Testing model accuracy when predicting pest susceptibility using expert-driven spatially-explicit models

30th August 2017
Justine Murray, Javier Navarro Garcia and Rieks van Klinken
Invasive species and pests

• Invasive species cause significant damage to the environment and economic assets
• Damage to productivity costs the agricultural sector millions of dollars each year.
• Adversely affect ecosystem function, e.g. destruction of habitat, competition
• Reduce agricultural productivity, e.g. crop damage, predation
• Threaten animal and human health and well-being, e.g. disease transmission, adverse human-pest interactions

Effective pest management benefits from scientifically-based recommendations
The problem........

...inconsistent results because data is not normally distributed.

Histogram of Data Values

<table>
<thead>
<tr>
<th>Calculation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>22.0</td>
</tr>
<tr>
<td>Min</td>
<td>8</td>
</tr>
<tr>
<td>Max</td>
<td>63.0</td>
</tr>
<tr>
<td>Range</td>
<td>55.0</td>
</tr>
<tr>
<td>Mode</td>
<td>22</td>
</tr>
<tr>
<td>Median</td>
<td>14.6</td>
</tr>
<tr>
<td>Std Dev</td>
<td>18.7</td>
</tr>
<tr>
<td>Variance</td>
<td>338</td>
</tr>
<tr>
<td>Coefficient</td>
<td>1.544 acres</td>
</tr>
</tbody>
</table>

22.0 - 18.7 = 3.3

40.7 = 22.0 + 18.7

68.3%

-3 SD       -2 SD       +2 SD       +3 SD

2.1%         13.6%       34.1%       34.1%       13.6%       2.1%

COMMUNICATION BREAKDOWN

COMMUNITY

SCIENCE
Increasing ‘buy-in’ for best management practices targeting pests
Modelling economic and environmental costs of pest species with different pest management strategies

New incentives are needed to encourage increasing adoption of effective strategies to pest management. Quantifying economic impacts of pests may influence solutions and greater participation in broader scale management strategies.

Introduction

Land management is an ongoing process for agricultural and natural environments with the focus on maximum benefit for strategic costs, especially with increasingly limited available resources. Developing models can be an integral part to the understanding of the functioning of a system (Murray et al., 2015). However, management models often fail short in users’ interpretation to end-users, especially if the primary target are land managers without statistical backgrounds. There is a need to convey the information in a manner that promotes literacy and willingness to participate through a sense of community empowerment and inclusivity.

The way to accomplish this is through engaging stakeholders through the entire model-building process. This emboldens ownership of the model from the initial concept to the end-product.

Methodology

- We used a Bayesian Network (BN) to incorporate the model framework consisting of three model components: habitat suitability, management and agricultural production (Figure 1).
- Potential pest impact, a proxy for potential pest density, was determined by experts from suitability of the habitat, a distance threshold and the management activities that are being undertaken. The distance threshold accounted for foraging behavior within a set distance away from suitable habitat, in this case 100 m for rabbits.
- Agricultural production loss (in percentage) was determined by experts given the potential pest impact and the commodity being produced, both of which can vary seasonally. The experts were asked to identify a production loss range with a mean and standard deviation, to capture the potential variability in loss.

Results

- The economic impact model allows the user to choose pest management options, individual agricultural commodities and land use to get an immediate economic response as to the effect on the agricultural loss and the resultant income loss.
- The agricultural loss alters with different management scenarios, thereby affecting the income lost from pest damage and ultimately net income (after also accounting for production costs and pest management costs) (Figure 1).
- To really engage the farming community on the impacts of a pest population, scenarios were run for major individual commodities to assess the potential impact on agricultural loss and overall income loss under habitat suitability was high (Table 1).

Table 1: Management scenarios for some major agricultural commodities, given high habitat suitability allowing for both losses, income costs that impact damage and costs to control after damage.

<table>
<thead>
<tr>
<th>Management Strategy</th>
<th>Average Agricultural Loss (€/ha)</th>
<th>Average Damage Loss Pest (€/ha)</th>
<th>Average Income after Damage &amp; Management Costs (€/ha)</th>
</tr>
</thead>
<tbody>
<tr>
<td>No Management</td>
<td>70-90</td>
<td>15,000 – 25,000</td>
<td>0-100</td>
</tr>
<tr>
<td>Fencing</td>
<td>50-50</td>
<td>7,800-15,000</td>
<td>5,000-10,000</td>
</tr>
<tr>
<td>Integrated</td>
<td>0-20</td>
<td>2,000-4,500</td>
<td>20,000-20,000</td>
</tr>
</tbody>
</table>

Management implications

- Our spatially-explicit economic impact model is able to give land managers a finer scale understanding of the impact of pests on individual land use categories and commodities.
- The use of scenarios allows land managers to weigh the relative risk of the different management options or the agricultural loss, income loss and resultant net income for each commodity and thereby more informed to make an economically sound decision towards pest management on their land.
- Revisiting the proximity of pest habitats to cultivation and conservation will hopefully drive community engagement towards effective integrated management. This tool endeavours to increase community engagement with participation in broad scale pest management planning.
Spatially-explicit risk modelling

• A risk map is a visual representation of the results of an analysis of potential risk within a landscape
• Risk maps allow multiple stakeholders to engage in the process
• Risk maps are not limited by the viewer’s statistical background
Expert elicitation

Benefits:
• One to multiple experts
• Encapsulating broad experiences over and management
• Quick to capture

Disadvantages:
• Individual vs group consensus
• Need to monitor for uncertainty
  – Natural variability (individual interpretations)
  – Linguistic uncertainty (confusion on definitions, context.....)
  – Epistemic uncertainty (limited or imperfect knowledge)
  – Anchoring (fixing on a set of beliefs)
  – Optimistic bias (tending toward positive outcomes)
  – Authority bias (being swayed by authority figure)
Predictive modeling: characterizing uncertainty

“Partitioning sources of error”

- Which environmental variables are most important? (sensitivity analysis)
- What uncertainty is there in conditional probability estimates?
- How sensitive are predictions to parameter uncertainty?
- Model structure (feedback & validation)

Sensitivity of 'Establishment'

<table>
<thead>
<tr>
<th>Node</th>
<th>Mutual Info</th>
</tr>
</thead>
<tbody>
<tr>
<td>Flood Frequency</td>
<td>0.094</td>
</tr>
<tr>
<td>Soil Type</td>
<td>0.074</td>
</tr>
<tr>
<td>Ground Cover</td>
<td>0.074</td>
</tr>
<tr>
<td>Grazing Management</td>
<td>0.014</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Environmental Variables</th>
<th>Establishment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Soil Type</td>
<td>Grazing management</td>
</tr>
<tr>
<td>High levels</td>
<td>Low levels</td>
</tr>
<tr>
<td>Good Good</td>
<td>90 % 10 %</td>
</tr>
<tr>
<td>Good Poor</td>
<td>30 % 70 %</td>
</tr>
<tr>
<td>Poor Poor</td>
<td>10 % 90 %</td>
</tr>
</tbody>
</table>
Validation

**Best:** balanced data set containing presence/absence covering all environmental gradients

However, we mostly only have access to records of pest presence or damage.
Testing for model accuracy

Classification and regression trees
- Robust
- Deals with different data sources

Testing model performance;
- R-squared plot
  - based on the variance explained
  - Comparison of tree size and split accountability against the size of the tree to determine which splits accounted for most of the data when varying the penalties for FP and FN errors.
- Cross-validation plot:
  - 10-fold cross-validation to estimate goodness of fit
  - uses the rsq.rpart function in Rpart.
Confusion matrix

• the best model predicts the probability of presence for the species in relation to presence and absence at actual sites ($P_{fair}$).
• threshold value for predicting the probability of presence where false positive (FP) and false negative (FN) error rates are equal.
Testing for model discrimination ability

ROC – AUC test
• Sensitivity (true positive rate)
• Specificity (true negative rate)
• Values of AUC were considered to reflect poor (0.5–0.7), acceptable (0.7–0.8), excellent (0.8–0.9) or outstanding (>0.9) model accuracy

Area under the ROC curve (AUC) 0.878
Standard Error 0.0311
95% Confidence interval 0.833 to 0.915
z statistic 12.162
Significance level P (Area=0.5) <0.0001

Youden index J 0.7768
Associated criterion >0.5
Sensitivity 97.95
Specificity 79.73
Generalised Linear Model

|                | Estimate | Std Error | Z Value | Pr(>|z|) |
|----------------|----------|-----------|---------|----------|
| (Intercept)    | -10.9073 | 2.8151    | -3.875  | 0.000107 |
| Water          | 1.1101   | 0.5264    | 2.070   | 0.038488 |
| Food           | 1.2247   | 0.4329    | 2.829   | 0.004665 |
| Shelter        | 1.5773   | 0.5078    | 3.106   | 0.059837 |
| Seclusion      | -0.9141  | 0.2962    | -3.086  | 0.016845 |
| Expected Value | 10.1305  | 1.2036    | 8.417   | 0.007962 |

Normal Q-Q Plot

Sorted residuals vs Normal scores

Residuals distribution

Expected Value
Suitability
Mapping classification errors
Summary

• All models results should be validated
• Getting the range of belief when collecting expert knowledge can identify the uncertainty around variables
• Pseudo-absences can be used to balance presence-only pest records
• Using a number of different statistical tests and running associated diagnostic plots can be used to assess discriminative ability of model outcomes and see if results are similar.
Acknowledgements:

We would like to thank Sama Low Choy from Griffith University for help with statistical code.

Justine Murray
CSIRO Health & Biosecurity
Ph: +61 7 3833 5679
Justine.Murray@csiro.au
Acknowledging the multiple sources of uncertainty

- Model structure: ✔✔✔
- Parameter estimation: ✔✔✔
- Validation data: ✔ ✔
- Environmental layers: ✔