Modelling orange hawkweed distribution

Nick Beeton

5 July 2019

Executive Summary

This report introduces a modelling approach to identify potential areas where orange hawkweed might occur in Tasmania's Central Plateau region, which can then be surveyed. The model uses available data to estimate which methods of spread are important (e.g. via wind, roads or waterways), then makes predictions on where new incursions of the weed are likely to be detected based on four current infestations (Shannon, Butlers Gorge, Tarraleah and Miena). Not surprisingly, the weed is most likely to be detected close to the boundaries of current infestations, and the model suggests that rapid long-range dispersal is unlikely. However, areas downwind via the prevailing winds (north-east to south-east), along roads and near waterways are all at least slightly more likely to be infested, providing a good starting point for survey efforts.

This model is ***preliminary*** and some of the results are ***likely to be unreliable*** until further analyses and checks can be performed. Work on improving the model, getting more expert information to inform the model, as well as collecting and incorporating more observational data into the model, are all likely to provide much clearer estimates.

1 Introduction

Orange hawkweed (OHW) is an invasive weed, present in the Australian Alps and in parts of Tasmania; most notably, the Central Plateau region and Hobart surrounds. It is considered a "sleeper weed": though it may currently seem benign, based on experience in other countries (e.g. New Zealand) it can potentially spread rapidly and without warning, causing serious damage to the native environment and agriculture. As such, early intervention via monitoring and eradication is vital.

The modelling presented in this report is designed to identify potential areas where the weed might occur, which can then be surveyed and managed. While some literature exists on OHW spread, the mechanisms of spread and their potential importance are not well understood. As such, we use observation data to estimate the relative importance of some potential modes of spread, then use these to inform model predictions. Due to time constraints, this model is only intended to be preliminary. In particular, the Markov Chain Monte Carlo process used to estimate the probability distributions of parameters is difficult to correctly calibrate, so results are likely to be unreliable. The model is thus best treated as a proof of concept pending further investigation.

2 Model summary

For this exploratory model, we have developed a simple mechanistic modelling framework for the spread of OHW at local scales, applied to several known incursions in Tasmania's Central Plateau region (see Figure 1). The known incursions are modelled independently. Each incursion is modelled on a spatial grid with 20m x 20m resolution. We have based our model framework on the available data and expert opinion on the modes of spread. Our spatio-temporal model relies on four distinct parts: initial distribution, potential distribution, population dynamics and statistical modelling.

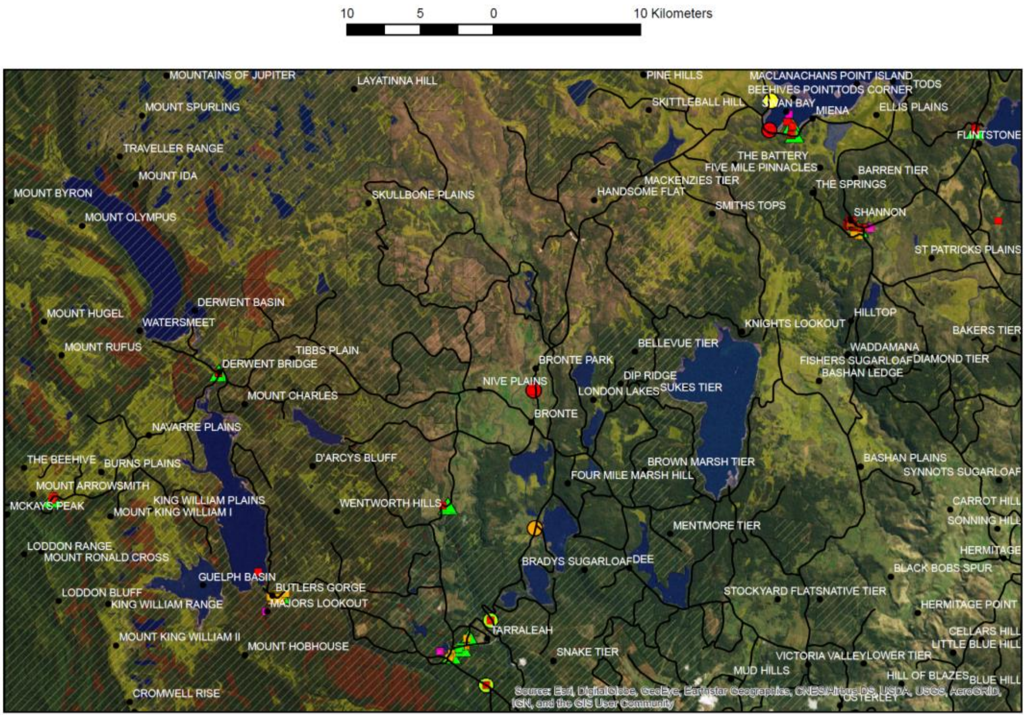


Figure 1 Map of the Central Plateau with OHW incursion sites. Triangles and circles represent locations of surveyed locations and smaller squares represent Natural Values Atlas observations, with colour denoting time of observation: pre-2011 in pink, 2011-2012 in red, 2013-15 in orange, 2016 in yellow, and 2017 onwards in green.
Roads are shown in black, water blue, and diagonal hatching represents reserve areas. Yellow and red areas represent grasslands and areas containing leptospermum respectively, both areas identified as potential OHW habitat.

Source: [ArcMap, Kathy Van Dullemen, Natural Values Atlas, TASVEG 3.0]

2.1 Initial distribution

The only data that were immediately available for estimating the distribution of early incursions of OHW came from observations from the Natural Values Atlas (NVA). We arbitrarily started our modelling in 2011, assigning any 20m x 20m cells that contains an observation at or before this time a 100% density of OHW and 0% elsewhere. In subsequent years in the model (see below for description), any cell that contains an observation for that year is also set to 100% density.

2.2 Potential distribution

Given the time and data constraints, we decided to apply the precautionary principle and assume that OHW will eventually populate any cell that doesn't contain water to 100% density.

2.3 Population dynamics

Based on knowledge of the species and expert opinion, we decided to model four potential mechanisms of OHW growth and spread:

- Wind-dispersed spread: we used the Bureau of Meteorology's wind rose data for the Shannon HEC station averaged over the summer months (9am and 3pm, December-

February) to determine the strength of each direction of spread in the 8 compass directions (call this $W(\theta)$ where θ is the direction of spread). Other directions were interpolated using a periodic spline. The resulting density of surrounding OHW based on a population at 100% density in a given cell, at each distance d from the cell, was calculated based on a dispersal kernel (Williams et al. 2007, page 32):

$$N(d,\theta) = \left(1 - \exp\left(-\left(\frac{d}{\boldsymbol{d_{av}}}\right)^{\boldsymbol{\alpha}}\right)\right)\left(\frac{W(\theta) + \boldsymbol{s}}{1 + d}\right) \tag{1}$$

with parameters $\boldsymbol{d_{av}}$ and $\boldsymbol{\alpha}$. This spreading mechanism was applied proportionally at each cell (e.g. a cell with 50% density would have half the spread).

- Additional dispersal by other mechanisms that are non-directional (animals, foot traffic, etc) with strength $\boldsymbol{s}$ (as seen in the formula above) – this is added to the wind-dispersed spread.
- Additional dispersal via any road or water cells with multipliers $\boldsymbol{r}$ and $\boldsymbol{w}$ respectively. These were applied for each cell using the algorithm

$$N^* = N_t + (N_{t+1} - N_t)(1 + \boldsymbol{r}\, R + \boldsymbol{w}\, W) \tag{2}$$

where N_t and N_{t+1} are the densities before and after the above dispersal mechanisms are applied; R is 1 in cells with roads and 0 elsewhere; W is 1 in cells with water and 0 elsewhere; and N^* is the resulting final dispersal.

- Growth of each cell's local population N towards the predefined carrying capacity within each cell using a logistic model with a given growth rate $\boldsymbol{k}$ over time t, measured in years**:**

$$\frac{dN}{dt} = kN(1 - N) \tag{3}$$

This was applied after the occurrence of dispersal each year.

2.4 Statistical modelling

The six parameters given in the above section – $\boldsymbol{d_{av}}, \boldsymbol{\alpha}, \boldsymbol{s}, \boldsymbol{r}, \boldsymbol{w}$, and $\boldsymbol{k}$ – are unknown, and so need to be estimated. We attempt this by testing how effective combinations of potential parameters are at predicting the known distribution of OHW at each site in January 2019 (supplied by Kathy Van Dullemen). To perform this test, we group (or "bin") cells based on their predicted density and estimate the probability of OHW presence or absence at those cells based on how many are actually present or absent in the group. For example, of the 200 cells with a predicted density between 0.1 and 0.2, OHW is present in 30 of these and absent in the remaining 170. Thus we predict cells with those predicted densities have a 30/200 = 15% chance of containing OHW.

After we convert each cell's density to a corresponding probability of presence in this way, a likelihood $\mathcal{L}$ is calculated of the data occurring given our parameters:

$$\log \mathcal{L} = \gamma \left(\textstyle\sum_{cells}(P \log p + (1 - P)\log(1 - p))\right) \tag{4}$$

where P is 1 in cells where OHW is present in 2019 and 0 otherwise, and p is the modelled probability of presence at each cell. We introduce a parameter γ which attempts to compensate for spatial autocorrelation (the fact that nearby cells are clearly related to each other). This parameter has been set by examining the typical distance and area over which spatial autocorrelation occurs in the models using a variogram (about 120m or 6 cell-length radius, so an area of about $36\pi \approx 113$ cells) and correspondingly setting $\gamma = 0.01$ to represent the lower contribution of individual cells to the overall likelihood.

This likelihood is then used in a Markov Chain Monte Carlo framework with a uniform prior (with 0 minimum for each parameter, and different maxima based on experimentation – see the later plots for these) to generate a Bayesian posterior distribution of parameters with 20,000 samples for each site. We then use this data to generate:

- Posterior estimates of probability of OHW presence in each cell for 2019

This gives an indication of how well the model is performing in predicting the known distribution in 2019. This can be assessed visually by examining how close the model prediction looks to the known distribution, or numerically by calculating the area under a Receiver Operating Characteristic (ROC) curve, which assesses the model's performance in classifying presence versus absence over a range of different cutoff values (e.g. setting anything under 50% probability as a predicted absence, and above as a predicted presence).

- Posterior estimates of probability of OHW presence in each cell for 2020

We then use the known distribution in 2019 as our new model starting point – setting density at 100% in present cells and 0% in absent cells – and run the model forward a year using our range of parameter estimates to generate a Bayesian estimate of OHW density in 2020. We can then use the same process as described above to convert this density to probability of presence, obtaining an estimate of the most likely places to find OHW outside of the known distribution.

3 Model results

3.1 MCMC performance

A Metropolis-Hastings MCMC was run for 10,000 samples for 2 separate chains for each site. A Gelmans-Rubin-Brook convergence diagnostic gave estimates between 1.00 and 1.19 (where the ideal score is 1), suggesting reasonable convergence. However, some 95% confidence intervals were as high as 1.68, so caution is recommended in interpreting these results. In addition, most posterior parameter estimates covered the range of the priors – while the priors were conservatively set to cover most biologically reasonable scenarios, extending these may give more accurate and detailed results.

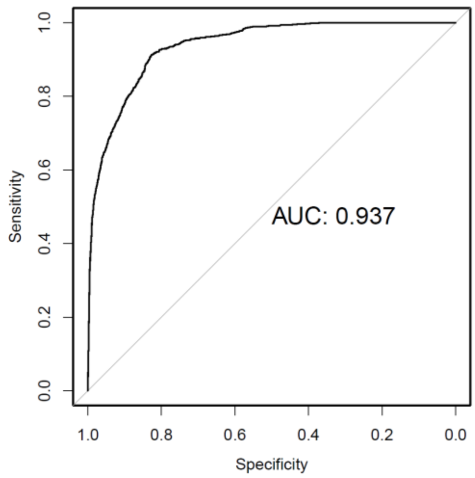
3.2 Description of figures

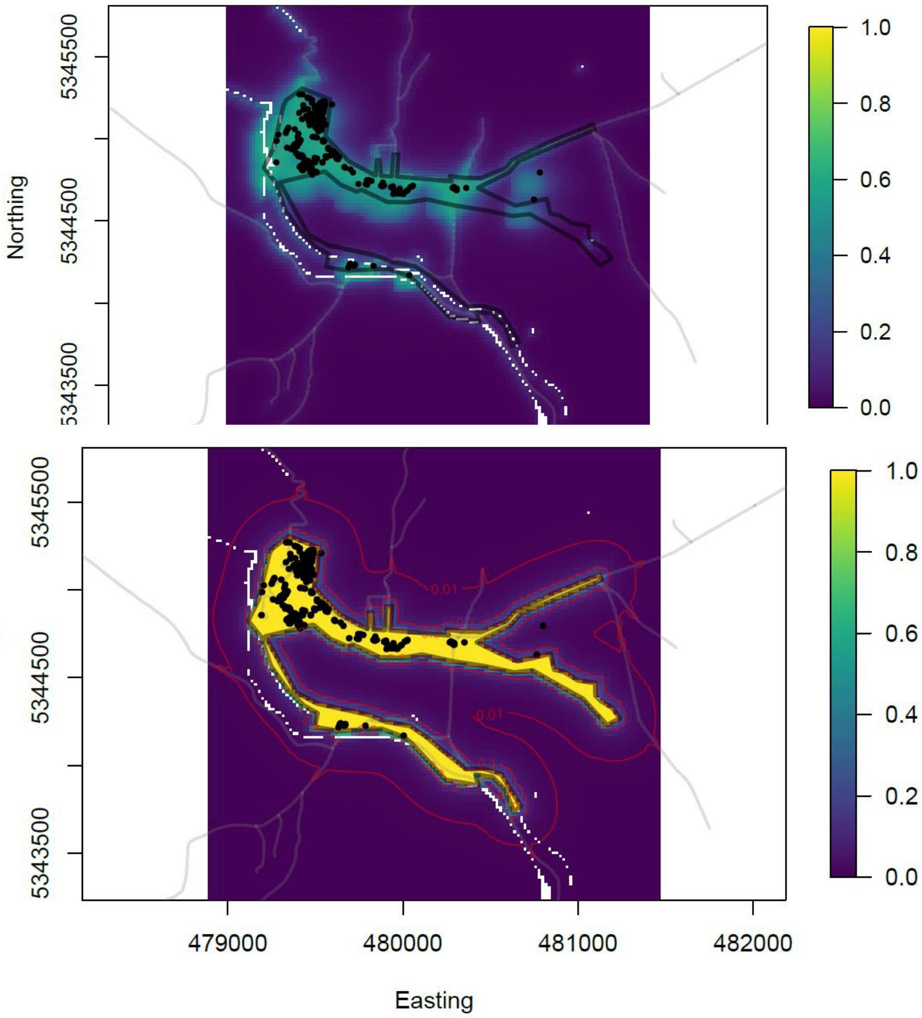
For each site, we give the corresponding model's ROC curve along with corresponding AUC (Area Under Curve) score – this gives a relative measure of how well each model did (1 being a perfect predictor, 0.5 being no better than random chance).

We then give the 2019 posterior probability estimates as described above, with the actual polygon of presence outlined in black, roads shown in grey, and any water features shown in white. Probabilities are graded from purple for 0% to yellow for 100%.

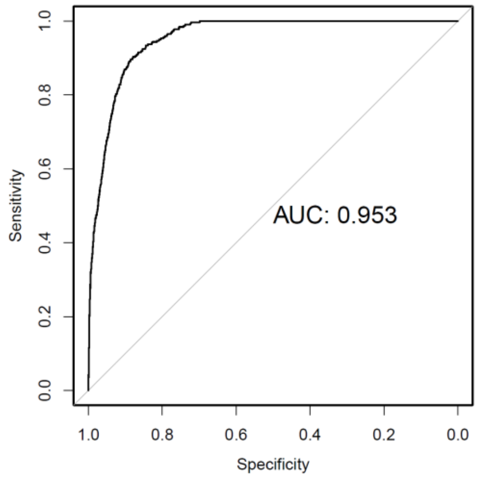
Below that, we give the predicted 2020 probability estimates, with contours to emphasise where the model predicts 0.5, 0.1 and 0.01 (or 50%, 10% and 1%) probability of presence respectively.

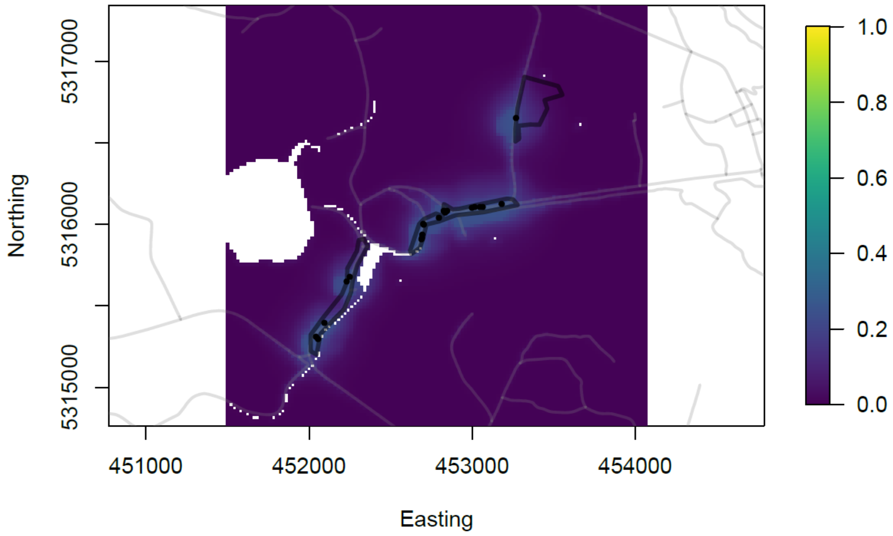
3.3 Shannon

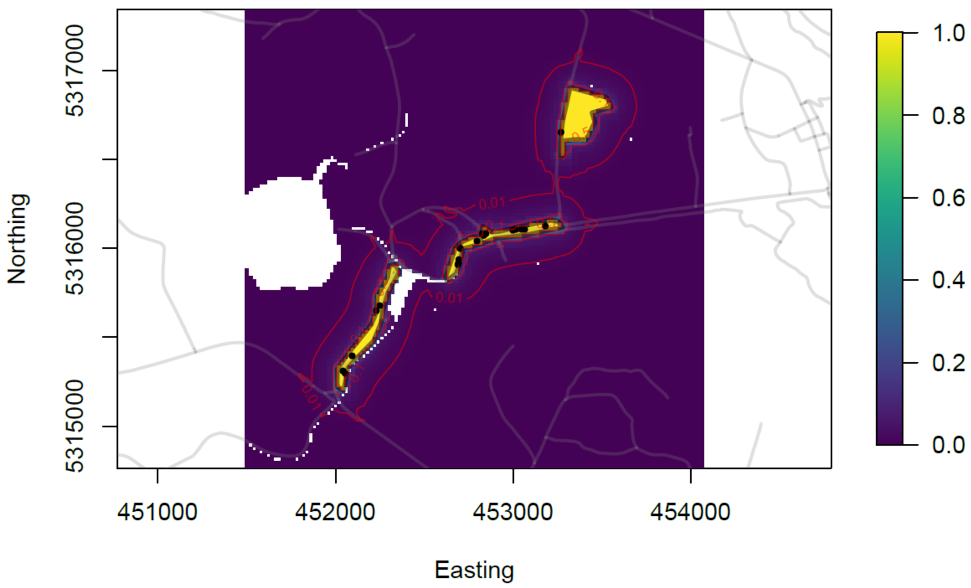




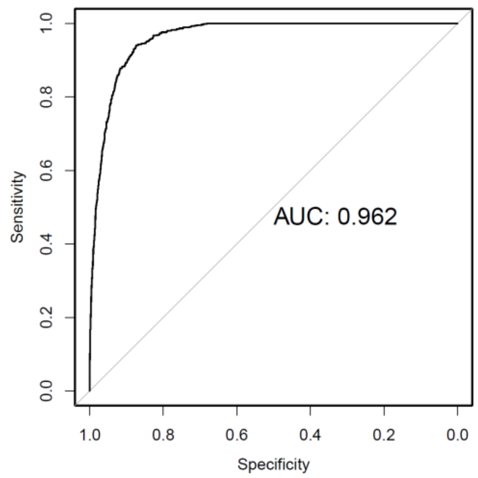
3.4 Tarraleah

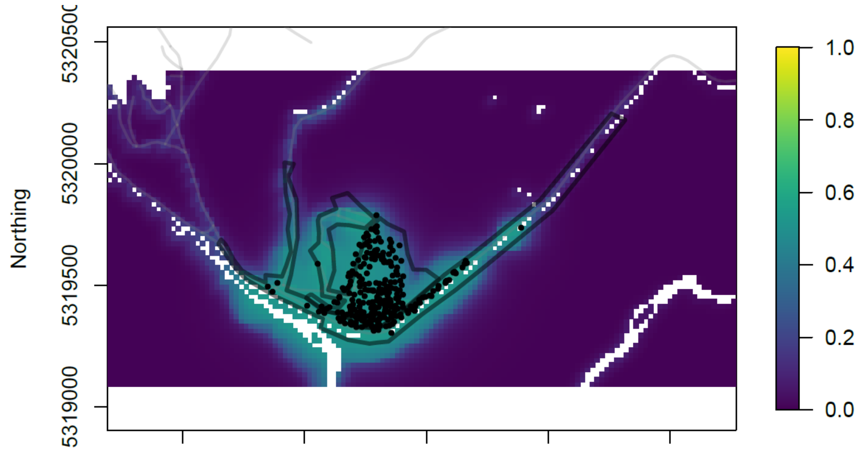


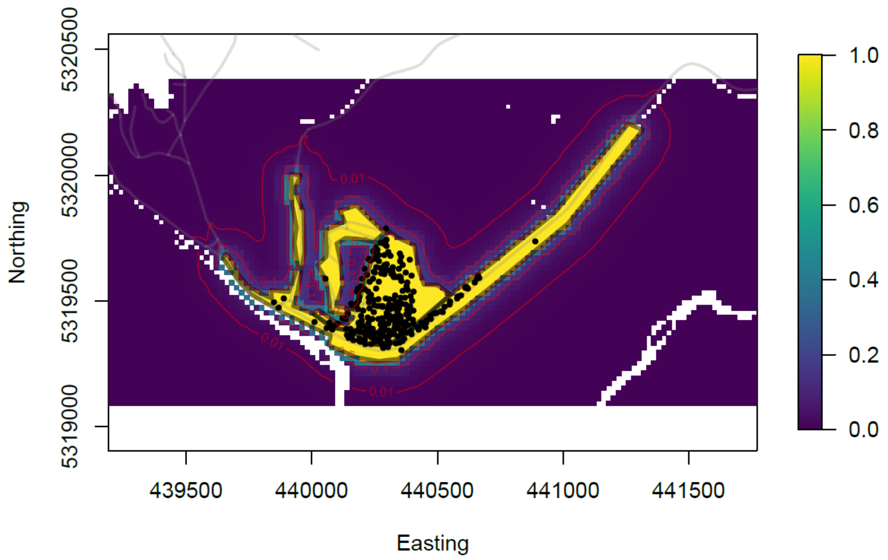




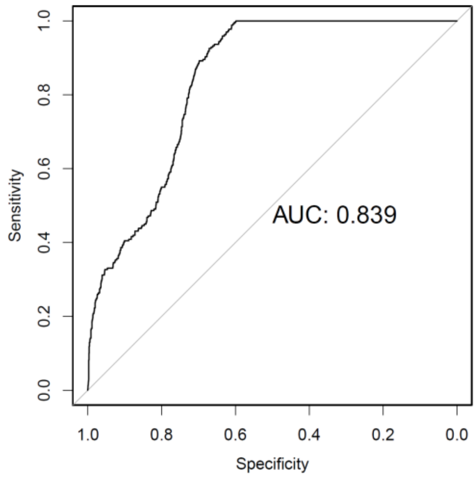
3.5 Butlers Gorge



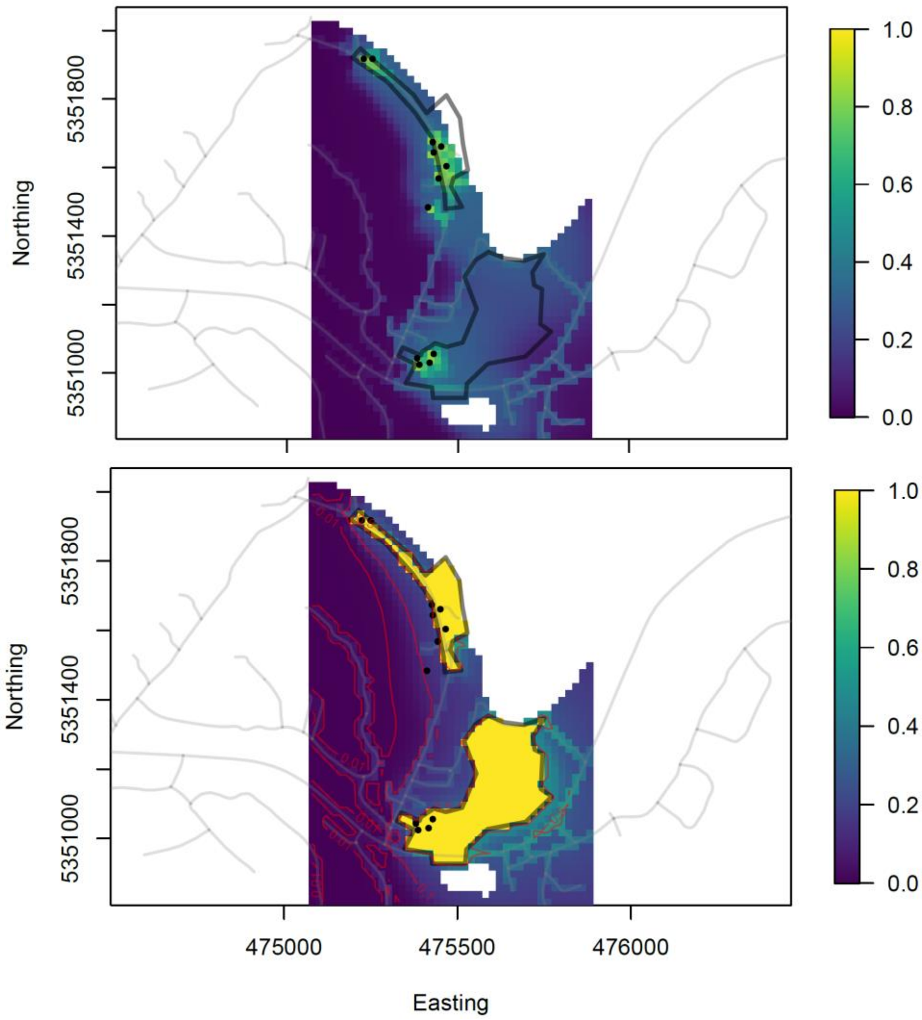




3.6 Miena



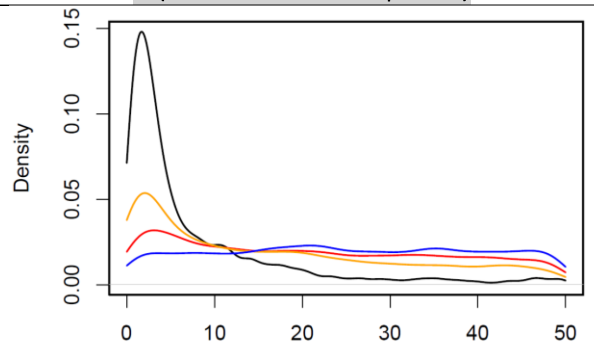
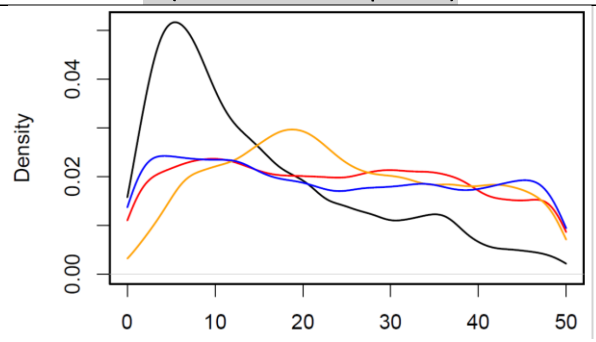
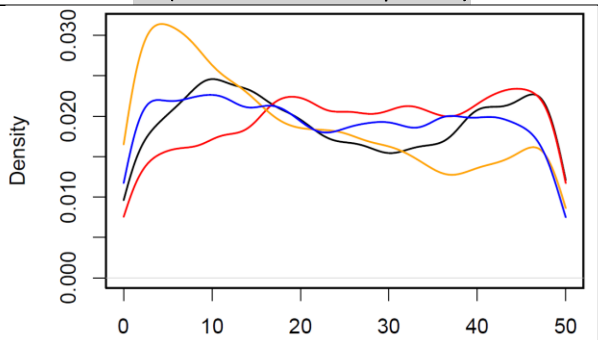
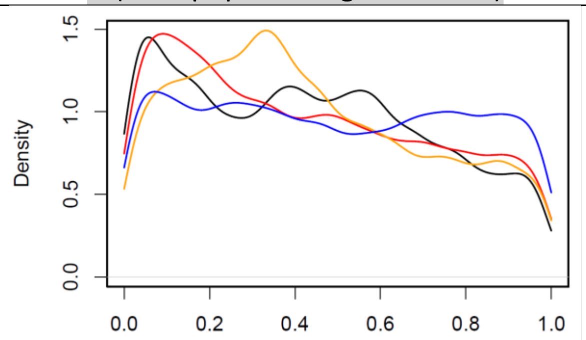
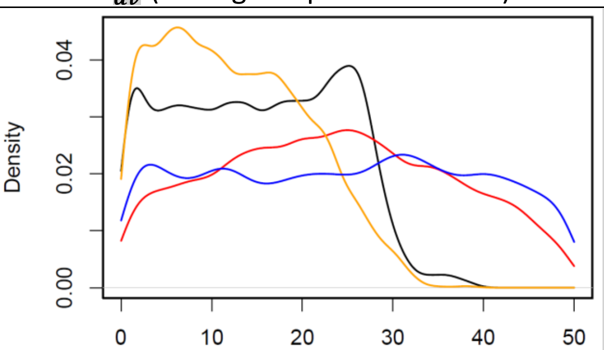
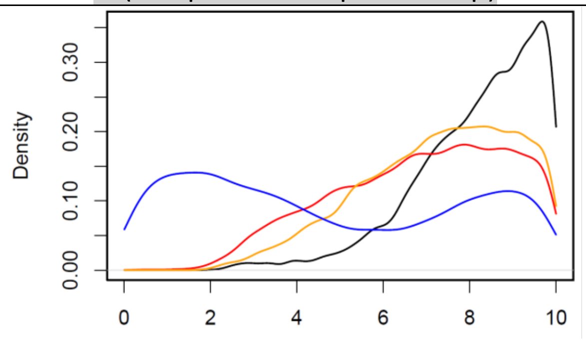
Sensitivity
Specificity
AUC: 0.839
1.0
0.8
0.6
0.4
0.2
0.0



Northing
Easting
5351000
5351400
5351800
475000
475500
476000
1.0
0.8
0.6
0.4
0.2
0.0

3.7 Parameter Estimates

These density plots show the result of the 20,000 Markov Chain Monte Carlo simulations separately for each parameter (labelled) and at each site (coloured **black for Shannon, red for Tarraleah**, **orange for Butlers Gorge**, and **blue for Miena**).

$\boldsymbol{s}$ (non-directional spread)	$\boldsymbol{r}$ (road-based dispersal)
	
$\boldsymbol{w}$ (water-based dispersal)	$\boldsymbol{k}$ (local population growth rate)
	
$\boldsymbol{d_{av}}$ (average dispersal distance)	$\boldsymbol{\alpha}$ (steepness of dispersal drop)
	

4 Discussion

In all but the case of Miena, most of the surveyed range of OHW was already covered by NVA observational data. This has inflated the model's apparent performance in both the AUC scores and the plots. Examining its behaviour in regions that have not been observed in the NVA data gives a more reliable assessment of its performance.

Many of these areas are along roads, as can be seen in Shannon and Butlers Gorge in particular. The model for Shannon predicts that road-based dispersal is likely to be limited in extent – namely, the model performs badly if roads have more than about a 20 times dispersal rate. This still allows for some road-based dispersal, supported by at least one road that it has clearly travelled along after the initial infestation, but reflects the fact that OHW has also not travelled along many of the other available roads. No clear evidence either way exists for the other sites – the plots for these are mostly flat, similar to the uniform prior, so the data has not changed these substantially.

Similarly, non-directional spread is also limited in the Shannon model - the model performs badly when this is greater than about 10 (units are percent of maximum possible). For reference, the strongest wind (north-westerly) is 11 percent, so this suggests that the mechanism of non-directional spread is weaker than wind-dispersed spread. Again, the other sites have less or no clear evidence of this.

The shape of the dispersal curve is also clearly important: Shannon and Butlers Gorge in particular show evidence of the importance of a smaller dispersal distance and a steeper drop in dispersal, suggesting that substantial long-range dispersal is not supported by the available evidence.

The other mechanisms have less clear support in the models based on the available data. There seems to be some suggestion of limited water-based dispersal in Butlers Gorge, possibly due to the fact that OHW has travelled along some watercourses but not others there (similarly to roads in Shannon). Finally, most sites seem to suggest that smaller local population growth fit slightly better to the model, but the signal is again not strong.

5 Conclusions

Running the model independently at different sites showed us that there is not sufficient data to get a clear overall picture of the effects of all mechanisms, as these varied strongly between sites. However, given the apparent overall slow spread of OHW on a spatial basis (as suggested by the fact that most surveyed locations were already mostly colonised), the model still may be of use in predicting localised spread away from the surveyed infestations.

While the model produces the obvious result that the closer to the existing infestations, the more likely we are to discover a new colony, it does to a greater or lesser extent predict higher probability of detection near roads and waterways, and downwind of the prevailing winds. This gives a good starting point for surveying – for example, south east of the south-eastern infestation in Shannon, in between the west-most two infestations at Tarraleah, around the local roads in Butlers Gorge, and east of the existing large infestation at Miena around the roads. The results of these surveys can then in turn inform further modelling, giving a better overall picture of the effects of each mechanism on spread.

6 Future work

This model is highly preliminary, so there is much that can be done to improve it. Most importantly, the MCMC process was run for a very small number of iterations, and convergence testing suggested that a much higher number of iterations would be required to provide reliable results. While even this small run did provide us with some signals, a clearer picture would likely emerge with increased power.

More exploratory testing of the model would provide a greater understanding of the interplay between mechanisms, and possibly allow the reworking of the model to decrease unnecessary correlation between parameters, or even a reduction in the number of parameters, making for faster analyses.

The model itself could readily be improved by such things as examining other mechanisms of spread (such as particular vegetation communities), tweaking the current mechanisms (e.g. differentiating between more and less popular roads, or between types of watercourses e.g. rivers vs lakes), developing a more complex picture of suitable habitat using niche modelling, or searching for other sources of data that might allow analysis from a different angle.

7 Acknowledgements

I would like to thank Cindy Hull (and the staff at NRM South) for approaching me to undertake this project and for her patient support throughout, Mark Wapstra for a useful discussion about vegetation (which may appear in later model iterations), and Geoff Hosack of CSIRO for a quick, thorough and helpful internal review.

CONTACT US

t 1300 363 400
+61 3 9545 2176
e csiroenquiries@csiro.au
w www.data61.csiro.au

WE DO THE EXTRAORDINARY EVERY DAY

We innovate for tomorrow and help improve today – for our customers, all Australians and the world.

WE IMAGINE
WE COLLABORATE
WE INNOVATE

FOR FURTHER INFORMATION

Nick Beeton
Research Scientist (Applied Biological Modeller)
t +61 3 6232 5318
e nick.beeton@data61.csiro.au
w www.data61.csiro.au