AUSTRALIAN MATHEMATICAL SCIENCES INSTITUTE

National Collaboration in the Mathematical Sciences

Bayesian invasive species modelling allowing for increasing public awareness

Thao Le
Supervisor: Dr Jonathan Keith
Monash University

February 2014

Contents

Postal Address: 111 Barry Street
c/- The University of Melbourne
Victoria 3010 Australia

Email: enquiries@amsi.org.au
Phone: +61 3 8344 1777
Fax: +61 3 9349 4106
Web: www.amsi.org.au

1 Introduction

Red imported fire ants (RIFA), also known as *Solenopsis Invicta*, are ranked as one of the world's 100 worst invasive alien species. [1] They were first detected in Brisbane, 2001, at two separated locations on the same day. Reasonably, this was met with alarm and the National Red Imported Fire Ant Eradication Program was launched. As the eradication program progressed, the number of detected nests decreased - it seemed that the invasion was being well curbed by the program. [2] Although nests were found further outwards from the expected infected sites, by 2008, there was an air of optimism about the progress of the eradication. [3] But it is now 2014 and the program continues to run.

Keith and Spring (2013) developed a model for the ants' invasion. The model found that the fire ants population was low in 2004 and subsequently increased. Furthermore, despite the fluctuations in the number of ant nests, the frontier of the invasion had been continuously expanding. [4]

Keith and Spring describe their model as being *agent-based Bayesian*. *Agent-based* is in contrast with using differential equations to model a process. Whilst differential equations computationally require spatial and temporal grids, agent-based models have no limit on temporal and spatial resolutions. The agents - the red imported fire ants nests - and their interactions are the basis of the model. *Bayesian* refers to a branch of probability and statistics. *Bayes Rule* is the underlying structure of the model's mathematics:

$$p(\theta|y) = \frac{p(y|\theta)\, p(\theta)}{p(y)} \tag{1}$$

The parameters of the model are represented by θ and the data by y. This is often reduced down to

$$p(\theta|y) \propto p(y|\theta)\, p(\theta) \tag{2}$$

$p(y|\theta)$ is known as the likelihood, and $p(\theta)$ is the prior density. The result, $p(\theta|y)$ is known as the posterior distribution.

The equation for the model is essentially a very long formula, and is a *non standard* distribution. Examples of standard probability distributions are the Normal (Gaussian) or the Binomial. Being non standard, the model draws it values indirectly instead, using *Markov Chain Monte Carlo* or MCMC. Markov Chain Monte Carlo describes an area of techniques that works by picking approximate values, such that the distribution of values converges to the distribution of the model. It is used to generate (potentially) thousands of samples, and the samples prior to convergence are discarded as *burn-in* - i.e. the samples that do not follow the distribution of the model.

2 Project Outline

Inside of this large model, my project was to add in time-dependent public awareness probabilities. Briefly, the public have some probability of finding an ant nest, and that probability depends on which area type (there are four classifications from urban to rural). The original model assumed that this was constant. However, one would expect that the public becomes better at recognising red imported fire ant nests as time passes. Thus, the project was to

1. Formulate how this time dependency works, and the resulting modified model.
2. Modify the existing code to allow for the time dependency.
3. Compile and run the software on the same datasets.
4. Then compare (if any), the changes to modelled invasion and predicted effectiveness of the eradication program.

The time dependency expansion and subsequent code changes were straight forward: the constant variables of the original model were replaced by functions that depended on time.

3 Results and Discussion

The original software generated 20000 samples – 20000 alternative invasion histories – and the parameters showed convergence after 10000 samples. However, this was not the case for the modified software with the time dependency. As such, the following results come from a run generating 40000 samples.

Since most of the α time values do appear converged after 20000 samples, the first 30000 samples were discarded as burn-in and the last 10000 samples were used for the analysis of α_2, α_3, α_4 and α_5. The median of the α time values are shown in Figure 1. The detection probabilities are low at the beginning, prior to the first discoveries in February 2001 (equivalently month 61) and increasing afterwards. The probabilities appear to increase sharply towards the more recent months, and is a marked difference from the static values in the original model.

Postal Address: 111 Barry Street
c/- The University of Melbourne
Victoria 3010 Australia

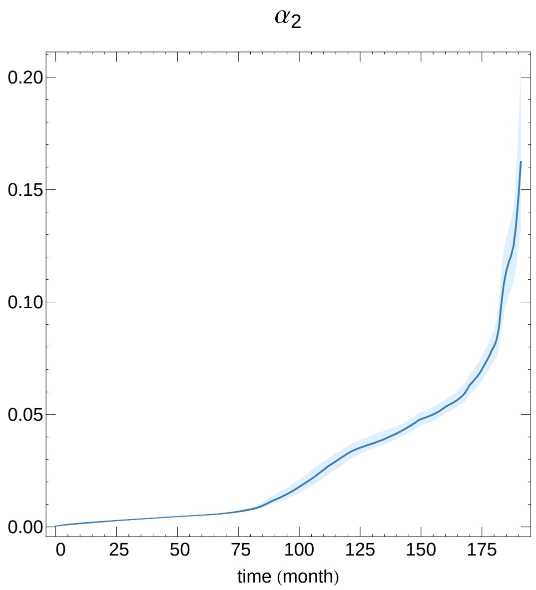
Email: enquiries@amsi.org.au
Phone: +61 3 8344 1777
Fax: +61 3 9349 4106
Web: www.amsi.org.au

However, whether or not the total number of nests converged, even after the 40000 samples, is unclear. Due to a simple lack of time to conduct a longer run, the last 10000 samples were also used to find the number of mature nests over time, in Figure 2. The estimated rise and fall of the numbers of fire ants nests follow the same patterns in the original model and in the modified model. The population decreased rapidly after the eradication began in 2001, and reached a minimum by the start of 2004. The nests then increased, peaking around 2010 and in both the original and modified model, appears to be declining since. The largest difference comes after 150 months: the modified model predicts that public detection becomes better at later times (Figure 1) and this is reflected in Figure 2 by the lower number of mature nests predicted at those later times.

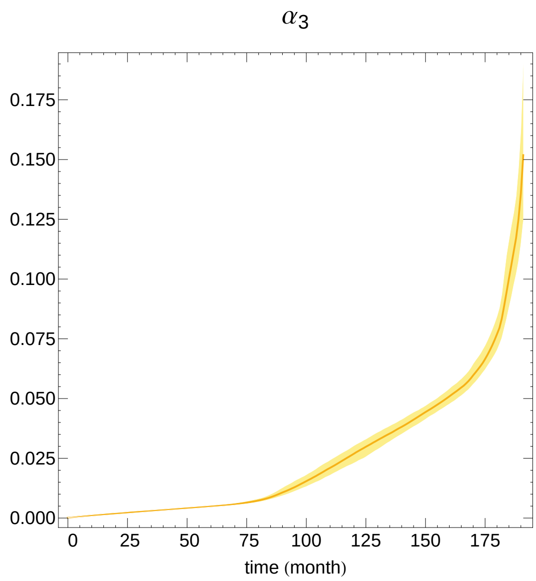
The result from the modified model reminds us of the importance of informing the public and harnessing their help to eradicate the red imported fire ants. This becomes especially important as the boundaries of the expansion increase and it becomes infeasible to professionally monitor large areas. The modified model predicted that public detection probabilities increased over time, sharply so towards the later months, and subsequently suggests that eradication program was more successful than previously predicted by the original model.

Postal Address: 111 Barry Street
c/- The University of Melbourne
Victoria 3010 Australia

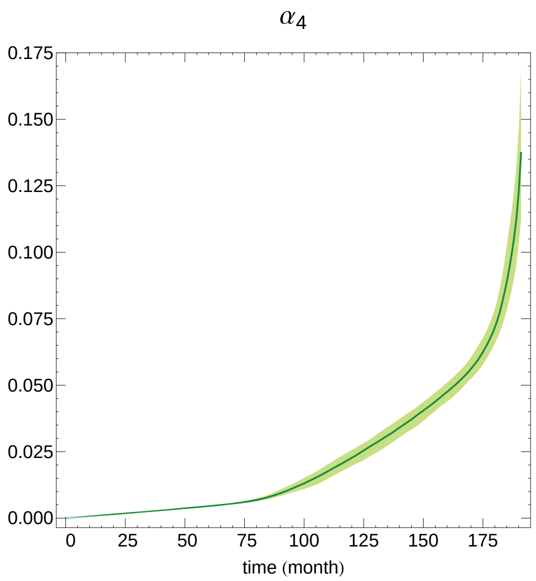
Email: enquiries@amsi.org.au
Phone: +61 3 8344 1777
Fax: +61 3 9349 4106
Web: www.amsi.org.au



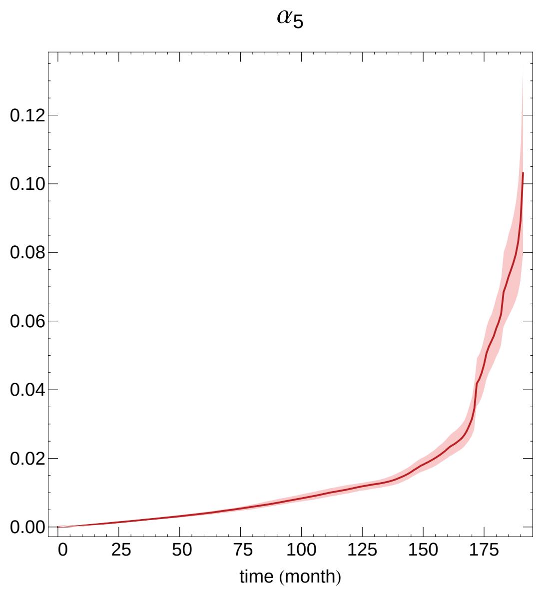
(a) α_2 public detection in *major urban* areas.



(b) α_3 public detection in *other urban* areas.



(c) α_4 public detection in *defined boundary* areas.



(d) α_5 public detection in *rural balance* areas.

Figure 1: The median α values over time. The solid shades are the central 95% credible intervals of the 10000 samples considered.

Postal Address: 111 Barry Street
c/- The University of Melbourne
Victoria 3010 Australia

Email: enquiries@amsi.org.au
Phone: +61 3 8344 1777
Fax: +61 3 9349 4106
Web: www.amsi.org.au

Comparison of the number of mature nests over time

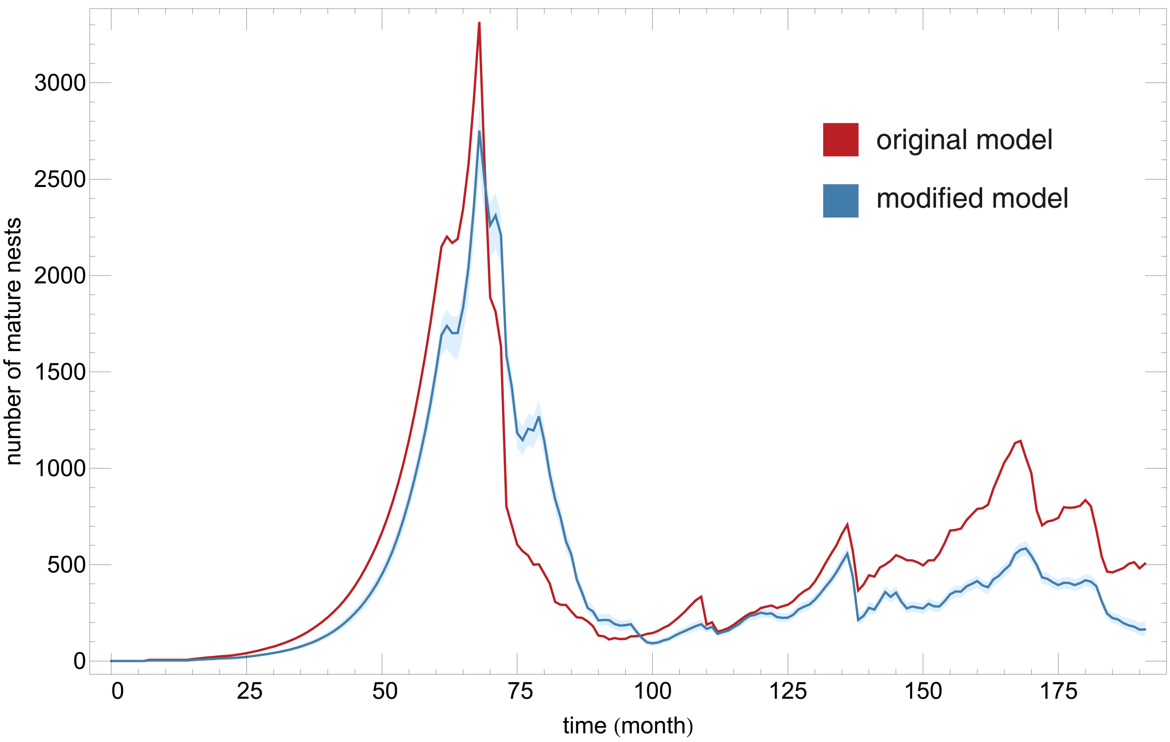


Figure 2: A comparison between the original model (red) and the modified model with time dependent public awareness (blue), of the number of mature nests over the months 0 to 191 (corresponding to January 1996 to December 2011). The blue shade represents the 95% credible interval of the number of mature nests determined in the modified model.

Postal Address: 111 Barry Street
c/- The University of Melbourne
Victoria 3010 Australia

Email: enquiries@amsi.org.au
Phone: +61 3 8344 1777
Fax: +61 3 9349 4106
Web: www.amsi.org.au

4 Materials and Methods

The data that was collected by the Biosecurity Queensland Control Centre, from 2001 to 2011.

A nest i is either observed or unobserved. Observed nests have known data of position (x_i, y_i) and month of death t_i. Month of death can be 0 to 191, corresponding to January 1996 to December 2011. If a nest is still alive by the end of December 2011, 9999 was assigned as the month of death. These values are unknown for the unobserved nests. All nests have an unknown month of founding (creation) f_i.

In the model, a nest has a discovery type d_i, with three possibilities. The first is discovery by the public, the second is discovery by targeted search and the third is the nest is undiscovered. We assume that professional searchers doing the targeted search have a constant probability of finding nests, denoted by α_1. The public's ability to detect ants related to the human population density in an area –in an urban area, there are more people and so it is expected that a nest there is more likely to be detected, than in a rural area. The detection values for the differing areas were α_2 for Major Urban, α_3 for Other Urban, α_4 for Defined Boundary and α_5 for Rural Balance. As part of including time dependence, α_2, α_3, α_4 and α_5 are all functions of time. α_0 is the probability that an unknown nest is destroyed by taking in toxic bait. Nests that are discovered are killed with probability 1.

There were two functions classifying the Brisbane area: $S(x, y)$ describing urban/rural land type given position (x, y); and $H(x, y)$ describing the habitat suitability for the fire ants. Both functions returned 1, 2, 3, or 4. It was assumed that there was no time dependence. Furthermore, there were two time dependent functions, $I_1(x, y, j)$ and $I_2(x, y, j)$ that returned 1 if the area $(x, y))$ was searched or treated during month j respectively, 0 otherwise.

Months 0 to 60, corresponding to January 1996 till January 2001 is the undetected growth phase in the model. No ant nests are detected during this time. The months 61 to 191, corresponding to February 2001 till December 2011 is the eradication phase. The eradication phase are four steps:

1. public search
2. targeted search
3. treatment step (treating areas with toxic bait and destroying discovered nests)
4. founding step (where mature nests can spawn more nests)

There is no founding step in December 2011 in the original and modified models.

Postal Address: 111 Barry Street
c/- The University of Melbourne
Victoria 3010 Australia

Email: enquiries@amsi.org.au
Phone: +61 3 8344 1777
Fax: +61 3 9349 4106
Web: www.amsi.org.au

4.1 Likelihood

The likelihood of the original model, as well as this modified model, is

$$\begin{aligned}\mathbb{P}\left(f,p,J,x,y,e,t,d|\lambda,\gamma,\sigma_X,\sigma_Y,\beta,\alpha,H,S'\right) = {} & \mathbb{P}\left(f,p|\lambda,t\right)\mathbb{P}\left(J|\gamma,\right) \\ & \times\mathbb{P}\left(x|p,J,\sigma_X\right)\mathbb{P}\left(y|p,J,\sigma_Y\right) \\ & \times\mathbb{P}\left(e|f,x,y,\beta,H\right)\mathbb{P}\left(t,d|f,x,y,\alpha,S'\right)\end{aligned}$$

with $S' = (S, I_1, I_2)$

The specific part of interest is $\mathbb{P}\left(t,d|f,x,y,\alpha,S'\right)$. For each nest i, we have $\mathbb{P}\left(t_i,d_i|f_i,x_i,y_i,\alpha,S'\right)$

With month j, let

$$\begin{aligned}W\left(t_i,f_i,x_i,y_i...\right) = {} & \left\{\prod_{j=\max\{61,f_i+6\}}^{t_i-1}\left(1-\alpha_{1+S(x_i,y_i)}\left(j\right)\right)\left(1-\alpha_1\right)^{I_1(x_i,y_i,j)}\right\} \\ & \times\left\{\prod_{j=\max\{61,f_i+1\}}^{t_i-1}\left(1-\alpha_0\right)^{I_2(x_i,y_i,j)}\right\}\end{aligned}$$

The probability that the public does not discover the ant nest is $\left(1-\alpha_{1+S(x_i,y_i)}\left(j\right)\right)$, and the probability that no targeted search discovers the ant nest is $\left(1-\alpha_1\right)^{I_1(x_i,y_i,j)}$. The product of these occur starting at $j = \max\{61, f_i + 6\}$ since month 61 is the first discoveries of the red imported fire ants nests, and it takes 6 months after founding f_i for a nest to be detectable. The probability that the ant nest does not take any of the toxic bait in the area is $\left(1-\alpha_0\right)^{I_2(x_i,y_i,j)}$. Since the eradication program only starts at month 61, the product starts at $j = \max\{61, f_i + 1\}$.

Thus, W is the probability of a nest surviving from month f_i to month $t_i - 1$.

For the total $\mathbb{P}\left(t_i,d_i|f_i,x_i,y_i,\alpha,S'\right)$, we then have to multiply in the probability of the type of death/discovery of that ant nest.

If a nest was reported by the public, $d_i = 0$, then

$$\mathbb{P}\left(t_i,d_i=0|f_i,x_i,y_i,\alpha,S'\right) = \alpha_{1+S(x_i,y_i)}\left(t_i\right)\times W$$

If a nest was found by targeted search, $d_i = 1$, then

$$\mathbb{P}\left(t_i,d_i=1|f_i,x_i,y_i,\alpha,S'\right) = \alpha_1\left(1-\alpha_{1+S(x_i,y_i)}\left(t_i\right)\right)\times W$$

If a the nest was not detected, $d_i = 2$, but nonetheless died by taking in toxic bait, then

$$\mathbb{P}\left(t_i,d_i=2|f_i,x_i,y_i,\alpha,S'\right) = \alpha_0\left(1-\alpha_{1+S(x_i,y_i)}\left(t_i\right)\right)\left(1-\alpha_1\right)^{I_1(x_i,y_i,t_i)}\times W$$

Postal Address: 111 Barry Street
c/- The University of Melbourne
Victoria 3010 Australia

Email: enquiries@amsi.org.au
Phone: +61 3 8344 1777
Fax: +61 3 9349 4106
Web: www.amsi.org.au

If the the nest was not detected $d_i = 2$, and not killed by the end of the data set $t_i = 9999$, then

$$\mathbb{P}\left(t_i = 9999, d_i = 2 | f_i, x_i, y_i, \alpha, S'\right) = \left(1 - \alpha_{1+S(x_i,y_i)}\left(t_i\right)\right)\left(1-\alpha_1\right)^{I_1(x_i,y_i,t_i)}\left(1-\alpha_0\right)^{I_2(x_i,y_i,t_i)} \times W$$

4.2 Priors

The prior densities remain the same for all other parameters, as detailed in the original paper [4]. For simplicity, we also take Uniform$[0, 1]$ priors for all the different $\alpha_k\left(j\right)$ for $k = \{2, 3, 4, 5\}$.

4.3 Conditional Posterior Density

The posterior distribution for the public detection probabilities α_k, for $k = 2, 3, 4, 5$ at month j, is

$$\mathbb{P}\left(\alpha_k\left(j\right) | e, t, d, f, x, y, S'\right) \propto \alpha_k\left(j\right)^{m_{k,j}}\left(1 - \alpha_k\left(j\right)\right)^{n_{k,j}}$$

where $m_{k,j}$ is the number of nests found by public search in month j and $n_{k,j}$ is the number of failed public searches in month j and where e, t, d, f, x, y, S' are some other parameters.

4.4 Constraints

The following constraints were applied in the modified model:

$$\alpha_2 \geq \alpha_3 \geq \alpha_4 \geq \alpha_5$$

for all time j, and

$$\alpha_k\left(j\right) \geq \alpha_k\left(j_0\right)$$

for all time $j > j_0$ and for $k = 2, 3, 4, 5$. This is based on the assumption that detection is more likely in urban areas than rural, and that detection ability would increase over time.

Postal Address: 111 Barry Street
c/- The University of Melbourne
Victoria 3010 Australia

Email: enquiries@amsi.org.au
Phone: +61 3 8344 1777
Fax: +61 3 9349 4106
Web: www.amsi.org.au

4.5 Sampling the distribution

The posterior distribution was sampled using MCMC, Markov Chain Monte Carlo. Specifically, the generalised Gibbs sampler was used, since there is an unknown number of parameters. [5]

The original code was written in C/C++ by Dr Jonathan Keith. The expansion to allow for time dependent public detection values was done by Thao Le.

The code was then run using the same data as in the original paper by Keith and Spring (2013), on the Monash Campus Cluster and took 12 days to generate 40000 samples.

5 Acknowledgements

I would like acknowledge and thank my supervisor Dr Jonathan Keith, for providing time for this project. Furthermore, I am grateful to AMSI for providing the scholarship, along with CSIRO for hosting the Big Day In.

Thao Le received a 2013/14 AMSI Vacation Research Scholarship.

References

[1] Lowe, S., Browne, M. and De Porter, M. (2000). *100 of the World's Worst Invasive Alien Species: A selection from the Global Invasive Species Database*, Invasive Species Specialist Group (ISSG), Auckland, New Zealand.

[2] Jennings, C. *Notes from the field: A brief history of the red imported fire ant eradication program*, The Australian Journal of Emergency Management, 19(3): 97-100.

[3] Taylor, J. (2008). *Once bitten, Stateline.* (ABC TV: http://www.abc.net.au/stateline/qld/content/2006/s2234162.htm)

[4] Keith, J. and Spring, D. (2013). *Agent-based bayesian approach to monitoring the progress of invasive species eradication programs*, Proceedings of the National Academy of Sciences. doi: 10.1073/pnas.1216146110

[5] Keith, J.M., Kroese, D. and Bryant, D. (2004) *A generalized Markov sampler*, Methodology And Computing In Applied Probability, 6(1): 29-53.