Biological Invasions (2006) 8: 277–286
DOI 10.1007/s10530-004-6703-z

Quantifying uncertainty in predictions of invasiveness

Peter Caley[1,2,3,*], W.M. Lonsdale[1,2] & P.C. Pheloung[2,4]
[1]*CSIRO Entomology, GPO Box 1700, Canberra, ACT 2601, Australia;* [2]*Cooperative Research Centre for Australia Weed Management, PMB 1, Waite Campus, Glen Osmond SA 5064, Australia;* [3]*Present Address: National Centre for Epidemiology & Population Health Australian National University, Canberra, ACT 0200, Australia;* [4]*Department of Agriculture, Fisheries and Forestry, GPO Box 858, Canberra ACT 2601, Australia;* **Author for correspondence (e-mail: Peter.Caley@anu.edu.au; fax: +61-2-6125-0740)*

Received 25 February 2004; accepted in revised form 24 November 2004

Key words: Bayesian, bootstrapping, decision support, invasion prediction, modelling uncertainty, risk assessment

Abstract

Using the Australian weed risk assessment (WRA) model as an example, we applied a combination of bootstrapping and Bayesian techniques as a means for explicitly estimating the posterior probability of weediness as a function of an import risk assessment model screening score. Our approach provides estimates of uncertainty around model predictions, after correcting for verification bias arising from the original training dataset having a higher proportion of weed species than would be the norm, and incorporates uncertainty in current knowledge of the prior (base-rate) probability of weediness. The results confirm the high sensitivity of the posterior probability of weediness to the base-rate probability of weediness of plants proposed for importation, and demonstrate how uncertainty in this base-rate probability manifests itself in uncertainty surrounding predicted probabilities of weediness. This quantitative estimate of the weediness probability posed by taxa classified using the WRA model, including estimates of uncertainty around this probability for a given WRA score, would enable bio-economic modelling to contribute to the decision process, should this avenue be pursued. Regardless of whether or not this avenue is explored, the explicit estimates of uncertainty around weed classifications will enable managers to make better informed decisions regarding risk. When viewed in terms of likelihood of weed introduction, the current WRA model outcomes of 'accept', 'further evaluate', or 'reject', whilst not always accurate in terms of weed classification, appear consistent with a high expected cost of mistakenly introducing a weed. The methods presented have wider application to the quantitative prediction of invasive species for situations where the base-rate probability of invasiveness is subject to uncertainty, and the accuracy of the screening test imperfect.

Introduction

Increasingly, countries are attempting to stem the rate of new biological invasions, and as many pests were deliberately introduced, there is a growing scrutiny of the procedures used for importing new species (Ruesink et al. 1995), and recognition of the need for better ways of evaluating the risks and benefits of deliberate introductions (Ewel et al. 1999). Quantitative models show promise for successfully predicting invasive species, as there appears to be some statistically identifiable characteristics of release events and species characteristics that influence the probability of a taxon becoming invasive (Scott and Panetta 1993; Kolar and Lodge 2001). The use of predictive models is particularly applicable in the case for plants, where historically numerous weeds have been imported at the behest of, for example, the pastoral industry (Lonsdale 1994). Currently there is high demand worldwide to import many hundreds of taxa as part of a burgeoning global horticultural trade, hence the need to accurately predict the invasive status of plants proposed for importation has never been higher. In response to the need for a publicly acceptable risk assessment system to predict the weediness, or invasive potential of plants being considered for importation, screening models such as the weed risk assessment (WRA) system (Pheloung et al. 1999) have been developed and implemented.

Despite some progress in identifying the characteristics of invasive species, successful discrimination between invasive and non-invasive species remains a difficult task, as it is widely accepted that the likelihood of an introduced organism making the transition to being invasive is low (Williamson and Fitter 1996; Mack et al. 2000; United States National Research Council 2002). For events that have a low prior probability of occurring, predicted probabilities of occurrence based on the results of imperfect screening tests alone tend to substantially overestimate the true probability of the event occurring (Gigerenzer 2002). In the context of invasive species, the 'prior' probability of invasiveness is the probability of invasiveness assigned to an individual taxon drawn at random from a group of species *before* consideration of additional information that may contribute to the likelihood of it being invasive, such as that gathered from a screening test. The prior probability is sometimes referred to as the 'base-rate probability'. The actual probability of a particular taxon being invasive in light of a screening test is obtained by 'revising' the prior probability using the screening test likelihood, and for this reason it is often referred to as the 'posterior' probability. The proportion of events predicted by a screening test that would be expected to actually occur based on this posterior probability is referred to as the 'positive predictive value' (PPV) of a test, and for example, in the case of the WRA model, could be in the order of 0.1 (Smith et al. 1999). That is, an estimated 90% of species effectively rejected are in fact not weeds – 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 and weather forecasting (Matthews 1996, 1997)] and medicine [e.g. cancer screening (Metz 1978; Gigerenzer 2002)], though has only more recently been addressed in issues of natural resource management [e.g. WRA (Smith et al. 1999)] or ecology [e.g. predicting species occurrence (Pearce and Ferrier 2000; 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. This is illustrated in Figure 1 in the case of using a screening test to identify weeds, and makes it clear that even for a test of exceptional accuracy [sensitivity = specificity = 99%], as the prior probability of being a weed becomes very small, the PPV declines dramatically.

One management approach for dealing with the imperfect nature of diagnostic tests and the resulting inaccuracies in prediction is the use of

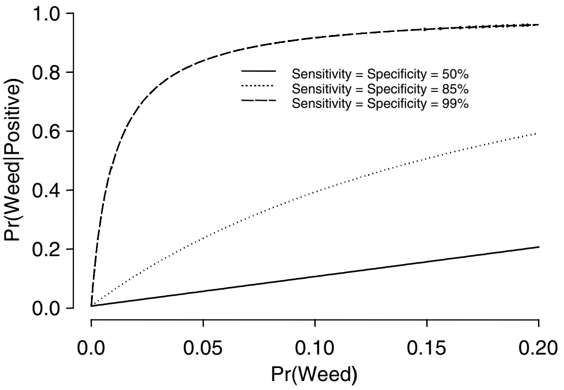

Figure 1. The relationship between the posterior probability of a taxon being a weed, given that a diagnostic test classifies it as a weed [Pr(Weed | Positive test)] and the prior probability of weediness [Pr(Weed)] for differing values of test sensitivity and specificity.

decision theory (Matthews 1997), and it has been suggested that WRA could be placed in this context (Smith et al. 1999). For such an approach, quantitative estimates of the probability of invasiveness are needed, along with estimates of the losses associated possible decisions such as importing an invasive plant or preventing importation of a useful plant etc. Classification-based screening models such as the WRA, whilst useful in terms of their simplicity and ease of interpretation, do not readily provide the required probabilities of invasiveness. Furthermore, viewing a classification at face value relegates uncertainty to being implicit or contained within the assessment system, rather than an explicit outcome of the assessment system. However, for classification systems based on summary scores, these scores may be converted to predicted probabilities of invasiveness by using, for example, logistic regression. For example, Hughes and Madden (2003) fitted a logistic regression model to the summary scores used to develop the WRA model. Note though that the resulting fitted probabilities of weediness in relation to WRA score are biased upwards, as the data contain an unrealistically high proportion (ca. 77%) of weeds. That is, there is verification bias (Begg and Greenes 1983).

A decision theory approach requires not only the estimated probability of a taxa becoming invasive, but also the uncertainty around that probability. As an extreme example for illustrative purposes, managers would view differently the application to import species A with estimated Pr(Invasive) = 0.1 and associated 95% uncertainty interval (U.I.) 0.09–0.11, than species B with estimated Pr(Invasive) = 0.1 and associated 95% U.I. 0.01–0.99. In deciding to import species A, the magnitude of risk being taken is very certain, whereas the risk associated with importing species B is largely unknown, ranging anywhere between near-zero risk and near-certain risk. Ruesink et al. (1995) suggest that species with unknown consequences should not be imported, unless additional information can reduce the uncertainty involved. Uncertainty in predicted probabilities may come from several main sources. The first is model selection uncertainty, whereby differing models differ in their predictions, and we are unsure which model is the correct on to use. Model-averaging is one approach for dealing with this (Burnham and Anderson 2002). The second source is the inherent variability in model predictions arising from observation and process uncertainty, 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. For example, a prediction confidence interval can be added to the logistic regression model of Hughes and Madden (2003) relating the probability of an imported plant being a weed as a function of WRA score (Figure 2). Lastly, there may be uncertainty surrounding structural parameters used within the modelling process. For example, while it is clear that the performance of a screening test is highly sensitive to the prior probability of the event being

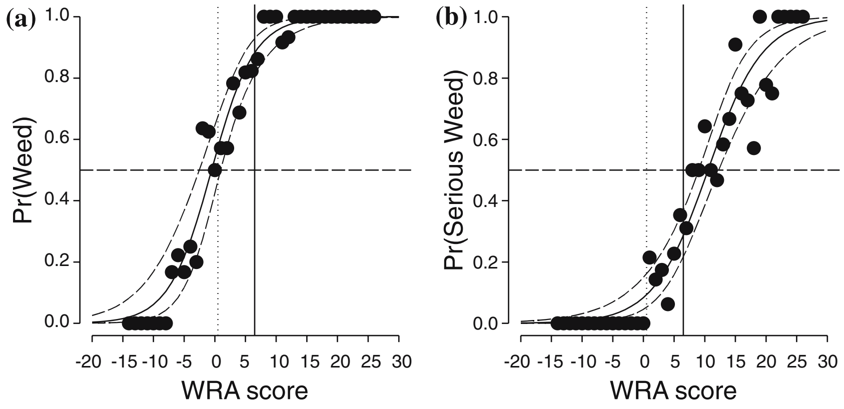

Figure 2. Fitted logistic regression model (—) relating the probability of being a weed to the WRA score for (a) all types of weeds; and (b) serious weeds. Dashed lines are 95% prediction confidence intervals around the fitted line. The vertical dotted and solid lines indicate the cut-off WRA score for taxa to be classified as 'further evaluate' or 'reject', respectively.

predicted, in the case of biological invasions, this quantity is poorly characterised. Indeed, the assumption of a very low prior probability of introduced organisms becoming invasive appears not to be universally true (e.g. Lonsdale 1994), and the emerging high rate of naturalisation of non-indigenous plants (Duncan and Williams 2002a) may indicate that in time this paradigm may change, as naturalisation is a necessary precursor to invasion. There is clearly a need to incorporate the uncertainty surrounding the prior probability of invasiveness when assessing the predictive capabilities of a screening model, and empirical Bayesian approaches provide a way of achieving this.

In this paper, using the WRA model of Pheloung et al. (1999) as an illustrative example, we demonstrate how screening scores arising from a predictive model of invasiveness may be re-expressed as posterior probability estimates of invasiveness, including uncertainty around these estimates that reflect uncertainty in the prior probability of invasiveness. We achieve this by utilising standard bootstrapping procedures (Efron and Tibshirani 1993) to generate bootstrapped datasets from the original WRA dataset that contain a more realistic prior probability of weediness, though we additionally use a Bayesian approach to explicitly incorporate uncertainty surrounding this prior probability. We then extend the logistic regression approach of Hughes and Madden (2003) to modelling our bootstrapped datasets. In doing so, we demonstrate how a screening system used to predictively classify taxa into various invasiveness categories may be modified to explicitly estimate risk in a probabilistic manner. We recognise that not all weeds would be classified as invasive plants under the definition of Richardson et al. (2000), however they would be under the definition of Heger and Trepl (2003). This slight confusion in notation does not affect the purpose of this paper in quantifying uncertainty around screening model predictions.

Materials and methods

The WRA model

The WRA model (Pheloung et al. 1999) has been operational in Australia since 1996, and is tasked with providing a measure of the weed potential of a plant species, and supports a decision-making policy consistent with relevant international treaties. The WRA model converts responses to questions relating to the plants climatic preferences, biological attributes, reproductive and dispersal method into a score, whose value determines whether to 'accept' (WRA score ≤ 0), 'further evaluate' ($1 \leq$ WRA score ≤ 6), or 'reject' (WRA score > 6) the taxon. It performs well in identifying weeds over a range of countries, with a test sensitivity of ca. 90% (Pheloung et al. 1999; Daehler and Carino 2000). Up until late 2003, about 1000 taxa have undergone valid assessment where the minimum required number of questions were answered, with 46% being accepted, 23% requiring further evaluation and the remaining 31% rejected. The WRA model is not a model in the statistical sense, as the coefficients and weightings to questions were determined manually to maximise the discriminatory performance between weeds and non-weeds, hence it is not amenable to standard cross-validation techniques such as jack-knifing (Kohavi 1995).

Specifying the prior probability of weediness

The prior probability of weediness of plants being considered for importation is a poorly quantified parameter. In a review of the available literature, Smith et al. (1999) considered the value to range from 0.01% (Williamson and Fitter 1996) to 17% (Lonsdale 1994) with a likely value of 2%. As a prior distribution for modelling this uncertainty we used a Beta distribution (Vose 2000) 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 (Figure 3). The Beta distribution is the common choice of a prior or mixing distribution for a probability parameter, as it is bounded by the interval (0,1) and its shape is quite flexible, ranging from near-uniform to highly skewed towards either zero or one.

Estimating posterior probability of weediness as a function of WRA score

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

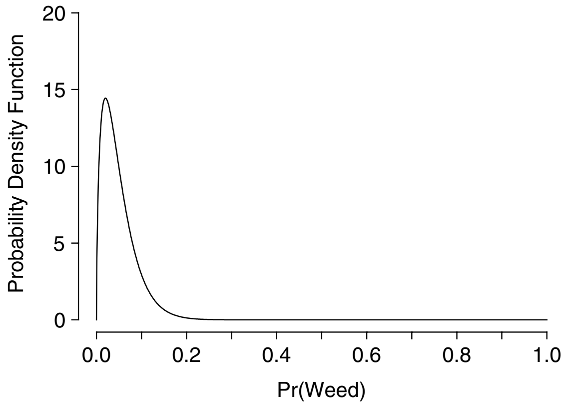

Figure 3. Probability density function of Beta(1.6, 31.4) distribution used to model the prior (or 'base-rate') probability of weediness.

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 and uncertainty around this prior probability. However, bootstrapping, whereby datasets are repeatedly resampled with replacement, provides a robust method of estimating this variability whilst accounting of the prior probability. Initially, to examine the sensitivity of the model predictions to the prior probability, we bootstrapped the original WRA dataset with the probability of selection of weeds and non-weeds calculated to produce an underlying *fixed* prior probability of weediness in the bootstrapped sample of either 0.001, 0.02, 0.1, or 0.2. We generated 1000 bootstrap samples for each prior probability. This resulted in four sets of data, each consisting of 1000 bootstrap samples from the WRA dataset. For each bootstrap sample, a logistic regression model was fitted, relating the predicted (posterior) probability of weediness to WRA score. Finally, over the range of WRA scores, we calculated the average of the logistic regression model curves over all bootstrap samples within each dataset, and the associated lower and upper 95% uncertainty intervals for each prior probability of weediness.

We then generated a further 1000 bootstrap samples from the WRA dataset, though with the probability of weed selection for each bootstrap sample drawn at random from a Beta (1.62, 31.4) distribution. Logistic regression was used to relate the posterior probability of weediness to WRA score as before. We also averaged the raw bootstrap probabilities as a function of WRA score over the 1000 bootstrap samples, to provide a 'raw bootstrap' estimate of the posterior mean probability of weediness as a function of WRA score, including uncertainty intervals.

Evaluating current thresholds

We evaluated the estimated posterior probabilities of weediness in relation to the thresholds used in the WRA model to distinguish between acceptable taxa (WRA score ≤ 0) from those requiring further evaluation ($1 \leq$ WRA score ≤ 6) from those that should be rejected (WRA score > 6). We concentrated on the results obtained for the 'further evaluate' class, as it is taxa in this class that are most problematic from a management point of view. For a given WRA score, the estimated posterior probability is equal to the PPVs of a screening test using that particular score as a cut-off value. Hence the PPVs of screening tests using different cut-offs may be read directly from a plot of posterior probabilities versus WRA score.

All analyses were undertaken using S-Plus version 6.1 for Windows (Insightful Co., Seattle, Washington).

Results

The posterior predicted probability of weediness as a function of WRA score was, as expected, highly sensitive to the prior probability of weediness (Figure 4). As the prior probability of weediness increased, the predicted probability of weediness increased for a given WRA score, and the uncertainty intervals around the predicted probability decreased as predictions became more certain. For a low prior probability of weediness (≤2%), the predicted probability of weediness was virtually zero up until a WRA score of ca. 5, and the uncertainty intervals were wide (Figures 4a and b). For modelled prior probabilities of weediness greater than 2%, the predicted probability approached the asymptote of 1 at high WRA scores (Figures 4b–d).

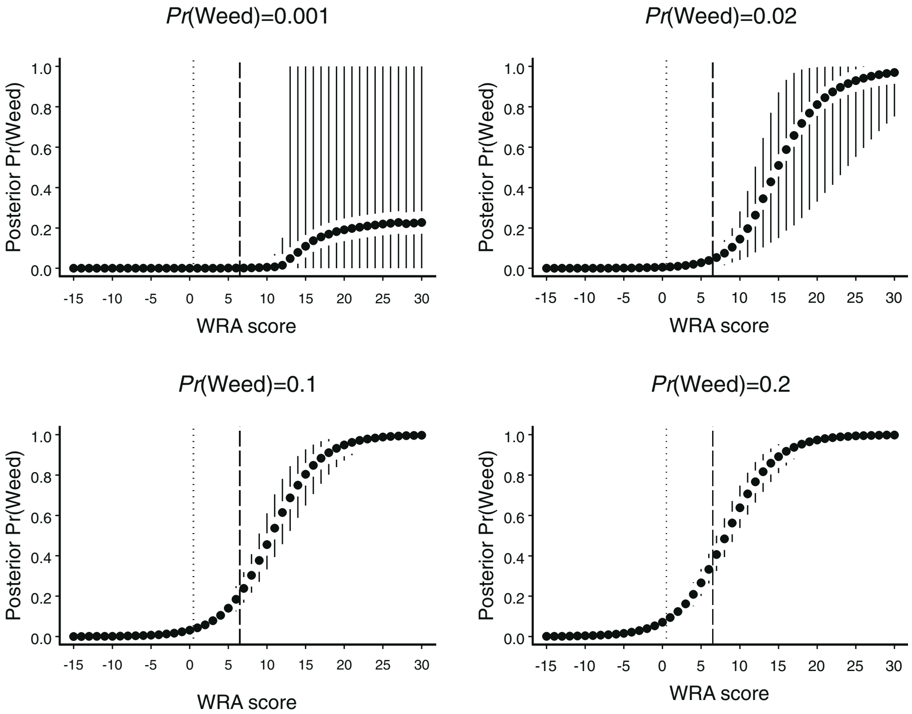

Figure 4. Bootstrapped results of logistic regression modelling the posterior probability of a plant being a weed (of any classification) as a function of WRA score assuming the prior probability of weediness *Pr*(Weed) takes on the values of either 0.001, 0.02, 0.1 or 0.2. Error bars represent 95% uncertainty intervals. Vertical dotted line represents cut-off WRA score for 'further evaluate' and vertical dashed line represents cut-off WRA score for 'reject'. Results are based on 1000 bootstrap samples from original 370 taxa survey.

Bootstrapping the dataset with the prior probability of weediness drawn from a Beta distribution provides our best estimates of how the probability of being a weed relates to the WRA score, and the estimated variability around this prediction (Figure 5). Comparing the uncertainty around the logistic regression models of Figure 2a (no uncertainty in prior probability) with Figure 5a (uncertainty in prior probability modelled as Beta) illustrates clearly how incorporating uncertainty in the prior probability of weediness substantially increases the uncertainty around the predicted probability of weediness for a given WRA score. The PPV of the WRA model using the current cut-offs is clearly low (<0.1) using both bootstrap approaches (Figure 5).

The predicted probability or weediness 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 approach used, as a general trend, the probability of weediness started to increase sharply from a WRA score of about zero for the logistic bootstrap and −2 for the raw bootstrap 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 received these WRA scores (Figure 5b). In contrast, the logistic regression recorded substantial variability around the predicted probability of weediness for high WRA scores (Figure 5a).

Within the 'further evaluate' classification applied to taxa with WRA scores in the range 1–6, the probability of weediness as a function of WRA score ranged from 2.0% for WRA

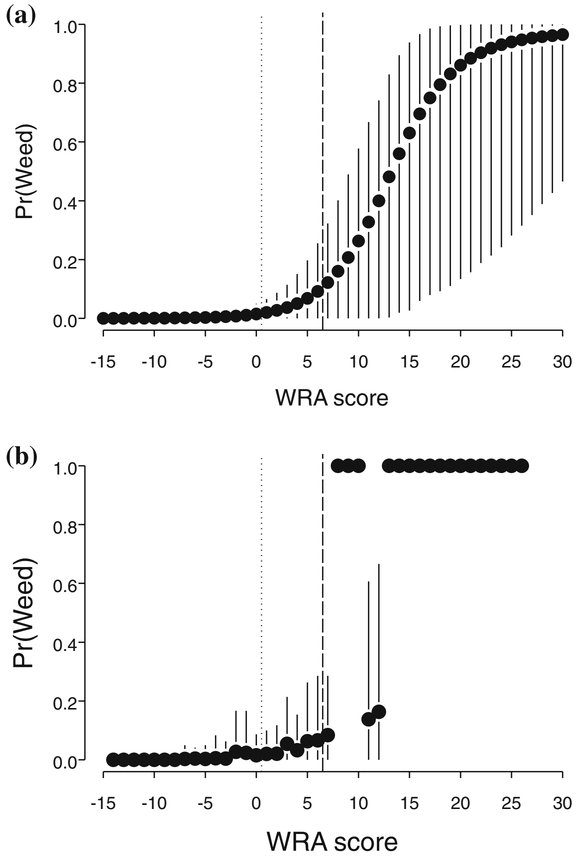

Figure 5. Bootstrapped posterior probabilities of a plant being a weed (of any classification) as a function of WRA score for (a) logistic regression bootstrap; and (b) raw proportions bootstrap, assuming a prior probability of weediness *Pr*(Weed) ~ Beta (1.62, 31.4). Error bars represent 95% uncertainty intervals. Vertical dotted line represents cut-off WRA score for 'further evaluate' and vertical dashed line represents cut-off WRA score for 'reject'. Results are based on 1000 bootstrap samples from original 370 taxa survey.

score = 1 to 6.4% for WRA score = 6 (Figure 6). The uncertainty intervals were proportionally large compared to the predicted probabilities (Figure 6). For example, the uncertainty interval for the probability of weediness for a WRA score of 5 ranged from 0 to 26.3% (Figure 6).

Discussion

Previous studies (Smith et al. 1999; Hughes and Madden 2003) have highlighted the effect of a low prior probability of invasiveness on the performance of screening tests to identifying invasive plant biota. We have extended this analysis by explicitly incorporating the uncertainty in this prior probability, in conjunction with uncertainty arising from the screening model itself. The resulting predicted probabilities of weediness as a function of WRA screening score include substantially more uncertainty than would be the case if uncertainty in the prior probabilities or weediness were ignored. As training datasets used to fit screening models will most commonly have a higher proportion of invasive 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 categorical and regression tree analyses (Reichard and Hamilton 1997; Kolar and Lodge 2002), regardless of their quantitative rigour. Indeed, the training dataset used by Reichard and Hamilton (1997) contained ca. 67% invasive 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

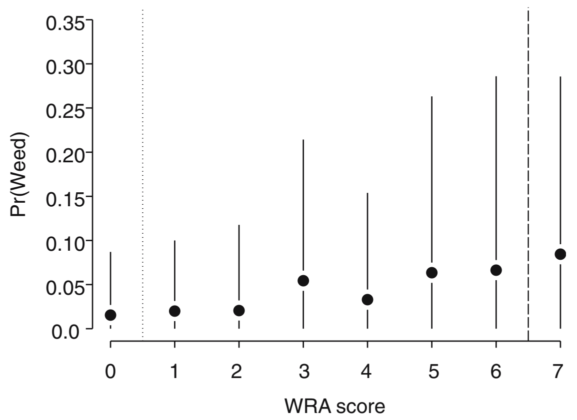

Figure 6. Bootstrapped posterior probabilities of a plant being a weed (of any classification) as a function of WRA scores in the 'further evaluate' category for a prior probability of weediness *Pr*(Weed) ~ Beta (1.62, 31.4). Error bars represent 95% uncertainty intervals. Vertical dotted line represents cut-off WRA score for 'further evaluate' and vertical dashed line represents cut-off WRA score for 'reject'. Results are based on 1000 bootstrap samples from original 370 taxa survey.

predictions of invasiveness. For example, the training dataset used by Daehler et al. (2004) contains a reasonably high (39%) proportion of weeds, hence the screening test will still undoubtedly perform at a level below what the sensitivity and specificity alone suggest, though possibly not to the extent of the previously mentioned studies.

Estimating the prior probability of invasiveness will remain problematical, and hence any estimate, will be subject to considerable uncertainty. Given the close relationship between the PPV of a screening test and this prior probability, it is important that this uncertainty is incorporated into model predictions. We consider our bootstrapping approach has considerable merit for addressing this problem, elucidating uncertainty arising from both the imperfect nature of the screening test, and the uncertainty in the prior probability of invasiveness. An advantage of the bootstrapping approach is that it can be applied to a wide range of models. For example, bootstrapping has previously been used to estimate uncertainty around predictions for non-parametric classification models such as classification trees (Kuhnert and Mengersen 2003). We note, however, that the two methods we used for analysing bootstrapped datasets, namely a raw summary of the bootstrapped datasets and the summary of repeated logistic regressions to each bootstrapped dataset, have both strengths and weaknesses. At the lower end of the WRA scale, the logistic regression approach clearly underestimates uncertainty in the probability of weediness compared to the raw bootstrap. This is due to the constraints on the shape of the logistic curve relating the probability of weediness to WRA score. At the upper end of the WRA scale, the raw bootstrap estimates no uncertainty in the probability of weediness for WRA scores of 7–9 and 13 and above. For WRA score of 7–9 at least, this lack of uncertainty in weediness is in all likelihood an illusion resulting from small sample sizes not containing any non-weeds that were assigned these WRA scores. In contrast, the logistic regression approach suggests substantial variation in the probability of weediness for higher WRA scores. We recommend that for estimating the probability of weediness for a given WRA score, it is best to use the raw bootstrap results for WRA scores less than 5 and the logistic regression results for WRA scores greater than or equal to six.

Taking our recommended approach for interpreting the results indicates that the probability of being a weed become noticeably non-zero at a score somewhat less than the current cut-off for 'further evaluation' (WRA score ≥ 1), and starts to increase rapidly for WRA scores at, or slightly greater than the current cut-off for 'reject' (WRA score > 6). For taxa with WRA scores classified in the 'further evaluate' category, the estimated posterior probabilities of weediness (2–6%), whilst low at first glance, become non-trivial when one considers the number of taxa being proposed for importation (P. Pheloung, unpubl. data), the uncertainty in the estimates, and the potentially high cost of importing a weed (Pimentel et al. 2000). For the current system to be cost neutral, the cost of mistakenly introducing a weed needs to far outweigh the benefit of introducing a useful plant.

In this paper, we have focused on estimating the likelihood of an introduced plant being a weed, and have not considered variation in the consequences between taxa. Hence we have not fully characterised risk as per the standard definition of $risk = likelihood \times consequences$. Given that the WRA model does not attempt to estimate the consequences of introducing a weed, this was beyond the scope of this study. We also have not considered other issues possibly affecting the performance of the WRA survey dataset and its interpretation, such as variations in weed definitions, possible biases in scores based on prior knowledge of a taxon's weediness, and variation in screening performance and/or weediness between taxonomic groups (Smith 1999; Duncan and Williams 2002b).

Unsurprisingly given the low base-rate of weediness and previous work (Smith et al. 1999), we found the WRA system to have a low PPV. This is likely true of other classification systems such as that presented by Reichard and Hamilton (1997), whose 'do not admit' classification category may in reality not have the very high probability of invasiveness that they suggest. Having a low PPV, 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 an invasive species is considered too high a risk [i.e. the rejection threshold is set at Pr(Invasive) = 0.05], and a taxon is rejected on the grounds that its predicted probability of invasiveness 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 taxon that may be of considerable use. Using the maximisation of PPVs as the sole criteria for determining cut-off values for screening tests is inappropriate, as the cut-off value that maximises the PPV will admit all but the highest scoring invasive species.

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 by as much as an order of magnitude depending on 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. For a taxon assigned a particular WRA score, the estimated probability of weediness, and associated uncertainty interval can be taken directly from our results. Any future attempt at bio-economic modelling of plant introductions should also incorporate uncertainty not only in predictions of invasiveness, but also in the expected costs and benefits of *future* plant introductions. Clearly the costs and benefits of past introductions are not necessarily a reliable guide to those of future introductions. We also note that the WRA model as implemented is the second component of a three-tiered system aimed at identifying and preventing the entry to Australia of environmental and agricultural weeds. Its current implementation is in no way intended to support a bioeconomic approach to plant importation decisions, although we have used it as a case study to illustrate how to generate the required outputs should this avenue be pursued.

In summary, we have demonstrated how screening test scores may be translated to quantitative probabilities of invasiveness, and argued that this form of screening information would enable managers to better assess the risk posed by a proposed introduction. Our results have demonstrated that a low and uncertain base-rate probability manifests itself in uncertainty surrounding predicted probabilities of invasiveness, and a low PPV of screening test positives. Clearly, better estimates of the prior probability of invasiveness and improved predictors of invasiveness will help to reduce this uncertainty and improve the accuracy of future screening models for invasive species.

Acknowledgements

Funding for P. Caley was provided by the Australian Grains Research and Development Corporation (Project No. CSE00010). The comments of Saul Cunningham, Keith Hayes, Dane Panetta, Peter Williams and two anonymous referees improved earlier versions of the manuscript.

References

Begg CB and Greenes RA (1983) Assessment of diagnostic tests when disease verification is subject to selection bias. Biometrics 39: 207–215

Burnham KP and Anderson DR (2002) Model Selection and Multimodel Inference: A Practical Information–Theoretic Approach, Springer-Verlag, New York

Daehler CC and Carino DA (2000) Predicting invasive plants: prospects for a general screening system based on current regional models. Biological Invasions 2: 93–102

Daehler CC, Denslow JS, Ansari S and Kuo H-C (2004) A risk-assessment system for screening out invasive pest plants from Hawaii and other Pacific islands. Conservation Biology 18: 360–368

Duncan RP and Williams PA (2002a) Darwin's naturalization hypothesis challenged. Nature 417: 608–609

Duncan RP and Williams PA (2002b) Taxonomic patterns in the naturalisation rate of plant species in New Zealand. In: Spafford JH, Dodd J and Moore JH (eds) Thirteenth Australian Weeds Conference. pp. 183–185. Plant Protection Society of Western Australia, Perth, Western Australia

Efron B and Tibshirani RJ (1993) An Introduction to the Bootstrap, Chapman and Hall, New York

Ewel JJ, O'Dowd DJ, Bergelson J, Daehler CC, D'Antonio CM, Gomez LD, Gordon DR, Hobbs RJ, Holt A, Hopper KR, Hughes CE, LaHart M, Leakey RRB, Lee WG, Loope LL, Lorence DH, Louda SM, Lugo AE, McEvoy PB, Richardson DM and Vitousek PM (1999) Deliberate introductions of species: research needs. BioScience 49: 619–630

Gigerenzer G (2002) Calculated Risks: How to Now When Numbers Deceive You, Simon and Schuster, New York

Heger T and Trepl L (2003) Predicting biological invasions. Biological Invasions 5: 313–321

Hughes G and Madden LV (2003) Evaluating predictive models with application in regulatory policy for invasive weeds. Agricultural Systems 76: 755–774

Kohavi R (1995) A study of cross-validation and bootstrap for accuracy estimation and model selection. In: Mellish CS (ed) Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence. pp 1137–1143. Morgan Kaufmann Publishers, Inc., Montréal, Québec

Kolar CS and Lodge DM (2001) Progress in invasion biology: predicting invaders. Trends in Ecology and Evolution 16: 199–204

Kolar CS and Lodge DM (2002) Ecological predictions and risk assessment for alien fishes in North America. Science 298: 1233–1236

Kuhnert PM and Mengersen K (2003) Reliability measures for local nodes assessment in classification trees. Journal of Computational and Graphical Studies 12: 1–19

Lonsdale WM (1994) Inviting trouble: introduced pasture species in northern Australia. Australian Journal of Ecology 19: 345–354

Mack RN, Simberloff D, Lonsdale WM, Evans H, Clout M and Bazzaz FA (2000) Biotic invasions: causes, epidemiology, global consequences, and control. Ecological Applications 10: 689–710

Manel S, Williams HC and Ormerod SJ (2001) Evaluating presence-absence models in ecology: the need to account for prevalence. Journal of Applied Ecology 38: 921–931

Matthews RAJ (1996) Base-rate errors and rain forecasts. Nature 382: 766

Matthews RAJ (1997) Decision-theoretic limits on earthquake prediction. Geophysics Journal International 131: 526–529

Metz CE (1978) Basic principles of ROC analysis. Seminars in Nuclear Medicine 8: 283–298

Pearce J and Ferrier S (2000) Evaluating the predictive performance of habitat models developed using logistic regression. Ecological Modelling 133: 225–245

Pheloung PC, Williams PA and Halloy SR (1999) A weed risk assessment model for use as a biosecurity tool evaluating plant introductions. Journal of Environmental Management 57: 239–251

Pimentel DL, Lach L, Zuniga R and Morrison D (2000) Environmental and economic costs associated with nonindigenous species in the United States. BioScience 50: 53–65

Reichard SH and Hamilton CW (1997) Predicting invasions of woody plants introduced into North America. Conservation Biology 11: 193–203

Richardson DM, Pysek P, Rejmanek M, Barbour MG, Panetta FD and West CJ (2000) Naturalization and invasion of alien plants: concepts and definitions. Diversity and Distributions 6: 93–107

Ruesink JL, Parker IM, Groom MJ and Kareiva PM (1995) Reducing the risks of nonindigenous species introductions: guilty until proven innocent. BioScience 45: 465–477

Scott JK and Panetta FD (1993) Predicting the Australian weed status of southern African plants. Journal of Biogeography 20: 87–93

Smith C (1999) Studies on Weed Risk Assessment. Master of Applied Science, University of Adelaide

Smith CS, Lonsdale WM and Fortune J (1999) When to ignore advice: invasion predictions and decision theory. Biological Invasions 1: 89–96

United States National Research Council (2002) Predicting Invasions of Nonindigenous Plants and Plant Pests, National Academy Press, Washington, DC

Vose D (2000) Risk Analysis: A Quantitative Guide, John Wiley and Sons, Chichester, West Sussex

Williamson M and Fitter A (1996) The varying success of invaders. Ecology 77: 1661–1665