hogweed in the study region for the year 2015 is shown in Figure 2. Note the relatively restricted high probability occupancy area for 2015, which would serve as an appropriate surveillance space for this region. In contrast, the large area grey shaded region shows projected areas of low probability Hogweed occupancy in the projected time period. Surveillance in these areas is unlikely to detect Hogweed in any region will be determined not only be decision support tools such as that presented here, but also by budget, cost of surveillance and other factors that may impact on management decisions.

Given the projection past the current date, no empirical validation could be conducted at this stage.

4. SUMMARY AND DISCUSSION

We have demonstrated here a powerful method in which data from a single point in time can be used to estimate the occupancy of an invasive species at an arbitrary point in time in the future. The results shown in Figure 2, for example, present the most likely expansion future for Hogweed in this region taking into account the spatial dynamics and the underlying heterogeneity in the region. Consequently this could be used to define the surveillance space for Hogweed to better plan and prioritise areas for management.

Using ABC in association with forward simulation presents an effective method for estimating the most like parameters of the range expansion and using them to forward simulate across a landscape. The advantages here are notable, since while external information can be used (in the form of prior information to limit the possible range of the parameters), it is also possible that the spatial dynamics of the invasive species will be different in a newly invaded environment. Consequently, the use of existing data from the invaded range in the analysis helps to limit any possible bias.
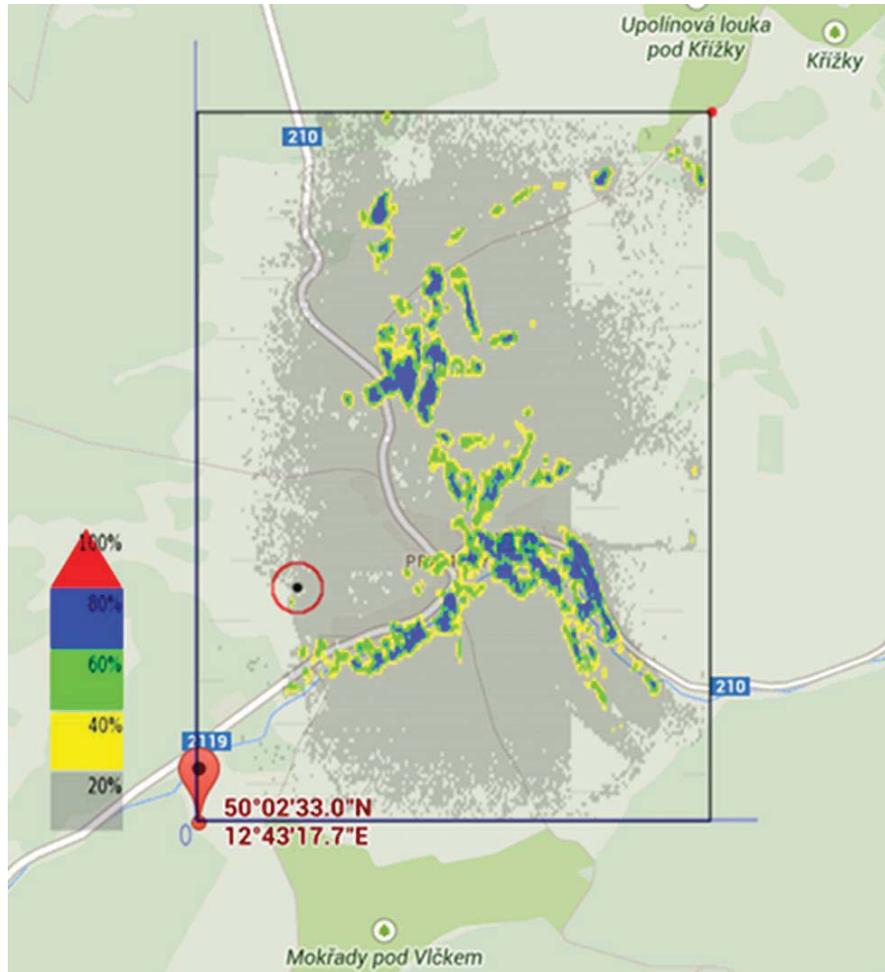
Figure 2. The estimated occupancy of Hogweed in 2015. The scale shows the probability of occupancy calculated as the proportion of times over 10,000 forward simulations that the cell is occupied.

The capture of data for invasive species presents challenges in its own right. We note here the possibility for capturing data using new technologies such as Unmanned Aerial Systems (UAS) at a regional scale. Notably, however, while these methods have generated much excitement, reliable methods to make use of the data for management will still be required. The generic nature of the approach presented here, which allows for the data to be interpreted more easily, should be sought after as the capacity to gather data easily increases.

The method should work well for established species of concern for which there are sufficient data. However as with any Bayesian method, the precision of results will in part depend on the availability of data in combination with the information that can be introduced through the priors. For newly detected species, such as would occur in a biosecurity response, the strength of the analysis may initially be limited, only improving as more data become available. For this, it may be useful to use an alternative and more generic analysis initially, introducing the method presented here as more information becomes available. Nonetheless, this method present a useful step forward in better defining the surveillance space in space and time for invasive species.

ACKNOWLEDGMENTS

were provided by the HPC and Research Support Group, Queensland University of Technology, Brisbane, Australia

REFERENCES

Beaumont, M. A., Zhang, W., and Balding, D. J. 2002. Approximate Bayesian computation in population genetics. Genetics, 162(4), 2025–2035

Beaumont, M. A., Cornuet, J. M., Marin, J. M., and Robert, C. P. 2009. Adaptive approximate Bayesian computation. Biometrika, 96(4), 983–990

Marjoram, P., Molitor, J., Plagnol, V., and Tavarè, S. 2003. Markov Chain Monte Carlo without Likelihoods. Proceedings of the National Academy of Sciences (PNAS), USA, 100(26), 15324– 15328

Müllerová, J., Pergl, J. and Pyšek, P. 2013. Remote sensing as a tool for monitoring plant invasions: Testing the effects of data resolution and image classification approach on the detection of a model plant species Heracleum mantegazzianum (giant hogweed). International Journal of Applied Earth Observation and Geoinformation, 25(1), 55-65

Pyšek, P., Chytrý, M., Pergl, J., Sádlo, J. and Wild, J. 2012. Plant invasions in the Czech Republic: current state, introduction dynamics, invasive species and invaded habitats. Preslia 84(3), 575-630.

Rasmussen, R. and Hamilton, G. 2012. An approximate Bayesian computation approach for estimating parameters of complex environmental processes in a cellular automata. Environmental Modelling and Software, 29(1), 1–10

Rasmussen, R. and Hamilton, G. 2013. ABC Rex url: http://hamiltonlab.wordpress.com/abc-in-ecology/abc-rex/

Skellam J.G. 1951. Random dispersal in theoretical populations Biometrika (1951) 38 (1-2): 196-218

Wilkinson,R.D. 2013. Approximate Bayesian computation (ABC) give exact results under the assumption of model error. Statistical applications in genetics and molecular biology 12(2):129-141