Report Cover Page

ACERA Project		
0901		
Title		
Demonstrating risk analysis capabilities		
Author(s) / Address (es)		
M. Burgman, M. Colyvan, R. Christian, B. Dobiecki, J. Elith, F. Fidler, N. Grant, C. Hauser, K. Hayes, S. Jones, K. Korb, P. Kuhnert, M. McBride, M. McCarthy, A. Nicholson, A. Robinson, T. Rout, A. Speirs-Bridge, T. Walshe		
Material Type and Status		
Final Report		
Summary		
This report implements a number of tools in the context of import risk analysis, some of which may be useful in training or in routine IRAs. Others may be useful in special cases, where particular features of the biology of a species or the circumstances of trade demand more detailed data collection and analysis. The purpose of this study was to demonstrate their capabilities, strengths and weaknesses for a hypothetical case study of insects on fruit. The tools demonstrated in this study include scenario trees, structured protocols for questioning experts, methods for combining judgements, Monte Carlo, interval arithmetic, Bayes nets, probability bounds, spatial habitat modelling, cognitive maps, structured decision making, and surveillance and inspection tools.		
ACERA Use only	Received By:	Date:
	ACERA / AMSI SAC Approval:	Date:
	DAFF Endorsement: ( ) Yes ( ) No	Date:

ACERA Project 0901

Demonstrating risk analysis capabilities

M. Burgman[1], M. Colyvan[2], R. Christian[3], B. Dobiecki[1], J. Elith[1], F. Fidler[1], N. Grant[3], C. Hauser[1], K. Hayes[4], S. Jones1, K. Korb[5], P. Kuhnert[6], M. McBride[1], M. McCarthy[1], A. Nicholson[5], A. Robinson[1], T. Rout[1], A. Speirs-Bridge[1] and T. Walshe[1]

1. ACERA, The University of Melbourne, Parkville
2. The University of Sydney, Sydney
3. Department of Agriculture, Fisheries and Forestry, Canberra
4. CMIS, CSIRO, Hobart
5. Monash University, Clayton
6. CMIS, CSIRO, Brisbane

Final Report

December, 2010

Acknowledgements

This report is a product of the Australian Centre of Excellence for Risk Analysis (ACERA). In preparing this report, the authors acknowledge the financial and other support provided by the Department of Agriculture, Fisheries and Forestry (DAFF), the University of Melbourne, Australian Mathematical Sciences Institute (AMSI) and Australian Research Centre for Urban Ecology (ARCUE). We are indebted to comments from the ACERA Scientific Advisory Committee and the Biosecurity Services Group in DAFF for comments that substantially improved the document.

Disclaimer

Contents

Executive Summary

This report assesses a range of scientific tools for risk analysis that may play a role in import risk assessment. They include:

- Exposure pathway diagrams, cognitive maps and event trees.
- Methods for selecting and questioning experts, and combining their judgements.
- Bayes nets for incorporating uncertainty in risk assessments.
- Numerical methods including Monte Carlo analysis and probability-bounds analysis.
- Spatial models for spread.
- Consequence estimation prior to establishment and spread.
- Inspection and surveillance systems.

Some of these tools are used extensively in other professional areas. Others have been suggested recently as solutions to some long-standing problems in risk analysis. The applications in this report identified a number of issues where new methods may improve the scientific consistency and reliability of risk analysis.

Cognitive maps and event trees will assist with evaluating systems approaches to risk mitigation and will improve the conditioning of likelihood assessments. Current approaches to consequence estimation may systematically underestimate risk because they discount consequences that emerge prior to establishment and spread of pests and diseases. Event trees and cognitive maps may play a useful role in some risk assessments, estimating these consequences.

Interval arithmetic provides a framework for ensuring that outputs from the combinations of likelihoods are consistent with inputs, and accounts for dependencies.

More generally, Bayes Nets provide a formal tool that can accommodate the existing rules for combining likelihoods exactly as currently specified. They provide an environment that would allow the system to be modified simply and consistently to account for uncertainty, to deal more coherently with likelihoods and to build a library of cases that could be used to inform future risk assessments.

Mathematical techniques such as Monte Carlo simulation may be useful in situations where data are relatively abundant. New developments in probability-bounds arithmetic will be useful in ensuring that any unstated or unjustified assumptions in a Monte Carlo simulation are not critical in determining decisions.

Statistical methods for predicting spread have developed rapidly over the decade. These tools are likely to be most beneficial when the biological factors governing habitat are unclear, when precise predictions are necessary for consequence assessment, and when effective monitoring and surveillance systems are important.

A large number of analytical tools to support surveillance and inspection have emerged recently. In the context of import risk assessments, the risk analyses could specify appropriate and effective surveillance recommendations. The data from these systems could be used, in turn, to validate the risk analysis assumptions, and provide a platform for periodic revision of the risk analyses.

In some situations, innovations such as those noted above may be un-necessary, because decisions are unlikely to be affected by their application. However, it is difficult to know a priori, in all but the most obvious cases, where their application will make an important difference. In all of the cases noted above, the tools for their implementation (software, guidelines and protocols) have been developed and tested.

1. Introduction

ACERA has developed and tested a range of tools that may assist biosecurity risk analysis. These tools have been trialled in a limited number of biosecurity case studies, although some have been tested more thoroughly in other technical and scientific domains. However, their applicability for import risk assessment (IRA) and related tasks needs to be assessed thoroughly. For example, they may be too cumbersome or require too many specialist skills to be useful in routine applications. The data required may typically be unavailable, or they may be particularly sensitive to uncertainties. Equally, the methods may not comply with the international rules that govern trade and the analyses of the risks associated with this trade.

This project aims to evaluate the usefulness of several methods to meet the operational needs of Australia's Biosecurity Services Group within the Department of Agriculture Fisheries and Forestry and their potential to support decision-making, using a case study that reflects many of the conditions experienced by those who conduct pest and import risk analysis. It will document the outcomes of the use of these tools, and will discuss their strengths and weaknesses in a biosecurity context. It will evaluate how these tools may be employed to support decisions to intervene to reduce risks to meet Australia's Appropriate Level of Protection (ALOP).

To evaluate the potential for a range of tools to contribute to IRAs and test their practical suitability within the biosecurity operational environment, the case study is based on a hypothetical scenario that uses an existing pest risk analysis (PRA). The evaluations are based on a side-by-side comparison with current approaches and consider whether the tool;

- is appropriate and practical,
- requires skills or resources that are currently available,
- improves the assessment of risk.

Some of analyses below illustrate simple tools such as tables of attributes and conceptual diagrams. Others illustrate more complex tools, such as spatial analyses and interval methods for uncertainty propagation. In all cases, more detailed background can be found in the supporting references.

1.1 Research Strategy

A staged approach was used to coordinate the development of the scenario and conventional analyses, and to allow the application of the range of techniques by several researchers independently. The stages for the project were:

Stage 1. Scenario outline and development

The project design requires a conventional analysis that can be used as a baseline against which to compare potential new approaches. For the conventional analysis, we adopted hypothetical IRA for an insect on fruit. The organism was considered to be a pest of quarantine concern and risk mitigation measures were required to reduce the risk from this pest to meet Australia's ALOP. We used this pest species and commodity as the model for the project. The research team were given the opportunity to comment on the conventional scenario, add details to help demonstrate aspects of the analysis, and otherwise suggest modifications that would improve the process.

Stage 2. Demonstration studies

The project aimed to identify methods that have been developed and tested by ACERA over the last three years and that are relevant to import risk analysis, with the intention of applying them to a relevant part of the scenario, using the same data and expertise as were necessary to conduct the standard and Monte Carlo-based risk assessments.

A workshop was held in November 2008, to discuss the tools that could be demonstrated in the project. Potential applications encompassed the full extent of the biosecurity continuum. The workshop agreed to include the following methods:

- Describing exposure pathway (exposure pathway diagrams and cognitive maps)
- Questioning experts (four-step questions and the expert procedures guide)
- Combining expert judgements (Bayesian aggregation, formal consensus)
- Monte Carlo simulation
- Bayes nets
- Uncertainty analysis (interval and probability bounds analysis)
- Estimating spread extent
- Applying cognitive maps to estimate consequences
- Surveillance and inspection specifications from IRAs

Stage 3. Implementation

Specialists from the ACERA research projects agreed to participate and to apply tools and methods to the common scenario, sharing the outputs of each step with others in the group so that they could be used as input to other stages of the analysis. Responsibilities, participation and authorship included

- Exposure pathway: S. Jones, K. Hayes
- Eliciting judgements: A. Speirs-Bridge, M. Burgman, M. McBride
- Combining judgements: M. McBride, P. Kuhnert, K. Hayes
- Monte Carlo: S. Jones, K. Hayes
- Bayes nets: A. Nicholson, K. Korb
- Uncertainty analysis: K. Hayes
- Spread extent: B. Dobiecki, J. Elith
- Consequences: T. Walshe, R. Christian
- Surveillance: C. Hauser, T. Rout, M. McCarthy
- Inspection: A. Robinson

Stage 4. Critical assessments

Each participant provided a summary of their results and an evaluation of the strengths, weaknesses and limitations of their analysis to the operational needs of BSG. These summaries and assessments are combined in this report. Each of the tools was implemented for the common case study, with the exception of the analysis of spread extent for which a study was conducted for a pest that was already present in Australia, the European Wasp, so that a retrospective validation of the prediction surface could be conducted.

Stage 5. Synthesis.

The research team drafted the final report. The discussion focuses on the potential for applying these approaches in routine risk analyses, and for developing them further so that they are better suited to biosecurity needs.

2. Exposure Pathway

2.1 The BA model

Biosecurity Australia's IRAs generally provide four estimates, one each for importation, distribution, establishment and spread (Figure 2.1). That is, IRAs classify the first five steps in Figure 2.2 into a judgement of the likelihood of importation (Figure 2.1).

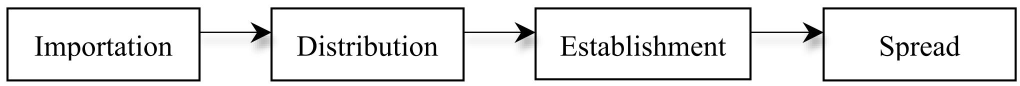


Figure 2.1. Steps in Biosecurity Australia's model.

The judgement is based on a subjective consideration of a range of factors including;

- distribution and incidence of the pest in the source area,
- occurrence of the pest in a life-stage that would be associated with the commodity,
- volume and frequency of movement of the commodity along each pathway,
- seasonal timing of imports,
- pest management, cultural and commercial procedures at the place of origin,
- speed of transport and conditions of storage compared with the duration of the life cycle of the pest,
- vulnerability of the life-stages of the pest during transport or storage,
- incidence of the pest likely to be associated with a consignment, and
- commercial procedures (e.g. refrigeration) during transport and storage in the country of origin, and during transport to Australia.

In this study, experts provided quantities for the more detailed structure (Figure 2.2).

2.2 Pathway diagrams for pest entry, establishment and spread

An exposure tree was constructed showing the exposure pathway for the insect (Figure 2.2). It was used by the project participants as a template for elicitation protocols, Bayes net construction, Monte Carlo simulations and probability bounds analysis. The exposure pathway in Figure 2.2 led to nine questions, one for each step in the chain (Figure 2.3).

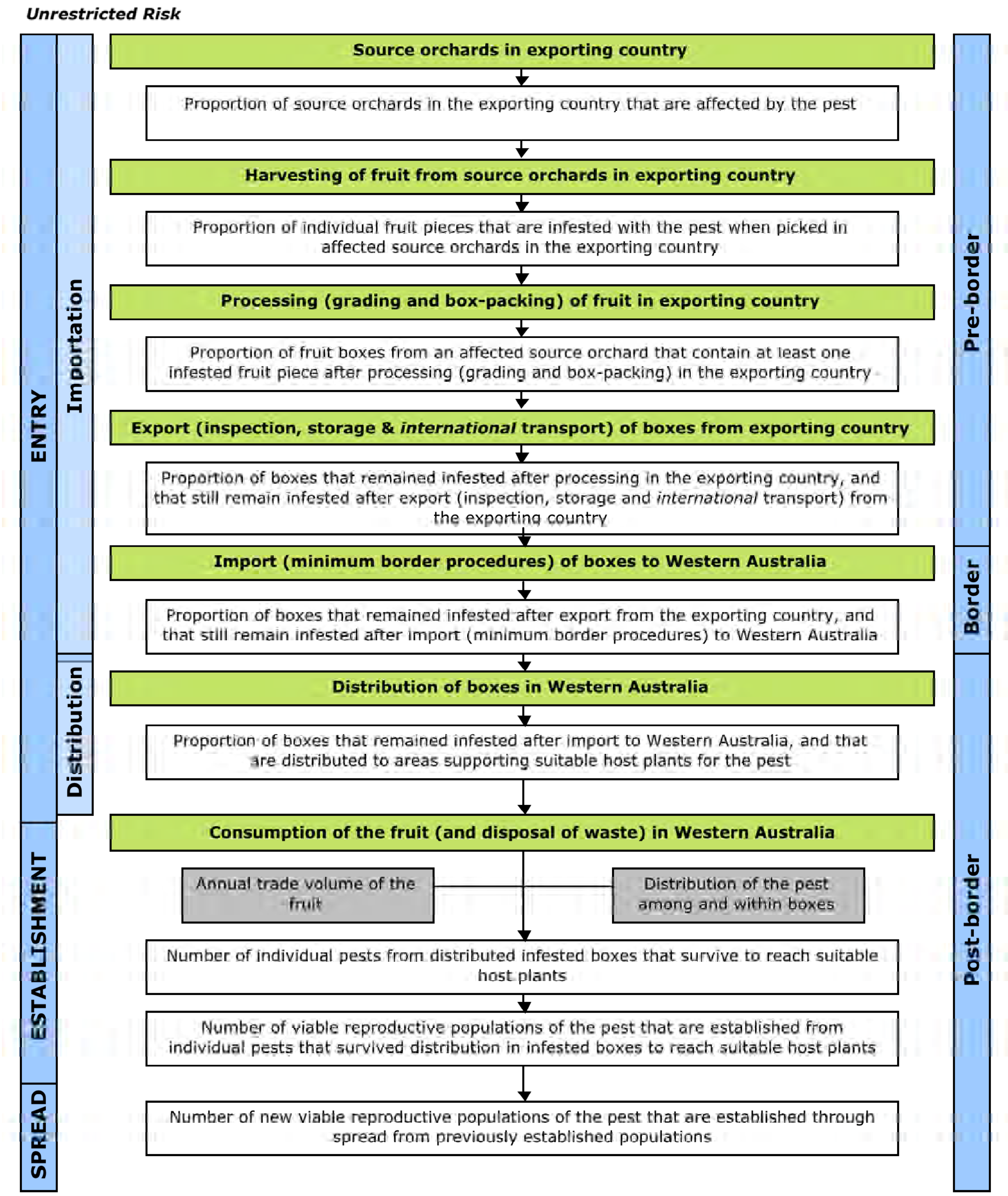


Figure 2.2. Steps in the exposure pathway for insects on imported fruit.

Notes:

- The terms 'affected' and 'infested' refer to the presence of at least one pest; i.e., an 'infested' fruit contains at least one pest, an 'infested' box contains at least one infested fruit and an 'affected' orchard contains at least one infested fruit.
- The 'proportion of fruit remaining infested' refers to the net effect of pest survivorship and cross-contamination (in this implementation, assumed = 0).
- It is assumed that, if infested fruit are found during inspection, the affected boxes are disposed and the remaining boxes are inspected further, but the entire shipment is not cancelled or the pathway otherwise changed.
- The details of distribution, establishment and spread have been bundled into general steps in the above diagram, covering the multiple steps and factors set out in IRAs.
- Each of the steps in the above diagram themselves represent bundled detailed steps, which are modelled implicitly in the minds of the elicited experts, rather than explicitly as model steps.
- Each of the steps in the above diagram is conditional on the previous step, except for the first step.

- It is assumed that the collection and/or elicitation of data is suitably conditioned to account for transitions from one unit to another (i.e., orchards to fruit, fruit to boxes, boxes to insects and insects to populations).

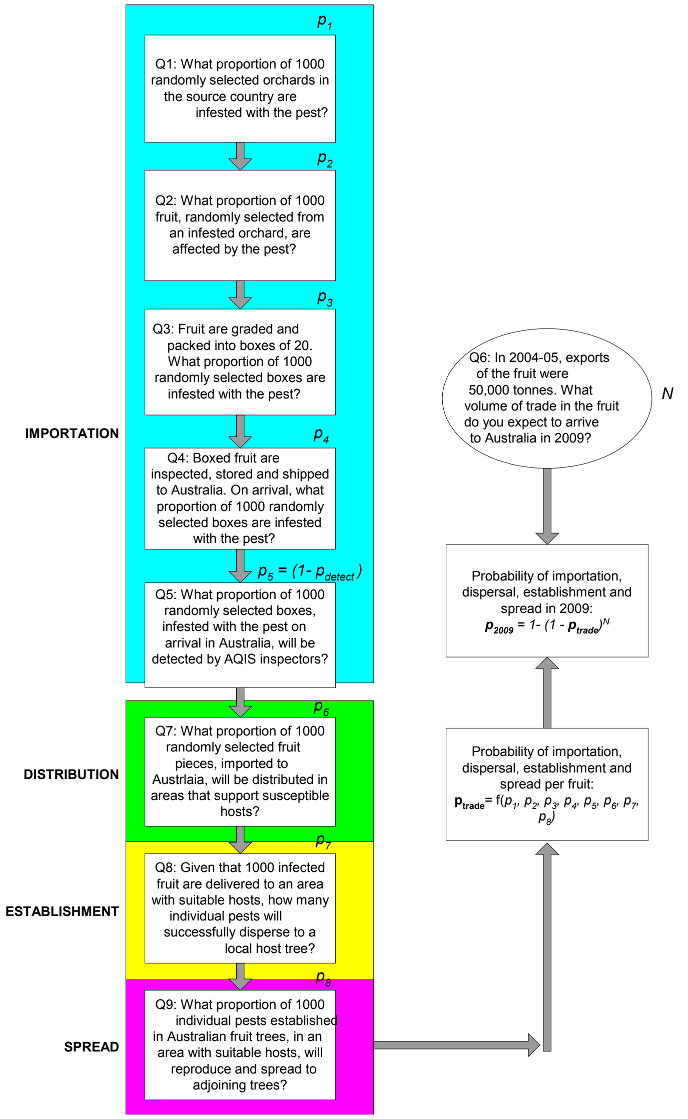


Figure 2.3. Questions associated with each step in the exposure pathway for the pest imported in fruit from the source country. The AQIS inspection assumes minimal on-arrival procedures that ensure the integrity of the consignment (i.e. no phytosanitary measures applied). Q9 includes expansion of geographic range as defined by ISPM5. The model implies the unrealistic assumption that an exact volume of trade would be known. Q3 asks for the proportion of boxes are infested, given that the boxes come from an infested orchard and Q4 assumes that the box was infested when it left the source country.

Trees showing exposure pathways have been used in some IRAs published by Biosecurity Australia and by similar agencies overseas.

2.3 Strengths, weaknesses, opportunities, threats

Trees representing exposure pathways reduce misunderstandings about what is meant by the terms and headings used to describe exposure pathways, and they make it easier to specify unambiguous, appropriately conditioned questions. Pathway diagrams are easy to construct. Even when the structure is simple, they may be useful in making the individual steps in a mental model clear. On the downside, a specific tree confines attention to one model of the system, reducing focus on other plausible structures. For example, the tree above diverts attention from the clustering that always occurs along importation pathways.

The procedures outlined below work equally well, whether the experts are asked a single question about importation (Figure 2.1) or several (Figure 2.3). If more steps are included, tree diagrams will become more important, as a way of ensuring the expert (and the reader of the report) understand what is meant by each estimate.

It is difficult for experts to estimate probabilities less than about 1 in 10,000, accurately. If an event is composed of several sequential steps, the task may be made easier by asking experts to estimate several higher probability events that combine to produce the estimate of the probability of the event in question. Disaggregation of importation into component steps is reputed to improve the estimation of low probability events (Cooke 1991). MacGregor (2001) warned that disaggregation should be only be used when the person making the estimates can make component estimates more accurately than the target estimate. The generality of this approach needs to be tested more thoroughly.

While none of the probabilities in the steps in this example is low enough to benefit from reducing the cognitive difficulties associated with small probabilities, Figures 2.2 and 2.3 illustrate how it might be done, using either numbers or words (probability terms tied to intervals).

The structure of recent IRAs encourages this thinking. The first five questions in Figure 2.3 reflect the logical steps outlined in Biosecurity Australia IRAs, which combine the independent components in a single assessment for introduction. Disaggregation and combination of the independent steps could form a useful part of routine training and internal validation. The answers that emerge from the combination of the disaggregated estimates should match the overall judgements, whether the assessments are made with numbers or with likelihood categories.

Exposure pathway diagrams may help clarify that the probabilities estimated at each step are conditional on the previous event. For example, in the model adopted by Biosecurity Australia (2008), the estimate of the probability of establishment assumes that entry has occurred. Readers could misinterpret the probabilities at each step to be the likelihood of the complete chain of events, from importation to the step under consideration. This issue is evaluated more fully below. Including a tree with associated commentary will reduce the potential for this misunderstanding.

Pathway diagrams are also important first steps in other analyses. Such diagrams are the first step in the development of Bayes nets, explored more fully below. In more complex situations than the example used here, there may be multiple exposure pathways, or it may be important to establish whether one control system could substitute for another. In the latter case, pathway diagrams are the first step in Hazard and Critical Control Point (HACCP) analysis, used effectively in the food industry in the US (Zach and Bier, 2009). HACCP can be time-consuming but may have a place in biosecurity in assessing existing operational procedures, identifying weaknesses and anticipating faults, especially when failures are critical. Pathway diagrams may therefore be may be an important element in assessing the implications of substituting one management system (or set of quarantine measures) for another, evaluating system equivalence and the potential for failures in the candidate system.

3. Selecting and questioning experts

3.1 Informal procedures

BA employs a range of scientific experts who, together with whatever external expertise is considered necessary and available, form consensus judgements of the likelihood of importation, distribution, establishment and spread of each pest of quarantine concern. Currently, people with appropriate training and experience are mostly drawn from Department staff and relevant State agencies, and occasionally from universities and industry, usually as part of small teams. Comments from reviewers and stakeholders on draft reports provide additional input on judgements of probabilities and consequences.

While there are no formal guidelines for selecting experts, government officers aim to identify the best-qualified and most experienced people to contribute understanding and data to the range of issues, including;

- production and quarantine practices in the source areas,
- capacity for managing pests and diseases off-shore,
- transport, packing and processing conditions and practices, and
- pest and host species biology, ecology, physiology, behaviour and taxonomy.

In general, experts integrate background information on the commodity, its treatment, and the interaction of these conditions with potential pest species. In doing so, they evaluate the credibility of source references, databases and observation records. They use this information to form an opinion about the likelihood of each of four steps on the exposure pathway. Species that have negligible risk of importation, distribution, establishment or spread are not considered further. Those pests that are present in the source area, could survive, be imported, establish and spread are evaluated further in a Pest Risk Analysis.

To illustrate the potential for differences of opinion among experts, the most experienced participant in the study described below was provided with the information in the IRA document and asked to classify the probabilities using the standard procedure.

Table 3.1. Expert judgements of the likelihoods of importation, distribution, establishment and spread of the hypothetical insect into Western Australia.

Q1. **Probability of importation**: The likelihood that insects will arrive in Western Australia with the importation of fruit is?
Expert 8: HIGH BA: HIGH

Q2. **Probability of distribution**: The likelihood that the insect will be distributed to Western Australia in a viable state, as a result of the processing, sale or disposal of fruit, is?
Expert 8: HIGH BA: LOW

Q3. **Probability of establishment:** The likelihood that the insect will establish in Western Australia, based on a comparison of factors in the source and destination areas considered pertinent to its survival and reproduction, is?
Expert 8: MODERATE BA: MODERATE

Q4. **Likelihood of spread**: The likelihood that the insect will spread within Western Australia, based on a comparison of those factors in source and destination areas considered pertinent to the expansion of the geographic distribution of the pest, is?
Expert 8: MODERATE BA: MODERATE

Expert 8 commented on the same scenario and disagreed with the assessment of the probability of distribution by Biosecurity Australia. Such disagreements between experts are common. Standard procedures provide a single judgement for each step, thereby masking disagreements between experts. It is rare to find levels of agreement between experts such that they would settle on a single risk category, especially if the underlying value lies close to the boundary between two categories.

3.2 Selecting experts

Where there is a diversity of opinions, groups may estimate facts better than individual experts (Clemen and Winkler 1985, Armstrong 2001). If a topic is contentious, from a scientific or social standpoint, it is helpful to involve experts from a range of positions on the issue, and to define 'expertise' broadly to include anyone with relevant, substantive knowledge (Franklin et al. 2008, Burgman et al. 2010).

The criteria for selecting experts may vary from case to case, but in all cases they should be relevant to establishing a person's expertise, or define one of the factors that underlie a contentious issue (e.g., for biosecurity issues, jurisdiction). Even though the topic of this study was not contentious, participants were included to cover a range of criteria. Some of the participants were inexperienced. Some were well versed in the technical detail of risk analysis, with limited experience of pests. Two were taxonomic specialists with substantial field experience. Several had relevant experience conducting pest risk assessments.

Eight experts with a range of skills and experience were identified through professional networks (Table 3.2). Seven were able to participate in an exercise in Canberra to estimate the likelihoods of entry, establishment and spread.

Table 3.2. Experts involved in assessment.

Expert	Highest qualification	Expertise	Years of relevant experience	Number of relevant publications
1	PhD	Pest risk analysis, systematics, entomology	29	20
2	PhD	Pest risk analysis, quarantine management	25	25
3	PhD	Genetics, plant pathology	30	60
4	PhD	Entomology, taxonomy, biosecurity risk analysis	35	20
5	PhD	Ecological risk analysis, quantitative methods	3	12
6	PhD	Entomology	28	120
7	BSc	Entomology, spatial modelling	1	0
8*	PhD	Entomology, fruit pest ecology	40	15

*Expert 8 participated by email and phone, but was not at the workshop

3.3 Questioning experts

Our approach to elicitation combines several elements that have been shown in empirical studies to outperform alternatives, in terms of providing accurate information efficiently. Integrating the components has made use of key cognitive theories, detailed in a range of ACERA publications. The components of the protocol are Delphi facilitated workshops (Burgman 2005), frequency question format (Gigerenzer et al. 1991), calibration questions (Cooke and Goosens 2004), four-step interval elicitation (Speirs-Bridge et al. 2010), visual feed-back (Lichtenstein and Fischhoff 1977; Lichtenstein et al. 1982; Tufte, 1983) and statistical aggregation of different judgements.

3.3.1 Delphi facilitated workshops and software

The Delphi method was created in the 1950's at the RAND Corporation to allow groups of experts to contribute to complex problems. The first Delphi application replaced a computer simulation by subjective estimates by experts (Dalkey and Helmer, 1951, cited by Turoff *et al*. 2006). It is a structured group interaction involving 'rounds' of opinion collection and feedback. It provides a framework that avoids or reduces many contextual and psychological issues (such as anchoring, underspecificity and social dominance) that affect unstructured approaches to gathering expert opinion (Kerr and Tinsdale 2004). The approach is especially useful where data are lacking, uncertainty is large and the primary source of information is informed judgement (Hess and Kin 2002, Turoff et al. 2006).

The project developed software that implements the Expert Procedures Guide (Cooke 1991) with the following elements;

- 'frequency' question format,
- calibration questions,
- four-step elicitation,
- visual feedback, group discussion, and revision of estimates, and
- aggregation.

The software is web-based and supports small, informal groups, face-to-face facilitated workshops, or remote web-based participation. The workshop procedure is described in greater detail below.

3.3.2 Frequency question format

Gigenenzer (2002, 2007) has established that in many situations, when people are asked questions about relative frequencies (probabilities), their answers are more reliable when questions are posed in a frequency format than when expressed as proportions, percentages or probabilities. People find it easier to visualise and reason with relative frequencies expressed as whole numbers. For example, experts can usually discriminate more easily and more reliably between '2 in 1000' and '15 in 1000' than between 0.002 and 0.015 (Spaetzler and von Holstein 1975). We use Gigerenzer's 'frequentist' approach in this procedure. For example, instead of asking for the probability that a fruit from an export region will contain a pest, we ask, given 1000 fruit collected randomly from the exporting region, how many will contain a pest?

3.3.3 Calibration questions

Cooke (1991) developed a protocol for testing and calibrating the accuracy of expert estimates of facts, such as probabilities, and applied it to industrial and power generation problems in Europe. We developed an approach based on Cooke (1991) and Cooke and Goosens (2004) in which, in addition to the questions generated by the scenario (Figure 2.3), we ask a number of additional so-called 'calibration' questions designed to test the skills and knowledge base of the participants. The results of these questions may provide a basis for weighting the experts’ opinions or excluding participants who perform relatively poorly. More importantly, the results may be fed back to the participants to assist them to improve their ability over time to estimate unknown quantities.

We prepared calibration questions suitable for a range of expert contexts, including human, animal and plant diseases and pests, quarantine systems, and transport and treatment of traded commodities. We select a subset of these questions to suit the context at hand. The suitability of the questions is evaluated by the workshop host or a senior expert with relevant experience, to ensure the question set is ‘reasonable’ (within the domain of expertise of the participants, representative of the kinds of knowledge they could be expected to have).

3.3.4 Four-step interval questions

We employ a four-step questioning (elicitation) procedure developed by Speirs-Bridge *et al.* (2010, Figure 3.1). This combines two approaches that have previously been successful in reducing overconfidence; the 3-point question format (Soll and Klayman 2004) and anticipated confidence levels (Teigen and Jorgensen 2005).

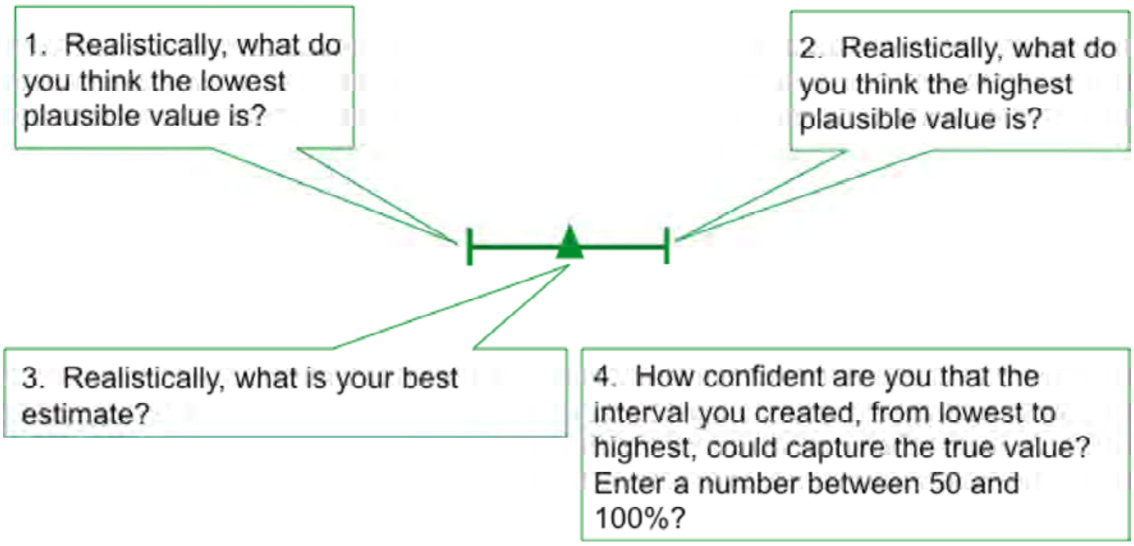


Figure 3.1. The four-step question format (Spiers-Bridge *et al.* 2010)

This approach makes use of Klayman *et al.*’s (2006) observation that by asking several questions, experts assess several different hypotheses and sources of information in constructing their estimate. Having produced an interval, the four-step procedure then asks the expert to estimate their level of confidence that it captures the true value. This step takes advantage of the observation by Teigan and Jorgensen (2005) and Klayman *et al*. (2006), that interval assessments encourage a more complete search of experience and knowledge.

3.3.5 Visual feed-back and revision

More than two decades of research in cognitive psychology has established that people interpret uncertainty more reliably and consistently when it is presented in visual form, rather than as numbers (e.g. Tufte 1983). Overconfidence is reduced when experts receive regular, systematic feedback (Lichtenstein and Fischhoff 1977, Lichtenstein et al. 1982). It is important to distinguish between feedback to reduce overconfidence, and feedback to revise estimates. The latter step is a verification of responses, a kind of double-check, to give experts a second opportunity to consider the data and their estimates (Kadane 1998, vonWinderfest and Edwards 1986). We use both kinds of feedback in our approach.

Our procedure uses a web-based program (‘ET’: http:/acera.unimelb.edu.au/software) to collect and display expert judgements of intervals, based on 4-point elicitation. Experts make individual judgements of intervals. Once complete, the intervals are recalibrated and displayed as 80% intervals (e.g., Figure 3.2). Participants discuss their judgements, introducing new information, reconciling the understanding of words and context and resolving differences of opinion. Once discussion is complete, participants have the opportunity to reassess their initial judgements and revise their intervals.

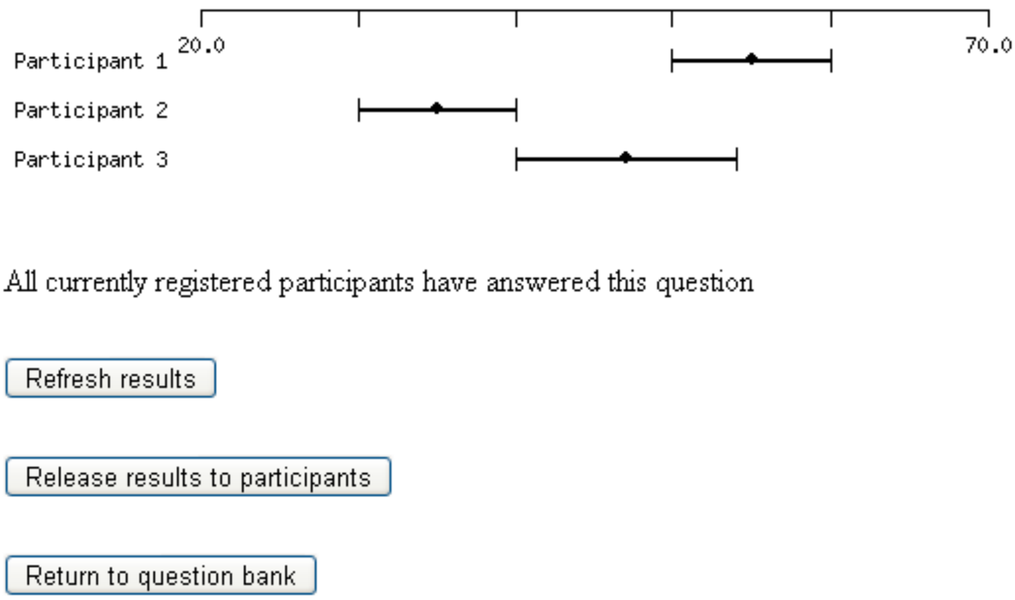


Figure 3.2. Screen capture of the software package that implements the experts procedure in facilitated workshops. The software is web-based and may be used in remote elicitation exercises.

The choice of a scale for display may affect participants’ judgements (a framing effect). We use a linear scale, but we could have employed a probability scale that expands and therefore emphasises the regions close to 0 and 1, or a log scale that emphasises values close to 0. Instead, this software employs the option of enlarging the linear scale in the region surrounding the combined interval, so that participants explore the further refinement of their judgements in iterations of the Delphi process using a rescaled display.

The final estimates provided by the experts, following discussion, represent the limit of behavioural consensus. Participants were unwilling to modify their estimates any further, on the basis of the discussion and the opinions of others in the group. This arises routinely in elicitation exercises. People reach a point at which they are no longer willing to adjust their assessments (e.g., Scholtz and Hannmann 2007). This creates a need for formal, numerical methods to resolve remaining differences.

3.4 Trials of the four-step elicitation procedure

Two types of questions are asked: elicitation questions where the answer to the question is unknown; and calibration questions where the answer to the question is known to the facilitator but not to the participants. In most workshops, we have included 10 calibration questions. Typically, the workshop takes place at a ‘neutral’ venue and each participant uses a computer with an internet browser and internet access. However, while these resources make the process easier to facilitate, they are not essential.

We have applied the four-step elicitation process in six workshops and a variety of expert contexts. They include human epidemiologists and medical practitioners, marine scientists, the workshop convened for this project and described in more detail below, and two State agencies responsible for biosecurity risk analysis (mixtures of animal and plant biosecurity specialists) (Figure 3.3).

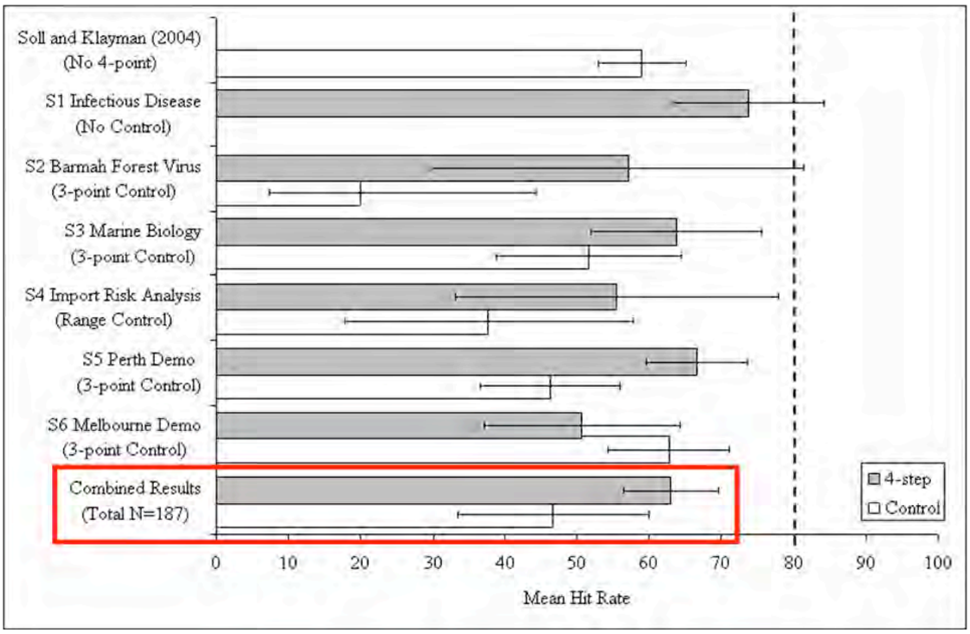


Figure 3.3. Results of responses to calibration questions, which are questions from the domain of the expert where the answers are known to the facilitator, but not to the participants. The hit rate is the relative frequency with which the expert interval encloses the truth within their specified interval. The bottom-most pair of bars represent the combined estimate of performance over all the studies (based on meta-analysis).

When the answer to the question is known by the facilitator, but not necessarily by the experts, in general, the four-step elicitation procedure results in 80% intervals that enclose the correct answer roughly 65% of the time. The next best procedure, the three-step elicitation procedure (used as a control), results in intervals that enclose the correct answer roughly 45% of the time. This represents a substantial improvement in the calibration of uncertainty in expert estimates.

3.5 A typical workshop agenda

A structured workshop may last from two hours to two days, depending on the number of questions (reflecting the complexity of the exposure pathway). We have run the procedure effectively in facilitated workshops involving as few as 5 and as many as 25 people with a broad variety of backgrounds, experience and training. Details of background information, preparation, and running a workshop are dealt with fully in the user’s manual and facilitator’s guide that accompany the web software noted above. A typical agenda for a workshop is:

(i) Welcome and overview.
(ii) Introductions, description of the context, discussion of background information.
(iii) Two to three practice questions, so that participants get used to the software and facilitation process.
(iv) Calibration and elicitation questions as follows.
   a. The underlying assumptions are agreed for each question, language-based misunderstandings are resolved and a shared understanding of the question is reached across the group.
   b. Each expert then enters their initial interval estimate.
   c. Errors are trapped and nonsensical entries are corrected.
   d. Each expert then reviews a visual representation of their interval, adjusted to an 80% confidence level. They have the opportunity to modify their estimate and once happy, they release their response for comment.
   e. Once all experts have provided an initial response to a question, the facilitator displays all initial estimates together on a single graph. Each interval is identified with a participant ID (Figure 3.3).
   f. A discussion is facilitated with the group, exploring the differences and/or similarities between the locations and the widths of the intervals.
   g. Following this discussion, the experts are provided with an opportunity to revise their 80% interval.
   h. The initial and final values for each expert are then stored but not displayed for a second time, preserving the right of each expert to submit a private decision.
(v) Wrap up and summarise workshop.
(vi) Export results for post-hoc analysis.
(vii) Provide feedback to experts on their performance on the calibration questions.

3.6 Strengths, weaknesses, opportunities, threats

3.6.1 Selecting experts

In biosecurity risk assessments in Australia, the pool of experts may be small, and some or all may have a stake in the outcome of a decision. In some analyses, all credible external experts may have a conflict of interest, either representing or being involved to some extent with a debate.

One solution to this problem is to limit involvement to people who have no stake in the decision, often, people from the Government agency itself. While this reduces conflict of interest, it risks complacency that results in available experts being considered adequate, simply because they are readily available. It also forgoes the substantial credible knowledge held by participants who have a stake in the outcome of the decision at hand.

Another solution is to ensure that a wide range of scientific opinions is included in an assessment. If a debate is divisive or polarised, it may be effective and informative to include experts from all sides. The agency or decision-maker can use the arguments from opposing positions to cross-examine evidence, before reaching a decision. This process is employed in the current protocols, through the opportunity for stakeholder input on draft IRAs. The information from stakeholders could be used earlier and to better effect, if tools were available for capturing that knowledge.

The range of opinions may be enhanced by widening the criteria for participation to include people without formal qualifications, but who have substantial direct experience. They may also include people with training but with relatively limited experience. Results documented below illustrate that people without formal training, or without extensive experience, can be very effective. This provides a platform for broadening the set of potential contributors to an IRA, alleviating the problem of the small pool, improving the range of insights to an issue, reducing the chances that something important will be overlooked, and reducing conflicting interest with no real reduction in accuracy of the assessment.

Documenting who contributes to the assessment in an IRA report, together with their background, experience, qualifications and any other relevant attributes, would provide readers with a sense of the breadth and credibility of opinions that have contributed to them. This may improve the acceptability of judgements, if the selection of expert input has been sufficiently broad and inclusive. This is done routinely in the United States and Canada. There is the potential for the disclosure to create problems for the people involved, when the issue is very divisive.

Ensuring input from a broad spectrum of experts will make the process of gathering experts together more difficult and time consuming. Remote contributions (by phone or the web) will alleviate these costs, to some extent. The trade off between convenience and credibility will be difficult to judge, and will vary from case to case. A broad and inclusive spectrum of experts may be most useful when the issue is contentious or the decision is ambiguous.

3.6.2 Questioning experts

While discussion and informal consensus is the most common approach to resolving differences of opinion, including in biosecurity contexts, the approach is susceptible to power imbalances between individuals, motivational and contextual biases, and a range of cognitive frailties (Burgman 2005). Informal approaches may also mask honest differences of opinion and borderline cases, which arise routinely in expert judgements.

A structured procedure may circumvent many of the well-documented weaknesses of unstructured and informal approaches. The elicitation system described above has four main strengths:

- It gathers relatively well-calibrated information about knowledge uncertainty;
- It reconciles differences in the understanding of words and context, reducing language-based misunderstandings that impede many risk analyses;
- It is designed to reduce the effects of the most pervasive cognitive difficulties in group-based, subjective risk assessments;
- Lastly, it will improve the performance of experts over time by providing them with feedback about their estimates of facts.

It could be used when questioning experts, for either numerical probabilities, or qualitative intervals.

Structured workshops are feasible in operational conditions because they need not require any more time, money, preparation or people than are currently employed to make expert judgements. The structured approach does, however, require that people know how to use the system, and have experience in the details of its operation. It will be more effective if there is robust, intuitive software, supported by appropriate guides for facilitators. Once operational, it would take no more time to gather estimates of the likelihood of entry, establishment and spread of a pest using the structured approach, than it would when using informal methods. The main limitation is that a facilitator needs (1) formal training or (preferably) experience in facilitation, (2) sound knowledge of the nature of the questions and how the outcomes will be used, and (3) sound knowledge of probability.

We have found some people have difficulty reasoning with numbers greater than 1000. This effectively limits responses about probabilities in some circumstances to the range [0.001, 1]. However, it is important to keep in mind that this problem exists, whether people are using a frequency format or are attempting to reason with probabilities or qualitative terms such as ‘very low’ of ‘negligible’. The frequency format simply makes this problem apparent.

On the other hand, we have observed experts make reliable assessments for probabilities of about 5 in 100,000 for disease rates in human populations, when the context was well understood and groups had an opportunity to exchange information, face to face. Thus, it is unclear how well people assess the probability of events, even when using a frequency format, when the probability is much less than 0.001, as is the case in many biosecurity problems.

It is important to have calibration questions well-matched to the task at hand, because calibration does not generalise well between disciplines. The more closely the calibration questions reflect the elicitation questions, the more useful they will be in scoring experts (Cooke 2004). Calibration questions and feedback may be confronting for some experts. However, we have not found anyone unwilling to participate (as yet).

Overall, an opportunity exists to improve the quality of expert judgements, either as quantities or as qualitative intervals, and to be able to capture and report their uncertainty, currently hidden in IRAs. The potential cost is that the information could be used to open new areas for criticism of risk assessments. The benefits (of more accurate and better calibrated estimates, validated by external data, and derived from a more inclusive set of experts) may not be worth the costs. Even if the structured methods are efficient and effective in the operational environment, greater transparency, especially in dealing with uncertainty, carries significant social and political risk, exposing the assessments to criticisms that presently are not possible. No other jurisdiction is explicit about the treatment of uncertainty.

These tools may have greatest value and cause least disruption, at least initially, if they were used in the training of risk analysts. They would allow analysts to explore the sensitivity of results, to evaluate the importance of uncertainty, and to test the effects of framing, anchoring, dominance and other inherent biases, so that they are better equipped to anticipate and deal with these issues in real evaluations. Staff would learn how to discriminate language-based misunderstanding, knowledge uncertainty and natural variation. It is valuable simply to observe the range of arbitrary differences of opinions about facts that arise because of language and context. Knowing how to discriminate between knowledge uncertainty and natural variation will assist people to anticipate these sources in routine risk analyses, and to manage informal procedures appropriately.

4. Combining expert opinions

Disagreement among experts can occur for a number of reasons including differences in:

- expertise,
- data available to each expert,
- interpretation of the data,
- theoretical biases,
- appreciation of the various uncertainties involved (Regan et al. 2002), and
- social and political agendas.

Differences can be resolved by voting or consensus. In voting methods, experts hold their views fixed and the voting system tries to deliver a group preference, which may or may not coincide with the preferences of the members of the group. Voting is a *compromise* method since the group outcome is typically not the unanimous opinion of the experts. There are many voting systems and they deliver different results (Saati 2001). They range from simple majority rule in which a disagreement is settled by a “straw poll”, to preferential voting systems and weighted scoring systems. They are not considered further here.

Consensus methods find a joint outcome that represents the position of the group as a whole. The methods can be formal or informal. While informal consensus is the most common approach to resolving differences of opinion, including in biosecurity contexts, the approach is susceptible to power imbalances between individuals, motivational and contextual biases, and cognitive frailties (Burgman 2005). Facilitated informal consensus uses an independent facilitator to negotiate an agreed position. A skilled facilitator may anticipate and circumvent many of these problems.

4.1 Informal consensus

In informal consensus, when differences of opinion arise, experts discuss differences of opinion, seek clarification of terms or understanding of systems, and acquire further data where possible or where they are available. In some informal consensus methods (such as with a jury, where unanimous approval is required from all members), the experts modify their views so that at the end they all agree. In others, experts reconcile differences as far as possible, and then defer to an arbitrator to make a final decision. BA uses this approach. Divergent opinions are not typically recorded or used in subsequent analyses, even though they may represent valuable information on the reliability (and uncertainty) of consensus opinions. Differences of opinion and uncertainties about classifications of likelihoods are not recorded by any biosecurity agency internationally.

4.2 Formal consensus

A variety of formal consensus methods use explicit algorithms to resolve differences of opinion (Lehrer and Wagner 1981, Regan *et al*. 2006, Steele *et al*. 2008). Statistical methods for aggregating opinions include linear pooling, weighted averaging and Bayesian aggregation. Some of these tools are outlined in more detail below. We compared four such methods—Lehrer-Wagner consensus, Regan consensus, linear pooling and Bayesian pooling—during the workshops and analysis described under section

4.2.1 Lehrer-Wagner and Regan consensus

The Lehrer-Wagner method requires that experts provide both their own judgements of a parameter or quantity, and judgements of the relative expertise of others in the group.

Expertise can be modelled with linear algebra, providing weights for the opinion of each expert that accords with the esteem in which they are held by their colleagues. This model guarantees a consensus outcome in any case where reasonable conditions are satisfied (Lehrer and Wagner 1981). We asked each participant in the elicitation exercise to give a score (out of 10) to each other participant, reflecting their assessment of how reliable each person would be in answering the biosecurity questions. This information was used to create weights for each person's responses.

The consensus convergence described above requires each individual in the group to assess all other group members and then assign a weight to each member according to their degree of respect for or agreement with that member's expertise or views on the issue at hand. It can be time consuming to obtain these estimates. Group members may also conceal their true agenda or distort their weights to achieve a preferred outcome. The assignment of a numerical value on a person's degree of respect for each of the other members in the group is abstract and may be provocative. To compensate for these potential problems, Regan et al. (2006) discovered that they could adapt the original convergence model to use a weight of respect based on the strength of the difference in the criteria weights assigned by individuals in the group. This method implements the consensus convergence model after the elicitation session, without requiring additional time from the experts.

4.2.2 Linear pooling

For numbers in the interval [0, 1], such as probabilities, we fitted a Beta distribution function to each of the expert's intervals by using a standard optimisation routine (the *optim* function in R). The optimisation routine finds the parameters (α, β) of a Beta distribution with 10th and 90th quantiles that match the lower and upper bounds (respectively) of the expert's second interval by minimising the following sum of squares function:

$$SS = \left[t_{q1} - pbeta(p_1, \alpha, \beta)\right] + \left[t_{q2} - pbeta(p_2, \alpha, \beta)\right]$$

where t_q is a vector of target quantiles (10th and 90th for an 80% confidence interval), $\boldsymbol{p}$ is a vector of the lower and upper bounds of the quantity in question (e.g. the proportion of infected orchards) and *pbeta* is the R function that returns the Beta distribution function. This may be done more easily using the the FitDistr function in library MASS (see http://rss.acs.unt.edu/Rdoc/library/MASS/html/fitdistr.html).

For variates that take values on the real line (e.g. the volume of trade in 2009) we identified the mean and standard deviation of a normal distribution function whose 10th and 90th quantiles matched the lower and upper bounds (respectively) of the expert's second interval using standard relationships between the mean, standard deviation and quantiles of a normal distribution.

It is important to note that in both cases the distribution fitting routines described above assume or require: (1) the expert provides a lower and upper bound; (2) a level of confidence is assigned to this bound; (3) and a standard distribution is assumed for the variate (e.g. Beta or Normal).

The distribution fitting procedure described above resulted in seven distribution functions (one for each expert) for each of the nine questions in the risk assessment model. For each question, we also fitted a pooled distribution using the simple linear pooling methodology described in O'Hagan et al (2006). In this approach, we pooled the distribution function (as opposed to the density function) but the function is otherwise identical:

$$F(\theta) = \sum_{i=1}^{n} w_i F_i(\theta)$$

where the subscript *i* represents each expert, *n* is the total number of experts, $F(\boldsymbol{\theta})$ is the distribution function with parameter vector $\boldsymbol{\theta}$, and w_i is the weights attributed to each expert's opinion. Note that in this example we have weighted each expert equally such that $w_i = 1/n$. The resulting pooled distribution is therefore a simple average of all of the experts’ individual distributions. The linear pooling method is intuitive and will capture the range of opinions expressed by each expert. The pooling procedure gives a non-standard distribution and cannot therefore be reproduced with standard methods. This is not, however, a serious impediment to any subsequent calculations. Linear pooling is sensitive to the selection of experts, making it important to span the breadth of opinions to provide a balanced (unbiased) consensus.

Having fitted and pooled the distributions for each question in the risk assessment model, the analysis was in a position to compare the results of a Monte Carlo Simulation (MCS) with a Probability Bounds Analysis (PBA). The Probability Bounds Analysis was performed using an R library (Scott Ferson, Applied Biomathematics, pers. comm.) calibrated against the software RiskCalc using the methods described in Ferson *et al*., (2004), Tucker and Ferson (2003) and Ferson (2002).

4.2.3 Bayesian pooling

We used a Bayesian hierarchical approach assuming normality to synthesise information from the group of experts after suitable transformation of the expert data. We used a probit transformation on the elicited probabilities and a log transformation on the volume of trade. The approach can be considered a “weighted evidence” approach where each expert’s response is weighted by their precision around their estimate. The approach assumes that the experts’ responses are comparable but independent. Unfortunately, most experts give answers that covary to some degree (e.g. Morris 1977, Winkler and Clemen 1999).

It also assumes that more confident experts carry greater weight. This assumption will be valid if experts are well calibrated – that is, if they have good understanding of the limitations of their own knowledge. Yaniv and Foster (1995, 1997) noted that people often give more precise intervals for questions where they are more knowledgeable (resulting in a constant hit rate, but with more precise intervals).

The practical advantage of this method is that where experts tend to agree, it generates a more precise aggregated interval, something the linear pool doesn’t take advantage of. A wider interval results when experts give very disparate responses, again something the linear pool doesn’t target.

The results for two of the nine questions that were asked of each expert are displayed below (Figure 4.1), one showing good agreement and the other showing substantial disagreement. For each plot, we show the expert predicted mean and corresponding 80% credible interval. The green line shows the pooled mean, $\hat{\mu}$ and the 80% credible interval for the pooled mean. The blue line shows the pooled interval $\tilde{\theta}$ and its corresponding 80% credible interval. The prior is shown beneath the *x*-axis, represented as the specification for the probability distribution.

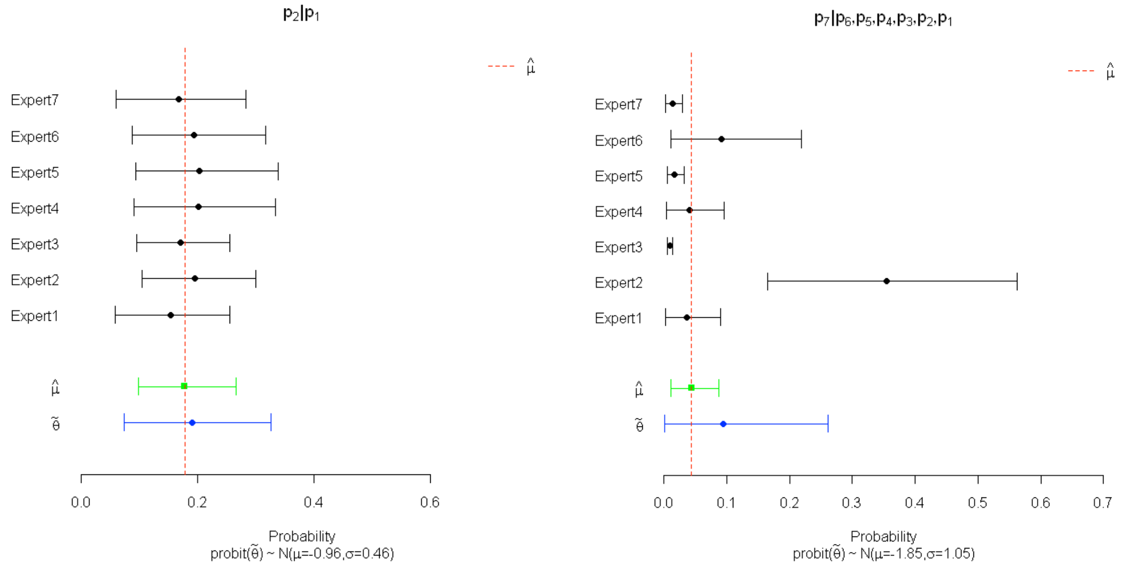


Figure 4.1. Responses from 7 experts to questions 2 and 7 (Figure 2.3), elicited using the 4-point elicitation procedure, put to the experts in a frequency format.

The assumption of independence between experts may be violated for several reasons. Often, the pool of experts is small and they share common sources of information, research projects, students, and / or work environments. In addition, the elicitation protocol encourages experts to share information and discuss differences of opinion. This will reduce irrelevant sources of variation such as misunderstanding of words and context, but at the cost of reducing the independence of opinions

4.3 Outcomes of the workshop

Seven experts (Table 3.2) participated in the workshop convened to estimate the probabilities of entry, establishment and spread of insects described in Figures 2.2 and 2.3. The expert procedures guide described above was employed. All the experts were provided with background information on the pest species, its taxonomy, biology and distribution.

Of the 10 calibration questions, six were quantities, and four were proportions. For the Lehrer-Wagner method outlined above, experts were asked to rank their own ability to answer questions relevant to plant biosecurity (self rating) and the abilities of other participants (peer rating) on a scale of 1 to 10. As participants were not all familiar with one another, weights were assigned after each expert had introduced themselves and described their knowledge and experience. Each person’s rank was calculated as the average weight assigned to them by the other six experts.

Expert performance was assessed using two metrics: calibration of intervals and error of the best guess (distance of the best guess from the truth). The distance of the best guess from the truth was standardized across questions using the inter-quartile range of the best guesses supplied by the seven experts, so that;

$$\textit{Distance from truth} = \text{abs}\{[\textit{Best guess} - \textit{True value}]\} / [Q_{75} - Q_{25}]$$

This value was then averaged across the 10 questions for each expert to give a final score.

Calibration was measured by the proportion of times (out of 10) that the specified 80% confidence interval enclosed the truth. A well-calibrated expert will enclose the truth 8 times out of 10. Fewer than eight ‘hits’ implies overconfidence; more than eight implies under-confidence.

Consensus estimates were calculated using a range of methods including the average of the best guesses, the median of the best guesses, and the Lehrer-Wagner consensus model. For the latter, we used the individual expert responses and the assigned average peer ranking, normalized to one to give a percentage weighting. Consensus estimates were also calculated using the Lehrer-Wagner consensus model, but with the individual expert responses and the weightings of respect calculated using the Regan et al. (2002) method based on distance between responses.

4.3.1 Results

Participants in the workshop rated their own expertise roughly in line with the judgements of their peers in the workshop (Figure 4.2). That is, participants had a good idea of how their level of expertise would be assessed by others. The assessment of expertise correlated closely with factors associated with standard measures of expertise, including qualifications, age, years of experience, publications and membership of professional organisations. The result in Figure 4.1 reflects the results obtained in several additional studies to date. Typically, the correlation between self-ranking and peer-ranking of expertise is about 0.6.

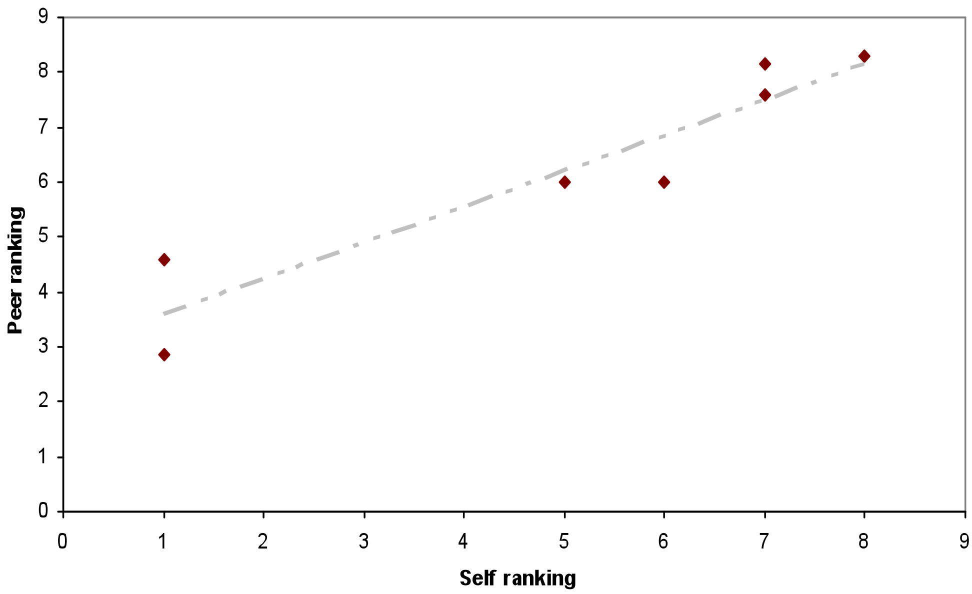


Figure 4.2. Average peer rating versus self rating (showing least squares line of best fit). Expertise was rated on a scale of 1 to 10. A high rank indicates the person was held in higher regard.

The ‘80% hit rate’ records the relative frequency with which the adjusted 80% intervals enclosed the truth, for each of the participants. There is no strong relationship (Figure 4.3) between peer rating of expertise and 80% hit rate, indicating that as people’s assessed expertise increased, their level of over-confidence in their answers did not change appreciably. That is, people whose expertise was assessed as being higher (in the opinion of their peers) and who tended to be older and have more relevant experience tended to enclose the truth about as often as those whose expertise was judged to be lower. If the left hand outlier is removed, the relationship is, in fact, negative. That is, older and more experienced people were more over-confident. The result here is based on a sample size that is too small to generalise. In the several studies such as this conducted to date, there was no important relationship between age or average per ranking and calibration. These results arise even though there is no 'penalty' for wide intervals; experts could include the truth 100% of the time, simply by specifying arbitrarily wide intervals.

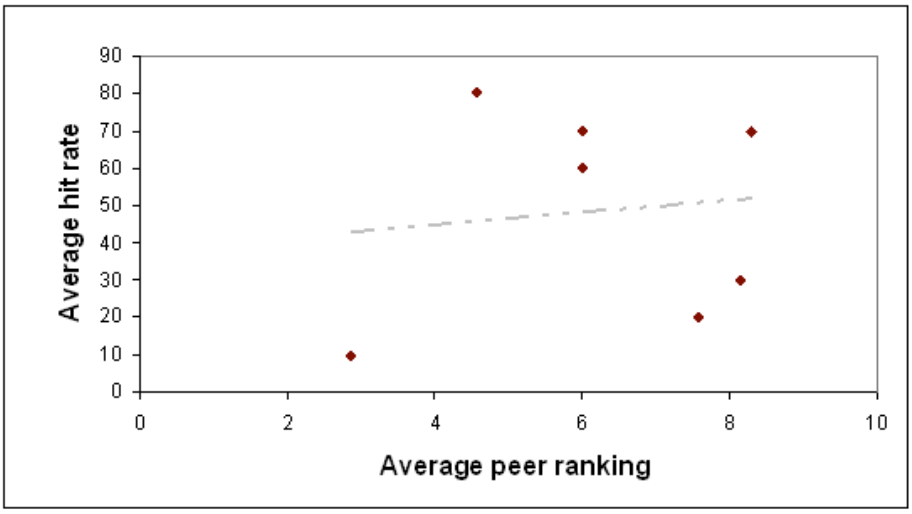


Figure 4.3. Average hit rate versus average peer rating of expertise (showing least squares line of best fit; the regression slope is not significantly different from zero). In contrast to a *proper* scoring system, such as the Brier score (Lichtendahl and Winkler 2007), this system has no penalty for wide intervals. Even though experts could get a 100% score by saying any probability is between 0 and 1, they routinely underestimate the width of intervals.

The critical question then arises: how did average peer rating of expertise relate to performance, in terms of distance from the truth? We would expect, if peer ratings of expertise have merit, that experts with greater experience should provide answers that are, on average, closer to the truth than less experienced and less well regarded experts.

In this example, we found a weak relationship between peer ranking and distance from the truth (Figure 4.4). According to the line in Figure 4.4, more experienced and better regarded experts provided answers that were on average, slightly closer to the truth than their less well regarded and less experienced counterparts. However, if the point at the left hand side is removed, the relationship is positive, suggesting more experienced people are less reliable. In the other experiments we have performed so far, the general result is that, once participants have enough experience or training to understand context and jargon, there is a very weak relationship between experience and accuracy, or none at all.

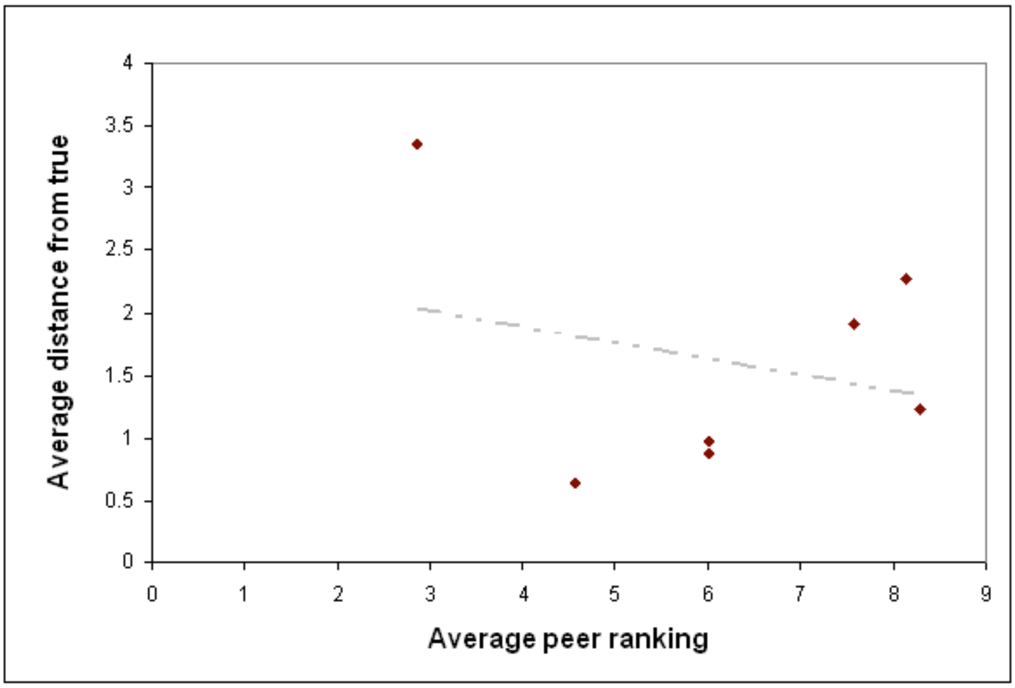


Figure 4.4. Distance of best guess from the true value versus average peer rating (showing least squares line of best fit; the regression slope is not significantly different from zero and is dominated by the point at the upper left of the diagram).

Above, we outlined linear and Bayesian pooling, alternative approaches to pooling expert judgements that capture both the central tendency and spread of the expert estimates. We also trialled several ways of combining the experts’ best guesses. We explored the average of the best guesses, the group median, the most highly regarded expert, consensus based on distance weights (Regan consensus) and consensus based on peer assessments (Lehrer-Wagner consensus). The judgments made by the most highly regarded and best performing experts are reported as a basis for comparison.

The median of the group gave the best results (Figure 4.5). Regan consensus based on distance weights performed next best. The most highly regarded expert performed at approximately the same level as the group average.

As above, this sample size is too small to support generalisations. However, the results are consistent with results from previous studies that generally find the mean and median of a group of experts perform well (Jose and Winkler 1992, Clemen and Winkler 1999). In the other experiments we have conducted, the group mean and median outperform the best-performing expert in the study, and substantially outperform the most highly regarded expert.

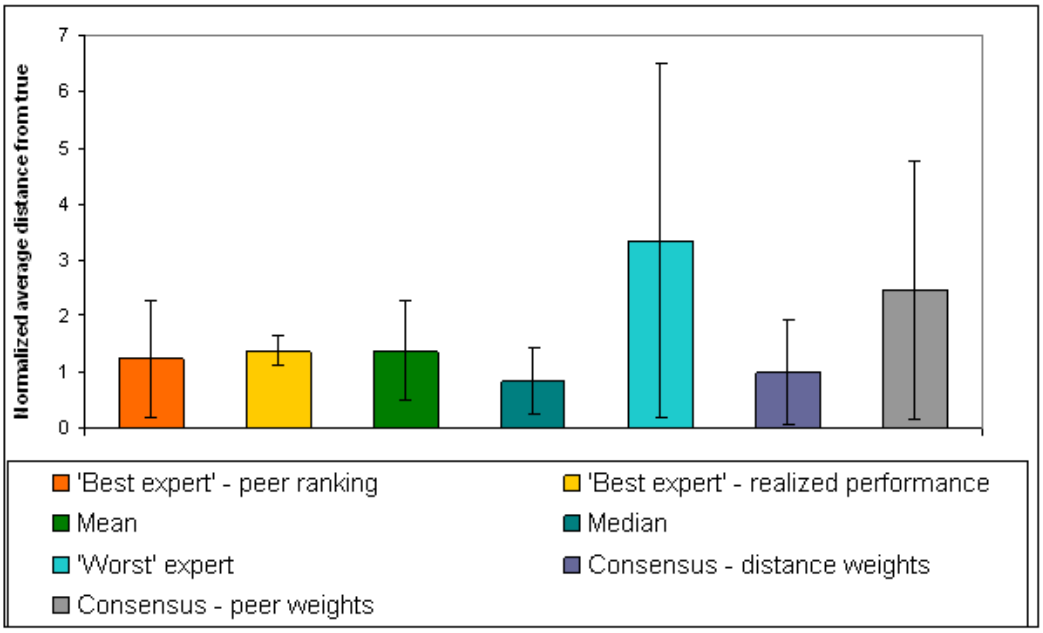


Figure 4.5. Performance of different consensus methods over the 10 calibration questions. Summary of average performance for best guess for the different methods. Error bars show the 95% confidence intervals based on the 10 questions.

These results are based on the average performance over all 10 calibration questions. There was substantial variation among the questions (Figure 4.6). Some were relatively easy to answer, in the sense that most people were close to the truth (questions 3 and 6), whereas some were more difficult (e.g., questions 1 and 5).

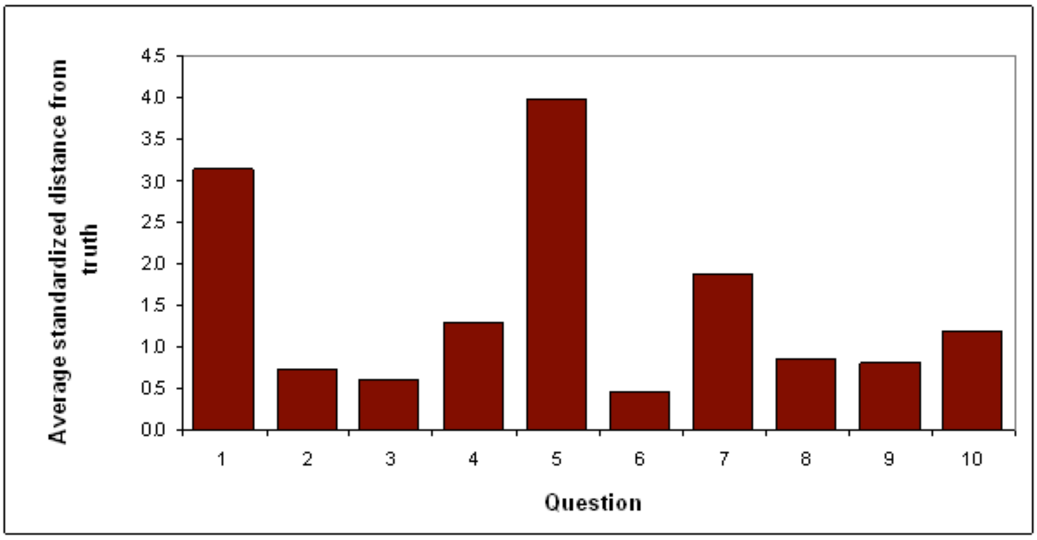


Figure 4.6. Average standardized distance from the truth for each of the calibration questions.

4.4 Strengths, weaknesses, opportunities, threats

When disagreements arise, participants can deal with them through informal discussion, leading to resolution of differences of opinion. Such deliberations may be effective, but they are susceptible to dominance effects and motivational pressures. Where experts disagree, even after discussion, it may provide an impetus to find more information, or to carry the range of opinions through to the end of the assessment, in case the decision is, in fact, ambiguous.

Formal methods (such as consensus modeling or averaging) avoid many of the pitfalls of ad hoc methods for consensus because they are inclusive of all group members, they use all the relevant information and not just information favorable to a particular point of view, they are blind to dominant personalities within the group and they allow for quantitative treatments of uncertainty in the decision-making process (Regan *et al.* 2006). Furthermore, individual estimates are recorded in formal decision-making method, enabling one to trace the ultimate decision back to its initial inputs.

We trialed both Lehrer-Wagner consensus and Regan consensus. In both cases, the weights made little difference to the consensus estimate generated by linear pooling with equal weights. Where there were differences, linear pooling out-performed these protocols. This is in contrast to the results of Cooke and his colleagues, who find that performance-weighted results outperform the best-performing individual and the unweighted group average. Further research would be required to establish whether Cooke’s (1991) method could provide better estimates than (unweighted) linear pooling.

Previous evaluations of peer rankings and weighted averages have found that they can be sensitive to relatively small estimation errors and can lead to negative weights or result in forecasts that are outside the range of values of the individual forecasts (Winker and Clemen 1992). We cannot recommend Lehrer-Wagner or Regan consensus methods because of the additional effort involved in using them. Linear pooling is simple to apply and gives intuitively reasonable outcomes. The simpler approaches are robust and efficient if couched in a protocol that guards against overconfidence, individual dominance behaviours and anchoring, such as the guide outlined above.

The structured elicitation protocol provides two important advantages. First, once a person has enough training or experience to understand the jargon and context of a problem, and if they have an opportunity to listen to others and to amalgamate their knowledge with background information, their judgements may be as sound as the most experienced person in the group. The evidence above is not sufficient to support this general conclusion, but we have repeated this experiment several times with the same qualitative result. Second, the group mean and median usually outperform the most highly regarded individual. These two advantages provide compelling reasons for contemplating the deployment of these tools, at least in training.

5. Interval arithmetic

The objective of interval analysis is to carry quantitative uncertainties through chains of calculations in a way that is guaranteed to enclose an estimate with at least the surety required. Interval arithmetic is a simple tool that can be used to carry uncertainties through chains of reasoning or calculations.

Occasionally, when confronted by uncertainty, analysts use ‘conservative’ judgements in place of their best estimates of a parameter. The conservative judgement reflects an upper or lower quantile of a distribution, even if it is not explicitly estimated. The thinking is that conservative judgements will protect environmental or economic values from unexpected eventualities. The trouble with this kind of strategy is that the level of protection resulting from the analysis is arbitrary and unknown. For example, estimation of the likelihood of entry, establishment and spread involves the product of four ‘quantities’. The product of the (say) 95th quantiles of a set of distributions is not equal to the 95th quantile of the product of the distributions. The degree to which these two quantities diverge, a measure of the conservatism of the estimate, depends on the number of arithmetic operations and the distributions involved. The results may be hyperconservative, much more protective of the economy or the environment than the data warrant. Hyperconservatism results in litigation, direct and indirect market costs, blocking potentially beneficial products and misdirecting scarce resources. Interval arithmetic avoids these issues because it calculates both the upper and lower bounds for a parameter.

5.1 Rules for combining intervals

Biosecurity Australia in many IRAs defines a set of intervals associated with words that reflect likelihoods of importation, distribution, establishment and spread (Table 5.1). These intervals are described above. We term them ‘fixed-interval’ bounds. They are combined with a set of rules, distilled into a matrix (Figure 5.1).

Table 5.1. Probability intervals employed by BA. These intervals are intended to be guides to the meanings of the words.

Language-based description of likelihood	Probability Interval
High	0.7 - 1.0
Moderate	0.3 - 0.7
Low	0.05 – 0.3
Very Low	0.001 – 0.05
Extremely low	10^{-6} – 0.001
Negligible	0 - 10^{-6}

	High	Moderate	Low	Very low	Extremely low	Negligible
High	High	Moderate	Low	Very low	Extremely low	Negligible
Moderate		Low	Low	Very low	Extremely low	Negligible
Low			Very low	Very low	Extremely low	Negligible
Very low				Extremely low	Extremely low	Negligible
Extremely low					Negligible	Negligible
Negligible						Negligible

Figure 5.1. Rules for combining probability intervals.

5.2 Rules for interval arithmetic

Intervals may be calculated from data, estimated from expert knowledge, or calculated using optimistic and pessimistic model assumptions. There are many kinds of intervals (see Burgman 2005), but the intervals elicited in the expert procedure outlined above are closest to a Bayesian credible interval. A Bayesian interval can be the shortest interval that contains a specified amount of a (posterior) probability distribution, or it may be the amount of a probability distribution contained within specified bounds (Jaynes 1976). An interval must be accompanied by *‘some indication of the reliability with which one can assert that the true value lies within it’* (Jaynes 1976, p. 179). We term these, ‘expert-prescribed’ intervals.

Box 5.1 shows the rules for interval arithmetic. Each operation ensures that the result encloses the true value with at least the level of confidence specified for the individual, expert-prescribed intervals.

Box 5.1. Rules for interval arithmetic.

Addition

$$[a_1, a_2] + [b_1, b_2] = [a_1 + b_1, a_2 + b_2]$$

Subtraction

$$[a_1, a_2] - [b_1, b_2] = [a_1 - b_2, a_2 - b_1]$$

Multiplication

$$[a_1, a_2] \times [b_1, b_2] = [a_1 \times b_1, a_2 \times b_2]$$

Division

$$[a_1, a_2] / [b_1, b_2] = [a_1 / b_2, a_2 / b_1]$$

Operations with a constant ($h \geq 1$)

$$h \times [a_1, a_2] = [ha_1, ha_2]$$
$$h + [a_1, a_2] = [a_1, a_2] + [h, h] = [a_1 + h, a_2 + h]$$

Powers ($b_1, b_2 \geq 1$)

$$[a_1, a_2]^{[b_1, b_2]} = [a_1^{b_1}, a_2^{b_2}]$$

The method is transparent and simple, but depends on assumptions about dependencies. If the results of interval arithmetic do not straddle a decision threshold, then dependencies and other uncertainties may be ignored because they do not affect the decision. If, however, the results do straddle a threshold, we need to find out more about sources of uncertainty and dependencies between variables, and make decisions that take into account the possibility of being wrong.

Interval arithmetic gives assurances about the reliability of the results by making conservative assumptions about dependencies. To compute the bounds, it assumes that quantities that are added or multiplied are perfectly, positively correlated, and it assumes that quantities that are subtracted or divided are perfectly negatively correlated.

5.3 Interval arithmetic

Applying the rules in Box 5.1 implies the following interval operations;

High [0.7, 1.0] x Low [0.05, 0.3] x Moderate [0.3, 0.7] x Moderate [0.3, 0.7]

= [0.0032, 0.147] (5.1)

Applying the rules in Figure 5.1 to these same values, we obtain

High x Low x Moderate x Moderate

= Low [0.05, 0.3]. (5.2)

This result is a general feature of the rules for combining intervals. That is, the intervals resulting from Figure 5.1 do not match precisely the interval operations for the same information and, in general, the results of applying the rules are more pessimistic (result in higher values) than those resulting from equivalent interval calculations.

One of the attributes of interval arithmetic (and probability arithmetic) is that operations are associative. That is, the order in which arithmetic operations is specified does not affect the result. This is true of the operations in Box 5.1, but not of those in Figure 5.1. For example, changing the order in (5.2) gives

Moderate x Moderate x High x Low

= Very Low [0.001, 0.05]

Thus, the rules for combining words in Figure 5.1 do not always result in conservative estimates. The system could be made more consistent and transparent by adjusting the rules in Figure 1 to comply with interval arithmetic. A more informative treatment would be provided by questioning experts about their precise judgements of likelihood and the associated intervals of uncertainty, combining them (by linear pooling), and expressing the result as a best guess and interval (i.e., by using the procedures outlined in Sections 3 and 4 above).

5.4 Strengths, weaknesses, opportunities, threats

The current method uses fixed-interval bounds for the likelihood estimates for each step (entry, establishment and spread). Intervals acknowledge uncertainty. The fixed-interval bounds are intended to capture the uncertainties inherent in all practical IRAs.

Interval arithmetic provides an opportunity to enhance thinking about uncertainty and to make the results of combining the likelihood estimates for entry, establishment and spread more informative. Specifically, interval arithmetic would generate the following benefits, if used in place of fixed-interval bounds and the associated rules;

- information that currently is lost in assigning judgements to a fixed-interval would be retained in an expert-prescribed interval,
- operations would be precisely consistent with the rules of probability,
- the results would enclose the truth with at least the prescribed degree of surety,
- improvements in knowledge and understanding could be reflected in successive reductions in the widths of intervals,
- expert performance could be improved by feeding back to individuals their accuracy and calibration on test questions (see Section 3).

Some of the costs in deploying expert-prescribed intervals include the need for skilled facilitation to question experts (Section 3), and explicit methods for combining judgements (Section 4).

It may seem that the assumptions about dependencies in interval arithmetic make the intervals as wide as possible. In fact, they are as narrow as possible, while remaining faithful to the specification that nothing is known about dependencies between variables. If something is known about such dependencies, the arithmetic should reflect it. For example, Monte Carlo simulation (section 6) can handle dependencies when they are specified exactly, and probability bounds analysis (section 8) can handle the arithmetic if dependencies are unknown, or are known only approximately.

A weakness of intervals is that they do not use all available information about a number. Using an interval loses information about the central tendency, standard deviation, sample size, distribution shape or any intuition an expert may have about these things. Lastly, intervals are only appropriate for numerical uncertainty. Many instances of linguistic and non-probabilistic uncertainty should be treated in the most appropriate manner for their sub-category (Regan et al. 2002).

Intervals have sharp boundaries. It may seem unreasonable to say nothing about a distribution within the limits, and yet to specify the boundaries as though they are exact. There are some solutions, not outlined here. For instance, fuzzy numbers are essentially stacks of intervals, each level of which represents a different degree of surety about the boundary (Kaufmann and Gupta 1985).

6. Monte Carlo simulation of entry-establishment-spread

Monte Carlo simulation (MCS) has been employed on a number of occasions to evaluate the probabilities of entry, establishment and spread for IRAs, both in Australia and elsewhere. MCS was undertaken in this study for completeness, and to provide input to the probability bounds analysis that follows. MCS was used here to characterise statistical uncertainty in estimates of the unrestricted probability of incursion of pests into Western Australia.

6.1 The BA model

The model used by BA is based on pathways, with estimates made for the likelihood of pests surviving undetected through the stages of entry, establishment and spread based on a synthesis of information about pests and the pathway environment. The only explicit pathway steps in the BA model are introduction, distribution, establishment and spread, with an implicit sub-structure within the model that is evident only in the sub-headings and information presented within the entry, establishment and spread sections of the IRA (see Chapter 2 for further discussion of the IRA model structure and information synthesis).

6.2 The Monte Carlo model

The pathway model used for the Monte Carlo simulation (MCS) is shown conceptually in Figure 2.2 and more explicitly in Figure 2.3. This pathway model makes explicit many of the sub-steps that are implicit in the BA model. Furthermore, Figure 2.3 shows the questions asked of experts to elicit estimates of probabilities (p_1 to p_8), volume of trade (N) and the equations used to combine these estimates (including I: $p_{2009} = 1-(1-p_{trade})^N$). Internal (individual) and external (group) data synthesis was undertaken by the experts at each of the model steps to generate probability estimates. That is, experts were able to reflect on their initial judgements, and to listen to the arguments and opinions of others, before making a final estimate. This approach was supported by a structured workshop (see Sections 3 and 4). Importantly, the elicitation questions were asked such that the expert probability estimates were conditional on the estimates for the preceding questions, an important element in ensuring that the model arithmetic (multiplication of probabilities) is sound.

6.3 Monte Carlo implementation and results

The MCS was implemented in Excel based on the exposure pathway (Figures 2.2 and 2.3) and parameter estimates supplied by experts. Beta distributions were fitted to individual estimates, which were pooled to produce input distributions[1] (see Section 4. Thirty thousand random numbers were generated for each of the input distributions. The random numbers were used to sample the distributions, generating 30,000 output values for the probability of entry, establishment and spread (I). The output histogram for I is shown in Figure 6.1 (linear scale) and Figure 6.2 (log scale). Summary statistics for I are shown in Table 6.1.

[1] Linear pooling was used. The distribution data are too extensive to include in this report but can be supplied on request.

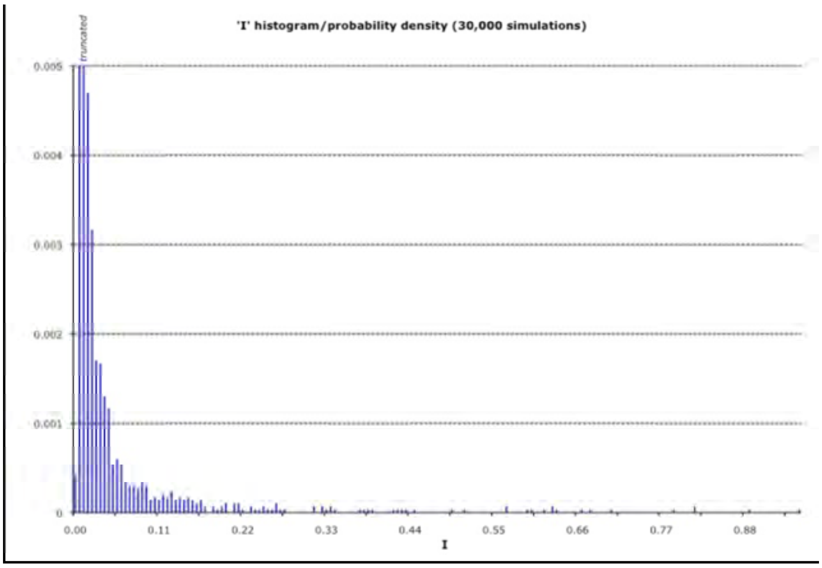


Figure 6.1. MCS output histogram for the overall probability of entry, establishment and spread (*I*) on a linear scale.

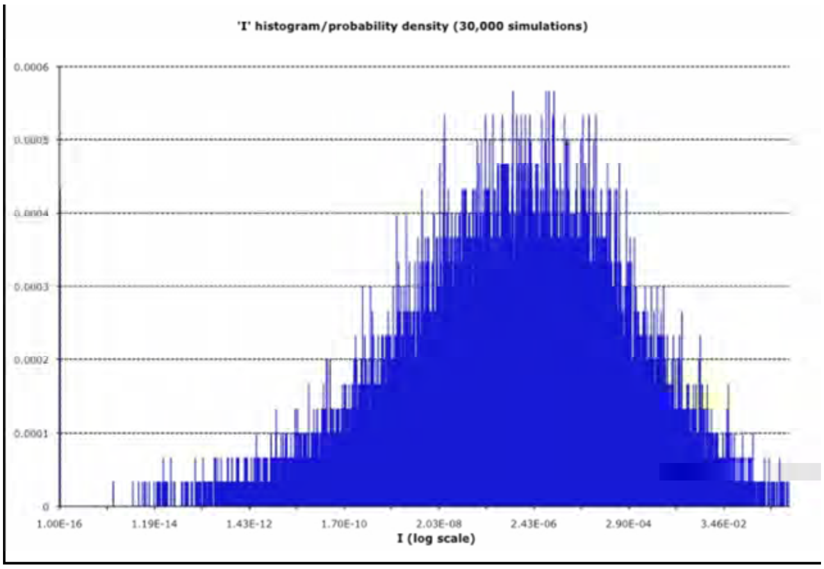


Figure 6.2. MCS output histogram for the overall probability of entry, establishment and spread (*I*) on a log scale.

Table 6.1. Comparison of MCS and IRA estimates of probabilities of entry, establishment and spread.

Step	Monte Carlo Simulation				Import Risk Analysis	
	5th %ile	Median	Mean	95th %ile	Category	Range
Entry	2.67×10^{-7}	8.47×10^{-4}	0.05	0.32	Low	0.05 to 0.30
Importation*	1.41×10^{-5}	0.03	0.23	1.00	High	0.70 to 1.00
Distribution*	0.12	1.00	0.89	1.00	Low	0.05 to 0.30
Establishment	0.15	1.00	0.88	1.00	Moderate	0.30 to 0.70
Spread	0.26	1.00	0.92	1.00	Moderate	0.30 to 0.70
Overall (I)	$\boldsymbol{1.85\times10^{-10}}$	$\boldsymbol{1.44\times10^{-6}}$	$\boldsymbol{1.86\times10^{-3}}$	$\boldsymbol{2.42\times10^{-3}}$	***Low***	***0.05 to 0.30***

* Note: individual importation and distribution probabilities are multiplied to calculate the probability of entry.

The results summarised in Table 6.1 reflect highly skewed uncertainty around a mean annual probability of overall entry, establishment and spread of the pest (*I*) of about 0.002. The median is three orders of magnitude smaller. As a result, the mean falls within the BA likelihood category ‘Very Low’ (0.001 to 0.050) while the median falls within the category ‘Extremely Low’ (0.000001 to 0.001). This is lower than the IRA estimate of ‘Low’ (0.05 to 0.30), although the uncertainty in the MCS results is extreme, with the minimum and maximum values straddling the entire range of BA likelihood categories (from ‘Negligible’ to ‘High’) and the 95% confidence intervals straddling ‘Negligible’ to ‘Very Low’ (Table 6.1).

These wide bounds reflect the uncertainties in the estimates of the parameters provided by the experts and the stochastic combination of these uncertainties across the nine input parameters. If the estimate for *I* were calculated by combining the estimates for the probabilities of entry (importation and distribution), establishment and spread using interval arithmetic instead of the matrix of rules, the estimate would range from 0.00315 to 0.147 (‘Very Low’ to ‘Low’), which is closer to the estimate arrived at by MCS (see Section 5).

Differences between the qualitative and MCS estimates of *I* are determined by the differences between the inputs. These differences (see Table 6.1) are:

- *Entry*: the median and mean of the MCS estimate for entry fall into the ‘Extremely Low’ and ‘Very Low’ categories, respectively, while the IRA estimate is ‘Low’. The MCS estimate for entry is the stochastic product of the importation and distribution estimates, as follows:
    - *Importation*: the median and mean of the MCS estimate for importation fall into the ‘Very Low’ and ‘Low’ categories, respectively, while the IRA estimate is ‘High’.
    - *Distribution*: the median and mean of the MCS estimate for distribution fall into the ‘High’ category, while the IRA estimate is ‘Low’.
- *Establishment*: the median and mean of the MCS estimate for establishment fall into the ‘High’ category, while the IRA estimate is ‘Moderate’.
- *Spread*: the median and mean of the MCS estimate for spread fall into the ‘High’ category, while the IRA estimate is ‘Moderate’.

Explanations for these differences may include incorrect conditioning of probabilities, and differences in analysis or opinion in synthesising the available data. The IRA makes a qualitative judgement to produce the probability estimates. In probability and risk literature this would be called a *subjective* judgement, though the intention is that the experts’ judgements are *objective*. The use of experts as described in Sections 3 and 4 above does not make this process significantly more explicit, but it does make systematic use of the available expert knowledge (even if the data synthesis still occurs in the heads of the experts).

An alternative approach would be to construct a more detailed model (perhaps a Bayesian network containing sub-networks – see Chapter 6), with expert input to the model construction and parameterisation (i.e., mathematical representation of the model structures in the experts’ heads). This approach would require explicit and fully-quantitative incorporation of the available data, accounting for dependencies and structural uncertainty.

6.4 Assumptions

The output distribution from the MCS (Figures 5.1 and 5.2) provides an estimate of the likelihood of incursion of insects that includes the pooled uncertainty of the sampled experts. It is expected that this expert uncertainty is comprised primarily of the experts’ beliefs about natural variation and their levels of confidence in their own abilities to make predictions about each question. The workshop provides an opportunity for experts to reduce linguistic uncertainty, even if it is not eliminated entirely.

The MCS assumes that:

- The model structure is an accurate and complete representation of the pathways for entry, establishment and spread of insects (i.e., the MCS does not account for structural uncertainty).
- The experts' responses to each question were conditioned on the outcome of the preceding question.
- Fitting Beta distributions to each expert's estimate for each question and combining these distributions by linear pooling produced input distributions that represent the consolidated opinion of the sampled experts (see Section 4).

Potential additional analyses within or parallel to MCS include:

- Basic sensitivity analysis – varying input parameters by a fixed amount, one at a time, to determine which inputs have the greatest influence on the output.
- Second order MCS – simultaneous stochastic sampling of the parameters for each input distribution across a fixed range to assess the sensitivity of the output.
- Variation of distribution shape – varying the shape of the input distributions using other feasible distributions to assess the sensitivity of the output.
- Variation of model structure – varying the structure of the model using other feasible structures to assess the sensitivity of the output.
- Variation of dependency assumptions – the model assumes each step in the process is independent of the preceding steps (through appropriate conditioning of expert judgements). This may not be so, and the assumption may be relaxed by adding dependencies of various strengths to the steps in the process. (This has been done using PBA in Chapter 7, below).

6.5 Strengths, weaknesses, opportunities, threats

The applications of MCS in IRAs in the past have illustrated clearly the strengths and weaknesses of the approach. MCS is simply a way of solving stochastic equations. As such, it depends on the structure of the model, choices of parameters, distributions and other model assumptions. All models (both qualitative and quantitative) are abstract simplifications and are no better than the input data that they rely on. MCS is no exception.

MCS has some unique qualities. It is broadly accepted as a risk analytic tool within the community of risk analysts in most scientific disciplines. Its strength lies in the explicit representation of the structure of a problem, and the requirement to specify all model elements. From a strictly scientific viewpoint, this makes the analyst's thinking transparent and results are (asymptotically) repeatable. This is a hallmark of the scientific method. International trade agreements exhort risk analysts to be consistent, scientific and transparent. MCS seems to achieve these goals, superficially at least.

Of course, the mathematical formalism makes the thinking inaccessible to many people. It also provides antagonists ample opportunity to question assumptions and model choices. Typical applications of MCS have several other weaknesses, including:

- MCS obliges the analyst to specify exact moments (parameters) for input distributions, even when these are unknown. This creates unrealistic precision in outputs. Uncertainty regarding distribution shape and moments can be accommodated within Bayesian approaches to simulation modelling, but the tools and skills for implementing these analyses are rarer than are those for MCS, and are less widely known and accepted
- MCS, like most methods, ignores structural (model) uncertainty. Structural and parameter uncertainty represent lack of knowledge and can be overcome by employing

two-dimensional analysis, but the computational overheads and burden of interpretation of such analyses are large.
- Most implementations of MCS ignore dependencies, assuming unrealistically that model inputs are independent. This generates output distributions that are too narrow.

These impediments mean that, typically, the bounds generated by MCS are narrower than they would be were all sources of uncertainty treated thoroughly. With additional effort these sources of uncertainty can be addressed, for instance by using sensitivity analysis discussed above (Section 6.4).

Monte Carlo has a number of limitations, apart from sometimes getting it wrong. There are a number of things that Monte Carlo simulation cannot do (easily) (Ferson 1996). In addition to the limitations outlined above, it cannot propagate non-statistical uncertainty (ignorance versus statistical uncertainty), cannot yield a realistic answer when;

- dependencies are unknown (dependency uncertainty),
- input distributions are unknown (parameter or shape uncertainty), and
- model structure is unknown (structural uncertainty).

Solutions such as two-dimensional Monte Carlo and trial-and-error back calculation are cumbersome and computationally costly.

In addition to the technical difficulties outlined above, implementation takes time, adequate software and hardware, and it requires some statistical and mathematical skills. Computer limitations have been largely overcome, unless two-dimensional MCS or Bayesian analysis is contemplated, but skills are rare and may be difficult to retain in-house.

Ultimately, MCS is best suited to models with relatively low structural uncertainty (or a few distinct alternative structures) and relatively high amounts of available data, allowing distribution shapes and moments to be represented with a degree of confidence. Where these conditions are met, MCS provides the opportunity to view the model predictions and associated uncertainties with higher resolution. However, where these data are not available this resolution can be misleading (as it represents a single set of assumptions) and, although this provides a useful mechanism for dialogue about model details (given the transparency of the assumptions), it may not be the best mechanism for communication with decision-makers and stakeholders without support from other analyses.

7. Bayes net analysis of entry-establishment-spread

This component of the demonstration applies Bayesian networks to the import risk assessment, summarising a longer report from Nicholson and Korb. It includes constructing a Bayesian network (BN) that models entry, establishment and spread. This part of the work:

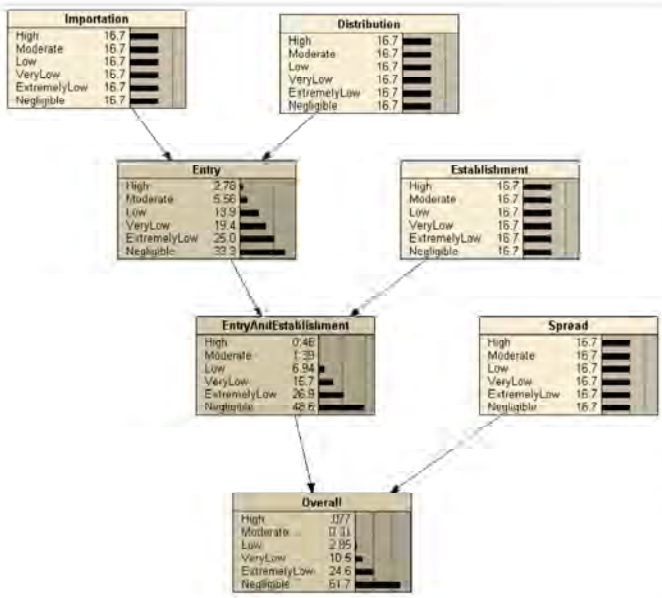
1. Demonstrates how a BN template can be constructed from an exposure pathways tree.
2. Incorporates the measures used in the current BA assessment method (e.g. negligible, extremely low, very low, low, moderate, high).
3. Extends the BN structure to incorporate uncertainty in these estimates.

7.1 Constructing a BN for the current risk assessment process

The current process can be modeled by a simple BN with deterministic relationships (Figure 7.1a). The nodes in this network are:

- “Root” nodes: Importation, Distribution, Establishment, Spread. These nodes are the input or evidence nodes.
- Intermediate nodes: Entry, EntryAndEstablishment.
- Target node: Overall “likelihood” of entry and establishment and spread.

Following current practice, the possible values for each node are: {High, Moderate, Low, VeryLow, ExtremelyLow, Negligible}, representing a qualitative “likelihood”. The entries in the deterministic conditional probability tables (CPTs) (see Figure 7.1b) were taken from the matrices showing how to combine qualitative likelihoods; e.g., Table 2.2 in Biosecurity Australia (2008; see Figure 5.1).



Node: Entry

Deterministic Function

Importation	Distribution	Entry
High	High	High
High	Moderate	Modera
High	Low	Low
High	VeryLow	VeryLo
High	ExtremelyLow	Extrem
High	Negligible	Negligil
Moderate	High	Modera
Moderate	Moderate	Low
Moderate	Low	Low
Moderate	VeryLow	VeryLo
Moderate	ExtremelyLow	Extrem
Moderate	Negligible	Negligil
Low	High	Low
Low	Moderate	Low
Low	Low	VeryLo
Low	VeryLow	VeryLo
Low	ExtremelyLow	Extrem
Low	Negligible	Negligil
VeryLow	High	VeryLo

Figure 7.1. (a) BN representing current BA Qualitative Risk Assessment for the Case Study with deterministic conditional probability tables for combining qualitative likelihoods. No input likelihoods added, deterministic nodes indicated by darker shade. (b) Deterministic CPT for Entry node (both are Netica BN software screen shots).

The model in Figure 7.1 has uniform priors for the root nodes. Priors for the root nodes for an individual case study should be obtained from experts. The overall assessments across these scenarios are uneven because of the threshold effect of the “rules” (for example, a combination of very low with any of high, moderate or low, gives very low). Further details of the functions of this Bayes net are provided below.

7.1.1 “Certain” Inputs

Other than the uncertainty in the categorisation of risk, the current method doesn’t use any distribution over that risk. We can represent this in a Bayes net by viewing the category ‘low’, for example, as a state (setting aside for the moment that it is indicative of a range of probabilities). If evidence is added as a fixed setting for each of Importation, Distribution, Establishment and Spread, there is a single Overall risk assessment. Figure 7.2 shows four such scenarios where all the risk inputs are considered “certain”, and hence are entered as a fixed 100% in the BN. Cases 3 and 4 are an example of how different combinations of risk can result in the same overall combination (‘very low’ in this example).

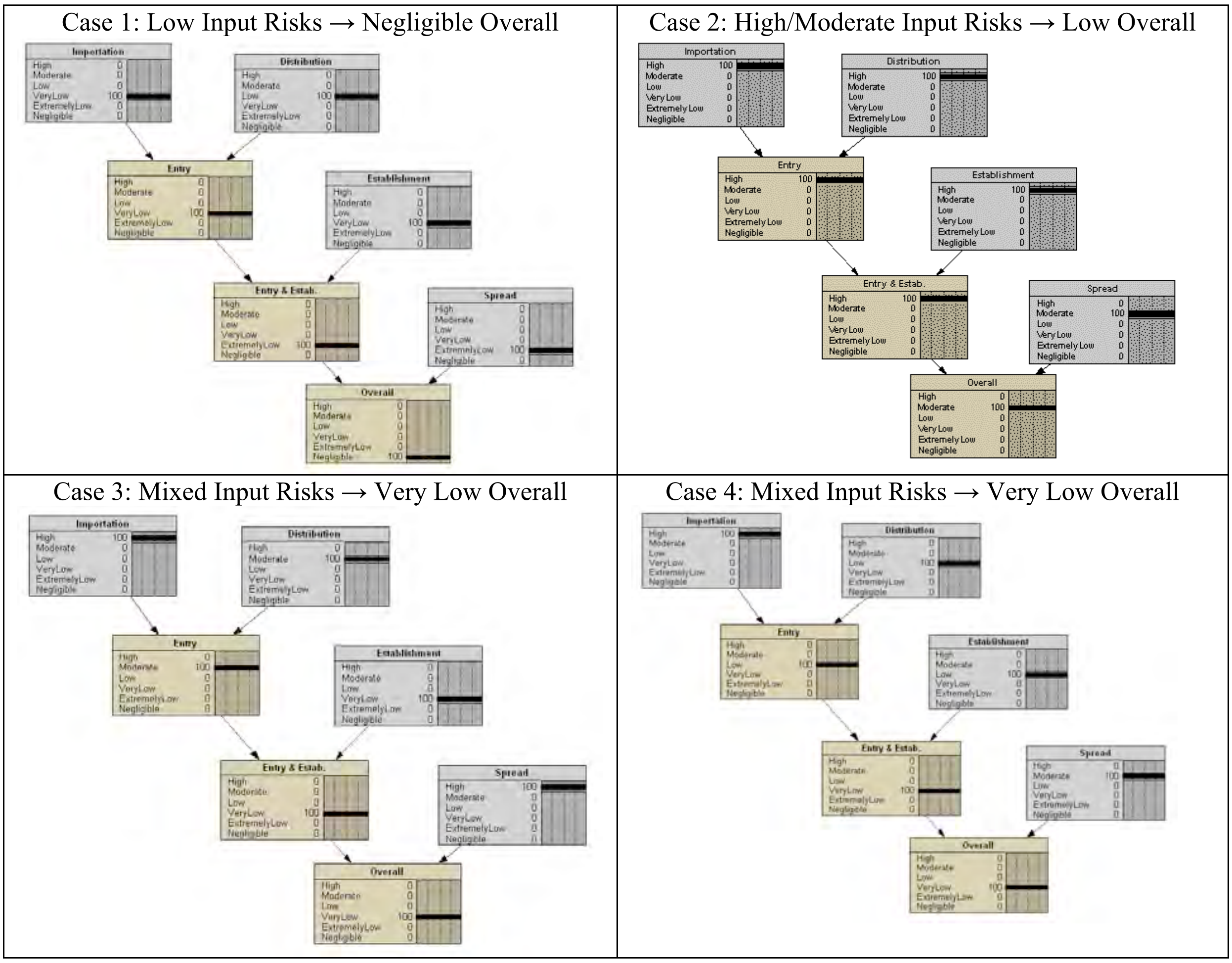


Figure 7.2. Four cases showing how the BN deterministically combines certain input likelihood for input nodes Importation, Distribution, Establishment and Spread, to give single output Overall likelihood. Some of these concepts are described elsewhere in this report, but are included here for completeness.

7.1.2 Representing uncertainty about input likelihoods

Adding uncertainty to the current process can be done in a number of ways, even within the existing risk assessment network structure. The input likelihoods (for Importation, Distribution, Establishment, Spread) can be entered as uncertain, either by giving a prior distribution or by adding so-called likelihood (uncertain) evidence. Some examples comparing certain and uncertain inputs are shown in Figure 7.3.

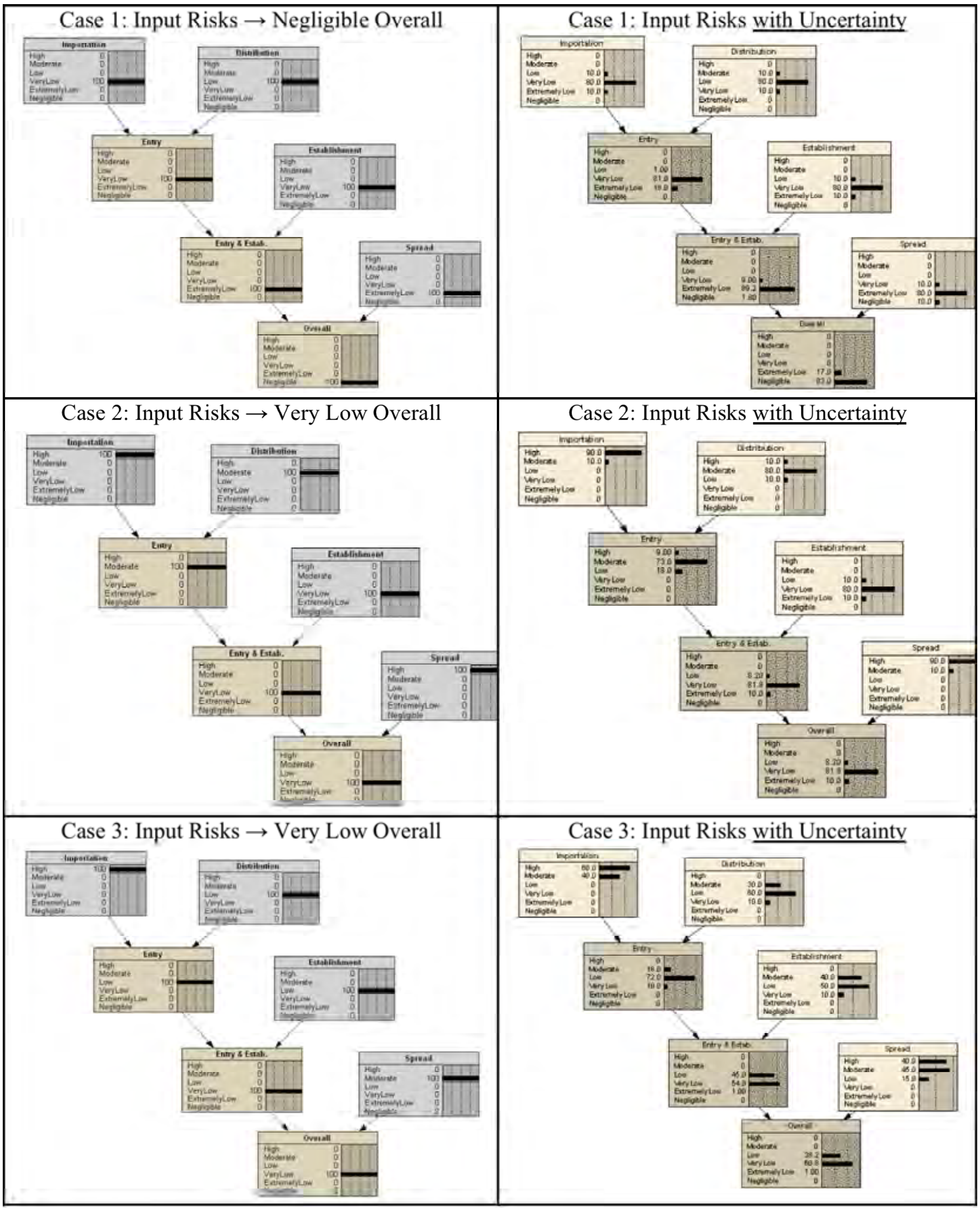


Figure 7.3. Three cases comparing the single overall likelihood based on certain inputs (left) with inputs including some uncertainty (right).

In each case the uncertainty in the inputs results in a distribution over the Overall likelihood, though the highest probability falls in the same category. In this example, we continue to think about input likelihoods as ‘states’. Uncertainties in the inputs represent uncertainty in beliefs that a particular input is in a given state.

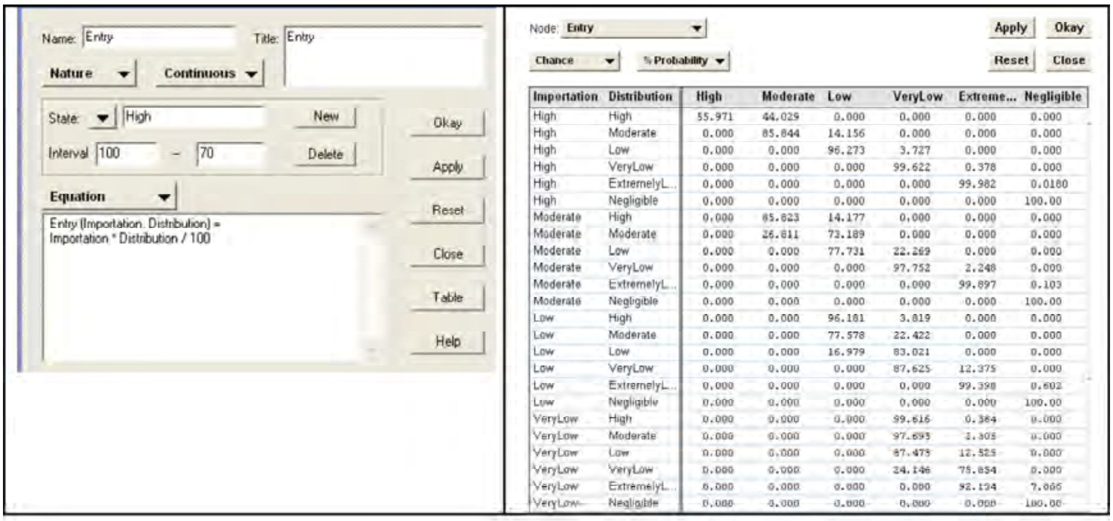
7.1.3 Adding uncertainty to the combinations of likelihoods

We can also represent the uncertainty in the way the likelihoods are combined. Each qualitative likelihood reflects an indicative probability range, as shown in Table 7.1. The deterministic combination of qualitative likelihoods is very coarse and is inaccurate, particularly when the probability range is large. For example, the deterministic table says that High x Low → Low. However multiplying the lower range values of 0.7 and 0.05 gives 0.035, which is in the Very Low range. This means that the deterministic BN, which gives P(Entry=Low|Importation=High,Distribution=Low)=1, is only approximating the distribution that would result if the actual ranges were used (see Section 5).

Table 7.1 From Biosecurity Australia (2008)

Likelihood	**Descriptive Definition**	**Indicative Prob. Range**
High	The event would be very likely to occur	$0.7 < P \leq 1.0$
Moderate	The event would occur with an even probability	$0.3 < P \leq 0.7$
Low	The event would be unlikely to occur	$0.05 < P \leq 0.3$
Very Low	The event would be very unlikely to occur	$0.001 < P \leq 0.05$
Extremely Low	The event would be extremely unlikely to occur	$0.000001 < P \leq 0.001$
Negligible	The event would almost certainly not occur	$0 < P \leq 0.000001$

It is straightforward to represent the uncertainty arising from combining discretised probability range in the BN by making the Likelihood nodes represent continuous variables, each ranging from 0 to 100. In the BN software (Netica), we discretise the nodes, using exactly the labels and the indicative ranges from Table 7.1. The arithmetic combination of likelihoods is done using an equation: for example, P(Entry | Importation, Distribution) = Importation x Entry / 100 (see Figure 7.4a). This equation is used to generate a CPT for the node (using a stochastic sampling method) (see Figure 7.4b).



Importation	Distribution	High	Moderate	Low	VeryLow	Extreme...	Negligible
High	High	55.971	44.029	0.000	0.000	0.000	0.000
High	Moderate	0.000	85.844	14.156	0.000	0.000	0.000
High	Low	0.000	0.000	96.273	3.727	0.000	0.000
High	VeryLow	0.000	0.000	0.000	99.622	0.378	0.000
High	ExtremelyL...	0.000	0.000	0.000	0.000	99.982	0.0180
High	Negligible	0.000	0.000	0.000	0.000	0.000	100.00
Moderate	High	0.000	85.823	14.177	0.000	0.000	0.000
Moderate	Moderate	0.000	26.811	73.189	0.000	0.000	0.000
Moderate	Low	0.000	0.000	77.731	22.269	0.000	0.000
Moderate	VeryLow	0.000	0.000	0.000	97.752	2.248	0.000
Moderate	ExtremelyL...	0.000	0.000	0.000	0.000	99.897	0.103
Moderate	Negligible	0.000	0.000	0.000	0.000	0.000	100.00
Low	High	0.000	0.000	96.181	3.819	0.000	0.000
Low	Moderate	0.000	0.000	77.578	22.422	0.000	0.000
Low	Low	0.000	0.000	16.979	83.021	0.000	0.000
Low	VeryLow	0.000	0.000	0.000	87.625	12.375	0.000
Low	ExtremelyL...	0.000	0.000	0.000	0.000	99.398	0.602
Low	Negligible	0.000	0.000	0.000	0.000	0.000	100.00
VeryLow	High	0.000	0.000	0.000	99.616	0.384	0.000
VeryLow	Moderate	0.000	0.000	0.000	97.695	2.305	0.000
VeryLow	Low	0.000	0.000	0.000	87.475	12.525	0.000
VeryLow	VeryLow	0.000	0.000	0.000	24.146	75.854	0.000
VeryLow	ExtremelyL...	0.000	0.000	0.000	0.000	92.134	7.866
VeryLow	Negligible	0.000	0.000	0.000	0.000	0.000	100.00

Figure 7.4. Netica screen shots for the Entry node, showing (a) the specification of the continuous range for High, and the equation for combining the likelihoods of the parent nodes, Importation and Distribution; (b) the resultant CPT, from stochastic sampling.

Figure 7.5a shows the resultant BN; the corresponding qualitative BN is given in the right column, for direct comparison. Note that there is a slight change in the Netica visualization for discretised continuous nodes, rather than simply discrete categories – there is an additional section at the bottom of each node, showing the mean and the standard deviation of the current calculated posterior distribution.

We can see that with no informative inputs, the resultant distribution for the Overall likelihood for the continuous model is not very different, as expected given the qualitative combination matrix was calibrated with the indicative probability ranges. The Overall likelihood Case 3 (Mixed inputs) is also similar, with P(Overall=Low | Case 3 Inputs) = 0.96.

However Case 2 (High/Moderate inputs) provides an example of how the Overall assessment can change when explicitly modeling the indicative probability ranges: here the highest probability outcome is P(Overall=Low | Case 2 Inputs) = 0.551, whereas the current qualitative method gives a Moderate overall combined likelihood.

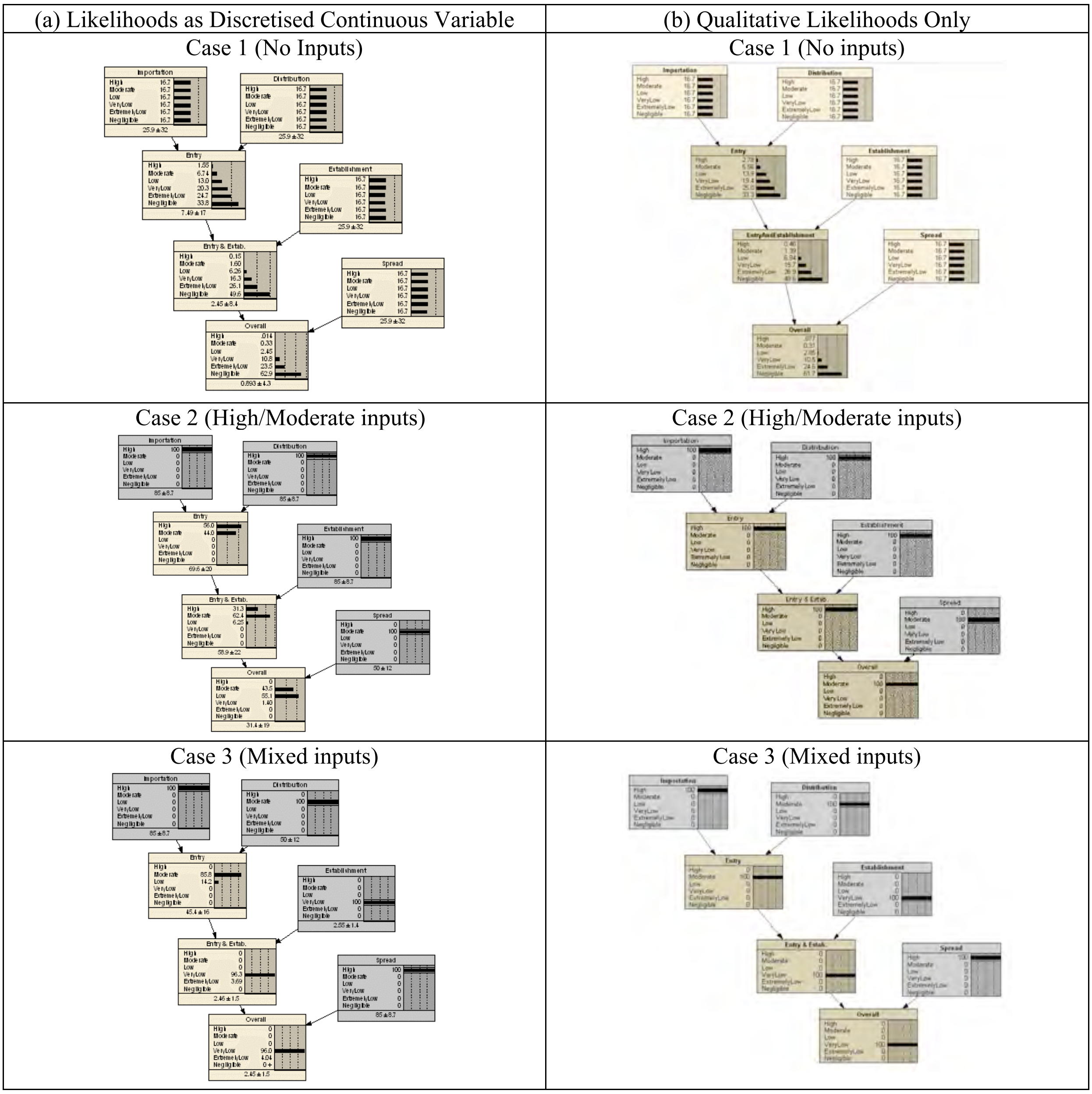


Figure 7.5. Comparison of the discretised continuous BN model (7.5a, left) with qualitative only (7.5b, right) for: no inputs, for a High/Moderate input scenario and Mixed inputs.

7.2 Improving the BN modelling

There are several limitations with the current process as modeled by the BNs in Figures 7.1 to 7.5. The following sections illustrate them.

7.2.1 Representing dependencies between likelihoods at different stages

The BN modeling above combines the risks for the different stages in a way that assumes they are independent. For example, in the BN, Importation and Distribution are independent parents of Entry:

P(Entry) = Prob(Importation) x Prob(Distribution).

However, this is *not* how the likelihood of distribution is assessed by the analyst; the qualitative assessment is the likelihood of Distribution *if* the pest has been imported. So the qualitative method is actually providing a <u>conditional</u> likelihood, which we haven't captured in the BN models above. The definition of Entry is then

P(Entry) = P(Distribution | Importation) x P(Importation).

7.2.2 Probabilities of Probabilities? Or degree of infestation? A "Pest Volume" BN

The second problem is that the BN modeling above is doing a kind of meta-reasoning, producing probability distributions over likelihoods, which represent indicative probability ranges. Given pest likelihood at one stage, the BN produces a distribution over the likelihood of infestation at the next stage. This improperly conflates aspects of volume (amount of fruit in a container; how prevalent the pest is on the fruit) with probability.

We suggest the following modelling changes to overcome these problems, shown in the example BN in Figure 7.6:

- Distribution is modelled explicitly as being conditional on Importation, given Entry.
- The node states explicitly represent the volume (number) of the pest for the Importation, Entry and Establishment nodes, while Spread models area.
- The risk level is the posterior distribution computed by the BN reasoning algorithm for each volume.
- The consequences are modelled as simple ternary nodes with values (though we equally could have stayed with the current qualitative levels A-G), with the cost of each consequence modelled by an associated utility node. The utility functions were obtained through methods such as described in Section 11 below.

Note that this version of the "Pest Volume" BN uses discrete levels (unspecified) rather than continuous volume units, as the additional volume-based elicitation was beyond the scope of the project; continuous volumes would be used in practice.

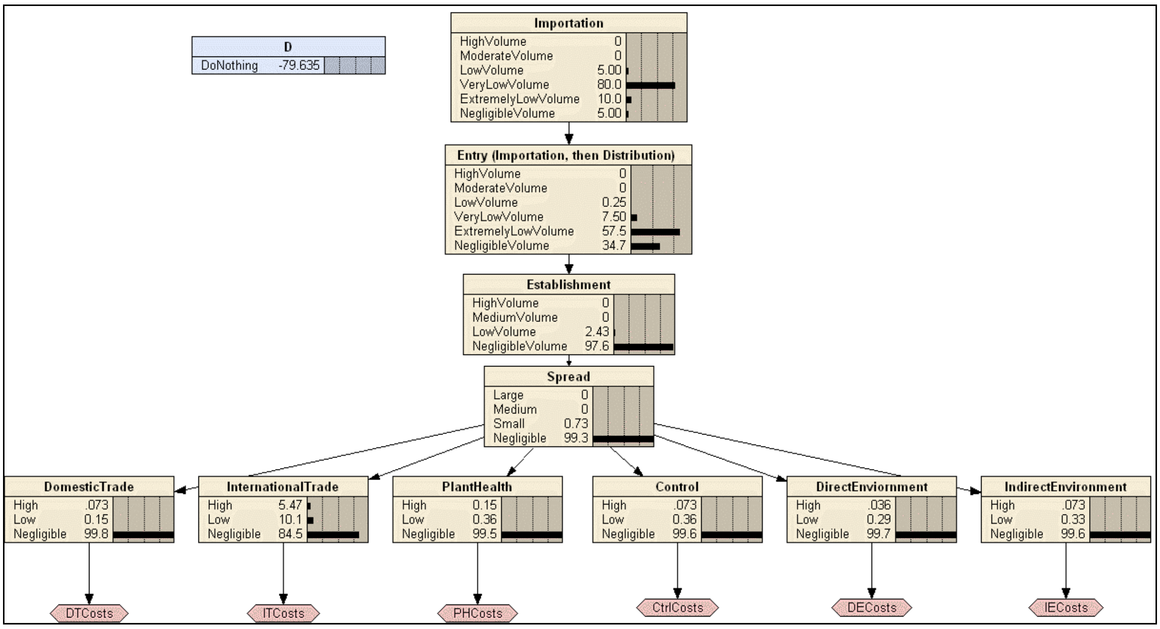


Figure 7.6. Alternative BN structure includes: correct dependencies; explicit representation of pest volume and spread area; representation of consequences, costs, and the “no nothing” action of unrestricted risk. The scenario starts with a distribution over a predominantly “Very Low Volume” Importation assessment.

7.2.3 Modelling risk at each stage

Thus far, we have modeled only the basic “stages” used previously in the BA qualitative risk assessment process. However, an advantage of BNs is that they can explicitly model the other information and factors being incorporated into the risk assessment. We do this by incorporating additional nodes which feed into the ``Stage'' nodes. For example (Figure 7.7), the Importation Assessment may depend on a large range of factors, including prevalence in the place of origin, inspection procedures at the export location, the ability of the pest to survive transportation, and so on. The layout of the process reflects the sequence of stages, with possibility of detection (say through inspection at source) and modeling whether or not the pest survives transportation.

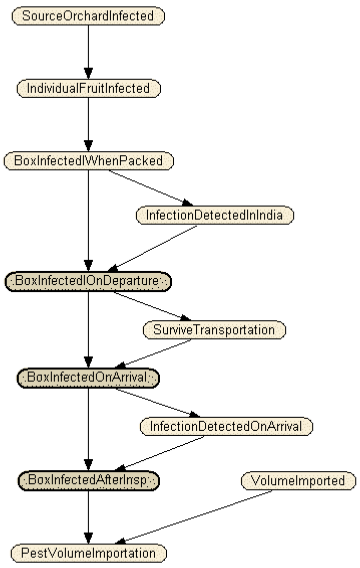


Figure 7.7. A possible subnet for an Import Risk Assessment

It is also possible to add additional nodes to the BN to represent the causal factors influencing the unmitigated risk, as shown in Figure 7.8. In this example, the LocalEnvironment node allows us to model the difference between the establishment of the pest, given that it has entered, depending on whether the local environment is benign or hostile. The second causal factor RegionalFruitTransport impacts on the spread of the pest.

7.2.4 Modelling risk mitigation actions

In the current qualitative risk assessment, if the unmitigated risk is high enough to warrant further modelling, the next stage is to look at risk mitigation actions and their associated costs. This is another straightforward extension of the unmitigated BN. In Figure 7.8 the possible actions modelled are the level of inspection on entry {High, Low, None) and whether or not to apply pesticides on entry or on local farms. Each action has an associated cost, modeled in the table for the utility nodes (InspCost, PCost1 and PCost2 respectively). This results in a so-called “Decision Network”.

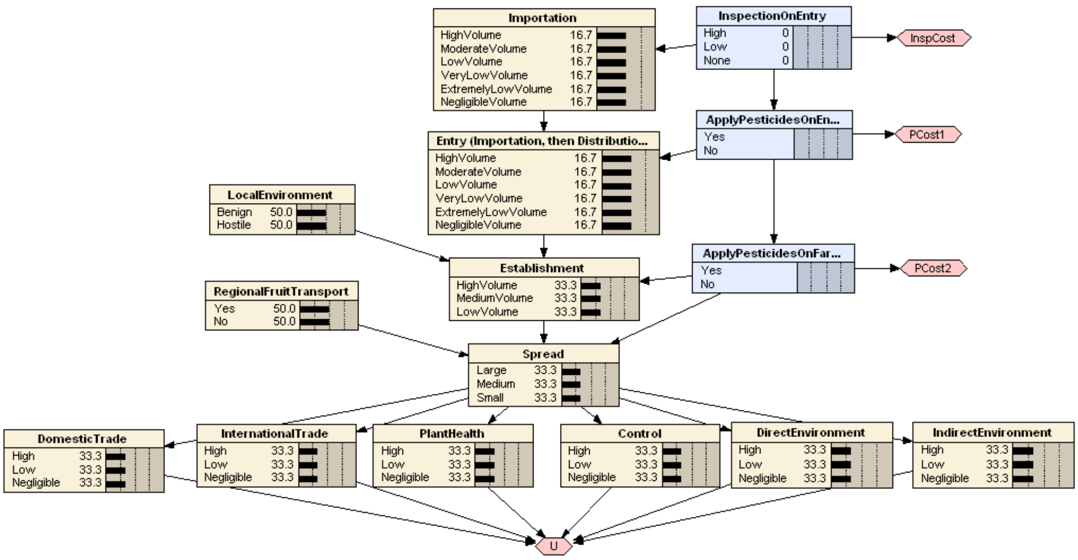


Figure 7.8. Modelling of causal factors for both unmitigated and mitigated risk.

7.3 Knowledge engineering: strengths, weaknesses, opportunities, costs

The development of useful, robust methodologies for constructing and using BNs – which we call the Knowledge Engineering Process – is still an active area of research. However, there are certain principles that *have* been established and should be followed if BNs were to become part of Biosecurity Australia’s risk assessment process.

The basic modelling steps are to select the nodes, determine the structure, elicit the probabilities and evaluate the outcome. The knowledge engineering process should be iterative. This was done informally for this case study. Thus, during the development of the models presented above there were a number of instances of revisiting past modelling choices following improved understanding of the problem domain. For example, during the elicitation of parameters from our expert, the values for the consequence nodes (DomesticTrade etc) were changed from [Yes, No] to [High, Low, Negligible].

In this project, as questions for the elicitation workshop (Sections 2 and 3) were done with non-BN purposes in mind, they did not provide all the conditional probability parameters needed for our example BN; the additional parameters were obtained by elicitation from a single expert.[2] While elicited *distributions* for parameters are not used by BNs (we used the ‘best-guess’ point probabilities in the CPTs), distributions can (and should, when available) be used to support analysis of sensitivity to parameters. For example, if a normal distribution were supplied, the mean would be used in the CPTs. However, values at, say, two standard deviations above and below the mean could be checked to see whether the resulting distributions under observations remain sensible, and whether any indicated decisions or utility values have made dramatic shifts.

We have illustrated how subnetworks could be used in the biosecurity context; a full evaluation of their use was beyond the scope of this project. Templates of subnetworks with the range of factors for each stage could be developed. Then, when a BN is to be developed for a new risk assessment, relevant factors would be selected and parameterised for the specific situation. We note that the Netica BN software used for our prototype BNs does not provide direct support for the use of subnetworks. As a result, the risk assessment BNs may quickly become large, unwieldy and difficult for the risk analyst to both build and validate. However other BN software (e.g. GeNIe[3]) does provides the type of support for templates and subnetworks needed for the flexible development of case specific BNs.

The routine deployment of Knowledge Engineering for import risk assessments would require investment in skill development. Naïve users of these tools often make several predictable mistakes that may compromise the legitimacy of the analyses. Appropriate skills may be developed in people with no quantitative background in a few days of dedicated training, given follow-up support and review.

While development of skills may be an initial hurdle, the approach holds the potential to capture current, experienced thinking in a more formal framework that can accommodate expert judgements and whatever data may become available. Over time, subnets that are developed for specific kinds of pests, diseases and commodities will form a library of templates that will inform and improve the efficiency of future efforts.

[2] We compared both direct elicitation and an elicitation tool based on verbal cues; there were no significant differences in the time for elicitation or the parameters elicited. See Nicholson & Korb report for details.
[3] GeNIe is made available in a compiled form, free of charge from the Decision Systems Laboratory, University of Pittsburgh (http://genie.sis.pitt.edu).

8. Probability Bounds Analysis

The term Probability Bounds Analysis (PBA) was coined by Ferson *et al*. (2004) to describe methods that allow risk assessment calculations to be performed with probability boxes, or p-boxes. These methods include the special case of dependency bounds analysis, developed by Williamson (1989) to calculate the bounds on the distribution of a function of random variables (such as a risk function) when only the marginal distributions of the variables are known. Probability bounds analysis and dependency bounds analysis belong to a class of methods that use imprecise probabilities to simultaneously represent epistemic uncertainty and variability. A probability box (p-box) is the class of distribution functions *F(x)* specified by a pair of upper and lower distribution functions $\overline{F}(x)$ and $\underline{F}(x)$ such that $\underline{F}(x) \le F(x) \le \overline{F}(x)$ for all x (Ferson *et al*. 2004). The intervals presented in section 5 are examples of p-boxes.

The most important practical characteristic of p-boxes is that they can be tailored to the available data in a manner that allows rigorous risk calculations without having to resort to assumptions about, for example, the moments or shape of a variable’s distribution function. This overcomes many of the disadvantages of Monte Carlo simulation. If assumptions can be justified, they can be incorporated into a p-box, typically narrowing or “pinching” the p-box, and in the extreme, reducing it to a precise distribution function whose moments and shape are assumed to be precisely correct. Ferson (2002) and Ferson *et al.*, (2003) provide comprehensive descriptions of the many different ways to construct p-boxes based on the various forms of (usually limited) data that are available to risk analysts.

8.1 P-bounds for IRA

In this case study two p-boxes were constructed: one that envelopes, and thereby contains, all the individual distribution functions fitted to each of the expert's beliefs; and another which represents (with a very small outward bounding error) the pooled distribution function (Figure 8.1). PBA allows us to develop two separate lines of enquiry, addressing:

- what is the difference in risk estimates between an approach that pools expert opinions compared to one that envelopes their opinions; and,
- what is the effect of assuming independence between each of the steps (probabilities) in the risk assessment model.

The results of the analysis (Figure 8.2) summarise the effects of both lines of enquiry, showing the risk assessment results for the pooled response and enveloped response, with and without the assumption of independence between the steps (probabilities) in the model. The analysis includes the MCS results of section 6 as a special case of the PBA (solid black line).

The qualitative assessment risk estimate for the annual probability of entry, establishment and spread was low, which is defined as a number in the interval [0.05, 0.3]. Comparing this result with the result of the MCS and PBA suggests that this estimate is conservative for most of the probability mass, as compared to the pooled MCS result and the pooled and enveloped PBA result if we assume the steps in the risk assessment model are independent. The result of ‘low’ slightly underestimates the probability of tail risks if we envelope, rather than pool the beliefs of the individual experts. If we make no assumptions about dependency between the steps in the risk model, then the upper bounds of the pooled and enveloped PBA are much higher than the qualitative assessment and the pooled MCS assessment.

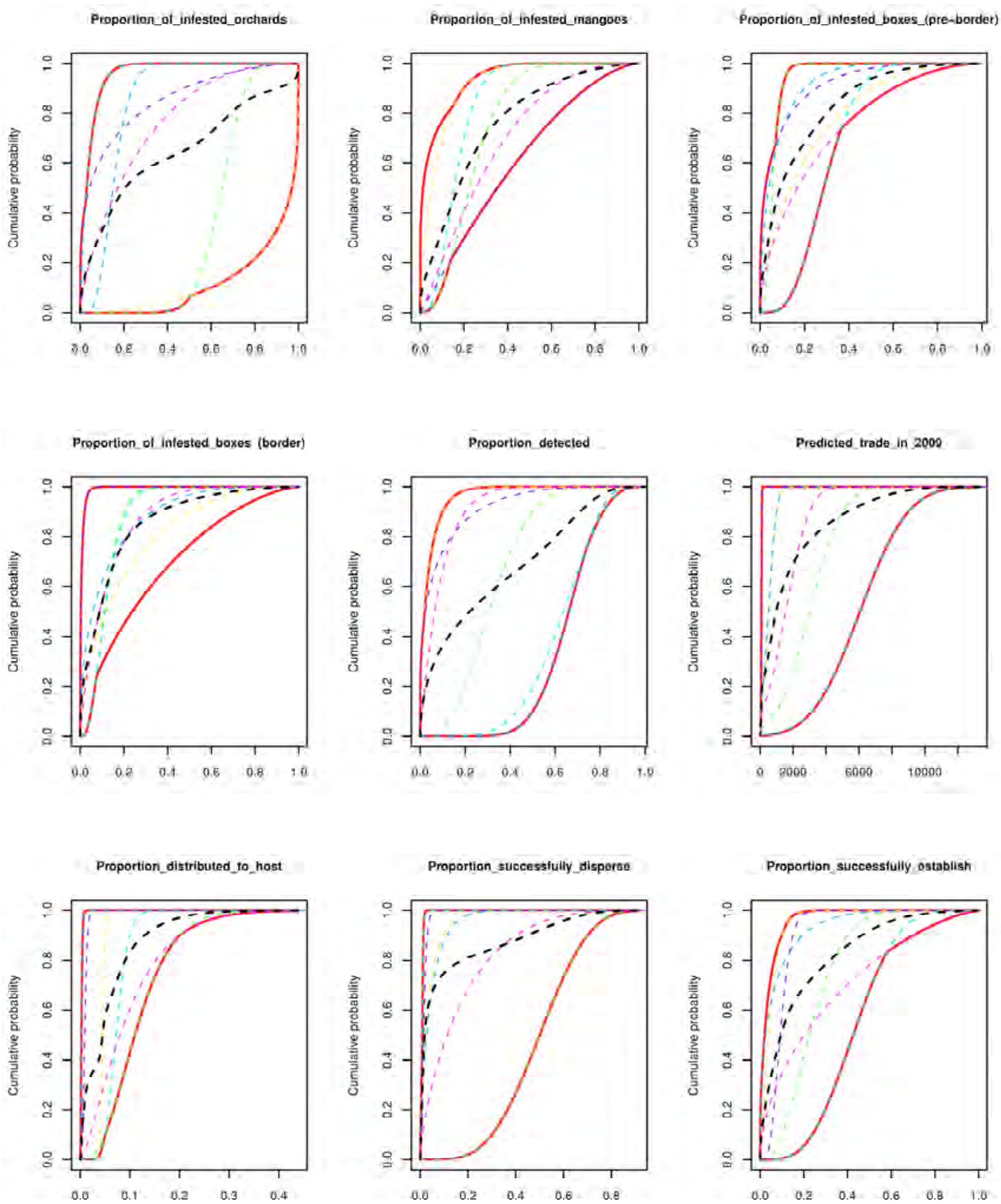


Figure 8.1. Probability boxes that envelope (red lines) each of the expert’s distribution functions (coloured lines), together with the pooled distribution function (black dotted line) for each of the steps in the risk assessment model.

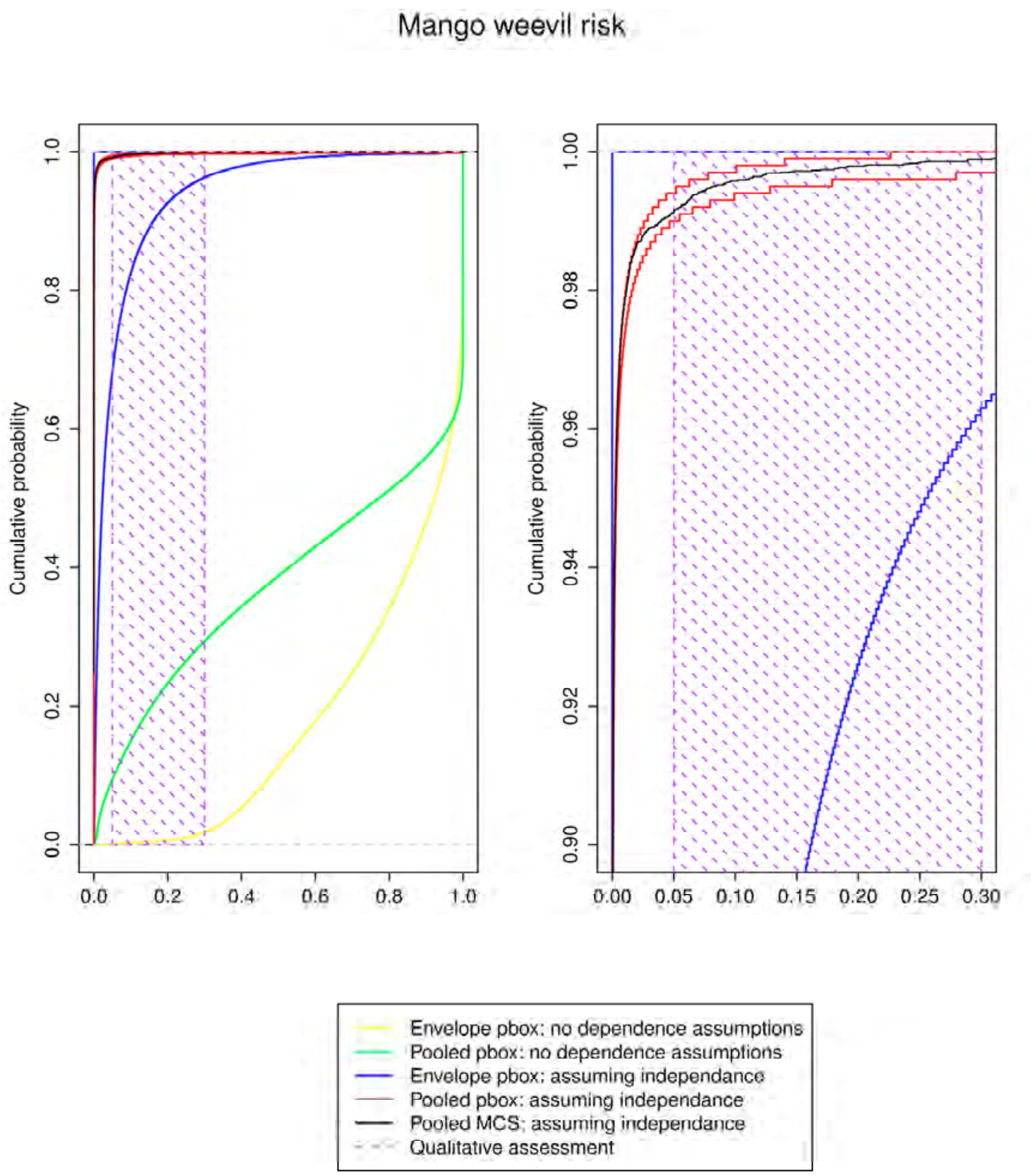


Figure 8.2. Results of the risk risk assessment showing the risk estimate of the qualitative risk assessment (purple dashed box), the MCS (black line) and the PBA (yellow, green, blue and red lines). The second plot (at right) shows detail from the upper left part of the first plot.

In other words if we assume that we know nothing about the dependency in the risk assessment model, then the 95th percentile of the annual probability of entry, establishment and spread could be as low as 0 and as high as 1 - i.e. the risk lies in the interval [0,1], from 'Negligible' to 'High', irrespective of whether we pool or envelope the distribution function's fitted to each of the expert's beliefs. If, however, we assume that the probability of each step in the risk assessment model is independent of the other steps, then the 95th percentile of the annual probability of introduction, establishment and spread, using MCS is 0.0024. All the results are summarised in Table 8.1 below.

Table 8.1. Results, expressed as the 95th percentile of the annual probability of introduction, establishment and spread, for the risk assessment

Method	Pool or envelope	Dependence assumption	95th percentile
Qualitative	NA	NA	[0.05, 0.3]
MCS	Pooled	Independent	0.0024
PBA	Pooled	Independent	[0.0024, 0.003]
PBA	Pooled	No assumption made	[0, 1]
PBA	Envelope	Independent	[0, 0.25]
PBA	Envelope	No assumption made	[0, 1]

These results illustrate that the effects of dependency are more important than the effects of pooling or enveloping the experts’ beliefs. It is important to note that these results could be improved (tightened) with additional information regarding the direction of dependency (positive, negative) between each of the steps in the risk assessment model.

8.2 Strengths, weaknesses, opportunities, threats

PBA was created as a kind of antidote to the unjustified assumptions that appear in most Monte Carlo simulations. It captures what is known about a problem, without making any unjustified or hidden assumptions. As such, it is perhaps the only tool available that deals explicitly with the full range of uncertainty in an analysis.

PBA is a new method and has not yet earned broad acceptance in the risk analysis community. There have been relatively few applications outside of a handful in environmental toxicology and engineering safety systems. It requires excellent skills in numerical analysis to understand and implement. Analysts need to consider carefully their choices of constraints about model parameter locations, distributional shapes, model structures and dependencies. Appropriate skills are rare in Australia and would be difficult to retain in-house.

If only range information is available, p-bounds provide the same answers as interval analysis. When information is sufficient to indicate precisely the distributions and dependencies for a problem, it gives the same answers as Monte Carlo. Thus, it generalizes both interval analysis and Monte Carlo (Ferson and Moore 2003). Possibilities within the bounds are not equally likely. Thus, bounding analyses do not replace the insight one might gain from a sensitivity analysis using a Monte Carlo model.

Where pathways are complex, it may be difficult or impossible to compute the best possible p-box, because some variables will be repeated. Repeated variables cause well-known problems with interval arithmetic, requiring mathematical or computational skills for their resolution (Ferson and Hajagos 2004).

9. Estimating extent of spread

Estimating the consequences of potential pests involves estimating the extent of their eventual spread. Within the extent of spread, the severity of consequences for several environmental, social and economic factors are evaluated at local, regional and national scales. Consequence estimation is explored in Section 11 but here we outline and discuss statistical methods that provide a reasonable approximation of the potential habitat of invading species.

There are several methods for estimating the extent of spread of an invasive pest. Here, we focus on two, Climate and Maxent, both members of the general class of models usually referred to as species distribution models, ecological niche models, or climate envelopes. These methods assume that all occurrence records have been collected at a time and place where the species is in equilibrium with its environment. The models then use the observed correlations between species records and environmental or geographic predictor variables to predict occurrence in new environments.

Climate has been developed by the Western Australian Department of Agriculture and Fisheries and the Australian Government, Bureau of Rural Sciences (BRS) (Pheloung 1996) and implements climate matching, to support inferences about potential habitat for a range of applications. Maxent has been developed more recently to create maps of species' distributions, specifically focussing on the problem of how to do this using presence-only data. It is increasingly a method of choice world-wide, and is used for modelling current distributions of species, and future distributions of invasives and other species undergoing range shifts.

9.1 Qualitative spatial prediction

BA currently base their estimates of potential distribution on the known distribution of a host species or a qualitative superimposition on Australia of the approximate climate of the pest’s native range. We will use the latter approach to compare qualitative approaches with the models described below.

9.2 Climate

‘Climate’ is the species modelling program most commonly used by BRS to model invasive species distributions. A web version called 'Climatch' is available at www.brs.gov.au/climatch (Crombie *et al*., 2008). The website explains: "*Climatch provides an interface for comparing climate characteristics between regions. It is typically used for predicting the potential spread of introduced or invasive species in applications such as risk assessments for live animal imports*". Climate comes with its own environmental data, shown in Table 8.1 (Barry 2006). Climate records are derived from meteorological stations closest to the nominal species location (Barry, 2006). The algorithm can also be run using environmental data and occurrence records from other sources, providing greater scope for comprehensive data sources.

Climate gives the option of three algorithms: Bioclim Euclidean or the Closest standard score (Barry 2006). Following Duncan *et al.* (2001) and Forsyth *et al*. (2004) this study used the Euclidean algorithm. The model was fitted in R, using species records and environmental data specifically collected for the study. This enables a fair comparison with the method described next.

Table 9.1. Predictors used in the study.

Predictor variable	Name Abbreviation[1]	Available in Climate?
Mean annual temperature	Bio 1	Y
Mean monthly temperature range	Bio 2	N
Isothermality	Bio 3	N
Temperature seasonality	Bio 4	N
Maximum temperature of the warmest month	Bio 5	Y
Minimum temperature of the coldest month	Bio 6	Y
Annual temperature range	Bio 7	Y
Mean temperature of the wettest quarter	Bio 8	Y
Mean temperature of the driest quarter	Bio 7	Y
Mean temperature of the warmest quarter	Bio 10	Y
Mean temperature of the coldest quarter	Bio 11	Y
Mean annual precipitation	Bio 12	Y
Mean precipitation of the wettest month	Bio 13	Y
Mean precipitation of the driest month	Bio 14	Y
Mean precipitation seasonality	Bio 15	Similar: Coeff of Var'n
Mean precipitation of the wettest quarter	Bio 16	Y
Mean precipitation of the driest quarter	Bio 17	Y
Mean precipitation of the warmest quarter	Bio 18	Y
Mean precipitation of the coldest quarter	Bio 19	Y
Mean NDVI of the greenest quarter	N	N
The log of human population density	P	N
Mean annual potential evapotranspiration	PE	N

[1] Source of variables or data:

Climate variables: Worldclim: www.worldclim.org/current.htm
NDVI: http://islscp2.sesda.com
Population density: http://sedac.ciesin.columbia.edu/gpw
Potential evapotranspiration: http://geodata.grid.unep.ch

9.2 Maxent

Maxent is a machine-learning method for modelling species distributions using presence-only data (Phillips *et al*. 2004, Phillips *et al*. 2006). It is a good candidate method for modelling potential distributions because it has been shown to perform effectively in several model comparison studies (Elith *et al*. 2006; Pearson *et al*. 2007) and has recently been applied in invasive species research (e.g., Ficeotla *et al*. 2009, Hinojosa-Díaz *et al*. 2009). Biosecurity examples are emerging. For instance, it has been used to map distributions of sheep scab outbreaks in Britain; the results, superimposed on sheep abundance records, identified risk foci and allowed targeted control measures (Rose *et al.* 2009).

Maxent can model complex responses and interactions between variables, though it controls complexity relative to sample size (Elith and Leathwick, *2009*). The underlying model focuses on constraints and entropy. Constraints are established by constructing ‘features’ (transformed functions of the predictor variables) and then predicting a distribution in which the mean of each feature under the predicted distribution is close to its mean across the observed ("empirical") sample. The constraints represent our knowledge of the distribution of the species. Amongst all distributions that satisfy the constraints, the most unconstrained (most spread out or closest to uniform) is chosen; this is the distribution of maximum entropy (Phillips *et al*. 2006).

The notable differences between Climate and Maxent are:

(i) Variables are equally weighted by Climate, whereas Maxent weights variables differently according to their effect on the species' distribution (Phillips and Dudík, 2008);

(ii) Maxent can fit more complex responses with its 'features' (Phillips and Dudík 2008);

(iii) Maxent uses a comparison with background data to build the model and is in that sense discriminative (Phillips *et al*. 2006), whereas Climate only relies on presence data (Duncan *et al*. 2001; Forsyth *et al*. 2004). Discriminative models tend to perform better for predicting distributions (Elith *et al.* 2006)

(iv) Maxent can give estimates of scaled probabilities whereas Climate can only give ranks.

(v) The environmental data in Climate is provided as a fixed set of 16 predictor variables; Maxent accepts any spatially distributed data, but it must be provided.

9.3 Case study

The case study makes use of an invasive species, the European wasp, which has now spread to most suitable environments in Australia except Western Australia, where it is under active and effective control, providing reasonable data for testing predicted distributions. Museum occurrence records for the wasp, *Vespula germanica*, were obtained from the 'Global Biodiversity Information Facility' (GBIF, 2008) from the website http://www.gbif.org/ (Downloaded on 23/07/2008) with all coordinates (latitude and longitude) in decimal degrees (Figure 9.1, red triangles).

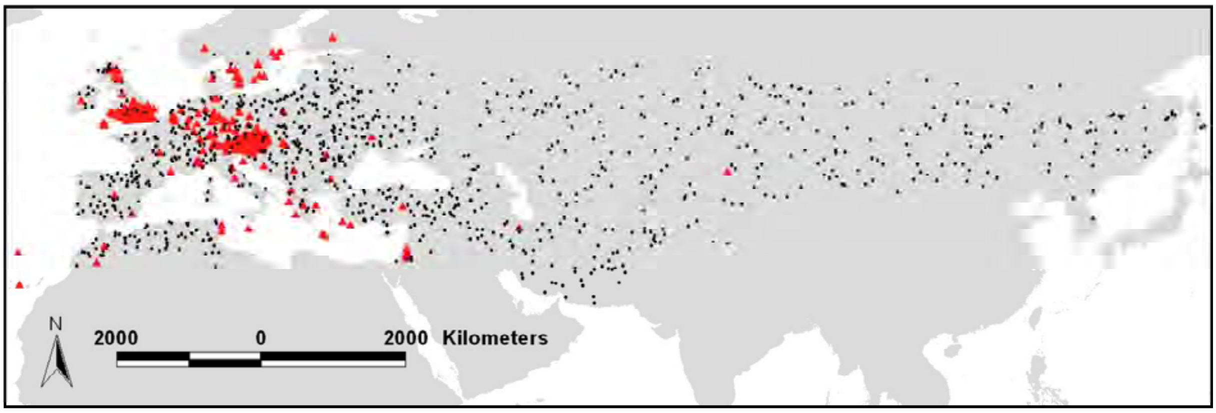


Figure 9.1. Occurrence records for *V. germanica*. Red triangles are museum records acquired from GBIF (1661 records). Black dots are randomly sampled points based on an 'expert' map, stratified to reflect abundance (1000 records). These points were not used in this study., but could be used to represent expert opinion about potential range.

These data nominally included textual information such as the species classification, locality, the identifier and collector, which museum the specimen belongs to and information on the precision of coordinates, etc. However, many of these fields were empty. Records missing coordinates but having text descriptions were assigned coordinates using 'Google

Earth' (http://earth.google.com/) and the most precise locality information available. Records only containing the country as a reference were deleted. The original dataset contained 2011 records and was reduced to 1661 after this 'cleaning'.

9.3.1 Background data

Maxent requires 'background' data as a reference, for the model to understand what environments are available in the region of interest (here, in the native range). They are often drawn uniformly at random from the whole study region, where this is sensible, avoiding points that by chance occur in the same grid cell as a presence record (Pearson, 2007). An exception might include data sets where only a subset of the region of occurrence has been sampled. Adjustments may need to be made for survey biases or dispersal constraints (Phillips et al. 2009). The background samples can be automatically generated within Maxent, and masks and bias files can be used to constrain sample locations and account for survey biases.

When used in a similar way to Climate, Maxent is being used to build a model for the native range and project the model (i.e., predict) to Australia. For this, the relevant background is those areas in and around the native range that could reasonably be expected to be sampled for the species — including those that are unhabitable by the species, but excluding those places to which the wasp would not have naturally dispersed (for instance, south of the Sahara Desert).

The background used in models here was limited to a latitude of about 25ºN (Figure 9.2), suggesting that the wasp could not naturally disperse into the tropics or across large oceans (Spradbery and Maywald 1992). The background sample consisted of a random sample of 10,000 points within the selected area.

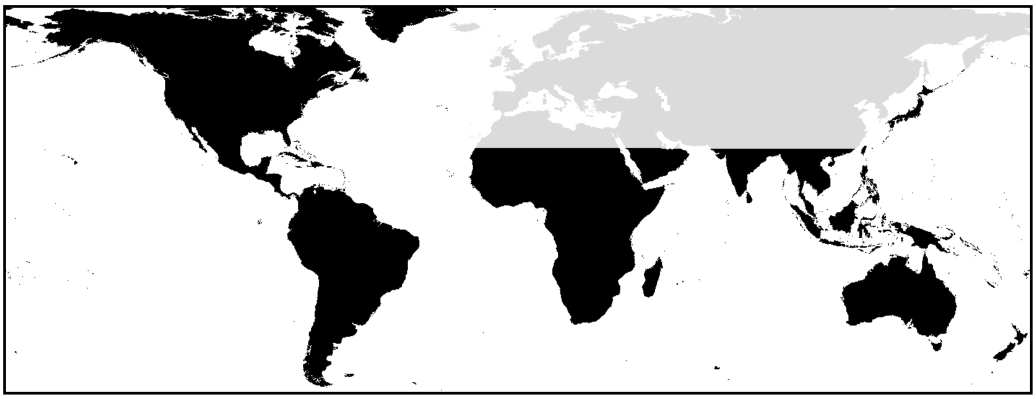


Figure 9.2. Spatial extent to which background points were selected (in grey). This map is the one used for the data presented in this report (3,611,404 grid cells).

9.3.2 Predictor variables

The aim in species distribution modelling is to select predictors that are ecologically relevant to the species. We focussed on climate, food and nest resources. Climatic data were downloaded from the WorldClim website (Hijmans *et al*. 2005) (Table 9.1). We also sourced data on human population density (CIESIN, 2005), normalised difference vegetation index (NDVI) (Los *et al*. 2000) and potential evapotranspiration (UNEP 2006). Population density grids were created from information on population densities in 2000, adjusted to match UN totals for persons per square kilometre. This was considered a potentially useful predictor variable as the wasp is a scavenger (Spradbery and Maywald 1992), utilising all the resources that come with the presence of humans. Densities were log-transformed to emphasise important differences in population densities.

The NDVI data sets provided monthly changes in the photosynthetic activity of terrestrial vegetation over seventeen years (Los *et al*. 2000). We calculated mean NDVI for the greenest quarter over five years to minimise the variation in recording NDVI (Maignan 2004). This was calculated by first averaging mean NDVI for each month, over the five years, then calculating running means of every three consecutive months. The three month block that possessed the maximum mean NDVI was chosen as the mean NDVI for the greenest quarter. This version of NDVI was considered to be the most ecologically relevant since ‘green-ness’ was considered a surrogate for vegetation density and this affects wasp abundance predominately during spring and summer when they are building and fixing their nests (the greenest quarter) (Spradbery 1973). This assumption may not hold in winter rainfall areas in Australia.

Mean annual potential evapotranspiration was calculated from Worldclim data, using the Thornthwaite formula (Thornthwaite 1948) and was used as a surrogate for the amount of energy present in an ecosystem (Schowalter 2006). Lower values of evapotranspiration imply less energy, which is likely to directly impact the suitability of the environment for the wasp.

Most predictor variables were available as grids with a spatial resolution of 2.5 arc minutes (~ 4.6 × 4.6 kilometres) for the entire world (for Maxent predictions), and the exceptions were resampled using bilinear interpolation, to form a consistent set. The scale chosen reflects a balance between the ecological relevance (Austin 2007) to *V. germanica*, the computation required for a continental prediction and data availability. The results presented later are averaged across various groupings of the available predictors (Table 9.1). The full study (Dobiecki 2009) tested the effect of different predictor choices; we only present the summary because it is representative of all results.

Many other predictors were potentially interesting (e.g. soil moisture, solar radiation, wind speed, etc.) but were unavailable. For example, Latimer *et al*. (2006) stressed that historical disturbances and ecological processes such as dispersal, reproduction and competition are rarely considered for use as predictor variables in species distributional models. It would be difficult to model these factors because most SDMs assume presence records are for a species in equilibrium with its environment. Data availability is always an issue for modelling, in the same way as expert knowledge is limiting for qualitative assessments. However, with increasing interest in, and application of, models for invasive species, data are constantly being updated and improved. At the moment several initiatives are exploring and developing relevant datasets for Australia and the world (e.g. Atlas of Living Australia (http://www.ala.org.au/); working group led by McGill et al. at the National Centre for Ecological Analysis and Synthesis, http://www.nceas.ucsb.edu/research/wg).

9.3.3 Evaluation methods

Reliable evaluation needs data ‘independent’ of the data used to build the model (Pearce and Ferrier 2000). This ensures that the evaluation results are not optimistic (Fielding and Bell 1997; Araújo *et al*. 2005; Hartley *et al*. 2006). Completely independent data are rare (Araújo *et al*. 2005), and modellers often resort to evaluating models with a subset of the original occurrence data through a form of data partitioning (Pearson 2007). This can be ineffective for invasive species because: (1) if the model is trained in the native range, performance in the native range is of some interest, but not the main aim of the modelling; (2) in the invaded range the final distribution is unknown.

The retrospective nature of this study allowed current independent data from Australia to be used instead. These data were obtained from occurrence records in the literature (Crosland 1991; Spradbery and Maywald 1992). This information was presence-only, therefore pseudo-absences were generated to allow test statistics to be calculated (following Phillips *et al*. 2006). Pseudo-absences were randomly sampled across the whole country (a sample size of 2000), excluding grid cells that contained a presence record, as was the case for data selected from the northern hemisphere. The full data set was treated as the true potential distribution, assuming the wasp was at equilibrium with the environment.

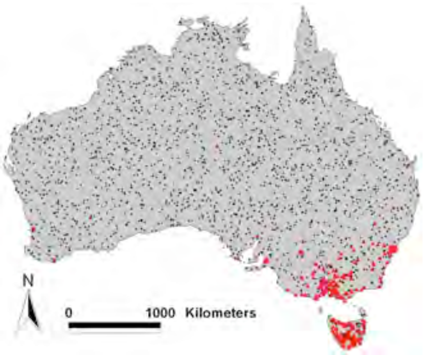


Figure 9.3. Presence-absence evaluation data set for *V. germanica* in Australia. Red points are occurrences. Black points are pseudo-absences.

9.3.4 Test statistics

Test statistics were used to measure the ability of a model to correctly distinguish between occupied and unoccupied sites — i.e. discrimination ability (Pearce and Ferrier 2000). Here we present results for Area Under the Receiver Operating Characteristic Curve (AUC). AUC is a rank-based discrimination measure, ranging from 0 to 1 (Pearson 2007), where 1 indicates perfect discrimination, 0.5 is no better than a random guess and less than 0.5 indicates performance worse than random (a model that fits the modelling data but predicts badly can cause a low test AUC) (Elith *et al*. 2006).

9.4 Results

The predictive performance of Maxent increased from about 0.9 on data set 1 to about 0.98 on data set 2. (Figure 9.4). In comparison, Climate’s predictive performance increased from about 0.8 to about 0.86. The predictive surface for Maxent is more tightly clustered around the southern coast than is the surface for Climate (Figure 9.5). We used 19 Worldclim variables in this comparison, rather than the 16 variables for Climate that it routinely assesses, to employ the best data available and to ensure the comparison between the two methods was not due to different input data.

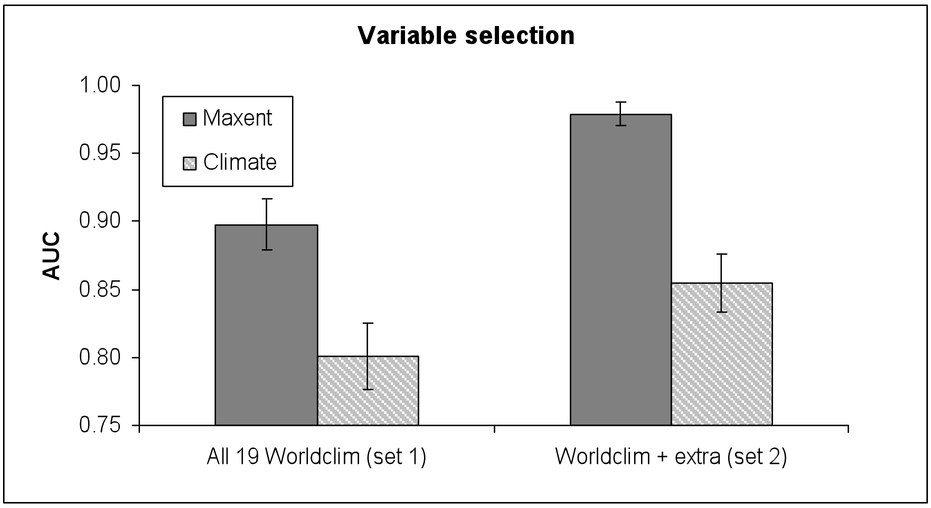


Figure 9.4. Bar graph showing the AUC_{aus} values of Maxent (grey) and Climate (stripes) models when built on predictor variables in set 1 and set 2 respectively. Error bars represent a 95% C.I. for AUC. Note the scale ranges from 0.75 to 1.0.

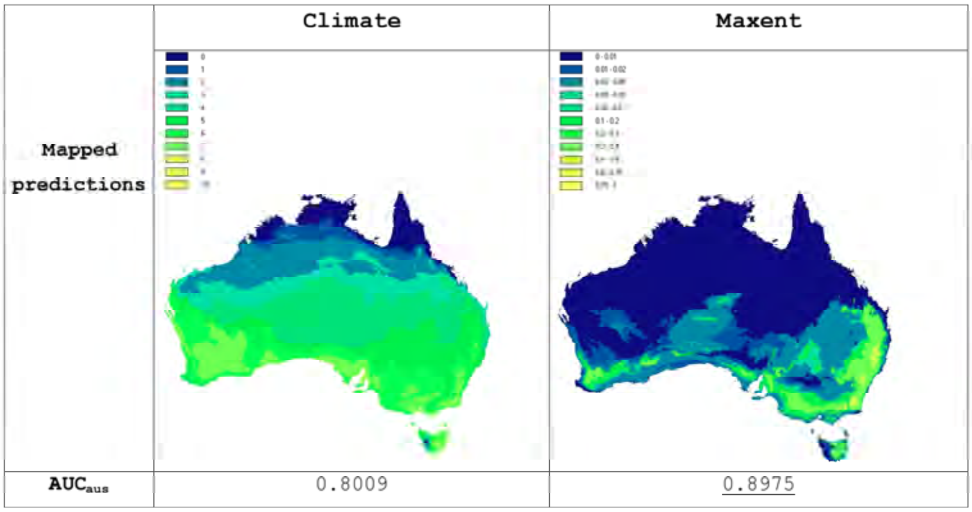


Figure 9.5. Predictions of the potential habitat of *V. germanica* in Australia as predicted by Climate and Maxent models, using nineteen Worldclim climatic variables.

Overall, the predictive performance of Maxent was superior to that of Climate (Figure 9.4, Figure 9.5). Thus, for the same investments in data acquisition, preparation and model building, Maxent better predicted the ‘future’ distribution of the invasive species in Australia. While this single result is not definite by itself, it is consistent with other comparative studies of the predictive capabilities of such tools using current species distributions (e.g., Elith *et al.* 2006; Pearson *et al.* 2007). The improvement in predictive performance, even if it holds generally, would have to be weighed against the additional software and new skills that would need to be acquired and maintained to run these analyses.

9.5 Qualitative spatial prediction

We estimated the wasp's potential distribution using information on climates in its native range. Using the Köppen-Geiger climate classification (http://koeppen-geiger.vu-wien.ac.at/index.htm, and see Figure 9.6), and expert advice from Dr Phillip Spradbery, world expert for the species (Figure 9.7), we estimated the potential distribution shown in Figure 9.8. The climate zones outlined in red (Figure 9.8) match those of the countries known to have records of the wasp. The prediction to central Australia is a poor match to the realised distribution of the species. Presumably BA (without knowledge of the final distribution of the wasp) would either include all these areas in the potential distribution, or find further locational or physiological information that might give reason to reduce the predicted extent.

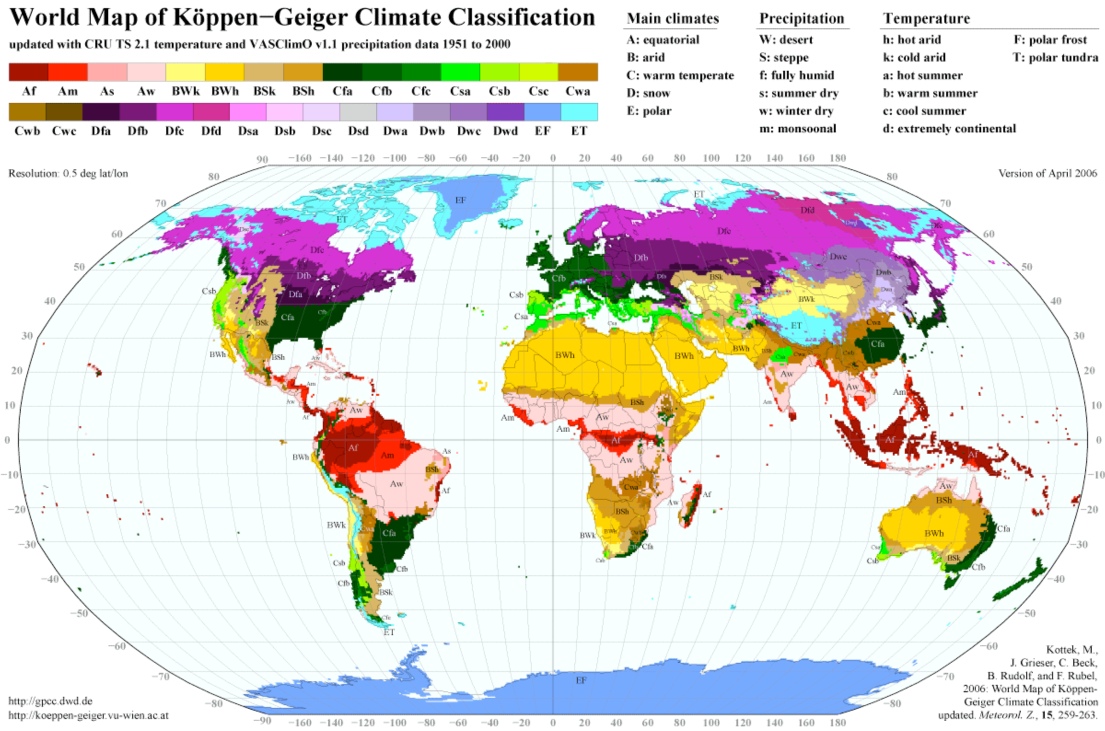


Figure 9.6 The world Köppen-Geiger climate classification, from http://koeppen-geiger.vu-wien.ac.at/present.htm#maps

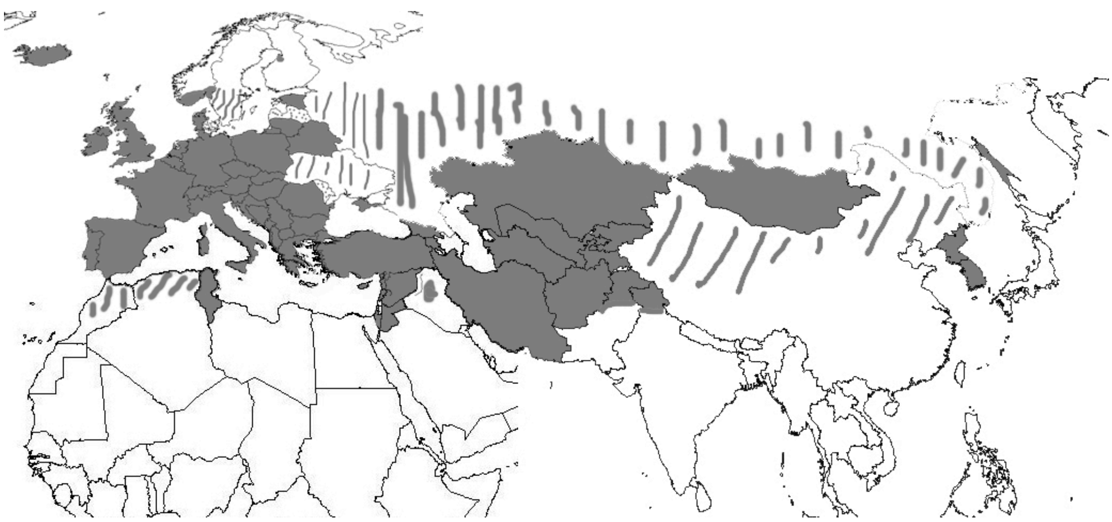


Figure 9.7 Endemic distribution of *V. germanica*. Provided by Dr. Phillip Spradbery. Dark grey represents countries with definite records; the hatched areas have some anecdotal records of occurrence.

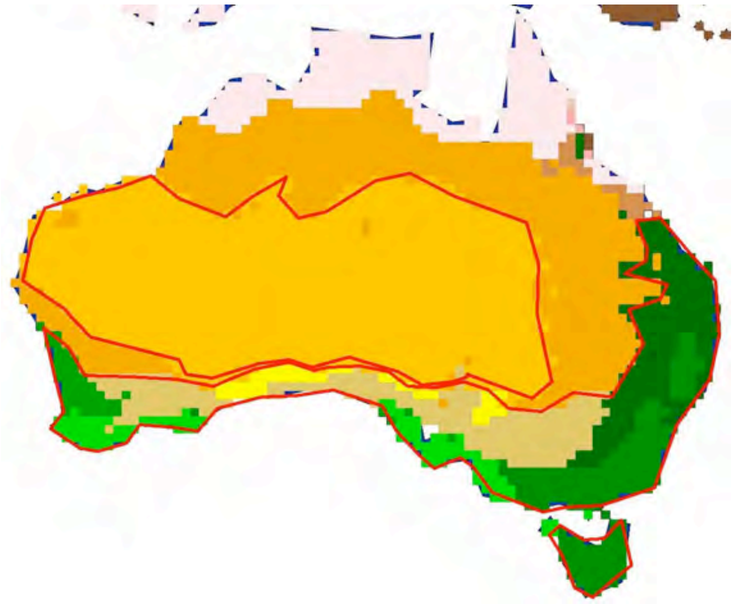


Figure 9.8. Koppen-Geiger climates in Australia. Areas outlined in red match those in countries with known occurrences of the wasp, identified in Figure 9.7. The colours represent different climate zones in Figure 9.6.

9.6 Strengths, weaknesses, opportunities, threats

This study primarily summarises a validation of the predictive performance of Climate and Maxent. There are no similar direct comparisons of these tools in the literature. However, a range of climate matching tools exist that operate in similar ways to Climate and that have been validated extensively, in comparison with statistical and machine learning methods. The result presented here is consistent with other comparative studies of predictive capabilities. Maxent and related machine learning tools consistently out-perform climate matching tools, when the objective is to predict current species distributions (e.g., Elith *et al*. 2006; Pearson *et al*. 2007). Simpler envelope methods such as Climate tend to predict less well, given the same data.

The study shows that both models can predict the potential distribution of this species reasonably well. In particular, the Maxent model with the extended data set had AUC values for this model greater than 0.95, indicating that the model can distinguish reliably between places where the species is likely to occur, and places it is not.

The Climate algorithm is currently used for many biosecurity decisions, in part because it is a simple algorithm and produces results that can easily be repeated. It has also been extensively tested using observed introductions of pest animals (see for example, Bomford 2006). This makes it a practical and relatively transparent tool for settling cost sharing arrangements for pests that have national impacts; alternatives including user-specified statistical models are more opaque for decision makers and may be more susceptible to 'gaming' by special interest groups. The maximum entropy algorithm implemented in the Maxent program shares the features of transparency and repeatability with Climate, but has produced more accurate predictions in this study and is likely to produce better results on theoretical grounds (Elith *et al*. 2006). This makes it a candidate for both risk analysis and cost sharing decisions.

Maxent is a free program with useful interface, and a tutorial that explains settings and results. It can be run in batch mode from a command line, enabling many analyses if required.

It does not include environmental or species data, and therefore all data must be obtained by the modeller. Maxent is being continually updated to include tools that enable the user to control and interrogate the modelled distributions; some of the current features include maps showing where extrapolation is occurring and methods for controlling the complexity of the model fit.

Maxent is especially valuable in a biosecurity context because it was developed for presence-only data. Much of the data available for predicting invasive species comes originally from museum and herbarium records, and other anecdotal information such as production audits.

In this case study, qualitative estimates might also have predicted a reasonable approximation of the distribution, depending on how the data beyond that presented above were incorporated. The importance of any differences depends on individual context, and will vary from case to case. For instance, in this case, the qualitative model did not predict the presence of the species in a small area around Alice Springs, and its absence elsewhere in Central Australia. Such details may be important if consequence assessments depended critically on expecting where the species could establish.

An advantage of explicit models is that they predict the likelihood of occurrence within the potential range of the invasive species, giving some idea of the most suitable areas. This information is the first step in designing effective surveillance systems (e.g. Section 12).

Quantitative models could be used alongside qualitative ones. Differences between the two provoke relevant questions, and allow iterative approaches that model data, probe and evaluate the output, then re-fit models in the light of new insight. Using an iterative approach is most likely to provide the best possible predictions, particularly given the often limited data and knowledge about species that pose threats.

Like many quantitative methods, routine deployment of Maxent would require skill development. Naïve use can lead to poor models and predictions. Appropriate training could be achieved for someone with knowledge in GIS systems and data manipulation, in one to two weeks.

10. Pre-establishment impacts in PRAs

This section focuses on one aspect of consequence assessment, namely, the identification and assessment of consequences that arise before a pest or disease has established and / or spread. This focus addresses a potential weakness of the current method used by BA for import risk assessment, namely the absence of explicit consideration of impacts that may arise prior to spread. The problem is illustrated using the case study. A remedy is suggested involving an event tree and cognitive mapping to structure thinking about expected consequences.

10.1 The BA model

The reported outcomes of the insect case study is shown in Table 10.1.

Table 10.1. Outcomes of the case study for likelihood and consequence.

LIKELIHOOD	
Event	**Outcome**
Importation	HIGH
Distribution	LOW
Entry	LOW
Establishment	MODERATE
Spread	MODERATE
EES	**LOW**

CONSEQUENCE	
Criterion	**Impact**
Plant life or health impacts	D
Direct environmental impacts	A
Domestic trade	B
International trade	E
Indirect environmental impacts	B
Eradication and control	D
Overall consequence rating	**MODERATE**

The current protocol appropriately characterises risk as likelihood × consequence. That is risk = *expected* consequence. Likelihood refers to the probability of entry AND establishment AND spread, Pr{EES}. Consequences are assessed against six criteria, assuming EES has occurred. The unrestricted risk under the current protocol is the combination of Pr{EES} and overall consequence, in the absence of phytosanitary measures. For the case study, LOW × MODERATE = LOW, exceeding Australia’s appropriate level of protection (ALOP) of VERY LOW (Table 10.2).

Table 10.2. Risk matrix used in Biosecurity Australia’s pest risk assessment. Likelihood is interpreted as the probability of entry and establishment and spread. Australia’s appropriate level of protection is ‘very low’. The import of commodities with potential pests considered higher risk is disallowed (Source: Commonwealth of Australia 2001).

Pr{EES}	***high***	negligible	very low	low	moderate	high	extreme
	moderate	negligible	very low	low	moderate	high	extreme
	low	negligible	negligible	very low	low	moderate	high
	very low	negligible	negligible	negligible	very low	**low**	moderate
	extremely low	negligible	negligible	negligible	negligible	very low	low
	negligible	negligible	negligible	negligible	negligible	negligible	very low
		negligible	***very low***	***low***	***moderate***	***high***	***extreme***
		Consequence					

This approach overlooks the possibility of impacts pre-entry, pre-establishment or pre-spread. It is not difficult to imagine scenarios where pests need only enter or establish (and not spread) for non-trivial impacts to be seen in control costs, domestic and international trade, or environmental values. Failure to consider the likelihood and consequences of states other than EES may result in systematic underestimation of risk.

10.2 An alternative approach

The current protocol encourages the characterisation of expected consequence shown in the cognitive map in Figure 10.1. Consequences arise only at the point of spread. The mitigating effects of eradication and control activities on other criteria are unspecified.

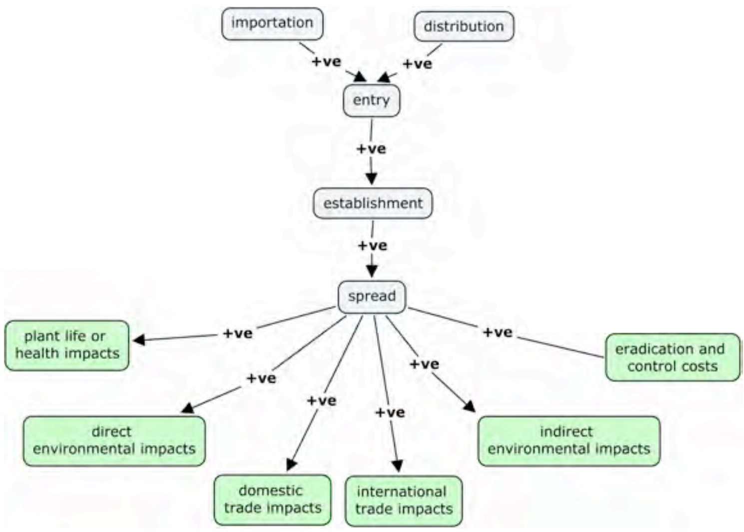


Figure 10.1. The cognitive map implied in the current protocol's characterisation of risk as the probability of entry, establishment and spread and the consequences associated with spread. The six consequence criteria are shaded green. Links between nodes describe the sign of association. Positive links indicate that an increase in the parent node leads to an increase in the child (or a decrease in the parent leads to a decrease in the child).

The ***event tree*** below (Figure 11.2) urges consideration of the possibility of consequences arising prior to entry (E'), at the point of entry only (EE'; i.e. entry but no subsequent establishment), or entry and establishment but no spread (EES'). Estimates of expected consequence need to consider each pathway in Figure 10.2.

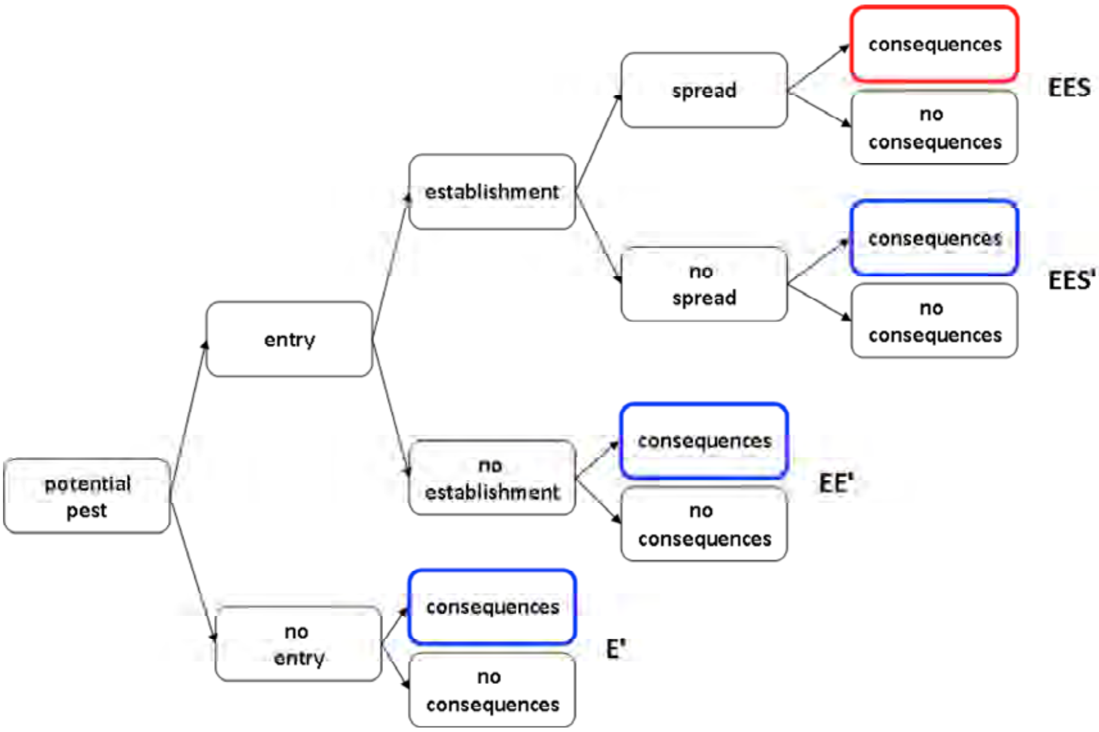


Figure 10.2. Event tree for potential consequences along the biosecurity continuum. There are four states that collectively represent the complete set of circumstances from which adverse consequences can potentially arise: E' - no entry, EE' - entry, but no establishment, EES' - entry, establishment, but no spread, and EES - entry, establishment and spread. Risk assessment protocols commonly consider consequences at the point of entry, establishment and spread (red box). A more complete characterisation of expected consequence would consider the possibility of impacts prior to entry, at the point of entry, or establishment (blue boxes).

The probability of observing the four states shown in Figure 10.2 is

$\Pr\{E'\} = 1 - \Pr\{\text{entry}\}$,

$\Pr\{EE'\} = \Pr\{\text{entry}\} \times (1 - \Pr\{\text{establishment}\})$,

$\Pr\{EES'\} = \Pr\{\text{entry}\} \times \Pr\{\text{establishment}\} \times (1 - \Pr\{\text{spread}\})$, and

$\Pr\{EES\} = \Pr\{\text{entry}\} \times \Pr\{\text{establishment}\} \times \Pr\{\text{spread}\}$.

Note that the four states are mutually exclusive, meaning that the expected consequences associated with any one state can be simply added to those of other states to characterise overall expected consequence.

Eradication and control in the case study involves (a) chemical control of infested orchards and (b) chemical quarantine treatment at the border for fruit exported from Western Australia. A more appropriate cognitive map for the case study is shown at Figure 10.3. It captures more detail, including a (hypothetical) understanding that:

- The impacts of pest *establishment* (without spread) on international and domestic trade are substantial. For all other criteria the impacts of establishment (without spread) are assumed negligible (and therefore not shown).
- The control of the pest in infested orchards is triggered at establishment and will mitigate impacts through reduction in the probability of spread, but will increase indirect environmental impacts through toxicological effects of pesticides.
- Chemical quarantine treatment of fruit exported from Western Australia does not reduce the probability of entry, establishment or spread. Rather, it mitigates impacts

on domestic trade. No mitigating effect is included for international trade because the quarantine treatment does not satisfy the ALOP of trading partners. We assume quarantine treatment will also be triggered at the time of establishment.

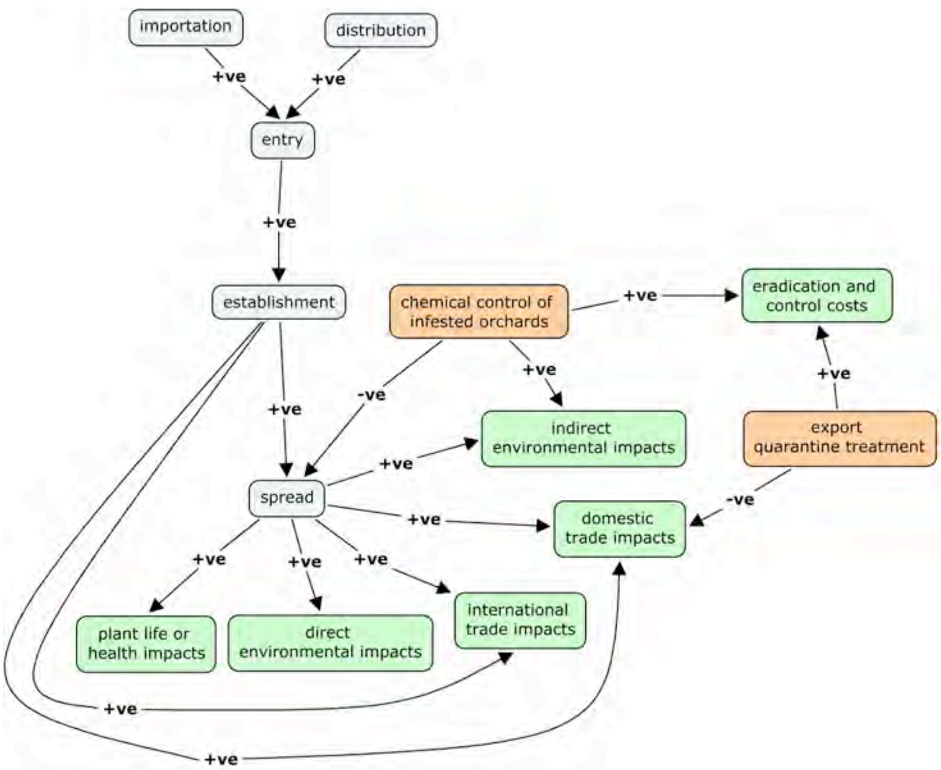


Figure 10.3. An improved cognitive map for risks associated with import of fruit. Consequence criteria are shaded green. Eradication and control actions are orange. Unlike Figure 10.1, the map includes trade impacts at the point of establishment. It also includes negative links describing the mitigating effect of eradication and control actions on other impacts.

Using Figure 10.3 to check the pathways in the event tree (Figure 11.2), it can be seen that the two states EES' and EES are relevant. No consequences are anticipated for E' or EE'. Assessment of the likelihood and consequence of the two relevant states is shown in Tables 10.3 and 10.4. Assessments were made using standard matrices employed in the current PRA protocol.

Table 10.3. Judgments of likelihood and consequence for EES', assisted by the cognitive map shown in Figure 10.3. All judgments are relative to those shown in Table 10.1.

LIKELIHOOD	
Event	**Outcome**
Importation	HIGH
Distribution	LOW
Entry	LOW
Establishment	MODERATE
not Spread	HIGH
EES'	**LOW**

CONSEQUENCE	
Criterion	**Impact**
Plant life or health impacts	A
Direct environmental impacts	A
Domestic trade	A
International trade	E
Indirect environmental impacts	C
Eradication and control	D
Overall consequence rating	**MODERATE**

Table 10.4. Judgments of likelihood and consequence for EES, assisted by the cognitive map shown in Figure 10.3. All judgments are relative to those shown in Table 10.1.

LIKELIHOOD	
Event	**Outcome**
Importation	HIGH
Distribution	LOW
Entry	LOW
Establishment	MODERATE
Spread	LOW
EES	**LOW**

CONSEQUENCE	
Criterion	**Impact**
Plant life or health impacts	D
Direct environmental impacts	A
Domestic trade	A
International trade	E
Indirect environmental impacts	C
Eradication and control	D
Overall consequence rating	**MODERATE**

The result is that the risk for EES' is LOW × MODERATE = LOW (Table 10.3), the same as EES (Table 10.4), and the same as the result reported in the assessment (Tables 10.1 and 10.2). But here overall risk is the sum of the expected consequences of the two relevant states: LOW + LOW. The PRA protocol currently has no convention for combining verbal descriptors of risk additively. Nevertheless it is clear that the more complete characterisation of expected consequence described here has led to an assessment reporting double the risk of that would have been made using existing procedures.

11.3 Strengths, weaknesses, opportunities, threats

The use of a specific model (the BA protocol) has restricted the focus of the analysis. Cognitive maps are one way of thinking about a range of possibilities. While they may offer a useful alternative, they are not the only way of representing the system. For example, one weakness of the cognitive maps drawn here is that it is hard to see how clustering/dependency and intersecting pathways could be handled. Nevertheless, explicit consideration of the four states along the biosecurity continuum provides a coherent logic for characterising expected consequence. This approach has the potential to characterise more completely the potential impacts of pests and diseases on the biosecurity continuum.

A simple but carefully constructed cognitive map can assist these judgments. Cognitive maps are intuitive devices that could be used to assist both consequence estimation and pathway assessment. ACERA has developed a software platform (www.acera.unimelb.edu.au) that implements this tool, and others are available.

The judgments required are very similar to those under the current protocol, with additional estimates for E', EE' and EES', where relevant. Thus, this approach could accommodate the existing expert-based approach to consequence assessment, or more formal and explicit economic analyses that may emerge in the future.

11. Surveillance

The information gained from surveillance systems should be used to validate import risk assessments. For instance, judgements at each step on the exposure pathway imply expected numbers of pests or infected individuals. The feedback of inspection and surveillance information should complete the risk assessment cycle, providing a platform for periodic revision of the risk assessment itself. The methods outlined in the following section provide a basis for specifying which data should be collected to verify estimates of entry made for unrestricted risk, or for the residual risk following implementation of quarantine measures. This is a critical step in ensuring that ALOP has been achieved, and may eventually provide a means to update and improve assessments over time.

Mitigation measures prescribed under an IRA should be least-trade-restrictive. Ideally, an IRA outlines the net effect of mitigation measures required such that the restricted risk meets the ALOP, and consultation with the exporting country decides the most economical or convenient way that this level of mitigation can be effectively achieved. For example, more emphasis may be placed on inspection, or treatments may be a more economical alternative. The necessary quarantine interventions are specified in the import conditions and will be the same for each consignment to which the PRA applies.

This section provides an example application of methods from the report by Robinson et al. (2008), using invented historical data. Robinson et al. (2008) provide background and context for the methods below. Applications need to consider the use of the tool and its output in the context of effectiveness, verification (of processes), validation (of systems) and reporting. The risk framework and analytical strategy uses historical data to identify high-risk import pathways and to prescribe candidate monitoring regimes based on the estimated risk.

11.1 Risk-return inspection methodology

The surveillance problem that we address is as follows. An inspectorate identifies a suite of independent pathways, each comprising sampling units (e.g., containers, packets of seed, vehicles, individual animals). Pathways may refer to importation of different products, via different transport routes, or different trading partners (e.g., Costello et al. 2007). In general, a pathway is an aggregation or stream of like items that can be interpreted as a population for statistical purposes.

Some of the units in the pathway are expected to contain a potentially damaging, invasive species, such units are referred to as ‘contaminated’ units. The inspectorate must identify as many of these contaminated units as possible across all the pathways. Identification of contaminated units is carried out by inspecting the units on the pathway. Inspection resources are finite. We define the ‘approach’ as the number or rate (number per sampling unit) of contaminated items arriving in a pathway, and ‘leakage’ as the number or rate of contaminated items remaining in a pathway after inspection and intervention.

In general, efficiency (the number of detections per unit effort) and effectiveness (the total number of detections) will be enhanced by reducing inspection rates in areas where the risks are smallest, and increasing inspection rates where the risks are highest. These enhancements are counter-balanced by the need to monitor trends in pathways. The key strategic questions are: Should we sample, or census (completely enumerate), a pathway? If we sample the pathway, then at what rate should it be sampled?

A key point is that inspections are the only means by which the degree of contamination of a pathway can be estimated, so inspections play the dual role of interception and education. The quality of information that has already been collected from pathways degrades as time passes.

Given exact information on the contamination rate for each pathway, it is known that the highest number of contaminated units will be identified by inspecting the pathways with the highest contamination rate (see Cannon 2009). Hence, the highest-rate pathways would be searched until all units have been inspected or until resources are exhausted. If resources remain, the pathway with the second highest contamination rate would be inspected. The procedure is iterated until resources are depleted. The lowest-rate pathways may not be inspected at all. We refer to this process as stratifying the pathways. This approach does not (yet) deal with aspects such as correlated inspections, multiple kinds of pests or the potential consequences of different kinds of pests.

Random fluctuations might make one pathway more or less contaminated than is expected based on historical data, so it is possible that pathways may be mis-stratified. The information that is used for the stratification is uncertain, and the uncertainty will increase as the information ages. This characteristic leads to the possibility that pathways with high contamination rates might be ignored because (i) the historical contamination rate was randomly low, or (ii) the contamination rate of the pathway was previously low but has changed in time.

A volume of items is expected on a pathway over some fixed period (a week, a month, or a year). We will denote the number of units expected to enter via the pathway as N, and the predicted probability that an individual unit will contain at least one contaminant will be p. We have historical data comprising n inspections, from which x is the observed number of contaminated items. Then, we assume that x is distributed as (~) a binomial,

$$x \sim \text{Binomial}(n, p) \qquad (11.1)$$

where n is the number of items inspected in the pathway, and p is the predicted (unknown) probability of contamination for each item.

We introduce the idea of 'predicted risk'. The predicted risk that an individual unit will contain at least one contaminant will be denoted f. This risk will be related to the predicted probability of contamination of an item from the pathway, p, inflated to reflect uncertainty about the quality of the prediction.

We argue for the need to consider both the estimate of the contamination probability p and the accuracy of the estimate when allocating resources. Given two commodities with the same contamination rate but different inspection rates, we would be less certain about the estimate of the contamination rate for the commodity that has the lower inspection rate. A risk-sensitive solution requires that we take account of the uncertainty of the estimate, effectively forcing us to pay for higher uncertainty in the same way that we pay for higher contamination rate, by allocating additional resources.

An effective way to achieve a risk-sensitive estimate is to use a quantile of a one-tailed confidence interval, to represent the predicted risk. That is, instead of choosing the value that is the best supported by the data (in this case, p), choose a value, f, that represents the upper limit of the interval with a specified probability of including the true rate under repeated sampling. For example, we might choose a predicted risk that corresponds to the upper 95% confidence interval of the true contamination rate. Informally, we could describe this by saying we are 95% confident that the true contamination rate is less than f (Figure 11.1).

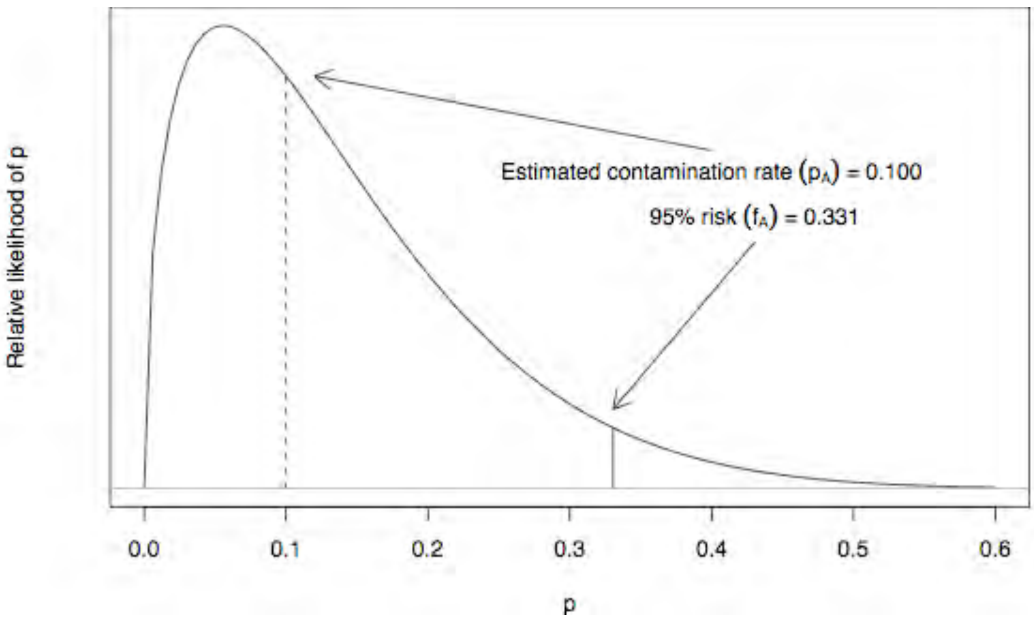


Figure 11.1. Demonstration of the inflation of risk to account for uncertainty. The underlying data are inspections of ten items, of which one was contaminated. The curved line shows the probability density function of the estimated rate, using Jeffrey's approach. The dashed vertical line shows the estimated contamination rate, the solid vertical line shows the one-sided, upper 95% confidence limit on the rate.

The behaviour of different confidence intervals for the binomial rate has been the subject of some research (see, e.g., Madden et al. 1996; Brown et al. 2001; Cai 2005). For one-tailed confidence intervals with robust statistical properties when the true rate is close to 0 or 1, Cai (2005) recommended Jeffrey intervals. Jeffrey intervals are Bayesian intervals, with a prior distribution on the binomial parameter equal to Beta(0.5, 0.5). Jeffrey intervals do not absolutely guarantee (1 – a)% coverage for all possible combinations of samples and a, but they represent a good compromise between coverage and parsimony. If absolute 95% coverage is necessary then a preferred alternative would be Clopper-Pearson intervals, which do guarantee (1 – a)% coverage but are also very conservative (Cai 2005, p. 81).

Jeffrey intervals are easily computable in standard spreadsheets. The upper one-tailed confidence interval is computed from the inverse of the cumulative density function of the Beta distribution. The Beta distribution requires two shape parameters; for our application they are computed as a = x + 0.5 and b = n – x + 0.5. Thus if we wished to use an upper 95% confidence interval then the function call in popular spreadsheet programs for *f*, the estimated 95% upper confidence interval for the predicted risk, would be:

= BETAINV(0.95; x + 0:5; n – x + 0:5); (11.2)

where x is the observed number of contaminated items, and n is the number of items inspected. It is worth noting that Clopper-Pearson intervals can also be easily computed using a version of equation 2 that replaces x + 0.5 by x + 1 and n – x + 0.5 by n – x (see, e.g., Krishnamoorthy 2006). This function is already employed by DAFF staff in the function BINOMCL(n, x) in PopTools (Hood 2009), which returns Jeffrey intervals using equation 11.2.

Strategic resource allocation relies on stratification of the pathways. The stratification approach we recommend follows hypothesis testing as advocated by Neyman and Pearson to guide behaviour (see, e.g., Neyman and Pearson 1933). The inspectorate nominates a risk cutoff, r , e.g. 1%, possibly with input from other stakeholders, and possibly varying across pathways as a function of expected consequence. Then,

- Every pathway with $f \geq r$ is censused.
- Every pathway with $f < r$ is sampled.

In this context, the decision on whether to sample or to census is the outcome of an hypothesis test, for which the null hypothesis is that the contamination rate for the pathway is higher than the cutoff, and the alternative hypothesis is that the contamination rate is less than or equal to the cutoff. A Type-1 error is to sample the pathway when it really should be censused, and a Type-2 error is to census the pathway when it really should be sampled. The determination of inspection rates for low-risk pathways using the predicted risk and the proposed cutoff amount to a power calculation. The inspection rate is chosen such that the probability that the future predicted risk will turn out to be, in fact, higher than r is, e.g., 0.05, if the underlying contamination rate does not increase. In other words, the inspection rate will be set at a sufficiently high level that the probability of mis-classifying the pathway in the future is low.

To find the preferred future monitoring rate m, we must predict the upper limit of the number of contaminated items that we would expect to get if we inspected x items and if the number of contaminated items does not increase during the period of the inspections. We note that the model for sampling, as opposed to censusing, a pathway as described by Equation 12.1 is called a partially-sampled binomial process. The exact probability density function for the total number of contaminated items in a partially-sampled binomial process is hypergeometric (see, e.g., Bain and Patel, 1993). Therefore we need to construct a prediction interval for a hypergeometric distribution. We are unaware of any methods for constructing prediction intervals for hypergeometric distributions with arbitrary parameters, so we deployed the following approximation.

Assume that in the historical data there were x contaminated items in n inspections. Then, the 95% upper confidence limit for the future contamination rate (the predicted risk) is computed using Equation 12.2, with the following changes. The expected number of contaminated items is now $m * x / n$ instead of x, and the number of inspections is (by definition) m instead of n, so, the predicted risk from a future sample of size x is:

$$k95 = \text{BETAINV}(0.95, m / n * x + 0{:}5, m - m / n * x + 0{:}5) * x \qquad (11.3)$$

Next, find the sample size m that will result in a predicted risk lower than the cutoff, r, if the contamination rate is no greater than the conservative value that was estimated from past data. We will call m the tentative future sampling rate. We obtain that sample size by finding the value of m that solves the following equation:

$$r = \text{BETAINV}(0.95, k95 + 0.5, m - k95 + 0.5), \qquad (11.4)$$

where r is the nominated risk cutoff (e.g. 1%). m is an unknown in both Equations 11.3 and 12.4, so they are both required for the solution. In the case of allocation under resource constraints, we fix either the total number of inspections that can be done (S m), or the nominated sampling rate (S m / S n), and find the value for r that satisfies Equation 11.4.

We note that the method suggested for choosing the sample size emphasizes the uncertainty that arises from small sample sizes. There are alternative ways of calculating a sample size that may cause more frequent switching between whether the group should be sampled at 100%, or at a lesser rate, but will also be less conservative in terms of penalizing the uncertainty.

11.2 Inspections

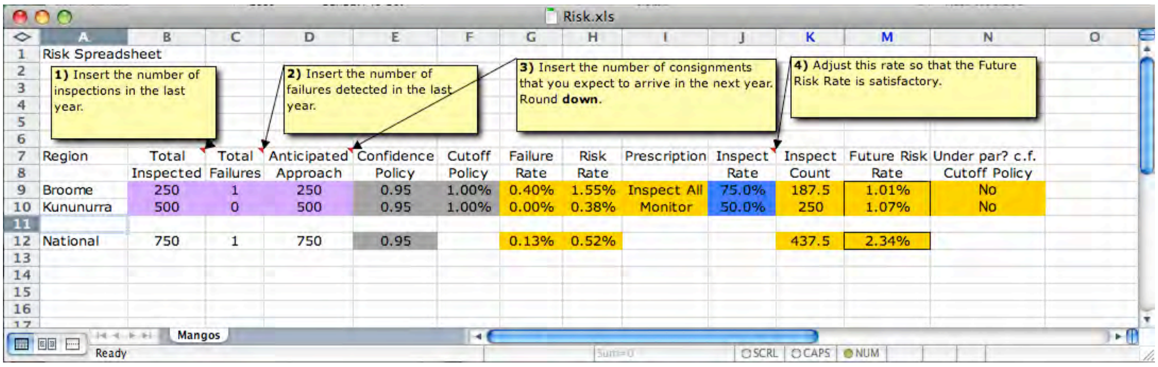
There are three important concepts. The *observed non-conformity rate* for a commodity is the number of observed non-conformities (also called failures) divided by the number of inspections. The *non-conformity risk rate* is the observed non-conformity rate inflated to represent uncertainty about the system. The *future non-conformity risk rate* is the non-conformity risk rate that is expected to be observed for future measurement periods.

Note that the activity levels reported below should be interpreted as a guide rather than as a prescription. Determining the levels of inspection to be used and timing of inspections depend on other operational factors.

The inspection guides are based on the following hypothetical data.

- Broome: 250 consignments inspected last year; 1 interception.
- Kununurra: 500 consignments inspected last years; 0 interceptions.

A screen capture of the spreadsheet is presented in Figure 11.2.



Region	Total Inspected	Total Failures	Anticipated Approach	Confidence Policy	Cutoff Policy	Failure Rate	Risk Rate	Prescription	Inspect Rate	Inspect Count	Future Risk Rate	Under par? c.f. Cutoff Policy
Broome	250	1	250	0.95	1.00%	0.40%	1.55%	Inspect All	75.0%	187.5	1.01%	No
Kununurra	500	0	500	0.95	1.00%	0.00%	0.38%	Monitor	50.0%	250	1.07%	No
National	750	1	750	0.95		0.13%	0.52%			437.5	2.34%	

Figure 11.2. Screen capture of risk spreadsheet.

Instructions for its use are as follows. The mauve cells are for general user input, the grey are for occasional policy input, the blue are for general policy input, and the goldenrod are for output. Note that national figures are computed by aggregating the regional totals.

Input by Users (Mauve cells, to be set each time a prescription is required.)

1. Record the total number of consignments of fruit inspected by region for the last decision period (e.g., year) in column B.
2. Record the total number of failures observed by region for the last decision period in column C.
3. Record the anticipated number of consignments of fruit that will be expected by region for the next period in column D.

Input by Policy Grey cells, to be set rarely, possibly once only.

Columns E and F reflect policy and do not need to be changed as part of the allocation process. They are used to define the policy regarding confidence intervals and the risk cutoff. For example, at 0.95 and 1% respectively, the spreadsheet will guide the user to determine the necessary inspection rate such that if the failure rate does not increase, then the user can be 95% confident that the future risk will be below 1%.

Output 1 in columns G - I, which will change with the user input.

1. Column G reports the observed failure rate.
2. Column H reports the observed risk rate.
3. Column I reports the prescription: monitor the region, or inspect all consignments of fruit in that region.

Planning Blue cells, to be set in response to the output cells.

The values in column J are the inspection rates that the user proposes. Changing these values has two effects: firstly, the number of consignments to be inspected will change (Column K), and secondly, the predicted future risk will change (Column M). Regions that have "Inspect All" in column I are recommended to be assigned 100% in column J. However, the value is at the discretion of the user.

Output 2 in columns K - N, which will change with the user input.

1. Column K reports the number of consignments of fruit to be inspected.

2. Column M reports the predicted future risk rate.

3. Column N reports whether or not the predicted future non{conformity risk rate is under the policy cutoff (Column F).

The goal is to find a compromise between sensible inspection rates and desirable future non-conformity risk rates. After entering the inspection data (Columns B, C, and D), the user should increase or decrease the inspection rates (Column J) until the future non-conformity risk rate (Column M) is at an acceptable level.

11.3 Post-border surveillance

As for border surveillance, information from post-border surveillance should be used to validate import risk assessments. The feedback should complete the risk assessment cycle, providing a platform for periodic revision of the risk assessment itself. Surveillance may also be necessary to support declarations of eradication. For example, on-going surveillance might be justified to maintain claims of pest-free status in WA. The application developed here focuses on how to allocate surveillance effort to detect the pest, assuming it has entered and established in Australia.

We use the surveillance resource allocation model developed by Hauser and McCarthy (2009), which relies on a risk map describing the probability that the species occupies a particular area (within a region). This model can be used to make spatial prioritisations using either a cost-benefit analysis (trading the costs of surveillance directly against the costs of an outbreak) or by maximising cost-effectiveness subject to a limited surveillance budget.

The risk map could be based on expert opinion, but in this case is based on relative probabilities of presence of the species derived from the spatial analysis outlined in section 10. We estimate the cost and efficiency of surveillance (e.g., the cost of establishing traps, and the chance of those traps detecting the species when it is present). This Section will also demonstrate the use of prior information and the validation of IRAs.

The use of a cost-benefit analysis requires the quantification of all likely consequences of outbreak using the same unit of measurement as is used to measure surveillance investment. It may be appropriate to incorporate the income losses to commercial growers in W.A. and/or the value of lost export markets in the face of an outbreak. Consequences for owners of non-commerical trees may be more difficult to quantify (and may or may not be negligible when compared to the consequences for commercial operations). In this report we will focus on surveillance prioritisation under fixed surveillance resources. In the following sections we consider the inputs required for the prioritisation.

Spatial grid. Hauser and McCarthy's method requires that the target heterogeneous landscape be divided into cells, each of which may be considered homogeneous with respect to the probability of occurrence, and each of equal area. For convenience and consistency, the 10 arc-minutes/20 km grid cells specified in the Maxent modelling report (Section 10) are used here.

Occurrence predictions. The surveillance prioritisation requires a predicted probability of occurrence for each grid cell. While the pest is absent from W.A. and the predicted probability is zero, habitat suitability indices provide worthwhile information for prioritising surveillance. In the event of an incursion, habitat suitability could be augmented with information on plausible spread from known sites of infestation.

Since the pest is highly host-specific, the most useful predictive maps would include data on host distribution.

Detection modelling. The relationship between search effort invested and the probability of detecting the pest where it is present is important for prioritisation. This report will focus on searches of trees and surrounding litter rather than sampling and cutting fruits to detect the pest.

Rates of detection are highly uncertain, and this report will use optimistic and pessimistic models to explore the effect of detection uncertainty on surveillance prioritisation (Figure 12.3). Accessing and searching domestic gardens may require different resources than searching commercial plantations; different detection models could be used to differentially prioritise searches of commercial and non-commercial trees.

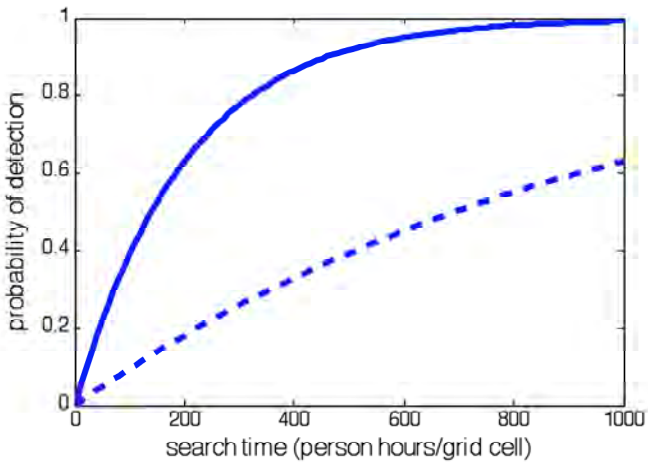


Figure 11.3. Possible probabilities of pest detection in infested grid cells. Solid line gives an optimistic model with surveillance efficacy λ = 0.005 grid cells/person hour, and dashed line gives a pessimistic model with surveillance efficacy λ = 0.001 grid cells/person hour.

Consequences of infestation. The consequences of infestation will be substantially higher for outbreaks occurring in commercial plantations compared to outbreaks of the same magnitude in domestic gardens. Therefore it would be appropriate to weight surveillance prioritisation differently between sites containing commercial and non-commercial trees, or using a measure of proximity to commercial trees. Only the highest risk non-commercial trees would be targeted for surveillance, unless the surveillance budget is very large. In the absence of information on tree distribution, this weighting has not been included in the report.

Hauser and McCarthy’s method finds the allocation of search effort over the landscape that minimises the expected costs of an infestation. Using methods of optimisation, they found that the cost-effectiveness of searching any given site in the landscape can be scored as

$$\text{Prioritisation score(site } i) = p_i \times \lambda_i \times R_i$$

where p_i is the predicted probability of pest occurrence (we use habitat suitability from Maxent modelling), λ_i is the efficacy of surveillance at the site (we consider the optimistic and pessimistic models in Figure 12.3), and R_i is the consequences of an infestation at the site. Since we have not estimated variation in consequences across the landscape, we set $R_i = 1$ for all sites.

Labelling sites in descending order of prioritisation score, the optimal amount of search effort to allocate to a site i is

$$x_i^* = \begin{cases} \dfrac{\ln\left[p_i\lambda_i R_i\right]}{\lambda_i} + \dfrac{\overline{\lambda}_K}{\lambda_i}\left[\dfrac{B}{k} - \overline{x}_K\right], & i = 1, 2, \ldots, k \\ 0, & i = k+1, k+2, \ldots, n, \end{cases}$$

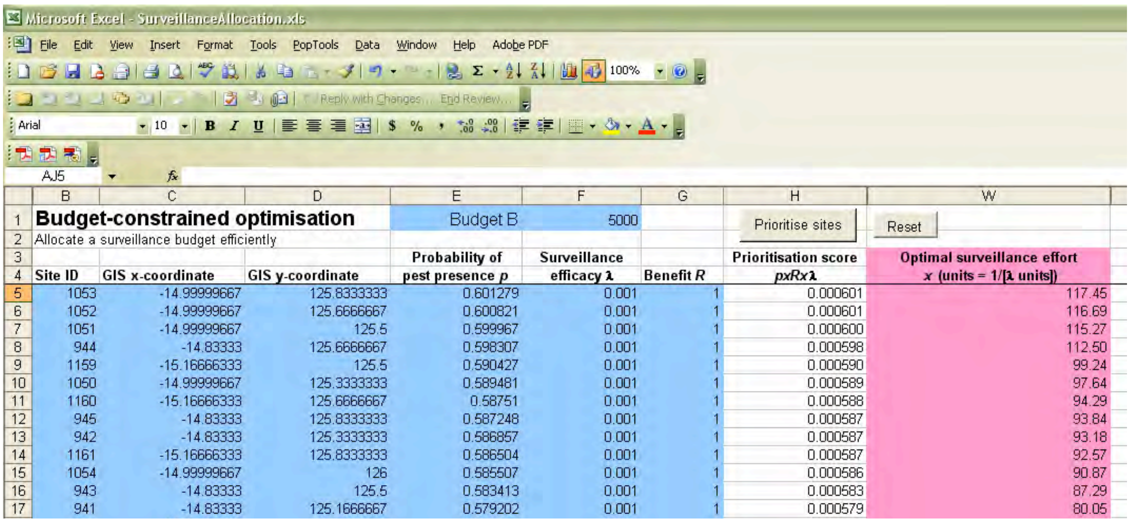
where

$$\overline{\lambda}_K = \frac{k}{\sum_{j=1}^{k} \lambda_j^{-1}} \text{ and } \overline{x}_K = \frac{1}{k}\sum_{j=1}^{k} \frac{\ln\left[p_j\lambda_j R_j\right]}{\lambda_j}$$

and B is the total budget available for surveillance. These equations demonstrate that sites with the highest prioritisation score should be searched, while sites with the lowest prioritisation scores will not be searched at all. The amount of survey effort allocated to the high priority sites is a logarithmic function of the prioritisation score; it is amended to account for the surveillance efficacy at this site i compared to other sites ($\overline{\lambda}_K / \lambda_i$) and the total budget available (B).

This cost-effectiveness analysis can be carried out in an Excel spreadsheet when the number of grid cells is sufficiently small (i.e. less than 65,536 grid cells[4]) and such a spreadsheet was used to prioritise the 9479 grid cells.

[4] Excel 2007 and later versions do not have this limitation.



Site ID	GIS x-coordinate	GIS y-coordinate	Probability of pest presence *p*	Surveillance efficacy λ	Benefit *R*	Prioritisation score $p \times R \times \lambda$	Optimal surveillance effort *x* (units = 1/[λ units])
1053	-14.99999667	125.8333333	0.601279	0.001	1	0.000601	117.45
1052	-14.99999667	125.6666667	0.600821	0.001	1	0.000601	116.69
1051	-14.99999667	125.5	0.599967	0.001	1	0.000600	115.27
944	-14.83333	125.6666667	0.598307	0.001	1	0.000598	112.50
1159	-15.16666333	125.5	0.590427	0.001	1	0.000590	99.24
1050	-14.99999667	125.3333333	0.589481	0.001	1	0.000589	97.64
1160	-15.16666333	125.6666667	0.58751	0.001	1	0.000588	94.29
945	-14.83333	125.8333333	0.587248	0.001	1	0.000587	93.84
942	-14.83333	125.3333333	0.586857	0.001	1	0.000587	93.18
1161	-15.16666333	125.8333333	0.586504	0.001	1	0.000587	92.57
1054	-14.99999667	126	0.585507	0.001	1	0.000586	90.87
943	-14.83333	125.5	0.583413	0.001	1	0.000583	87.29
941	-14.83333	125.1666667	0.579202	0.001	1	0.000579	80.05

Figure 11.2. Screen capture of the prioritisation spreadsheet. Some of details of the calculations are outlined in the text.

Recommended allocations of search effort are plotted for the pessimistic and optimistic detection models for a range of budgets (Figure 11.3). If detection is relatively poor, then intense searches should be conducted in only the highest risk grid cells (habitat suitability > 0.5, Fig. 11.3a). When detection rates are higher, effort can be spread across a larger number of grid cells with sufficient probability of infestation detection within those cells (Fig. 11.3b). Increases in surveillance investment (budget) should be used to incorporate some new grid cells into the plan, as well as increasing search intensity at already-targeted cells.

Optimisations within the spreadsheet are calculated quickly, so it is possible to explore a range of scenarios when input parameters are not known with certainty. The objective function is *expected* cost and has a linear relationship with probability of occurrence p_i and consequences R_i. It follows that

1. Only the *expected* (mean) values of p_i and R_i are required for each site (not necessarily known point values)
2. For the budget-constrained case, only the relative values of p_i and R_i among sites are required (e.g. knowledge that the pest is twice as likely to occur at site 1 as at site 2 is sufficient; the two probability values are not necessary).

Therefore, if the probability of occurrence or the consequences of infestation across the landscape are not known with certainty, they can be represented by probability distributions. Their mean values can be used to determine the most cost-effective allocation, but their full uncertainty distributions can be retained and updated as new information comes to hand. In particular, the survey itself is likely to provide information on which sites are suitable for infestation. Updated distributions can be used to allocate effort for the next iteration of surveys.

A more risk-averse strategy would be to use upper percentiles of the uncertainty distributions in the analysis. This allows for higher rates of occurrence and/or larger consequences of infestation across the landscape. The optimal surveillance allocation produced by Hauser and McCarthy’s method will consequently spread surveillance effort more broadly across the landscape, with less effort allocated to each site, than if the distribution means are used.

When surveillance efficacy (λ_i) at each site is uncertain, it is likewise possible to explore a range of scenarios or substitute percentile values taken from uncertainty distributions. However, the means of those distributions are not direct substitutes for known parameter values. Surveillance allocation strategies (either risk-neutral or risk-averse) have not yet been investigated when detection rates are uncertain.

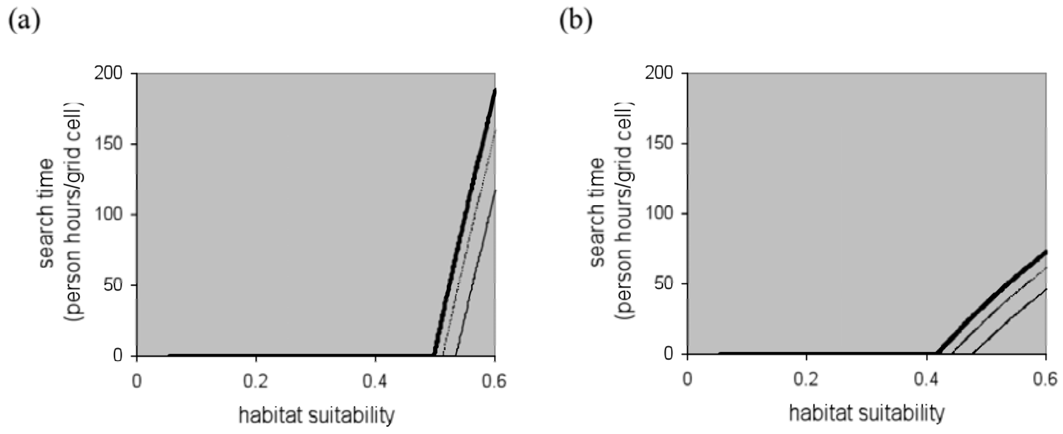


Figure 11.3. Optimal budget-constrained allocations of search effort as a function of habitat suitability: (a) for pessimistic detection model λ = 0.001, and (b) for optimistic detection model λ = 0.005. Total search budgets are 5000 (thin line), 10000 (dotted line) and 15000 (thick line) person hours per grid cell.

11.4 Strengths, weaknesses, opportunities, threats

It is best to consider operational constraints and opportunities when determining the inspection rate. Choosing a precise but low inspection rate might not be as useful, overall, as choosing a coarser and higher rate that is easy to implement. For example, if fruit consignments arrive 7 days per week, then multiples of 15%, corresponding to inspection on approximately 1 day per week, should be entertained.

In Figure 12.2, we adopt inspection rates of 50% and 75%. However, these activity levels do not include prescriptions for the timing of inspections. That is, a prescription of 50% could refer to inspecting, for example:

1. every second consignment,
2. a single random consignment out of every two,
3. all consignments on every second day, or
4. all consignments on one randomly-chosen day per two days, or
5. all consignments on fifteen randomly-chosen days per month.

This list is not intended to be exhaustive, and there may be alternatives that prove better suited to the circumstances at hand.

The different scenarios listed above have different implications for security and efficiency, and their suitability will depend on the rate at which consignments arrive, and the flexibility that managers have to allocate personnel resources to inspection. For example, a systematic pattern would be easier to implement, but a random pattern would be more secure.

The results of inspection at the border can be used to confirm that the pest levels and pest types are consistent with Australia’s ALOP; this is especially useful in the first few years of the trade. For instance, an excessive number of quarantine pests would indicate that the mitigation measures should be reviewed. The type and frequency of quarantine pests detected may provide useful background information for PRA for similar commodities and pests.

There are also opportunities to improve the efficiency of quarantine intervention, in particular with respect to auditing and non-commodity inspection. Building a profile of the risk posed by various combinations of commodity, country of origin, exporter, etc. allows inspectors to target higher risk commodities by either a more thorough or more frequent inspection. These concepts apply to the inspection and management of both imports and exports.

The risk tool distinguishes between different monitoring scenarios put forward by the program, but the program needs to consider the use of the tool and its output in the context of effectiveness, verification (of processes), validation (of systems), and reporting. Implementation and feedback to import risk assessment will depend on the development of systems to record appropriate data and feed it back in a timely fashion to those who are concerned with the estimation of the likelihood, of entry, establishment and spread of pests and diseases.

Surveillance systems work most effectively when conditioned by informative prior information. Import risk analyses incorporate substantial ecological and operational information that could be used to target observational effort to greatest benefit.

As noted above, the information from surveillance systems should be used to validate and update import risk assessments. There have been significant advances in analytical tools for surveillance internationally, especially in tools relevant to biosecurity. A substantial ACERA review (ACERA Project 1002) provides a relatively complete overview and identifies those that are useful in routine biosecurity settings in Australia

12. Conclusions

In Australian IRAs, the assessment of the probability of entry, establishment and spread[5] is described as ‘*qualitative*’ and for entry is said to be ‘*based on pathway scenarios depicting necessary steps in the sourcing of the commodity for export, its processing, transport and storage, its utilisation in Australia and the generation and disposal of waste*’ (e.g., BA 2008, p.14).

The claim that the risk analysis is qualitative reinforces a false dichotomy between quantitative and so-called qualitative methods. The objectives of the analysis – to measure the overall probability and consequences of entry, establishment and spread – and the overall goal – to determine whether the unrestricted risk exceeds Australia’s acceptable level of protection (ALOP) – do not describe qualities, such as colour of fruit or wellbeing of farmers, but estimate quantities (i.e., probabilities, consequence metrics and an overall risk score).

The probabilities (also referred to as likelihoods) are described with words, such as ‘High’ or ‘Low’, but these words are explicitly quantified through the use of a table that defines the nomenclature using a numerical interval for each word (see Section 2.2.2 of BA (2008)). Furthermore, these interval estimates of the likelihood of the occurrence of ‘the event’[6] are combined in a structured way (roughly concordant with the rules of probability) to produce an estimate of the overall probability of entry, establish and spread (Section 5). Ultimately, through combination of these interval estimates with a composite metric of estimated consequence, an estimate of overall risk is produced. This estimate of unrestricted risk is compared to the decision threshold (Australia’s ALOP) in a manner that should be described as semi-quantitative and model-based.

12.1 Reasons for model-based biosecurity risk analysis

Import risk analyses are model-based and it would be best, from a scientific perspective, if the models were as explicit as possible (e.g., through diagrams, appropriately constructed, consistent rules and, where possible, numerical or mathematical expressions), for a number of reasons, including:

- *Consistency:* the use of explicit diagrams, written descriptions and numbers to model connections, dependencies and parameters allows a model to be tested for consistency and compared with other models.
- *Repeatability:* published, explicit models can be reproduced and tested by third parties, opening methods and assumptions for discussion.
- *Avoidance of ambiguity:* where model structure (e.g., pathways), parameters, assumptions, values, thresholds and objectives are explicit, the potential for misunderstandings of the meaning of estimates of probability, consequence and overall risk is reduced.
- *Treatment of dependency:* where a model is explicitly stated, any known or suspected dependencies between steps can be taken into consideration. Dependencies can be

[5] ‘Entry’, ‘establishment’ and ‘spread’ are steps on the general pest invasion pathway used for all pest risk analyses and are defined in FAO (2004).
[6] This term is not defined and is postulated to refer either to the survival/entry of at least one weevil or to the proportion of weevils that survive (or the proportion of mangos that remain infested).

addressed through design of model structure, conditioning of experts during elicitation or mathematical correlation of parameters in the model.

- *Quantification of uncertainty:* explicit presentation of the data, their sources and uncertainties (known or estimated), and the model used to combine them is necessary for uncertainties associated with outputs to be quantified.

- *Analysis of sensitivity:* an explicit model structure is a necessary precursor for sensitivity analysis.

- *Logical combination of intervals:* where intervals are used as input data instead of point estimates or distributions, interval arithmetic ensures the output is consistent with the inputs.

- *Communication with stakeholders:* communication with some stakeholders, particularly those who wish to test or verify the model estimates, is more likely to be satisfactory following full disclosure of methods, assumptions or values used in the model.

- *Decision making under uncertainty:* Should discussion reveal disagreement about facts (e.g., alternative parameter estimates that are consistent with data), then decisions should record these uncertainties together with the decision-maker's attitude towards uncertainty. Should the discussion reveal disagreement about values (e.g., the calculation of consequences, or the definition of ALOP), these disagreements can be made explicit and resolved through alignment of policy, while complying with legislation and international agreements.

13.2 What issues have the demonstrations highlighted?

The demonstration report documents a range of applications and evaluates their strengths and weaknesses for biosecurity risk analysis applications. The most important of these issues in current IRAs include:

1. There is no explicit link between the prose describing the ecology and behaviour of the species, and the analyst's judgement of risk. That is, the pathway on which the risk analysis is based was described through the statements of 'factors' and the bullet point summaries of information about pests. These factors and summaries are clustered under the broad steps of importation, distribution, establishment and spread but are not explicitly connected to the estimate of likelihood. For example, there is no systematic method for combining data (e.g., number of days of survival with/without food) to arrive at the estimates of likelihood.

2. Information is potentially incomplete. All risk analyses operate with incomplete information. However, where information is provided in written form without an explicit model, it is difficult to ascertain where the data gaps are and, therefore, to undertake a meaningful sensitivity analysis to assess the impact of the data gaps on the model estimates and resulting decisions. Furthermore, much of the information depends on the volume and timing (seasonality) of trade. For example, areas of distribution within Western Australia depend on the total volume of trade. However, it is not clear whether the estimates of probability or, particularly, the method of combining probabilities, take these dependencies into account.

3. The matrix rules are only approximately consistent with formal logic for combining intervals (interval arithmetic). They differ from exact rules in a number of potentially important ways.

4. Methods for gaining estimates from experts are informal and unstructured, making them relatively susceptible to dominance, anchoring, motivational bias and psychological effects.

5. The rules for combining likelihood judgements do not account for disagreement among experts, or for other sources of uncertainty.

6. Subjective methods for estimating spread are, in fact, approximate methods for climate matching – explicit spatial models could supply more accurate information efficiently and repeatably.

7. Consequence estimation underestimates the risks associated with hazards that express themselves prior to establishment or spread.

More generally, we have shown (sections 2 and 10) that model-based analysis can be unduly restrictive, diverting attention from events and pathways that are important in some IRAs, but which are not central for generic analyses. Where results cannot be reproduced precisely as originally derived, examined and tested, many alternative (and equally valid) results can be derived from the information presented. This creates a risk that the results can be disputed or dismissed as contrived.

13.3 Will the alternative tools improve risk assessments?

The evidence from these applications indicates some areas where gains in the quality of risk analysis may be achieved for modest cost.

1. An example pathway model is presented in Figure 2.2. This figure is a conceptual diagram of the pathway that is further developed quantitatively in subsequent sections of this report (see Figures 2.2 and 5.1 to 5.4). Explicit, mathematical models such as these identify and account for incomplete information (directing sensitivity analysis and subsequent data collection) and appropriately handle dependent information (through model structure and conditioning of experts). Cognitive maps and event trees will assist with correct conditioning of likelihood assessments, estimating consequences prior to spread, and evaluating systems alternatives when designing risk mitigations. The pathways in some current Australian IRAs are more detailed and concrete than those shown in section 2.

2. The procedures for choosing and questioning experts, and combining their opinions, provide more reliable answers than unstructured techniques. They offer the possibility of improving the quality of expert judgements (in terms of both accuracy and reliability), both within the existing framework, and if used in conjunction with other techniques.

3. The rules for combining probabilities could be revised to be consistent with interval arithmetic, guaranteeing that outputs are consistent with inputs. This may lead to some

assessments bridging the ALOP, resulting in a requirement to obtain additional information to reduce uncertainty.

4. Bayes Nets offer a natural framework for incorporating the current procedures, for modifying them to deal with uncertainty, to deal consistently with likelihood estimates, and to build a library of examples that could guide the development of future risk assessments.

5. Monte Carlo simulations may be of some use in contentious contexts, and when data are relatively abundant. In general, they should be accompanied by p-bounds analysis, to ensure that unjustified assumptions do not have a critical bearing on decisions. The transparency this engenders does not guarantee acceptance, and may even encourage greater disagreement.

6. Statistical models for predicting spread are likely to be useful for cost-sharing discussions, when biological factors governing the distribution of a species are uncertain, when relatively precise predictions are important for consequence assessment, or when effective surveillance strategies are required.

7. Risk assessments could be used to specify inspection and surveillance systems. Data collected from these systems could then be used to validate the risk analysis assumptions, and provide a basis for periodic revision and reassessment of the risks.

The tools demonstrated above could move the organisation into a ‘learning’ mode. That is, this demonstration does not establish unequivocally that they will be useful, or useful enough to justify the investment to implement them. Rather, those that appear to have some utility may become the focus for training materials, providing platforms for new risk analysts to experience alternative ways of analysing problems, test assumptions, evaluate criticisms, and experience the technical overheads required for implementation.

References

Armstrong, J.S. (Ed.) (2001) *Principles of forecasting: a handbook for researchers and practitioners.* Boston: Kluwer Academic Publishers.

Anderson, R.P., Lew, D. and Peterson, A.T. (2003) Evaluating predictive models of species' distributions: criteria for selecting optimal models. *Ecological Modelling*, **162:** 211 - 232.

Araújo, M.B., Pearson, R.G., Thuiller, W. and Erhard, M. (2005) Validation of species –climate impact models under climate change. *Global Change Biology,* **11:** 1504 - 1513.

Arriaga, L., Castellanos, A.E., Moreno, E. and Alarcón, J. (2004) Potential ecological distribution of alien invasive species and risk assessment: a case study of buffel grass in arid regions of Mexico. *Conservation Biology*, **18:** 1504 - 1514.

Arrignon, F., Deconchat, M., Sarthou, J., Balent, G. and Monteil, C. (2007) Modelling the overwintering strategy of a beneficial insect in a heterogeneous landscape using a multi-agent system. *Ecological Modelling*, **205:** 423 - 436.

Arrow, K.J. 1951. *Social Choice and Individual Values*, New York: Wiley.

Austin, M.P. (2002) Spatial prediction of species distribution: an interface between ecological theory and statistical modelling. *Ecological Modelling*, **157:** 101 - 118.

Austin, M.P. (2007) Species distribution models and ecological theory: a critical assessment and some possible new approaches. *Ecological Modelling,* **200:** 1 - 19.

Austin, M.P., Nicholls, A.O. and Margules, C.R. (1990) Measurement of the realised qualitative niche: environmental niches of five Eucalyptus species. *Ecological Monographs*, **60:** 161 - 177.

Australian Government Department of Agriculture, Fisheries and Forestry (2007) Import Risk Analysis Handbook 2007, Canberra. ISBN 0-97504-730-2. Available at http://www.daff.gov.au/ba/ira/process-handbook

Axelrod, R. (1976). Structure of Decision: The Cognitive Maps of Political Elites. Princeton University Press, Princeton, NJ.

Barlow, N.D., Beggs, J.R. and Barron, M.C. (2002) Dynamics of common wasps in New Zealand beech forests: a model with density dependence and weather. *Journal of Animal Ecology*, **71:** 663 - 671.

Barron, M.C. (2007) Retrospective modelling indicates minimal impact of non-target parasitism by *Pteromalus puparum* on red admiral butterfly (*Bassaris gonerilla*) abundance. *Biological Control*, **41:** 53 - 63.

Barry, S. (2006) Climate software manual (PC version). Bureau of Rural Sciences, Canberra, http://affashop.gov.au/product.asp?prodid=13506

Barry, S. and **Elith,** J. (2006) Error and uncertainty in habitat models. *Journal of Applied Ecology,* **43:** 413 - 423.

Ben-David, A. (2008) About the relationship between ROC curves and Cohen's kappa. *Engineering Applications of Artificial Intelligence*, **21:** 874 - 882.

BenDor, T.K., Metcalf, S.S., Fontenot, L.E., Sangunett, B. and Hannon, B. (2006) Modeling the spread of the Emerald Ash Borer. *Ecological Modelling*, **197:** 221 - 236.

Bio, A.M.F., Becker, P., Bie, E., Huybrechts, W. and Wassen, M. (2002) Prediction of plant species distribution in lowland river valleys in Belgium: modelling species response to site conditions. *Biodiversity and Conservation*, **11:** 2189 - 2216.

Bomford, M. (2006) Risk assessment for the establishment of exotic vertebrates in Australia: recalibration and refinement of models. Canberra: Department of the Environment and Heritage

Boyce, M.S., Vernier, P.R., Nielsen, S.E., and Schmiegelow, F.K.A. (2002) Evaluating resource selection functions. *Ecological Modelling,* **157:** 281 - 300.

Burgman et al 2010. Redifining expertise. In Review (in Conservation Letters).

Burgman, M.A. (2005) *Risks and Decisions for Conservation and Environmental Management*, Cambridge: Cambridge University Press.

Cawsey, E.M., Austin, M.P. and Baker, B.L. (2002) Regional vegetation mapping in Australia: a case study in the practical use of statistical modelling. *Biodiversity Conservation*, **11:** 2239 - 2274.

Centre for International Earth Science Information Network (CIESIN), Columbia University; and Centro Internacional de Agricultura Tropical (CIAT) (2005) Gridded Population of the World Version 3 (GPWv3): Population Density Grids. Palisades, NY: Socioeconomic Data and Applications Center (SEDAC), Columbia University. Available at http://sedac.ciesin.columbia.edu/gpw (4/11/2008 - date of download).

Clemen, R. T. (1989). Combining forecasts: A review and annotated bibliography. International Journal of Forecasting, 5, 559−583.

Clemen, R. T. and R. L. Winkler (1999). Combining probability distributions from experts in risk analysis. *Risk Analysis* **19**(2): 187-203.

Clemen, R.T. and Winkler, R.L. (1985). Limits for the precision and value of information from dependent sources. *Operations Research*, **33**: 427-442.

Coelho, J.R. and Ross, A.J. (1996) Body temperature and thermoregulation in two species of yellowjackets, *Vespula germanica* and *V. maculifrons*. *Journal of Comparative Physiology B*, **166:** 68 - 76.

Collingham, Y.C., Wadsworth, R.A., Huntley, B. and Hulme, P.E. (2000) Predicting the spatial distribution of non-indigenous riparian weeds: issues of spatial scale and extent. *Journal of Applied Ecology,* **37:** 13 - 27.

Cook, D.C., Thomas, M.B., Cunningham, S.A., Anderson, D.L. and De Barro, P.J. (2007) Predicting the economic impact of an invasive species on an ecosystem service. *Ecological Applications*, **17:** 1832 - 1840.

Cooke, R. M. 1991. Experts in uncertainty: opinion and subjective probability in science. Oxford University Press, Oxford.

Cooke. R. M. and Goosens, L. H. J. 2004. Expert judgement elicitation for risk assessments of critical infrastructures. *Journal of Risk Research, 7* (6), 643–656.

Crespi, B.J. (1994) Three conditions for the evolution of eusoiciality: are they sufficient? *Insectes Sociaux*, **41:** 395 - 400.

Crombie, J., Brown, L., Lizzio, J. and Hood, G. (2008) Climatch user manual. Bureau of Rural Sciences, Canberra. http://adl.brs.gov.au:8080/Climatch/

Crosland, M.J.W. (1991) The spread of the social wasp, *Vespula germanica,* in Australia. *New Zealand Journal of Zoology*, **18:** 375 - 388.

Crozier, R.H. (2008) Advanced eusociality, kin selection and male haploidy. *Australian Journal of Entomology*, **47:** 2 - 8.

D’Adamo, P., Sackman, P. and Corley, J.C. (2002) The potential distribution of German wasps (*Vespula germanica*) in Argentina. *New Zealand Journal of Zoology*, **29:** 79 - 85.

Dalkey, N. and Helmer, O. 1951. The use of experts for the estimation of bombing requirements – a project Delphi experiment, Rand Report RM-7272-PR, November 1951.

Donovan, B.J. (1984) Occurrence of the common wasp, *Vespula vulgaris* (L.) (Hymenoptera:Vespidae) in New Zealand. *New Zealand Journal of Zoology,* **11:** 417 - 427.

Dormann, C.F. (2007) Promising the future? Global change projections of species distributions. *Basic and Applied Ecology,* **8:** 387 - 397.

Duncan, R.P., Bomford, M., Forsyth, D.M. and Conibear, L. (2001) High predictability in introduction outcomes and the geographical range size of introduced Australian birds: a role for climate. *Journal of Animal Ecology*, **70:** 621 - 632.

Eeva, T., Sorvari, J. and Rinne, V. (2006) A likely German wasp (*Vespula germanica*) breeding colony in continental Finland. *Sahlbergia*, **11:** 53 - 54.

Ehler, L.E. (1998) Invasion Biology and Biological Control. *Biological Control,* **13:** 127 - 133.

Elith *et al*., 2006

Elith, J. and Graham, C. (2009) Do they? How do they? WHY do they differ? On finding reasons for differing performances of species distribution models. *Ecography*, **32:** 66 - 77.

Elith, J. and Leathwick, J.R. (2007) Predicting species’ distributions from museum and herbarium records using multiresponse models fitted with multivariate adaptive regression splines. *Diversity and Distributions,* **13:** 165 - 175.

Elith, J. & Leathwick, J.R. (2009) Conservation prioritization using species distribution models. *Spatial conservation prioritization: quantitative methods and computational tools* (eds A. Moilanen, K. A. Wilson & H. P. Possingham), pp. 70-93. Oxford University Press.

Elith, J., Ferrier, S., Huettmann, F. and Leathwick, J.R. (2005) The evaluation strip: a new and robust method for plotting predicted responses from species distribution models. *Ecological Modelling*, **186:** 280 - 289.

Elith, J., Graham, C.H., Anderson, R.P., Dudík, M., Ferrier, S., Guisan, A., Hijmans, R.J., Huettmann, F., Leathwick, J.R., Lehmann, A., Li, J., Lohmann, L.G., Loiselle, B.A., Manion, G., Moritz, C., Nakamura, M., Nakazawa, Y., Overton, J.McC., Peterson, A.T., Phillips, S.J., Richardson, K.S., Scachetti-Pereira, R., Schapire, R.E., Soberón, J., Williams, S., Wisz, M.S. and Zimmermann, N.E. (2006) Novel methods improve prediction of species’ distributions from occurrence data. *Ecography*, **29:** 129 - 151.

ESRI (1999) ArcView GIS 3.2. Environmental Systems Research Institute, Redlands, California.

ESRI (2007) ArcInfo GIS 9.2. Environmental Systems Research Institute, Redlands, California.

Estrada-Peña, A.E., Acedo, C.S., Quílez, J. and Del Cacho, E. (2005) A retrospective study of climatic suitability for the tick *Rhipicephalus (Boophilus) microplus* in the Americas. *Global Ecology and Biogoegraphy*, **14:** 565 - 573.

FAO 2004, ‘International standards for phytosanitary measures Publication No. 11: Pest risk analysis for quarantine pests, including analysis of environmental risks and living modified organisms’, Report, Secretariat of the International Plant Protection Convention, Food and Agriculture Organisation of the United Nations, Rome.

Ferson, S. and Hajagos, J.G. (2004) Arithmetic with uncertain numbers: rigorous and (often) best possible answers. *Reliability Engineering and System Safety*, 85: 135-52.

Ficeotla *et al*., 2009;

Fielding, A.H. and Bell, J.F. (1997) A review of methods for the assessment of prediction errors in conservation presence/absence models. *Environmental Conservation*, **24:** 38 - 49.

Fitzpatrick, M.C., Weltzin, J.F., Sanders, N.J. and Dunn, R.R. (2007) The biogeography of prediction error: why does the introduced range of the fire ant over-predict its native range? *Global Ecology and Biogeography,* **16:** 24 - 33.

Forsyth, D.M., Duncan, R.P., Bomford, M. and Moore, G. (2004) Climatic suitability, life-history traits, introduction effort, and the establishment and spread of introduced mammals in Australia. *Conservation Biology*, **18:** 557 - 569.

Frey, J.K. (2009) Distinguishing range expansions from previously undocumented populations using background data from museum records. *Diversity and Distributions*, **15:** 183 - 187.

Gates, D.M. (1980) Biophysical Ecology. Springer-Verlag, New York.

Gigerenzer, G. (2007). *Gut feelings: The intelligence of the unconscious.* New York: Viking.

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

Gilbert, M., Guichard, S., Freise, J., Grégoire, J.C., Heitland, W., Straw, N., Tilbury, N. and Augustin, S. (2005) Forecasting *Cameraria ohridella* invasion dynamics in recently invaded countries: from validation to prediction. *Journal of Applied Ecology,* **42:** 805 - 813.

Global Biodiversity Information Network (GBIF) (2008) Field Museum of Natural History, Museum of Vertebrate Zoology, University of Washington Burke Museum, and University of Turku (Accessed through GBIF Data Portal, www.gbif.net, 2008-07-23).

Goodisman, M.A.D., Matthews, R.W. and Crozier, R.H. (2001a) Hierarchical genetic structure of the introduced Wasp *Vespula germanica* in Australia. *Molecular Ecology*, **10:** 1423 - 1432.

Goodisman, M.A.D., Matthews, R.W., Spradbery, P.J., Carew, M.E. and Crozier, R.H. (2001b) Reproduction and recruitment in perennial colonies of the introduced wasp *Vespula germanica. The Journal of Heredity*, **92:** 346 - 349.

Goutte, C. (1997) Note on free lunches and cross-validation. *Neural Computation*, **9:** 1211 - 1215.

Guichón, M.L. and Doncaster, C.P. (2008) Invasion dynamics of an introduced squirrel in Argentina. *Ecography*, **31:** 211 - 220.

Guisan, A. and Thuiller, W. (2005) Predicting species distribution: offering more than simple habitat models. *Ecology Letters,* **8:** 993 - 1009.

Guisan, A. and Zimmerman, N.E. (2000) Predictive habitat distribution models in ecology. *Ecological Modelling,* **135:** 147 - 186.

Hanley, J.A. and McNeil, B.J. (1982) The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology*, **143:** 29 - 36.

Harris, R.J. (1991) Diet of the wasps *Vespula vulgaris* and *V. germanica* in honeydew beech forest of the South Island, New Zealand. *New Zealand Journal of Zoology,* **18:** 159 - 170.

Harris, R.J. (1996) Frequency of over wintered *Vespula germanica* (Hymenoptera: Vespidae) colonies in scrubland- pasture habitat and their impact on prey. *New Zealand Journal of Zoology,* **23:** 11 - 17.

Hartley, S., Harris, R. and Lester, P.J. (2006) Quantifying uncertainty in the potential distribution of an invasive species: climate and the Argentine ant. *Ecology Letters,* **9:** 1068 - 1079.

Hastings, A., Cuddington, K., Davies, K.F., Dugaw, C.J., Elmendorf, S., Freestone, A., Harrison, S., Holland, M., Lambrinos, J., Malvadkar, U., Melbourne, B.A., Moore, K., Taylor, C. and Thomson, D. (2005) The spatial spread of invasions: new developments in theory and evidence. *Ecology Letters*, **8:** 91 - 101.

Hauser, C.E. and McCarthy, M.A. (2009) Streamlining 'search and destroy': cost-effective surveillance for invasive species management. *Ecology Letters* 12: 683—692.

Heinrich, B. (1974) Thermoregulation in endothermic insects. *Science,* **185:** 747 - 756.

Hernandez, P.A., Graham, C.H., Master, L.L. and Albert, D.L. (2006) The effect of sample size and species characteristics on performance of different species distribution modeling methods. *Ecography,* **29:** 773 - 785.

Hess, G.R. and King T.J. 2002. Planning open spaces for wildlife I. Selecting focal species using a Delphi survey approach. *Landscape and Urban Planning*, 58: 25-40.

Hijmans, R.J., Cameron, S.E., Parra, J.L., Jones, P.G. and Jarvis, A. (2005) Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology,* **25:** 1965 - 1978.

Hinojosa-Díaz, I.A., Feria-Arroyo, T.P. and Engel, M.S. (2009) Potential distribution of orchid bees outside their native range: the cases of *Eulaema polychroma* (Mocsáry) and *Euglossa viridissima* Friese in the USA (Hymenoptera: Apidae). *Diversity and Distributions,* **15:** 421 - 428.

Hochberg, M.E. and Gotelli, N.J. (2005) An invasions special issue. *Trends in Ecology and Evolution,* **20:** 211 - 211.

Hood, G. (2009). PopTools, Version 3.1. [url] http://www.cse.csiro.au/poptools

Horwood, M.A., Toffolon, R.B. and Brown, G.R. (1993) Establishment and spread of *Vespula germanica* (F.) (Hymentoptera: Vespidae) in New South Wales and the influence of rainfall on its abundance. *Journal of the Australian Entomological Society*, **32:** 241 - 248.

Inglis, G.J., Hurren, H., Oldman, J. and Haskew, R. (2006) Using habitat suitability index and particle dispersion models for early detection of marine invaders. *Ecological Applications*, **16:** 1377 - 1390.

Jose V. R. R and Winkler, R. L. (1992). Simple robust averages of forecasts: Some empirical result. International Journal of Forecasting 24 (2008) 163–169

Juliano, S.A. and Lounibos, L.P. (2005) Ecology of invasive mosquitoes: effects on resident species and on human health. *Ecology Letters*, **8:** 558 - 574.

Kadane, J.B. and Wolfson, L.J., 1998. Experiences in elicitation. *Journal of the Royal Statistical Society Series D – the Statistician*, **47**: 3-19.

Kasper, M.L., Reeson, A.F., Mackay, D.A. and Austin, A.D. (2008) Environmental factors influencing daily foraging activity of *Vespula germanica* (Hymenoptera, Vespidae) in Mediterranean Australia. *Insectes Sociaux*, DOI 10.1007/s00040-008-1004-7.

Kasper, M.L., Reeson, A.F., Mackay, D.A. and Austin, A.D. (2008) Environmental factors influencing daily foraging activity of *Vespula germanica* (Hymenoptera, Vespidae) in Mediterranean Australia. *Insectes Sociaux*, DOI 10.1007/s00040-008-1004-7.

Kearney, M. (2006) Habitat, environment and niche: what are we modelling? *Oikos*, **115:** 186 - 191.

Kearney, M.R. and Porter, W.P. (2004) Mapping the fundamental niche: Physiology, climate and the distribution of a nocturnal lizard. *Ecology,* **85:** 3119 - 3131.

Kearney, M.R. and Porter, W.P. (2008) Mechanistic niche modelling: combining physiological and spatial data to predict species ranges. *Ecology Letters*, **12:** 334 - 350.

Keith, D.A., Akçakaya, H.R., Thuiller, W., Midgley, G.F., Pearson, R.G., Phillips, S.J., Regan, H.M., Araújo, M.B. and Rebelo, T.G. (2008) Predicting extinction risks under climate change: coupling stochastic population models with dynamic bioclimatic habitat models. *Biology Letters*, doi:10.1098/rsbl.2008.0049.

Kenis, M., Auger-Rozenberg, M., Roques, A., Timms, L., Péré, C., Cock, M., Settele, J., Augustin, S. and Lopez-Vaamonde, C. (2009) Ecological effects of invasive alien insects. *Biological Invasions*, **11:** 21 - 45.

Kenis, M., Auger-Rozenberg, M., Roques, A., Timms, L., Péré, C., Cock, M., Settele, J., Augustin, S. and Lopez-Vaamonde, C. (2009) Ecological effects of invasive alien insects. *Biological Invasions*, **11:** 21 - 45.

Kerr and Tinsdale 2004. Group performance and decision making. *Annual Review of Psychology*, **55**: 623-655.

Latimer, A.M., Wu, S., Gelfand, A.E. and Silander Jn, J.A. (2006) Building statistical models to analyze species distributions. *Ecological Applications*, **16:** 33 - 50.

Legendre, P. (1993) Spatial autocorrelation: trouble or new paradigm? *Ecology*, **74:** 1659 - 1673.

Lehrer, K. and Wagner, C. 1981. *Rational Consensus in Science and Society*, Dordrecht: D. Reidel Publishing.

Levine, J.M. and D’Antonio, C.M. (2003) Forecasting biological invasions with increasing international trade. *Conservation Biology*, **17:** 322 - 326.

Lichtendahl, K.C. and Winkler, R.L. (2007). Probability elicitation, scoring rules, and competition among forecasters. Management Science 53, 1745-1755.

Lichtenstein, S. & Fischhoff, B. (1977). Do those who know more also know more about how much they know? *Organizational Behaviour and Human Decision Processes, 20*, 159-183.

Lichtenstein, S., Fischhoff, B., & Phillips, L. (1982). Calibration of probabilities: The state of the art 1980. In D. Kahneman, P. Slovic & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases* (306-334). New York: Cambridge University Press.

Lobo, J.M., Jiménez-Valverde, A. and Real, R. (2008) AUC: a misleading measure of the performance of predictive distribution models. *Global Ecology and Biogeography,* **17:** 145 - 151.

Loiselle, B.A., Howell, C.A., Graham, C.H., Goerck, J.M. Brooks, T., Smith, K.G. and Williams, P.H. (2003) Avoiding pitfalls of using species distribution models in conservation planning. *Conservation Biology,* **17:** 1591 - 1600.

Los, S.O., Collatz, G.J., Sellers, P.J., Malmstršm, C.M., Pollack, N.H., DeFries, R.S., Bounoua, L., Parris, M.T., Tucker, C.J. and Dazlich, D.A. (2000) A global 9-year biophysical land-surface data set from NOAA AVHRR data. *Journal of Hydrometeorology*, **1:** 183 - 199.

MacGregor, D.G. 2001. Decomposition for judgmental forecasting and estimation. In, J. S. Armstrong (ed.) Principles of Forecasting. Pp 107-123. Kluwer, Boston.

Morris P.A 1977. Combining expert judgments: a Bayesian approach. *Management Science*, 23 679-693.

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

Pearson, R.G. (2007) Species’ distribution modeling for conservation educators and practitioners. Synthesis. American Museum of Natural History. Available at http://ncep.amnh.org.

Pearson, R.G., Raxworthy, C., Nakamura, M. and Peterson, A.T. (2007) Predicting species distributions from small numbers of occurrence records: a test case using cryptic geckos in Madagascar. *Journal of Biogeography*, **34:** 102 - 117.

Perrings, C., Williamson, M., Barbier, E.B., Delfino, D., Dalmazzone, S., Shogren, J., Simmons, P. and Watkinson, A. (2002) Biological invasion risks and the public good: an economic perspective. *Conservation Eco*logy, **6:** 1 - 7.

Peterson, A.T. (2003) Predicting the geography of species’ invasions via ecological niche modelling. *The Quarterly Review of Biology*, **78:** 419 - 433.

Peterson, A.T. and Vieglais, D.A. (2001) Predicting species invasions using ecological niche modelling: new approaches from bioinformatics attack a pressing problem. *Bioscience*, **51:** 363 - 371.

Peterson, A.T., Papeş, M. and Soberón, J. (2008) Rethinking receiver operating characteristic analysis applications in ecological niche modelling. *Ecological Modelling*, **213:** 63 - 72.

Pheloung, P.C. (1996) CLIMATE: a system to predict the distribution of an organism based on climate preferences. Agriculture, Western Australia, Perth.

Phillips, B.A., Chipperfield, J.D. and Keaney, M.R. (2008) The toad ahead: challenges of modelling the range and spread of an invasive species. *Wildlife Research*, **35:** 222 - 234.

Phillips, S. (2008) Response to “Transferability and model evaluation in ecological niche modeling”. *Ecography*, **31:** 272 - 278.

Phillips, S.J. and Dudík, M. (2008) Modeling of species distributions with Maxent: new extensions and a comprehensive evaluation. *Ecography,* **31:** 161 – 175.

Phillips, S.J., Anderson, R.P. and Schapire, R.E. (2006) Maximum entropy modeling of species geographic distributions. *Ecological Modelling*, **190:** 231 - 259.

Phillips, S.J., Dudik, M. and Schapire, R.E. (2004) A maximum entropy approach to species distribution modeling. In: Proceedings of the Twenty-First International Conference on Machine Learning. ACM Press, New York, pp. 472 - 486.

Phillips, S.J., Dudik, M. and Schapire, R.E. (2005) Maxent software for species distribution modeling. http://www.cs.princeton.edu/~schapire/maxent/

Phillips, S. J., M. Dudik, J. Elith, C. H. Graham, A. Lehmann, J. Leathwick, and S. Ferrier. 2009. Sample selection bias and presence-only distribution models: implications for background and pseudo-absence data. *Ecological Applications* **19**:181-197.

Plunkett, G.M., Moller, H., Hamilton, C., Clapperton, B.K. and Thomas, C.D. (1989) Overwintering colonies of German (*Vespula germanica*) and common wasps (*Vespula vulgaris*) (Hymenoptera: Vespidae) in New Zealand. *New Zealand Journal of Zoology*, **16:** 345 - 353.

Porter, W.P., Mitchell, J.W., Beckman, W.A. and DeWitt, C.B. (1973) Behavioural implications of mechanistic ecology – thermal and behavioral modeling of desert ectotherms and their microenvironment. *Oecologia*, **13:** 1 - 54.

Pyron, R.A., Burbrink, F.T. and Guiher, T.J. (2008) Claims of potential expansion throughout the U.S. by invasive python species are contradicted by ecological niche models. *PLoS ONE*, **3:** e2931.

R Development Core Team (2005) R: A language and environment for statistical computing, reference index version 2.5.1. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org.

Rose, H., J. Learmount, M. Taylor, and R. Wall. 2009. Mapping risk foci for endemic sheep scab. *Veterinary Parasitology* **165**:112-118.

Regan, H.M., Colyvan, M. and Markovchick-Nicholls, L. 2006. A Formal Model for Consensus and Negotiation in Environmental Management, *Journal of Environmental Management*, 80(2): 167–76.

Robinson, A., Burgman, M., Atkinson, W., Cannon, R., Miller, C., and Immonen, H. (2008). AQIS Import Clearance Data Framework. Technical Report 0804, Australian Centre of Excellence for Risk Analysis.

Rodríguez, J.P., Brotons, L., Bustamante, J. and Seoane, J. (2007) The application of predictive modelling of species distribution to biodiversity conservation. *Diversity and Distributions,* **13:** 243 - 251.

Rout, T.M., Salomon, Y. and McCarthy, M.A. (2009) Using sighting records to declare eradication of an invasive species. *Journal of Applied Ecology* 46: 110—117.

Savage, L.J. (1954). The Foundations of Statistics. New York, Wiley.

Scanlan, J.C. and Vanderwoude, C. (2006) Modelling the potential spread of *Solenopsis invicta* Buren (Hymenoptera: Formicidae) (red imported fire ant) in Australia. *Australian Journal of Entomology,* **45:** 1 - 9.

Schowalter, T.D. (2006) *Insect ecology: an ecosystem approach*. 2nd edition.

Segurado, P., Araújo, M.B. and Kunin, W.E. (2006) Consequences of spatial autocorrelation for niche-based models. *Journal of Applied Ecology,* **43:** 433 - 444.

Simberloff, D. (2003a) How much information on population biology is needed to manage introduced species? *Conservation Biology*, **17:** 83 - 92.

Simberloff, D. (2003b) Confronting introduced species: a form of xenophobia? *Biological Invasions,* **5:** 179 - 192.

Soll, J. B., & Klayman, J. (2004). Overconfidence in interval estimates. *Journal of Experimental Psychology Learning Memory and Cognition, 30* (2), 299-314.

Speirs-Bridge, A., Fidler, F., McBride, M., Flander, L., Cumming, G., Burgman,M. (In press). Reducing overconfidence in the interval judgments of experts, *Risk Analysis*.

Spetzler, C.S. and Staelvonholstein, C.A.S. (1975) Probability encoding in decision analysis. *Management Science*, 22:340-358.

Spradbery, J.P. (1973) Wasps: an account of the biology and natural history of solitary and social wasps. London: Sidgwick & Jackson.

Spradbery. J.P. and Maywald, G.F. (1992) The distribution of the European or German wasp, *Vespula germanica* (F.) (Hymenoptera: Vespidae), in Australia: Past, Present and Future. *Australian Journal of Zoology*, **40:** 495 - 510.

Stephenson, C.M., MacKenzie, M.L., Edwards, C. and Travis, J.M.J. (2006) Modelling establishment probabilities of an exotic plant, *Rhododendron ponticum*, invading a heterogeneous, woodland landscape using logistic regression with spatial autocorrelation. *Ecological Modelling*, **193:** 747 - 758.

Suárez-Seoane, S., García de la Morena, E.L., Morales Prieto, M.B., Osborne, P.E. and de Juana, E. (2008) Maximum entropy niche-based modelling of seasonal changes in little bustard (*Tetrax tetrax*) distribution. *Ecological Modelling*, **219:** 17 - 29.

Sutherst, R.W. and Maywald, G.F. (2005) A climate model of the Red Imported Fire Ant, *Solenopsis invicta* Buren (Hymenoptera: Formicidae): Implications for invasion of new regions, particularly Oceania. *Environmental Entomology*, **34:** 317 - 335.

Sutherst, R.W., Maywald, G.F. and Bourne, A.S. (2007) Including species interactions in risk assessments for global change. *Global Change Biology*, **13:** 1 - 7.

Taber, R. (1991). Knowledge processing with fuzzy cognitive maps. Expert Systems with Applications. 2: 83 – 87.

Teigen, K. H., & Jorgensen, M. (2005). When 90% confidence intervals are only 50% certain: On credibility of credible intervals. *Applied Cognitive Psychology, 19*, 455-475.

Thornthwaite, C.W. (1948) An approach toward a rational classification of climate. *The Geographical Review*, **38:** 55 - 94.

Thuiller, W., Richardson, D.M., Pyšek, P., Midgley, G.F., Hughes, G.O., and Rouget, M. (2005) Niche-based modelling as a tool for predicting the risk of alien plant invasions at a global scale. *Global Change Biology,* **11:** 2234 - 2250.

Tribe, G.D. and Richardson, D.M. (1994) The European wasp, *Vespula germanica* (Fabricus) (Hymenoptera: Vespidae), in southern Africa and its potential distribution as predicted by ecoclimatic matching. *African entomology,* **2:** 1 - 6.

Tufte, E. R. (1983). *The visual display of quantitative information*. Cheshire, CN: GraphicsPress.

Turoff, M., Hiltz, S.R., Li, Z., Wang, Y. And Cho, H.-K. 2006. Online collaborative learning enhancement through the Delphi method. Turkish Online Journal of Distance Education-TOJDE April 2006, Volume: 7 Number: 2 Article: 6.

UNEP (2006) The GEO Data Portal. Global mean annual potential evapotranspiration. United Nations Environment Programme. Available at http://geodata.grid.unep.ch (Downloaded on 10/10/2008).

VanDerWal, J., Shoo, L.P., Graham, C. and Williams, S.E. (2009) Selecting pseudo-absence data for presence-only distribution modeling: How far should you stray from what you know? *Ecological Modelling*, **220:** 589 - 594.

Ward, D., Honan, P. and Lefoe, G. (2002) Colony structure and nest characteristics of European wasps *Vespula germanica* (F.) (Hymenoptera: Vespidae) in Victoria, Australia. *Australian Journal of Entomology*, **31:** 306 - 309.

Ward, D.F. (2007) Modelling the potential geographic distribution of invasive ant species in New Zealand. *Biological Invasions,* **9:** 723 - 735.

Warren. I. and Statham, S. (2002) Control of European wasps (*Vespula germanica*) by baiting. Hobart: Tasmanian Institute of Agricultural Research. Available at http://www.dpiw.tas.gov.au

Wharton, T.N. and Kriticos, D.J. (2004) The fundamental and realized niche of the Monterey Pine aphid, *Essigella Californica* (Essig) (Hemiptera: Aphididae): implications for managing softwood plantations in Australia. *Diversity and Distributions,* **10:** 253 - 262.

Williamson, M. (1999) Invasions. *Ecography*, **22:** 5 - 12.

Wilson, P.R., Karl, B.J., Toft, R.J., Beggs, J.R. and Taylor, R.H. (1998) The role of introduced predators and competitors in the decline of kaka (*Nestor meridionalis*) populations in New Zealand. *Biological Conservation,* **83:** 175 - 185.

Winkler, R. L., & Clemen, R. T. (1992). Sensitivity of weights in combining forecasts. Operations Research, 40(3), 609−614.

Wintle, B.A., Elith, J. and Potts, J.M. (2005) Fauna habitat modelling and mapping: A review and case study in the Lower Hunter Central Coast region of NSW. *Austral Ecology,* **30:** 719 - 738.

Wintle, B.A., McCarthy, M.A., Volinsky, T.A. and Kavanagh, R.P. (2003) The use of Bayesian model averaging to better represent uncertainty in ecological models. *Conservation Biology*, **17:** 1579 - 1590.

Worner, S.P. and Gevrey, M. (2006) Modelling global insect pest species assemblages to determine risk of invasion. *Journal of Applied Ecology,* **43:** 858 - 867.

Yaniv, I. and D. P. Foster (1995). Graininess of Judgment under Uncertainty - an Accuracy Informativeness Trade-Off. *Journal of Experimental Psychology-General* **124**(4): 424-432.

Yaniv, I. and D. P. Foster (1997). Precision and accuracy of judgmental estimation. *Journal of Behavioral Decision Making* **10**(1): 21-32.

Zach, L. S. and V. M. Bier (2009). Risk-based approaches to import safety. Paper presented at the Society for Risk Analysis Conference, December 9, 2009. Baltimore, USA.

Zalucki, M.P. and Furlong, M.J. (2005) Forecasting Helicoverpa populations in Australia: a comparison of regression based models and a bioclimatic based modelling approach. *Insect Science,* **12:** 45 - 56.