Estimating uncertainty in weed risk assessment predictions

Peter Caley[1,2], Paul Pheloung[2,3] and Mark Lonsdale[1]
[1] CSIRO Entomology, GPO Box 1700, Canberra, Australian Capital Territory 2601, Australia
[2] CRC for Australia Weed Management, PMB 1, Waite Campus, Glen Osmond, South Australia 5064, Australia
[3] Department of Agriculture, Fisheries and Forestry, GPO Box 858, Canberra, Australian Capital Territory 2601, Australia

Summary Using the Australian Weed Risk Assessment (WRA) model as an example, we demonstrate how screening scores arising from risk assessment models may be translated into probabilities. The approach estimates uncertainty arising from both the risk assessment model itself, and uncertainty in the likely base-rate probability for the characteristic being screened. The results confirm the high sensitivity of the posterior probability of weediness to the prior base-rate of weediness of plants subjected to screening. Results provide a quantitative estimate of the weediness probability posed by taxa classified using the WRA model, enabling bio-economic modelling to contribute to the decision process, should this avenue be pursued.

Keywords Risk assessment, bootstrapping, Bayesian, WRA, modelling uncertainty.

INTRODUCTION

Successfully predicting invasiveness is a difficult task, as it is widely accepted that the likelihood of an introduced organism making the transition to being invasive is low (e.g. Mack *et al*. 2000). For events that have a low prior probability of occurring (sometimes referred to as having a low 'base-rate probability'), predicted probabilities of occurrence based on the results of screening tests alone tend to substantially overestimate the true probability of the event occurring. The true probability of the event occurring in light of a screening test is obtained by 'revising' the prior probability using the screening test likelihood—for this reason it is often referred to as the posterior probability. The proportion of predicted events that would be expected to actually occur based on this posterior probability is referred to as the 'Positive Predictive Value' (PPV) of a test. For example, in the case of mammogram screening for breast cancer, the PPV could be in the order of 0.1 (Gigerenzer 2002). That is, an estimated 90% of women with positive mammograms do not have breast cancer—these cases are referred to as false positives. This high rate of false positives may at first glance invite criticism of the screening test; however a low PPV is not necessarily a result of a screening system being sub-standard, but more often a phenomenon of trying to predict uncommon events with imperfect discriminatory tests. This problem, sometimes referred to as the 'base-rate effect', also occurs within disciplines such as engineering e.g. earthquake forecasting (Matthews 1997), though has only more recently been addressed in issues of natural resource management e.g. weed risk assessment (Smith *et al*. 1999), or ecology e.g. predicting species occurrence (Manel *et al.* 2001). The exact value of the PPV depends heavily on the prior probability of the event in question, along with the sensitivity and specificity of the screening test, particularly the latter, and can be calculated by direct application of Bayes' Theorem.

One management approach for dealing with the imperfect nature of diagnostic tests and the resulting inaccuracies in prediction is the use of decision theory (Matthews 1997), and it has been suggested that weed risk assessment could be placed in this context (Smith *et al.* 1999). For such an approach, quantitative estimates of the probability of weediness are needed, along with estimates of the costs e.g. importing a weedy plant or preventing importation of a useful plant, and benefits e.g. preventing the importation of a weedy plant or introducing a useful plant, of different actions. Classification-based screening models such as the Australian Weed Risk Assessment (WRA) system (Pheloung *et al*. 1999) do not readily provide estimated probabilities of weediness. However, summary scores may be converted to predicted probabilities of weediness by using, for example, logistic regression (Hughes and Madden 2003). Unfortunately, if the base-rate effect is ignored, the resulting fitted probabilities of weediness in relation to WRA score are biased upwards, as the training dataset contained an unrealistically high proportion of weeds.

A decision theory approach requires not only the estimated probability of a taxon becoming invasive, but also the uncertainty around that probability. Uncertainty in predicted probabilities may come from several main sources. The first is model selection uncertainty, whereby differing models differ in their predictions, and there is uncertainty as to the correct model. The second source is the inherent variability in model predictions, which may be estimated from the statistical properties of the model being used, or by computer intensive methods such as bootstrapping (Efron and Tibshirani 1993), or a mix of both. Lastly, there may be uncertainty surrounding 'fixed' parameters within the model. For example, while it is clear that the performance of a screening test is highly sensitive to the prior probability of the event being predicted, in the case of biological invasions, this quantity is poorly characterised. In fact, the widely held assumption that only a very small proportion of introduced taxa will become invasive is not universally true e.g. pasture species (Lonsdale 1994), and the emerging high rate of naturalisation (a necessary precursor to a taxon becoming invasive) of non-indigenous plants (Duncan and Williams 2002) may indicate that in time this paradigm may change.

In this paper, using the WRA system as an example, we demonstrate how screening scores arising from a predictive model of weediness may be re-expressed as posterior probability estimates of weediness, including uncertainty around these estimates that reflect uncertainty in the prior probability of weediness. We achieve this by extending the logistic regression approach of Hughes and Madden (2003), utilising standard bootstrapping procedures to account for the base-rate effect, and using a Bayesian approach to incorporate prior uncertainty in the base-rate probability of weediness. In doing so, we demonstrate how a screening system used to predictively classify taxa into various weediness categories may be modified to explicitly estimate risk in a probabilistic manner.

MATERIALS AND METHODS

The Weed Risk Assessment (WRA) model The Weed Risk Assessment (WRA) model has been operational in Australia since 1996 as the second component of a three-tiered system aimed at identifying and preventing the entry to Australia of environmental and agricultural weeds. Briefly, the WRA model converts responses to questions relating to the plant's climatic preferences, biological attributes, reproductive and dispersal method into a score, whose value determines whether to 'accept' (WRA Score ≤0), 'further evaluate' (1 ≤ WRA Score ≤ 5), or 'reject' (WRA Score ≥6) the taxon.

Specifying the prior probability of weediness In a review of the available literature, Smith *et al.* (1999) considered the prior probability (or base-rate) of weediness of plants to range from 0.01% (Williamson and Fitter 1996) to 17% (Lonsdale 1994) with a likely value of 2%. Hence, as a prior distribution for modelling this uncertainty, we used a Beta distribution with parameters $\alpha = 1.62$ and $\beta = 31.4$, that correspond to a mode of 0.02, a mean of 0.05 and a 99% quantile of 0.17. The Beta distribution is the standard distribution used for modelling prior uncertainty in proportion data. In the current context it refers to the probability of a plant being presented for importation being a weed, independent of its subsequent WRA score.

Estimating posterior probability of weediness The original 370 taxa data set analysed by Pheloung *et al.* (1999) contains 286 species classified as weeds, and 84 species classified as non-weeds. As the WRA model is not a model in the statistical sense, variability around the predicted outcomes cannot be investigated by standard parametric means. The issue is further complicated by the need to incorporate the effect of the prior probability of weediness. However, bootstrapping (repeatedly resampling with replacement) provides a robust method of estimating this variability whilst accounting for the prior probability. We bootstrapped this data set 1000 times, with the probability of selection for weeds and non-weeds drawn from a Beta (1.62, 31.4) distribution. For each bootstrap sample, a logistic regression model was fitted, relating the predicted (posterior) probability of weediness to WRA score. We calculated the average of the model predictions, and associated lower and upper 95% uncertainty intervals for each prior probability of weediness.

RESULTS

Bootstrapped posterior probabilities of weediness as a function of WRA score are shown in Figure 1. The predicted probability was similar for both the logistic regression and raw bootstrap up until a WRA score of about 6, after which the raw bootstrap was no longer a smooth function. Regardless of the model examined, as a general trend, the probability of weediness started to increase sharply from a WRA score of about zero and upwards. The estimates of uncertainty around the predicted probabilities for the two models differed depending on WRA score. Uncertainty in the raw bootstrap was proportionally greater for WRA scores less than 7, with the reverse occurring for scores greater than 7. In fact, for the raw bootstrap, no variability was estimated around the predicted probability of weediness for WRA scores greater than 7 (other than 12 and 13), as weeds alone were assigned to these WRA scores (Figure 1b). In contrast, the logistic regression recorded substantial variability around the predicted probability of weediness for high WRA scores (Figure 1a).

Within the 'further evaluate' class (WRA scores 1–5), the probability of weediness as a function of

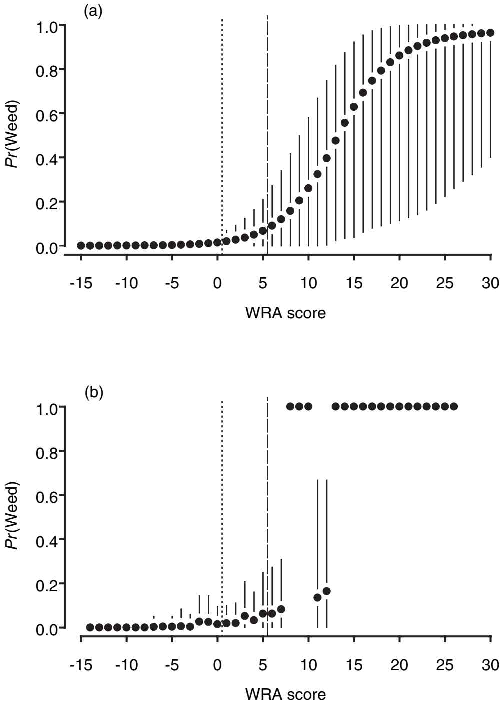

Figure 1. Bootstrapped posterior probabilities of a plant being a weed as a function of WRA score for (a) logistic regression bootstrap; and (b) raw proportions bootstrap, assuming a prior probability of weediness *Pr*(Weed) ~ Beta (1.62, 31.4). Error bars represent 95% uncertainty intervals. Vertical dotted line represents cut-off WRA score for 'further evaluate' and vertical dashed line represents cut-off WRA score for 'reject'.

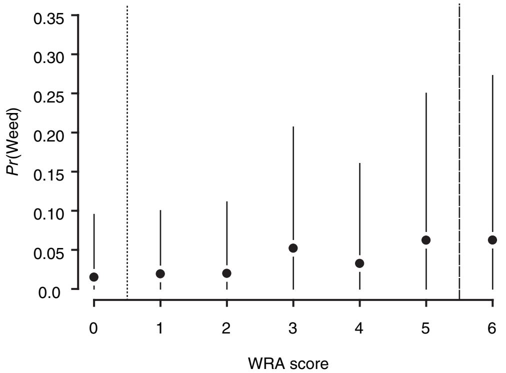

Figure 2. Bootstrapped posterior probabilities of a plant being a weed as a function of WRA scores in the 'further evaluate' category for *Pr*(Weed) ~ Beta (1.62, 31.4). Error bars represent 95% uncertainty intervals. Vertical dotted line represents cut-off WRA score for 'further evaluate' and vertical dashed line represents cut-off WRA score for 'reject'.

WRA score ranged from 2.0% for WRA score = 1 to 6.4% for WRA score = 5 (Figure 2). The uncertainty intervals were proportionally large compared with the predicted probabilities (Figure 2). For example, the uncertainty interval for the probability of weediness for a WRA score of 5 ranged from zero to 26.3% (Figure 2).

DISCUSSION

Previous work has highlighted the effect of a low base-rate probability of weediness on the performance of screening tests for identifying weedy plant biota (Smith *et al*. 1999). We have extended this analysis by incorporating uncertainty in this base-rate, in conjunction with uncertainty arising from the screening model itself. As training datasets used to fit screening models will most commonly have a higher proportion of weedy taxa than the environment in which they are required to predict in, there will always be a need to correct for the prior probability when evaluating the performance of invasive screening models. This statement holds true for all types of predictive models, such as the categorical and regression tree analyses (Reichard and Hamilton 1997), regardless of their quantitative rigour. Indeed, the training dataset used by Reichard and Hamilton (1997) contained *ca*. 67% weedy species, a similar proportion to the dataset analysed here. Obviously, the closer the true proportion of invaders in the suite of species being evaluated is to that contained in the training dataset, the less the bias in the model predictions of invasiveness. We suggest our bootstrapping approach has considerable merit, elucidating uncertainty arising from both the imperfect nature of the screening test, and the uncertainty in the prior probability of invasiveness.

Unsurprisingly, given the low base-rate of weediness, we found the WRA system to have a low PPV, however the uncertainty in the base-rate resulted in considerable uncertainty in the PPV. This is likely true of other classification systems—for example Reichard and Hamilton (1997), whose 'do not admit' classification category may in reality not have the very high probability of weediness that they suggest. Having a low positive predictive value, although undesirable in the context of risk assessment, is not necessarily a problem within the context of risk management. For example, if a 1 in 20 chance of introducing a weedy species is considered too high a risk (i.e. the rejection threshold is set at *Pr*(Weedy) = 0.05), and a taxon is rejected on the grounds that its predicted probability of weediness exceeds this, there is no inconsistency. Rather, it is ignorance of the PPV of a screening test that could bias management decisions, particularly where there is a cost associated with implementing the screening test outcome – for example, a plant that may be of considerable use. Our results indicate a greater demarcation in the probability of weediness between taxa whose WRA scores lie either side of the current threshold score for rejection (WRA Score ≥6), supporting the contention that taxa with WRA model scores above this threshold pose a considerable risk of becoming weeds. The estimated posterior probabilities of weediness for taxa classified in the 'further evaluate' category (2–6%), whilst low at first glance, become non-trivial when one considers the number of taxa being proposed for importation (P. Pheloung unpublished data), the uncertainty in the estimates, and the potentially high cost of importing a weed (Pimentel *et al.* 2000).

The bio-economic modelling of Smith *et al.* (1999) pooled all ultimately rejected taxa ('further evaluate' and 'reject') when calculating the proportion of false positives arising from the WRA model. The current analysis shows that within the group either rejected or in need of further evaluation, the probability of a false positive varies greatly (an order of magnitude) depending on the WRA score. Hence the approach of Smith *et al*. (1999) is overly simplistic, in that it pools good predictions with bad. Logically the analytical structure presented by Smith *et al*. (1999) could be modified to account for this, and also incorporate uncertainty surrounding predictions of weediness. Clearly, better estimates of the prior base-rate probability of weediness will help to reduce uncertainty in predictions of weediness.

ACKNOWLEDGMENTS

Funding for P. Caley was provided by the Australian Grains Research and Development Corporation (Project No. CSE00010).

REFERENCES

Daehler, C.C. and Carino, D.A. (2000). Predicting invasive plants: prospects for a general screening system based on current regional models. *Biological Invasions* 2, 93-102.

Duncan, R.P. and Williams, P.A. (2002). Darwin's naturalization hypothesis challenged. *Nature* 417, 608-9.

Efron, B. and Tibshirani, R.J. (1993). 'An introduction to the bootstrap.' (Chapman and Hall, New York).

Gigerenzer, G. (2002). 'Calculated risks: how to know when numbers deceive you.' (Simon and Schuster, New York).

Hughes, G. and Madden, L.V. (2003). Evaluating predictive models with application in regulatory policy for invasive weeds. *Agricultural Systems* 76, 755-74.

Lonsdale, W.M. (1994). Inviting trouble: introduced pasture species in northern Australia. *Australian Journal of Ecology* 19, 345-54.

Mack, R.N., Simberloff, D., Lonsdale, W.M., Evans, H., Clout, M. and Bazzaz, F.A. (2000). Biotic invasions: causes, epidemiology, global consequences, and control. *Ecological Applications* 10, 689-710.

Manel, S., Williams, H.C. and Ormerod, S.J. (2001). Evaluating presence-absence models in ecology: the need to account for prevalence. *Journal of Applied Ecology* 38, 921-31.

Matthews, R.A.J. (1997). Decision-theoretic limits on earthquake prediction. *Geophysics Journal International* 131, 526-9.

Pheloung, P.C., Williams, P.A. and Halloy, S.R. (1999). A weed risk assessment model for use as a biosecurity tool evaluating plant introductions. *Journal of Environmental Management* 57, 239-51.

Pimentel, D.L., Lach, L., Zuniga, R. and Morrison, D. (2000). Environmental and economic costs associated with non-indigenous species in the United States. *BioScience* 50, 53-65.

Reichard, S.H. and Hamilton, C.W. (1997). Predicting invasions of woody plants introduced into North America. *Conservation Biology* 11, 193-203.

Smith, C.S., Lonsdale, W.M. and Fortune, J. (1999). When to ignore advice: invasion predictions and decision theory. *Biological Invasions* 1, 89-96.

Williamson, M. and Fitter, A. (1996). The varying success of invaders. *Ecology* 77, 1661-5.